CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Markov chain Monte Carlo for a hyperbolic Bayesian inverse problem in traffic flow modeling

Jeremie Coullon[1,*] and Yvo Pokern[2]

[1]Mathematics and Statistics, Lancaster University, Bailrigg, Lancaster LA1 4YW, United Kingdom
[2]Statistical Science, University College London, Gower Street, London WC1E 6BT, United Kingdom
*Corresponding author. E-mail: jeremie.coullon@gmail.com

Abstract

As a Bayesian approach to fitting motorway traffic flow models remains rare in the literature, we empirically explore the sampling challenges this approach offers which have to do with the strong correlations and multimodality of the posterior distribution. In particular, we provide a unified statistical model to estimate using motorway data both boundary conditions and fundamental diagram parameters in a motorway traffic flow model due to Lighthill, Whitham, and Richards known as LWR. This allows us to provide a traffic flow density estimation method that is shown to be superior to two methods found in the traffic flow literature. To sample from this challenging posterior distribution, we use a state-of-the-art gradient-free function space sampler augmented with parallel tempering.

Impact Statement

Accurately modelling motorway traffic is essential in understanding and predicting such features as traffic jams and travel time. Consequently, models have been developed which use conservation of mass expressed through partial differential equations. These are difficult to fit to traffic flow data making it hard to assess their accuracy and to compare them. One difficulty is in estimating traffic density which is a variable in the models but cannot be directly measured. Here, we provide a statistical method to estimate the parameters jointly with traffic density in the Lighthill–Whitham–Richards model. Our Bayesian methodology allows us to quantify the inferences' uncertainties and results in a density estimation method that is superior to two methods from the traffic flow literature.

1. Introduction Cl

Fitting differential equations to data—a type of inverse problem—is an essential part of modelling in the sciences and engineering. It allows researchers to model complex systems to be able to understand and predict their behavior. We can find applications in such varied fields as Geophysics (Sambridge, 2014), Hydrogeology (Iglesia et al., 2014), and fluid mechanics (Cotter et al., 2009).

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

*The online version of this article has been updated since original publication. A notice detailing the change has also been published.

CrossMark

In this article, we consider the Bayesian framework as outlined in Stuart (2010) for partial differential equations (PDEs). We consider observations y generated by an observation operator $\mathcal{G}$ polluted with noise η:

$$y = \mathcal{G}(\psi) + \eta. \tag{1}$$

Starting from a prior belief on the parameter ψ, the objective is to update this distribution based on observations represented by the likelihood.

For nonlinear PDEs such as the one studied in this article, the posterior is unavailable analytically, and so one must resort to numerical methods such as sampling methods (Markov chain Monte Carlo [MCMC] and Sequential Monte Carlo [SMC]) (see Liu, 2001 or Brooks et al., 2011). In this article, we use MCMC.

We consider as application vehicular traffic flow on motorways. Many approaches have been considered in the traffic flow literature to model such a system: systems of ordinary differential equations (Bando et al., 1995; Gasser et al., 2004), stochastic differential equations (Mahnke et al., 2005), or PDEs (Lighthill and Whitham, 1955; Aw and Rascle, 2000; Zhang, 2000). We focus on the most well-known model, the Lighthill–Whitham–Richards (LWR) model, which is a conservation law with a nonlinear flux function:

$$\rho_t + (\rho V_e(\rho))_x = 0. \tag{2}$$

With density ρ and $V_e(\rho)$ the equilibrium velocity. We use subscripts to denote partial derivatives. Using the fundamental relationship relating flow q to density ρ and speed v: $q = \rho v$, we can define the fundamental diagram (FD) $Q_e(\rho) = \rho V_e(\rho)$.

The main contributions of the article are as follows: as a rigorous Bayesian treatment of motorway traffic flow models is rare in the literature, we empirically explore the sampling challenges this offers. In particular, we provide a unified statistical model to estimate both boundary conditions (BCs) and FD parameters in LWR. This allows us to provide a traffic flow density estimation method that is shown to be superior to two methods found in the traffic flow literature. Finally, we highlight how fitting these models is now tractable due to recent advances in gradient-free function space MCMC samplers.

The structure of the article is as follows. In Section 2, we give an overview of traffic flow theory, the LWR model, the traffic data, and previous work in this area. In Section 3, we fit the FD parameters directly to data using MCMC without considering the PDE model which is the usual approach in the traffic flow literature. In Section 4, we sample from the FD parameters and the BCs within LWR. To overcome the sampling difficulties of this inverse problem we use a state-of-the-art gradient-free function space MCMC sampler and parallel tempering (PT). We also compare this methodology to other methods both in the traffic engineering and statistical literature.

2. Motorway Traffic Flow and the LWR Model

2.1. Traffic flow on the motorway

We start by showing in Figure 1 traffic density (the number of vehicles per km) in space and time for a 5 km stretch of M25 motorway in the UK for 49 min. In this plot, time is denoted on the horizontal axis and space (distance on the road) on the vertical axis. Vehicles move from distance 0 to 5 km (namely upward in the plot) and forward in time (namely to the right in the plot) so therefore move diagonally (upward and to the right). We can see the vehicles' movement in the first few minutes (approximately between minutes 381 and 405) where they take a few minutes to cover the 5 km section of road. This is consistent with a vehicle speed of approximately 120 km/hr (i.e., 2 km/min). This vehicle movement corresponds to the free flow waves. We also observe backward moving waves in the second half of the $x - t$ plot (from around minute 405 until the end) which are high-density waves. These backward moving waves correspond to the experience of needing to brake sharply when driving on motorways to avoid crashing into a traffic jam: congested flow waves move upstream in traffic.
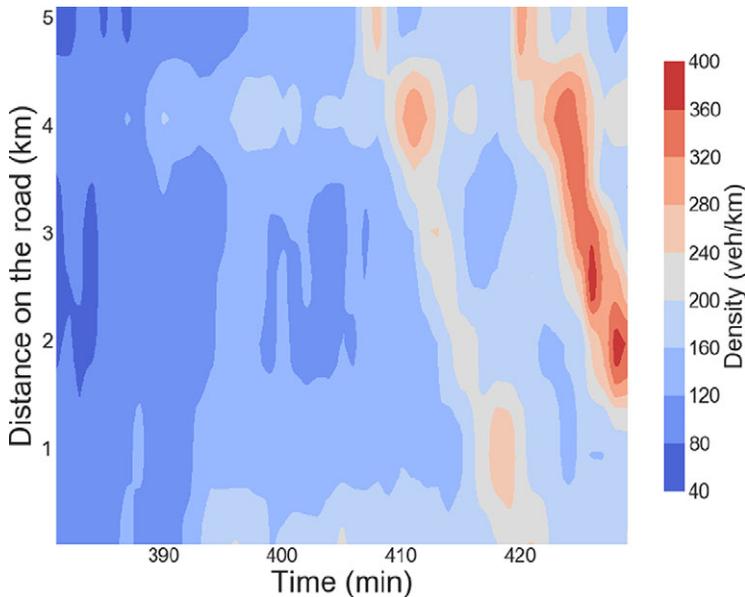
Figure 1. *Density estimated from occupancy for the section of M25 on January 8, 2007 between 6:21 am and 7:09 am. We observe forward moving free flow waves between minutes 381 and 405, which correspond to the movement of vehicles. We also observe backward moving high-density waves in the second half of the $x - t$ plane.*
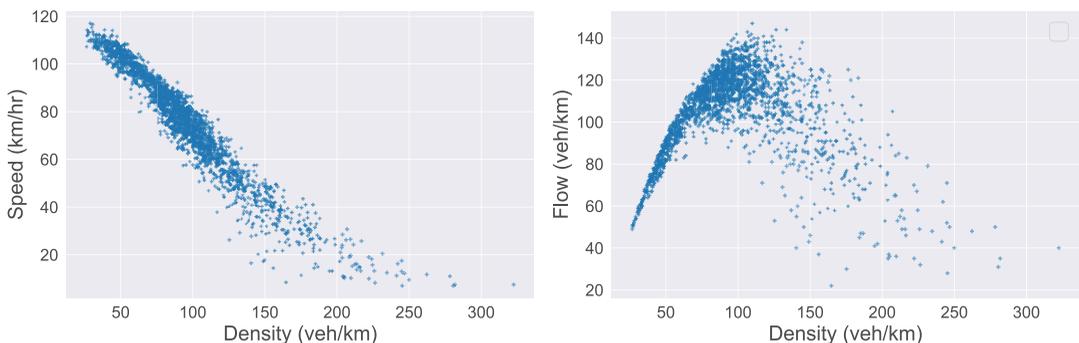

Figure 2. *Figures showing the empirical relationship between vehicle flow q, density ρ, and velocity v. The data is taken from detectors on the M25 measuring various quantities averaged every minute.*

2.2. The fundamental diagram

We now observe flow and density traffic data without the time component (flow being the number of vehicles per minute) as can be seen in Figure 2, and noting that for the flow-density plot (the right-most plot) the measurements on the left of the plot can be well described by a curve with positive slope. We define traffic following this curve to be in the state of *free flow.* The rest of the points are scattered to the right of this curve which we call *congested flow.* A main difference is that vehicles do not interact much in free flow; adding a vehicle does not decrease the speed of the others. In contrast, adding a vehicle to congested flow will decrease the speed of the other vehicles as they are forced to slow down to avoid collisions. We can see the effect of increasing density on average vehicle speed in the left-most plot of Figure 2.

The FD, first studied in Greenshields (1934) is the function used to describe the relationship between density and flow as seen in Figure 2. It is used to close the conservation of mass equation in PDE models of traffic, as will be described later.

Three variables are usually considered in the PDE approach of modeling traffic flow: density ρ which describes the number of vehicles per unit length, speed v which describes the average speed of vehicles at a point, and flow q which is the number of vehicles that pass a point in space during a unit of time. These three variables are linked through the following relationship:

$$q = \rho v. \tag{3}$$

A popular example is Daganzo's *Triangular FD* (also called the bilinear FD) from Daganzo (1994) given in Equation (4) and plotted in Figure 3a. This FD has important traffic flow quantities as parameters: the capacity q_c (maximum possible value of flow), the critical density ρ_c which separates free flow from congested flow, and the jam density ρ_j (maximum possible value of density). This FD is popular for its simplicity as well as for its computational efficiency when used in PDE models such as LWR (introduced in Section 2.2.1). We will see later how the wave speed propagation of traffic flow models depends on the shape of the FD.

$$q(\rho) = \begin{cases} \dfrac{q_c}{\rho_c}\rho & \rho < \rho_c \\ q_c \dfrac{\rho_j - \rho}{\rho_j - \rho_c} & \rho \geq \rho_c \end{cases} \tag{4}$$

We will focus in this article on the FD introduced by del Castillo (2012) called the negative power model (which we will simply call *del Castillo's FD*). This FD has four parameters (all in $(0, \infty)$): flow scaling term Z, jam density ρ_j, shape parameter γ, and parameter u relating to the critical density. The shape parameter γ determines the tightness of the peak and in the limit of $\gamma \to \infty$ we obtain the Triangular FD defined above. The parameter Z determines the vertical scaling of the FD, and u is related to the critical density via the following relation:

$$\rho_c = \frac{1}{1 + u^{\gamma/(\gamma+1)}}. \tag{5}$$
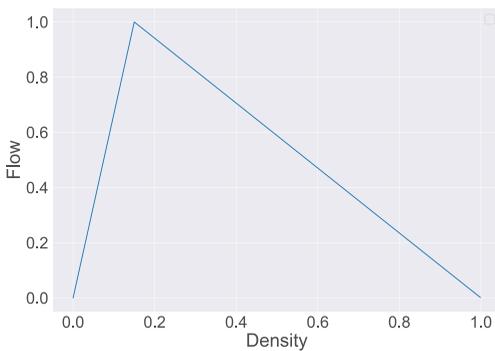
(a) The Triangular FD
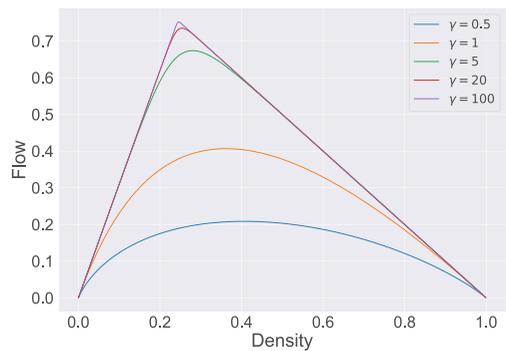
(b) Del Castillo's FD

Figure 3. *(a) Daganzo's triangular fundamental diagram (FD) plotted for dimensionless flow and density with $(q_c, \rho_c, \rho_j) = (1, 0.15, 1)$. (b) Del Castillo's FD plotted for dimensionless flow and density with $(Z, \rho_j, u) = (1, 1, 3.1)$ and for $\gamma \in [0.5, 1, 5, 20, 100]$.*

We give the FD in Equation (6) below and plot the FD for several values of γ in Figure 3b to illustrate how del Castillo's FD includes the Triangular FD as a limiting case.

$$q(\rho) = Z \left[\left(u \frac{\rho}{\rho_j} \right)^{-\gamma} + \left(1 - \frac{\rho}{\rho_j} \right)^{-w} \right]^{-\frac{1}{\gamma}} \tag{6}$$

2.2.1. Lighthill–Whitham–Richards

As mentioned in Section 1, one of the ways to model traffic flow is to use PDEs. This approach describes the macroscopic behavior of traffic flow emerging from the individual interactions between vehicles, so we consider traffic flow as a continuum without representing the individual behavior of vehicles. The differences between these models will therefore be in the assumptions made in how the system acts. We point to Hoogendoorn and Bovy (2001) and Bonzani and Gramani (2009) for a critical review on the topic.

An important class of these PDE models is those that consist of a single equation, which corresponds to a conservation law of the form:

$$\rho_t + (\rho v)_x = 0. \tag{7}$$

As mentioned earlier, we use the subscripts ρ_t to denote $\frac{\partial \rho}{\partial t}$ and ρ_x to denote $\frac{\partial \rho}{\partial x}$. This conservation equation is derived from the obvious fact that vehicles are conserved on a length of road (assuming there are no on or off-ramps) (see Leveque, 2004 for a derivation). Different assumptions in how the system acts will lead to different ways of closing this conservation equation.

A standard way of doing this is to use the following form for the velocity: $v = V_e(\rho)$ (see Coullon, 2019 for an overview of PDE models in traffic flow). This assumes that vehicles adapt their speed instantaneously to a change in local density; that is, traffic flow is always in equilibrium described by the function $V_e(\rho)$. Multiplying this function by the density gives the FD $q = \rho V_e(\rho)$. This model is named the LWR model after Lighthill and Whitham (1955) and Richards (1956) who developed it independently. Note that although these authors used a particular parametric form for the FD, the traffic flow literature considers the name LWR to denote the conservation law while being agnostic of the FD used, see Hoogendoorn and Bovy (2001).

A property of the LWR model is that it allows the formation of *shock waves* (discontinuities in the solution); more generally, nonlinear hyperbolic PDEs are prone to forming shocks. As a result, LWR with this choice of FD can model the propagation of upstream and downstream shock fronts on the highway. We can see these shock fronts in Figure 1: the forward moving waves occur for free-flow traffic (namely for low density) and the backward moving waves occur for congested flow traffic (namely for high density).

In hyperbolic PDEs, these shock fronts will move at speeds depending on the density and this is given by the gradient of the FD at that value of density. More generally, a disturbance in the initial condition of the PDE moves along *characteristics*: curves in the $x - t$ plane where the solution is constant (with speed also given by the gradient of the FD). So using the FDs in Figure 5, we can see that in free flow we have $q'(\rho) > 0$ so characteristics have positive speed, while in the congested regime they have negative speed.

To solve LWR we need to choose the FD along with its parameters, and choose some initial conditions (density along the road at time 0) and BCs (density at the inlet and outlet of the road for the entire time period of the simulation).

2.3. Numerical method

We would like to solve LWR for arbitrary FD parameters, initial conditions, and BCs. However, we cannot solve it analytically in the general case. We will use the open-source software Clawpack (Clawpack Development Team, 2018) which is a package for solving conservation laws using finite volume methods

(see Mandli et al., 2016 for information about the 5.0 release). Furthermore, an introduction to these finite volume methods along with an overview of the software can be found in Leveque (2004).

We give here an overview of the Godunov method (a first-order finite volume method) and point out a limitation. The idea of this method is to solve the PDE in conservation form (as in Equation (8)), which ensures that the method behaves correctly in the presence of shock waves:

$$\rho_t + f(\rho)_x = 0. \tag{8}$$

We discretize space and time into cells, and consider methods of the form:

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{\Delta x}\left(F_{i+1/2}^n - F_{i-1/2}^n\right). \tag{9}$$

with

- ρ_i^n density at cell i, time n,
- $F_{i-1/2}^n$ flux (flow) at the left boundary of cell i at time n (computed according to equation (12.4) in Leveque, 2004),
- $F_{i+1/2}^n$ flux (flow) at the right boundary of cell i at time n,
- Δt and Δx the time and space discretization.

If we consider the density to be constant in each cell, we obtain a Riemann Problem at each boundary: an initial value problem with piecewise constant data and a single discontinuity. We then solve this Riemann Problem—which can be done analytically—at each boundary to find $F_{i+1/2}^n$ and $F_{i-1/2}^n$.

In the special case of $f''(q)$ not changing sign, there are two main cases in the Riemann Problem (see Leveque, 2004 for dealing with the more general case):

Case 1: $\rho_{i-1}^n < \rho_i^n$

In this case, we will have a shock wave as seen in *Figure 5*: the discontinuity will simply be advected with speed: $v_{\text{shock}} = \frac{f(\rho_i)-f(\rho_{i-1})}{\rho_i-\rho_{i-1}}$.

Case 2: $\rho_{i-1}^n > \rho_i^n$

Here, we will have a rarefaction wave as seen in *Figure 4*.

The case of linear PDEs is straightforward: methods can be developed that update each cell by looking in which direction information is coming from (by using the characteristic speed) and updating the flux of the cell accordingly. However for nonlinear models such as LWR, information can come from both sides (i.e., characteristics can cross), so this approach needs to be extended. In this case, Godunov's method can deal with characteristics crossing to resolve this issue (see Leveque, 2004 for details).

To deal with the BCs we simply extend the domain with an extra cell (called a "ghost cell") and set a chosen value of density on this ghost cell. The benefit of this approach is that no special method is needed to deal with the BCs; the ghost cells are used to update the cells at the edge of the domain.

Creating methods that operate on the PDE in conservation form ensures that we can accurately model the position of the shock wave at any point in time and space. However, they introduce a great deal of numerical viscosity that smooths out the solution. One can then extend Godunov's method to create second-order numerical methods such as the Lax–Wendroff method (see Leveque, 2004 for a detailed account) that models smooth solutions more accurately than the first-order Godunov scheme but fails at discontinuities. To correct this, one needs to add so-called *flux limiters*. Using the Clawpack software, one only needs to solve the Riemann problem at each cell and the software automatically uses a Lax–Wendroff method with a flux limiter (the default one used is called the *minmod* limiter).

However, these methods still do exhibit some numerical viscosity. To illustrate this effect, we consider a Riemann problem (case 1 above) and compare the true solution of LWR with del Castillo's FD for two different times to the output of Clawpack in Figure 5. This numerical viscosity is therefore expected to cause the posterior to be slightly different than if LWR was solved exactly. On the other hand, even though we observe jumps in density in motorway traffic, we do not expect them to be truly discontinuous: we
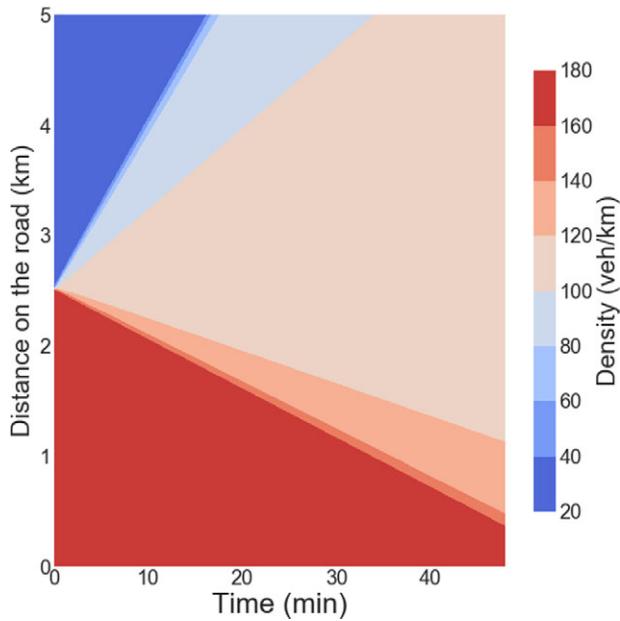
Figure 4. *The Riemann Problem with $\rho_{i-1} > \rho_i$ causes a rarefaction wave. The initial condition (namely at $t = 0$) consists of a constant value of high density for $x \in [0, 2.5]$ and a constant value of low density for $x \in [2.5, 5]$. As the simulation moves forward in time we observe a rarefaction wave, or a fanning out of density values between the low and high values of the initial condition.*
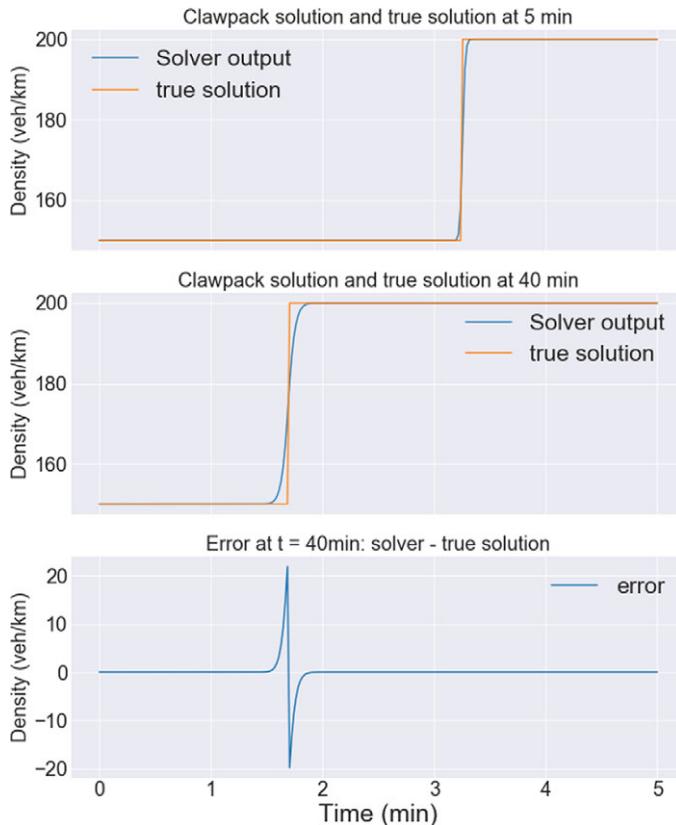

Figure 5. *We plot the analytic solution to the Riemann Problem along with its numerical solution using Clawpack. As time progresses we observe that the discontinuity is smoothed slightly. However, we notice that the position of the shock wave remains accurate.*

rather expect them to be rapid but smooth transitions of density. Formally investigating the effect of this error would be an interesting area of research.

When defining the solver we must also choose the resolution Δx and Δt such that the *CFL condition* (Courant–Friedrichs–Lewy)—a necessary condition for convergence—is satisfied (see Leveque, 2004). To understand this condition, first, we recall that as information in hyperbolic PDEs propagates with finite speed (along its characteristics) the solution $\rho(x,t)$ is only affected by an interval of the initial condition. Points outside this interval do not affect the solution $\rho(x,t)$. We define this interval to be the *domain of dependence* of the solution $\rho(x,t)$. The CFL condition then states that the numerical domain of dependence must contain the true domain of dependence (the numerical domain of dependence is similarly defined). For methods that only use adjacent cells to compute the solution at the next time step (e.g., Equation (9)), this means that information must only come from the adjacent cells and not from cells further away. We then define $\lambda_{\max}$ to be the fastest wave speed of the PDE and we require $\Delta t \leq \frac{\Delta x}{|\lambda_{\max}|}$, that is:

$$\left| \frac{\Delta t \lambda_{\max}}{\Delta x} \right| \leq 1. \tag{10}$$

This condition means we require the time resolution Δt to be smaller than the time it takes for the fastest wave speed to cross a cell of size Δx. Thus information cannot propagate from nonadjacent cells to the current cell when calculating the density at the next time step.

The Clawpack software automatically chooses the time resolution for the solver based on the chosen spatial resolution and the CFL condition. We choose a spatial resolution of 19 m (as we have 259 cells for a 5 km length of road), and pass in a new BC density value every 1.5 s (which corresponds to the time resolution required by CFL for typical FD parameter values found in data). Clawpack therefore chooses $\Delta t \leq 1.5$ s depending on the FD parameter values. We used this resolution for the so-called *square wave test* in Figure 5. This test corresponds to using the solver to in a simple case for which the true solution is known analytically: a Riemann Problem with a shock wave moving toward the left (case 1 as defined above). The initial condition is $\rho_{i-1}^0 = 150$ and $\rho_i^0 = 200$. This test was performed using del Castillo's FD with parameter values $(Z, \rho_j, u, \gamma) = (15, 300, 4, 100)$. The wave speed is given by $v_{\text{shock}} = \frac{f(\rho_i) - f(\rho_{i-1})}{\rho_i - \rho_{i-1}} = -0.05 \text{ km/min}$.

2.4. Data

We use MIDAS data from the Highways Agency on the M25 in 2007. The data is measured on loop detectors spaced every 500 m on the road which take measurements averaged every minute. These loops measure count, occupancy, headway, and average speed. Count is the number of vehicles that have passed the detector in a minute and therefore corresponds to flow (number of vehicles per unit time). Occupancy (ω_{occ}) is the percentage of time in a minute that the detector was recording the passing of a vehicle (so 100% is gridlock, and 0% means that no vehicles passed over the detector), and headway is the time difference between a vehicle leaving the detector and another one arriving.

As density is a variable that must be included in models but is not directly measured by MIDAS detectors, we must estimate its value. One way to do this is to estimate it from speed data: for each lane, multiply the count by 60 and divide it by average speed (which is in km/hr). We obtain the following estimate:

$$\rho_{speed} := \frac{q}{v}. \tag{11}$$

The limitation of this approach is that it does not consider the size of vehicles (which can vary greatly; from small cars to lorries for example).

Another approach is to estimate density from occupancy. To do this we use the relationship between occupancy and density (Heydecker and Addison, 2011) $\rho = \frac{\omega_{occ}}{L}$, with L the average vehicle length. We use this to obtain density estimated from occupancy:

$$\rho_{occ} := \frac{\omega_{occ}}{L}. \tag{12}$$

There are several ways to estimate the average vehicle length L (or its reciprocal L^{-1}): in Heydecker and Addison (2011) the authors outline several methods to estimate L or L^{-1} and find that these have different trade-offs.

However, as we have the count data by vehicle type in the MIDAS data, we will use it to estimate the vehicle length at every minute. The flow by type is the overall count (for all lanes) of the number of vehicles that have passed the detector in the minute classified by type (type 1: 4 m vehicles, type 2: 6 m, type 3: 9 m, and type 4: 16 m). As this count data is given over all lanes (rather than for each lane), we will have to assume that vehicles types are evenly distributed across lanes. Although this assumption allows us to estimate the average vehicle length in a practical way, it is unrealistic as motorways have lanes designated for slower vehicles (which include longer vehicles). Using q_i to denote the count data for vehicles of type i and q the total count data we have: $L = (4q_1 + 6q_2 + 9q_3 + 16q_4)/q$. We then use this value (calculated at every minute) to estimate density from occupancy for each lane: $\rho_{occ} = \frac{\omega_{occ}}{L}$.

We plot flow versus density using these two methods in Figure 6 to show that they give very different results. In particular the congested flow wave speeds vary greatly between the two methods, while the free flow wave speeds are approximately the same. In Section 5, we will estimate density in the BCs (as well as estimate the FD parameters) and compare the resulting wave speeds to those arising from density from speed and occupancy.

We will need to choose an appropriate section of road for analysis. As we are dealing with a single-lane model, we need to choose a section of road with no in/out flows (junctions), the same number of lanes (which we will aggregate variables over), few detector faults, and a consistent flow-density relationship. The chosen road is a 5 km section of the M25, and as the detectors are spaced every 500 m there are 11 detectors (the endpoints of the section are included). However some of the detectors have faults, so we use eight detectors in our inference at the following locations (in km): $[0, 1, 2, 2.5, 3, 4, 4.5, 5]$. We use the detector measurements between 6:21 am and 7:09 am (including both endpoints) on January 8, 2007; we therefore use 48 min of data which corresponds to 49 time points.
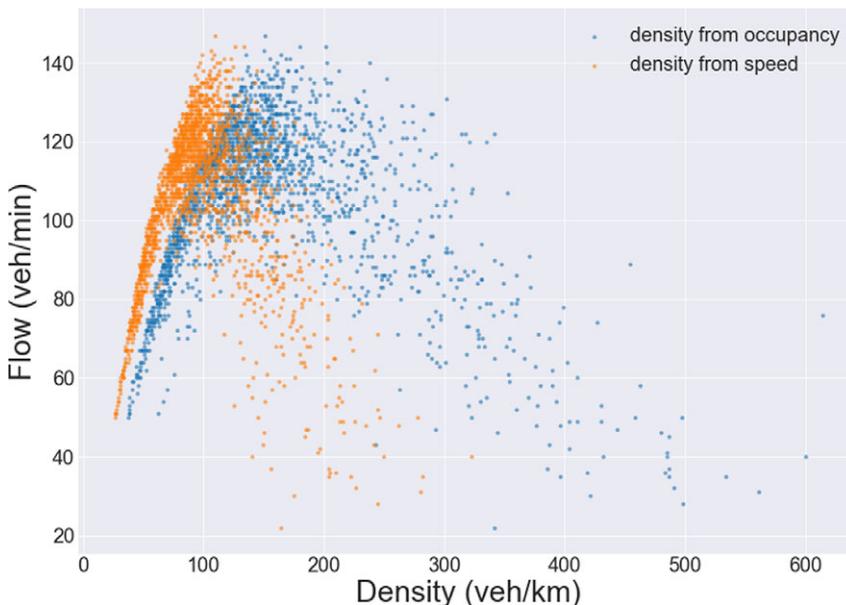

Figure 6. *Section of M25 on January 8, 2007 between 6 am and 10 am. We plot flow versus density for two estimation methods: density from occupancy (Equation (12)) and density from speed (Equation (11)) summed over all lanes. These methods give very different estimates. In particular, we note that the congested flow wave speeds vary greatly between methods, while the free flow wave speed is approximately the same.*

2.5. Previous related work

A study close to the topic of this article is a Bayesian analysis of traffic flow tested on motorway data by Polson and Sokolov (2015). The objective of the article is to develop a methodology (using particle filtering) to estimate traffic density and parameters in the LWR model in real-time. This allows real-time estimation of road capacity (maximum possible flow on the road) and critical density (density at which flow is maximized) that can adapt to the drop in capacity due to an accident on the road. We go over key points in their methodology and provide a critical review.

The motorway data they use includes occupancy ω_{occ} which they use to estimate density (see formula (12)). As explained in Section 2.4, they use this quantity to estimate density with the formula $\rho = \frac{\omega_{occ}}{L}$ (see Heydecker and Addison, 2011) with ρ traffic density and L the average vehicle length. They use a constant average vehicle length, but it is unclear from the paper how it has been obtained.

They discretize the road into M cells and assume that the BCs and initial conditions are known (they use density data from occupancy to construct these). Defining $\psi_t = (\rho_{1t}, \ldots, \rho_{Mt})$ to be the hidden state vector of traffic densities for each cell, the model they use in the particle filter is:

$$\begin{cases} y_{t+1} = H_{t+1}\psi_{t+1} + \varepsilon_{t+1}^V & \text{with} \quad \varepsilon_{t+1}^V \sim \mathcal{N}(0, V_{t+1}) \\ \psi_{t+1} = f_\theta(\psi_t) + \varepsilon_{t+1}^W & \text{with} \quad \varepsilon_{t+1}^W \sim \mathcal{N}(0, W_{t+1}) \end{cases}. \tag{13}$$

where V_t and W_t are evolution and equation error respectively, y_{t+1} is the vector of measured traffic density, f_θ is the LWR evolution equation with FD parameters θ calculated using a Godunov scheme (see Leveque, 2004 for a comprehensive account of these numerical methods or recall Section 2.3). The observation matrix H_{t+1} picks out the cells with the measurements. The objective of the methodology is to sample from the posteriors $p(\psi_t|y^t)$ and $p(\theta|y^t)$ with $y^t = (y_1, \ldots, y_t)$ the current history of data.

They consider a length of road on an interstate outside Chicago which seems to have two detectors separated by 845 m. They discretize this distance into four cells (so each cell corresponds to a distance of 211 m) and use a time discretization of 5 min (which seems to be the resolution of data available). They run the analysis for 24 hr worth of data and estimate the drop in capacity due to a traffic accident. They also test their methodology on simulated data assuming known initial and BCs. For this simulated data, they use a spatial resolution of 300 m and a time resolution of 5 s over a total road length of 1.5 km and a time horizon of 1600 s.

As their objective is real-time estimation of certain traffic flow quantities (such as traffic state and capacity) they use a particle filter rather than MCMC as it is more appropriate for real-time analysis. In contrast, we develop in this article a more general methodology for estimating parameters in hyperbolic PDEs with a more rigorous treatment of the boundary and initial conditions. Indeed, they assume that the boundary and initial conditions are known using density estimated from occupancy. However, estimating density from occupancy has problems (like all methods) which we summarize in Section 2.4; we will therefore impute the BCs rather than estimate them directly from data.

Another issue with the methodology is the coarse discretization of the LWR model (211 m and 5 min for space and time respectively for the analysis of real motorway data); at this resolution, the numerical solver will rather coarsely approximate the underlying PDE. Moreover, the shock waves which are important features of these nonlinear PDEs will be slightly smoothed out, as we point out in Section 2.3. Furthermore, each time step in the PDE solver includes a Gaussian error term as seen in Equation (13); this could smooth out shock waves which could help the sampling methodology. However, we point out that the numerical method which converges to the PDE (described in Leveque, 2004) does not include a random term. Adding a Gaussian error at each time step seems more like the discretization of a stochastic PDE (SPDE) - the LWR model has no stochastic component - although it is unclear whether this discretization would indeed converge to a specific SPDE under refinement of the grid. This may depend on the scaling of the variance in the Gaussian error with discretization step. Moreover, adding a Gaussian error at each time step amounts to adding or removing a random fraction of a vehicle thus violating the conservation of the total number of vehicles (i.e., conservation of mass). The model used in the article

should therefore be considered as an ad hoc discrete model inspired by the Godunov method for LWR rather than a discretization of the PDE.

The methodology developed in this article attempts to remedy these issues and be a more rigorous treatment of the Bayesian inverse problem. However, we reiterate that the objective of the article described above is to develop a methodology for *practical* real-time estimation of certain traffic quantities. Given this objective, these simplifications and assumptions may be justified.

Another relevant paper is by Würth et al. (2021) where the authors consider a general 2 equation model of traffic flow along with LWR and fit the scalar parameters to motorway data using MCMC. Furthermore, they consider a bias term that accounts for the model misfit; this bias is modeled as a Gaussian Process whose hyperparameters are optimized at each iteration. They find that using a two equation model results in a better fit of the PDE to observations, which is consistent with traffic flow theory. However, they use fixed density estimates which are known to have problems as discussed in previous sections. In contrast, we estimate density in the BCs nonparametrically which allows us to bypass this problem.

The use of individual observed trajectories to fit a flow model, that is, the transition from microscopic data to macroscopic models is the main focus of Gomes et al. (2019), in which a large number of individual pedestrian trajectories are used and the flow is modeled with a McKean–Vlasov PDE. The transition is achieved by employing a stochastic differential equation to describe individual pedestrian trajectories. The evolution of the associated transition probability densities is given by a Fokker–Planck-like PDE with BCs allowing inflow and outflow thus generalizing beyond the usual Fokker–Planck case. These inlet and outlet BCs are modeled as constant over time and a generalization to modeling these—like many other—parameters as functions rather than constants is mentioned as an abstract possibility.

3. Parameter Estimation in the FD: Direct Fit

In this section, we consider fitting the FD directly to motorway data, namely without the PDE model. This corresponds to the usual approach in the engineering literature (see, e.g., del Castillo, 2012).

3.1. Statistical model

As the two methods to estimate density discussed in Section 2.4 give very different results, we would like to build our likelihood based on a quantity that has fewer assumptions built in, namely flow (which is simply vehicle counts per minute). We use a Poisson model for the statistical error, which is a standard model for count data. We choose this model because it has the correct domain, is unimodal, and because of its simplicity. A drawback of this model is that interarrival times in a Poisson process are exponentially distributed, whereas we expect to not have a vehicle immediately following another one (especially for high speeds). However as the detectors count vehicles over a minute, many vehicles will have passed before the next count and the model misfit for small time resolutions should not be apparent. Finally, the chosen section of road has four lanes, but our model is only for single-lane roads. We therefore sum flow values over all lanes to obtain the total number of vehicles, and if the individual flows are independent Poisson random variables, then the sum of flows also follows a Poisson distribution.

Letting θ be the vector of FD parameters in LWR, and let (ρ_i, q_i) (with $i \in \{1, 2, \ldots, N\}$) be observed density and flow. We assume that data is iid with a Poisson error model which leads to the log-likelihood:

$$l(\theta) \propto \sum_{i=1}^{N} (-\widehat{q}_i(\theta) + q_i \log(\widehat{q}_i(\theta))). \tag{14}$$

With $\widehat{q}_i(\theta)$ the predicted flow for the ith observation resulting from the observation operator $\mathcal{G}(\theta)$ from Equation (1). In this section, we sample from the parameters in del Castillo's (Equation (6)), namely (z, ρ_j, u, γ). High values of γ corresp in the limit to the triangular FD (Equation (4)) which we would like to include in our model. We, therefore, switch to the more convenient parameterization $\omega := \frac{1}{\gamma}$.

Based on domain knowledge of the realistic shape the FD can take, we set uniform priors for the FD parameters shown in Equation (15) below:

$$\begin{cases} z \sim \mathcal{U}(100,400) \\ \rho_j \sim \mathcal{U}(300,800) \\ u \sim \mathcal{U}(1,10) \\ w \sim \mathcal{U}(0.004,10) \end{cases}. \tag{15}$$

Code for reproducing the sampling results in this article can be found at https://github.com/jeremie coullon/BIP_LWR-paper.

3.2. *Direct fit*

Here we fit del Castillo's FD directly to flow-density data. So the predicted flow $\widehat{q}_i(\theta)$ in Equation (14) is simply flow predicted from observed density ρ_i using del Castilo's FD with parameters $\theta = (z, \rho_j, u, \omega)$ (so without using the LWR model). This is a usual approach in the traffic flow literature (see, e.g., del Castillo, 2012).

We sample from the posterior using a random walk Metropolis-Hastings algorithm. We run three chains for 5 K iterations with covariance matrix given in the Appendix. We obtain acceptance rates of 24.4, 23.1, and 23.8%, and show the trace plots for the parameters in Figure 7 which suggest good mixing, as expected in such a low-dimensional model.
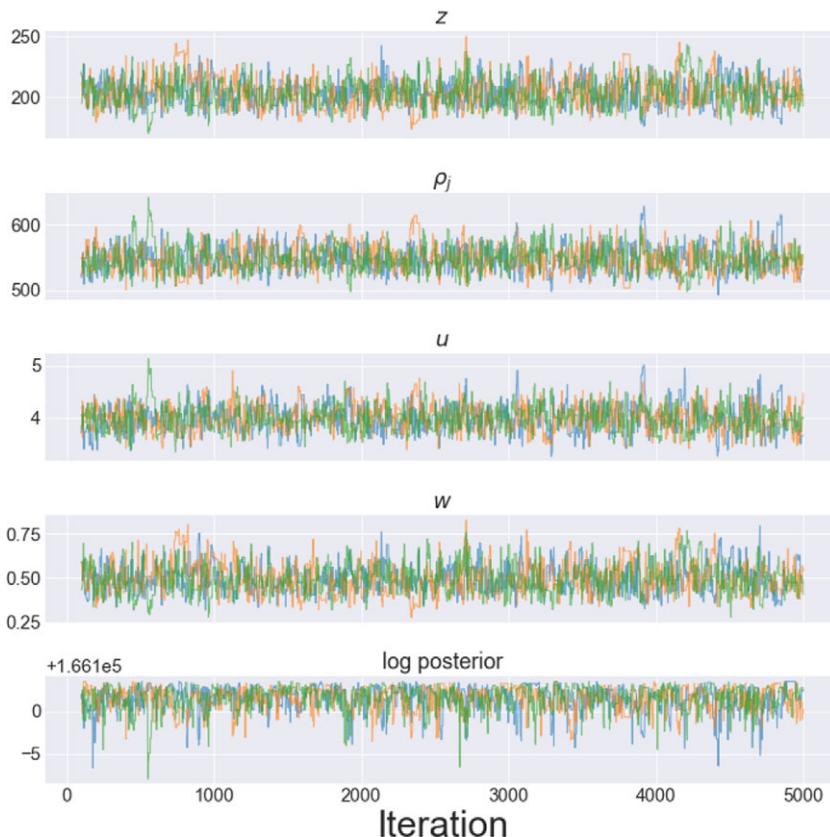

Figure 7. *Posterior samples from a direct fit of del Castillo's fundamental diagram to M25 data. Trace plots of sampled parameters against flow-density data. The three colors correspond to the three Markov chain Monte Carlo chains.*
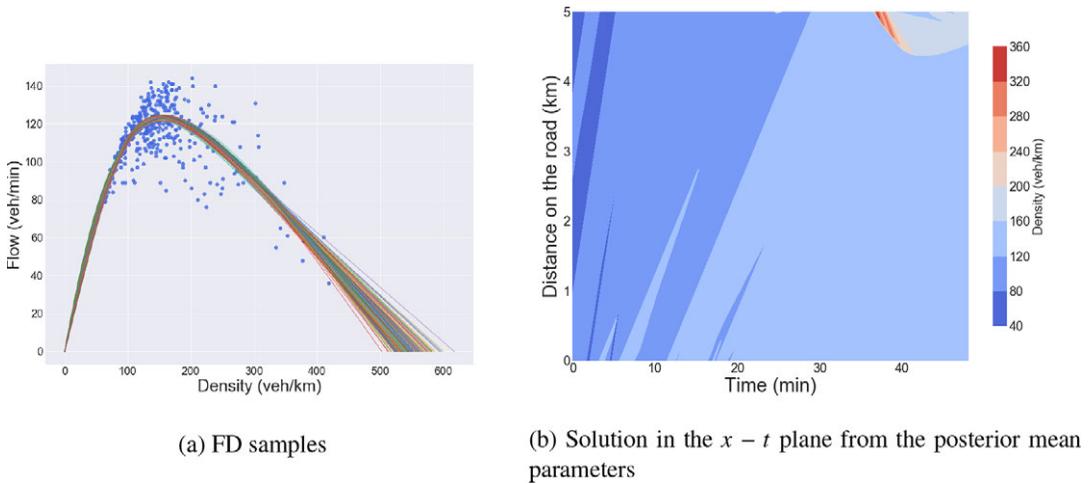
(a) FD samples

(b) Solution in the $x - t$ plane from the posterior mean parameters

Figure 8. *Posterior samples from a direct fit of del Castillo's fundamental diagram (FD) to M25 data. (a) Plotted FDs using the samples. (b) Density in the $x - t$ plane from Lighthill–Whitham–Richards. Parameters used are the posterior mean from samples. We notice that the congested flow waves do not cross the domain as they do in the data.*

We show FD plots with sampled parameters against flow-density data (using density from occupancy) in Figure 8a. We then show the output from LWR in the $x - t$ plane using the posterior mean samples in Figure 8b. Comparing this plot to the real data in Figure 1 we can clearly see that the congested flow wave speed is incorrect (namely, the congested flow waves in the case of the direct fit do not cross the domain at all). This suggests a misfit of the model. Further experiments with other FDs and more detailed investigations of mixing are available in Coullon (2019).

4. Inverse Problem Methodology

In this section, we infer the parameters in the LWR model to predict flow values q_i in Equation (14). The observation operator $\mathcal{G}$ solves the LWR model given some BCs, initial conditions, and the FD parameters θ, maps the density output to flow using the FD with parameters θ, and picks out the flow at the (x,t) values corresponding to observations.

Estimating the parameters in this way captures the dynamic behavior of the PDE compared to the direct fit. Furthermore, estimating the BCs accounts for the uncertainties in the density estimates which were discussed in Section 2.4. We therefore aim to estimate the four FD parameters along with the two BCs.

The BCs are functional parameters so function space samplers are required for this. This is an active area of research that has yielded many samplers and methods such as random walk (Cotter et al., 2013) ∞-Metropolis-Adjusted Langevin Algorithm, and ∞-Hamiltonian Monte Carlo (Beskos et al., 2017). There is also a relevant class of samplers and methods, which involve splitting the space into a low-dimensional part informed by the likelihood and an infinite-dimensional complementary subspace (Cui et al., 2014, 2016; Law, 2014). However, most efficient samplers require gradients, which are not available in our application; we must therefore restrict ourselves to gradient-free samplers. The first gradient-free sampler to try is preconditioned Crank-Nicholson (pCN) (Cotter et al., 2013), which is the function space extension of random walk. This algorithm is simple to implement but will mix slowly if the parameters are correlated or if the posterior is multimodal. This is the case in our application (see Section 5.2 for a discussion): the FD and BC parameters are highly correlated and the BCs exhibit multimodality. We, therefore, use the functional

ensemble sampler (FES) (Coullon and Webber, 2021), which is a gradient-free sampler that can handle correlations in the parameters. We extend it with a PT scheme to mix efficiently within the modes.

In the following sections, we describe the sampler along with the BC prior and the treatment of the initial conditions in LWR.

4.1. Functional ensemble sampler

The FES (Coullon and Webber, 2021) combines pCN and the affine invariant ensemble sampler (AIES) (Goodman and Weare, 2010). For exposition purposes, we describe the sampler for a posterior distribution that only includes a single functional parameter. The extension to our application (two functional parameters and a finite dimensional parameter) is straightforward.

To apply this sampler, we first calculate the Karhunen–Loève (KL) expansion for the Bayesian prior distribution, assumed to be a Gaussian process prior. We then use a Metropolis-within-Gibbs sampler that uses AIES to sample the posterior distribution on the low-wavenumber KL components and uses pCN to sample the posterior distribution of the high-wavenumber KL components. Alternating between AIES and pCN updates, we obtain an efficient functional sampler without requiring detailed knowledge of the target distribution.

We give the pseudocode for the algorithm below:

Algorithm 1 (FES)

To sample a distribution $\pi(d\psi) \propto \exp(\varphi(\psi)) \pi_0(d\psi)$ where $\pi_0 = \mathcal{N}(0, C)$, perform the following steps:

1. Identify a matrix J whose columns are the first M eigenvectors of C. Set $P = JJ^T$ and $Q = I - JJ^T$.
2. Initialize an ensemble of walkers $(X_1^0, \dots X_L^0)$.
3. For $\tau = 0, 1, \dots$:
    (a) For $i = 1, \dots, L$:
        (i) Randomly choose a walker $X_j^{2\tau} \neq X_i^{2\tau}$.
        (ii) Propose the update
        $$\widetilde{X}_i^{2\tau} = X_i^{2\tau} + (1 - Z) P\left(X_j^{2\tau} - X_i^{2\tau}\right), \tag{16}$$
        where $Z \in [\frac{1}{a}, a]$ has density $g(z) \propto \frac{1}{\sqrt{z}}$.

        (iii) Set $X_i^{2\tau} = \widetilde{X}_i^{2\tau}$ with probability
        $$\min\left\{1, Z^{M-1} \frac{\pi\left(\widetilde{X}_i^{2\tau}\right)}{\pi\left(X_i^{2\tau}\right)}\right\}. \tag{17}$$

    (b) Set $(X_0^{2\tau+1}, \dots, X_L^{2\tau+1}) = (X_0^{2\tau}, \dots, X_L^{2\tau})$.
    (c) For $i = 1, \dots, L$:
        (i) Propose the update
        $$\widetilde{X}_i^{2\tau+1} = P X_i^{2\tau+1} + Q\left(\sqrt{1 - \omega^2} X_i^{2\tau+1} + \omega \xi\right), \tag{18}$$
        where $\xi \sim \mathcal{N}(0, C)$.
        (ii) Set $X_i^{2\tau+1} = \widetilde{X}_i^{2\tau+1}$ with probability
        $$\min\left\{1, \exp\left(\varphi(\widetilde{X}_i) - \varphi(X_i)\right)\right\}. \tag{19}$$
    (d) Set $(X_0^{2\tau+2}, \dots, X_L^{2\tau+2}) = (X_0^{2\tau+1}, \dots, X_L^{2\tau+1})$.

The main tuning parameter in FES is M, which controls how many KL coordinates are included in the AIES sampling. Coullon and Webber (2021) recommend keeping this parameter less than 20 as the performance of AIES tends to deteriorate for high dimensional distributions. The authors of Coullon

and Webber (2021) also explore how sensitive the algorithm is to different values of M and find that values below 20 tend to work well. The other parameter to tune is the pCN step size ω which should be tuned to obtain a reasonable acceptance rate (aiming for $20 - 50\%$ works well in our experience). Finally a is also a tunable parameter but as recommended in Coullon and Webber (2021) we fix it to $a = 2$.

4.2. Parallel tempering

PT, also known as Replica Exchange MCMC (Liu, 2001; Brooks et al., 2011), is an algorithm used to sample from multimodal distributions. We augment the state space by introducing an additional discrete parameter $\beta_{\text{temp}} \in I_{\text{temp}} = \{\beta_1 = 1 < \beta_2 < \ldots < \beta_L\}$ with $\beta_i \in [0, 1]$. Defining the unnormalized target posterior as $\pi(\psi) \propto \exp\{-k(\psi)\}$ we can augment the state space by tempering the likelihood (with $-\varphi(\psi)$ the log-likelihood, $-\varphi_0(\psi)$ the log-prior, and $c(\beta_{\text{temp}})$ the pseudo-prior (as defined in Brooks et al., 2011)):

$$\pi\left(\psi, \beta_{\text{temp}}\right) \propto \exp\left\{-\beta_{\text{temp}}\varphi(\psi) - \varphi_0(\psi)\right\} c\left(\beta_{\text{temp}}\right). \tag{20}$$

We therefore run L chains in parallel to target the joint distribution $\Pi(\psi_1, \ldots, \psi_L) = \pi_1(\psi_1), \ldots, \pi_L(\psi_L)$ where $\pi_i(\psi)$ is the posterior tempered using the inverse temperature β_i.

We define α_0 to be the probability of making a within-temperature move, sample $U \sim \mathcal{U}(0, 1)$, and run the algorithm:

- if $\alpha_0 > U$, perform a within-temperature move; this can be done using any MCMC sampling scheme and
- if $\alpha_0 \leq U$, choose uniformly a pair of posteriors $\pi_i(\psi_i)$ and $\pi_j(\psi_j)$ (usually chosen so that the inverse temperatures are adjacent) and swap the states ψ_i and ψ_j with probability $\alpha = \min\left\{1, \frac{\pi_i(\psi_j)\pi_j(\psi_i)}{\pi_i(\psi_i)\pi_j(\psi_j)}\right\}$.

Many extensions of these tempering algorithms have been proposed in the literature, such as ones generalizing the between-temperature moves in Tawn and Roberts (2018) or ones defining a continuous temperature schedule in Graham and Storkey (2017).

A benefit of this algorithm is that the chains can be run on parallel cores rather than sequentially. However, one needs to choose a temperature schedule I_{temp}. To tune this schedule we use the following iterative tuning procedure (based on the one outlined in Atchade et al., 2011): we raise the likelihood to the power β_{temp} with $\beta_{\text{temp}} \in [0, 1]$, and find a value of β_{temp} that allows the chain to mix well. We then find a colder temperature such that the swap acceptance rate is approximately 23%. We then fit a geometric schedule between these two values and extrapolate to find the other temperatures. We then check that the acceptance rate for temperature swaps is around 23% for these spacings.

4.3. FES with PT

The sampler used in our application merges FES and PT and includes some modifications on FES as shown in algorithm 1. The details of our implementation are as follows:

- We set $M = 4$ and so the low dimensional space is 12 dimensional: 4 for each BC and 4 for the FD.
- The Metropolis-within-Gibbs sampler alternates four steps rather than two: the AIES update for the BCs and FD, the pCN update for the inlet BC, the pCN update for the outlet BC, and the temperature swaps.
- We use a random scan Metropolis-within-Gibbs sampler rather than deterministic scan, and we tune the move probabilities to obtain good mixing. The tuned parameters can be found in the Appendix.
- We use 13 walkers and use the parallelized version of FES (see Coullon and Webber, 2021).
- We use a uniform prior for the FD parameters and a "log-OU" prior for the BCs as described in Section 4.4.

4.4. BC prior elicitation

In this section, we describe how we chose a prior for the two BCs. As density from occupancy (see Section 2.4) is considered an appropriate method for estimating density, we will use it to elicit the prior. As discussed previously, it is estimated using the average vehicle length (unlike density estimated from speed). Eliciting the prior in this way will encode our prior belief that the estimated BCs should not deviate too far from density estimated from occupancy.

   We will use time series of density from our section of road to fit a Gaussian process prior for the BCs (of course discarding the day that we use in inference).

   We choose as prior a "log-OU" process. By that, we mean that the logarithm of the centered BCs follow an Ornstein Uhlenbeck (OU) process. We choose this prior for three reasons:

- We would like density to always be positive.
- We would like the prior to allow sudden excursions in density corresponding to high-density waves. Indeed, with a log-OU prior we model the logarithm of the centered BCs with an OU process. As a result, a high value of density will a priori have a higher variance which enables high-density waves, and a low value of density will a priori have a low variance.
- We would like to be able to sample easily from the prior as well as evaluate the probability of a sample under the prior. This is computationally inexpensive to do with a log-OU prior.

We first give a succinct overview of the OU process and then estimate the parameters of this process from traffic flow data.

   For a given BC (i.e., inlet or outlet), let Y_t be the logarithm of the BC at time t and let $X_t := Y_t - \mu(t)$ with $\mu(t)$ be the time-dependent mean. Then we choose X_t to be the unique solution of the stochastic differential equation $dX_t = -\beta X_t dt + \sigma dW_t$ (with W_t a Wiener process), with $\beta > 0$ and $\sigma > 0$ the mean-reversion parameter and diffusivity parameter respectively (see Iacus, 2008).

   We fit an OU process to centered log-BCs for the inlet and outlet BC together, as fitting them separately yielded similar OU parameters. The inlet BC is a function of time that returns density, and corresponds to the inlet of the studied stretch of road (i.e., $x = 0$). Similarly, the outlet BC corresponds to the outlet of the road (i.e., $x = 5$km).

   For the inlet and outlet detector data we keep only weekdays, discard January 1, and keep only the 75 and 65 first days for the inlet and outlet detectors respectively. We removed these last days as they have unusual density curves. We also removed January 8, as this is the dataset used in the inference. We fit the mean $\mu(t)$ to the logarithm of traffic curves Y_t, and then fit the OU parameters β and σ to X_t (we fix $\Delta t = 1$ to define a unit of time to be 1 min). We apply a very slight smoothing to the log-BC means to ensure that they are smooth.

   We estimate the parameters for the inlet and outlet together using MCMC. Defining X^i to be the ith measured density curve and Λ the precision matrix for the OU process, we write the likelihood as:

$$l(\Lambda) \propto -\frac{N}{2} \log |\Lambda| - \frac{1}{2} \sum X^i \Lambda X^i. \tag{21}$$
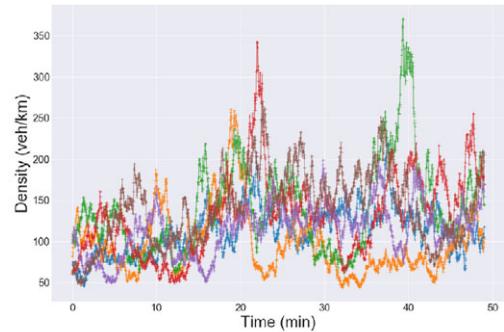
   We use a flat prior for the parameters and use a random walk Metropolis sampler to sample from the posterior. The posterior mean is $\beta = 0.22$ and $\sigma = 0.256$.

   We plot in Figure 9a inlet BCs from data (data used to fit the OU process) along with prior samples. To allow comparison to the BCs from data, the prior samples here have the same resolution: one point per minute. We can visually check that the log-OU prior fits fairly well the inlet BCs from data (prior samples for the outlet BCs are also similar to BCs from data).

   We also plot in Figure 9b samples from the inlet BCs at full resolution: one point every 1.5 s, which is the resolution that we use in the inference. We use this resolution as it allows for a detailed description of the density waves that will get propagated by LWR. Indeed comparing the two figures, we can see that a resolution of 1 min does not capture the details of the high-density waves.
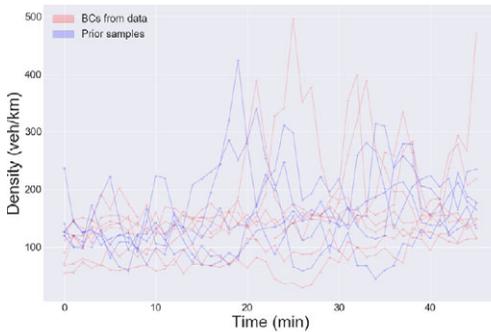
(a) Inlet BC prior samples vs data   (b) Inlet BC prior samples - high resolution

Figure 9. *(a) Inlet boundary conditions from data (using density from occupancy) along with prior samples at 1 min resolution. (b) Samples from the prior for the inlet at full resolution: one point every 1.5 s.*

4.5. Treatment of the initial conditions

LWR requires the initial condition as well as the BCs. The density will be propagated with either free flow or congested flow wave speed and so only the density measured for the first few minutes will be influenced by the initial condition. To avoid having to infer this initial condition, we simply do not use these first few detector times to build the likelihood; as a result, the likelihood is unaffected by the initial condition and is only affected by the choice of BCs and FD parameters. To be able to only remove a small number of points in the $x - t$ plane (i.e., just the first few minutes) we assume that the FD parameters are such that density for these initial times corresponds to free flow (which is a reasonable assumption as can be seen in Figure 1). We further assume that the free flow wave speed lies within a reasonable range of speeds. We remove the influence of the initial condition on the likelihood in this way for all further inferences in the paper.

5. Results and Discussion

5.1. Sampling results

We run the sampler for 102,000 iterations and thin the samples by 100.

We show in Figure 10 the trace plots for the FD parameters and for a few of the walkers which show good mixing. We show in Figure 11 the FD samples plotted with M25 flow data against three density estimates: density from speed, from occupancy, and estimated in the BCs. To obtain the latter we used the mean BCs (both inlet and outlet) and picked out the time points that correspond to measurements. We then plotted the M25 flow data at those time points against the density in the BC means. We first observe that for a given value of flow, the densities estimated in the BCs do not agree with densities estimated from occupancy but somewhat agree with the densities estimated from speed. In terms of the wave speeds, the free flow wave speeds implied by all three density estimates seem to agree, but the congested wave speeds do not. The congested flow wave speed in the fitted model seems to be in between the wave speed implied by the two other density estimates.

As the congested flow waves in the fitted model (in Figure 12a) seem to agree with the waves in M25 data (in Figure 1), this suggests that estimating the density in LWR rather than estimating it in a preprocessing step yields a better fit of the wave speeds. Finally, we show the residuals in Figure 12b which suggests a good model fit.
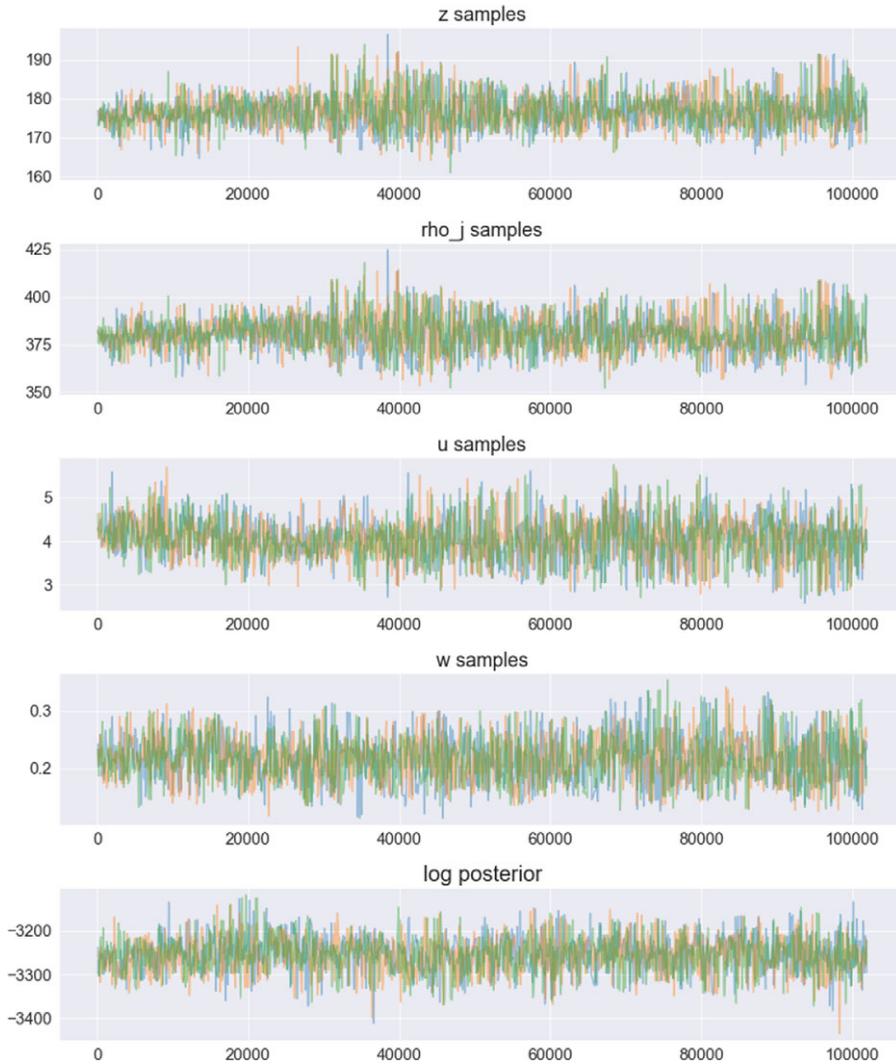
Figure 10. *Trace plots for the fundamental diagram parameters for a parallel tempering functional ensemble sampler sampler, which show good mixing.*

5.2. Discussion on the correlations and multimodality

The slow mixing speed can in part be explained by the strong correlations between the outlet and inlet BC parameters: we can see in Figure 1 that a forward moving free flow wave leaving the inlet BC (with wave speed given by the slope of the FD) will hit the outlet BC (and vice versa for congested flow waves). Changing one of the BCs, therefore, requires changing the other one in a way that is compatible with the first. Furthermore, as hyperbolic PDEs are prone to shock formation, the BCs will have sharp changes in density: a small translation of the BC to the left or right (say) will therefore cause a large drop in likelihood. In contrast, PDEs that exhibit diffusion will smooth out such discontinuities quickly, so such a misfit will be penalized much less by the likelihood. An algorithm such as pCN—which is simply a random walk proposal—will therefore mix slowly.

To explain the multimodality visible in Figures 13a and 14, we recall that the likelihood is built from flow. This means that different values of density that map to the same value of flow will be equally likely.
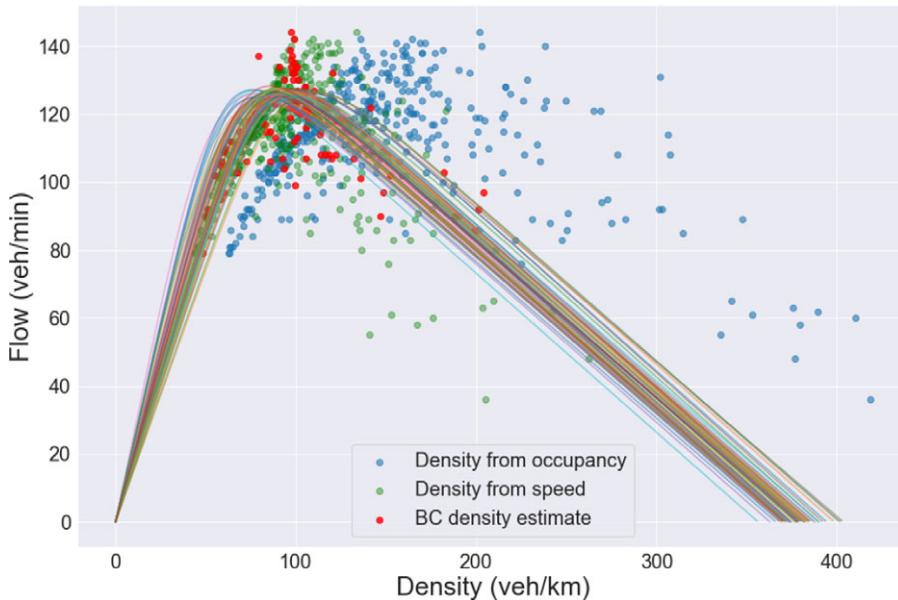
Figure 11. *Fundamental diagram (FD) samples plotted with M25 flow data and three density estimation methods: from occupancy, from speed, and from boundary conditions (BCs). The samples are from FD and BC sampling for del Castillo's FD for M25 data. The density estimated in the BCs seems to agree with density from speed, but the congested flow wave speed in the fitted model seems to be different from the wave speeds implied by the other two density estimation methods.*

To illustrate this, we plot in Figure 15 del Castillo's FD with the same parameters values used in the sampler, and we plot two vertical lines for two values of density ($\rho_1 = 90$ and $\rho_2 = 195$) that map to the same value of flow (the horizontal line). If we then inspect the sections of the outlet and inlet BCs that exhibit multimodality, we observe that some of the pairs of density branches approximately correspond to these two values. Of course, there is dynamic behaviour in the $(x - t)$ plane so this explanation is an approximation.

6. Conclusion

After having given a brief review of motorway traffic flow modelling, we fit the FD directly to flow-density data, but found that estimating the FD and BCs with LWR as the forward model resulted in a superior fit in terms of wave speed. The fitting procedure required a state-of-the-art gradient-free sampler augmented with PT to sample from the highly correlated and multimodal posterior.

We have provided a unified statistical model to estimate both BCs and FD parameters while respecting the character of LWR as a conservation law. Furthermore, we compared the density estimated in the BCs to two density estimates in the engineering literature (density from occupancy and speed) and find that although the free flow wave speeds implied by the three methods agree, only the congested flow wave speeds in the density from BCs (namely, in the fitted model) agree with the congested flow waves in M25 data. When inserted into LWR, the BCs estimated using our method provide a fit superior to that obtained from BCs using engineering methods.

Furthermore, the sampling methodology developed in this article could be used to fit more sophisticated traffic flow models to data. Examples of these are systems of PDEs that include conservation of
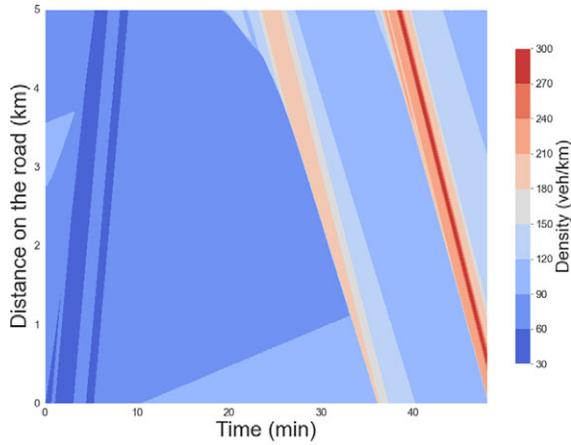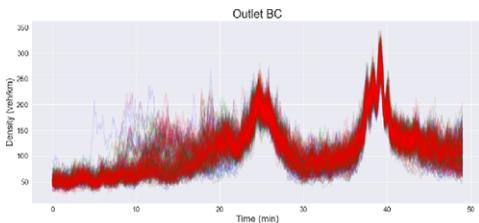
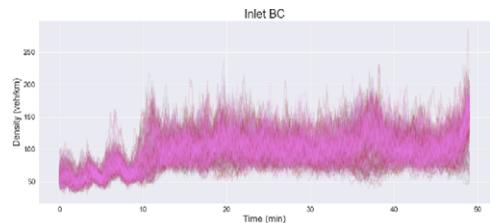(a) $x - t$ plane from posterior mean
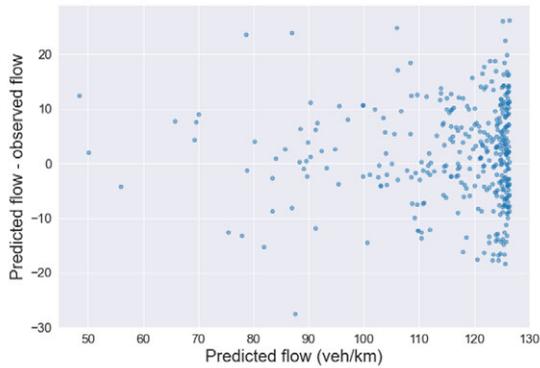
(b) Residuals from posterior mean

Figure 12. *Using the posterior mean parameters from fundamental diagram and boundary condition sampling with a parallel tempering functional ensemble sampler sampler, we plot the output of Lighthill–Whitham–Richards in the $x - t$ plan (a) and the residuals (b).*

(a) outlet BC trace plots

(b) Inlet BC trace plots
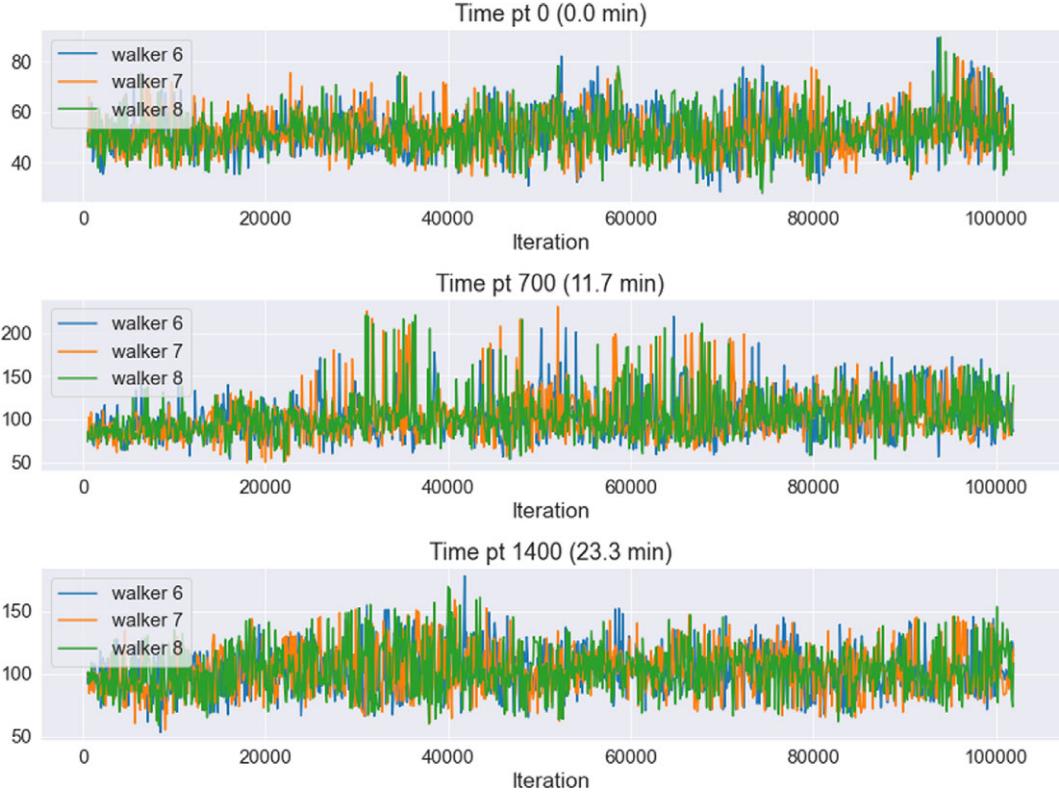
Figure 13. *Outlet and inlet boundary condition samples.*

Figure 14. *Trace plots for three time points in the outlet boundary condition which show some of the multimodality.*
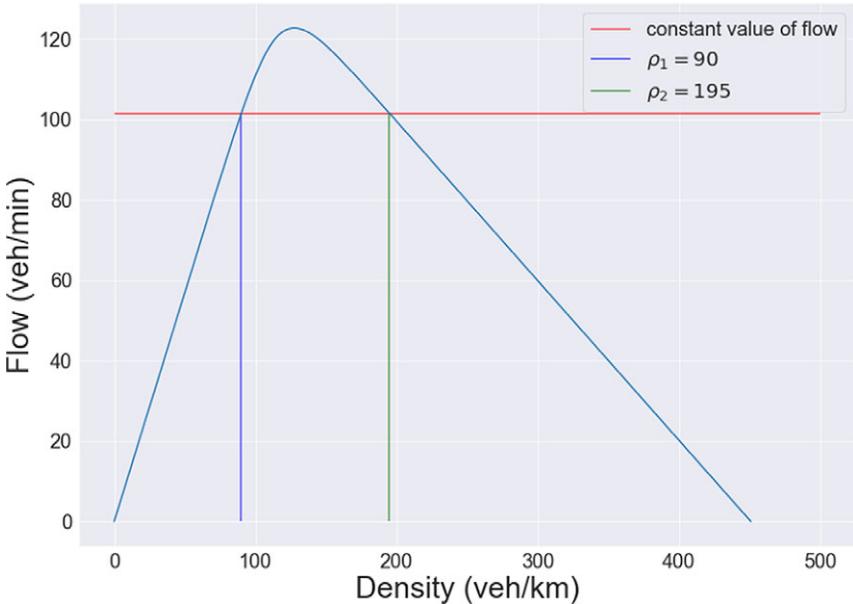

Figure 15. *Del Castillo fundamental diagram. The two vertical lines correspond to two values of density ($\rho_1 = 90$ and $\rho_2 = 195$) that map to the same value of flow. As the likelihood is built from flow, these two values of density are equally likely and therefore the posterior exhibits multimodality.*

momentum as well as mass (namely, two equation models), which are discussed in Coullon (2019). These classes of models could be quantitatively compared after fitting the parameters and BCs for each of them. This would allow for a rigorous assessment of the strengths and weaknesses of these models.

References

Atchadé YF, **Roberts GO and Rosenthal JS** (2011) Towards optimal scaling of metropolis-coupled Markov chain Monte Carlo. *Statistics and Computing* 21(4), 555–568.

Aw A and Rascle M (2000) Resurection of "second order" models of traffic flow. *SIAM Journal on Applied Mathematics* 60(3), 916–938.

Bando M, **Hasebe K**, **Nakayama A**, **Shibata A and Sugiyama Y** (1995) Dynamic model of traffic congestion and numerical simulation. *Physical Review E* 51, 1035–1042. https://doi.org/10.1103/PhysRevE.51.1035

Beskos A, **Girolami M**, **Lan S**, **Farrell PE and Stuart AM** (2017) Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics* 335, 327–351. https://doi.org/10.1016/j.jcp.2016.12.041

Bonzani I and Gramani CLM (2009) Critical analysis and perspectives on the hydrodynamic approach for the mathematical theory of vehicular traffic. *Mathematical and Computer Modelling* 50, 526–541. https://doi.org/10.1016/j.mcm.2009.03.007

Brooks S, **Gelman A**, **Jones GL and Meng X** (2011) *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL: CRC Press.

Clawpack Development Team (2018) Clawpack Version 5.4.0. Available at http://www.clawpack.org; https://doi.org/10.5281/zenodo.262111 (accessed 14 February 2020).

Cotter SL, **Dashti M**, **Robinson JC and Stuart AM** (2009) Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Problems* 25, 115008. https://doi.org/10.1088/0266-5611/25/11/115008

Cotter SL, **Roberts GO**, **Stuart AM and White D** (2013) MCMC methods for functions: Modifying old algorithms to make them faster. *Statistical Science* 28(3), 424–446. https://doi.org/10.1214/13-STS421

Coullon J (2019) *MCMC for a Hyperbolic Bayesian Inverse Problem in Motorway Traffic Flow.* PhD Dissertation, University College London.

Coullon J and Webber RJ (2021) Ensemble sampler for infinite-dimensional inverse problems. *Statistics and Computing* 31, 28. https://doi.org/10.1007/s11222-021-10004-y

Cui T, **Law KJH and Marzouk YM** (2016) Dimension-independent likelihood-informed MCMC. *Journal of Computational Physics* 304, 109–137.

Cui T, **Martin J**, **Marzouk YM**, **Solonen A and Spantini A** (2014) Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems* 30(11), 114015.

Daganzo CF (1994) The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological* 28(4), 269–287.

del Castillo J (2012) Three new models for the flow-density relationship: Derivation and testing for freeway and urban data. *Transportmetrica* 8(6), 443–465.

Gasser I, **Sirito G and Werner B** (2004) Bifurcation analysis of a class of "car following" traffic models. *Physica* D 197, 222–241.

Gomes SN, **Stuart AM and Wolfram M-T** (2019) Parameter estimation for macroscopic pedestrian dynamics models from microscopic data. *SIAM Journal on Applied Mathematics* 79(4), 1475–1500.

Goodman J and Weare J (2010) Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Scienc* 5(1), 65–80.

Graham MM and Storkey AJ (2017) *Continuously Tempered Hamiltonian Monte Carlo. arxiv* Available at https://arxiv.org/abs/1704.03338v1 (accessed 5 March 2019).

Greenshields BD (1934) *The photographic method of studying traffic behaviour. In Proceedings of the 13th Annual Meeting of the Highway Research Board*, pp. 382–399.

Heydecker BG and Addison JD (2011) Measuring traffic flow using real-time data. In Kuehne R and Gartner NA (eds), *Transportation Research Circular, E-C149: 75 Years of the Fundamental Diagram for Traffic Flow Theory*. Washington, DC: Transportation Research Board, pp. 109–120.

Hoogendoorn SP and Bovy PHL (2001) State-of-the-art of vehicular traffic flow modelling. *Proceedings of the Institution of Mechanical Engineers, Part 1. Journal of Systems and Control Engineering 215*, 283. https://doi.org/10.1177/095965180121500402

Iacus MS (2008) *Simulation and Inference for Stochastic Differential Equations*. New York: Springer.

Iglesia MA, **Lin K and Stuart AM** (2014) Well-posed Bayesian geometric inverse problems arising in subsurface flow. *Inverse Problems 30*, 114001. https://doi.org/10.1088/0266-5611/30/11/114001

Jasra A, **Stephens DA and Homes CC** (2007) On population-based simulation for static inference. *Statistics and Computing 17*(3), 263–279.

Law KJH (2014) Proposals which speed up function-space MCMC. *Journal of Computational and Applied Mathematics 262*(15), 127–138.

LeVeque RJ (2004) *Finite Volume Methods for Hyperbolic Problems*. Cambridge: Cambridge University Press.

Lighthill MH and Whitham GB (1955) On kinematic waves 2: A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London A 229*, 317–345.

Liu JS (2001) *Monte Carlo Strategies in Scientific Computing*. New York: Springer-Verlag.

Mahnke R, **Kaupuzs J and Lubashevsky I** (2005) Probabilistic description of traffic flow. *Physics Reports 408*(2005), 1–130.

Mandli KT, **Ahmadia AJ**, **Berger MJ**, **Calhoun DA**, **George DL**, **Hadjimichael Y**, **Ketcheson DI**, **Lemoine GI and LeVeque RJ** (2016) Clawpack: Building an open source ecosystem for solving hyperbolic PDEs. *PeerJ Computer Science 2016*, e68. https://doi.org/10.7717/peerj-cs.68

Martino L, **Elvira V**, **Luengo D**, **Corander J and Louzanda F** (2016) Orthogonal parallel MCMC methods for sampling and optimization. *Digital Signal Processing 58*(2016), 64–84.

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

Richards PI (1956) Shock waves on the highway. *Operations Research 4*(1), 42–51.

Sambridge M (2014) A parallel tempering algorithm for probabilistic sampling and multimodal optimization. *Geophysical Journal International 196*, 357–374. https://doi.org/10.1093/gji/ggt342

Stuart AM (2010) Inverse problems: A Bayesian perspective. *Acta Numerica 19*, 451–559. https://doi.org/10.1017/S0962492910000061

Tawn NG and Roberts GO (2018) Accelerating Parallel Tempering: Quantile Tempering Algorithm (QuanTA). *arxiv* [preprint] Stat.ME. Available at https://arxiv.org/abs/1808.10415v1 (accessed 5 March 2019).

Würth A, **Binois M**, **Goatin P and Göttlich S** (2021) Data Driven Uncertainty Quantification in Macroscopic Traffic Flow Models. *Hal*, hal-03202124v2.

Zhang HM (2000) A non-equilibrium traffic model devoid of gas-like behaviour. *Transportation Research Part B 36*(2002), 275–290.

A. Appendix

A.1. Direct fit: FD sampling

Covariance matrix for the direct fit to data for $(z\, \rho_j\, u\, w)^T$:

$$\begin{pmatrix} 182.292318 & -288.07905 & -2.34389543 & 1.21897887 \\ -288.07905 & 561.808314 & 5.26749447 & -1.7470824 \\ -2.34389543 & 5.26749447 & 0.08204741 & -0.00839764 \\ 1.21897887 & -1.7470824 & -0.00839764 & 0.00931977 \end{pmatrix}.$$

A.2. FES with PT

- Metropolis-within-Gibbs move probabilities for AIES, pCN outlet, and pCN inlet swap: $[0.25, 0.125, 0.125, 0.5]$.
- Inverse-temperatures: $[1, 0.76, 0.58, 0.44]$.
- pCN step sizes for outlet (for each inverse-temperature): $[0.078, 0.09, 0.11, 0.15]$.
- pCN step sizes for inlet (for each inverse-temperature): $[0.155, 0.17, 0.2, 0.25]$.
- FES truncation: $M_{\text{trunc}} = 4$.