Research Article

Metamodel-Based Optimization Method for Traffic Network Signal Design under Stochastic Demand

Wei Huang, Xuanyu Zhang, Haofan Cheng, and Jiemin Xie

School of Intelligent Systems Engineering, Sun Yat-Sen University, Guangzhou 510006, China
School of Intelligent Systems Engineering, Shenzhen Campus of Sun Yat-Sen University, Guangdong 518107, China

Correspondence should be addressed to Jiemin Xie; xiejm28@mail.sysu.edu.cn

Received 13 March 2023; Revised 20 April 2023; Accepted 24 April 2023; Published 27 May 2023

Copyright © 2023 Wei Huang 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.

Traffic network design problems (NDPs) play an important role in urban planning. Since there exist uncertainties in the real urban traffic network, neglecting the uncertainty factors may lead to unreasonable decisions. This paper considers the transportation network signal design problem under stochastic origin-destination (OD) demand. In general, solving this stochastic problem requires a large amount of computational budget to calculate the equilibrium flow corresponding to a certain demand distribution, which limits its real applications. To reduce the computational time in calculating the equilibrium flow under stochastic demand, this paper proposes a metamodel-based optimization method. First, a combined metamodel that integrates a physical modeling part and a model bias generic part is developed. The metamodel is used to approximate the time-consuming average equilibrium flow solution process, hence to improve the computational efficiency. To further improve the convergence and the solution optimality performance of the metamodel-based optimization, the gradient information of traffic flow with respect to the signal plan is incorporated in the optimization model. A gradient-based metamodel algorithm is then proposed. In the numerical example, a six-node test network is used to examine the proposed metamodel-based optimization method. The proposed combined metamodel is compared with the benchmark method to investigate the importance of incorporating a model bias generic part and the traffic flow gradient information in the combined metamodel. Although there is a reduction in solution optimality since the metamodel is an approximation of the original model, the metamodel methods greatly improve the computational efficiency (the computational time is reduced by 4.84 to 13.47 times in the cases of different initial points). By incorporating the model bias, the combined metamodel can better approximate the original optimal solution. Moreover, incorporating the gradient information of the traffic flow in the optimization search algorithm can further improve the solution performance. Numerical results show that the gradient-based metamodel method can effectively improve the computation efficiency while slightly reducing the solution optimality (with an increase of 0.09% in the expected total travel cost).

1. Introduction

Traffic network design is a basic means to improve traffic flow distribution and alleviate traffic congestion in urban traffic networks. The network design problem (NDP) is usually to determine optimal network supply decisions, such as adding new links or improving the capacity of existing ones, with certain objectives (e.g., maximizes social benefits or minimizes total travel cost), while considering users’ route choice behavior [1–6]. There exist different network supply decisions for the traffic network design problems, including road network expansion design, signal control optimization design, and road tolling design. For network traffic signal control, it focuses on determining optimal signal timing plans that can trigger better equilibrium flow patterns with optimal network performance. It is also called the combined traffic assignment and signal control problem [7–12] or anticipatory network traffic control [13–16] because the signal control anticipates the effect of route choice response.

The traffic network signal design problem has been extensively explored in the literature. Allsop [17] first proposed the concept of combined signal control and traffic assignment and developed an iterative optimization method.
to achieve the equilibrium solution by alternately modifying the signal timings and the equilibrium flows. However, it is reported that the solution of the iterative optimization method highly depends on the initial point (initial assignment), and the equilibrium solution is generally not necessarily the optimal solution [18–20]. In view of the drawbacks of the iterative approach, Yang and Yagar [21] established the network signal control model from the perspective of global optimization and proposed a global optimization approach, which is usually formulated as a bilevel programming optimization model. The upper level problem is the signal control optimization problem, which optimizes the network performance with flow constraints. The lower level problem solves the user equilibrium (UE) problem [22–25] under the given signal timing plan. The global optimization approach needs to simultaneously consider the traffic flow equilibrium and signal control optimization, which makes it time-consuming and difficult to solve. The computational budget increases especially for large-scale road network problems.

Traditional traffic network design problems usually assume fixed or deterministic traffic conditions, such as fixed traffic demand. However, the transportation system is generally affected by many uncertain factors on demand and supply, for instance, OD demand fluctuations, link capacity variations, special events, and random route choice behavior. Ignoring the uncertainty effects in the decision-making process may result in inaccurate evaluations and suboptimal control plans [26–28]. Li et al. [29] dealt with NDP under stochastic demand and reported that demand stochasticity affects the reliability of the optimal solution and its real application. Lv and Liu [30] also showed that the stochastic features of traffic demand will significantly affect the optimal signal control settings as well as the associated equilibrium flow pattern of the transportation network, leading to suboptimal network performance. This paper focuses on network traffic control under stochastic demand. To account for the impact of demand stochasticity and ensure the reliability of the solution, it is required to calculate the equilibrium flows under a large number of random demand scenarios, which substantially adds to the computation complexity of the control optimization problem. The high computational budget of the control method that addresses the demand uncertainty limits its real-time and large-scale network applications.

Metamodel (or surrogate model) is a common method to solve nonlinear problems with high computational budget. It typically makes use of simple analytical models, which are called metamodels to approximate the original time-expensive analysis or models, so as to improve the overall computation efficiency [31]. In general, metamodels can be classified in two types: physical metamodels and functional metamodels. The physical metamodel usually develops problem-specific model to approximate the original problem from first principle. Their functional form and parameters have a physical or structural interpretation. Osorio and Bierlaire [32] considered simulation-based optimization approach for traffic signal control and developed a simplified analytical queueing network model to approximate the complex queue network in the simulation. To improve the accuracy of the physical metamodel, it is necessary to conduct a model parameter fitting. A conventional two-step approach was usually applied to reduce the errors between the physical model and the real system [33]. However, for complex transportation system, it is difficult to calibrate model parameters and establish an accurate physical model.

Another is the functional metamodel, which is composed of generic functions with general purposes. The functional metamodel is usually developed based on analytical tractability, following a data-driven regression analysis approach. Hence, it does not include physical information regarding the underlying problem. A common way is to apply low-order polynomials for constructing the functional metamodel. In recent years, the functional metamodel has been gradually applied to the domain of traffic network design. Chen et al. [34] introduced Kriging surrogates to solve the network design problem under dynamic traffic assignment. Li et al. [35] proved the convergence of solving the continuous network design problem with the surrogate model and showed the advantages of the surrogate model in computation efficiency. However, the functional metamodel relies highly on data, and the approximation performance outside the range of sample data is often unsatisfactory. It is typically that the data-driven method has a rigid requirement on the sample data and parameter fitting in order to achieve a good approximation performance.

To overcome the shortcomings of the physical and functional metamodels, Osorio and Bierlaire [32] proposed a metamodel that combines a functional component with a physical component to approximate the traffic queueing process. The purpose of the functional component is to provide a more accurate local approximation, and the physical component is to provide a good global approximation. It has been shown that the combined metamodel method has a faster convergence speed and better fitting performance [36,37]. The above-mentioned studies focus on local intersection signal control, which does not consider the travelers’ route choice behavior. Moreover, the demand uncertainty is not explicitly addressed.

Following the combined metamodeling approach, this paper proposes a metamodel-based optimization method for traffic network signal design under stochastic demand. Taking account of the stochastic features of traffic demand, a global optimization model is established with the goal of minimizing the expected total travel cost of the road network. Therefore, it needs to calculate the equilibrium flows under random demand scenarios and derive the expected performance. In order to improve the computational efficiency of solving the average equilibrium traffic state, a metamodel that consists of a traffic assignment model (physical modeling) and a model bias (generic function) is constructed to approximate the expected equilibrium traffic flow. This paper further proposes to incorporate the gradient information of traffic flow with respect to the decision variable (the signal timing plan in our case) in the combined metamodel. By incorporating the gradient information, it is able to improve the parameter fitting performance and
hence the solution optimality. A gradient-based metamodel algorithm is then developed to solve the network signal control optimization problem. The main contributions of this paper are summarized as follows:

1. A metamodel-based optimization method is developed for traffic network signal design under stochastic demand. To explicitly address the stochasticity in traffic demand and improve the computation efficiency, a combined metamodel that consists of a physical modeling part and a model bias generic part is proposed to approximate the time-consuming average equilibrium flow solution process.

2. A gradient metamodel scheme is further developed to make use of the gradient information of traffic flow to improve solution performance.

3. A gradient-based metamodel algorithm is proposed to solve the network signal control optimization problem.

The rest of the paper is organized as follows. Section 2 elaborates the problem formulation and methodology of the metamodel-based optimization for traffic network signal design. Section 3 presents the numerical example on a test network. Insights into the properties of the proposed metamodel method and the solution performance of the method are demonstrated. Concluding remarks are discussed in Section 4.

2. Metamodel-Based Optimization Method for Traffic Network Signal Design

2.1. Traffic Network Design Problems under Stochastic Demand. In view of the inherent variations in traffic demand, in the traffic network design problem, the stochastic features of traffic demand need to be explicitly addressed in the optimization model to ensure reliable decisions. For the traffic network signal design problem under stochastic demand, it can be expressed as the problem of minimizing the expected total travel cost of the road network as follows:

\[
\min_{\mathbf{g}} \mathbb{E} \left[ \sum_{i \in L} Z_i(\mathbf{g}, \mathbf{x}) \right], \quad (1)
\]

s.t. \( \mathbf{x} = \mathbf{x}^{eq} (\mathbf{g}, d_k) \quad k = 1, 2, \ldots, K, \) \( (2) \)

\( f(\mathbf{g}) = 0, \) \( (3) \)

\( g_{\text{min}} \leq \mathbf{g} \leq g_{\text{max}}, \) \( (4) \)

where \( Z_i \) represents the travel cost of link \( i \), which is a function of signal settings \( \mathbf{g} \) (such as green splits) and link flow \( \mathbf{x} \). \( L \) represents the total number of links in the road network. Equation (1) is the objective function, i.e., minimizing the average of the total travel cost of all links. Constraint condition (2) represents the equilibrium flow constraint. The equilibrium link flow pattern \( \mathbf{x} \) is related to the signal settings \( \mathbf{g} \) and the stochastic traffic demand \( d_k, k = 1, 2, \ldots, K \), represents the sample size of stochastic demand. The equilibrium flow is derived by the traffic assignment model \( \mathbf{x}^{eq} (\mathbf{g}, d_k) \). Equation (3) is the signal timing constraint. Equation (4) sets the upper and lower limits of the signal control variables. According to the discussion above, in the presence of demand uncertainty, it is necessary to calculate the equilibrium link flow under a certain demand distribution. In other words, calculations of the traffic assignment model and the total travel cost function are repeated a large number of times, leading to a computational-intensive optimization problem. Therefore, the computational budget restricts the application of the stochastic network design in real-time or large-scale problems.

2.2. Metamodel-Based Optimization Method for Network Signal Control. In order to improve the efficiency of calculation, this paper proposes a metamodeling approach. As shown in equation (1), the objective is to minimize the expected total travel cost, which requires calculating the equilibrium flow under different demand scenarios. It usually involves a large number of scenarios (sample size) in order to achieve a comparable accuracy level, leading to a time-intensive calculation process. Therefore, the metamodel is developed as a surrogate of the expensive calculation process to improve computational efficiency. First, we assume that the expected total travel cost is associated with the expected equilibrium link flow \( \mathbf{x}^{ave} = \mathbb{E}[\mathbf{x}^{eq}(\mathbf{g}, d)] \) under stochastic demand. In general, calculating the expected equilibrium flow takes most of the computation time.

To reduce the computation time, we introduce a metamodel \( \mathbf{x}^{meta} (\mathbf{g}, \bar{d}; \beta, \theta) \) as a surrogate of \( \mathbf{x}^{ave} \), to approximate the expensive calculation of the expected equilibrium flow with different demand scenarios. \( \bar{d} \) is the average traffic demand. \( \beta \) and \( \theta \) are parameters of the metamodel, whose feasible regions are \( \beta \) and \( \Theta \), respectively.

Based on the metamodel, traffic network signal design problem under stochastic demand can be written as follows:

\[
\min_{\mathbf{g}} \mathbb{E} \left[ \sum_{i \in L} Z_i(\mathbf{g}, \mathbf{x}) \right], \quad (5)
\]

s.t. \( \mathbf{x} = \mathbf{x}^{meta} (\mathbf{g}, \bar{d}; \beta, \theta), \) \( (6) \)

\( f(\mathbf{g}) = 0, \) \( (7) \)

\( g_{\text{min}} \leq \mathbf{g} \leq g_{\text{max}}, \) \( (8) \)

where in equation (6) we calculate the expected equilibrium flow with the metamodel. Other constraints are the same of the original problem. In order to improve the approximation accuracy and make the approximate result of the metamodel closer to the actual average equilibrium flow, a suitable parameter set should be determined. The parameter fitting can be formulated as a general least square error problem:

\[
\min_{\beta, \theta} \sum_{i \in L} \mathbb{E} \left[ \mathbf{x}^{eq}(\mathbf{g}, d_i) \right] - \mathbf{x}^{meta}(\mathbf{g}, \bar{d}; \beta, \theta) \right]^{2}, \quad (9)
\]
where \( t \) is the iteration indicator and \( \mathbf{g}_t \) is signal settings at iteration \( t \).

As discussed, the metamodel is an analytical approximation of the expensive calculation process of the original optimization, i.e., the calculation of average equilibrium flow under stochastic demand. The metamodel-based optimization method then iterates over two main steps, including a metamodel fitting step and a signal control optimization step (i.e., the traffic network signal design). Figure 1 shows the interaction between different modules in the metamodel-based optimization framework. The metamodel is constructed based on a sample of calculation results of the average equilibrium flow. Given the signal settings and stochastic demand, we can calculate the average equilibrium flow which involves solving the equilibrium flow for each demand and taking the average value. In the metamodel fitting step, based on the current sample of average equilibrium flow, the metamodel is fitted by solving the optimization problem (9). Then, the signal control optimization step uses the fitted metamodel as constraint (6) to solve the signal control design problem and derive the optimal signal settings \( \mathbf{g}_t \). Further, the updated signal settings are implemented in the expensive calculation process, which leads to a new calculation result of the average equilibrium flow \( \mathbf{x}^\text{ave}_t \).

As the new sample becomes available, the metamodel is fitted again, leading to a more accurate metamodel. The two steps iterate until convergence. At convergence, an accurate metamodel that approximates the original model can be obtained, and ultimately the optimal control scheme derived based on the metamodel should be close to the solution of the original traffic network design problem under stochastic demand.

### 2.3. Equilibrium Flow and Metamodel Fitting

The traffic network signal design considers the equilibrium flow constraint. From a network planning perspective, the signal control involves the interaction between the controller and travelers. The controller anticipates travelers’ route choice response when determining the signal settings, while travelers make route choice based on traffic conditions depending on the signal settings \([13–15]\). Hence, the route choice response and the resulting equilibrium flow pattern, which is derived by solving a traffic assignment problem, are taken as constraints in the network signal design process. In general, finding the solution of traffic assignment problem can be represented as a fixed-point problem. The link flow determines the link travel cost, and the route travel cost calculated from the link travel cost will affect the route selection and hence the traffic flow assignment. This can be formulated as the following equations:

\[
\begin{align*}
\mathbf{c} &= \mathbf{C} (\mathbf{x}, \mathbf{g}), \\
\mathbf{x} &= \mathbf{F} (\mathbf{c}) \\
&= \mathbf{B} \mathbf{h} (\mathbf{c}),
\end{align*}
\]  

(10)

where \( \mathbf{c} \) is the link cost vector, which is calculated as a function of link flows \( \mathbf{x} \) and signal settings \( \mathbf{g} \), \( \mathbf{h} (\mathbf{c}) \) represents the route flow, and \( \mathbf{B} \) is incidence matrix of link-route flow, which can transform the route flow function into the link flow function \( \mathbf{F} (\mathbf{c}) \). Finding the solution of equations (10) represents a fixed-point problem, for which there exist different solution algorithms \([38]\). Assuming that the link cost function \( \mathbf{C} (\mathbf{x}, \mathbf{g}) \) is continuous and strictly increasing with \( \mathbf{x} \) and the link flow function \( \mathbf{F} (\mathbf{c}) \) is continuous and monotonically decreasing with \( \mathbf{c} \), the existence and uniqueness of the fixed-point solution is guaranteed \([39]\). The solution of the fixed-point problem is the equilibrium flow. The signal settings will affect the equilibrium state because the travel cost highly depends on the signal settings. Given a set of signal settings \( \mathbf{g}_0 \), the equilibrium flow can be expressed as follows:

\[
\mathbf{x} = \mathbf{F} (\mathbf{C} (\mathbf{x}, \mathbf{g} = \mathbf{g}_0)).
\]

(11)

The solution of this fixed-point problem depends on the link travel cost function and link flow function. Equation (11) shows that the equilibrium flow is related to the signal settings.

In this paper, the metamodel is used to approximate the average value of equilibrium flow under stochastic demand. As mentioned above, the metamodel that combines a functional component with a physical component is considered. The purpose of the functional component is to provide a detailed local approximation and that of the physical component is to provide a good global approximation. This study develops a combined metamodel to approximate the average equilibrium flow. We formulate the traffic assignment model \( \mathbf{F} (\mathbf{g}; \theta) \) as the physical modeling part, \( \mathbf{g} \) is the set of signal settings, and \( \theta \) is the set of model parameters to be calibrated. The generic function is expressed as \( \Phi (\mathbf{g}; \beta) \). Then the metamodel can be written as follows:

\[
\mathbf{x}^\text{meta} (\mathbf{g}; \beta, \theta) = \mathbf{F} (\mathbf{g}; \theta) + \Phi (\mathbf{g}; \beta),
\]

(12)

where \( \beta \) is the parameter of the generic function. In this paper, we consider the low-order polynomials function and define \( \Phi \) as follows:

\[
\Phi (\mathbf{g}; \beta) = \beta_0 + \sum_{i=1}^N \beta_i \mathbf{g}_i,
\]

(13)

where \( N \) is the number of signal control variables. Therefore, the objective function of the metamodel fitting problem (9) can be written as follows:

\[
\min_{\beta, \theta} \sum_i \left( \mathbf{x}_i^{\text{ave}} - \mathbf{x}_i^{\text{meta}} (\mathbf{g}; \beta, \theta) \right)^2 + \sum_{i=1}^N \beta_i^2.
\]

(14)

The first term is the error between the approximate result of the metamodel and the actual average equilibrium flow \( \mathbf{x}_i^{\text{ave}} \), and the second term is the ridge penalty term.

Next, we elaborate on the development of the metamodel. First, the physical metamodel that only considers the simplified problem-specific model (the traffic assignment model in our case) is established. Then, the concept of model bias is introduced, and a combined metamodel with the model bias as the generic part is proposed. To improve the solution performance, this paper further integrates the traffic flow gradient information into the combined metamodel.
2.3.1. Physical Metamodel. The physical metamodel only considers the simplified physical model, that is, the traffic assignment model $F(g; \theta)$, as shown in equation (11). Generally, a two-step scheme is used to iteratively update model parameters $\theta$ and determine the optimal signal settings as follows:

$$
\theta_t = \arg \min_{\theta} \| \bar{x}_{\text{ave}} - F(g_t; \theta) \|,
$$

(15)

$$
g_{t+1} = \arg \min_{g} z(g, F(g_t; \theta_t)).
$$

(16)

where $t$ is the iteration indicator, and constraints of the optimization problem are not included for simplicity. Equation (15) represents the problem of model parameter estimation, which minimizes the distance between the approximate metamodel and the average equilibrium flow by updating the model parameters. Equation (16) represents the signal optimization problem. Based on the fitted metamodel, the optimal signal setting is calculated by minimizing the total travel cost, and these two steps iterate until convergence.

2.3.2. Combined Metamodel with Model Bias. In view of the physical modeling error, this paper introduces a concept of model bias, which is defined as the error between the traffic assignment model and the average equilibrium link flow of the system as follows:

$$
b_t = \bar{x}_{\text{ave}} - F(g_t; \theta).
$$

(17)

At iteration step $t$, the model bias is calculated by the average equilibrium flow and the traffic assignment model with the corresponding signal settings $g_t$ as follows:

$$
b_t = \bar{x}_{\text{ave}}(g_t) - F(g_t; \theta).
$$

(18)

With the help of model bias, a combined metamodel is developed as follows:

$$
x^{\text{meta}}(g_t; \theta) = F(g_t; \theta) + b_t.
$$

(19)

The combined metamodel consists of two terms. The first term is the traffic assignment model, which is the physical modeling part. The second term of model bias $b_t$ corresponds to the generic function part, which is updated by using the data during the iteration process. The signal optimization design problem is formulated as follows:

$$
x^{\text{meta}}(g_t; \theta) = F(g_t; \theta) + b_t,
$$

(20)

$$
g_{t+1} = \arg \min_{g} z(g, x^{\text{meta}}(g_t; \theta)).
$$

(21)

The model bias $b_t$ is updated with the data obtained during the iteration process.

2.3.3. Gradient-Based Metamodel. In this paper, a combined metamodel considering gradient information of traffic flow is proposed. In general, gradient is an important information for finding the descending direction of the optimization problem. In the traffic assignment model, the gradient reflects the variations of the traffic flow when the signal settings change. Incorporating gradient information generally improves solution performance in terms of convergence and solution optimality, i.e., faster convergence and better solution point [14]. Patwary et al. [40] proposed a metamodel method with traffic flow gradient for an efficient calibration of large-scale traffic simulation models. For calculating the gradient of the equilibrium flow, this paper makes use of a finite difference (FD) approach, which requires perturbing each signal control variable and calculates the corresponding derivative component.

$$
\frac{\partial F(g)}{\partial g_i} = \frac{F(g_1, g_2, \ldots, g_i + h, \ldots, g_N) - F(g_1, g_2, \ldots, g_i, \ldots, g_N)}{h},
$$

(22)

where $F(g)$ is the equilibrium flow function (i.e., the traffic assignment model) and $h$ is a small perturbation on signal control variable $g_i$. Calculating the gradient of the traffic assignment model, i.e., $\nabla F(g; \theta)$, is trivial. However, it is computationally intensive to estimate the gradient of the actual average equilibrium flow, i.e., $\nabla x^{\text{ave}} = \nabla E[x^{\text{ave}}(g_t, d_t)]$. This is because for each changed signal control variable, derivatives of equilibrium flows under different demand scenarios are required, which involve repeatedly solving the traffic assignment model. Regarding the computational budget on calculating the
gradient information, this paper applies a finite difference approximation method [41], which uses the results recorded in previous iterations to estimate the Jacobian matrix of the average equilibrium flow. In each iteration, the Jacobian matrix is updated by the average equilibrium flow obtained in the historical iterations. Assuming that the number of control variables is \( n_p \), then \( n_p + 1 \) control parameters and corresponding values of the average equilibrium flow are required, i.e., \( \{ g_0, \ldots, g_{n_p} \} \) and \( \{ x_{i,ave}^0, \ldots, x_{i,ave}^{n_p} \} \). The Jacobian matrix of iteration \( t \) can be calculated by the following formula:

\[
\frac{\partial x_{i,ave}}{\partial g} \bigg|_{g_t} = (\Delta g_t)^{-1} \begin{bmatrix}
(x_{i,ave}^t - x_{i,ave}^{t-1})^T \\
\vdots \\
(x_{i,ave}^1 - x_{i,ave}^{t-n_p})^T
\end{bmatrix},
\]

where \( \Delta g_t = [g_t - g_{t-1}, \ldots, g_t - g_{n_p}]^T \).

Therefore, the Jacobian matrix estimation based on the average equilibrium flow recorded in the historical iterations can be implemented as follows:

Step 1: set a set of initial signal settings and the corresponding values of the average equilibrium flow under stochastic demand, i.e., \( \{ g_0, \ldots, g_{n_p} \} \) and \( \{ x_{0,ave}, \ldots, x_{n_p,ave} \} \); calculate the initial Jacobian matrix \( (\partial x_{i,ave}/\partial g)|_{g_0} \) from equation (21), i.e., \( \nabla x_{ave}^0 \); then solve the metamodel-based optimization; and derive \( g_1 \).

Step 2: apply \( g_t \), calculate the average equilibrium flow, update the set of signal control settings and flows, i.e., \( \{ g_{t+1}, \ldots, g_{t+n_p} \} \) and \( \{ x_{t+1,ave}, \ldots, x_{t+n_p,ave} \} \), and calculate \( \nabla x_{ave}^t = (\partial x_{i,ave}/\partial g)|_{g_t} \).

Step 3: input \( \nabla x_{ave}^t \) to solve the metamodel-based optimization and update \( g_{t+1} \).

Step 2 and Step 3 are iterated until the convergence condition is satisfied. During the process, the gradient of the traffic assignment model \( \nabla F(g; \theta) \) and the gradient of the average equilibrium flow \( \nabla x_{ave}^t \) should be calculated for each iteration step. Considering the gradient information of traffic flow, the following combined metamodel is constructed:

\[
x_{meta}^t(g; \theta) = F(g; \theta) + x_{i,ave}^t - F(g; \theta)
+ (\nabla x_{ave}^t - \nabla F(g; \theta)) (g - g_t).
\]

The gradient information is added to the calculation of model bias. As shown in equation (24), the model bias part is updated with \( x_{i,ave}^t - F(g; \theta) + (\nabla x_{ave}^t - \nabla F(g; \theta)) (g - g_t) \), corresponding to the generic function \( \Phi(g; \theta) \) in equation (12).

Now at iteration step \( t \), the combined metamodel can be determined by the average equilibrium flow \( x_{ave}^t \), the equilibrium flow calculated by traffic assignment model \( F(g; \theta) \), the gradient of the average equilibrium flow \( \nabla x_{ave}^t \), and the gradient of traffic assignment model \( \nabla F(g; \theta) \). Compared with the metamodel with model bias in equation (19), the gradient-based metamodel (24) not only considers the value of model bias but also takes account of the gradient information, i.e., the first-order derivative information. This gradient-based metamodel method, which incorporates the gradient information of the metamodel at each local point \( g_t \), can ensure the first-order optimality at convergence.

Applying the gradient-based metamodel, the optimal signal setting \( g_{t+1} \) at \( (t+1) \) th iteration is determined by solving the following optimization problem:

\[
g_{t+1} = \arg \min_g z(g, x_{meta}^t(g; \theta)),
\]

\[
s.t. \quad x_{meta}^t(g; \theta) = F(g; \theta) + x_{i,ave}^t - F(g; \theta)
+ (\nabla x_{ave}^t - \nabla F(g; \theta)) (g - g_t),
\]

\[
x_{meta}^t(g; \theta) \geq 0,
\]

\[
g_{min} \leq g \leq g_{max}.
\]

Choose a control step size \( \mu \) and update the signal settings by

\[
g_{t+1} = g_t + \mu (g_{t+1} - g_t),
\]

where \( \mu \) represents the control step size with a range of \( [0, 1] \).

Algorithm 1 summarizes the solution process of the traffic network signal design problem under stochastic demand by using the combined metamodel method considering gradient information.

3. Numerical Examples

3.1. Simulation Setup. In this paper, a combined metamodel considering gradient information of traffic flow is proposed. This section establishes a simulation network to test the performance of the proposed method. Figure 2 is the test network, which consists of one OD pair (from node 1 to node 6), 8 links, and 5 routes. Link travel cost is calculated using a linearized Bureau of Public Roads (BPR) function. The signal control plans of intersection node 3 and node 4 are decision variables. Assuming that the intersections operate in a two-phase timing plan, the green split is to be optimized. The signal loss time is not considered in this case, i.e., \( g_3 = g_4 = 1 \) at intersection 3 and \( g_4 = g_5 = 1 \) at intersection 4.

Assuming that the travelers follow a nested logit (NL) structure for making route choice decisions, the probability of choosing route \( i \) can be expressed as follows:

\[
p_{meta}[i] = \frac{e^{-w_{t,i}}}{\sum_{j \neq i} e^{-w_{t,j}}} \cdot \frac{e^{\xi_{Y_t}}}{\sum_{k \neq i} e^{\xi_{Y_k}}},
\]

where the route travel cost is denoted by \( w \). The route choice set is divided into subsets \( J_1, \ldots, J_k \). \( \xi \) is the probability of dispersion parameters of the two-layer structure of NL, associated with the first and second choice levels, respectively.
Step 1: initialization. Set the parameters \( \theta \) of traffic assignment model \( F(\mathbf{g}; \theta) \). Set a set of initial signal settings \( \{\mathbf{g}_0, \ldots, \mathbf{g}_{n_s}\} \).

Step 2: apply the initial signal setting. Based on the initial signal settings \( \{\mathbf{g}_0, \ldots, \mathbf{g}_{n_s}\} \), calculate the average equilibrium flow \( \mathbf{x}^\text{ave} = E[\mathbf{x}_i^\theta(\mathbf{g}, d_i)] \), and get the corresponding \( \{\mathbf{x}_0^\text{ave}, \ldots, \mathbf{x}_{n_s}^\text{ave}\} \); calculate the gradient of the traffic assignment model and the gradient of the initial average equilibrium flow, construct the combined metamodel as equation (24), and apply it into equations (25)–(28) to solve the signal control optimization problem, obtain the control \( \mathbf{g}_1 \), and update the iteration step \( t = 1 \).

Step 3: calculate the average equilibrium flow. Implement \( \mathbf{g}_1 \) to derive \( \mathbf{x}^\text{ave} = E[\mathbf{x}_i^\theta(\mathbf{g}, d_i)] \) and update the set of signal settings and the corresponding average equilibrium flow, i.e., \( \{\mathbf{g}_1, \ldots, \mathbf{g}_t\} \) and \( \{\mathbf{x}_1^\text{ave}, \ldots, \mathbf{x}_t^\text{ave}\} \).

Step 4: update the gradient-based metamodel. Calculate the gradient of the metamodel \( \nabla F(\mathbf{g}; \theta) \) according to equation (22); calculate the Jacobian matrix \( (\partial \mathbf{x}^\text{ave} / \partial \mathbf{g})_t \) based on equation (23) to obtain \( \nabla \mathbf{x}_t^\text{ave} \); update the combined metamodel at the current iteration, i.e., \( \mathbf{x}^\text{ave}(\mathbf{g}; \theta) = F(\mathbf{g}; \theta) + \mathbf{x}_t^\text{ave}_t + F(\mathbf{g}; \theta) + (\nabla \mathbf{x}_t^\text{ave} - \nabla F(\mathbf{g}; \theta))(\mathbf{g} - \mathbf{g}_t) \).

Step 5: update the signal setting. Calculate \( \mathbf{g}_{t+1} \) by solving the control optimization (25)–(28) with the updated metamodel and update the signal setting \( \mathbf{g}_{t+1} = \mathbf{g}_t + \mu(\mathbf{g}_{t+1} - \mathbf{g}_t) \).

Step 6: check termination. Stop if the termination condition is satisfied; otherwise, set \( t = t + 1 \) and go to Step 3.

**Algorithm 1**: Gradient-based metamodel algorithm.

The link travel cost \( c \) is derived by a linearized BPR function [42]. Defining the free-flow travel time \( c_0 \), saturation flow \( s \), and a coefficient \( \alpha \), the link travel cost is expressed as a function of the link flow \( x \) and signal settings \( \mathbf{g} \) as follows:

\[
c = C(x, g) = c_0 + \alpha \frac{x}{gs}.
\]  

(31)

For nonsignalized links, signal settings \( g \) are equal to 1. The equilibrium link flow that can be obtained by solving the fixed-point problem depends on the link travel cost and link flow under a given traffic demand. The above calculations need to be carried out many times under the stochastic traffic demand to obtain the average equilibrium flow and then calculate the average total travel cost of the network.

In this paper, the metamodel method is introduced to simplify the traffic assignment calculation process and approximate the average equilibrium flow. In general, we cannot derive an accurate model of route choice behavior. In this case study, we assume that a multinomial logit (MNL) model with the dispersion parameter \( \theta \) is used to describe the route choice and construct the metamodel of average equilibrium flow. The probability is calculated by the model as follows:

\[
p_{\text{model}}[i] = \frac{e^{-wi\theta}}{\sum_{j \in I} e^{-wi\theta}}.
\]

(32)

The link travel cost is also represented by the BPR function (31). The equilibrium flow is derived by solving the fixed-point problem with MNL and BPR function, which is used as the physical modeling part in the combined metamodel to approximate the average equilibrium flow. The total travel time \( z \) is formulated as a function of the equilibrium flow and signal setting:

\[
z = z(\mathbf{g}, x) = \sum_{i} C(x, g) \cdot x.
\]

(33)

Signal control decisions are to be made based on the metamodel, and the objective is to minimize the expected total travel cost on this network. All optimization problems in this numerical example are solved using the Python optimization toolbox. Characteristics of the network and model parameters are listed in Table 1.

3.2. Sensitivity Analysis of the Model Parameter. As discussed, the traffic assignment model is used as the physical modeling part in the metamodel. In order to evaluate the role of model parameters and examine whether the model performance is sensitive to the parameters, we first conduct...
To illustrate the performance of the proposed method, we compare three metamodel schemes, i.e., the proposed gradient-based metamodel method (GD), the combined metamodel with model bias (bias), and a traditional physical metamodel method (two-step). By comparing with the physical metamodel method, we test the value of adding a model bias generic part in the combined metamodel. Furthermore, by comparing the GD method and the bias method, we validate the role of gradient information in improving solution optimality. Select different initial control points and analyze the convergence performance of the three methods. The initial points are \((g_1, g_2, g_3, g_4, g_5) = (0.3, 0.7, 0.73, 0.27), (0.8, 0.2, 0.8, 0.2), \) and \((0.2, 0.8, 0.2, 0.8), \) respectively, and the control step size \(\mu = 0.7. \)

3.3.2. Solution Performance of the Metamodel Method. To illustrate the performance of the proposed method, we compare three metamodel schemes, i.e., the proposed gradient-based metamodel method (GD), the combined metamodel with model bias (bias), and a traditional physical metamodel method (two-step). By comparing with the physical metamodel method, we test the value of adding a model bias generic part in the combined metamodel. Furthermore, by comparing the GD method and the bias method, we validate the role of gradient information in improving solution optimality. Select different initial control points and analyze the convergence performance of the three methods. The initial points are \((g_1, g_2, g_3, g_4, g_5) = (0.3, 0.7, 0.73, 0.27), (0.8, 0.2, 0.8, 0.2), \) and \((0.2, 0.8, 0.2, 0.8), \) respectively, and the control step size \(\mu = 0.7. \)

Figures 6 and 7 illustrate the convergence performance and the optimal solutions of three methods under different initial points. The selection of initial points typically affects the convergence process of the algorithm. The results show that, compared with the physical metamodel, the combined metamodel greatly improves the optimal solution (in terms of reducing the expected total travel cost) with the help of the model bias. Moreover, by explicitly incorporating the gradient information of traffic flow, the gradient-based metamodel method further improves the solution performance and
converges to a smaller total travel cost (i.e., convergence to a lower contour in Figure 7), which is closer to the original optimal solution.

3.3.3. Analysis of the Computation Time and Solution Optimality. Solving the network signal control problem under stochastic demand requires carrying out the fixed-point problems multiple times to obtain the corresponding equilibrium flow and the expected total travel cost, leading to a computationally expensive process. Therefore, this paper proposes a gradient-based metamodel method to approximate the average equilibrium flow function, replacing the time-consuming part of the signal control design problem. In this regard, the metamodel method can be evaluated from two aspects, namely, computational efficiency and solution optimality (i.e., whether the optimal solution derived from the metamodel method is close to the optimal solution of the original problem).

Tables 3–5 list the results of three metamodel methods with different initial points, including the computation time and the optimal solution performance (the expected total travel cost). In this example, different initial points have little influence on the optimal solutions. The entire process of metamodel-based optimization includes metamodel fitting, solving the optimal control, and calculating the sample average equilibrium traffic flow. The time for metamodel fitting and solving the optimal control problem with the metamodel methods is in total approximately 0.04 s. The time to obtain the average equilibrium traffic is about 0.28 s. Therefore, in terms of computation efficiency, the time to solve the average equilibrium flow problem accounts for approximately 85% of the total calculation time in the metamodel optimization method. This shows that the time-consuming process in the iteration is the multiple runs of the traffic assignment model under stochastic demand, which in turn validates the need for a more efficient surrogate for the calculation of the average equilibrium flow. An improvement factor (defined as the ratio of the computation time of the benchmark optimal control scheme to the computation time of the metamodel method) is introduced to capture the improvement of the computation time. The results show that although there is a small reduction in solution optimality, the metamodel methods can significantly reduce the computation time (the computation time is reduced by 4.84 to 13.47 times under different initial points). With the help of the model bias, the combined metamodel can better approximate the original optimal solution. As indicated in

Figures 3: The impact of route choice parameter θ on (a) link flow, (b) total travel cost, and (c) optimal control.
Tables 3–5, compared with the traditional physical metamodel method, the combined metamodel method with model bias improves the total travel cost. Moreover, by incorporating the gradient information, the gradient-based method further improves the optimal solution. The numerical results show that the proposed gradient-based metamodel method can effectively improve the computation efficiency while slightly increasing the total travel cost (i.e., 0.09%, 0.09%, and 0.06% under the three initial points). The influence of control step size on the gradient-based metamodel method is further analyzed. The step size adjustment methods with different optimization descent
directions are considered. The commonly used step size update methods include Adam, Momentum, and RMSprop algorithms. We select the initial point (0.45, 0.55, 0.5, and 0.5) and compare these step size update methods, as shown in Figure 8. Adam and RMSprop converge slowly at the beginning because they limit the update within a certain range, which however makes the convergence process more stable. Therefore, different control steps will also affect the convergence process of the gradient-based metamodel method. In the solution process, we should carefully select
Figure 7: The total travel cost under different initial points: (a) (0.3, 0.7, 0.73, 0.27), (b) (0.8, 0.2, 0.8, 0.2), and (c) (0.2, 0.8, 0.2, 0.8).

### Table 3: Computation time and optimal solution with initial points (0.3, 0.7, 0.73, and 0.27).

<table>
<thead>
<tr>
<th>Method</th>
<th>Computation time (s)</th>
<th>Improvement factor (iteration)</th>
<th>Total travel cost (veh/h)</th>
<th>Increase in total travel cost (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The original problem under stochastic demand</td>
<td>10.236</td>
<td></td>
<td>2008.76</td>
<td>2052.96</td>
</tr>
<tr>
<td>Physical metamodel</td>
<td>0.8067</td>
<td>12.69 (5)</td>
<td>2013.22</td>
<td>2.20</td>
</tr>
<tr>
<td>Combined metamodel method with model bias</td>
<td>0.7598</td>
<td>13.47 (5)</td>
<td>2010.59</td>
<td>0.22</td>
</tr>
<tr>
<td>Gradient-based metamodel</td>
<td>1.8562</td>
<td>5.51 (10)</td>
<td></td>
<td>0.09</td>
</tr>
</tbody>
</table>

### Table 4: Computation time and optimal solution with initial points (0.8, 0.2, 0.8, and 0.2).

<table>
<thead>
<tr>
<th>Method</th>
<th>Computation time (s)</th>
<th>Improvement factor (iteration)</th>
<th>Total travel cost (veh/h)</th>
<th>Increase in total travel cost (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The original problem under stochastic demand</td>
<td>9.6482</td>
<td></td>
<td>2008.76</td>
<td>2052.92</td>
</tr>
<tr>
<td>Physical metamodel</td>
<td>0.8985</td>
<td>10.73 (5)</td>
<td>2013.18</td>
<td>2.20</td>
</tr>
<tr>
<td>Combined metamodel method with model bias</td>
<td>0.7878</td>
<td>12.25 (5)</td>
<td>2010.57</td>
<td>0.22</td>
</tr>
<tr>
<td>Gradient-based metamodel</td>
<td>1.9916</td>
<td>4.84 (11)</td>
<td></td>
<td>0.09</td>
</tr>
</tbody>
</table>
the initial point and the control step under a specific problem setting.

4. Conclusion

This paper developed a metamodel-based optimization method for traffic network signal design under stochastic OD demand. Solving the network design problem considering uncertainty typically involves an expensive calculation process to derive the equilibrium flows with a certain demand distribution. This paper applied a metamodeling approach and used a metamodel as a surrogate of the expensive calculation process of the average equilibrium flow, so as to enhance the overall computational efficiency. More specifically, based on the concept of model bias, a combined metamodel was developed, which integrates a physical modeling part (i.e., the traffic assignment model) and a model bias generic function. In order to further improve the solution performance, i.e., convergence and solution optimality, of the metamodel-based optimization method, the gradient information of traffic flow was incorporated in the metamodel, which provides a better descent direction of searching for the optimal solution. We tested the proposed gradient-based metamodel method on an example network. Three methods were compared, including our proposed gradient-based metamodel, the combined metamodel with model bias, and the physical metamodel. The comparison was conducted to investigate the importance of incorporating a model bias generic part and the traffic flow gradient information in the combined metamodel. Numerical results showed that there is a trade-off between computation time and solution optimality. Although there is a reduction in solution optimality, the metamodel methods significantly reduce the computation time (by 4.84 to 13.47 times under different initial points). By incorporating the model bias, the combined metamodel is able to better approximate the original optimal solution. Moreover, incorporating the traffic flow gradient information in the search algorithm further improves the solution performance. Comparison results indicated that the proposed gradient-based metamodel method can effectively improve the computation time with a small increase of 0.09% in the expected total travel cost.

In this paper, we apply the linear model to construct the generic function part of the combined metamodel. In future study, more functional forms including higher-order functions can be explored to improve the fitting performance of the method. Moreover, methods that can handle a larger amount of data should be explored. In addition, this paper focuses on developing the methodology and we test the effectiveness of the proposed metamodel method on a small example network. Our further research work will consider applications on larger road networks, probably based on certain traffic simulation models.

Data Availability

The numerical example data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

<table>
<thead>
<tr>
<th>Method</th>
<th>Computation time (s)</th>
<th>Improvement factor (iteration)</th>
<th>Total travel cost (veh/h)</th>
<th>Increase in total travel cost (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The original problem under stochastic demand</td>
<td>9.8356</td>
<td></td>
<td>2008.76</td>
<td></td>
</tr>
<tr>
<td>Physical metamodel</td>
<td>0.8975</td>
<td>10.96 (5)</td>
<td>2052.52</td>
<td>2.18</td>
</tr>
<tr>
<td>Combined metamodel method with model bias</td>
<td>0.7639</td>
<td>12.88 (5)</td>
<td>2012.95</td>
<td>0.21</td>
</tr>
<tr>
<td>Gradient-based metamodel</td>
<td>1.6565</td>
<td>5.94 (9)</td>
<td>2009.94</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Figure 8: Convergence performance under different step sizes.
Acknowledgments

The study was supported by the National Natural Science Foundation of China (No. 52102401), the Guangdong Basic and Applied Basic Research Foundation (No. 2023A1515012886), and the Shenzhen Science and Technology Program (Grant No. 202206193000001, 20220817201129001).

References


