# Data driven uncertainty quantification in macroscopic traffic flow models

Alexandra Würth, Mickael Binois, Paola Goatin, Simone Göttlich

# Data driven uncertainty quantification in macroscopic traffic flow models

Alexandra Würth[1] Mickaël Binois[1] Paola Goatin[1] Simone Göttlich[2]

November 17, 2021

**Abstract**

We propose a Bayesian approach for parameter uncertainty quantification in macroscopic traffic flow models from cross-sectional data. A bias term is introduced and modeled as a Gaussian process to account for the traffic flow models limitations. We validate the results comparing the error metrics of both first and second order models, showing that second order models globally perform better in reconstructing traffic quantities of interest.

*2020 Mathematics Subject Classification: 35L65, 35R30, 62F15, 76A30*

**Key words:** Macroscopic traffic flow models; Godunov scheme; loop detector traffic data; Bayesian calibration; parameter estimation; optimization.

## 1 Introduction

Macroscopic traffic flow models, consisting in hyperbolic partial differential equations based on the mass conservation principle, are employed since several decades to describe the spatio-temporal evolution of traffic aggregate quantities such as density and mean velocity on road networks. Compared to microscopic approaches, they offer the advantage of being computationally less expensive, and therefore adapted to large road networks. Moreover, their analytical properties make them suitable for solving optimal control problems motivated by traffic management issues. Last but not least, they involve a small number of parameters, thus reducing calibration cost. Yet, the inherent simplification of the dynamics induced by the models, their non-linearity and the data noise are all sources of challenging difficulties when dealing with parameter identification.

In this work, we focus in particular on the comparison between first order models, that are models consisting only in the mass conservation equation, here represented by the celebrated Lighthill-Whitham-Richards (LWR) model [36, 48], and second order ones, including a second equation accounting for the speed evolution.

Classically, macroscopic traffic models are calibrated either by fitting the so-called fundamental diagram, i.e., the density-flow or density-speed mapping described by the model flux

[1]Inria Sophia Antipolis - Méditerranée, Université Côte d'Azur, Inria, CNRS, LJAD, 2004 route des Lucioles - BP 93, 06902 Sophia Antipolis Cedex, France. E-mail: {alexandra.wuerth, mickael.binois, paola.goatin}@inria.fr

[2]University of Mannheim, Department of Mathematics, B 6, 28-29, 68159 Mannheim, Germany. E-mail: goettlich@uni-mannheim.de

function (see e.g. [7, 15, 16, 17, 18]), or by minimizing some error measure of the simulation output, against either data provided by loop detectors at fixed locations [42, 50, 53] or trajectory data [44, 55]. Also, data information can be leveraged to improve existing models or design new ones, as proposed by [14, 16, 17, 18, 28, 40]. Nevertheless, up to our knowledge few works have been devoted to evaluate the inherent uncertainty of both models and data and its impact on model-based predictions [3, 21, 29]. Yet, this is a fundamental aspect to improve decision and control strategies based on mathematical models. To this end, we propose here to follow a *Bayesian approach*, which allows quantification of uncertainty in parameter estimates through their posterior probability distribution [33].

Statistical calibration techniques have been recently used for pedestrian models [10, 23, 24, 25] against microscopic trajectory data. In this paper, we only consider aggregate measurements provided by detectors at fixed locations. Without the ability to closely follow the microscopic dynamics, we are forced to use non-intrusive methods, combined with comparatively coarser grained data. Consequently, following Kennedy-O'Hagan [34], we introduce a bias term to better account for possible discrepancies between the mathematical models and reality, which also need to be estimated. This generic framework has been applied in a variety of fields, e.g. ranging from physics [30] to engineering [2, 32] or biology [45]. See also, e.g. [5] for a recent review of the methods and [4] for a discussion on the model discrepancy.

We remark that the present work refers to off-line traffic state reconstruction, and not to real-time traffic state estimation and prediction, which is usually addressed with other data assimilation techniques such as extensions of Kalman filter (see e.g. [54]) or more general particle filtering [46]. Specifically, [46] focuses on quantifying uncertainty of the road capacity, considering solely the LWR model, with a triangular fundamental diagram, where the bias correction is over fixed calibration parameters and not the model itself.

The article is organized as follows. In Section 2, we introduce the class of Generic Second Order Models, which will be the object of the study. Section 3 describes the data sets considered for model calibration. The statistical model for this data is then presented in Section 4. The calibration results are discussed in Section 5 and some perspectives are drawn in Section 6.

## 2 Macroscopic traffic flow models

Generic Second Order traffic flow Models (GSOM in short) were introduced in [35] and consist in $2 \times 2$ hyperbolic systems of the form

$$\begin{cases} \partial_t \rho + \partial_x(\rho v) = 0, \\ \partial_t w + v \partial_x w = 0, \end{cases} \qquad t > 0,\ x \in \mathbb{R}, \tag{2.1}$$

where the average speed of vehicles is a function of the density $\rho = \rho(x,t)$ and a Lagrangian vehicle property $w = w(x,t)$, namely $v = \mathcal{V}(\rho, w)$ for some speed function $\mathcal{V}$ satisfying the following hypotheses [17]:

$$\mathcal{V}(\rho, w) \geq 0, \quad \mathcal{V}(0, w) = w, \quad \mathcal{V}(\rho, 0) = 0, \tag{2.2a}$$

$$\frac{\partial^2 Q}{\partial \rho^2}(\rho, w) < 0 \text{ for } w > 0, \quad \text{where } Q(\rho, w) := \rho \mathcal{V}(\rho, w) \tag{2.2b}$$

$$\frac{\partial \mathcal{V}}{\partial w}(\rho, w) > 0, \tag{2.2c}$$

$$\forall w > 0 \quad \exists\, R_w > 0 : \quad \mathcal{V}(R_w, w) = 0. \tag{2.2d}$$

In particular, condition $\mathcal{V}(0, w) = w$ in (2.2a) identifies $w$ as the driver dependent *empty road velocity*. As in [8, 17], we observe that (2.2b) implies $\dfrac{\partial \mathcal{V}}{\partial \rho}(\rho, w) < 0$ for $w > 0$, if $\mathcal{V}$ is a $C^2$ function in $\rho$. We also remark that in (2.2d) we can have $R_w = R$ for all $w > 0$.

Setting $y := \rho w$ for $\rho > 0$, system (2.1) can be rewritten in conservative form as

$$\begin{cases} \partial_t \rho + \partial_x(\rho v) = 0, \\ \partial_t y + \partial_x(y v) = 0, \end{cases} \qquad x \in \mathbb{R},\ t > 0, \tag{2.3}$$

with $v = \widetilde{\mathcal{V}}(\rho, y) := \mathcal{V}\left(\rho, \dfrac{y}{\rho}\right)$.

System (2.1), respectively (2.3), is strictly hyperbolic for $\rho > 0$, with eigenvalues

$$\lambda_1(\rho, w) = \mathcal{V}(\rho, w) + \rho \mathcal{V}_\rho(\rho, w), \qquad \lambda_2(\rho, w) = \mathcal{V}(\rho, w), \tag{2.4}$$

with the first characteristic field being genuinely non-linear and the second linearly degenerate. The associated Riemann invariants are

$$z_1(\rho, w) = \mathcal{V}(\rho, w), \qquad z_2(\rho, w) = w.$$

Since shock and rarefaction curves of the first family coincide, the system belongs to the Temple class [51]. Notice that, setting $\mathcal{V}(\rho, w) = w - p(\rho)$ for a suitable "pressure" function $p$, system (2.3) corresponds to the celebrated Aw-Rascle-Zhang (ARZ) model [1, 56]. We also remark that, taking $w = const$, we recover the classical Lighthill-Whitham-Richards (LWR) model [36, 48].

In the present setting, we are interested in the Initial Boundary Value Problem (IBVP) for (2.3), namely

$$\begin{cases} \partial_t \rho + \partial_x(\rho v) = 0, \\ \partial_t(\rho w) + \partial_x(\rho w v) = 0, \end{cases} \qquad x \in ]x_{in}, x_{out}[ \subset \mathbb{R},\ t > 0, \tag{2.5a}$$

$$(\rho, w)(x, 0) = (\rho_0, w_0)(x), \qquad x \in ]x_{in}, x_{out}[, \tag{2.5b}$$

$$(\rho, w)(x_{in}, t) = (\rho_{in}, w_{in})(t), \qquad t > 0, \tag{2.5c}$$

$$(\rho, w)(x_{out}, t) = (\rho_{out}, w_{out})(t), \qquad t > 0, \tag{2.5d}$$

with values in an invariant domain of the form

$$\Omega_{\mathcal{V}} := \left\{ (\rho, w) \in \mathbb{R}^2_+ : w \in [w_{min}, w_{max}] \right\} \tag{2.6}$$

for some fixed $0 \le w_{min} \le w_{max} < +\infty$ being respectively the admissible bounds of the vehicle property $w$.

Solutions to (2.5) have to be intended in the weak sense as in [9, Definition 2.2]. In particular, we remark that boundary conditions (2.5c) and (2.5d) may not be satisfied in the strong sense, i.e., we may have

$$(\rho_{in}, w_{in})(t) \neq \lim_{x \to x_{in}+} (\rho, w)(x, t),$$

$$(\rho_{out}, w_{out})(t) \neq \lim_{x \to x_{out}-} (\rho, w)(x, t).$$

Note also that general IBVP well-posedness results provided in the literature, see e.g. [9, Theorem 2.3], hold under the hypothesis of strict hyperbolicity, which is not satisfied by (2.5a) at $\rho = 0$. The well-posedness of (2.5) will make the object of a separate study.

**Remark 1.** *In the implementation, we derive the initial and boundary conditions directly from real traffic data. Since the quantity $w$ is not directly provided by the measured data, we need to invert the velocity function $\mathcal{V}$. This gives us a new function $\mathcal{W}$ which reads as*

$$\begin{aligned} \mathcal{W} : \Omega_{\mathcal{W}} &\to [w_{min}, w_{max}] \\ (\rho, v) &\mapsto \mathcal{W}(\rho, v) \end{aligned} \tag{2.7}$$

*where the domain is defined as $\Omega_{\mathcal{W}} = \{(\rho, v) \colon \rho \in [0, R_{w_{max}}], v \in [\mathcal{V}(\rho, w_{min}), \mathcal{V}(\rho, w_{max})]\}$. We note that providing the initial data and boundary conditions for both the density and the empty-road velocity, the second order GSOM model is equipped with more information compared to the first order LWR model.*

In this paper, we consider a speed function of the form

$$\mathcal{V}(\rho, w) = w \left( 1 - \exp\left( \frac{C}{V} \left( 1 - \frac{R}{\rho} \right) \right) \right), \tag{2.8}$$

which is derived from Newell-Franklin [20, 41] and satisfies (2.2). In (2.8), the parameters to be identified are $\theta = (V, C, R)$, with $V > 0$ is the maximal speed, $R > 0$ is the maximal density and $C > 0$ is the wave propagation speed in congestion. The first order LWR model is obtained fixing $w = V$ in (2.8).

## 2.1 Numerical solution

To efficiently compute approximate solutions of (2.5), we use the Harten-Lax-van Leer (HLL) Riemann solver [27]. Given a (possibly non-uniform) spatial discretization $\{x_0, \dots, x_M\}$ of the interval $]x_{in}, x_{out}[$ (with $x_0 = x_{in}$ and $x_M = x_{out}$), we set the cell sizes $\Delta x_j := x_j - x_{j-1}$ for $j = 1, \dots, M$, and a suitable time step $\Delta t$ satisfying the Courant-Friedrichs-Lewy (CFL) stability condition

$$\Delta t \leq \frac{\min_j \Delta x_j}{\max_{(\rho,w)\in\Omega_{\mathcal{V}}} \{ |\lambda_1(\rho, w)|, |\lambda_2(\rho, w)| \}}. \tag{2.9}$$

Denoting by

$$U := \begin{pmatrix} \rho \\ \rho w \end{pmatrix} \qquad \text{and} \qquad F(U) := \begin{pmatrix} \rho v \\ \rho w v \end{pmatrix}$$

the vectors of the conserved quantities and fluxes respectively, we approximate the initial data (2.5b) as

$$\mathbf{U}_j^0 := \frac{1}{\Delta x_j} \int_{x_{j-1}}^{x_j} U^0(y)\, dy.$$

Approximate solutions are then computed iteratively by the finite volume formula

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x_j} \left( \mathbf{F}_j^n - \mathbf{F}_{j-1}^n \right), \qquad j = 1, \dots, M, \tag{2.10}$$

where the numerical fluxes $\mathbf{F}_j^n := F_{HLL}(\mathbf{U}_j^n, \mathbf{U}_{j+1}^n)$ are computed using the HLL Riemann solver:

$$F_{HLL}(\mathbf{U}_L, \mathbf{U}_R) := \begin{cases} F(U_L) & \text{if } S_L \geq 0, \\ F(U_{HLL}) & \text{if } S_L < 0 \leq S_R, \\ F(U_R) & \text{if } S_R < 0, \end{cases} \tag{2.11}$$

where $S_L < S_R$ are the approximate wave speeds. Following [13, 17], we employ the following definition of characteristic speeds

$$\begin{aligned} S_L &:= \min\{\lambda_1(\mathbf{U}_L), \lambda_1(\mathbf{U}_R)\}, \\ S_R &:= \min\{\lambda_2(\mathbf{U}_L), \lambda_2(\mathbf{U}_R)\}, \end{aligned} \tag{2.12}$$

and we set

$$F(U_{HLL}) := \frac{S_R F(\mathbf{U}_R) - S_L F(\mathbf{U}_L) - \big(F(\mathbf{U}_R) - F(\mathbf{U}_L)\big)}{S_R - S_L}.$$

Note that, since $\lambda_2(\rho, w) = v \geq 0$ for all $(\rho, w) \in \Omega_{\mathcal{V}}$, the case $S_R < 0$ in (2.11) never occurs.

Boundary conditions (2.5c) and (2.5d) are taken into account by (2.10) setting

$$\begin{aligned} \mathbf{U}_0^n &:= \frac{1}{\Delta t} \int_{t^{n-1}}^{t^n} U_{in}(s)\, ds, \\ \mathbf{U}_{M+1}^n &:= \frac{1}{\Delta t} \int_{t^{n-1}}^{t^n} U_{out}(s)\, ds, \end{aligned}$$

where $t^n = n\Delta t$ and $n = 1, \ldots, N$.

When considering the scalar LWR equation, we use the classical Godunov scheme [22] in its supply-demand implementation [12]. We note that, for the first order LWR model, the boundary conditions can be given in terms of flows $F_{in}$ and $F_{out}$, instead of densities. Thus, the IBVP (2.5) reads

$$\begin{aligned} &\partial_t \rho + \partial_x(\rho v) = 0, && x \in ]x_{in}, x_{out}[,\ t > 0, \\ &\rho(x, 0) = (\rho_0)(x), && x \in ]x_{in}, x_{out}[, \\ &q(x_{in}, t) = F_{in}(t), && t > 0, \\ &q(x_{out}, t) = F_{out}(t), && t > 0. \end{aligned}$$

**Remark 2.** *Since initial and boundary conditions provided by data, but also approximate values computed by finite volume schemes [6] can lie outside the domain of the inverse mapping $\mathcal{W}$, we will perform a projection following [17, Section 3.3]:*
*Given a density $0 \leq \rho \leq R_{w_{max}}$ and any speed $v > 0$ (can be outside of the domain $\Omega_{\mathcal{W}}$) we consider*

$$\tilde{\mathcal{V}}(\rho, v) = \min\Big\{\max\{v, \mathcal{V}(\rho, w_{min})\}, \mathcal{V}(\rho, w_{max})\Big\} \tag{2.13}$$

*where $\mathcal{V}(\rho, w_{min})$ (resp. $\mathcal{V}(\rho, w_{max})$) is the lower (resp. upper) bound of the speed curves. This allows to define $\tilde{D}_{\mathcal{W}} := \{(\rho, v) \colon \rho \in [0, R_{w_{max}}], v \geq 0\}$.*

# 3 Data and simulation setting

In this section, we describe the two sensor data sets that will be used for the calibration tests in Section 5.

One data set was provided by the *Direction Interdépartementale des Routes Méditerranée* (DIRMED) and covers 3 months, from September 1st to November 30th, 2015. The considered set of data (DirMed data) comes from 135 loop detectors placed on the freeways in the North of Marseille (A7, A51, A50 and A55). Given the 6 minute averages of the traffic flow and speed measured by the loop detectors, the traffic density is a derived quantity.
In this paper, we focus on a 1.25 km long rampless road stretch on the freeway A50 from Aubagne to Marseille. The road consists of three lanes and the speed limit is 90 km/h. The data for this freeway section come from 4 loop detectors (IDs 305, 304, 303 and 302). All the details about the considered traffic data and the loop detector locations are available in the git repository `https://gitlab.inria.fr/acumes/dduq-traffic`.
The fundamental diagrams in Figure 1 illustrate the collected flow and speed data together with their corresponding densities for loop 303. Most of the data points are located in the free flow phase which is visible by the clear functional relationship between density and flow or speed. The free flow phase corresponds roughly to speed values greater than 60 km/h. The more widely distributed data points in the higher density region correspond to congested regimes.

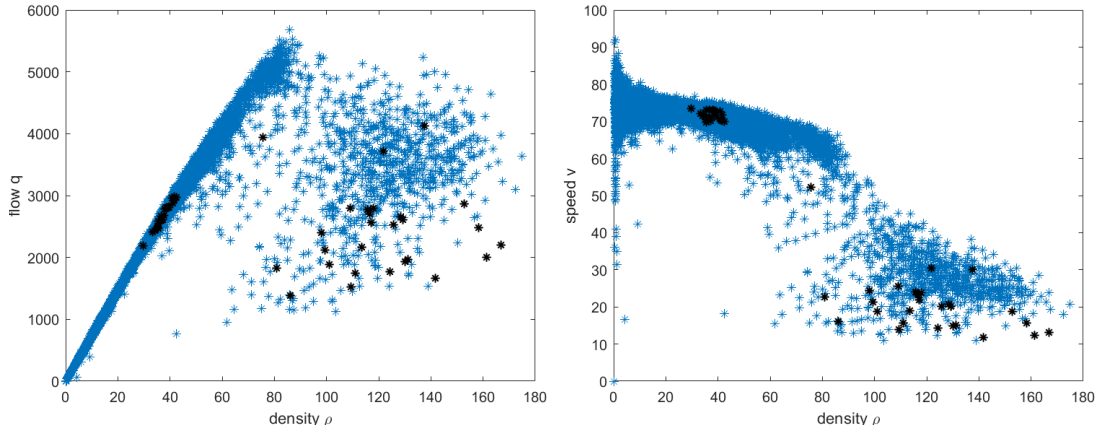



Figure 1: Fundamental flow and speed diagrams for DirMed data corresponding to loop 303. Data from 09/01/2015 to 11/30/2015. In black: data corresponding to the selected DirMed scenario (see Figure 3a).

For the calibration and validation in Section 5, we will focus on Tuesday 11/10/2015 for a five hours time slot between 7am and 12pm. The density evolution is illustrated in Figure 3a.

The other data set is referred to as the RTMC data set [39], which is provided by the Minnesota Department of Transportation (MnDOT). Analogously to the DirMed data set, the MnDOT data are 6 minute averages obtained by single loop detectors measuring the traffic flow and the occupancy (fraction of time that a vehicle is occupying a detector) along several (interstate) highways (while in the DirMed data set, speeds were measured instead of the occupancy). For the tests, we consider a 1.1 km long road stretch on the northbound direction

of the interstate highway I-35W equipped with 4 loop detectors ($S60, S61, S1708, S62$). The rampless road stretch has five lanes and the speed limit is 55 miles/hour ($\approx$ 90 km/h).
The fundamental flow and speed diagrams for the inner loop $S1708$ are depicted in Figure 2. In contrast to Figure 1, the congested area is equipped with more data points, which are again widely spread. However, the functional relationship in the free flow region is still visible, especially in the flow diagram. We emphasize that both the maximum density and the maximum flow value are higher than in the DirMed data case, since the road has five lanes (instead of three).

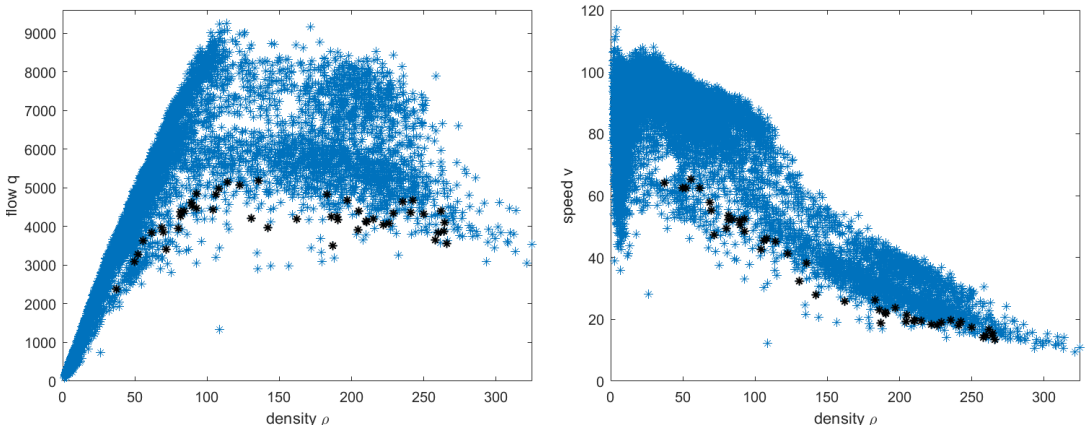
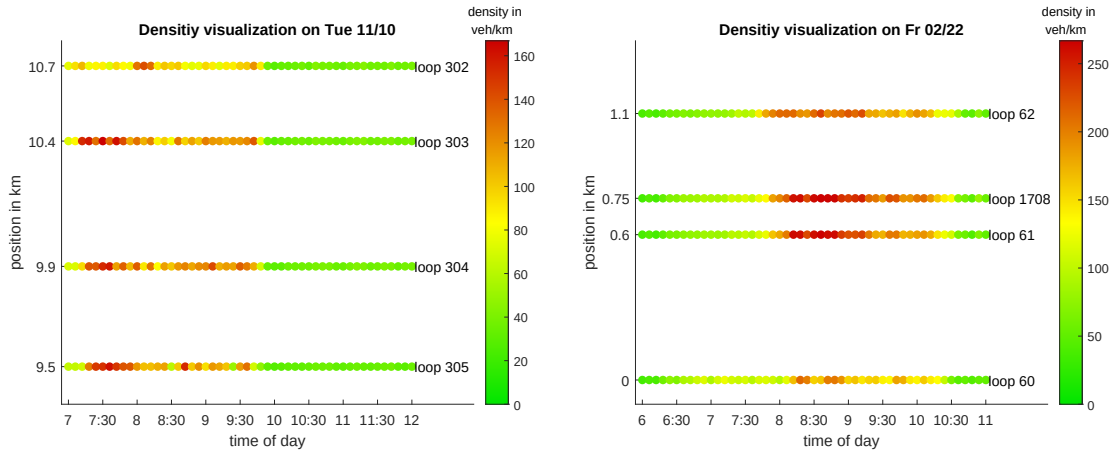



Figure 2: Fundamental flow and speed diagrams for RTMC data corresponding to loop $S1708$. Data from weekdays from 01/01/2013 to 04/12/2013. In black: data of the selected RTMC scenario (see Figure 3b).

In our analysis, we will consider the data on Friday 02/22/2013 in a five hours time slot between 6am and 11am. The density evolution is illustrated in Figure 3b.
Figure 3 emphasizes that we take both congestion (orange and red colored densities) and




(a) DirMed data

(b) RTMC data

Figure 3: Density visualization for two data scenarios in a 5 hours time slot.

free flow (green colored densities) phases into account. The rush hour for the DirMed data case takes place at the beginning of the considered time period, between 7:30am and 9:30am, whereas the congested phase for the RTMC data starts after two hours, at 8:00am.

**Remark 3.** *In order to overcome the difficulty of precisely reconstructing the initial density condition from loop measurements, we run the traffic model through an initialization phase of 6 minutes (see [17]). Thus, the validation of the data will be finally done on a 4 hours 54 minutes time slot.*

# 4 Kennedy-O'Hagan calibration framework

This section describes the calibration approach we follow to identify the traffic model parameters from real data. Instead of a least-squares approach applied to fundamental diagrams, as generally performed in this context, see e.g. [14, 17], we adopt the statistical framework proposed by [30, 34]. The benefits are two-fold. First, accounting for and estimating a bias term between the mathematical model and data will improve the prediction results. Second, additionally relying on a Bayesian estimation procedure, uncertainty quantification is directly available in the form of a posterior distribution on the calibration parameters, rather than scalar values. Looking at the fundamental diagrams of Figures 1 and 2, it is clear that a single parameter $\theta$ value is insufficient to properly fit the spread of point measures.

In the following, we denote by $P$ the real process under study (e.g. the space-time evolution of the density $\rho$), $F$ the so-called "field" where $P$ is physically observed, both depending on $x$ and $t$. Denote by $M$ the mathematical model, which depends on $(x, t, \theta)$, with $\theta$ the additional calibration parameter(s), and by $(X, T)$ all the $(x, t)$ points where observations have been recorded. Denote by $y^F(x, t)$ the field observations under conditions $x, t$ of the real output $y^P(x, t)$. It is generally assumed that $P$ and $F$ are related by

$$y^F(x,t) = y^P(x,t) + \varepsilon, \text{ where } \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2).$$

Kennedy-O'Hagan (KOH) [34] proposed to additionally take into account the inadequacy between the mathematical model with optimal parameters $\theta^*$ and reality, via an additional discrepancy (or bias) term $b(x, t, \theta)$: $y^P(x, t) = y^M(x, t, \theta^*) + b(x, t, \theta^*)$ resulting in

$$y^F(x,t) = y^M(x,t,\theta^*) + b(x,t,\theta^*) + \varepsilon.$$

We rely on Gaussian process regression for estimating the bias term [30, 34], which amounts to assume a multivariate normal distribution for the errors. KOH also models $y^M$ with a Gaussian process as $y^M$ is computationally expensive in their setup. This is not necessary here since the mathematical model evaluation is cheap, thus corresponding to the framework described in [30].

## 4.1 Gaussian process modeling of the bias

Given a set of observations of the bias $\mathbf{b}_n = \Big(b\big((x_1, t_1), \ldots, (x_n, t_n)\big)\Big)$ (dropping $\theta^*$ for notational simplicity) at $n$ observation points $\mathcal{X}_n = \big((x_1, t_1), \ldots, (x_n, t_n)\big)$, this GP assumption amounts to consider that $\mathbf{b}_n$ is a realization of a (zero-mean) multivariate normal (MVN) distribution:

$$\mathbf{b}_n \sim \mathcal{N}(\mathbf{0}_n, \mathbf{K}_n) \text{ with } \mathbf{K}_n = \sigma^2(\mathbf{C}_n + g\mathbf{I}_n) \text{ and } g = \frac{\sigma_\varepsilon^2}{\sigma^2}.$$

Here, $\mathbf{K}_n$ (resp. $\mathbf{C}_n$) denotes the covariance (resp. correlation) matrix between the observed biases; the (hyper-)parameter $\sigma^2$ stands for the process variance. More precisely, the matrix entries of $\mathbf{C}_n$ are computed by $\mathbb{C}orr\Big(b\left((x,t)\right), b((x',t'))\Big) = c\left((x,t),(x',t')\right)$ where $c(\cdot,\cdot)$ is a positive definite function, typically from a parametric family, such as the Gaussian kernel:

$$c\left((x,t),(x',t')\right) = \exp\left(-\frac{(t-t')^2}{l_1^2}\right) \times \exp\left(-\frac{(x-x')^2}{l_2^2}\right),$$

among others like Matérn kernels, see e.g. [49]. The (hyper-)parameters $l_1$ and $l_2$ denote, respectively, the length-scales for the time and space variables.

The GP predictive equations of the bias at $\hat{n}$ new locations $\hat{\mathcal{X}}_{\hat{n}} = \Big((\hat{x}_1,\hat{t}_1),\ldots,(\hat{x}_{\hat{n}},\hat{t}_{\hat{n}})\Big)$, denoted $\mathbf{b}(\hat{\mathcal{X}}_{\hat{\mathbf{n}}}) \mid \mathbf{b}_n$, are the so-called "kriging" equations:

$$\mathbf{b}(\hat{\mathcal{X}}_{\hat{\mathbf{n}}}) \mid \mathbf{b}_n \sim \mathcal{N}(m_n(\hat{\mathcal{X}}_{\hat{n}}), s_n^2(\hat{\mathcal{X}}_{\hat{n}},\hat{\mathcal{X}}_{\hat{n}})), \text{with} \tag{4.1a}$$

$$m_n(\hat{\mathcal{X}}_{\hat{n}}) := \mathbb{E}[\mathbf{b}(\hat{\mathcal{X}}_{\hat{\mathbf{n}}})|\mathbf{b}_n] = \mathbf{k}_n(\hat{\mathcal{X}}_{\hat{n}})^\top \mathbf{K}_n^{-1}\mathbf{b}_n, \tag{4.1b}$$

$$s_n^2(\hat{\mathcal{X}}_{\hat{n}},\hat{\mathcal{X}}_{\hat{n}}) := \mathbb{C}ov[\mathbf{b}(\hat{\mathcal{X}}_{\hat{\mathbf{n}}}),\mathbf{b}(\hat{\mathcal{X}}_{\hat{\mathbf{n}}})|\mathbf{b}_n] = k(\hat{\mathcal{X}}_{\hat{n}},\hat{\mathcal{X}}_{\hat{n}}) - \mathbf{k}_n(\hat{\mathcal{X}}_{\hat{n}})^\top \mathbf{K}_n^{-1}\mathbf{k}_n(\hat{\mathcal{X}}_{\hat{n}}) \tag{4.1c}$$

where $k(\cdot,\cdot) = \sigma^2 c(\cdot,\cdot)$, $\mathbf{k}_n(\hat{\mathcal{X}}_{\hat{n}}) := (k(\hat{\mathcal{X}}_{\hat{n}}^{(j)}, \mathcal{X}_n^{(i)}))_{1\le j\le \hat{n}, 1\le i\le n}$ and the predictive variance is the diagonal of $s_n^2(\hat{\mathcal{X}}_{\hat{n}},\hat{\mathcal{X}}_{\hat{n}})$. Formulas (4.1) describe the best (minimizing Mean Square Prediction Error) linear unbiased predictor (BLUP).

To estimate the hyperparameters of the kernel function, we maximize the likelihood, i.e. the probability density of the observations given the parameters: $\mathbb{P}[\mathbf{b}(\mathcal{X}_n) = \mathbf{b}_n]$. As $\mathbf{b}_n \sim \mathcal{N}(0_n, \mathbf{K}_n)$, the likelihood $\mathcal{L}$ is given by the MVN density. Taking the logarithm, this gives

$$\log \mathcal{L}(l_1,l_2,\sigma^2,g) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\log|\mathbf{C}_n + g\mathbf{I}_n| - \frac{1}{2\sigma^2}\mathbf{b}_n^\top(\mathbf{C}_n + g\mathbf{I}_n)^{-1}\mathbf{b}_n. \tag{4.2}$$

We can compute the optimal variance $\hat{\sigma}^2$ by differentiating the resulting expression so that

$$\frac{\partial \log\mathcal{L}}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\mathbf{b}_n^\top(\mathbf{C}_n + g\mathbf{I}_n)^{-1}\mathbf{b}_n = 0$$

which gives

$$\hat{\sigma}^2(l_1,l_2,g) = \frac{\mathbf{b}_n^\top(\mathbf{C}_n + g\mathbf{I}_n)^{-1}\mathbf{b}_n}{n}. \tag{4.3}$$

Plugging $\hat{\sigma}^2$ in (4.2), we obtain the concentrated likelihood $\tilde{\mathcal{L}}$:

$$\log\tilde{\mathcal{L}}(l_1,l_2,g) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat{\sigma}^2(l_1,l_2,g) - \frac{1}{2}\log|\mathbf{C}_n + g\mathbf{I}_n| - \frac{n}{2}. \tag{4.4}$$

We note that the correlation matrix $\mathbf{C}_n$ also depends on the hyperparameters $l_1, l_2$. Moreover, since the bias $\mathbf{b}_n$ depends on the unknown calibration parameter $\theta$, the process variance $\hat{\sigma}^2$ and finally $l_1, l_2$ and $g$ also depend on $\theta$. Since it is generally not possible to compute these remaining hyperparameters explicitely, numerical optimization methods are mandatory to conclude. Furthermore, due to the grid structure of the data (regular measurement at the same loop locations), the computational cost of fitting GPs (in $\mathcal{O}(n^3)$) can be reduced by exploiting the resulting Kronecker structure (see, e.g. [11] for a detailed application example). We refer to [26, 47] for further details on GP modeling.

### 4.2 Bayesian parameter identification by MCMC

In [34], the authors further proposed a Bayesian framework to estimate the best calibration parameter $\theta^*$, along with $\sigma_\varepsilon$ and $b(\cdot,\cdot,\cdot)$. The outcome is a posterior probability distribution instead of a single optimal value. Following [30], besides defining priors on the calibration parameters, one needs to be able to estimate the posterior. Bayes' rule expresses it as

$$\pi(\theta^* \mid y^F) = \frac{\mathcal{L}(y^F \mid \theta^*) \times \pi(\theta^*)}{\pi(y^F)},$$

which says that the posterior is given by the product of the likelihood and the prior, divided by the marginal likelihood. The marginal likelihood (a.k.a. evidence) is a normalizing constant that does not depend on $\theta$. Analytical expressions of posterior distributions are seldom available, leading to a variety of estimation techniques. In this paper, we use a standard Markov Chain Monte Carlo (MCMC) method, the Metropolis algorithm, see Algorithm 1. More details can be found in [30]. The samples obtained by MCMC are thinned out, keeping only one out of $p$ samples, to reduce autocorrelation in the Markov chain (see [31]).

**Algorithm 1** Metropolis algorithm with symmetric proposal distribution

**Require:** proposal covariance matrix $\mathbf{\Sigma}^p$, prior distribution $\pi(\cdot)$
1: initialize $\theta_0^*$ (sample from the prior distribution $\pi(\theta^*)$)
2: initialize $l_1(\theta_0^*), l_2(\theta_0^*), g(\theta_0^*)$ (by maximizing (4.4)) and subsequently $\hat{\sigma}^2(\theta_0^*)$ (by (4.3))
3: **for** each $i$ in $\{1, ...N\}$ **do**
4: generate $\hat{\theta}^*$ from a symmetric distribution, e.g. $\hat{\theta}^* \sim \mathcal{N}(\theta_{i-1}^*, \mathbf{\Sigma}^p)$
5: compute $l_1(\theta_i^*), l_2(\theta_i^*), g(\theta_i^*)$, $\hat{\sigma}^2(\theta_i^*)$
6: compute the ratio $\alpha := \min\left\{1, \frac{\pi(\hat{\theta}^*|y^F)}{\pi(\theta_{i-1}^*|y^F)}\right\}$
7: generate a uniform random number $u \sim \mathcal{U}([0,1])$
8: **if** $u \leq \alpha$ **then**
9: $\theta_i^* = \hat{\theta}^*$
10: **else**
11: $\theta_i^* = \theta_{i-1}^*$
12: **end if**
13: **end for**
14: Return $\theta^* = (\theta_1^*, \ldots, \theta_N^*)$

We remark that Bayesian calibration via MCMC estimation can be time consuming. Several simplifications have been proposed, such as relying on optimization and on modularization, as in [26, 37]. That is, compartmentalizing the hyperparameter optimization of the GP fitting can prevent these parameters from taking over the computer model. Moreover, it reduces the number of parameters to calibrate by MCMC.

## 5 Validation and comparison

In this section, we present our calibration results for the two data sets introduced in Section 3. Our quantity of interest ($y$) will be the flow observations since we expect flow measurements be more precise compared to the speed or density data.

### 5.1 Optimization approach

We start with considering a calibration approach leading to a single optimized value instead of a posterior probability distribution, which will be referred to as the *optimization approach*. It consists in the maximization of the concentrated log-likelihood function (4.4):

$$\max_{l_1,l_2,g,\theta} \log \tilde{\mathcal{L}}(l_1(\theta), l_2(\theta), g(\theta)) \tag{5.1}$$

Following the principle of modularization [37], we apply a 2-step optimization to isolate the GP fitting and thus preventing confounding effects in the calibration. In the so-called *inner-level*, we maximize the concentrated log-likelihood function dependent on $\theta$, thus obtaining the hyperparameters $l_1(\theta), l_2(\theta), g(\theta)$. These hyperparameters are then inserted into (4.4) and the concentrated likelihood is maximized with respect to $\theta$, giving the optimized calibration parameter $\theta^*$. We call the second step the *higher-level*.

The inner optimization level is executed by the local MATLAB optimization solver `fmincon`. For the higher-level we compare both `pso` and `fmincon` solvers, choosing the one leading to the better result. Since our quantity of interest are the flow observations, the optimization is done on the flow errors. The applied admissible intervals for the parameters are listed in Table 1.

| | | inner-level | | | higher-level | | |
|---|---|---|---|---|---|---|---|
| | | $l_1$ | $l_2$ | $g$ | $V$ | $C$ | $R$ |
| DirMed | lower bound | 0.1 | 0.1 | 0.005 | 70 | 10 | 180 |
| | upper bound | 5 | 1.2 | 5 | 120 | 70 | 390 |
| RTMC | lower bound | 0.1 | 0.1 | 0.005 | 70 | 10 | 250 |
| | upper bound | 5 | 1.1 | 5 | 120 | 70 | 750 |

Table 1: Upper and lower bounds for the optimization approach.

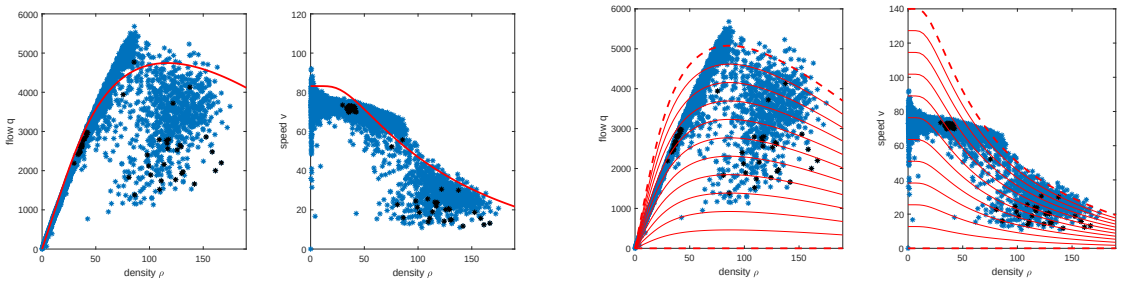
**Remark 4.** *Since the calibration of the lower and upper bound of the empty-road velocity $w$ in* (2.6) *did not lead to an improvement of the predictions in preliminary tests, we set them to reasonable values ($w_{min} = 0$ and $w_{max} = 140$) to reduce the dimension of the calibration problem. In order to limit the influence of the projection procedure described in Remark 2, giving simulated values lying on the lower or upper fundamental curves (corresponding to $w_{min}$ or $w_{max}$ respectively), we limit the number of projection occurrences in every time step $\Delta t$ of the numerical solution not to exceed 5% of the number $M$ of spatial discretization points.*
*On the contrary, unlike [16, 17], we do not set* a-priori *the value of the maximum density $R$, but we calibrate it.*

The optimization results for the calibration parameters are summarized in Table 2. We observe that the two traffic flow models (LWR and GSOM) lead to different optimal calibration parameters but lying in similar ranges. However, none of the parameters lie on the boundary in the second order model (unlike the maximum density of 390 for the LWR model on DirMed data). We note that, for both models, the maximum density $R^*$ for the DirMed data is smaller since there are only three lanes instead of five. Moreover, for these scenarios, the second order model leads to higher $V^*$ values but lower $C^*$ values, knowing that different scenarios would lead to different values.

| | LWR | | | GSOM | | |
|---|---|---|---|---|---|---|
| | $V^*$ | $C^*$ | $R^*$ | $V^*$ | $C^*$ | $R^*$ |
| DirMed | 83.51 | 23.84 | 390.00 | 94.74 | 17.02 | 348.22 |
| RTMC | 87.37 | 29.15 | 422.51 | 91.56 | 18.39 | 478.23 |

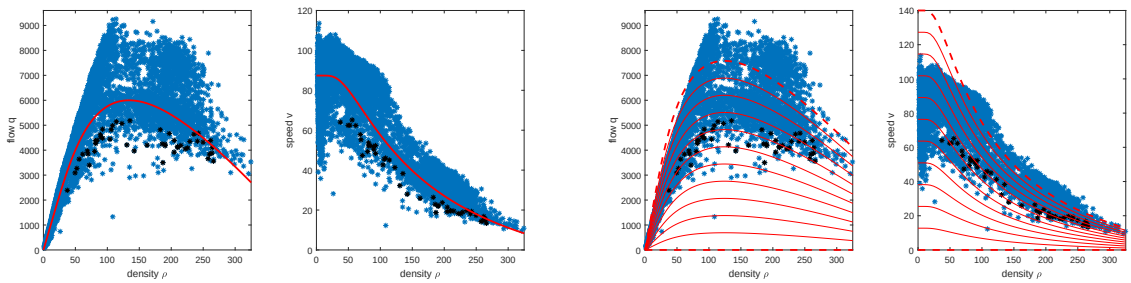
Table 2: Optimization results for the optimization approach.

The fundamental diagrams in Figures 4 and 5 report the same data as Figures 1 and 2, where we superpose the flow and speed curves obtained corresponding to the optimal calibration parameters from Table 2 for the function (2.8). Figure 4b and Figure 5b show that the GSOM model can capture the spread of the data in the congested part better, compared to the LWR model. This is due to the fact that the second order model considers a family of fundamental diagrams corresponding to different $w$ values (inserted in Equation 2.8).




(a) LWR model (b) GSOM model

Figure 4: Fundamental flow and speed diagrams for DirMed data corresponding to loop 303. In red, equilibrium curve for LWR (Figure 4a) and family of flow and speed curves for GSOM (Figure 4b), based on the optimal calibration parameters of Table 2. In black: data of the selected scenario (see Figure 3a).

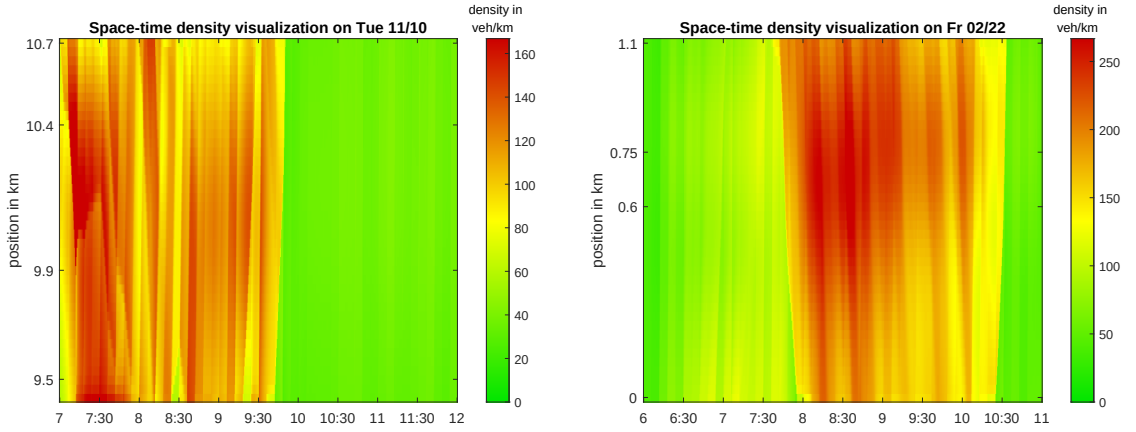



(a) LWR model (b) GSOM model

Figure 5: Fundamental flow and speed diagrams for RTMC data corresponding to loop $S1708$. In red, equilibrium curve for LWR (Figure 5a) and family of flow and speed curves for GSOM (Figure 5b) based on the optimal calibration parameter of Table 2. In black: data of the selected scenario (see Figure 3b).

In contrast to the real data, the numerical solution enables us to reconstruct the traffic density along a finer space-time dimension. This is also of interest from an application point of view (e.g. for travel time estimation), since we obtain a detailed information on traffic

dynamics (compared to Figure 3). The space-time density evolution is depicted in Figure 6.




(a) DirMed data

(b) RTMC data

Figure 6: Space-time density visualization for the simulated data.

Figure 7 depicts the flow and speed profiles for the DirMed scenario and Figure 8 shows the flow and density profiles for the RTMC scenario. We observe that for the GSOM model the numerical solution (blue squares) follows the profile of the measured data (red stars) better than the LWR model (green circles). Even jumps and drop regions are well captured, though being more difficult to predict. At this point, we want to emphasize that, although the optimization is done only on the flow errors, the GSOM is able to recover the density and speed quantities much better than the LWR model. However, by comparing the flow profiles, we observe a similar performance for the two traffic flow models. In both cases, the simulated values reflect the real data well.

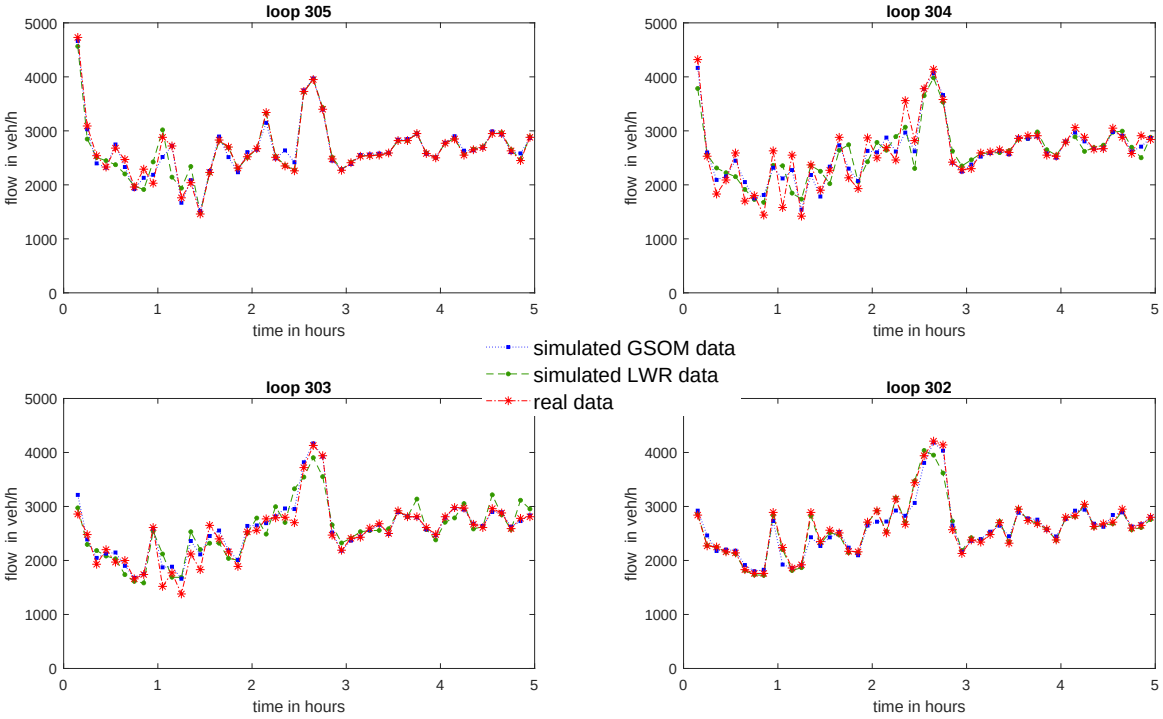
In order to compare the predictive accuracy of the results between the two traffic flow models numerically, we define an error metric $\mathbf{E}$ which consists of the sum of normalized flow, speed and density errors. Denoting $y^F_{\text{flow}}$ (resp. $y^F_{\text{speed}}$, $y^F_{\text{density}}$) the measured flow (resp. speed, density) data and $y^M_{\text{flow}}$ (resp. $y^M_{\text{speed}}$, $y^M_{\text{density}}$) as the simulated flow (resp. speed, density) data, including the correction by their predictive kriging means (4.1b), the error function is defined as the sum of

$$E^k = \frac{1}{T_f \cdot (x_{out} - x_{in})} \frac{1}{\Delta q} \sum_{(x,t)\in(X,T)} |y^F_k(x,t) - \hat{y}^M_k(x,t,\theta^*)|,$$

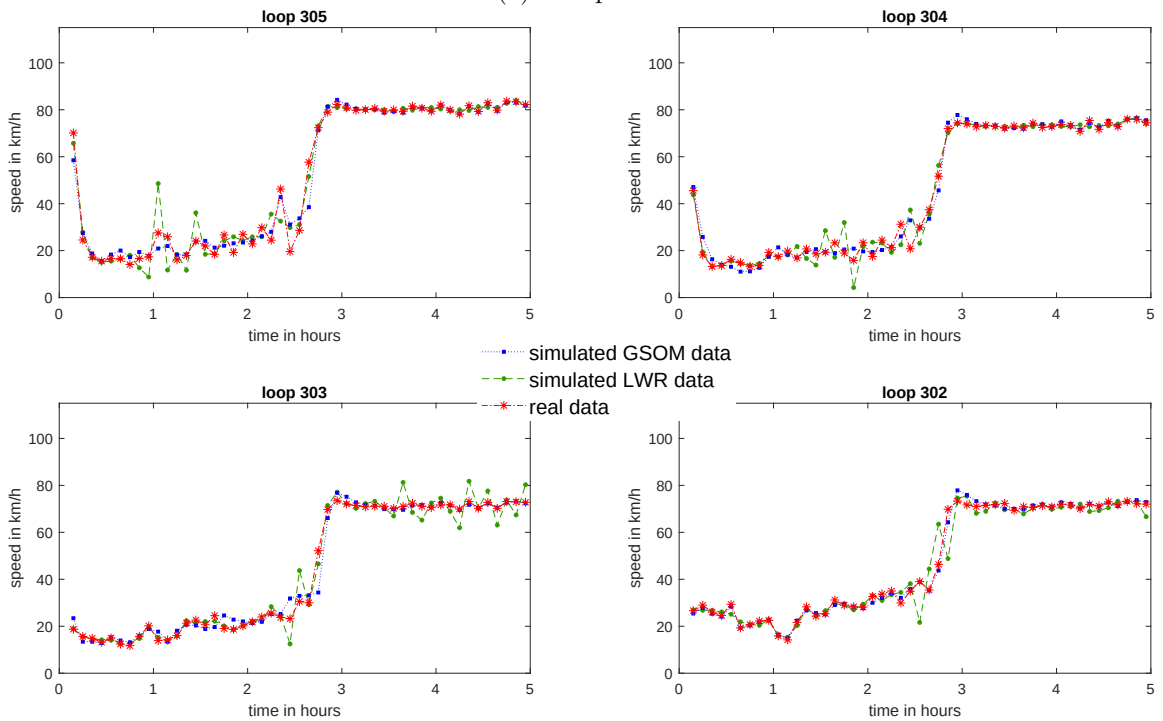
for $k \in \{\text{flow}, \text{speed}, \text{density}\}$. Thus

$$\mathbf{E} = E^{\text{flow}} + E^{\text{speed}} + E^{\text{density}}, \tag{5.2}$$

where $\hat{y}^M_k(x,t,\theta^*)$ is defined as the output of the simulation code evaluated at the optimal calibration parameter $\theta^*$ at time $t$ and loop position $x$, corrected by adding its kriging mean $m_n$. Above, $T_f$ stands for the time horizon (4 hours 54 minutes) and $\Delta q$, (resp. $\Delta v$, $\Delta \rho$) represents normalization constants to overcome the biases induced by the different physical units. For the choice of the normalization constants, we follow [16] and define $\Delta q$ (resp. $\Delta v$,

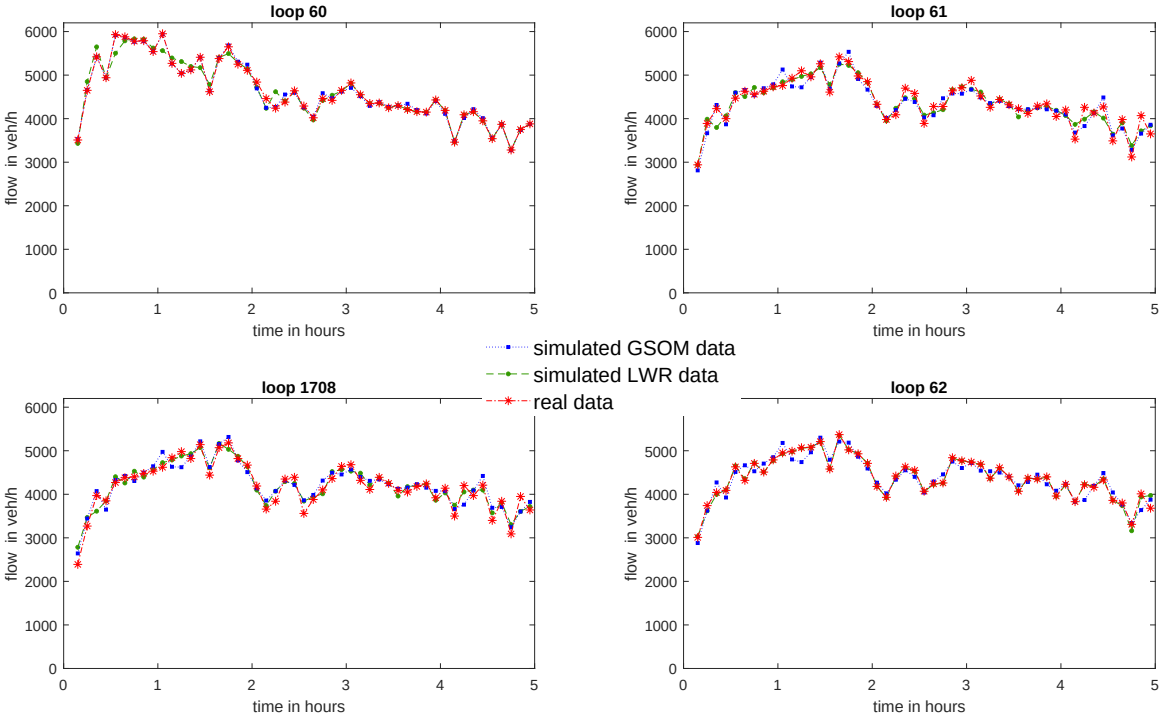



(a) flow profile

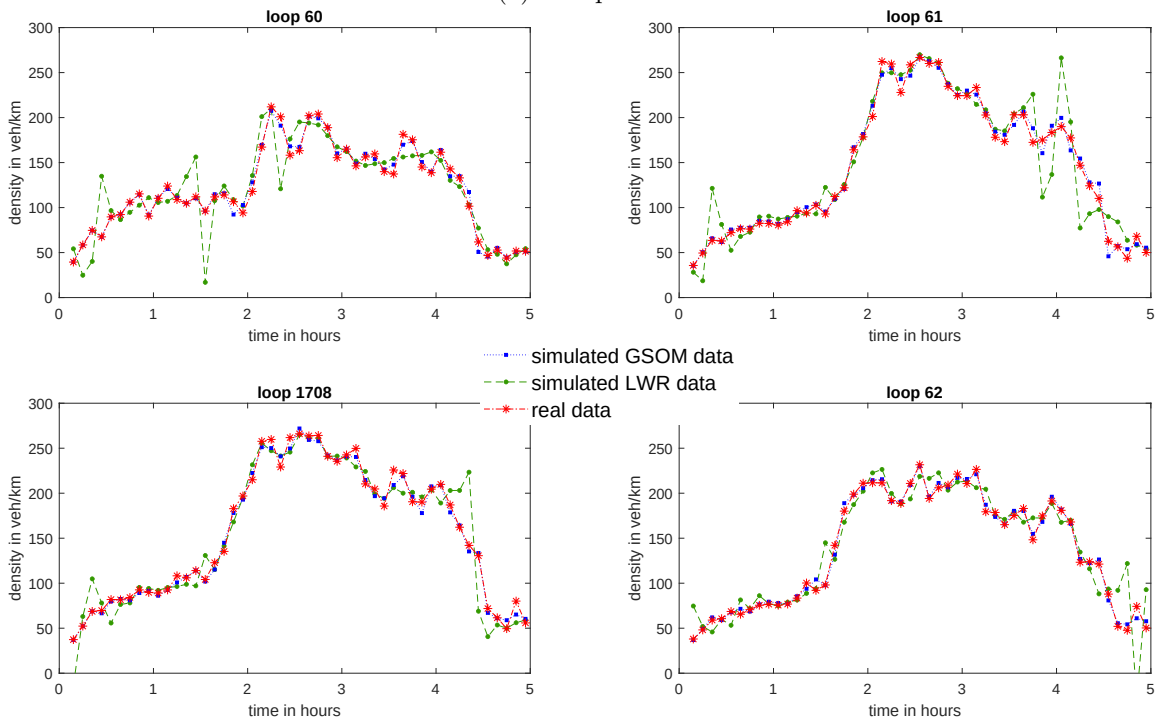



(b) speed profile

Figure 7: Comparison of real and simulated flow and speed profiles for LWR and GSOM models: optimization approach for the DirMed data.




(a) flow profile




(b) density profile

Figure 8: Comparison of real and simulated flow and density profiles for LWR and GSOM models: optimization approach for the RTMC data.

$\Delta\rho$) as the ranges of the real measured traffic flow (resp. speed, density) data. Thus, we obtain

$$\Delta q = Q^{\max} - Q^{\min},\ \Delta v = V^{\max} - V^{\min},\ \Delta\rho = \mathrm{R}^{\max} - \mathrm{R}^{\min},$$

where $Q^{\max}$ and $Q^{\min}$ (resp. $V^{\max}$ and $V^{\min}$ , $\mathrm{R}^{\max}$ and $\mathrm{R}^{\min}$ ) stands for the maximum and minimum observed traffic flow (resp. speed, density) in the considered data set.

| | LWR | | | | GSOM | | | |
|---|---|---|---|---|---|---|---|---|
| | $E^{\text{flow}}$ | $E^{\text{speed}}$ | $E^{\text{density}}$ | **E** | $E^{\text{flow}}$ | $E^{\text{speed}}$ | $E^{\text{density}}$ | **E** |
| DirMed | 1.2452 | 1.3700 | 3.6954 | 6.3106 (+85%) | **0.9173** | **0.8785** | **1.6186** | **3.4144** |
| RTMC | **0.8833** | 1.9285 | 2.5484 | 5.3601 (+93%) | 1.1836 | **0.8630** | **0.7303** | **2.7769** |

Table 3: Error results for the optimization approach. In bold, the lowest flow, speed, density and total errors per data scenario.

Table 3 shows the error values for both data sets and both traffic flow models. The GSOM model has the smallest total error **E** for both data scenarios, whereas the error in the LWR model is at least 85% higher. As we already observed in Figures 7 and 8, the main error contributions for the LWR model come from the speed and density errors. These values are strikingly higher than the corresponding errors from the GSOM model. However, the two models give alternate results regarding the error with respect to the flow variable.

**Remark 5.** *Without bias estimation, the resulting estimation procedure reverts to the least squares approach. Appendix A presents the results of this more classical approach. However, the optimization approach outperforms this rather naive approach since every single error quantity in Table 3 is lower than the corresponding one from Table 8. This supports our decision to equip the model with a discrepancy term.*

## 5.2 Bayesian approach

We now apply the the Bayesian approach introduced in Section 4.2 to obtain posterior probability distributions of the model parameters. In the following, we introduce the formulas used in the Metropolis Algorithm 1.

- The prior distribution $\pi(\theta_{\text{VCR}})$ for the calibration parameters $\theta_{\text{VCR}} = (V, C, R)$ is given by a multivariate normal distribution, i.e.,

$$\pi(\theta_{\text{VCR}}) \propto \frac{1}{\sqrt{|\Sigma_{\theta_{\text{VCR}}}|}} \exp\Big\{-0.5\left(\theta_{\text{VCR}} - \mu_{\theta_{\text{VCR}}}\right)^{\top} \Sigma_{\theta_{\text{VCR}}}^{-1}\left(\theta_{\text{VCR}} - \mu_{\theta_{\text{VCR}}}\right)\Big\}$$

  with mean $\mu_{\theta_{\text{VCR}}}$ and covariance matrix $\Sigma_{\theta_{\text{VCR}}}$.

- The proposal distribution for $\hat{\theta}_{\text{VCR}}$ is defined as a multivariate normal distribution, i.e.,

$$\hat{\theta}_{\text{VCR}} \sim \mathcal{N}((\theta_{\text{VCR}})_{i-1}, \Sigma_{\text{VCR}}^{p})$$

  with covariance matrix $\Sigma_{\text{VCR}}^{p}$. In accordance with the optimization approach, we only consider proposal parameters $\hat{\theta}_{\text{VCR}}$ for which the number of projection occurrences in the numerical scheme is lower than 5% of the number of spatial discretization points.

- The sampling model for $y^F$ reads as

$$\mathcal{L}(y^F \mid \theta^*) \propto \frac{1}{\sqrt{|\mathbf{K}_n(\theta^*)|}} \exp\Big\{-0.5\left(b(\theta^*)^\top \mathbf{K}_n(\theta^*)^{-1} b(\theta^*)\right)\Big\}$$

with $\mathbf{K}_n(\theta^*) = \hat{\sigma}^2(\theta^*)\left(\mathbf{C}_n\left(l_1(\theta^*), l_2(\theta^*)\right) + g(\theta^*)\mathbf{I}_n\right)$
and $\hat{\sigma}^2(\theta^*) = \frac{b(\theta^*)^\top\left(\mathbf{C}_n\left(l_1(\theta^*), l_2(\theta^*)\right) + g(\theta^*)\mathbf{I}_n\right) b(\theta^*)}{n}$.
Inserting the process variance $\hat{\sigma}^2(\theta^*)$ into the covariance matrix $\mathbf{K}_n$, appearing in the exponential, yields the following simplification:

$$\mathcal{L}(y^F \mid \theta^*) \propto |\mathbf{K}_n(\theta^*)|^{-1/2}.$$

Thus, the posterior distribution $\pi(\theta^* \mid y^F)$ is computed by means of Bayes Theorem, i.e.,

$$\pi(\theta^* \mid y^F) \propto \mathcal{L}\left(y^F \mid \theta^*\right) \times \pi(\theta_{\text{VCR}}).$$

The choice of the mean vectors, covariance matrices and bounds for the uniform distributions can be found in Table 4.

| | $\mu_{\theta_{\text{VCR}}}$ | $\text{diag}(\Sigma_{\theta_{\text{VCR}}})$ | $\text{diag}(\Sigma^p_{\text{VCR}})$ | |
|---|---|---|---|---|
| | | | LWR | GSOM |
| DirMed | $(90, 30, 350)$ | $(100, 100, 2500)$ | $(6, 1.5, 90)$ | $(10, 5, 120)$ |
| RTMC | $(90, 40, 450)$ | $(56.3, 225, 5652)$ | $(8, 4, 100)$ | $(25, 12, 260)$ |

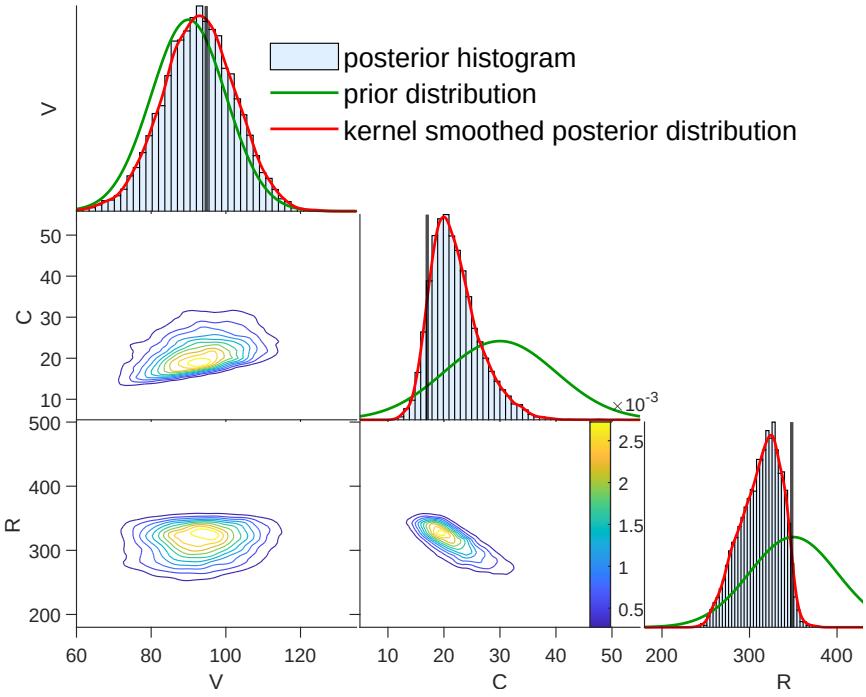
Table 4: Prior and proposal specifications applied in the Metropolis Algorithm 1 for the two data sets.

To apply the Metropolis Algorithm 1, we set the number of iterations to $N = 10^5$. It is a common approach to remove the first MCMC outputs in order to reduce the dependence of the proposal distribution on the initial guess. We set this *burn-in phase* to 10% of the $N$-iterations. Next, we reduce the sample chain to a number $\hat{N}$ minimizing autocorrelations. To this end, we use the multivariate effective sample size (ESS) function, `multiESS` in the `R` package `mcmcse` [19]. The effective sample sizes are listed in Table 5.

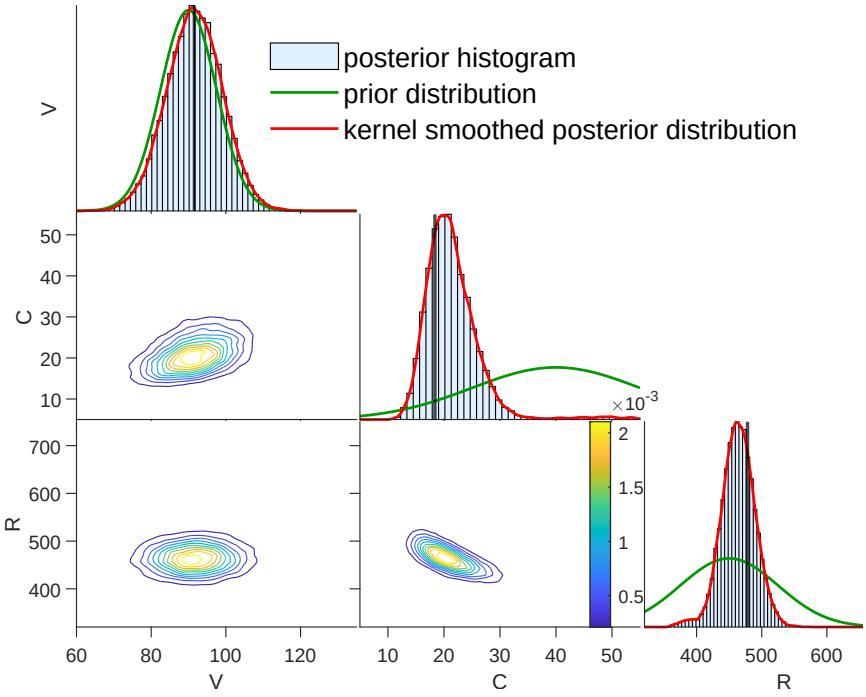
| | LWR | GSOM |
|---|---|---|
| DirMed | 720 | 11250 |
| RTMC | 1407 | 22500 |

Table 5: Effective sample sizes $\hat{N}$ for the two data sets.

For a graphical representation of the results of the MCMC method, we consider in Figure 9 both the histograms and the two-dimensional density contour plots, which are smoothed by a kernel density estimator. In the histogram graphics, we additionally add the probability density of the prior distribution for the calibration parameters (green line) and the kernel smoothed posterior distribution which is computed by the MATLAB command `fitdist` (red line). This operator fits a kernel probability distribution object to the sample data. The parameters for the kernel distribution object in MATLAB are chosen to be "normal" by default. Additionally, we visualize the parameter estimates obtained by the optimization approach (see Table 2 in Section 5.1) by a black vertical line.




(a) DirMed data




(b) RTMC data

Figure 9: Histograms and 2-dimensional density contour plots for the GSOM model. Black vertical line: parameter estimates obtained by the optimization approach (see Table 2).

We observe that, for both data sets, the speed calibration parameter $V$ has the largest variance among the V, C, R parameters (despite being the most intuitive one to specify *a-priori*) and the kernel smoothed posterior distribution is close to the prior distribution. In comparison, the posterior distributions for $C$ and $R$ are more peaked and shifted in one direction compared to the prior ones. In general, we detect a negative correlation between $C$ and $R$, visible by the diagonal shape of the $C - R$ contours. We conclude that lower $C$ values correspond to higher $R$ values and vice-versa, whereas the $V$ parameter seems uncorrelated with the other ones. Moreover, we detect a second less pronounced peak for the $C$ and $R$ parameter in the RTMC data case. Finally, comparing the loglikelihood-value of the histogram's mode parameter with the one computed by the optimization approach, we obtain values in a similar range (absolute difference is lying in the range of $10^{-3}$).

For a performance comparison between the two traffic flow models and also between the two calibration approaches, we again define an error metric for the simulated MCMC outputs. In formulas, this reads for the flow error

$$E^{\text{flow}}_{\text{MCMC}} = \frac{1}{T_f \cdot (x_{out} - x_{in})} \frac{1}{\Delta q} \sum_{(x,t)\in(X,T)} \left( \left| y^F_{\text{flow}}(x,t) - \mathbb{E}[\hat{y}^M_{\text{flow}}(x,t)] \right| \right)$$

$$\text{with } \mathbb{E}[\hat{y}^M_{\text{flow}}(x,t)] = \mathbb{E}_\theta \left[ \mathbb{E}[\hat{y}^M_{\text{flow}}(x,t) \mid \theta^*] \right] \approx \frac{1}{\hat{N}} \sum_{i=1}^{\hat{N}} \left( y^M_{\text{flow}}(x,t,\theta^*_i) + m_n(x,t,\theta^*_i) \right),$$

where $\hat{y}^M_{\text{flow}}(x,t,\theta^*_i)$ is defined as the output of the simulation code evaluated at the $i^{th}$ optimal calibration parameter $\theta^*_i$ (for $i \in \{1,\dots,\hat{N}\}$) at time $t$ and loop position $x$, corrected by adding its kriging mean $m_n$. This error metric both quantifies the accuracy of the model and is coherent with the one used for the optimization approach.

Analogously, we define the speed (resp. density) error $E^{\text{speed}}_{\text{MCMC}}$ (resp. $E^{\text{density}}_{\text{MCMC}}$) by using the measured and simulated speed (resp. density) values and by using the normalization constant $\Delta v$ (resp. $\Delta\rho$). Again, we emphasize that we correct all simulated traffic quantities $\hat{y}^M$ by their predictive means (4.1b).
Finally, the total MCMC cost error is given by the sum of the previously defined errors, i.e.,

$$\mathbf{E}_{\text{MCMC}} = E^{\text{flow}}_{\text{MCMC}} + E^{\text{speed}}_{\text{MCMC}} + E^{\text{density}}_{\text{MCMC}}.$$

The error values for the two traffic flow models are listed in Table 6.

| | LWR | | | | GSOM | | | |
|---|---|---|---|---|---|---|---|---|
| | $E^{\text{flow}}_{\text{MCMC}}$ | $E^{\text{speed}}_{\text{MCMC}}$ | $E^{\text{density}}_{\text{MCMC}}$ | $\mathbf{E}_{\text{MCMC}}$ | $E^{\text{flow}}_{\text{MCMC}}$ | $E^{\text{speed}}_{\text{MCMC}}$ | $E^{\text{density}}_{\text{MCMC}}$ | $\mathbf{E}_{\text{MCMC}}$ |
| DirMed | 1.0560 | 1.2720 | 4.3358 | 6.6639 (+101%) | **0.9290** | **0.8265** | **1.5666** | **3.3221** |
| RTMC | **0.8961** | 2.5066 | 2.2182 | 5.6209 (+111%) | 1.1662 | **0.6874** | **0.8130** | **2.6666** |

Table 6: Time-space error results for the Bayesian approach. In bold, the lowest flow, speed, density and total errors per data scenario.

We clearly see that the total error $\mathbf{E}_{\text{MCMC}}$ is at least 101% higher for the first order model. Additionally, the speed and density predictions of the GSOM model are more accurate for all considered scenarios. As observed with the optimization approach, we remark again that although the calibration is based only on the flow data, the corrected simulated speed and density outputs lead to good prediction accuracy results in the second order model, whereas the flow error is similar for the two traffic models.

**Remark 6.** *In [16, 17], the authors consider only the error on the speed and the density. For these quantities, they come to the same conclusion, namely that GSOM outperforms the LWR model. However, the flow error is neglected in their evaluations, although they construct their data-fitted model on the measured flow-density fundamental diagram.*

We have seen that the Bayesian approach provides us with more information than the direct single output approach. Regarding the numerical error computations, we can compare the entries of Tables 3 (optimization approach) and 6 (Bayesian approach). We observe that the total error in the Bayesian approach $\mathbf{E}_{\text{MCMC}}$ is slightly lower for the GSOM model but a bit higher for the LWR model. As for the optimization approach, the flow error does not perform better for any of the two models, and we cannot detect a clear pattern on the performance accuracy: the error is lower in some cases for the optimization approach and in other cases for the Bayesian approach.

# 6 Conclusion

We have applied a Bayesian calibration technique for parameter identification and uncertainty quantification in macroscopic road traffic models, exploiting different loop detector data sets. The study has highlighted the globally better performances of second order compared to first order models.
The proposed approach results in better reconstruction performances than direct calibration techniques commonly used in practice, which moreover do not consider parameter uncertainty. In particular, our results point out the benefit of introducing a bias term to compensate model limitations in reproducing real data.

From the traffic modeling point of view, further investigations should consider more complex situations including the presence of on- and off-ramps and road junctions, traffic lights, etc. Also, time or space dependence could be considered for some parameters [43], as well as local variations of the bias on the road, possibly depending on the traffic regimes [32].

Regarding the calibration techniques, in this paper we have applied a modularized version of KOH calibration [37, 52], alleviating some of the shortcomings of the original approach. Given the flexibility of this framework, it can be improved in many ways: either by reducing identifiability issues with orthogonality constraints as in [45], or by increasing the scalability with the deployment of deep GPs in a variational inference scheme [38]. Rather than reconstructing the traffic conditions like here, predicting future conditions will also require improving the prior information on the model discrepancy, as discussed by [4]. Refining the model comparison, e.g. via efficient computation of Bayes factors, is another topic for further research.

# A Least square approach

A commonly used approach to calibrate the optimal parameters is the minimization of a least square cost function taking both the real data and simulated data into account, see e.g. [42, 50, 53]. Accordingly, the calibration is based on the minimization of the following cost function

$$C(\theta) = \sum_{(x,t)\in(X,T)} \left| y^F(x,t) - y^M(x,t,\theta) \right|^2 .$$

Thus, the optimal parameter $\theta^*$ is given by

$$\theta^* = \underset{\theta \in \Theta}{\operatorname{argmin}}\, C(\theta).$$

The bounds for the three dimensional parameter space $\Theta$ are those defined in the right columns of Table 1. The optimization results for the calibration parameters are summarized in Table 7.

| | LWR | | | GSOM | | |
|---|---|---|---|---|---|---|
| | $V^*$ | $C^*$ | $R^*$ | $V^*$ | $C^*$ | $R^*$ |
| DirMed | 86.74 | 20.05 | 389.80 | 85.78 | 22.54 | 390.00 |
| RTMC | 76.18 | 32.35 | 435.25 | 106.51 | 21.42 | 488.19 |

Table 7: Optimization results for the least squares approach.

Comparing the errors reported in Table 8 with those of Tables 3 and 6, we conclude that both the optimization and the Bayesian approaches greatly outperform this basic calibration procedure, thus evidencing the benefit of introducing a bias term.

| | LWR | | | | GSOM | | | |
|---|---|---|---|---|---|---|---|---|
| | $E^{\text{flow}}$ | $E^{\text{speed}}$ | $E^{\text{density}}$ | **E** | $E^{\text{flow}}$ | $E^{\text{speed}}$ | $E^{\text{density}}$ | **E** |
| DirMed | **1.3365** | 10.2647 | 24.1959 | 35.7971 (+303%) | 2.2918 | **3.0850** | **3.4981** | **8.8749** |
| RTMC | **1.5926** | 10.6936 | 6.8826 | 29.1688 (+294%) | 2.9600 | **1.8807** | **2.5602** | **7.4009** |

Table 8: Time-space error results for the least squares approach. In bold, the lowest flow, speed, density and total errors per data scenario.

# References


[1] A. Aw and M. Rascle. Resurrection of "second order" models of traffic flow. SIAM J. Appl. Math., 60(3):916–938, 2000.

[2] M. J. Bayarri, J. O. Berger, M. C. Kennedy, A. Kottas, R. Paulo, J. Sacks, J. A. Cafeo, C.-H. Lin, and J. Tu. Predicting vehicle crashworthiness: Validation of computer models for functional and hierarchical data. Journal of the American Statistical Association, 104(487):929–943, 2009.

[3] E. Bertino, R. Duvigneau, and P. Goatin. Uncertainty quantification in a macroscopic traffic flow model calibrated on GPS data. Math. Biosci. Eng., 17(2):1511–1533, 2020.

[4] J. Brynjarsdóttir and A. O'Hagan. Learning about physical parameters: The importance of model discrepancy. Inverse problems, 30(11):114007, 2014.

[5] M. Carmassi, P. Barbillon, M. Chiodetti, M. Keller, and E. Parent. Bayesian calibration of a numerical code for prediction. Journal de la société française de statistique, 160(1):1–30, 2019.

[6] C. Chalons and P. Goatin. Transport-equilibrium schemes for computing contact discontinuities in traffic flow modeling. Commun. Math. Sci., 5(3):533–551, 2007.

[7] N. Chiabaut, C. Buisson, and L. Leclercq. Fundamental diagram estimation through passing rate measurements in congestion. IEEE Transactions on Intelligent Transportation Systems, 10(2):355–359, 2009.

[8] F. A. Chiarello, J. Friedrich, P. Goatin, and S. Göttlich. Micro-macro limit of a nonlocal generalized Aw-Rascle type model. SIAM J. Appl. Math., 80(4):1841–1861, 2020.

[9] R. M. Colombo and A. Groli. On the initial boundary value problem for Temple systems. Nonlinear Anal., 56(4):569–589, 2004.

[10] A. Corbetta, A. Muntean, and K. Vafayi. Parameter estimation of social forces in pedestrian dynamics models via a probabilistic method. Math. Biosci. Eng., 12(2):337–356, 2015.

[11] I. Crandell, A. J. Millican, S. Leman, E. Smith, W. N. Alexander, W. J. Devenport, R. Vasta, R. Gramacy, and M. Binois. Anomaly detection in large-scale wind tunnel tests using Gaussian processes. In 33rd AIAA Aerodynamic Measurement Technology and Ground Testing Conference, page 4131, 2017.

[12] C. F. Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Res. Part B, 28(4):269–287, 1994.

[13] S. F. Davis. Simplified second-order Godunov-type methods. SIAM J. Sci. Statist. Comput., 9(3):445–473, 1988.

[14] M. L. Delle Monache, K. Chi, Y. Chen, P. Goatin, K. Han, J.-M. Qiu, and B. Piccoli. A three-phase fundamental diagram from three-dimensional traffic data. Axioms, 10(1):17, 2021.

[15] G. Dervisoglu, G. Gomes, J. Kwon, R. Horowitz, and P. Varaiya. Automatic calibration of the fundamental diagram and empirical observations on capacity. In Transportation Research Board 88th Annual Meeting, volume 15, pages 31–59. Citeseer, 2009.

[16] S. Fan. Data-fitted generic second order macroscopic traffic flow models. ProQuest LLC, Ann Arbor, MI, 2013. Thesis (Ph.D.)–Temple University.

[17] S. Fan, M. Herty, and B. Seibold. Comparative model accuracy of a data-fitted generalized Aw-Rascle-Zhang model. Netw. Heterog. Media, 9(2):239–268, 2014.

[18] S. Fan, Y. Sun, B. Piccoli, B. Seibold, and D. B. Work. A collapsed generalized Aw-Rascle-Zhang model and its model accuracy, 2017.

[19] J. M. Flegal, J. Hughes, D. Vats, and N. Dai. mcmcse: Monte Carlo Standard Errors for MCMC. Riverside, CA, Denver, CO, Coventry, UK, and Minneapolis, MN, 2020. R package version 1.4-1.

[20] R. Franklin. The structure of a traffic shock wave. Civil Engineering Pulb. Wks. Rev, 56:1186–1188, 1961.

[21] S. Gerster, M. Herty, and E. Iacomini. Stability analysis of a hyperbolic stochastic Galerkin formulation for the Aw-Rascle-Zhang model with relaxation. arXiv preprint arXiv:2102.09359, 2021.

[22] S. K. Godunov. A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. Mat. Sb. (N.S.), 47 (89):271–306, 1959.

[23] S. N. Gomes, A. M. Stuart, and M.-T. Wolfram. Parameter estimation for macroscopic pedestrian dynamics models from microscopic data. SIAM J. Appl. Math., 79(4):1475–1500, 2019.

[24] S. Göttlich and S. Knapp. Artificial neural networks for the estimation of pedestrian interaction forces. In L. Gibelli, editor, Crowd Dynamics, Volume 2: Theory, Models, and Applications, pages 11–32. Springer International Publishing, Cham, 2020.

[25] S. Göttlich and C. Totzeck. Optimal control for interacting particle systems driven by neural networks. arXiv preprint arXiv:2101.12657, 2021.

[26] R. B. Gramacy. Surrogates: Gaussian process modeling, design, and optimization for the applied sciences. CRC Press, 2020.

[27] A. Harten, P. D. Lax, and B. van Leer. On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. SIAM Rev., 25(1):35–61, 1983.

[28] M. Herty, A. Fazekas, and G. Visconti. A two-dimensional data-driven model for traffic flow on highways. Netw. Heterog. Media, 13(2):217–240, 2018.

[29] M. Herty, A. Tosin, G. Visconti, and M. Zanella. Reconstruction of traffic speed distributions from kinetic models with uncertainties. arXiv preprint arXiv:1912.03706, 2019.

[30] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne. Combining field data and computer simulations for calibration and prediction. SIAM Journal on Scientific Computing, 26(2):448–466, 2004.

[31] P. D. Hoff. A first course in Bayesian statistical methods, volume 580. Springer, 2009.

[32] J. Huang, R. B. Gramacy, M. Binois, and M. Libraschi. On-site surrogates for large-scale calibration. Applied Stochastic Models in Business and Industry, 36(2):283–304, 2020.

[33] M. Iglesias and A. M. Stuart. Inverse problems and uncertainty quantification. SIAM News, pages 2–3, 2014.

[34] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(3):425–464, 2001.

[35] J.-P. Lebacque, S. Mammar, and H. H. Salem. Generic second order traffic flow modelling. In Transportation and Traffic Theory 2007. Papers Selected for Presentation at ISTTT17Engineering and Physical Sciences Research Council (Great Britain) Rees Jeffreys Road FundTransport Research FoundationTMS ConsultancyOve Arup and Partners, Hong KongTransportation Planning (International) PTV AG, 2007.

[36] M. J. Lighthill and G. B. Whitham. On kinematic waves. II. A theory of traffic flow on long crowded roads. Proc. Roy. Soc. London Ser. A, 229:317–345, 1955.

[37] F. Liu, M. Bayarri, J. Berger, et al. Modularization in Bayesian analysis, with emphasis on analysis of computer models. Bayesian Analysis, 4(1):119–150, 2009.

[38] S. Marmin and M. Filippone. Variational calibration of computer models. arXiv preprint arXiv:1810.12177, 2018.

[39] Minnesota Departement of Transportation. Mn/Dot Traffic Data. Website: `http://data.dot.state.mn.us/datatools/`.

[40] S. Mollier, M. L. Delle Monache, C. C. de Wit, and B. Seibold. Two-dimensional macroscopic model for large scale traffic networks. Transportation Research Part B: Methodological, 122:309–326, 2019.

[41] G. Newell. A theory of traffic flow in tunnels. Theory of Traffic Flow, pages 193–206, 1961.

[42] D. Ngoduy and S. Hoogendoorn. An automated calibration procedure for macroscopic traffic flow models. IFAC Proceedings Volumes, 36(14):263–268, 2003. 10th IFAC Symposium on Control in Transportation Systems 2003, Tokyo, Japan, 4-6 August 2003.

[43] M. Pereira, P. B. Baykas, B. Kulcsár, and A. Lang. Parameter and density estimation from real-world traffic data: A kinetic compartmental approach. arXiv preprint arXiv:2101.11485, 2021.

[44] B. Piccoli, K. Han, T. L. Friesz, T. Yao, and J. Tang. Second-order models and traffic data from mobile sensors. Transportation Research Part C: Emerging Technologies, 52:32–56, 2015.

[45] M. Plumlee. Bayesian calibration of inexact computer models. Journal of the American Statistical Association, 112(519):1274–1285, 2017.

[46] N. Polson and V. Sokolov. Bayesian analysis of traffic flow on interstate I-55: The LWR model. The Annals of Applied Statistics, 9(4):1864 – 1888, 2015.

[47] C. E. Rasmussen and C. Williams. Gaussian Processes for Machine Learning. MIT Press, 2006.

[48] P. I. Richards. Shock waves on the highway. Operations Res., 4:42–51, 1956.

[49] M. L. Stein. Interpolation of spatial data: some theory for kriging. Springer Science & Business Media, 2012.

[50] G. Strofylas, K. Porfyri, I. Nikolos, A. Delis, and M. Papageorgiou. Using synchronous and asynchronous parallel differential evolution for calibrating a second-order traffic flow model. Advances in Engineering Software, 125:1–18, 2018.

[51] B. Temple. Systems of conservation laws with invariant submanifolds. Trans. Amer. Math. Soc., 280(2):781–795, 1983.

[52] R. Tuo and C. Jeff Wu. A theoretical framework for calibration in computer models: parametrization, estimation and convergence properties. SIAM/ASA Journal on Uncertainty Quantification, 4(1):767–795, 2016.

[53] P. Wagner. Fluid-dynamical and microscopic description of traffic flow: a data-driven comparison. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 368(1928):4481–4495, 2010.

[54] Y. Wang and M. Papageorgiou. Real-time freeway traffic state estimation based on extended Kalman filter: a general approach. Transportation Research Part B: Methodological, 39(2):141–167, 2005.

[55] D. B. Work, S. Blandin, O.-P. Tossavainen, B. Piccoli, and A. M. Bayen. A traffic model for velocity data assimilation. Appl. Math. Res. Express. AMRX, pages 1–35, 2010.

[56] H. M. Zhang. A non-equilibrium traffic model devoid of gas-like behavior. Transportation Res. Part B, 36(3):275–290, 2002.