

Bayesian Calibration of Traffic Flow Fundamental Diagrams Using Gaussian Processes

ZHANHONG CHENG^{1,2} (Member, IEEE), XUDONG WANG^{1,2}, XINYUAN CHEN³, MARTIN TRÉPANIÉ^{2,4},
AND LIJUN SUN^{1,2} (Senior Member, IEEE)

¹Department of Civil Engineering, McGill University, Montreal, QC H3A 0C3, Canada

²Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Montreal, QC H3T 1J4, Canada

³College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

⁴Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montreal, QC H3T 1J4, Canada

CORRESPONDING AUTHOR: L. SUN (e-mail: lijun.sun@mcgill.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada and in part by the Canada Foundation for Innovation (CFI).

ABSTRACT Modeling the relationship between vehicle speed and density on the road is a fundamental problem in traffic flow theory. Recent research found that using the least-squares (LS) method to calibrate single-regime speed-density models is biased because of the uneven distribution of samples. This paper explains the issue of the LS method from a statistical perspective: the biased calibration is caused by the correlations/dependencies in regression residuals. Based on this explanation, we propose a new calibration method for single-regime speed-density models by modeling the covariance of residuals via a zero-mean Gaussian Process (GP). Our approach can be viewed as a generalized least-squares (GLS) method with a specific covariance structure (i.e., kernel function) and is a generalization of the existing LS and the weighted least-squares (WLS) methods. Next, we use a sparse approximation to address the scalability issue of GPs and apply a Markov chain Monte Carlo (MCMC) sampling scheme to obtain the posterior distributions of the parameters for speed-density models and the hyperparameters (i.e., length scale and variance) of the GP kernel. Finally, we calibrate six well-known single-regime speed-density models with the proposed method. Results show that the proposed GP-based methods (1) significantly reduce the biases in the LS calibration, (2) achieve a similar effect as the WLS method, (3) can be used as a non-parametric speed-density model, and (4) provide a Bayesian solution to estimate posterior distributions of parameters and speed-density functions.

INDEX TERMS Fundamental diagram, Gaussian processes, generalized least-squares, traffic flow theory.

I. INTRODUCTION

THE FUNDAMENTAL diagram is a mathematical representation of speed-density or flow-density relationships, which is one of the underpinnings of traffic flow theory. Numerous works have been dedicated to modeling the relationships since the first fundamental diagram Greenshields model [1] was proposed. Thanks to the simple but efficient formulas, the single-regime speed-density models with a few meaningful parameters [e.g., [2], [3], [4], [5], [6]] have been widely used in various applications, such as determining the

road capacity [7] and modeling traffic control strategies [8]. However, as discussed in Drake et al. [4], Edie [9], single-regime models usually cannot perfectly fit the empirical data ranging from the light-traffic/free-flow regime to the congested/jam regime. Although multi-regime models (piecewise functions consisting of multiple single-regime models) can better fit empirical data, they lack the simplicity and the nice mathematical properties of single regime models (e.g., hard to determine the breakpoints [10] and not differentiable at breakpoints).

Fundamental diagrams are often calibrated by the data from loop detectors. Recent methods also tried to calibrate

The review of this article was arranged by Associate Editor Meng Li.

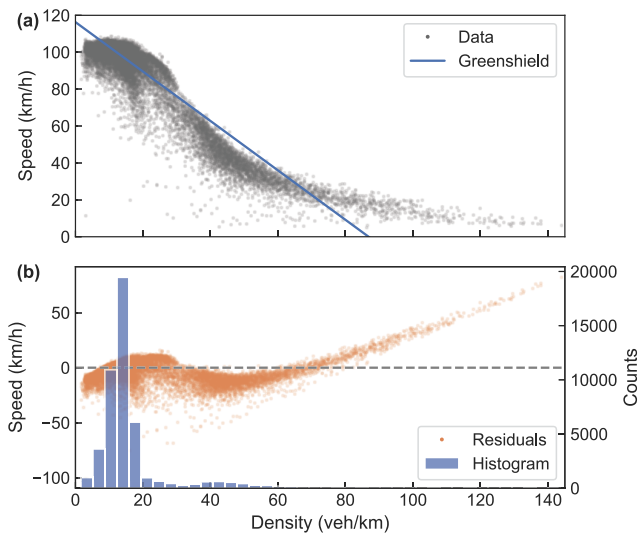


FIGURE 1. (a) Speed-density data and the least-squares fit of the Greenshields model; (b) Regression residuals and the data histogram in terms of density.

fundamental diagrams by floating vehicle data [11], [12]; this paper pertains to the more common approach that uses detector data. Qu et al. [13] pointed out that the calibration method, e.g., the least-squares (LS) method, can cause inaccurate/biased parameter estimations for single-regime models due to the uneven distribution of speed-density observations. We illustrate this issue using a dataset collected by loop detectors from 76 stations on the Georgia State Route 400 (referred to as the GA400 dataset). The average speed, flow, and occupancy are collected for each station with a 20-second sampling interval and aggregated every 5 minutes. The GA400 dataset contains 47,815 observations and has been widely used in studying fundamental diagrams [6], [13], [14], [15], [16]. Figure 1 (a) shows the speed-density observations of the GA400 dataset and a Greenshields model calibrated by the LS method; Figure 1 (b) shows the regression residuals and the data histogram in terms of density. Several issues can be found from the figure. (1) Skewed data: the distribution of the data is highly skewed (this is a general issue to most speed-density datasets); most observations have a small density (the densities of 86.8% of the observations are in the range of being less than 20 veh/km). (2) Biased calibration: the Greenshields model calibrated by the LS method has much smaller absolute values of residuals for regions with more observations than regions with fewer observations. (3) Correlated residuals: there are strong correlations in the residuals.

To address the biased calibration issue, Qu et al. [13] proposed to calibrate single-regime speed-density models by a weighted least-squares (WLS) method. The weights in the WLS are determined by the distance in density between adjacent observations. Therefore, data from regions with fewer observations have larger weights to balance the biases. In a closely related work by Zhang et al. [15], the authors developed a method to generate uniformly distributed

samples from non-uniformly distributed observations and then calibrated the fundamental diagram by the LS method. This method has a similar performance to the WLS method. The above two works both attribute the biased calibration problem to the unevenly distributed data. However, we explain this issue from a statistical perspective and claim that the biased calibration of the LS is caused by the correlations in residuals. The LS method can be derived from the maximum likelihood estimation (MLE), where a fundamental assumption is that residuals are independent and identically distributed (i.i.d.) Gaussian noise. The LS method becomes biased when there are strong correlations in the residuals, such as in Figure 1 (b).

Based on our new explanation, first, this paper proposes a new calibration method for single-regime speed-density models, which reduces the biases by modeling the covariance of residuals using a zero-mean Gaussian Process (GP). Our approach can be viewed as a generalized least-squares (GLS) method with a specific covariance structure and is a generalization of the existing LS and WLS methods. Next, we use a sparse approximation to address the scalability issue in estimating GPs and apply a Markov chain Monte Carlo (MCMC) sampling scheme to obtain the posterior distributions of parameters for speed-density models. Finally, we test our GP-based method by calibrating six well-known single-regime speed-density models on the GA400 dataset. Results show that the proposed methods (1) significantly reduce the biases in the LS calibration, (2) achieve a similar effect as the WLS method, (3) can be used as a non-parametric speed-density model, and (4) provide a Bayesian solution to estimate posterior distributions of parameters and speed-density functions. It is necessary to clarify that the posterior distribution of parameters/speed-density functions in this paper is different from the stochastic fundamental diagram models [14], [16], [17], [18]. The stochastic fundamental diagram aims to use a set of speed-density functions to capture the distribution (uncertainty) of data, while the posterior distribution in this paper captures the distribution (uncertainty) of the estimation for the speed-density function.

Non-i.i.d. noise is a general problem in model calibration. Back in 2001, Kennedy and O'Hagan [19] have proposed to use GP to model the correlations in residuals. Later, similar ideas were used in various fields [20], [21], [22]. In the transportation domain, a recent closely related work by Würth et al. [23] also introduced a GP bias term in calibrating traffic flow models. Würth et al. [23] focus on calibrating second-order traffic flow models, while this paper estimate model parameters for speed-density functions and adopt a scalable algorithm using sparse GP.

There are four major contributions of this paper. First, we provide a statistical explanation for the biases of the LS method in calibrating speed-density models. Second, we propose to use a GP to resolve the biased calibration problem, and our method unites the existing LS and WLS methods into a generalized framework. Third, we introduce two solutions

(sparse GP and MCMC sampling) to estimate the proposed GP-based model, making our method is scalable to large datasets and providing posterior distributions for the estimation. Lastly, the proposed GP regression can be used as a new non-parametric speed-density model.

The remainder of the papers is organized as follows. Section II is the methodology part for the GP-based calibration method for speed-density models; the theory, the connections with existing models, and two solutions for the GP will be introduced in this section. Section III is about experiments, where we will calibrate six single-regime speed-density models by the GP-based methods and compare the results with the LS and the WLS methods. Section IV is left for conclusions and discussions.

II. METHODOLOGY

Consider a list of n observations for speed $\mathbf{v} = [v_1, v_2, \dots, v_n]^\top$ and density $\mathbf{k} = [k_1, k_2, \dots, k_n]^\top$. We assume these paired observations can be explained by an unknown function f and a random noise term

$$v_i = f(k_i) + \varepsilon_i. \quad (1)$$

Conventional approaches directly replace the unknown function with a given-form single-regime speed-density model $m(k)$ and calibrate model parameters by the LS method. However, most single-regime speed-density models are not a perfect fit for the speed-density relationship, and the residuals ($v - m(k)$) are, therefore, not independent. The correlations in regression residuals violate a fundamental assumption— independent noise—in the LS estimation,¹ which causes biases in parameter calibrations, particularly in the case of unevenly distributed observations [13]. Although the generalized least-squares (GLS) method [25] can overcome the biases of the LS method, the estimation for the GLS in our case is difficult because of the unknown residual covariance matrix, non-linear speed-density function, and the large number of observations.

We propose a new calibration method by Gaussian Process (GP) regression to resolve the above issue. Section II-A elaborates on the foundation of the GP-based method and its connections with the existing WLS and LS methods. We further introduce a sparse GP method in Section II-B to resolve the scalability issue in the GP regression. Finally, Section II-C introduces an MCMC sampling method for variational sparse GP [26] to obtain parameters' posteriors in speed-density models.

A. GAUSSIAN PROCESS REGRESSION

We impose a GP prior to the unknown function f :

$$f(k) \sim \mathcal{GP}(m(k), c(k, k')), \quad (2)$$

The LS method can be derived from a maximum likelihood estimation when assuming i.i.d. Gaussian noise [24, Ch. 3]. Although, when the function is linear, according to the Gauss-Markov theorem, the LS estimation is unbiased when the noise is independent, zero-mean, and homoscedastic with finite variance (could be non-Gaussian). The noise should be independent in either case.

$$c(k, k') = \sigma^2 \exp\left(-\frac{(k - k')^2}{2\ell^2}\right), \quad (3)$$

where the mean function $m(k)$ is a speed-density model; the covariance function (a.k.a. kernel function) $c(k, k')$ captures the covariance of the residuals. The function value at observed points follows a multivariate Gaussian distribution $p(\mathbf{f}) = \mathcal{N}(m(\mathbf{k}), \mathbf{C}_{nn})$, where \mathbf{C}_{nn} is a covariance matrix with elements $\mathbf{C}_{nn}[i, j] = c(k_i, k_j)$. The squared exponential (SE) kernel in Eq. (3) is one of the most commonly used kernels in GP (although using other kernels is possible), characterized by smooth function values; the variance σ^2 and the length scale ℓ are hyperparameters that should be estimated from data. To better understand our method, an equivalent form for Eq. (1)–Eq. (2) is

$$v = m(k) + g(k) + \varepsilon, \quad (4)$$

$$g(k) \sim \mathcal{GP}(0, c(k, k')), \quad (5)$$

where $g(k)$ can be regarded as a systematic error component that cannot be modeled by the speed-density function $m(k)$, and we model the systematic error component by a zero-mean GP. To avoid ambiguity, we refer to $(v - m(k))$ the regression residual and the ε the noise thereafter.

Assuming an i.i.d. Gaussian noise term of ε brings great convenience to calibrating a GP. In this case, the joint distribution of n speed observations follows a multivariate Gaussian distribution $p(\mathbf{v}) = \mathcal{N}(m(\mathbf{k}), \mathbf{C}_{nn} + \sigma_\varepsilon^2 \mathbf{I})$, where σ_ε^2 is the variance of ε , and \mathbf{I} is an identity matrix. We denote by $\boldsymbol{\beta}$ the parameters in the mean function and by $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \ell, \sigma^2, \sigma_\varepsilon^2\}$ the parameters of the GP. We can estimate model parameters by Maximum marginal Likelihood Estimation (MLE), which is equivalent to minimizing the following negative log marginal likelihood with respect to $\boldsymbol{\theta}$:

$$\begin{aligned} -\log p(\mathbf{v}|\boldsymbol{\theta}) &= -\log \mathcal{N}(\mathbf{v}|m(\mathbf{k}), \mathbf{C}_{nn} + \sigma_\varepsilon^2 \mathbf{I}) \\ &= \frac{1}{2}(\mathbf{v} - m(\mathbf{k}))^\top (\mathbf{C}_{nn} + \sigma_\varepsilon^2 \mathbf{I})^{-1} (\mathbf{v} - m(\mathbf{k})) \\ &\quad + \frac{1}{2} \log(|\mathbf{C}_{nn} + \sigma_\varepsilon^2 \mathbf{I}|) + \frac{n}{2} \ln(2\pi). \end{aligned} \quad (6)$$

This minimization can be solved numerically by gradient-based methods.

Indeed, the above GP-based calibration can be viewed as a generalized least-squares (GLS) method with a structured residual covariance matrix $\boldsymbol{\Sigma} = (\mathbf{C}_{nn} + \sigma_\varepsilon^2 \mathbf{I})$. The quadratic term $\frac{1}{2}(\mathbf{v} - m)^\top \boldsymbol{\Sigma}^{-1}(\mathbf{v} - m)$ in Eq. (6) plays a role of minimizing the regression residuals in squared Mahalanobis length [27]. When $\mathbf{C}_{nn} = \mathbf{0}$ and $\sigma_\varepsilon = 1$, minimizing Eq. (6) becomes an LS estimation. When $\mathbf{C}_{nn} = \mathbf{0}$ and allowing different σ_ε along the diagonal of $\boldsymbol{\Sigma}$ (i.e., heteroscedasticity), Eq. (6) is equivalent to the WLS method adopted by Qu et al. [13]. In this paper, we use the SE kernel in Eq. (3) to construct \mathbf{C}_{nn} , accounting for the dependencies in the residuals. Meanwhile, the second term $\frac{1}{2} \log(|\boldsymbol{\Sigma}|)$ in Eq. (6) penalizes the complexity of the covariance matrix, leaving spaces for the mean function to explain more variance of

the data. Therefore, the GP-based calibration is a generalization of a wide range of calibration methods with different covariance structures. Besides, although the main purpose of this paper is to calibrate a speed-density model, the GP itself can function as a non-parametric model for data-driven fundamental diagrams.

B. SPARSE GP

The GP regression introduced in Section II-A does not scale to large datasets because of the $\mathcal{O}(n^3)$ time complexity and the $\mathcal{O}(n^2)$ storage complexity. Fortunately, various scalable GP methods have been developed to overcome this difficulty [28]. We applied a sparse GP method using inducing points to address the computational issue.

The computational bottleneck of a GP is calculating the inverse and the determinant of the covariance matrix Σ . The idea of the sparse GP is to approximate the covariance matrix using a small set of u auxiliary inducing points at \mathbf{k}_u . The function values at these inducing points follow the same GP prior $p(\mathbf{f}_u) = \mathcal{N}(m(\mathbf{k}_u), \mathbf{C}_{uu})$, and it assumes that the function values \mathbf{f} at observed points and \mathbf{f}_* at new locations are conditionally independent given \mathbf{f}_u , i.e., $p(\mathbf{f}_* | \mathbf{f}, \mathbf{f}_u) = p(\mathbf{f}_* | \mathbf{f}_u)$. Using inducing points, the function covariance matrix is approximated with a low-rank representation $\mathbf{C}_{nn} \approx \mathbf{C}_{nu} \mathbf{C}_{uu}^{-1} \mathbf{C}_{nu}^\top$, where $\mathbf{C}_{nu} \in \mathbb{R}^{n \times u}$ is the covariance matrix between \mathbf{f} and \mathbf{f}_u . Next, the inverse and the determinant in Eq. (6) can be simplified by the matrix inversion lemma (also known as the Sherman-Morrison-Woodbury formula) and the matrix determinant lemma, respectively; it follows

$$\begin{aligned} & (\mathbf{C}_{nu} \mathbf{C}_{uu}^{-1} \mathbf{C}_{nu}^\top + \sigma_\varepsilon^2 \mathbf{I})^{-1} \\ &= \sigma_\varepsilon^{-2} \mathbf{I} - \sigma_\varepsilon^{-4} \mathbf{C}_{nu} (\sigma_\varepsilon^{-2} \mathbf{C}_{nu}^\top \mathbf{C}_{nu} + \mathbf{C}_{uu})^{-1} \mathbf{C}_{nu}^\top, \quad (7) \end{aligned}$$

$$\begin{aligned} & |\mathbf{C}_{nu} \mathbf{C}_{uu}^{-1} \mathbf{C}_{nu}^\top + \sigma_\varepsilon^2 \mathbf{I}| \\ &= |\sigma_\varepsilon^{-2} \mathbf{C}_{nu}^\top \mathbf{C}_{nu} + \mathbf{C}_{uu}| |\mathbf{C}_{uu}^{-1}| \sigma_\varepsilon^{2n}. \quad (8) \end{aligned}$$

With Eq. (7)–Eq. (8), the time and storage complexity of estimating a sparse GP using the MLE reduces to $\mathcal{O}(nu^2)$.

Although there are methods for choosing the number and the locations for inducing points, we use evenly spaced twenty points between the minimum and the maximum observed density as inducing points. This simple choice for inducing points will suffice for our problem since the shape of the speed-density relationship is pretty simple. Besides, because this paper focuses on the mean function $m(k)$ rather than the function value $f(k)$, many other sparse GP methods in the literature will have the same estimation results for our problem. For example, the Nyström GP [29], the subset of regressors [30], and the deterministic training conditional approximation [31], [32]. Interested readers are referred to an article by Quinonero-Candela and Rasmussen [33] for the details of these sparse GP models.

C. MCMC FOR VARIATIONAL SPARSE GP

The estimation methods in Section II-A and Section II-B only works when the noise term ε follows a Gaussian

distribution. However, as we can see from Figure 1 (a), there are a fair amount of outliers in the data and using a long-tailed distribution (e.g., Student- t distribution) for the noise term is more robust in such a case. Moreover, traffic practitioners often have prior knowledge about model parameters (e.g., free flow speed and jam density); knowing the uncertainty (posterior distribution) of parameters is much more interesting than a point estimation. Therefore, a Bayesian estimation based on an MCMC sampling method developed by Hensman et al. [26] for variational sparse GP is introduced in this section to resolve the above issues.

The sparse GP in Section II-B uses approximate GP prior (in the covariance matrix) but exact inference. In contrast, the variational sparse GP [26], [34] uses exact GP prior but approximate inference. Following the notations in Section II-B, unknown values in a sparse GP include \mathbf{f} , \mathbf{f}_u , \mathbf{f}_* , and θ . The idea of the variational sparse GP by Hensman et al. [26] is to approximate the intractable true posterior $p(\mathbf{f}, \mathbf{f}_u, \mathbf{f}_*, \theta | \mathbf{y})$ by another distribution $q(\mathbf{f}, \mathbf{f}_u, \mathbf{f}_*, \theta)$ by minimizing their Kullback-Leibler (KL) divergence $\text{KL}(q(\mathbf{f}, \mathbf{f}_u, \mathbf{f}_*, \theta) \| p(\mathbf{f}, \mathbf{f}_u, \mathbf{f}_*, \theta | \mathbf{y}))$. Hensman et al. [26] found that the minimum KL divergence is reached with an optimal variational distribution:

$$\begin{aligned} \log \hat{q}(\mathbf{f}_u, \theta) &= \mathbb{E}_{p(\mathbf{f} | \mathbf{f}_u, \theta)} [\log p(\mathbf{y} | \mathbf{f})] \\ &+ \log p(\mathbf{f}_u | \theta) + \log p(\theta) - \log C, \quad (9) \end{aligned}$$

where $p(\theta)$ is the prior distribution of parameters, C is a normalizing constant. The first term on the right side of Eq. (9) is the expected log-likelihood for noise ε . This expectation can be factorized across data when assuming i.i.d. noise; next, this term can be analytically solved for Gaussian or Poisson noise, and can be approximated by Gauss-Hermite quadrature for other types of noise. The time complexity for computing Eq. (9) is $\mathcal{O}(nu^2)$.

Direct sampling from Eq. (9) using the MCMC is inefficient because the function value \mathbf{f}_u and GP parameters θ are usually strongly coupled [35]. Therefore, the following reparameterization (known as whitening the prior) is introduced to improve the sampling efficiency. Define an ancillary random variable $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, such that $\mathbf{f}_u = \mathbf{R}\mathbf{v}$ with $\mathbf{R}\mathbf{R}^\top = \mathbf{C}_{uu}$. The optimal log probability density for \mathbf{v} and θ is thus

$$\begin{aligned} \log \hat{q}(\mathbf{v}, \theta) &= \mathbb{E}_{p(\mathbf{f} | \mathbf{f}_u = \mathbf{R}\mathbf{v})} [\log p(\mathbf{y} | \mathbf{f})] \\ &+ \log p(\mathbf{v}) + \log p(\theta) - \log C. \quad (10) \end{aligned}$$

We apply No-U-Turn sampler [36], a type of Hamiltonian MCMC [37] method that automatically tunes step size, to efficiently draw samples from Eq. (10). Next, parameters' posterior distributions can be obtained from the samples.

III. TESTS

This section tests the effect of using the GP in calibrating single regime speed-density models. Section III-A introduces the settings of the experiment and models. The calibration results of different models are compared in Section III-B.

TABLE 1. Single-regime speed-density models.

Model	Function	Parameters
Greenshields [1]	$v = v_f (1 - \frac{k}{k_j})$	v_f, k_j
Greenberg [2]	$v = v_0 \ln(\frac{k_j}{k})$	v_0, k_j
Underwood [3]	$v = v_f \exp(-\frac{k}{k_0})$	v_f, k_0
Northwestern [4]	$v = v_f \exp(-\frac{1}{2}(\frac{k}{k_0})^2)$	v_f, k_0
Newell [5]	$v = v_f (1 - \exp(-\frac{\lambda}{v_f}(\frac{1}{k} - \frac{1}{k_j})))$	v_f, k_j, λ
Three-parameter logistic (3PL) [6]	$v = \frac{v_f}{1 + \exp(k - k_c/\theta)}$	v_f, k_c, θ

Section III-C will show the posterior distributions obtained by the MCMC solution for our GP method. Section III-D discusses how prior knowledge improves the model calibration under a different sample size.

A. EXPERIMENT SETTINGS

We calibrate the six speed-density models in Table 1 on the GA400 dataset using the GP regression. The GA400 dataset is a representative dataset that encompasses a wide range of traffic flow characteristics (e.g., skewed-distributed samples, outliers, traffic hysteresis, and sufficient sample points) and has been extensively used in previous research [13], [14], [15], [16]. The calibration results of the proposed method are compared with the WLS [13] and the LS methods. We use the sparse approximation for the MLE because of the large number of observations (47,581 points). As mentioned in Section II-B, twenty inducing points are evenly spaced from minimum to maximum observed densities.

We also test using the MCMC sampling for variational sparse GP to obtain the posterior distributions of model parameters. To make the estimation more robust to the outliers in the data, we assume the noise term ε follows an i.i.d. zero-mean Student- t distribution. The choice of priors in the MCMC sampling are as follows:

- Gaussian priors are used for all parameters regarding the speed-density functions. For a parameter x , its prior mean is set as the WLS estimation x_{WLS} , and the standard deviation is set by $\max(\frac{x_{WLS}}{6}, 10)$.
- Other hyperparameters are positive values, including σ^2 and ℓ in the kernel functions and scale and degree of freedom for the Student- t noise. Therefore, we set their prior by Half-Cauchy distribution $p(x) = \frac{2}{\pi(1+x^2)}$, $x \in [0, +\infty)$.

We use a Python package named GPflow [38] to set up the variational sparse GP introduced in Section II-C. Next, the No-U-Turn sampling routine implemented in TensorFlow Probability [39] is used to draw samples from Eq. (10). The first 2,000 steps are taken as the “burn-in” period to tune the step size and converge the sequence; we collect the next 3,000 steps to calculate posterior distributions.

B. COMPARISON WITH OTHER MODELS

The six speed-density models calibrated by the four methods are compared in Figure 2, where the result of the GP-MCMC

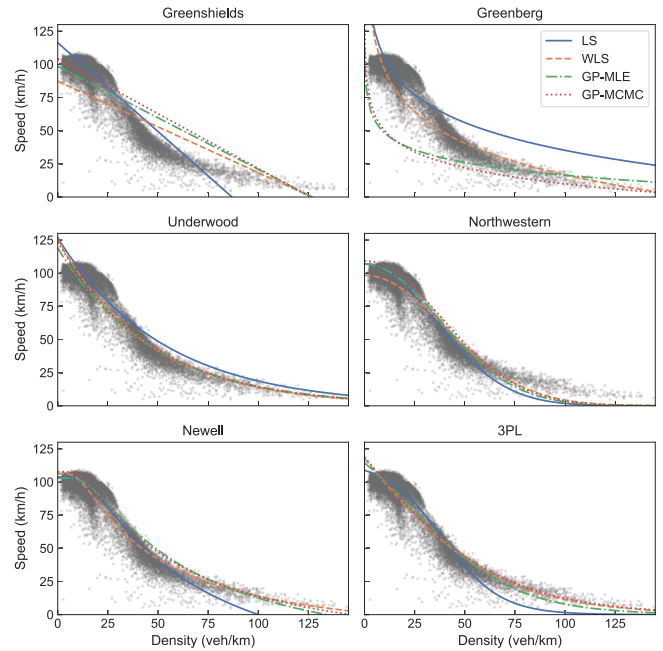


FIGURE 2. Speed-density models calibrated by LS, WLS, sparse GP by MLE, and the mean of MCMC sampling for variational sparse GP.

is the average of posterior samples. We can see that the LS method results in speed-density models that fit the low-density region (≤ 50 veh/km) well but are inappropriate for the high-density region (> 50 veh/km). In contrast, the WLS and the two GP methods generally produce more “unbiased” speed-density models in the entire density range. Two GP methods have very similar calibration results to the WLS method for the Underwood, the Newell, and the 3PL speed-density models. For the Greenshields and the Northwestern speed-density model, the free flow speed (v at $k = 0$) estimated by GP methods is higher than which estimated by the WLS method.

An exception for the GP method is the Greenberg speed-density model, where the mean function does not fit the data well. This is caused by a limitation of the Greenberg model: $v \rightarrow +\infty$ when $k \rightarrow +0$. A Greenberg model that fits most observations well (e.g., the WLS estimation) has a significantly higher estimation for speed at the near zero-density area; this “abrupt surge in the regression residual” is very unlikely to happen in the GPs because the SE kernel that we use favors a smooth function. Therefore, a GP-based estimation produces a flat mean function $m(k)$ and a significant systematic error component $g(k)$ in the Greenberg model. The GP essentially uses a different way than the other methods to reveal the inappropriateness of the Greenberg model. Many measures can be taken to improve the GP estimation for the Greenberg model, including (1) adding a small positive shift parameter k_s , so $\lim_{k \rightarrow 0^+} v_0 \ln(\frac{k_j}{k+k_s})$ is not infinity, (2) fitting the inverse function $k = f^{-1}(v)$, (3) using other kernels in the GP (e.g., non-stationary kernels), and (4) discarding data with small densities.

TABLE 2. Kernel hyperparameters estimated by the MLE and the MCMC methods.

Method	Hyperparameters	Greenshields	Greenberg	Underwood	Northwestern	Newell	3PL
MLE	Kernel length scale ℓ	11.27	11.16	9.42	10.36	10.31	10.12
	Kernel variance σ^2	84.26	89.32	24.62	34.01	24.11	38.06
MCMC	Kernel length scale ℓ	12.23	12.97	10.37	10.85	10.75	9.91
	Kernel variance σ^2	188.54	363.03	33.72	34.90	20.14	19.27

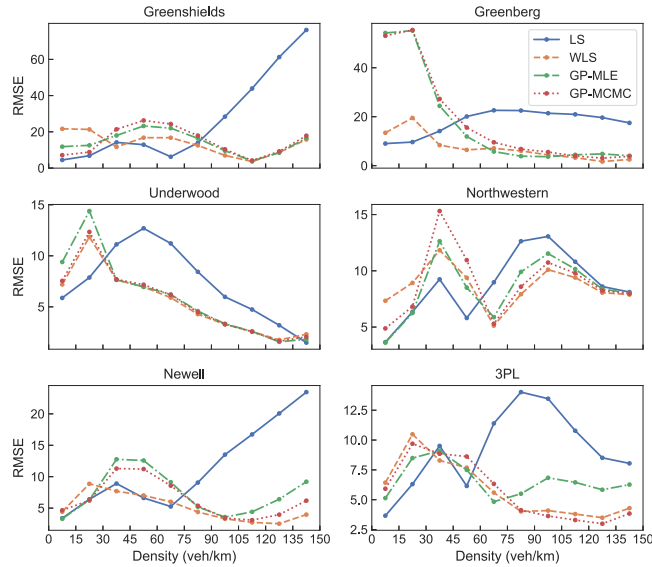


FIGURE 3. The RMSE of the four calibration methods on the six single-regime models at different density groups.

In Figure 3, we divide data into ten groups by density from $k = 0$ to $k = 150$ at an interval of 15 to better understand the fitting performance. We calculate the root-mean-square error (RMSE) for each density group between speed-density models and data. We can find from Figure 3 that the WLS and the two GP methods alleviate the biased calibration of the LS method and greatly reduce the RMSE for high-density (around $k > 75$) groups. Except for the Greenberg model, it is hard to conclude which one between the GP and WLS method has a better calibration result. Besides, the RMSE of the GP using the MCMC sampling is a little lower than the MLE methods for the Newell and the 3PL models, which could be a benefit from assuming the Student- t noise and using the priors.

The main purpose of this paper is to calibrate parameters in speed-density models, but it is also interesting to see how well a GP can fit the data. Figure 4 shows the function values ($f(k)$, or equivalently $m(k) + g(k)$) fitted by the MLE of GPs under different speed-density mean functions. We can see the function values are almost identical under different mean functions (except for the two edge parts). More importantly, GPs fit the empirical data much better than their mean functions, which means we can use the GP as a non-parametric fundamental diagram. Although it lacks the simplicity and the mathematical properties of single-regime models, the non-parametric fundamental diagram is

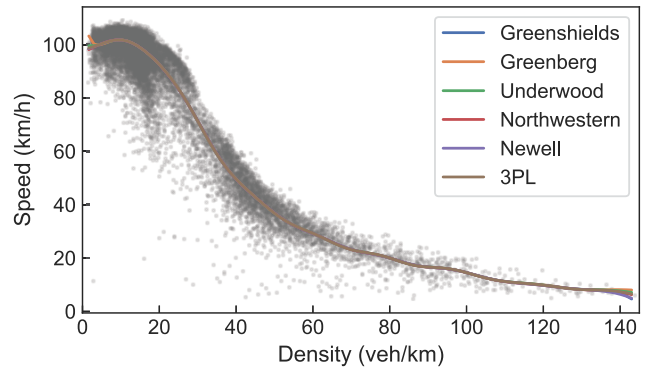


FIGURE 4. The function values $f(k)$ fitted by the MLE of sparse GP under different speed-density mean functions.

more close to empirical data. Moreover, we can use the GP confidence intervals to screen outliers and use the remaining observations to better calibrate speed-density models.

Moreover, hyperparameters in the SE kernel have practical implications and can assist in diagnosing the goodness of fit of a mean function. The length scale ℓ determines the length of the “waves” in the residuals; a large ℓ indicates long-distance correlations in the residuals, and a small ℓ means locally correlated residuals. The kernel variance σ^2 determines the average distance of the function f away from its mean; a smaller σ^2 indicates better goodness of fit. Table 2 shows the hyperparameters estimated by the MLE and the MCMC method, where the values of the MCMC method are sample means. The Greenshields and the Greenberg model have the largest kernel variance and length scale, indicating poor fitness. In contrast, the 3PL model estimated by the MCMC has the smallest kernel length scale and variance, meaning the best fitting result. The conclusions drawn from the hyperparameters are consistent with the observations in Figure 2 and Figure 3.

C. POSTERiors OF ESTIMATIONS

We use the Underwood model as an example to show samples drawn by the MCMC. The steady traces in Figure 5 (a) and (b) indicate the MCMC sampling has converged. The top and right panel in Figure 5 are empirical distributions of v_f and k_0 , respectively. The gray dots in Figure 5 (c) are samples and the contour is the joint distribution of v_f and k_0 . We can find that samples of v_f and k_0 are negatively correlated. We can also obtain the credible intervals for each parameter using the samples’ distribution. The MCMC sampling helps to understand the uncertainty of the estimation.

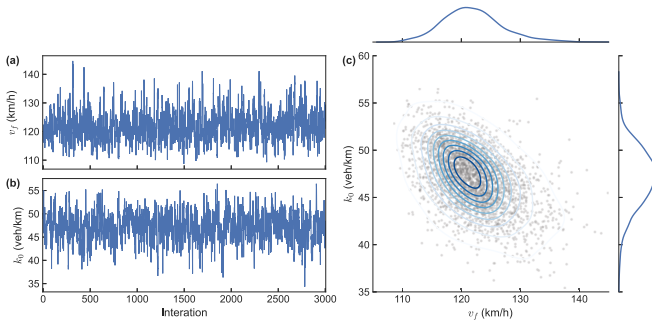


FIGURE 5. MCMC sampling traces and the posterior distributions of parameters for the Underwood model.

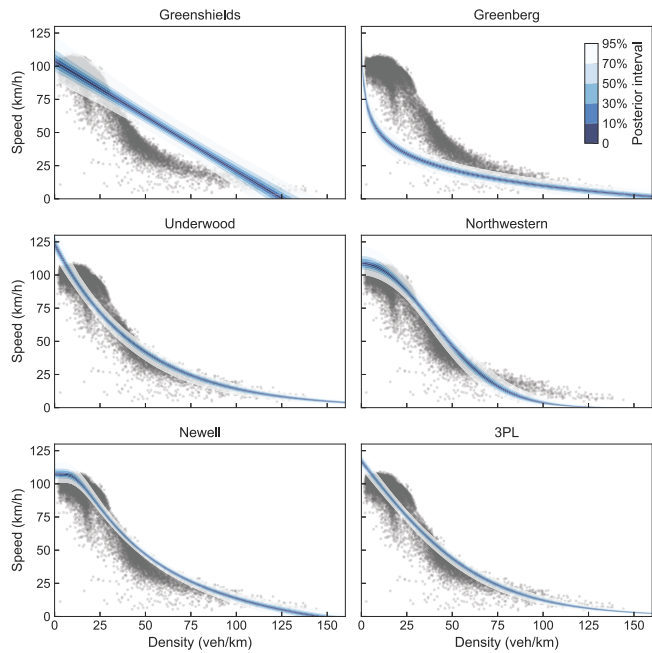


FIGURE 6. Posterior distributions of speed-density functions.

A more intuitive way to examine the uncertainty in the calibration is to visualize the posterior distributions of speed-density functions. Figure 6 shows the posterior distributions of speed-density functions by different equal-tailed intervals [40, ETI]. For example, a 95% ETI has 2.5% of the samples on either side of this interval. We can see the posterior of the Greenshields models has the widest ETI, meaning a large uncertainty. In contrast, the Newell and the 3PL models, both with three parameters, have relatively narrow ETIs. The ETI of a model has varying widths at different density levels, such as the free flow region of the Newell model has a larger ETI than the other part of the curve. In addition to the speed-density function, we can also examine the posterior distribution of the flow-density function (obtained by $k \times v$). From Figure 7, we can find that the estimation of 3PL flow-density function has relatively good fitness and narrow ETI. Overall, examining posterior distributions in Figure 6 and Figure 7 helps understand the calibration uncertainty and

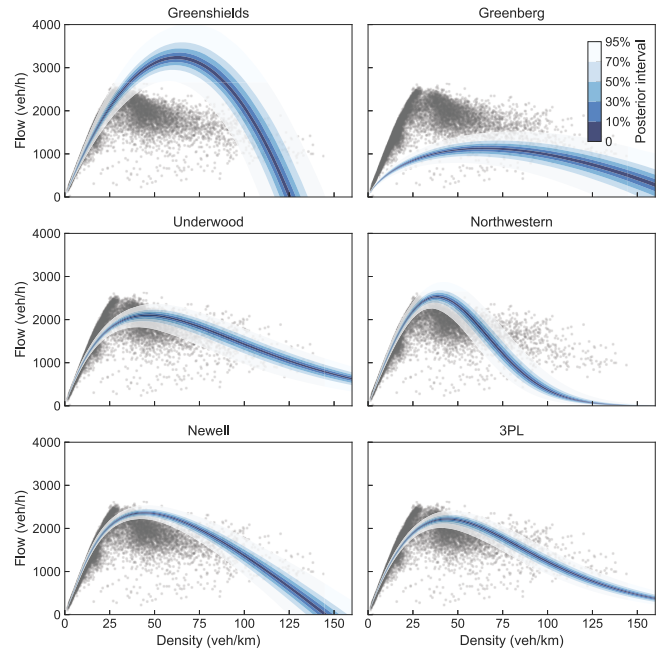


FIGURE 7. Posterior distributions of flow-density functions.

choose among different speed-density models, particularly when there are insufficient observations.

D. IMPACTS OF PRIORS AND SAMPLE SIZE

Transportation practitioners usually have strong prior knowledge of traffic flow parameters, such as the range of free flow speed and jam density. In a Bayesian setting, prior knowledge plays an essential role in improving model calibration, especially in the case of small sample size and skewed sample distribution. In fact, data points with jam density are very rare in the GA400 dataset, and it is, therefore, difficult to quantify the goodness of jam density estimation with limited observations. Using an appropriate prior distribution can alleviate this problem.

We randomly sample 100, 500, and 5000 data points from the GA400 dataset to test how the model calibration changes with different sample sizes. Assume the prior distributions for the free flow speed and the jam density are $v_f \sim \mathcal{N}(100, 15^2)$ and $k_j \sim \mathcal{N}(120, 15^2)$; these are very broad ranges for parameters. The parameter calibration results for the Greenshields model are shown in Figure 8.

From Figure 8, we can see the WLS and the GP-MLE are likely to produce similar estimations as the LS method when the sample size is small (e.g., 100). The estimations of the WLS and the GP-MLE methods vary significantly with the increase of the sample size, and relative change in k_j even surpasses 50% of its value; this is because there are few data points with a high density and the parameter estimation is very sensitive to these observations. In contrast, the GP-MCMC using the prior distribution produces more consistent posterior estimations of v_f and k_j over different sample sizes. Interestingly, the posterior distribution of v_f

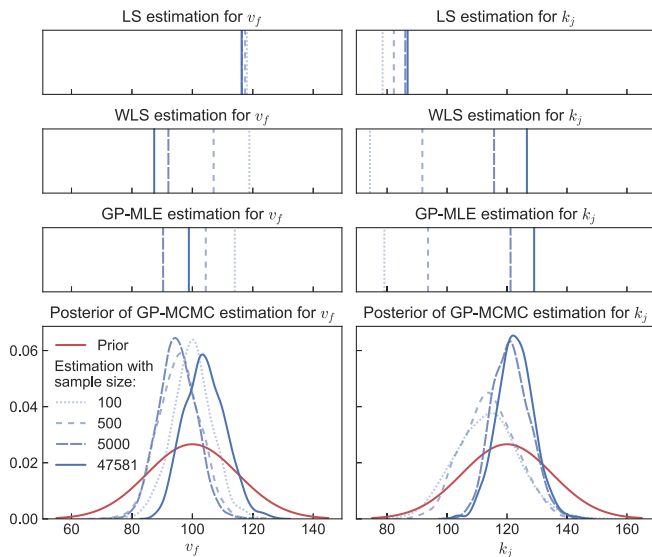


FIGURE 8. Parameters of the Greenshields model calibrated by the four methods under different sample sizes.

does not change much from small to large sample size, meaning even 100 data points already give a reasonably good estimation of the free flow speed. The peak of the posterior k_j becomes significantly higher with the increase of the sample size, indicating a good estimation for the jam density requires a large sample size. Overall, the example in Figure 8 demonstrate the importance of prior knowledge and posterior uncertainty quantification, especially when the sample size is small.

IV. CONCLUSION AND DISCUSSIONS

The speed-density relationship is an essential concept in traffic flow theory. This article proposes a new low-biased calibration method for single-regime speed-density functions by modeling the residual dependencies using a zero-mean GP. Our approach based on GP has a solid statistical explanation and is a generalization of the existing WLS and LS methods. Experiments show that GP-based methods significantly reduce the biases in the LS calibration and achieve a similar effect as the WLS method. The GP can also be used as a non-parametric way to model speed-density relationships. Moreover, we put the GP regression in a Bayesian setting, leveraging different prior distributions and noise assumptions, and using an MCMC sampling algorithm to get the posterior distributions of parameters. Besides the problem in this paper, the GP-based method can also be used to calibrate other more general problems, such as macroscopic fundamental diagrams [41], [42] and speed-crash severity relationship [43]. Particularly, our method is beneficial for calibration problems with correlated residuals, biased sample distributions, and when knowing parameters' uncertainties is essential (e.g., too few samples).

This research has some limitations and spaces for improvement. First, this paper only considers single-regime fundamental diagrams. Multi-regime models are usually not

differentiable at the breakpoints, which would be inappropriate to use a GP with the SE kernel to model its residual correlations. Second, we assume i.i.d. noise but the data could be heteroscedastic. An improvement can be made by using GP with heteroscedastic noise [e.g., [44], [45]]. Lastly, we can use GP with non-stationary kernels [46] to better model the residuals in some extreme cases (e.g., the Greenberg model).

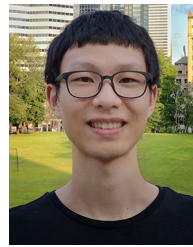
ACKNOWLEDGMENT

X. Wang would like to thank FRQNT for providing the B2X Doctoral Scholarship.

REFERENCES

- [1] B. Greenshields, J. Bibbins, W. Channing, and H. Miller, "A study of traffic capacity," in *Proc. Highway Res. Board*, vol. 14, 1935, pp. 448–477.
- [2] H. Greenberg, "An analysis of traffic flow," *Operat. Res.*, vol. 7, no. 1, pp. 79–85, 1959.
- [3] R. T. Underwood, "Speed, volume, and density relationships," in *Proc. Qual. Theory Traffic Flow. A Symp. Yale Univ. Bur. Highway Traffic*, 1961, pp. 141–188.
- [4] J. Drake, J. Schofer, and A. May, "A statistical analysis of speed-density hypotheses," in *Proc. Traffic Flow Transp.*, 1965, pp. 53–87.
- [5] G. F. Newell, "Nonlinear effects in the dynamics of car following," *Operat. Res.*, vol. 9, no. 2, pp. 209–229, 1961.
- [6] H. Wang, J. Li, Q.-Y. Chen, and D. Ni, "Logistic modeling of the equilibrium speed–density relationship," *Transp. Res. A Policy Pract.*, vol. 45, no. 6, pp. 554–566, 2011.
- [7] N. Wu and H. Rakha, "Derivation of van aerde traffic stream model from tandem-queueing theory," *Transp. Res. Rec.*, vol. 2124, no. 1, pp. 18–27, 2009.
- [8] Y. Wang, E. B. Kosmatopoulos, M. Papageorgiou, and I. Papamichail, "Local ramp metering in the presence of a distant downstream bottleneck: Theoretical analysis and simulation study," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2024–2039, Oct. 2014.
- [9] L. C. Edie, "Car-following and steady-state theory for noncongested traffic," *Operat. Res.*, vol. 9, no. 1, pp. 66–76, 1961.
- [10] E. Kidando et al., "Novel approach for calibrating freeway highway multi-regimes fundamental diagram," *Transp. Res. Rec.*, vol. 2674, no. 9, pp. 561–574, 2020.
- [11] T. Seo, Y. Kawasaki, T. Kusakabe, and Y. Asakura, "Fundamental diagram estimation by using trajectories of probe vehicles," *Transp. Res. B Methodol.*, vol. 122, pp. 40–56, Apr. 2019.
- [12] R. Yin, N. Zheng, and Z. Liu, "Estimating fundamental diagram for multi-modal signalized urban links with limited probe data," *Physica A Stat. Mechan. Appl.*, vol. 606, Nov. 2022, Art. no. 128091.
- [13] X. Qu, S. Wang, and J. Zhang, "On the fundamental diagram for freeway traffic: A novel calibration approach for single-regime models," *Transp. Res. B Methodol.*, vol. 73, pp. 91–102, Mar. 2015.
- [14] X. Qu, J. Zhang, and S. Wang, "On the stochastic fundamental diagram for freeway traffic: Model development, analytical properties, validation, and extensive applications," *Transp. Res. B methodol.*, vol. 104, pp. 256–271, Oct. 2017.
- [15] J. Zhang, X. Qu, and S. Wang, "Reproducible generation of experimental data sample for calibrating traffic flow fundamental diagram," *Transp. Res. A Policy Pract.*, vol. 111, pp. 41–52, May 2018.
- [16] S. Wang, X. Chen, and X. Qu, "Model on empirically calibrating stochastic traffic flow fundamental diagram," *Commun. Transp. Res.*, vol. 1, Dec. 2021, Art. no. 100015.
- [17] A. Muralidharan, G. Dervisoglu, and R. Horowitz, "Probabilistic graphical models of fundamental diagram parameters for simulations of freeway traffic," *Transp. Res. Rec.*, vol. 2249, no. 1, pp. 78–85, 2011.
- [18] S. E. Jabari and H. X. Liu, "A stochastic model of traffic flow: Gaussian approximation and estimation," *Transp. Res. B Methodol.*, vol. 47, pp. 15–41, Jan. 2013.
- [19] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *J. Roy. Stat. Soc. B*, vol. 63, no. 3, pp. 425–464, 2001.

- [20] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafo, and R. D. Ryne, "Combining field data and computer simulations for calibration and prediction," *SIAM J. Sci. Comput.*, vol. 26, no. 2, pp. 448–466, 2004.
- [21] M. J. Bayarri et al., "Predicting vehicle crashworthiness: Validation of computer models for functional and hierarchical data," *J. Amer. Stat. Assoc.*, vol. 104, no. 487, pp. 929–943, 2009.
- [22] M. Carmassi, P. Barbillon, M. Chiodetti, M. Keller, and É. Parent, "Bayesian calibration of a numerical code for prediction," *J. de la société française de statistique*, vol. 160, no. 1, pp. 1–30, 2019.
- [23] A. Würth, M. Binois, P. Goatin, and S. Göttlich, "Data driven uncertainty quantification in macroscopic traffic flow models," Nov. 2021. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-03202124>
- [24] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
- [25] A. C. Aitken, "IV.—On least squares and linear combination of observations," in *Proc. Roy. Soc. Edinburgh*, vol. 55, 1936, pp. 42–48.
- [26] J. Hensman, A. G. Matthews, M. Filippone, and Z. Ghahramani, "MCMC for variationally sparse gaussian processes," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1648–1656.
- [27] P. C. Mahalanobis, "On the generalized distance in statistics," in *Proc. Nat. Inst. Sci.*, vol. 2, no. 1, 1936, pp. 49–55.
- [28] H. Liu, Y.-S. Ong, X. Shen, and J. Cai, "When gaussian process meets big data: A review of scalable GPs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4405–4423, Nov. 2020.
- [29] C. Williams and M. Seeger, "Using the Nyström method to speed up kernel machines," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 13, 2000, pp. 682–688.
- [30] A. Smola and P. Bartlett, "Sparse greedy gaussian process regression," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 13, 2000, pp. 619–625.
- [31] L. Csató and M. Opper, "Sparse on-line gaussian processes," *Neural Comput.*, vol. 14, no. 3, pp. 641–668, 2002.
- [32] M. W. Seeger, C. K. Williams, and N. D. Lawrence, "Fast forward selection to speed up sparse gaussian process regression," in *Proc. Int. Workshop Artif. Intell. Statist.*, 2003, pp. 254–261.
- [33] J. Quinero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *J. Mach. Learn. Res.*, vol. 6, no. 65, pp. 1939–1959, 2005.
- [34] M. Titsias, "Variational learning of inducing variables in sparse gaussian processes," in *Proc. Artif. Intell. Statist.*, 2009, pp. 567–574.
- [35] I. Murray and R. P. Adams, "Slice sampling covariance hyperparameters of latent gaussian models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 23, 2010, pp. 1732–1740.
- [36] M. D. Hoffman, A. Gelman et al., "The no-U-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1593–1623, 2014.
- [37] M. Betancourt, "A conceptual introduction to hamiltonian Monte Carlo," 2017, *arXiv:1701.02434*.
- [38] A. G. d. G. Matthews et al., "GPflow: A gaussian process library using TensorFlow," *J. Mach. Learn. Res.*, vol. 18, no. 40, pp. 1–6, 2017.
- [39] J. V. Dillon et al., "Tensorflow distributions," 2017, *arXiv:1711.10604*.
- [40] D. Makowski, M. S. Ben-Shachar, and D. Lüdtke, "BayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework," *J. Open Source Softw.*, vol. 4, no. 40, p. 1541, 2019.
- [41] C. F. Daganzo, "Urban gridlock: Macroscopic modeling and mitigation approaches," *Transp. Res. B Methodol.*, vol. 41, no. 1, pp. 49–62, 2007.
- [42] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings," *Transp. Res. B Methodol.*, vol. 42, no. 9, pp. 759–770, 2008.
- [43] L. Aarts and I. Van Schagen, "Driving speed and the risk of road crashes: A review," *Accid. Anal. Prevent.*, vol. 38, no. 2, pp. 215–224, 2006.
- [44] K. Kersting, C. Plagemann, P. Pfaff, and W. Burgard, "Most likely heteroscedastic gaussian process regression," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 393–400.
- [45] F. Rodrigues and F. C. Pereira, "Heteroscedastic Gaussian processes for uncertainty modeling in large-scale crowdsourced traffic data," *Transp. Res. C Emerg. Technol.*, vol. 95, pp. 636–651, Oct. 2018.
- [46] M. Heinonen, H. Mannerström, J. Rousu, S. Kaski, and H. Lähdesmäki, "Non-stationary gaussian process regression with hamiltonian monte carlo," in *Proc. Artif. Intell. Statist.*, 2016, pp. 732–740.



ZHANHONG CHENG (Member, IEEE) received the B.S. and M.S. degrees from the Harbin Institute of Technology, Harbin, China, and the Ph.D. degree from McGill University, Montreal, QC, Canada, where he is currently a Postdoctoral Researcher with the Department of Civil Engineering. His research interests include public transportation, travel behavior modeling, spatiotemporal forecasting, and machine learning in transportation.



XUDONG WANG received the M.S. degree from Beihang University, Beijing, China. She is currently pursuing the Ph.D. degree with the Department of Civil Engineering, McGill University, Montreal, QC, Canada. Her research interest is spatiotemporal data analysis including traffic state estimation, traffic pattern discovery, and traffic anomaly detection.



XINYUAN CHEN received the Ph.D. degree from Monash University, Australia. He is a Lecturer with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, China. He worked as a Postdoctoral Fellow with The Hong Kong Polytechnic University. His research interests include urban transport network modeling, parallel computing in transport system analysis, multimodal transport system modeling and optimization, stochastic fundamental diagram, and maritime search and rescue.



MARTIN TRÉPANIÉRIE is a Civil Engineer and a Professor with the Department of Mathematics and Industrial Engineering, École Polytechnique de Montréal, an engineering school affiliated to the Université de Montréal. He is the Cotitular of the Chair in the transformation of transportation. Since January 2019, he has been the Director of the Interuniversity Research Center on Enterprise Networks, Logistics and Transportation. He conducts many research projects about intelligent transport systems, urban mobility, transportation

planning, operations logistics, data mining and processing of large datasets: smart cards, carsharing transactions, bikesharing, and GPS traces. He sits on the boards of directors of CargoM and Jalon Montreal.



LIJUN SUN (Senior Member, IEEE) received the B.S. degree in civil engineering from Tsinghua University, Beijing, China, in 2011, and the Ph.D. degree in civil engineering (transportation) from the National University of Singapore in 2015. He is currently an Assistant Professor with the Department of Civil Engineering, McGill University, Montreal, QC, Canada. His research centers on intelligent transportation systems, machine learning, spatiotemporal modeling, travel behavior, and agent-based simulation.