Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 1829104, 14 pages
https://doi.org/10.1155/2022/1829104

Research Article

Differential Congestion Pricing Strategies for Heterogeneous Users in the Mixed Traffic Condition

Yifan Chen,[1] **Yuliang Zhang**,[2] **and Ziyuan Gu**[3]

[1]*Institute of Transport Studies, Department of Civil Engineering, Monash UniversityClayton, VIC 3800, Australia*
[2]*Hangzhou Institute of Advanced Technology, Hangzhou 310018, China*
[3]*Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, Nanjing 210096, China*

Correspondence should be addressed to Yifan Chen; yifan.chen@monash.edu

Received 15 October 2021; Accepted 23 December 2021; Published 19 January 2022

Academic Editor: Zhihong Yao

Copyright © 2022 Yifan Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Congestion pricing is one effective demand management strategy to alleviate traffic congestion. This work investigates pricing schemes for mixed traffic flow systems where the human-driven vehicles (HVs) and autonomous vehicles (AVs) coexist. The emerging and integration of autonomous vehicles can help improve the overall transportation efficiency and safety. Given the coexistence of HVs and AVs in the near future, there is need to adjust the existing traffic management strategies to adapt to the mixed traffic conditions. In this study, congestion pricing is imposed on the HVs and the AVs differently, that is, a distance-based toll to the HVs while a delay-based toll to the AVs. We consider six user groups based on the value of time (VOT) and the vehicle types. Compared with the unified distance-based toll, the advantage of delay-based toll is demonstrated first. Then, a surrogate-based optimization framework, namely the regressing Kriging (RK) model, is formulated. Three pricing schemes are investigated and compared: equity-oriented (EQ), environment friendliness-oriented (EN), and revenue-oriented (RE) schemes. Results show that the RE scheme collects the highest revenues; however, its cost-efficiency is weakened. The EQ scheme reduces the variance in the average travel costs among user groups, thus solving the equity issue.

1. Introduction

With the rapidly increasing population, traffic congestion has become a major problem, especially in urban areas. Due to the limited land resources and high costs, it is unsustainable to address this issue by merely building new infrastructures. Given this situation, congestion pricing, one of the promising demand-oriented strategies, is widely advocated to mitigate congestion. As an economic lever, it originates from the Pigou's theory where extra traveler entering the network should be charged an additional fee for his negative impact imposed on other road users [1]. This type of pricing model is known as the first-best pricing. However, the first-best pricing is difficult to be implemented because it is unrealistic to charge all links in a network strictly by the marginal cost surplus of traffic [2]. Instead, the second-best pricing model was proposed wherein just a subset of links are charged based on the traffic conditions [3]. A branch of congestion pricing practices emerged since 1975, when the area licensing scheme (ALS) was successfully operated in Singapore. However, quite a few schemes were aborted at the stage of trial or referendum. Low public acceptance was the main reason which led to the failure of these cases [4].

Equity is a vital factor influencing public acceptance towards the congestion pricing policy [5]. As a result, vast literature discussed the equity issue on the design of congestion pricing schemes from different perspectives. Lucas et al. pointed out that drivers' places of residence will affect their willingness to pay. Under a cordon-based pricing scheme, people living outside the cordon and commuting towards it may become losers while the winners are those living and conducting activities inside the cordon [6]. Gaunt et al. investigated residents' attitudes towards congestion

pricing in Edinburgh (Scotland). He found that people with high car dependency were less supportive of congestion tolls [7]. Chen at al. conceived a novel pricing scheme considering drivers' perceived level of service (LOS). People who enjoyed higher LOS on their trips would pay more [8].

Value of time (VOT) is an important direction when we deal with the equity issue. According to Arnott's research, the heterogeneity in VOT affected the calculation of welfare effects. The congestion toll favored people with high VOT relative to the schedule delay [9]. Later, more studies were dedicated to examine the impact of congestion pricing on users with different VOTs [10–12]. However, few researches designed the equitable toll rates directly based on the difference in VOTs. Zheng et al. carried out a seminal study where the group-based tolls were determined according to the distribution of VOTs. They argued that when the total time savings after the implementation of congestion pricing were monetized, drivers with higher VOT would gain more. Hence, it is unfair to ask all road users to pay the same amount of toll [13]. However, the VOT categorization in their research is only related to income levels. Obviously, the vehicle type also influences VOT. The advent of autonomous vehicles (AV) has the potential to impact travel behaviour as well as the transportation system. The AV technologies enable drivers to do other activities on their trips, thus making travel time more productive [14–16]. Zhong et al. applied a mixed logit model to quantify the changes in VOT if taking AVs. The results showed that the VOT would be reduced by 30% at most [17]. On the contrary, the majority of actualized congestion pricing schemes are cordon-based or distance-based [5]. The drawback of the cordon-based scheme is that drivers are charged equally regardless of the actual distance travelled inside the cordon. As for the distance-based case, it may induce people to choose the shortest paths, thus making the distribution of congestion more uneven. Gu et al. once put forward the joint distance and delay toll (JDDT) scheme to remedy the limitation of the distance-based congestion pricing policy [2]. But the novel scheme is difficult to be put into practice in the era of human-driven cars. With the embedded vehicle-to-vehicle communication technology, AVs can collect rich information for the real-time traffic condition. As a result, it provides opportunities to apply advanced congestion pricing approaches.

The traditional way to solve the congestion pricing problem is formulating a bi-level mathematical program with equilibrium constraints (MPECs) [18–20]. Nonetheless, it remains challenging to deal with MPEC in a dynamic large-scale transportation network. As the dynamic congestion pricing (DCP) problem often involves high-dimension decision variables and a nonconvex, nonlinear, and nonclosed form objective function [21–23]. In recent years, with the help of mature dynamic traffic assignment (DTA) simulators, simulation-based optimization (SBO) is recognized as an alternative way to handle this problem. It does not require an explicit mathematical formulation of the objective function which is usually unavailable under stochastic traffic dynamics. In general, SBO methods can be classified into four types: (1) direct search; (2) gradient-based approach; (3) feedback control; (4) surrogate-based method. The direct search partitions the search space into hyper rectangles. By comparing the objective function values, the potential optimal rectangle is determined for further partition [24]. Given its exhaustive partitioning characteristics, it is time-consuming to improve the current best solution. The gradient-based method is derivative-based, which tries to approximate the gradient by finite difference. However, it is not suitable for objective functions where the gradient does not exist everywhere [25]. The feedback control is an efficient optimization method. It adjusts the input variables iteratively to let the system output approach the set point [26]. But this method is not applicable when it comes to high-dimension problem with complex constraints [27]. In this context, the surrogate-based methods gain popularity for their ability in approximating expensive-to-evaluate functions. With limited number of evaluations, they can approximate simulation input-output mapping [28]. The regressing Kriging (RK) model outperforms other surrogate models for its excellent performance in prediction [29]. By constructing a probability model, it generates not only an interpolated spatial correlation but also an estimate of the uncertainty. Meanwhile, the feature of regressing sample points overcomes the influence of simulation noise [30].

In order to monitor the effects of the congestion pricing schemes, it is essential to understand the traffic dynamics at the network level. Geroliminis and Levinson proposed the macroscopic fundamental diagram (MFD), which related the network average density to the average flow [31]. The congestion can be identified when the flow decreases with the density, and it indicates the oversaturated state of the network. The critical density is the one at which the maximal production (flow) is achieved [32]. After the implementation of the pricing scheme, we hope the network density would be around the critical value.

Although plenty of studies attempted to design appropriate congestion pricing schemes, there are still some research gaps: (1) few works considered the mixed traffic of human-driven vehicle (HV) and AVs. The inserted communication technologies of AVs enable us to apply the advanced pricing strategy, which can be different from that of HVs; (2) reduction in VOT for AV users was not well considered; and (3) existing studies rarely compared the pricing schemes with different objectives. In this paper, we propose a surrogate-based optimization to solve the dynamic pricing problem for mixed traffic with drivers of heterogeneous VOTs. The HVs are charged by distance while AV users pay according to the average link delay. The VOT-based tolls are adopted among different groups. The results of three pricing schemes, namely equity-oriented (EQ), environment friendliness-oriented (EN), and revenue-oriented (RE) schemes are compared. To the best of the authors' knowledge, the congestion pricing optimization under mix traffic with heterogeneous users has never been proposed before and hence offers significant contributions to the literature.

2. Methodology

2.1. Simulation-Based Dynamic Assignment. The stochastic route choice (SRC) model is a kind of nonequilibrium-seeking DTA model. It has been extensively used for transportation operations owing to the lower computational cost compared to the dynamic user equilibrium (DUE) model [33]. Table 1 summarizes the notations used in SRC model. Let $G = (N, L)$ denote the object network. The time-varying distance-based toll rates for the HV user group i can be represented as a decision vector $\mathbf{v}_i = [v_{1i}, v_{2i}, \ldots, v_{ni}, \ldots, v_{Ni}]$. And the delay-based toll rates for AV user group j is expressed as $\mathbf{w}_j = [w_{1j}, w_{2j}, \ldots, w_{nj}, \ldots, w_{Nj}]$.

The generalized travel cost for HV user group i to choose link a in the nth interval is as follows:

$$C_{ai}(n) = \bar{t}_a(n-1) + \frac{l_a v_{ni}}{\mathrm{VOT}_i}. \tag{1}$$

The generalized travel cost for AV user group j to choose link a in the nth interval is as follows:

$$C_{aj}(n) = \bar{t}_a(n-1) + \frac{\bar{d}_a(n-1) w_{nj}}{\mathrm{VOT}_j}. \tag{2}$$

Note that the communication is only among AVs. Hence $\bar{d}_a(n-1)$ is the average value acquired from AVs which crossed link a during the $(n-1)$th time interval. Given the mixed traffic, it can approximately represent the average delay for all groups. The C-logit model is used to simulate the route choice behavior. The probability of choosing path r among the path set R^{OD} for a given OD pair in the nth interval is as follows:

$$P_r(n) = \frac{\exp\left(-C_r(n) - CF_r\right)}{\sum_{m \in R^{OD}} \exp\left(-C_m(n) - CF_m\right)}, \tag{3}$$

where the term CF_r (CF_m) is the commonality factor of path r (m) which describes the degree of overlapping with other alternative paths [34]. $C_r(n)$ is the sum of the travel costs of all links in path r.

2.2. Proposed Framework. In this paper, travelers are categorized according to the income level and the vehicle type (HV or AV). Given the same vehicle type, high-income travelers have higher VOT due to high productivity [35]. Given the same income level, AV users will have a smaller VOT because they can spend in-vehicle time engaged in other activities. The differentiated VOTs influence the route choice behavior. When paying the same amount of toll, travelers of high VOT will perceive fewer costs (according to equation (1) and (2)), thus resulting in inequity. As a result, let $v_{n,\mathrm{low}}$ denote the distance-based toll rate for the HV user group of the lowest VOT (HV reference group) in the nth time interval. Then, the toll rates for other HV user groups are set based on the VOT difference from the HV reference group. Similarly, denote the delay-based toll rate for the AV user group of the lowest VOT (AV reference group) in the nth time interval as $w_{n,\mathrm{low}}$. Then, the toll rates for other HV user groups are set based on the VOT difference from the AV reference group. We focus on the optimization of congestion pricing for the congested central area of the city. Three different congestion pricing schemes are investigated: (1) EQ scheme; (2) EN scheme; and (3) RE scheme.

As we adopt the distance-based and delay-based tolls for HV and AV user groups separately, the average travel costs (AUD/trip) for different groups would be disperse after the congestion pricing. Hence, the objective of the EQ scheme is to minimize the variance in average travel costs:

$$\min Z_1 = \sigma^2\left(\mathrm{ATC}_g\right), \tag{4}$$

where Z_1 is the objective function for the EQ scheme and ATC_g is the average travel cost for the user group g after the implementation of congestion pricing. Note that ATC_g includes the trip travel time (in monetary unit) and toll paid.

As for the EN scheme, we aim to minimize the total emission in the pricing zone. The QUARTET pollution emission model is utilized [36]. In the mesoscopic simulation, we can only acquire the average speed for a link during each time interval. The impact of the acceleration and deceleration processes has been considered via the average speed:

$$\min Z_2 = \sum_{g=1}^{G} EM_g, \tag{5}$$

where Z_2 is the objective function of the EN scheme; EM_g is the total emission of the user group g during the toll period; and G is the number of user groups.

The RE scheme is devoted to maximizing the total toll paid. It is equivalent to minimize the negative version of this term. The collected revenues can be redistributed to road users or used to maintain the transport infrastructures:

$$\min Z_3 = -\sum_{g=1}^{G} \mathrm{REV}_g, \tag{6}$$

where Z_3 is the objective function of the RE scheme; REV_g is the revenues collected from the user group g; and G is the number of user groups.

There are several constraints considered for all three schemes:

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left|\overline{K}_n - K_{cr}\right| < e,\\
&\left|v_{n,\mathrm{low}} - v_{n+1,\mathrm{low}}\right| \le \alpha, \quad n = 1, 2, \ldots, N-1,\\
&\left|w_{n,\mathrm{low}} - w_{n+1,\mathrm{low}}\right| \le \beta, \quad n = 1, 2, \ldots, N-1,\\
&v_{\min} < v_{n,\mathrm{low}} \le v_{\max}, \quad n = 1, 2, \ldots, N,\\
&w_{\min} < w_{n,\mathrm{low}} \le w_{\max}, \quad n = 1, 2, \ldots, N.
\end{aligned} \tag{7}$$

Constraint 7 controls the network performance. $\overline{K}_n$ is the network average density in the nth time interval; K_{cr} is the critical density of the network; and e is the preset threshold.

Table 1: Notations used in the SRC model.

Notation	Description
N	Node set for the object network
L	Directed link set for the object network
v_{ni}	Distance-based toll rate for HV user group i in the nth time interval
w_{nj}	Delay-based toll rate for AV user group j in the nth time interval
N	Number of time intervals
VOT_i	Value of time for HV user group i
VOT_j	Value of time for AV user group j
$\bar{d}_a(n-1)$	Average delay in link a in the $(n-1)$th interval
$\bar{t}_a(n-1)$	Average travel time for link a in the $(n-1)$th interval
CF_r	Commonality factor of path r

This constraint makes sure that the network density is close to the critical value over time. Constraint 8 and 9 ensure that the toll rates between adjacent time intervals do not change sharply. $v_{n,\mathrm{low}}$ $(v_{n+1,\mathrm{low}})$ is the distance-based toll rate for the HV reference group in the nth $((n+1)\mathrm{th})$ time interval; $w_{n,\mathrm{low}}$ $(w_{n+1,\mathrm{low}})$ is the delay-based toll rate for the AV reference group in the nth $((n+1)\mathrm{th})$ time interval; and α and β are the smoothing parameters. Constraint 10 and 11 set the upper and lower bound for the toll rates. $v_{\min}$ $(v_{\max})$ is the lower and upper bound for the distance-based toll rate of the HV reference group. $w_{\min}$ $(w_{\max})$ is the lower and upper bound for the delay-based toll rate of the AV reference group. Note that the toll rates of other groups are set proportionally. Once the toll rate ranges for reference groups are confirmed, the ranges for other groups are also determined.

2.3. Surrogate-Based Optimization. The RK model is employed to do the constrained optimization. Figure 1 shows the procedure of the surrogate optimization. Some initial points (or toll plans) are generated by a certain sampling method. In our work, Latin hypercube sampling (LHS) is adopted. It is a space-filling approach which stratifies each dimension of the decision variables into an equal number of partitions [37]. At least $(m+1)$ initial points are needed for m-dimension problems [38]. Considering the high dimension of our optimization problems, totally $2(m+1)$ points are generated. The we run the network simulation to get the objective function values for these initial inputs. The simulation input-output mapping is used to build a preliminary surrogate model. In order to enhance the model, additional infill points are required. The additional samples can provide more information in the potential regions where the good solution may exists. The infill process will not terminate until the stop criteria are met. Finally, the accuracy of the model is checked. When the accuracy is validated, we can declare that the optimum is achieved. Otherwise, the surrogate model should be redesigned.

2.3.1. Model Construction. For the RK model, the prediction for a nonsampled point $\mathbf{x}_{n+1}$ is written as follows:

$$\hat{y}(\mathbf{x}_{n+1}) = \hat{\mu} + \varepsilon(\mathbf{x}_{n+1}), \tag{8}$$

where $\hat{\mu}$ is the constant mean; $\varepsilon(\mathbf{x}_{n+1})$ is a normally distributed and independent estimation error, and it can be expanded as follows:

$$\varepsilon(\mathbf{x}_{n+1}) = \boldsymbol{\psi}^{\mathrm{T}}(\mathbf{R} + \lambda\mathbf{I})^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}), \tag{9}$$

where $\mathbf{y}$ is the output vector for n sampled points $(\mathbf{y} = [y_1, y_2, \ldots, y_n]^{\mathrm{T}})$; $\mathbf{I}$ is the identity matrix; $\mathbf{R}$ is the correlation matrix between the sampled points; and λ $(\lambda \in (0, 10])$ is the regression constant. With the positive λ, the RK model regresses the data to reduce the influence from the simulation noise. Ψ is the correlation vector between n sampled points and the new point. The correlation between point $\mathbf{x}_i$ and $\mathbf{x}_j$ is given as follows:

$$\mathrm{Corr}[\mathbf{x}_i, \mathbf{x}_j] = \exp\left[-\sum_{h=1}^{k} \hat{\theta}_h \left|x_{i,h} - x_{j,h}\right|^{\hat{p}_h}\right], \tag{10}$$

where k is the dimension of variables; $\hat{\theta}_h$ and $\hat{p}_h$ is the scaling and smoothness coefficient for the hth dimension, separately; and $\hat{\theta}_h$ $(\hat{\theta}_h \in [0.001, 100])$ denotes the sensitivity to the objective function value. A larger $\hat{\theta}_h$ means the certain dimension is more sensitive than others. The use of $\hat{p} \in [0, 1]$ is usually advocated. However, $\hat{p} = 2$ proves to be suitable for the engineering-based problem [39].

Then, the vector of scaling coefficients θ, the variance σ^2 and the mean μ are estimated by maximizing the likelihood probability [40]:

$$\frac{1}{2\pi^{1/2}\left(\sigma^2\right)^{1/2}|\mathbf{R} + \lambda\mathbf{I}|^{1/2}} \exp\left(\frac{-(\mathbf{y} - \mathbf{1}\mu)^T\mathbf{R} + \lambda\mathbf{I}^{-1}(\mathbf{y} - \mathbf{1}\mu)}{2\sigma^2}\right). \tag{11}$$

The limited memory Broyden–Fletcher–Goldfarb–Shanna (LBFGS) algorithm is employed to solve the maximization problem. This iterative method is appropriate for problems with large numbers of variables [41].

2.3.2. Infill Strategy. In order to improve the accuracy of the surrogate model by the augmented data set, the acquisition function is used to guide the search direction. The expected improvement (EI) function is a widely-used acquisition function, which estimates the magnitude of improvement at a new point [42]. It can explore the unvisited region and exploit the domain of interest at the same time. Let $y_{\min}$

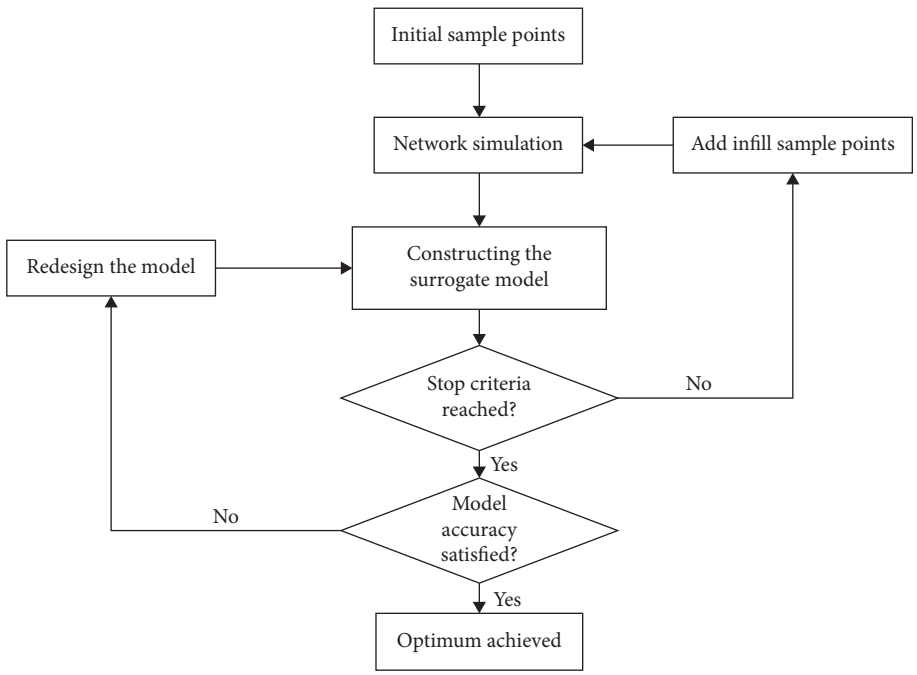


Figure 1: Flowchart of the surrogate-based optimization.

denote the current best solution. The improvement of a new point $\mathbf{x}$ can be expressed as $I(\mathbf{x}) = \max(0, y_{\min} - \hat{y}(\mathbf{x}))$. Then, the formulation of EI is demonstrated as [43]:

$$E[I(\mathbf{x})] = \begin{cases} [y_{\min} - \hat{y}(\mathbf{x})]\Phi\left(\dfrac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{s}_{ri}(\mathbf{x})}\right) + \hat{s}_{ri}(\mathbf{x})\phi\left(\dfrac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{s}_{ri}(\mathbf{x})}\right), & \text{if } \hat{s}_{ri}(\mathbf{x}) > 0, \\ 0, & \text{if } \hat{s}_{ri}(\mathbf{x}) = 0, \end{cases} \tag{12}$$

where $\Phi(.)$ is the normal cumulative distribution function; $\phi(.)$ is the probability density function; and $\hat{s}_{ri}(\mathbf{x})$ is the re-interpolation prediction error for the objective:

$$\hat{s}_{ri}{}^2(\mathbf{x}) = \hat{\sigma}_{ri}{}^2\left[1 - \boldsymbol{\psi}^T(\mathbf{R} + \lambda\mathbf{I})^{-1}\boldsymbol{\psi}\right], \tag{13}$$

where $\hat{\sigma}^2{}_{ri}$ is the estimate of σ^2 for the re-interpolation:

$$\hat{\sigma}_{ri}^2 = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu})^T(\mathbf{R} + \lambda\mathbf{I})^{-1}\mathbf{R}(\mathbf{R} + \lambda\mathbf{I})^{-1}(y - \mathbf{1}\hat{\mu})}{n}. \tag{14}$$

Note that by replacing $\hat{\sigma}^2$ with $\hat{\sigma}_{ri}^2$, $E[I(\mathbf{x})]$ stays at zero for all the existing points. Hence, the EI sampling will not be trapped at existing sample points. The new points are obtained with the aim of maximizing the $E[I(\mathbf{x})]$.

The EI function is sufficient for unconstrained optimization. However, when it comes to the constrained optimization. A new point that has a good objective value may not meet the expensive-to-evaluation constraint at the same time. As such, another acquisition, namely the probability of improvement (PI) function is utilized. The mechanism of PI is to research for a new point which has the highest probability to meet the complex constraint. We need to construct a new surrogate model for the constraint separately. Denote $\hat{c}(\mathbf{x})$ as the estimated constraint value. Then, we can formulate the PI function as follows:

$$\mathrm{P}\left[\hat{c}(\mathbf{x}) > c_{\min}\right] = \begin{cases} \Phi\left(\dfrac{c_{\max} - \hat{c}(\mathbf{x})}{\hat{s}_{c,ri}(\mathbf{x})}\right), & \hat{s}_{c,ri}(\mathbf{x}) > 0, \\ 0, & \hat{s}_{c,ri}(\mathbf{x}) = 0, \end{cases} \tag{15}$$

where $c_{\max}$ is the preset threshold as in Constraint 7 and $\hat{s}_{c,ri}(\mathbf{x})$ is the re-interpolation prediction error for the constraint. The constrained EI function $E[CI(\mathbf{x})]$ can be written as follows:

$$E[CI(\mathbf{x})] = \begin{cases} E[I(\mathbf{x})] \cdot P\left[\hat{c}(\mathbf{x}) \le c_{\min}\right], & \hat{c}(\mathbf{x}) \le c_{\min}, \\ 0, & \hat{c}(\mathbf{x}) > c_{\min}. \end{cases} \tag{16}$$

2.3.3. Model Validation. The leave-one-out cross-validation is a method frequently used to measure the accuracy of surrogate models [44]. This approach leaves out one observation and reconstructs the surrogate model with the

remaining points. The estimated output of the removed point from the re-fitted model will be compared with the actual output. The key metric for the leave-one-out cross-validation is the standardized cross-validated residual (SCVR). It is calculated as follows:

$$\frac{y(\mathbf{x}_i) - \hat{y}_{-i}(\mathbf{x}_i)}{\hat{s}_{-i}(\mathbf{x}_i)}, \tag{17}$$

where $y(\mathbf{x}_i)$ is the actual output for $\mathbf{x}_i$; $\hat{y}_{-i}(\mathbf{x}_i)$ is the estimated output of $\mathbf{x}_i$ without $\mathbf{x}_i$; and $\hat{s}_{-i}(\mathbf{x}_i)$ is the prediction error of $\mathbf{x}_i$ from the re-fitted model without $\mathbf{x}_i$. The model is validated when the SCVRs for all points roughly fall in the interval [−3,3]. It means the model proves to be 99.7% confident that $y(\mathbf{x}_i)$ lies in the interval $[\hat{y}_{-i}(\mathbf{x}_i) - 3\hat{s}_{-i}(\mathbf{x}_i), \hat{y}_{-i}(\mathbf{x}_i) - 3\hat{s}_{-i}(\mathbf{x}_i)]$ [41].

3. Results and Discussion

3.1. Experiment Setup. In this paper, we perform a mesoscopic simulation-based DTA in the Melbourne city, Australia in the traffic simulation software AIMSUN. The travel demand is collected via the loop detector and changes every 15 min between 6 : 00 AM and 10 : 00 AM. Figure 2(a) shows the congested central business district (CBD) (inside the red box). We choose it as the pricing zone. There are 75 nodes and 218 links in the object network. The information of daily income of residents is extracted from the Victorian Integrated Survey of Travel and Activity (VISTA). We assume that people work 8 hours a day. As a result, the VOTs equate to the daily income divided by eight. Travelers are classified into three groups based on the distribution of VOT initially. The VOT of each group is the median of the corresponding population. Since the mass market penetration of AVs is challenging in the near future, 30% AV penetration rate is used in our work. Meanwhile, we assume that more high-income travelers can afford AVs. Finally, there are totally six user groups shown in Table 2. Note that AV users will have a reduction in their VOTs by 30% [14].

The agents of different user groups are uniformly distributed in the network. In order to avoid simulation randomness, we run three replications with different random seeds. The MFDs for the three replications (R1, R2, and R3) and the average result (AVE) are presented in Figure 2(b). According to the MFDs, the flow starts to drop at 25 veh/km/lane. Hence, we set the critical density as 25 veh/km/lane. The network density over time in the no-toll scenario is demonstrated in Figure 2(c). Observe that the density exceeds the critical value around 8:30. So the toll period is from 8:30 to 10:00.

In our study, $v_{\min}$ is set as 0 AUD/km, and $v_{\max}$ is set as 1 AUD/km. $w_{\min}$ is set as 0 AUD/h, and $w_{\max}$ is set as 15 AUD/h. The smoothing parameter α is set as $1/3 \times (1 - 0) = 0.3$, while β is set as $1/3 \times (15 - 0) = 5$. One can refer to the study by Gu et al. [34], wherein the analysis on the smoothing parameters is conducted. The threshold e is set as 2.45 veh/km/lane which is nearly one-fourth of the value (9.8 veh/km/lane) in the no-toll scenario. The toll period is divided into six 15-minute intervals. Hence, the decision variable vector has 12 dimensions, six of which are distance-based toll rates for the HV reference group, while the rest are delay-based toll rates for the AV reference group.

3.2. Comparison between Unified Delay-Based Toll and Distance-Based Toll. In order to unveil the advantage of the delay-based toll, we compare the network performance under the unified delay-based toll and distance-based toll. First of all, let us introduce a term, namely the spatial spread of density [45]. It measures the heterogeneity of congestion distribution and is calculated as follows:

$$\gamma = \sqrt{\frac{\sum_i l_i n_i (k_i - K)^2}{\sum_i l_i n_i}}, \tag{18}$$

where l_i is the length of link i; n_i is the number of lanes of link i; k_i is the density of link i; and K is the network density.

The relationship between the spatial spread of density and the network density can be expressed as the third-polynomial function [34]:

$$\gamma(K) = aK^3 + bK^2 + cK. \tag{19}$$

We collect 1000 data points to plot Figure 3(a). The lower envelop of Figure 3(a) is fitted to Equation (19). Parameters are estimated as: $a = -0.0004$, $b = 0.0278$, and $c = 1.2345$. According to Figure 3(a), the spatial spread of density will naturally increase with the network density. Simoni et al. further proposed an advanced metric: the deviation from spread [46]:

$$\Delta = \gamma - \gamma(K). \tag{20}$$

Smaller Δ indicates the lower heterogeneity of congestion distribution, thus leading to the higher network flow. We select a delay-based toll scheme with the unified delay toll rate of 13 AUD/h and a distance-based toll scheme with the unified distance toll rate of 0.5 AUD/km. According to Figure 3(b), the network density under the distance-based toll scheme (DI_density) is smaller than that of the delay-based toll (DE_density) in most of the time. However, the result of the network flow is the opposite. When we look into the average deviation from spread during the toll period, we find that it is much smaller under the delay-based toll scheme (1.93 veh/km/lane vs. 4.93 veh/km/lane). This conclusion is in line with Gu's research [27]. It indicates that the delay-based toll can reduce the heterogeneity of congestion distribution to a great extent. This is also one motivation of our research to implement the delay-based toll on AVs which are equipped with the perfect technology to realize it. The inserted advanced communication technology of autonomous vehicles enables us to collect data such as the vehicle position and the average delay of queues.

3.3. Results of Optimal Toll Designs with Different Objectives. The stop criterion in this paper is that 30 infill points are collected or $EI = 0$ for consecutive 10 points. Figures 4(a)–4(f) show the results for the EQ scheme. When we track the convergence history of EI, the value decreases gradually

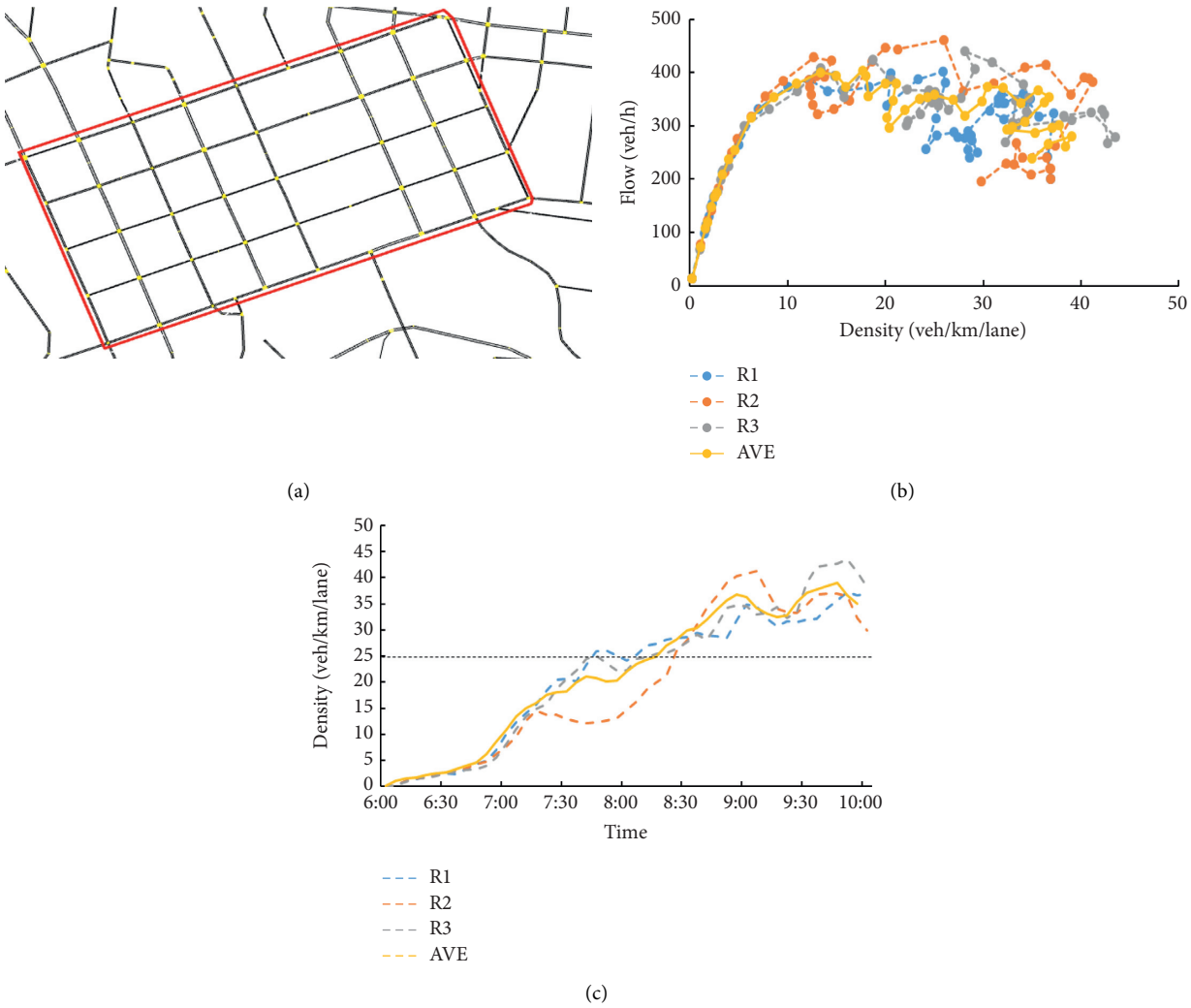
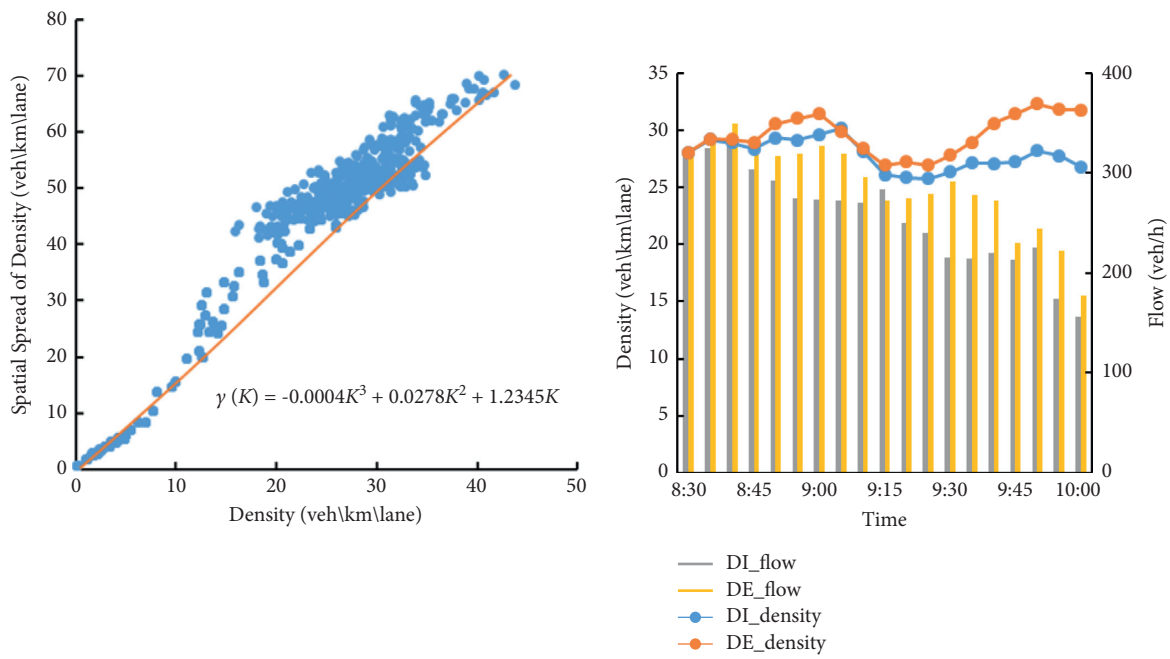


Figure 2: (a) Pricing zone; (b) MFD for the pricing zone in no-toll scenario; (c) density over time in no-toll scenario.

Table 2: The VOTs and percentages of different user groups.

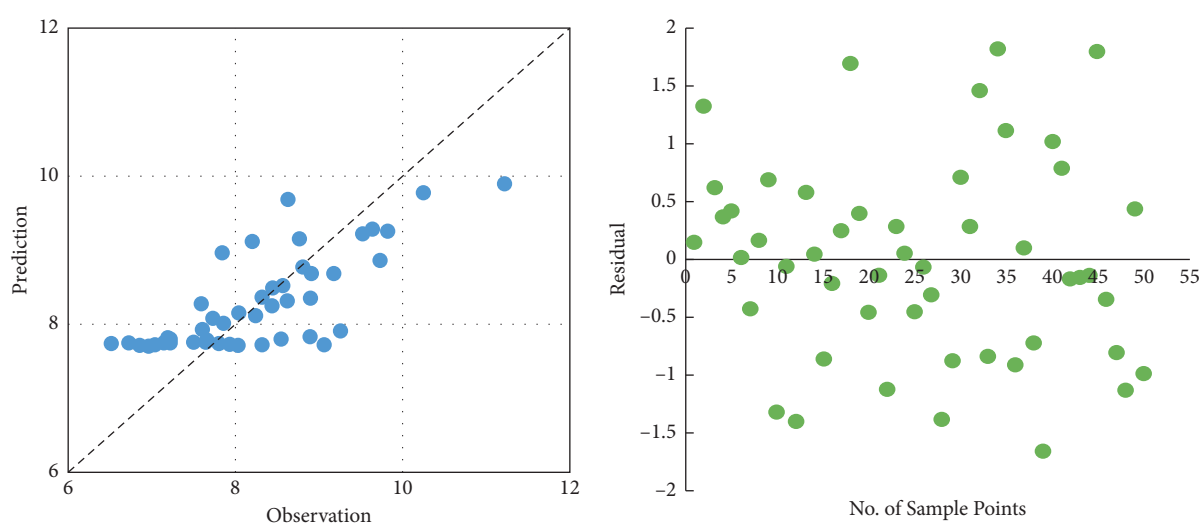
Group name	VOT (AUD/h)	Percentage	AV penetration (%)
*AV*_low	8	15	20
*AV*_medium	19	10	50
*AV*_high	28	5	85
*CV*_low	12	59	0
*CV*_medium	28	10	0
*CV*_high	42	1	0

(shown in Figure 4(e)). The infill process terminates after the 24th infill point because EI equates to zero for 10 function evaluations successively. It means, it is unlikely to find a new point which improves the current best solution. According to Figures 4(a) and 44(b), the accuracy of the objective and constraint models are both validated, seeing all the sample points lie in the interval [−3,3]. The distribution of sample points is demonstrated in Figure 4(f). Note that only the infill points which have $EI = 0$ are plotted. As the limited memory LBFGS algorithm will randomly select a point when it cannot find a new point with $EI > 0$. As a result, the infill points with $EI = 0$ are not representative in terms of the objective and constraint values. The X-axis is the constraint value while Y-axis is the objective value. The dashed red line is the constraint limit. Most of the initial points violate the constraint while only one infill point is beyond the constraint limit, which proves that the infill strategy works well. The point inside the red circle is the optimal solution for the EQ schemes. We solve the EN (Figures 5(a)–55(f)) and RE scheme (Figures 6(a)–66(f)) in the same way.



(a) (b)

Figure 3: (a) Relationship between the spatial spread of density and density and (b) comparison of flow and density between unified delay toll scheme and distance toll scheme.



(a) (b)

Figure 4: Continued.

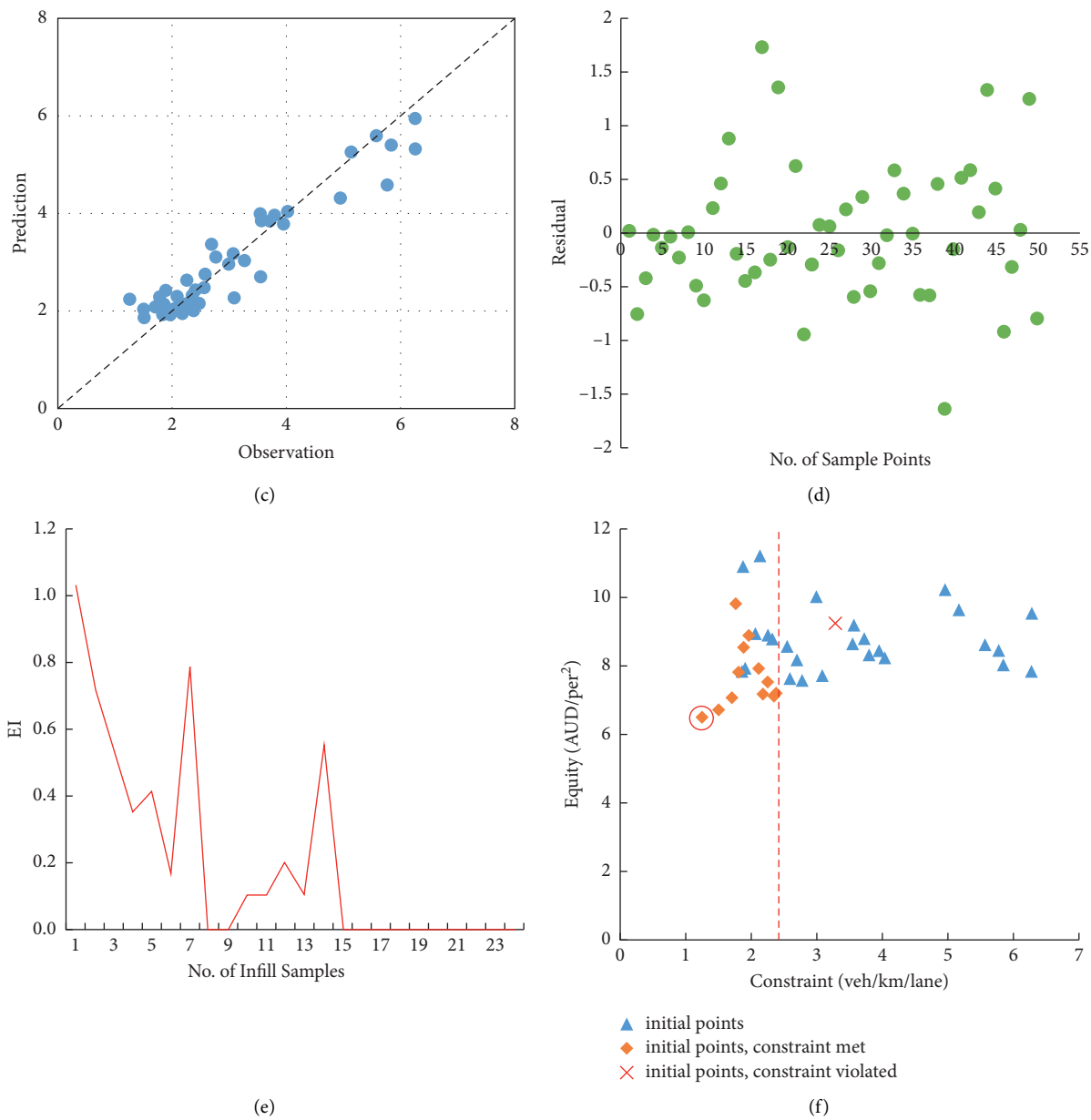


Figure 4: Results for EQ scheme: (a) comparison between predictions and observations for the objective model; (b) residuals of sample points for the objective model; (c) comparison between predictions and observations for the constraint model; (d) residuals of sample points for the constraint model; (e) convergence history of EI; (f) distribution of sample points.

3.4. Comparison among Three Schemes. In this section, we compare the results of three schemes. The optimal toll plans for three schemes are illustrated in Figures 7(a)–7(c). The primary Y-axis is the distance toll rate for HV groups, while the secondary Y-axis is the delay toll rates for AV groups. According to Figure 7(d), all schemes control the network density over time around the critical value (25 veh/km/lane) well. The constraint values under all three schemes are within 2.45 veh/km/lane (shown in Table 3). The RE scheme lets road users to pay 15.79 (10 × 3 AUD) which is 40% higher than the revenues collected in the EQ or EN scheme. However, when we look into the total travel time reduction and the constraint value, the network performance under RE scheme is similar to those under other two schemes. It indicates that RE scheme is less cost-effective. The extremely high toll may induce the low marginal effect.

The EN scheme produces least emissions (7.02 kg). According to the QUARTET pollution emission model, the emission rate goes up with the increase of the cruising speed. Moreover, the optimal toll plan for the EN scheme reduce the emissions by controlling the total time travelling inside the pricing zone, seeing the least total travel time (7726 min.veh) among three schemes. It suggests that the EN scheme can encourage individuals to travel less time in the pricing zone. Nonetheless, since the HVs and AVs are charged differently in our work, in order to increase the

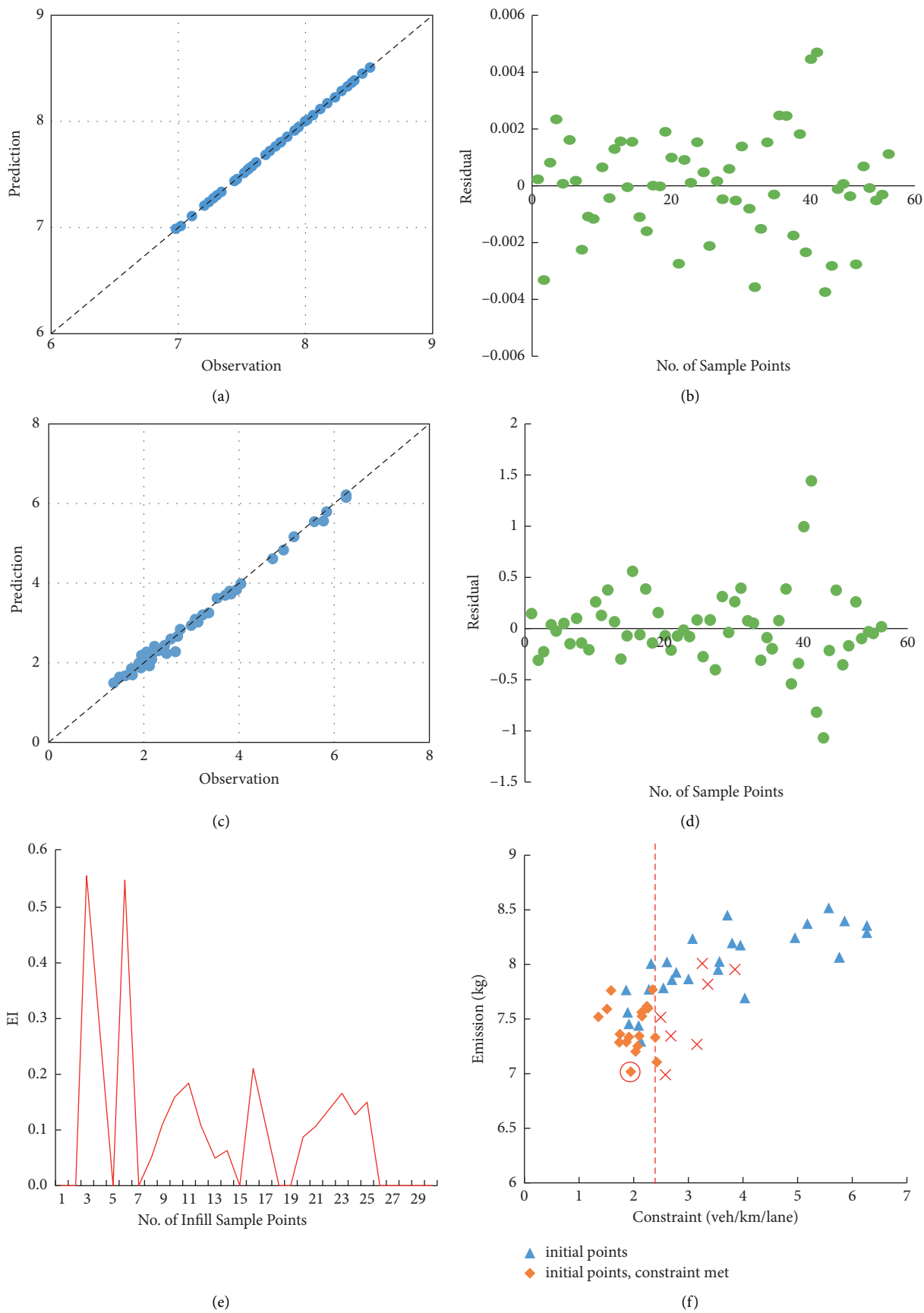


Figure 5: Results for EN scheme: (a) comparison between predictions and observations for the objective model; (b) residuals of sample points for the objective model; (c) comparison between predictions and observations for the constraint model; (d) residuals of sample points for the constraint model; (e) convergence history of EI; (f) distribution of sample points.

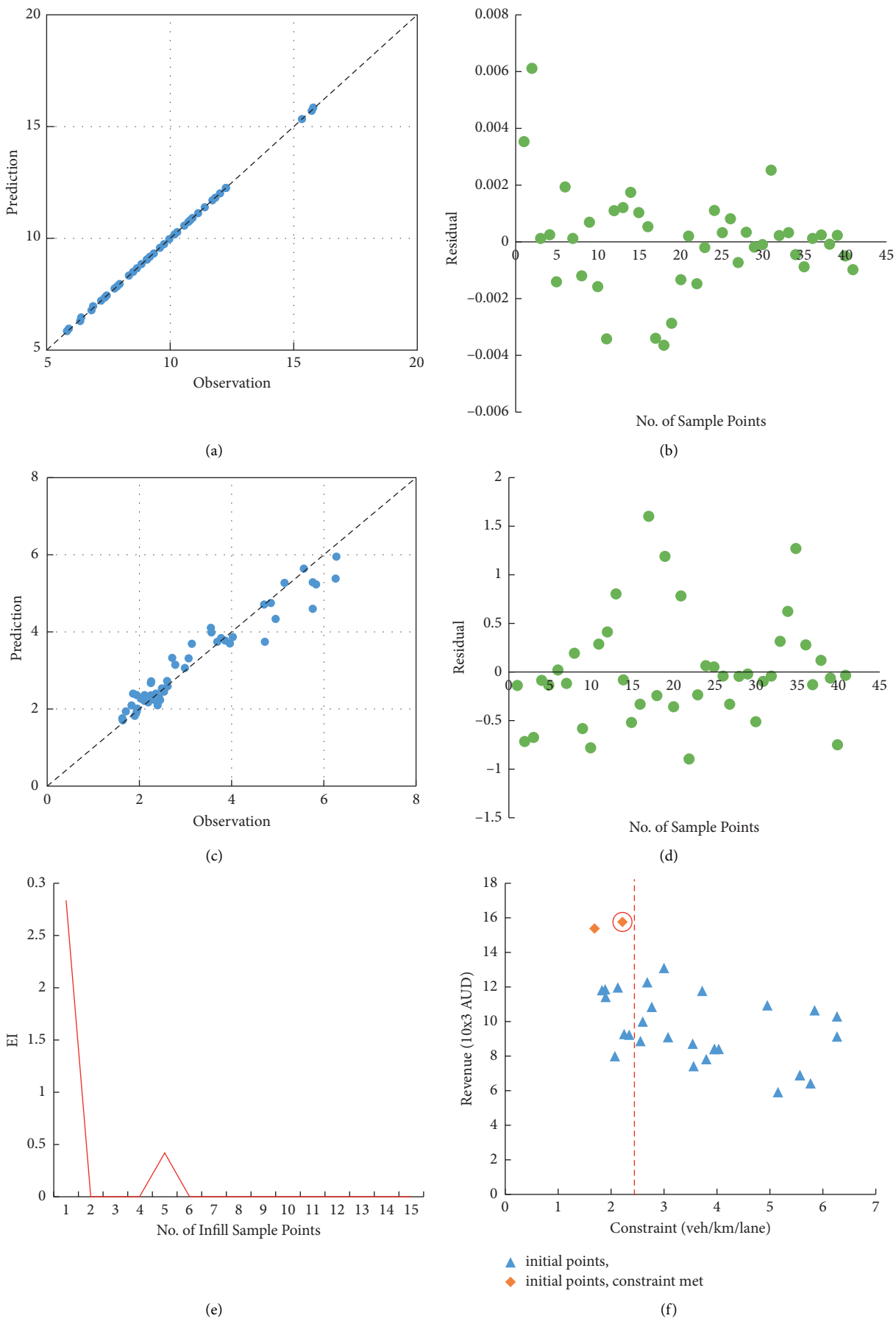


Figure 6: Results for RE scheme: (a) comparison between predictions and observations for the objective model; (b) residuals of sample points for the objective model; (c) comparison between predictions and observations for the constraint model; (d) residuals of sample points for the constraint model; (e) convergence history of EI; (f) distribution of sample points.

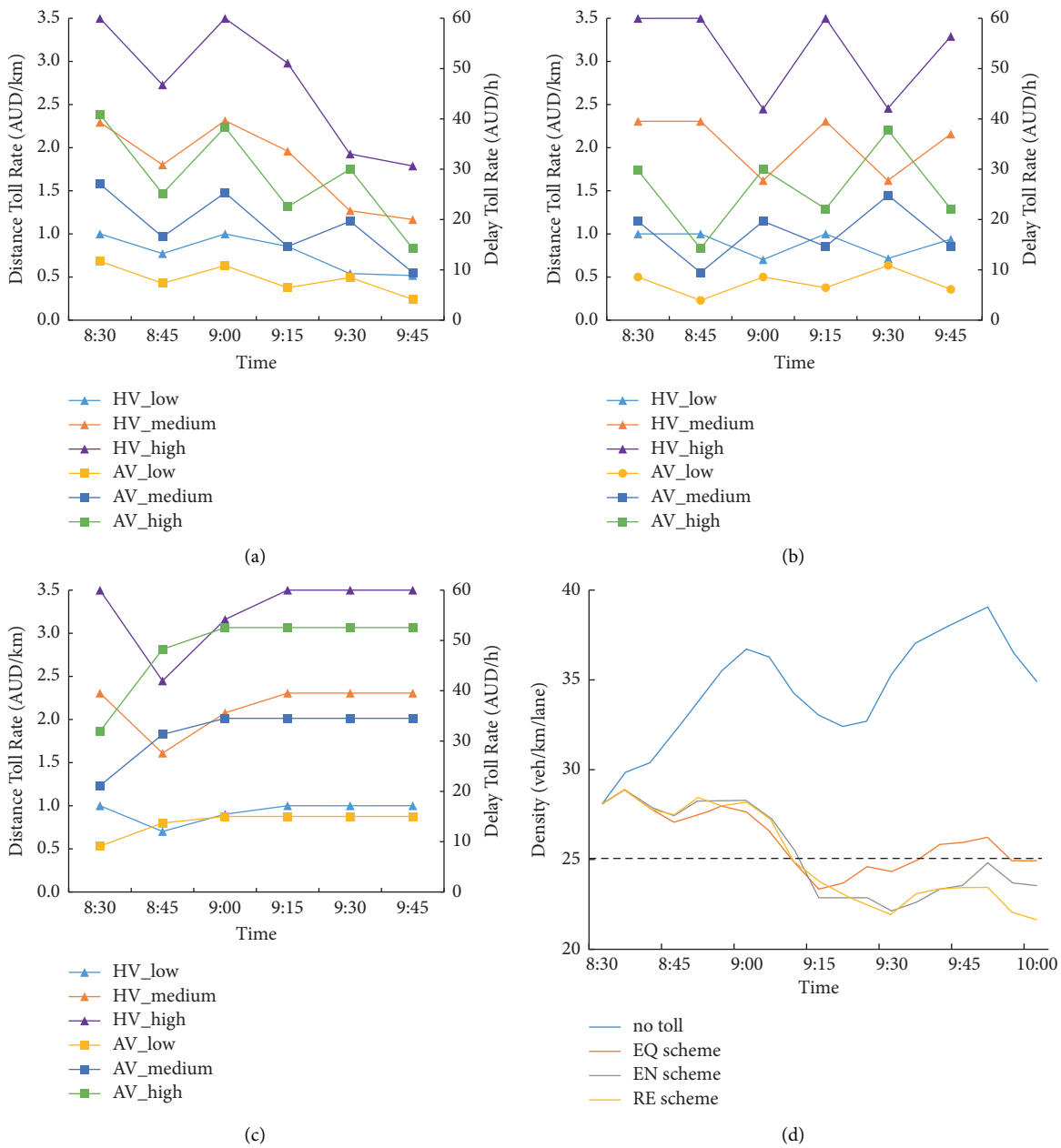


Figure 7: (a) Optimal toll plan for EQ scheme; (b) optimal toll plan for EN scheme; (c) optimal toll plan for RE scheme; (d) density over time under different schemes.

Table 3: Metric comparison among different schemes.

Scheme name	No-toll	EQ scheme	EN scheme	RE scheme
Equity (AUD/per^2)	9.14	**6.53**	8.00	9.50
Emission (kg)	9.35	7.64	**7.02**	7.63
Revenue (10 × 3 AUD)	0	11.92	11.20	**15.79**
Constraint value (veh/km/lane)	9.8	1.25	1.94	2.23
Total travel time (min.veh)	10,665	8589	**7726**	7829
Average speed (km/h)	27.69	30.30	30.47	30.62

The bold ones are the most important values.

acceptance of the policy, $\sigma^2(\mathrm{ATC}_g)$ should be as small as possible. From this perspective, the EQ scheme outperforms other schemes with $\sigma^2(\mathrm{ATC}_g) = 6.53$ AUD/per^2. In addition, compared with the no-toll scenario, the EQ scheme reduces the NOx emissions by 18.3% and collects 11.92 (10 × 3 AUD), which is appreciable.

4. Conclusion

The formulation of DCP optimization involves high-dimension decision variables as well as the expensive-to-evaluate objective function. Meanwhile, with the emerging of autonomous technology, HVs and AVs will coexist in the transportation network in the near future. How to design appropriate congestion pricing schemes under mixed traffic is challenging. In this paper, we divide the road users into six groups according to the vehicle type and VOTs. Three congestion pricing schemes with different objectives are investigated. The network performance is considered as a complex constraint to keep the network density close to the critical value. The RK model is adopted to estimate the input-output mapping. Initial and infill sampling strategies are combined to search for optimal toll plans.

A large-scale simulation-based DTA model of Melbourne, Australia, is used to demonstrate the results of proposed pricing schemes. The RE scheme is not advocated because it charges road users too much money without achieving much better network performance. The phenomenon illustrates that demand originates from the pricing zone is evitable. High tolls may obtain low marginal effect. The EN scheme produces least emissions and maintains the relatively high average speed. In this study, AVs are charged by delay while HVs are charged by distance. Meanwhile, people with higher VOTs will have high toll rate. As a result, we should guarantee that the proposed pricing scheme can be accepted by all groups. Hence, the EQ scheme is superior to EN and RE schemes for its smallest $\sigma^2(\mathrm{ATC}_g)$. To ensure the efficiency of our pricing schemes.

We should also guarantee that travelers report their VOTs honestly. Nan et al. [15] conducted an in-depth discussion on how to stimulate people to tell truth. For instance, the mechanism design which does not require travelers to provide confidential information. In addition, punishments such as fines or banning the liars from the pricing zone for a certain period are considerable. As long as the government takes a lead role to boost the group-based discrimination, the proposed pricing schemes can be put into practice.

Data Availability

The simulation-based traffic model is available from GitHub through https://github.com/meeadsaberi/dynamel.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

The first author was supported by the China Scholarship Council grant scheme.

References

[1] A. Pigou, *The Economics of Welfare*, MacMillan and Co., Ltd, London, UK, 1st edition, 1932.

[2] Z. Y. Gu, S. Shafiei, Z. Y. Liu, and M. Saberi, "Optimal distance- and time-dependent area-based pricing with the network fundamental Diagram," *Transportation Research Part C*, vol. 95, pp. 1–28, 2018.

[3] M. Merchand, "A note on optimal tolls in an imperfect environment," *Econometrica*, vol. 36, pp. 575–581, 1968.

[4] Z. Y. Gu, Z. Y. Liu, Q. X. Cheng, and M. Saberi, "Congestion pricing practices and public acceptance: a review of evidence," *Case Studies on Transport Policy*, vol. 6, no. 1, pp. 94–101, 2018.

[5] D. A. Hensher and Z. Li, "Referendum voting in road pricing reform: a review of the evidence," *Transport Policy*, vol. 25, pp. 186–197, 2013.

[6] L. M. D. Freitasm, "Equity effects of congestion charges: an exploratory analysis with MATSim," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2670, pp. 75–82, 2017.

[7] M. Gaunt, T. Rye, and S. Allen, "Public acceptability of road user charging: the case of Edinburgh and the 2005 referendum," *Transport Reviews*, vol. 27, no. 1, pp. 85–102, 2007.

[8] Y. F. Chen, N. Zheng, and H. Vu, "A novel urban congestion pricing scheme considering travel cost perception and level of service," *Transportation Research Part C*, vol. 125, 2021.

[9] R. Arnott, A. D. Palma, and R. Lindsey, "Schedule delay and departure time decisions with heterogeneous commuters," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1197, pp. 56–67, 1988.

[10] H. Yang and X. Zhang, "Multi-class network toll design problem with social and spatial equity constraints," *Journal of Transportation Engineering*, vol. 128, no. 5, pp. 420–428, 2002.

[11] F. Xiao, Z. Qian, and M. Zhang, "The morning commute problem with coarse toll and non-identical commuters," *Networks and Spatial Economics*, vol. 11, no. 2, pp. 343–369, 2011.

[12] V. V. D. Berg and E. Verhoef, "Congestion tolling in the bottleneck model with heterogeneous values of time," *Transportation Research Part B*, vol. 45, pp. 60–78, 2011.

[13] N. Zheng and N. Geroliminis, "Area-based equitable pricing strategies for multimodal urban networks with heterogeneous users," *Transportation Research Part A*, vol. 136, pp. 357–374, 2020.

[14] D. J. Fagnant and K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," *Transportation Research Part A*, vol. 77, pp. 167–181, 2015.

[15] Z. H. Yao, H. R. Jiang, Y. Cheng, Y. Jiang, and B. Ran, "Integrated schedule and trajectory optimization for connected automated vehicles in a conflict zone," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2020.

[16] Z. H. Yao, T. R. Xu, Y. Jiang, and R. Hu, "Linear stability analysis of heterogeneous traffic flow considering degradations of connected automated vehicles and reaction time," *Physica A: Statistical Mechanics and Its Applications*, vol. 561, 2021.

[17] H. T. Zhong, W. Li, M. Burris, A. Talebpour, and K. Sinha, "Will autonomous vehicles change auto commuters' value of travel time?" *Transportation Research Part D*, vol. 83, 2020.

[18] L. N. Liu and J. F. McDonald, "Economic eficiency of second-best congestion pricing schemes in urban highway systems," *Transportation Research Part A*, vol. 33, no. 3, pp. 157–188, 1999.

[19] Z. Liu, Q. Meng, and S. Wang, "Speed-based toll design for cordon-based congestion pricing scheme," *Transportation Research Part C*, vol. 31, pp. 83–98, 2013.

[20] Z. Liu, S. Wang, and Q. Meng, "Optimal joint distance and time toll for cordon-based congestion pricing," *Transportation Research Part C*, vol. 69, pp. 81–97, 2014.

[21] X. Chen, C. Xiong, H. Xiang, Z. Zheng, and Z. Lei, "Time-of-day vehicle mileage fees for congestion mitigation and revenue generation: a simulation-based optimization method and its real-world application," *Transportation Research Part C*, vol. 63, pp. 71–95, 2016.

[22] Z. Tan, H. Yang, and R. Y. Guo, "Dynamic congestion pricing with day-to-day flow evolution and user heterogeneity," *Transportation Research Part C*, vol. 61, pp. 87–105, 2015.

[23] S. V. Ukkusuri, L. Han, and K. Doan, "Dynamic user equilibrium with a path based cell transmission model for general traffic networks," *Transportation Research Part B*, vol. 46, no. 10, pp. 1657–1684, 2012.

[24] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, "Lipschitzian optimization without the lipschitz constant," *Journal of Optimization Theory and Applications*, vol. 79, no. 1, pp. 157–181, 1993.

[25] S. Amaran, N. V. Sahinidis, B. Sharda, and S. J. Bury, "Simulation optimization: a review of algorithms and applications," *Annals of Operations Research*, vol. 240, pp. 351–380, 2016.

[26] N. Zheng, G. R´erat, and N. Geroliminis, "Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment," *Transportation Research Part C*, vol. 62, pp. 133–148, 2016.

[27] Z. Y. Gu and M. Saberi, "Simulation-based optimization of toll pricing in large-scale urban networks using the network fundamental Diagram: a cross-comparison of methods," *Transportation Research Part C*, vol. 122, 2021.

[28] X. Chen, L. Zhang, X. He, C. Xiong, and Z. Li, "Surrogate-based optimization of expensive-to-evaluate objective for optimal highway toll charges in transportation network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 29, pp. 359–381, 2013.

[29] X. He, X. Q. Chen, C. F. Xiong, Z. Zhu, and L. Zhang, "Optimal time-varying pricing for toll roads under multiple objectives: a simulation-based optimization approach," *Transportation Science*, vol. 51, pp. 412–426, 2017.

[30] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statistical Science*, vol. 4, pp. 409–423, 1989.

[31] N. Geroliminis and D. Levinson, "Cordon pricing consistent with the physics of overcrowding," in *Proceedings of the 18th International Symposium on Transportation and Traffic Theory*, Springer, Berlin, Germany, July 2009.

[32] N. Zheng, R. A. Waraich, K. Axhausen, and N. Geroliminis, "A dynamic cordon pricing scheme combining the macroscopic fundamental Diagram and an agent-based traffic model," *Transportation Research Part C*, vol. 46, pp. 1291–1303, 2012.

[33] M. Ashfaq, Z. Y. Gu, S. T. Waller, and M. Saberi, "Comparing dynamic user equilibrium and noniterative stochastic route choice in a simulation-based dynamic traffic assignment model: practical considerations for large-scale networks," *Journal of Advanced Transportation*, vol. 2021, Article ID 6667335, 16 pages, 2021.

[34] Z. Y. Gu, S. T. Waller, and M. Saberi, "Surrogate-based toll optimization in a large-scale heterogeneously congested network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, pp. 638–653, 2019.

[35] M. Börjesson, J. Eliasson, M. Hugosson, and K. Brundell-Freij, "The stockholm congestion charges - 5 Years on effects, acceptability and lessons learnt," *Transport Policy*, vol. 20, pp. 1–12, 2012.

[36] Quartet Deliverable, *Assessment of Current Tools for Environmental Assessment in QUARTET, DRIVE II Project V2018*, QUARTET, London, UK, 1992.

[37] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, pp. 239–245, 1979.

[38] J. Ekstrom, I. Kristoffersson, and N. H. Quttineh, "Surrogate-based optimization of cordon toll levels in congested traffic networks," *Journal of Advanced Transportation*, vol. 50, pp. 1008–1033, 2016.

[39] A. I. J. Forrester, A. J. Keane, and N. W. Bressloff, "Design and analysis of "noisy" computer experiments," *AIAA Journal*, vol. 44, pp. 2331–2339, 2006.

[40] H. M. Gutmann, "A radial basis function method for global optimization," *Journal of Global Optimization*, vol. 19, no. 1, pp. 201–227, 2001.

[41] Y. H. Xiao, Z. X. Wei, and Z. G. Wang, "A limited memory BFGS-type method for large- scale unconstrained optimization," *Computers and Mathematics with Applications," Computers & Mathematics with Applications*, vol. 56, pp. 1001–1009, 2008.

[42] E. Brochu, V. M. Cora, and N. D. Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," 2014, https://arxiv.org/abs/1012.2599.

[43] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, no. 4, pp. 455–492, 1998.

[44] A. Molinaro, R. Simon, and R. Pfeiffer, "Prediction error estimation: a comparison of resampling methods," *Bioinformatics*, vol. 21, no. 15, pp. 3301–3307, 2005.

[45] V. L. Knoop and S. P. Hoogendoorn, "Empirics of a generalized macroscopic fundamental Diagram for urban freeways," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2391, no. 1, pp. 133–141, 2013.

[46] M. D. Simoni, A. J. Pei, R. A. Waraich, and S. P. Hoogendoorn, "Marginal cost congestion pricing based on the network fundamental Diagram," *Transportation Research Part C*, vol. 56, pp. 221–238, 2015.