Bayesian Stochastic Kriging Metamodel for Active Traffic Management of Corridors

Lei Zhang, Xiqun (Michael) Chen, Xiang He, Chenfeng Xiong and Zheng Zhu
Department of Civil and Environmental Engineering, University of Maryland
1173 Glenn Martin Hall, College Park, MD 20742, United States

Abstract

The transportation system, particularly commuting corridors of freeways and major arterials in metropolitan areas are congested. The purpose of implementing active traffic management (ATM) strategies, e.g. high-occupancy/toll lane management and congestion warning ahead of freeway diversion points, is primarily to improve the corridor-wide performance. Utilizing a simulation-based dynamic traffic assignment model, this paper proposes a Bayesian stochastic Kriging metamodel to optimize integrated planning and operational ATM strategies for corridors. Since transportation simulations are influenced by a variety of uncertainties, e.g. random seed, signal timing, route choice behaviors and etc., we observe stochastic outputs given the same input setting. The developed approach accounts for model uncertainty raised by stochastic travel behaviors and the induced heteroscedasticity in random simulation errors. The simulation based optimization approach is tested both in a synthetic network and a real-world corridor of I-270 (freeway) and MD-355 (arterial) in the State of Maryland. Field measurements by fixed traffic flow detections are used to calibrate the travel demand and supply model. Results show that the joint optimization of travel demand management and operational strategies is promising to reduce the corridor-wide average travel time and enhance the vehicle throughput.

Keywords
Simulation based optimization, Bayesian stochastic Kriging, active traffic management, high-occupancy/toll, freeway diversion rate

1. Introduction

Corridor active traffic management (ATM) is a pro-active approach to corridor travel demand management and traffic operations. ATM will reallocate traffic in transportation infrastructure, provide real-time congestion information and additional capacity to peak-hour traffic, improve the detection and response to work zones, reduce delays resulted from recurrent congestion and/or incidents, and thus enhance the corridor’s performance in efficiency, safety and reliability. Successful implementations of ATM can be found worldwide and especially in Europe and U.S. Promising ATM techniques currently under operation in many metropolitan areas in U.S. include managed lanes, advanced traveler information provision (e.g. congestion warning, dynamic re-routing guidance, etc.) and other strategies.

Taking a freeway work zone as an example, this paper considers two types of corridor management strategies from perspectives of travel planning and traffic operations:

(I) High-occupancy/toll (HOT) lanes have become an increasingly prevalent congestion pricing measure in the U.S. over the past few decades. Benefits of HOT lanes in travel time reduction, freeway efficiency improvement, and bottleneck congestion mitigation have been widely studied in multiple region-specific case studies [1, 2].

(II) Dynamic message sign (DMS) is the most common way to provide real time travel information to drivers and encourage diversion before they approach the congested areas. In previous research, the optimization of diversion rate has been analyzed through heuristic methods [3-5]. In this paper, we assume that the relationship between diversion rate and the information provided through DMS is already clear, and the objective of this paper is to search for the optimal diversion rate that creates the best system performance.
Although both the HOT rate and the diversion rate can be dynamically controlled separately, which are well studied in the literature, it is challenging to jointly optimize different corridor ATM strategies in practice. Since transportation simulations are influenced by a variety of uncertainties, e.g. random seed, signal timing, route choice behaviors and etc., we observe stochastic outputs given the same integrated ATM strategies. Different from the existing studies, we aim to develop an approach accounts for model uncertainty raised by stochastic travel behaviors and the induced heteroscedasticity in random simulation errors. This paper proposes a Bayesian stochastic Kriging metamodel to optimize integrated ATM for corridors utilizing a simulation-based dynamic traffic assignment model. The objective is to reduce average trip travel time, which is mainly from the government’s perspective. Taking advantage of the simulation, evaluation and optimization of combined strategies of transportation planning and traffic operations can be achieved, such as travel demand policies (e.g. HOV, HOT) and advanced traveler information systems (e.g. congestion warning, en-route diversion). This paper aims to jointly optimize the HOT rate in peak hours and the freeway diversion rate under the congestion warning information provision.

The rest of the paper is organized as follows: Section 2 presents the Bayesian stochastic Kriging model formulation, which is then tested by a numerical example of a toy network in Section 3. Section 4 demonstrates the field application in a corridor-level freeway/arterial network in Maryland followed by optimization result discussions. Finally, Section 5 concludes the whole paper.

2. Bayesian Stochastic Kriging Surrogate Model

In this section, a Bayesian stochastic Kriging model that incorporates both parameter uncertainties and heteroscedastic simulation noises is developed based on the regressing Kriging we applied in our previous study [6]. The proposed model is derived in a Bayesian analysis framework which endeavors to estimate parameters of an underlying distribution based on the observed data.

Since transportation simulations are influenced by a variety of uncertainties, e.g. random seed, actuated signal timing, route choice behaviors and etc., we assume that the simulation outputs (observations) are random with heteroscedastic variances. The stochastic Kriging method predicts a response by summarizing a linear model and a high frequency variation component that represents fluctuations around the trend. In this study, we consider the following stochastic Kriging model

\[ y_r(x) = q^T(x)\beta + Z(x) + \epsilon_r(x), \quad r = 1, 2, \ldots \]  

where \( y_r(x) \) is the observed response obtained from the \( r \)th simulation replication at point \( x \), \( q(x) = [q_1(x), \ldots, q_m(x)]^T \) is an \( m \times 1 \) vector of known regression basis functions (e.g. polynomial functions) of \( x \), and \( \beta \) is an \( m \times 1 \) vector of unknown weight parameters of each regression basis function. The stochastic nature of \( Z(x) \) is referred as extrinsic uncertainty [7] because it is imposed to reconstruct the metamodel. The covariance of variation components between two arbitrary points is given by the correlation function, i.e.

\[ \text{Cov}(Z(x^{(i)}), Z(x^{(j)})) = \sigma^2 \psi(\|x^{(i)} - x^{(j)}\|), \quad i, j = 1, \ldots, n, \]  

where \( \sigma^2 \) is the variance of \( Z(x) \). \( \psi(\|x^{(i)} - x^{(j)}\|) \) is the Kriging basis function (extrinsic spatial correlation function) that depends only on \( x^{(i)} - x^{(j)} \) (spatial dependence), given by \( \psi(x^{(i)}, x^{(j)}) = \exp\left(-\sum_{l=1}^{k} \theta_l (x^{(i)_l} - x^{(j)_l})^2 \right), \quad i, j = 1, \ldots, n \), where \( \theta = [\theta_1, \ldots, \theta_k]^T \) is a vector of scaling coefficients that allow different widths of the basis function for each dimension. \( \epsilon_r(x) \) is the intrinsic uncertainty that is only associated with simulation random noises. A series of simulation replications \( \{\epsilon_r(x), \epsilon_r(x), \ldots\} \) at the same point \( x \) are assumed to be independent and identically distributed.

The Kriging surrogate approach assumes that the joint distribution of \( f_0 = f(x^{(o)}) \) and \( f = [f(x^{(1)}), f(x^{(2)}), \ldots, f(x^{(n)})]^T \) is a \((n+1)\)-multivariate Gaussian distribution given by

\[ \begin{bmatrix} f_0 \\ f \end{bmatrix} \sim \mathcal{N}_{n+1}[q_0^T Q \beta, \sigma^2 \Psi], \]  

where \( q_0 = q^T(x^{(0)}) \) is an \( m \times 1 \) vector of regressions at \( x^{(0)} \in \mathbb{R}^d \). Take the \( m \times n \) matrix of regression basis functions as \( Q = [q_1(x^{(0)}), q_2(x^{(0)}), \ldots, q_m(x^{(0)})]^T \), the \((i, j)\)th element of which is \( q_j(x^{(i)}) \) for \( 1 \leq i \leq n, 1 \leq j \leq m \).
Zhang, Chen, He, Xiong and Zhu

Similar to the $\times nn$ extrinsic covariance matrix of $\sigma^2 \Psi$, let $\Sigma_\epsilon$ be the $\times nn$ intrinsic covariance matrix implied by the simulation noise. To simplify the covariance structure of simulation noises and keep its heteroscedasticity feature, the estimation of $\Sigma_\epsilon$ can be given by

$$\hat{\Sigma}_\epsilon = \begin{pmatrix} s_z^2(x^{(1)})/R_1 & 0 & \cdots & 0 \\ 0 & s_z^2(x^{(2)})/R_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_z^2(x^{(n)})/R_n \end{pmatrix}$$

(3)

where $R_i$ is the number of simulation replications at $x^{(i)}$, and $s_z^2(x^{(i)})$ is the unbiased sample variance of $R_i$ simulation replications.

**Lemma 1 (Conditional distribution of the multivariate Gaussian distribution).** Suppose that $[f_0, f]^T$ follow the $(n+1)$ dimensional multivariate Gaussian distribution $\mathcal{N}_{n+1}$, i.e.

$$\begin{pmatrix} f_0 \\ f \end{pmatrix} \sim \mathcal{N}_{n+1}\left( \begin{pmatrix} q_0^T \\ Q \end{pmatrix}, \sigma_z^2 \begin{pmatrix} 1 & \psi^T \\ \psi & \Psi + \Sigma_\epsilon / \sigma_z^2 \end{pmatrix} \right)$$

(4)

then the conditional distribution of $f_0$ given $f$ is a univariate Gaussian distribution $\mathcal{N}_1$ given by

$$f_0 | f \sim \mathcal{N}_1\left( q_0^T \beta + \psi^T (\Psi + \Sigma_\epsilon / \sigma_z^2)^{-1}(f - Q \beta), \sigma_z^2 [1 - \psi^T (\Psi + \Sigma_\epsilon / \sigma_z^2)^{-1} \psi] \right)$$

(5)

**Lemma 2.** For any $m \times m$ symmetric and positive definite matrix $\beta \Sigma$, and $m \times 1$ vector $v$, if the probability density function and an $m \times 1$ multivariate random variable $\beta \Sigma$ satisfies

$$p(\beta \Sigma) \propto \exp \left[ -\frac{1}{2} (\beta \Sigma v)^T \beta \Sigma \right]$$

then $\beta \Sigma \sim \mathcal{N}_m(\Sigma v, \Sigma)$.

**Lemma 3.** Suppose that $\sigma_z^2$, $\Psi$, $\Sigma_\epsilon$ are known, then for an arbitrary $\beta$ priori, the best linear unbiased predictor, which minimizes the mean squared prediction error between the linear predictor of the response at $x^{(0)}$, is

$$\hat{f}_0 = q_0^T \hat{\beta} + \psi^T (\Psi + \Sigma_\epsilon / \sigma_z^2)^{-1}(\bar{y} - Q \hat{\beta})$$

(6)

where $\hat{\beta} = [Q^T (\Psi + \Sigma_\epsilon / \sigma_z^2)^{-1} Q]^{-1} Q^T (\Psi + \Sigma_\epsilon / \sigma_z^2)^{-1} \bar{y}$.

Specifically, when $\beta$ has a non-informative priori, i.e. $p(\beta) \propto 1$, the best linear unbiased predictor of the response at $x^{(0)}$ is

$$\hat{f}_0 = q_0^T \hat{\beta} + \psi^T (\Psi + \Sigma_\epsilon / \sigma_z^2)^{-1}(\bar{y} - Q \hat{\beta})$$

(7)

**Theorem 1.** When $\sigma_z^2$ is known, (I) if $p(\beta) \propto 1$ on $\mathbb{R}^m$, then the predictive distribution of the response at $x^{(0)} \in \mathbb{R}^k$ belong to a Gaussian distribution, i.e. $p(f_0 | f) = \mathcal{N}_1(\mu_{f_0}, \sigma_{f_0}^2)$, where

$$\mu_{f_0} = q_0^T \beta + \psi^T (\Psi + \Sigma_\epsilon / \sigma_z^2)^{-1}(\bar{y} - Q \hat{\beta})$$

(8)

$$\sigma_{f_0}^2 = \sigma_z^2 \left[ 1 - (q_0^T, \psi^T)^T \begin{pmatrix} 0_{m \times m} & Q^T \\ Q & \Psi + \Sigma_\epsilon / \sigma_z^2 \end{pmatrix}^{-1} \begin{pmatrix} q_0^T \\ \psi \end{pmatrix} \right]$$

(9)

(II) if $p(\beta) = \mathcal{N}_m(\beta_0, B)$ on $\mathbb{R}^m$, then the predictive distribution of the response at $x^{(0)} \in \mathbb{R}^k$ belong to Gaussian distribution, i.e. $p(f_0 | f) = \mathcal{N}_1(\mu_{f_0}, \sigma_{f_0}^2)$, where

$$\mu_{f_0} = q_0^T \mu_{f_0} + \psi^T (\Psi + \Sigma_\epsilon / \sigma_z^2)^{-1}(\bar{y} - Q \mu_{f_0})$$

(10)

where $\mu_{f_0} = [Q^T (\sigma_z^2 \Psi + \Sigma_\epsilon)^{-1} Q + B^{-1}]^{-1} [Q^T (\sigma_z^2 \Psi + \Sigma_\epsilon)^{-1} \bar{y} + B^{-1} \beta_0]$, and

$$\sigma_{f_0}^2 = \sigma_z^2 \left[ 1 - (q_0^T, \psi^T)^T \begin{pmatrix} -\sigma_z^2 B^{-1} Q^T \\ Q & \Psi + \Sigma_\epsilon / \sigma_z^2 \end{pmatrix}^{-1} \begin{pmatrix} q_0^T \\ \psi \end{pmatrix} \right]$$

(11)
Since the (I) in Theorem 1 is a special circumstance of (II) when \(|B| \to \infty\), we only use (II) in this paper. Note that the priori of \(\beta\) can be estimated by the first-round simulations across all design points. The least squares error estimations of \((\hat{\beta}, B)\) are
\[
\hat{\beta} = (Q^T Q)^{-1} Q^T y, \quad \text{and} \quad \hat{B} = (y_i - Q \hat{\beta}_i)^T (y_i - Q \hat{\beta}_i)(Q^T Q)^{-1} / (n - m),
\]
where \(y_i = [y_{i1}, y_{i2}, \ldots, y_{in}]^T\) is the first-round simulation output vector of all \(n\) points.

For evaluating the predictive distributions of the stochastic surrogate models, we use the Kullback-Leibler divergence \([8]\) for the performance measure. The analytical derivation of the Kullback-Leibler divergence of Gaussian distributions is given by
\[
D_{KL}(N_{\text{Est}} \mid N_{\text{True}}) = \frac{1}{2} \left[ \text{tr}(\Sigma_{\text{True}}^{-1} \Sigma_{\text{Est}}) - (\mu_{\text{True}} - \mu_{\text{Est}})^T \Sigma_{\text{True}}^{-1} (\mu_{\text{True}} - \mu_{\text{Est}}) - n - \ln \left( \frac{\det(\Sigma_{\text{Est}})}{\det(\Sigma_{\text{True}})} \right) \right]
\]
where \(D_{KL}\) is the Kullback-Leibler divergence, \(\mu_{\text{True}}\) and \(\mu_{\text{Est}}\) are the mean values of the two Gaussian distribution \(N_{\text{True}}\) and \(N_{\text{Est}}\), respectively, \(\Sigma_{\text{True}}\) is the nonsingular covariance matrix of \(N_{\text{True}}\), and the estimated covariance matrix of \(N_{\text{Est}}\) is given by \(\Sigma_{\text{Est}} = \hat{\sigma}_e^2 \hat{\Psi} + \hat{\Sigma}_e\)

3. Numerical Example
To illustrate the methodology developed in this paper and compare it with existing models, we set up a toy network to simulate how travelers’ value of time (VOT) influence the route choice behaviors. Our purpose in this section is three-fold: To provide some intuition about what the Bayesian stochastic Kriging method does on approximating the black-box function of a simulation based dynamic traffic assignment problem; to assess the parameter uncertainty; and to evaluate the estimation robustness by comparing the heteroscedastic errors in our model.

3.1 Toy network
The toy network is depicted in Fig. 1, and the configuration of the links is illustrated in Table 1.

![Fig. 1. Numerical illustration network.](image)

<table>
<thead>
<tr>
<th>Link</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length (mile)</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>30</td>
<td>30</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Lanes</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Speed limit (mph)</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Free flow travel time (min)</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Capacity (veh/lane/hour)</td>
<td>1500</td>
<td>1500</td>
<td>1500</td>
<td>1500</td>
<td>1500</td>
<td>1500</td>
<td>1500</td>
<td>1500</td>
<td>1500</td>
<td>1500</td>
</tr>
</tbody>
</table>

This network is coded into DynusT, which is a simulation-based dynamic traffic assignment (DTA) model. Five-hour dynamic travel demand between the OD pair is set up as in Table 2. To allow all vehicles to dissipate, we simulate the network for 7 hours, with no travel demand during the last two hours.

In this numerical example, the average travel time for all vehicles is selected as the dependent variable, and the surface to be approximated is the response of average travel time to VOT. To create the real response surface, we first generate a uniformly distributed sample of VOT, and then run 100 replications of simulation for each specific input of VOT. The mean of the output from the 100 replications is assumed to be the true response. For the estimation of the Bayesian stochastic Kriging model, results from the first 10 replications of simulation for each corresponding VOT value are utilized. Moreover, to verify the advantages of incorporating parameter uncertainties and heteroscedastic simulation noise into the surrogate model, we estimate a regressing Kriging model with the same input as that for the Bayesian stochastic Kriging model, and compare their performance on approximating the real response surface.
Table 2. Dynamic travel demand.

<table>
<thead>
<tr>
<th>Time (hour)</th>
<th>SOV</th>
<th>HOV</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 1</td>
<td>3000</td>
<td>300</td>
</tr>
<tr>
<td>1 - 2</td>
<td>5000</td>
<td>500</td>
</tr>
<tr>
<td>2 - 3</td>
<td>5800</td>
<td>850</td>
</tr>
<tr>
<td>3 - 4</td>
<td>4700</td>
<td>950</td>
</tr>
<tr>
<td>4 - 5</td>
<td>2800</td>
<td>350</td>
</tr>
</tbody>
</table>

*a SOV: single occupancy vehicles.

3.2 Numerical results

To compare the Bayesian stochastic Kriging model developed in Section 3 with existing surrogate-based optimization approaches, e.g. quadratic polynomial response surface method, ordinary Kriging for deterministic input-output relationship, and regressing Kriging, we generate 10 replications to estimate the heterocedastic simulation errors at different design points and 100 replications to approximate the true response distributions, respectively.

Table 3 compares results of four models in terms of six measures of goodness-of-fit. It should be pointed out that the Kullback-Leibler divergence measures the difference of two probability density functions. It is an important performance measure to evaluate the surrogate function of a stochastic simulation optimization problem. Fig. 2 shows the surrogates of regressing Kriging and Bayesian stochastic Kriging corresponding to Table 3.

Table 3. The goodness-of-fit of surrogate models at 30 design points.

<table>
<thead>
<tr>
<th>Goodness-of-fit</th>
<th>RSMª</th>
<th>Ordinary Kriging</th>
<th>Regressing Kriging</th>
<th>Bayesian stochastic Kriging</th>
</tr>
</thead>
<tbody>
<tr>
<td>KL D</td>
<td>N/A</td>
<td>N/A</td>
<td>65.58</td>
<td>38.33</td>
</tr>
<tr>
<td>Root Mean Square Error</td>
<td>1.90</td>
<td>0.19</td>
<td>0.87</td>
<td>0.20</td>
</tr>
<tr>
<td>Maximum Absolute Error</td>
<td>5.67</td>
<td>0.59</td>
<td>2.70</td>
<td>0.56</td>
</tr>
<tr>
<td>Normalized Root Mean Squared Error</td>
<td>1.89%</td>
<td>0.19%</td>
<td>0.86%</td>
<td>0.20%</td>
</tr>
<tr>
<td>Normalized Maximum Absolute Error</td>
<td>3.10</td>
<td>0.33</td>
<td>1.52</td>
<td>0.31</td>
</tr>
<tr>
<td>Pearson correlation coefficient</td>
<td>0.00</td>
<td>0.99</td>
<td>0.98</td>
<td>0.99</td>
</tr>
</tbody>
</table>

ª RSM: response surface method using quadratic polynomial function; ª N/A: not applicable.

Fig. 2. Comparison of surrogate models: (a) Regressing Kriging; (b) Bayesian stochastic Kriging.

4. Integrated Corridor Planning and Operational Optimization

4.1 Study area

The study freeway/arterial corridor is along a 15.50-mile freeway segment of Interstate 270 in the state of Maryland. The left lane on each side is used as an HOV lane in the northbound direction between 15:30 and 18:30 weekdays and in the southbound direction between 6:00 AM and 9:00 AM weekdays. We set up the urban network of arterials and freeways using DynusT. All freeways and major arterials illustrated in Fig. 3 are included in the network, which has 61 traffic analysis zones, 435 nodes and 766 links. Three modes of dynamic OD matrices, i.e. SOV, HOV and trucks, were estimated based on demand data from the regional planning model [9, 10].
In this paper, the planning policy (i.e. HOT toll) and operational strategy (i.e. DMS) are jointly optimized to enhance the corridor level transportation system performance (i.e. average travel time for all finished trips). The optimization problem is given by

\[
\min_{x \in \mathbb{R}^3} \mathbb{E}[f(x)] = \mathbb{E}[f(x_1, x_2, x_3)] \\
\text{s.t. } x_{\text{min}} \leq x \leq x_{\text{max}}
\]

where \( f(x) \) represents the unknown true average trip travel time of the corridor given the input \( x \), \( x_1 \) is the HOT toll rate, \( x_2 \) is the diversion rate of the DMS next to the work zone, guiding travelers to Montrose Pkwy, \( x_3 \) is the diversion rate of the DMS at the off-ramp to MD 187. So \( x \) is a three-dimensional decision variable vector. The box constraints are \( x_{\text{min}} = [0, 0, 0]^T \) and \( x_{\text{max}} = [\text{US$ 5.00}, 100\%, 100\%]^T \), which are lower and upper boundaries for planning and operational strategies, respectively.

### 4.2 Simulation demand and supply calibration

In this paper, the model simulates travels during the whole 24-hour weekday. A total number of 1,053,052 vehicles are loaded in the network during 24 hours. Field collections of urban street signal timing are also included in the network. To calibrate the simulation model, we collect traffic flow data of 35 detector stations from January 1, 2013 to June 30, 2013 along the I-270 corridor, including 11 detectors on I-270 general-purpose lanes, 6 detectors on I-270 HOV lanes, 10 detectors on I-495, 6 detectors on MD 187, and 2 detectors on MD 355. We use 6-month (January 1 to June 30, 2013) empirical loop/microwave data of the freeway network, including lane-by-lane speed, occupancy and volume extracted from fixed detectors [11]. Since simultaneous demand–supply calibration was
Zhang, Chen, He, Xiong and Zhu

found to be superior to demand-only calibration in accuracy [12-14], we will use the lane aggregate traffic data to support the demand and supply calibration.

The mesoscopic traffic simulation in DynusT is based on the anisotropic mesoscopic simulation model, which moves vehicles according to the speed-density relationship [15]. The modified two-regime Greenshield’s type equation is used to quantify the relationship as follows

\[
v = \begin{cases} 
  v_r & \text{if } 0 \leq \rho \leq \rho_c \\
  v_r + (v_t - v_r)(1 - \rho / \rho_r) & \text{if } \rho_r < \rho \leq \rho_t
\end{cases}
\]

where \( v \) is the space-mean speed, \( \rho \) is the density.

Fig. 4(a) shows the comparison of the default traffic flow model setting and the calibrated speed-density relationship for one of the detectors, i.e. station ID 3392 that locates at I-270 NB 0.23 Mile North of Grosvenor Ln. Fig. 4(b) shows simulation results of the freeway network average speed, density and flow by comparing them with 6-month traffic flow data. The 90% confidence interval (CI) is estimated by the 130-weekday data.

4.3 Simulation based optimization

We simulate the 5-hour PM peak from 14:00 to 19:00 of the corridor. The HOT rate takes effects from 15:30 to 18:30, while DMS provides congestion warning information to the work zone for the whole simulation period. A space filling design of experiments, i.e. Latin hypercube sampling [16], is applied to generate initial design points to fit the Bayesian stochastic Kriging surrogate model. We obtain an sample plan by means of an heuristic search [17], i.e. \( X \), including 3^3 = 27 design points plus upper and lower bounds, and the baseline (neither HOT nor DMS implementations). To reduce the influence of random simulation outputs, we run 5 replications for each design point, and each simulation run includes 10 iterations of DTA to achieve the convergence. DynusT obtains valid results when the convergence is achieved after several times of assignments and vehicular platoon simulations. The average simulation takes around 63 min for each replication, and the relative gaps between two adjacent iterations for the DTA were found to be below 7% for all experiments. So the total computational time spent is 158 hours.

Fig. 5 shows the predictive distributions of the baseline and the optimal solution, i.e. \( \hat{\mathbf{x}} = \begin{bmatrix} \text{US$1.42, 100%, 0} \end{bmatrix}^T \), which corresponds to the estimated global mean value of the average trip travel time. The predictive distribution of the average travel time belongs to \( N_{ \text{optima} } (12.01, 0.03^2) \), so the mean value is estimated to reduce from 12.32 min to 12.01 min. We further run 5-replication simulations of the optimal inputs, we find the mean value of simulation outputs is 11.97 min that is very close to the predictive mean value. More network-wide statistics of the 5-hour simulation such as throughput, average overall trip time (including the demand loading time, stop/queueing time, and travelling time) are listed in Table 4.
Table 4. Comparison of the baseline and optima for PM peak simulation results.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Baseline</th>
<th>Optima</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete trips</td>
<td>302,475</td>
<td>302,918</td>
<td>0.15%</td>
</tr>
<tr>
<td>Average overall trip time (min)</td>
<td>12.32</td>
<td>11.97</td>
<td>2.84%</td>
</tr>
<tr>
<td>Average trip distance (mile)</td>
<td>4.94</td>
<td>4.95</td>
<td>-0.20%</td>
</tr>
<tr>
<td>Average travel speed (mph)</td>
<td>24.07</td>
<td>24.82</td>
<td>3.12%</td>
</tr>
</tbody>
</table>

a The negative value indicates no improvement.

Then the mean values of statistics of these replications are utilized to demonstrate performance improvements. Fig. 6(a) illustrates the average trip travel time in every 5 minutes for vehicles that complete their journeys over the entire corridor. The network average travel time is reduced in the optimal case than the baseline. The largest reduction in the average travel time occurs during 17:00 and 18:00, which is a part of the HOV/HOT operation hours. Thus the optimal HOT rate together with DMS implementations successfully help alleviate peak-hour congestion throughout the network. Fig. 6(b) compares the total corridor throughputs of the optimal solutions and the baseline.

As shown in Table 4, the optima increase the cumulative throughput slightly by 0.15% during the study period. The small improvement in the average travel time of all users in corridor (2.84% reduction) corresponds to more than 26 thousand dollars saved when we use the value of time as US$ 15/hour for the 5-hour PM peak simulations. The benefits can be even larger if we consider 24 hours or a long-term effect.
5. Conclusions
Both travel demand management and traffic control measures are effective ways to promote the transportation system performance in real world. However, the joint optimization of these strategies at different analysis levels was seldom discussed in the existing literature. The optimal travel demand management policies (e.g. congestion pricing, parking control, fares for public transit and etc.) were usually investigated through solving analytical problems with macro level user equilibrium as the constraint. While for the analysis of meso/micro level traffic control measures, formulating analytical models at such detailed levels is somewhat inappropriate. Moreover, the stochasticity plays an important role in modelling the travel behavior at this resolution. Thus, there are usually no closed forms for the meso/micro level models, and simulation is the conventional way to evaluate the transportation system performance when a certain traffic control measure is applied. Due to the difference in analysis resolution, the coordinate optimization of travel demand management and traffic control strategies has not been investigated in the past.

Surrogate models can intelligently mimic simulation based objective function evaluation and reduce computational times. This paper proposes to evaluate the transportation system performance under joint application of travel demand management and traffic control measures with simulation. As the computation burden for a medium to large scale network would be very heavy, a surrogate approach is utilized for the simulation based optimization. The major contribution at the methodology front is that the heteroscedasticity of the stochastic simulation output is taken into account. On the basis of the regressing Kriging model, a Bayesian stochastic Kriging metamodel is developed, which assumes a quadratic form of global trend and different variance levels of the random simulation error at different position of the input domain.

A synthetic network is built in DynusT and used to test the performance of the proposed Bayesian stochastic Kriging model. Through comparing the goodness-of-fit of the proposed model with three other surrogate models, i.e. quadratic polynomial response surface method, ordinary Kriging and regressing Kriging, we find that the Bayesian stochastic Kriging model outperforms the other three models in both estimating the mean values and standard errors for a heteroscedastic simulation input-output relationship.

The proposed Bayesian stochastic Kriging model is then applied for coordinate optimization of the HOT toll rate and the freeway diversion rates at two upstream off-ramps in a work zone scenario for a real world freeway/arterial corridor in Maryland, i.e. I-270 and MD 355. Optimization results show that the SOV is allowed to use the HOT lane by paying the US$ 1.42 toll during the operational time. The optima case increases the cumulative throughput slightly by 0.15% and reduces the average travel time of all users by 2.84%, which corresponds to more than 26 thousand dollars saved for the 5-hour PM peak simulations.

Although a Bayesian stochastic Kriging model is proposed to optimize a real-world heteroscedastic problem, we recommend the following further research:

(I) How travelers respond to congestion warning information? It needs more investigations to reveal the compliance rate of en-route travelers to search alternative routes ahead of the work zone in this study area. The compliance rate here is defined as the proportion of complied drivers who make route choices based on their own perception in total travelers. It is obvious that the benefits of DMS implementations critically depend on how drivers respond to the system.

(II) Implementation of more uncertainties in the simulation based optimization. This study demonstrates the application of Bayesian stochastic Kriging model in a real-world corridor management problem with fixed demands. On one hand, due to traveler behavior adjustments to work zone and HOT operations, the elastic travel demand modeling should be integrated into the simulation tool to reveal heterogeneous travelers’ responses/adjustments to different planning and operational strategies. On the other hand, more supply uncertainties can be learned by simulations beyond work zone. Such scenarios include noncurrent capacity reduction induced by traffic incidents or natural disasters, and capacity increase after the road construction.

(III) Coordination of HOT rate and congestion warning with other ATM strategies, e.g. arterial signal timing and ramp metering. A work zone of freeway mainline forms a physical bottleneck for its upstream demands, which can be partially diverted to the parallel arterial by DMS. Thus, the coordinated arterial signal timing would be important to maintain the corridor throughput capability and should be jointly optimized to deal with the vehicle reallocation from freeways to arterials. On the contrary, ramp metering regulates traffic flow entering freeways to decrease traffic congestion on the freeway mainline. As a complete corridor management, the interactions between freeways and arterials will be studied in the future.
Acknowledgements
This research is financially supported by National Science Foundation CAREER Award Project “Reliability as an Emergent Property of Transportation Networks”, the U.S. Federal Highway Administration Exploratory Advanced Research Program and the Maryland State Highway Administration. The opinions in this paper do not necessarily reflect the official views of NSF, FHWA and SHA. They assume no liability for the content or use of this paper. The authors are responsible for all statements in the paper.

References