Bayesian calibration of traffic flow fundamental diagrams using Gaussian processes

Zhanhong Cheng [1] Xudong Wang [1] Xinyuan Chen [2] Martin Trépanier [3] Lijun Sun [1*]

[1]Department of Civil Engineering, McGill University, Montreal, Canada [2]College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[3]Department of Mathematics and Industrial Engineering, Polytechnique Montreal, Montreal, Canada

Introduction

Using the least-squares (LS) regression to calibrate single-regime speed-density models is biased towards the region with more data points, as shown in the example in Figure 1. The biased calibration is caused by the **autocorrelations in the regression residuals** (LS estimation requires residuals to be i.i.d.). Therefore, we propose a better calibration method for single-regime speed-density models by modeling the covariance of residuals via a zero-mean Gaussian Process (GP).

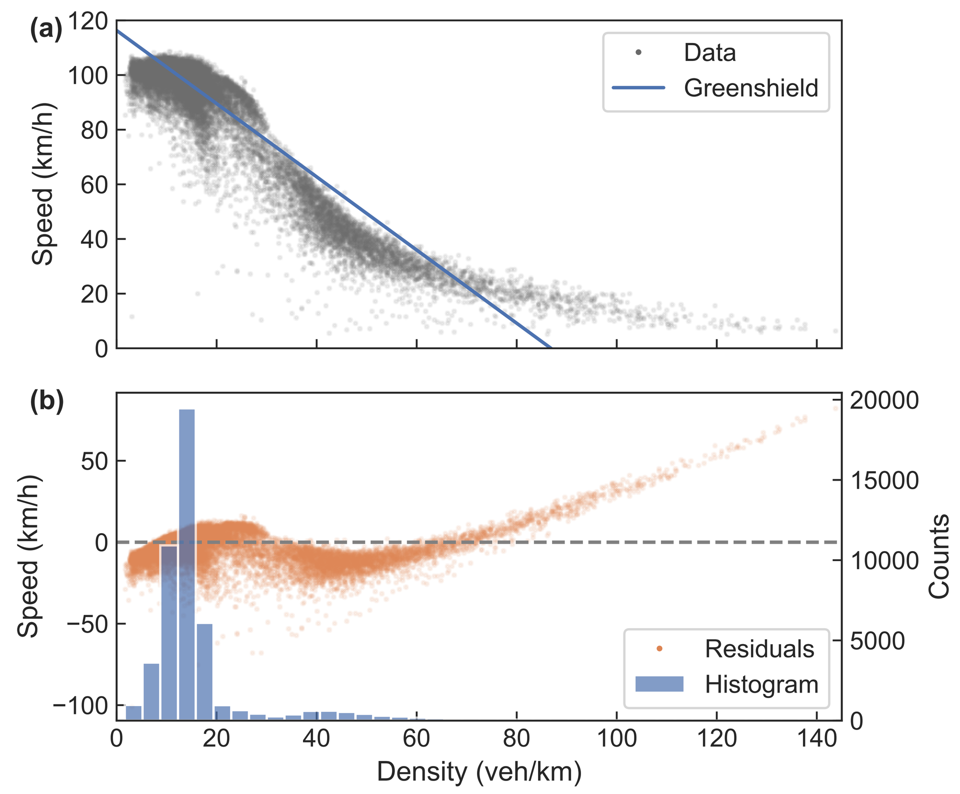


Figure 1. An illustration of the biases in the least-squares regression. (a) Speed-density data and the least-squares fit of the Greenshields model; (b) Regression residuals and the data histogram in terms of density.

Research highlights

- We provide a statistical explanation for the biases when using the least-squares in calibrating speed-density models.
- We propose to use a GP to resolve the biased calibration problem, and our method unites the existing LS and weighted least squares (WLS [2]) methods into a generalized framework.
- 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 (uncertainty quantification) for the estimation.
- The proposed GP regression can be used as a new non-parametric speed-density model.

Gaussian Process regression

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

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

Conventional approaches replace the unknown function with a given-form single-regime speed-density model $m(k)$ and calibrate model parameters by the LS method. However, as show in Figure 1, the correlations in regression residuals violate a fundamental assumption—independent noise—in the LS estimation, causing biases in parameter calibrations. Therefore, we impose a GP prior to the unknown function f:

$$f(k) \sim \mathcal{GP}\left(m(k), c\left(k, k'\right)\right),$$
$$c(k, k') = \sigma^2 \exp\left(-\frac{(k-k')^2}{2\ell^2}\right),$$

where the mean function $m(k)$ is a speed-density model; the covariance function (a.k.a. kernel function) $c\left(k, k'\right)$ captures the covariance of the residuals.

Parameter estimation

Maximum marginal Likelihood Estimation (MLE)

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. The Maximum marginal Likelihood Estimation (MLE) is equivalent to minimizing the following negative log marginal likelihood with respect to $\boldsymbol{\theta}$:

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

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

Sparse GP for large-scale problems

The GP regression requires $\mathcal{O}(n^3)$ time complexity and the $\mathcal{O}(n^2)$ storage complexity. Therefore, we use the sparse GP that approximates the covariance matrix using a small set of u auxiliary inducing points, the function covariance matrix is approximated with a low-rank representation $\mathbf{C}_{nn} \approx \mathbf{C}_{nu}\mathbf{C}_{uu}^{-1}\mathbf{C}_{nu}^\top$, and the inverse and the determinant in Eq. (1) can be simplified:

$$\begin{aligned} \left(\mathbf{C}_{nu}\mathbf{C}_{uu}^{-1}\mathbf{C}_{nu}^\top + \sigma_\varepsilon^2 \mathbf{I}\right)^{-1} &= \sigma_\varepsilon^{-2}\mathbf{I} - \sigma_\varepsilon^{-4}\mathbf{C}_{nu}\left(\sigma_\varepsilon^{-2}\mathbf{C}_{nu}^\top \mathbf{C}_{nu} + \mathbf{C}_{uu}\right)^{-1}\mathbf{C}_{nu}^\top, \\ |\mathbf{C}_{nu}\mathbf{C}_{uu}^{-1}\mathbf{C}_{nu}^\top + \sigma_\varepsilon^2 \mathbf{I}| &= |\sigma_\varepsilon^{-2}\mathbf{C}_{nu}^\top \mathbf{C}_{nu} + \mathbf{C}_{uu}||\mathbf{C}_{uu}^{-1}|\sigma_\varepsilon^{2n}. \end{aligned}$$

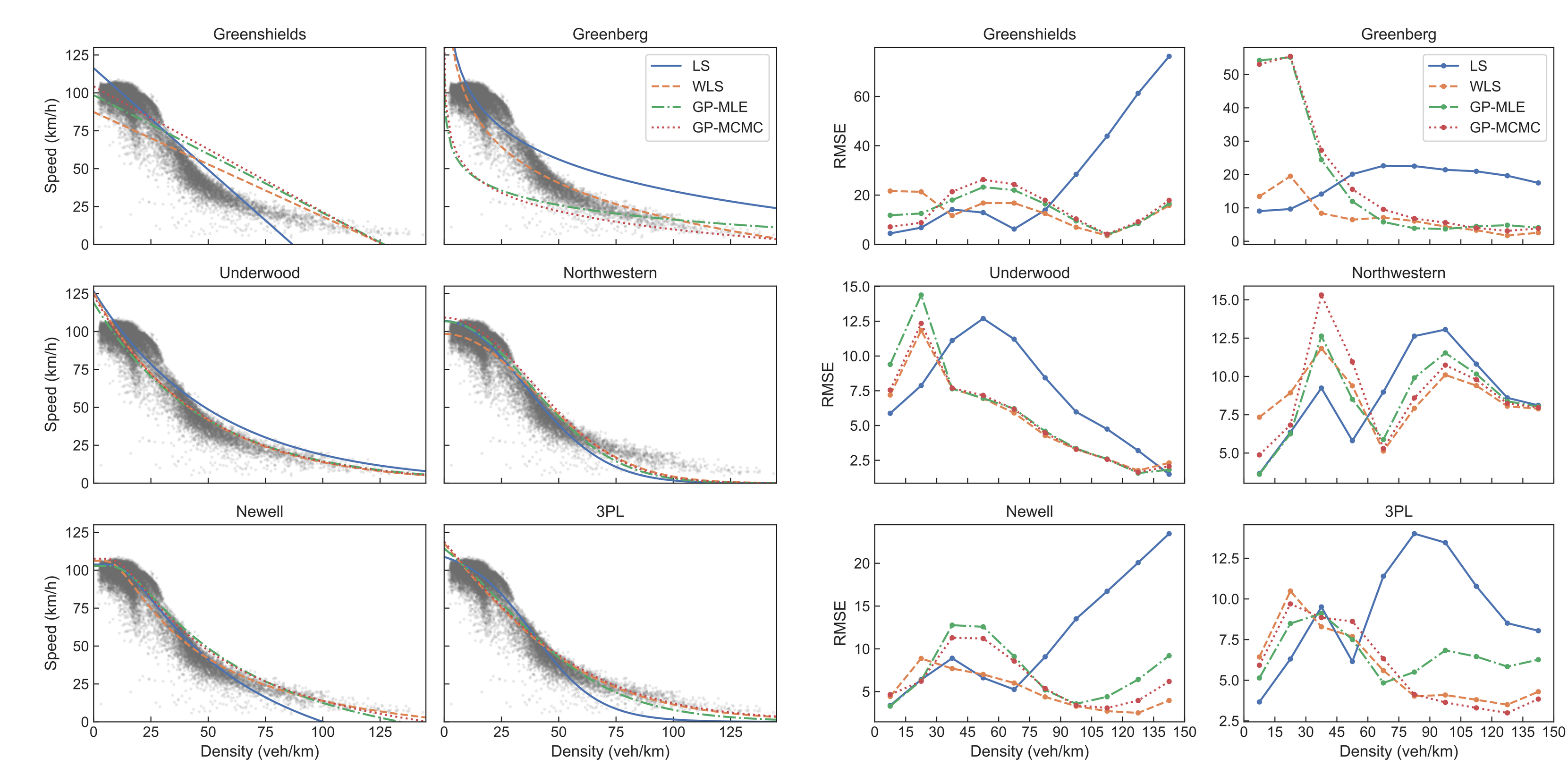
The time and storage complexity of estimating a sparse GP using the MLE reduces to $\mathcal{O}(nu^2)$.

MCMC for variational sparse GP

We adopted a fully Bayesian estimation based on an MCMC sampling scheme [1]. Because:

- A long-tailed distribution (e.g., Student-t) for the noise ε is more robust to outliers, which requires MCMC to solve.
- Traffic practitioners often have strong 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.

Comparison with other calibration methods



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

(b) The RMSE of the four calibration methods on the six single-regime models at different density groups.

Figure 2. Comparison with other calibration methods on the GA400 dataset.

Posteriors of the MCMC estimations

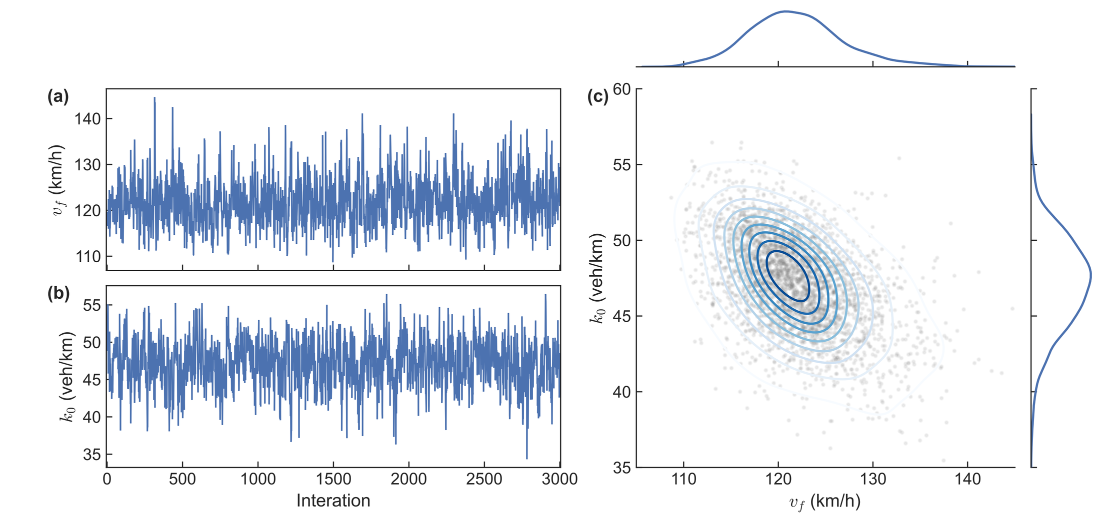
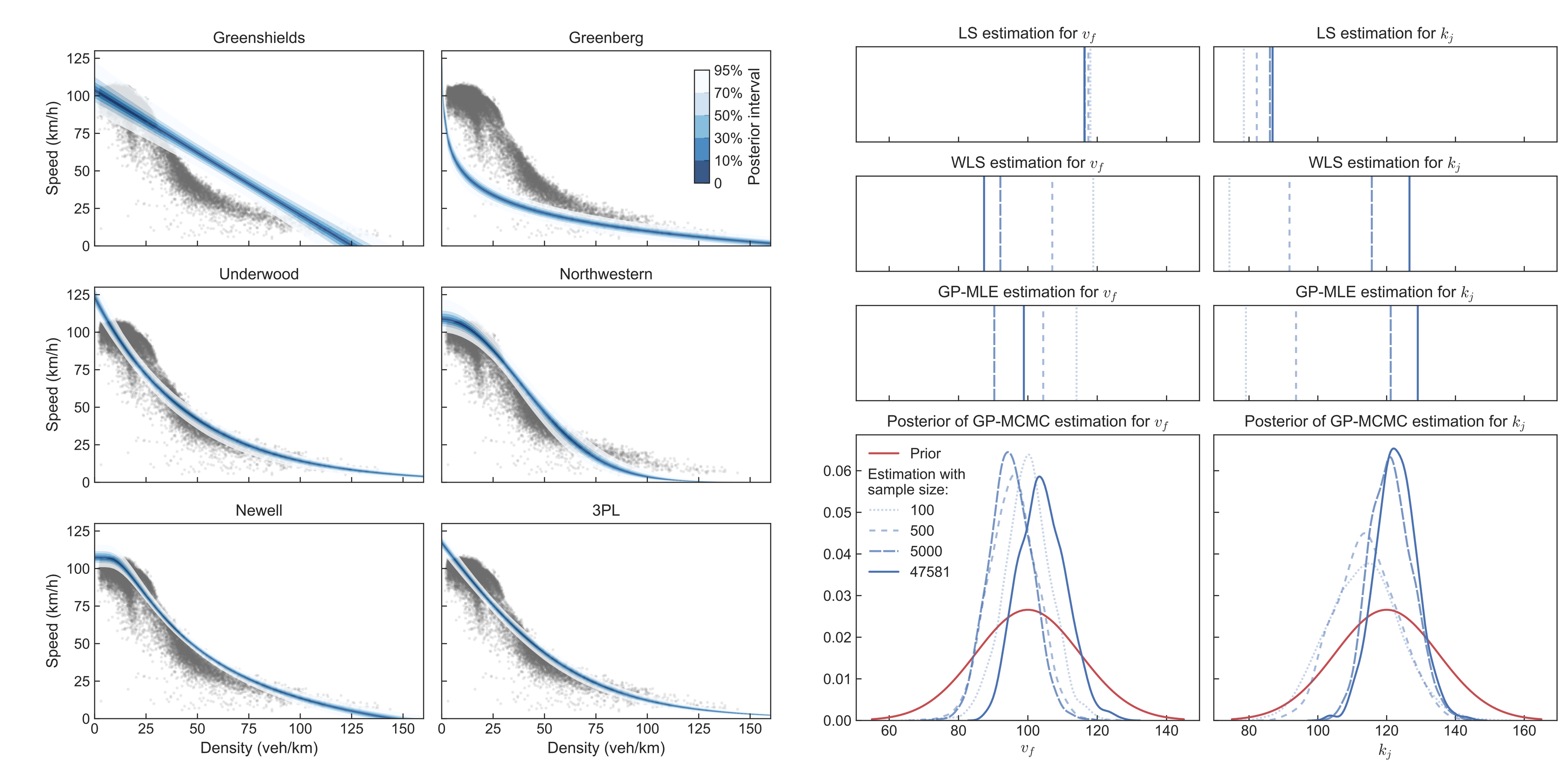


Figure 3. MCMC sampling traces and the posterior distributions of parameters for the Underwood model.



(a) Posterior distributions of speed-density functions.

(b) Parameters of the Greenshields model calibrated by the four methods under different sample sizes. The GP MCMC with a prior is more robust under different sample sizes.

Figure 4. Posteriors of the MCMC estimations

Acknowledgement

- This research is supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canada Foundation for Innovation (CFI). X. Wang would like to thank FRQNT for providing the B2X Doctoral Scholarship.
- Code available at `https://github.com/chengzhanhong/gaussian_process_calibration`.
- If you want to cite this work, please cite: Cheng, Z., Wang, X., Chen, X., Trépanier, M., & Sun, L. (2022). Bayesian calibration of traffic flow fundamental diagrams using Gaussian processes. IEEE Open Journal of Intelligent Transportation Systems.

References

[1] James Hensman, Alexander G Matthews, Maurizio Filippone, and Zoubin Ghahramani. Mcmc for variationally sparse gaussian processes. *Advances in Neural Information Processing Systems*, 28, 2015.

[2] Xiaobo Qu, Shuaian Wang, and Jin Zhang. On the fundamental diagram for freeway traffic: A novel calibration approach for single-regime models. *Transportation Research Part B: Methodological*, 73:91–102, 2015.