MCMC for a hyperbolic Bayesian inverse problem in motorway traffic flow

Jeremie Coullon

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Civil, Environmental, and Geomatic Engineering

University College London

June 20, 2019

I, Jeremie Coullon, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

We study the LWR model: a hyperbolic conservation law used to model traffic flow on motorways. This is an old model dating back to the 1950s, but has been shown to be robust and is parametrised by the so-called Fundamental Diagram (FD) which provides the relationship between flow and density. We consider the boundary conditions as nuisance parameters to be estimated but neglect the initial conditions as their effect on data is quickly washed out.

The data we use to estimate the parameters in the model is MIDAS data on a section of motorway that does not include any on/off ramps, thus conforming with the nature of the model as a conservation law. Little statistically sound work has been done so far on this inverse problem to estimate the FD parameters as well as the boundary conditions.

We consider two families of FDs, Del Castillo's FD and the exponential FD – which have 4 and 2 parameters respectively – and perform inference for these along with the boundary conditions. We assume as prior that the boundary conditions follow a log Ornstein Uhlenbeck process which corresponds surprisingly well to practitioners' prior belief.

We use standard MCMC methods (Gibbs, RWMH, parallel tempering, functional preconditioned RWMH) to sample from the posterior distribution. For some models, the posterior is highly correlated, multimodal and non-Gaussian, so we introduce novel proposals and find that while these are underpinned by clear intuition and show great promise in preliminary studies, they do not seem to appreciably accelerate mixing judging from the studies carried out so far.

Impact Statement

This thesis developed statistical methodology for fitting mathematical models to motorway traffic and tested the methodology on data from the M25 motorway. We distinguish two ways in which the insights and methodology developed in the thesis could be put to beneficial use: benefits inside academia and benefits outside of it.

Considering firstly the benefits inside academia, there has not been much statistically sound work on hyperbolic Bayesian problems and in particular for nonlinear models. As a result, the insights gained from this class of problems is of interest to the statistics community working on Bayesian inverse problems. Furthermore, this research has opened up several promising avenues to explore for both the area of hyperbolic Bayesian inverse problems but also for MCMC methodology.

In terms of benefits outside academia, we recall that the application we focused on this thesis was motorway traffic. As a result this research furthers the understanding of the dynamic behaviour of traffic, which in term can lead to a more efficiently managed motorways. Indeed, our study used data from the M25 motorway, and the insights gained about estimating traffic density could in principle be applicable to any motorway.

More generally, the fluid flow models studied here have a wide range of applications. For example, very similar models are used to describe the flow of gas in pipe networks. This methodology could therefore be applied to further the understanding of this system and to ensure the optimality of the design and implementation of these networks. This is one example among a wide array of applications of such models in the fields of physics and engineering. Furthermore, a key component of our methodology is quantifying uncertainty in a statistically rigorous way. An essential aspect in making engineering decisions based on mathematical models is to be able to understand the uncertainty inherent in the system.

Acknowledgements

I would like to express my deepest gratitude to my family and my supervisors for the support in writing this thesis.

Contents

List of Figures

0.1 Introduction

The fields of applied statistics, engineering, and physics have always interacted to understand and control systems ranging from electrical engineering, geophysics, and transportation. In these applications, mathematical models are developed using experts' domain knowledge resulting from experiments and research. It is then natural to wish to constrain these models and theories by fitting them to data while still considering the accumulated knowledge of these systems. Bayesian inference is a paradigm that aims to explicitly incorporate prior knowledge of the system in the modeling process via *prior distributions* and update them with data summarised in the *likelihood* to obtain the *posterior distribution*. The early 1990s saw an explosion in the use of Markov Chain Monte Carlo to sample from analytically intractable posterior distributions. As computation became cheaper and more powerful, these methods have been extended and used to tackle more and more complicated problems. In this thesis we explore the problem of inferring scalar and functional parameters from posteriors arising from a nonlinear hyperbolic PDE.

In chapter one, we examine the array of models developed in engineering and their classification in order to position our work in the engineering context. We then give an overview of doing MCMC in practice and review some MCMC methods for Bayesian inverse problems. The chapter two we introduce the inverse problem for the advection equation and as this model is similar mathematically to the LWR model we use this model to experiment with sampling methods that will come into play in the later chapters.

In chapter three we introduce the M25 motorway data used to test the sampling methodology, review various traffic density estimation methods, and introduce the LWR model of traffic along with a numerical method to compute its solution. We then create simulated data using the LWR model and sample the parameters of two fundamental diagrams (FDs) before testing the sampler on M25 data. In chapter 4 we elicit a prior for the boundary conditions (BCs) and use blocking along with parallel tempering to sample BCs given a fixed FD. We find that the posterior exhibits bimodality which can be understood by examining the properties of LWR along with the statistical model used.

Finally in chapter 5 we merge the methodology from chapters 3 and 4 to sample

from both the FD parameters and the BCs and test it on both simulated and M25 data. We analyse the correlations between the FD parameters and the BCs and use these to propose a joint update. However this proposal does not seem to improve mixing in the datasets studies so far. Furthermore, we compare traffic density estimated in our methodology with two density estimation methods in the engineering literature and find that our estimate delivers the best fit to data.

Chapter 1

Literature review

1.1 Traffic flow modelling

1.1.1 Empirical features in traffic

In this section we give a brief overview of the empirical features that can be found in traffic. The aim is to to be able to replicate and explain these features using traffic flow models.

We start by observing flow and density traffic data (flow being the number of vehicles per minute and density the number of vehicles per km) as can be seen in figure (1.1)), and noting that for the flow-density plot (plot (c) in the figure) the measurements on the left of the plot can be well described by a curve with a 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 other ones. 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 plot (b) of figure (1.1).

As well as observing the relationship between flow and density, researchers have studied patterns in traffic in space in time. Indeed Schönhof and Helbing in [48] distinguish 5 main kinds of spatio-temporal congestion patterns.

- Pinned Localized Cluster (PLC): The length of the section of congested traffic stays approximately stable or stablizes over time, and stays at a fixed location on

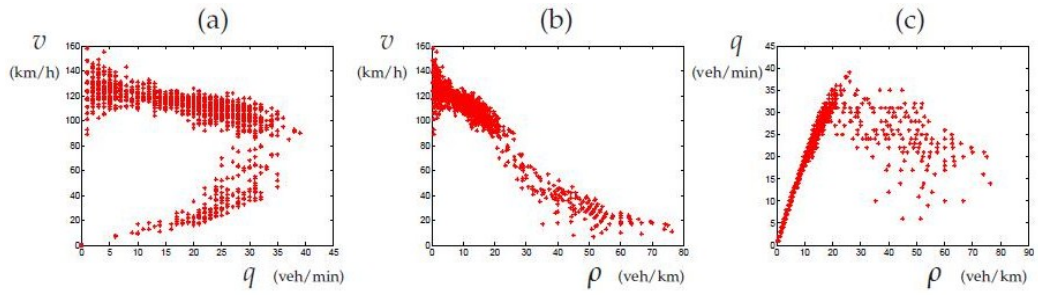

Figure 1.1: Figures showing the empirical relationship between vehicle flow, density, and velocity. The data is taken from inductance loops over one minute averages (data and figure from [61])

the road.

- Moving Localized Cluster (MLC): Same as the PLC, except that it propagates upstream with a fixed characteristic speed.

- Homogeneous Congested Traffic (HCT): congested traffic that is fixed at its downstream front (at a bottleneck) but that grows in size in the upstream direction.

- Oscillating Congested Traffic (OCT): Similar to HCT, but exhibits regular oscillations with a frequency and amplitude that is approximately constant over time.

- Stop and Go Waves (SGW): consists of a series of traffic jams with no particular wavelength and with free traffic in between. They can be thought of as series of MLCs, or as OCT with free traffic between the congested waves.

Furthermore, the authors identify two regions of metastability of traffic in the flow-density plane. These are regions where a small perturbation will decay but a large one can amplify and generate features such as a MLC (see [48] for detailed theoretical and empirical phase diagrams). They find evidence of this in traffic: a downstream moving cluster of vehicles can grow and cause an upstream moving wave (a MLC). They call this the ”boomerang effect”. This evidence supports models that can be unstable. They also use this classification to explain the different dynamics that can occur at bottleneck consisting of successive on and off-ramps.

However, different classifications have been proposed in the literature (for example in [35]) and researchers have trouble agreeing on which specific phases or features

are found in traffic. An intuitive explanation for this disagreement is that traffic flow is composed of human drivers who, although following certain rules of the road, all act slightly differently, have different types of vehicles, and adapt their driving style to changing conditions. This means that trying to extract features from traffic that appear universally might be difficult, as they could depend on the driving culture of the country, the road conditions (road gradient, quality of road), and the environmental conditions (weather). It is also possible that nearby on/off ramps, the proximity to a city, the season, or day of the week could affect driving condition. These different conditions can also affect the composition of the road (for example: more or less trucks). With these points in mind, the classification briefly described above seems appropriate: it defines some basic features in traffic that can act as building blocks to describe more complicated spatiotemporal features.

In reaction to the many factors (human and environmental) that can influence traffic flow and the difficulty in finding a ”true” classification or model of traffic flow, Papageorgiou in [50] (p 324) suggests that “the best one can do (and this is challenging enough) is to look for useful theories (mathematical models) of traffic flow, i.e. theories with sufficient descriptive power”. In the next section we give an overview of the different modelling approaches for motorway traffic flow.

1.1.2 Traffic flow models

Traffic flow models have been developed since the beginning of the twentieth century, and fit within the context of mathematical modelling of complex systems. This approach of expressing a system as a set of mathematical equations dates back to Newton who mathematically described the interactions between gravity and mass. It has flourished and has been used to describe many systems in physics, biology, economics, and more recently social sciences. We will present a review of the various types of traffic flow models, focusing on the so-called macroscopic models.

There exists three important types of traffic flow models: macroscopic models exploiting the analogy between traffic flow and fluid flow, microscopic models which describe traffic flow at the level of individual vehicles, and mesoscopic models which attempt to link the two by considering probability distributions of vehicle position and velocity. Following [8], two main questions to consider when classifying traffic flow

models are:

1. the representation scale used in the model (vehicle or flow representation)

2. the assumptions made in how the system acts

The above classification into micro, meso, and macroscopic refer to the scale with which the model is representing the dynamics of traffic flow, while the assumptions behind the various models can differ within this classification.

Traffic flow models historically originate from the fundamental diagram, which is based on the assumption that there exists a relationship between the average speeds of vehicles and flow describing equilibrium conditions.

1.1.2.1 The fundamental diagram

The fundamental diagram (FD), first studied by Greenshields [24] is the analytic function used to described the relationship between density and flow. It is used to close the conservation of mass equation in macroscopic models, as will be described later. Figure (1.1) shows how this relationship can be seen from traffic flow data.

Three variables are usually considered in this macroscopic modelling framework (the different modelling frameworks will be discussed later): 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:

$$q = \rho v \tag{1.1}$$

A large portion of traffic flow research has been to find models $q(\rho)$ that best explain the relationship between flow, density, and speed in data. As mentioned above, the earliest known model was proposed by Greenshields (sometimes called the *Parabolic FD*) and takes the form $q(\rho) = \rho V_e(\rho)$ with $V_e(\rho) = v_{max}(1-\rho)$ (plotted in figure (1.2)). Since then, many other different relationships have been suggested. A popular example is Daganzo's *Triangular FD* (also called the bi-linear FD) from [16] given in equation (1.2) and plotted in figure (1.3). 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 (1.1.2.4)). We will see later how the wave speed propagation of traffic flow models depend on the shape of the fundamental diagram.

$$q(\rho) = \begin{cases} \frac{q_c}{\rho_c}\rho & \rho < \rho_c \\ q_c \frac{\rho_j - \rho}{\rho_j - \rho_c} & \rho \geq \rho_c \end{cases} \tag{1.2}$$

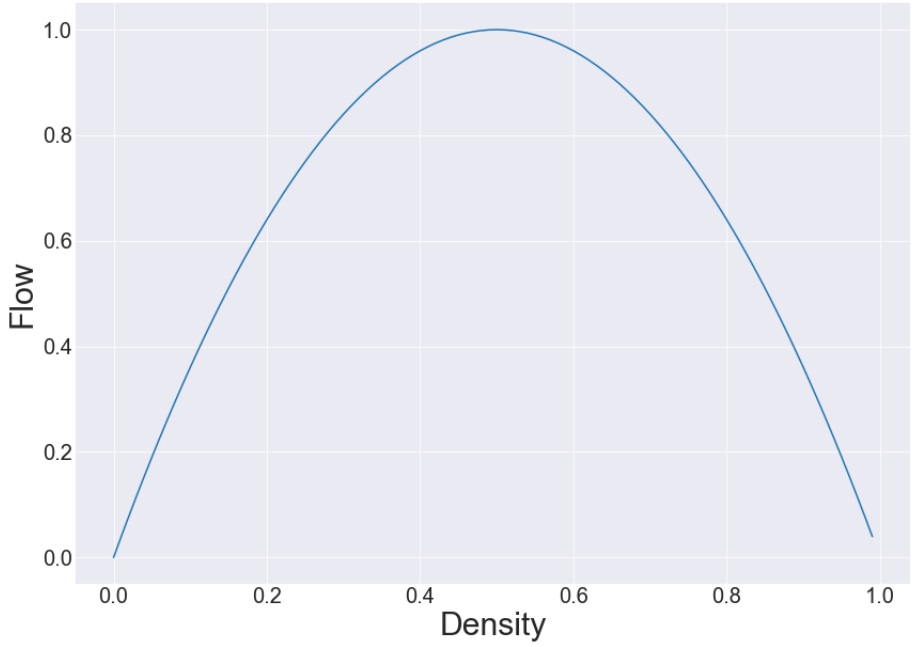

Figure 1.2: Parabolic fundamental diagram due to Greenshields plotted for dimensionless flow and density with $v_{max} = 4$

Another FD as used in [44] and introduced in [59] is the *Exponential FD* seen in figure (1.4). The model is given in equation (1.3):

$$q(\rho) = \alpha\rho \exp^{-\beta\rho} \tag{1.3}$$

In an attempt to derive Fundamental Diagrams in a principled way, Del Castillo in [18] proposed 6 properties that a sound Fundamental Diagram must satisfy:

- $0 < v < v_{max}$: speed is positive and below a maximum value v_{max}

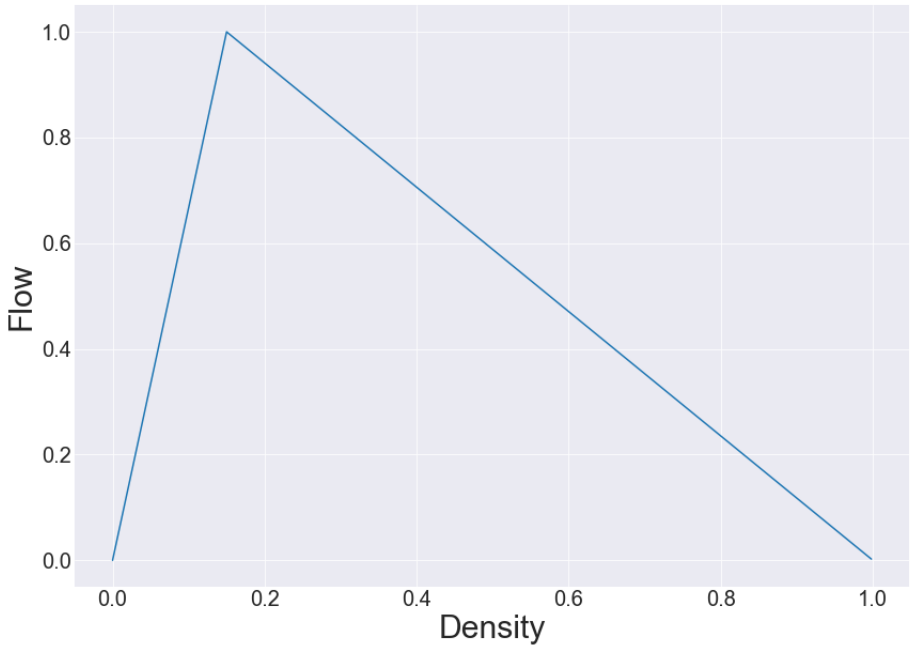

Figure 1.3: Daganzo's triangular fundamental diagram plotted for dimensionless flow and density with $(q_c, \rho_c, \rho_j) = (1, 0.15, 1)$

- $0 < \rho < \rho_j$: density is positive and below the maximum density ρ_j
- $v(0) = v_{max}$ and $v(\rho_j) = 0$: speed at extreme density values are 0 and v_f
- $q(0) = q(\rho_j) = 0$: flow at extreme density values are 0
- $\frac{dq}{d\rho}|_{\rho=0} = v_{max}$ and $\frac{dq}{d\rho}|_{\rho=\rho_j} = C_j$ with C_j the kinematic wave speed at jam density.
- $\frac{d^2q}{d\rho^2}(\rho) > 0$ for $0 < \rho < \rho_j$: the FD should be strictly concave

If the above requirements are satisfied, then the capacity q_c is unique, and we obtain $q_c = \rho_c v_c$ (for v_c speed at capacity). Furthermore, Del Castillo also introduced in [18] a Fundamental Diagram called the negative power model (which we will simply call *del Castillo's FD*). This FD has 4 parameters: flow scaling term Z, jam density ρ_j, shape parameter w, and parameter u relating to the critical density. The shape parameter w determines the tightness of the peak and in the limit of $w \to \infty$ 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:

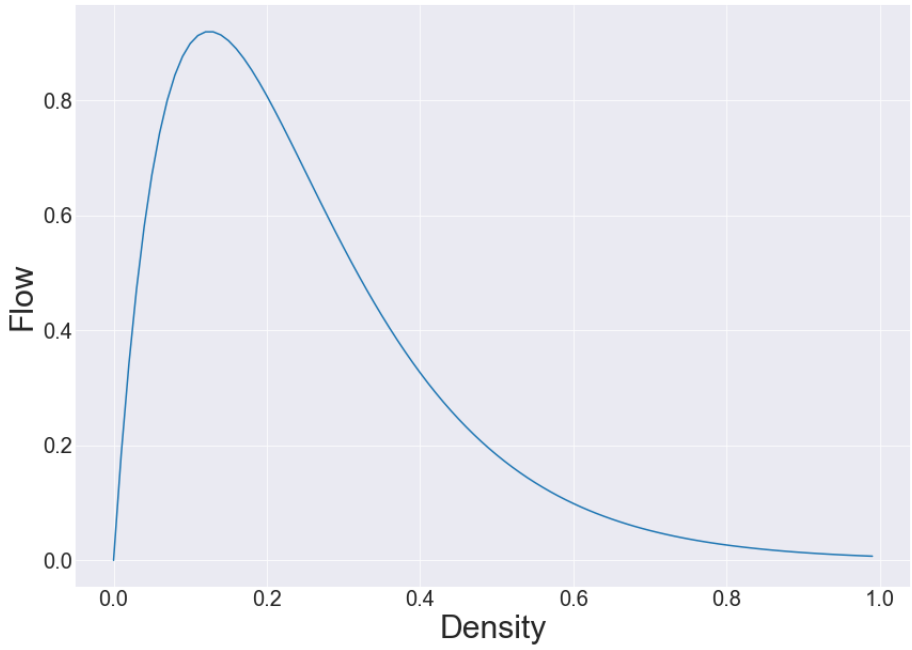

Figure 1.4: Exponential fundamental diagram plotted for dimensionless flow and density with $(\alpha, \beta) = (20, 8)$

$$\rho_c = \frac{1}{1 + u^{w/(w+1)}} \tag{1.4}$$

We give the FD in equation (1.5) below and plot the FD for several values of w in figure (1.5) to illustrate how del Castillo's FD includes the Triangular FD as a limiting case.

$$q(\rho) = Z\left[(u\frac{\rho}{\rho_j})^{-\omega} + (1 - \frac{\rho}{\rho_j})^{-w}\right]^{-\frac{1}{\omega}} \tag{1.5}$$

Although the properties listed by del Castillo seem reasonable, the last one is often not satisfied (as in the case of the Exponential FD introduced above). Another important example of this requirement not being satisfied is the class of fundamental diagrams that attempt to describe the scatter in the empirical flow-density plot, such as in figure (1.6) (figure from [60]).

We can see in this figure that the free flow curve continues past the critical density; there is a region where there are two possible values of flow for a given density. This shape is commonly known as the 'inverse-λ' shape. The possibility of flow 'dropping'

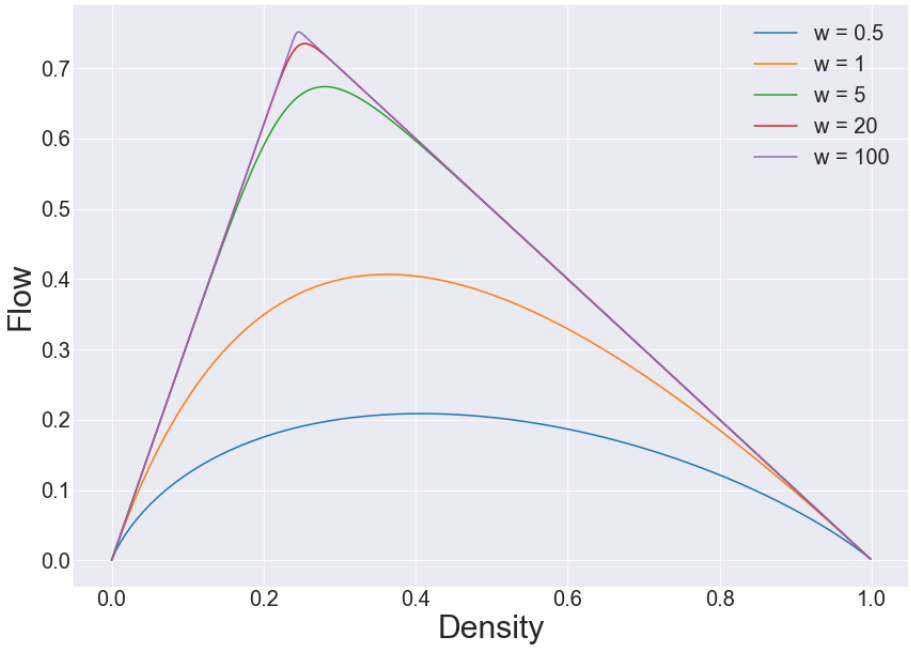

Figure 1.5: Del Castillo's fundamental diagram plotted for dimensionless flow and density with $(Z, \rho_j, u) = (1, 1, 3.1)$ and for $w \in [0.5, 1, 5, 20, 100]$

suddenly from the free flow curve to the congested flow is known as capacity drop. Researchers such as Cassidy and Bertini in [12] find evidence that the flow immediately downstream of a queue is lower than the flow before the queue's formation, which supports this type of fundamental diagram (even though it does not satisfy del Castillo's 6th requirement of strict concavity).

Fundamental diagrams are usually derived from theoretical justification mixed with empirical evidence (as above: evidence in [12] support the 'inverse-λ' shape) and then inserted into a model (we will see how this happens later). Some authors however claim that this relation should emerge from the model or from postulated microscopic mechanisms. For example, Bonzani and Mussone in [6] derive the fundamental diagram from individual driver behaviours. The problem remains that we must choose what is an 'appropriate' fundamental diagram (what shape? how to model the scatter?). Deriving it from microscopic assumptions seems to simply reduce the problem to choosing driver mechanisms that lead to the properties of the fundamental diagram that the researcher thinks are important to reproduce.

Whatever the approach taken however, the assumption in traffic flow models is

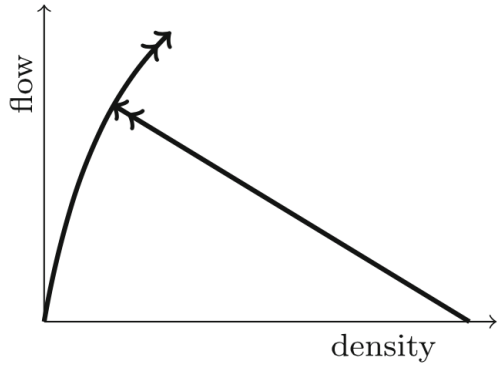

Figure 1.6: Inverse lambda FD

always that traffic is in a state described by the fundamental diagram (or in a state tending towards the FD). This relationship is therefore fundamental in how a model will act.

1.1.2.2 Microscopic Models

Following Ward ([61]) we identify two main types of microscopic models: car following models and cellular automata.

The setup of a car-following model is as follows (see figure 1.7): to describe the movement of individual vehicles in time, we consider a single lane of vehicles following each other labelled 1, 2, etc. in the upstream direction. We define the position and velocity of each vehicles to be x_n and $\dot{x}_n := v_n > 0$ respectively. We can then define the front-to-front spacing between vehicles to be $h_n := x_{n-1} - x_n$, and the speed difference between a vehicle and its leader as $\Delta v_n = v_{n-1} - v_n$. (from [63])

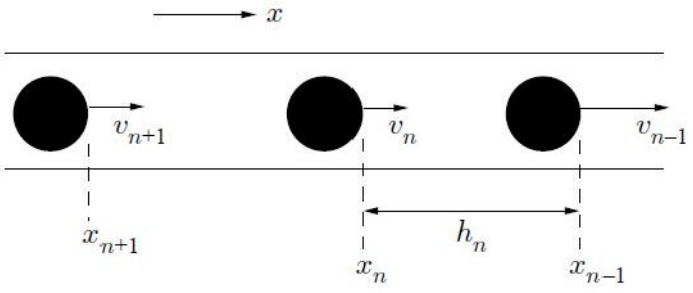

Figure 1.7: car-following model setup.

Bando et al. in [3] defined an *optimal velocity function* $V(h)$ based on the observation from the fundamental diagram that there exists a relationship between a vehicles speed and the spacing to the preceding vehicle. This function is increasing in spacing

to represent that one chooses a higher speed for larger spacings. They proposed in their paper the following S-shaped function:

$$V(h_n) = tanh(h_n - 2) + tanh(2) \tag{1.6}$$

Their model, named the Optimal Velocity Model (OVM) represents the acceleration function of the n^{th} vehicle as follows (with dot denoting differentiation with respect to time):

$$\dot{v}_n = \alpha(V(h) - v_n) \tag{1.7}$$

The parameter $\alpha := \frac{1}{\tau}$ is to be calibrated to observations, and can be interpreted as "sensitivity", or the inverse of the speed adaptation time τ. This latter variable models how quickly a vehicle will adapt its present velocity to its desired velocity.

Although this model has a very simple form it is able to replicate qualitatively stop-and-go waves (as defined in section (1.1.1)). However, this model has some drawbacks: Treiber and Kesting in [58] performed a simulation of OVM and found that they were obliged to use a speed adaptation time that was much too small compared with experimental values by a factor of about 10. Increasing it only slightly would result in *negative* spacings (which correspond to accidents), and decreasing it slightly would cause the system to not be able to exhibit stop-and-go waves. Thus the simulation outcome is not robust to the fine tuning of the parameter. Another result of this unrealistic parameter value is that the periods of the resulting stop and go waves are much too low. This behaviour is mainly due to the fact that drivers in this model are short sighted: they only react to space headways and not the difference in speed with the preceding vehicle, and thus need to react unreasonably quickly to avoid accidents.

Nonetheless, the simulation outcomes are qualitatively correct: the model reproduced stop an go waves in simulations as is discussed in [58], and a bifurcation analysis of the model reveals some very interesting dynamics [20] which can provide an explanation for stable stop and go traffic. This suggests that the assumptions behind the model, namely the existence of a relationship between velocity and spacing and the relaxation of a vehicle to its "optimal velocity" is a fundamentally correct one.

To correct for the short sighted quality of drivers in the OVM, one can include the speed difference Δv_n between a vehicle's speed and the speed of the preceding vehicle. This results in the following acceleration function ([58]):

$$\dot{v_n} = \alpha(V(h_n) - v_n) + \beta \Delta v_n \tag{1.8}$$

One can further refine the model by letting the additional parameter β depend on the headway rather than be constant. An example of such a model is the *Intelligent Driver Model* [57]. Its acceleration function is given by:

$$\dot{v}_n = a\left[1 - \left(\frac{v_n}{v_{max}}\right)^4 - \left(\frac{h_{stop} + v_n T_{gap} - v_n \Delta v_n / \sqrt{4ab}}{h_n}\right)^2\right] \tag{1.9}$$

In this model, a is the maximum acceleration, b is the comfortable deceleration, v_{max} is the desired maximum velocity, h_{stop} is the desired stopping distance, T_{gap} is the desired time gap. As well as considering the velocity difference in a more sophisticated way than the FVDM, this model has an intelligent breaking strategy which models how drivers desire to break with a certain *comfortable deceleration*, but will decelerate more if they are in a so-called "critical situation" (for example if breaking with comfortable deceleration would not be enough to avoid a collision). The parameters of this model therefore have a clear meaning and the results fit empirical data as expected. However, it does have some drawbacks: a platoon of identical drivers will disperse more than in observed traffic. Furthemore, if the time gap between two vehicles is too small (due to a vehicle overtaking for example) the deceleration of the following vehicle will be too strong and unrealistic (as most drivers in real traffic would not consider a vehicle merging into their lane just in front of them as a critical situation). Extensions to this model have been proposed to address these issues, and are used in adaptive cruise controls (ACC) in vehicles. However, IDM is on the whole a model that reproduces observed spatio-temporal dynamics well (a discussion can be found in [58]).

All these different models can be considered as a special case of the following general formulation [58]:

$$\dot{x_n} = v_n \qquad \dot{v}_n = f(h_n, v_n, \Delta v_n) \tag{1.10}$$

The acceleration function should model the fact that drivers accelerate if their front-to-front spacing or relative velocity increases, but that they decelerate if their velocity increases. This is reflected in the following general criteria for sensible microscopic models (f_x denoting the partial derivative of f with respect to the variable x):

$$f_v(h, v, \Delta v) < 0 \tag{1.11}$$

$$f_h(h, v, \Delta v) \geq 0 \tag{1.12}$$

$$f_{\Delta v}(h, v, \Delta v) \geq 0 \tag{1.13}$$

An important property of these models is that they accept a *uniform flow* in which the velocities and headways are time independent:

$$h_n(t) = h_e \quad \Delta v_n(t) = 0 \quad v_n(t) = v_e \tag{1.14}$$

Other extensions of car-following models not considered here include delay terms and anticipation terms (the headway of several preceding vehicles is considered).

The second main branch of microscopic models, cellular-automata models consider a discretised space and velocity (and sometimes time). They are considered microscopic as they are representing the movement of individual vehicles, and are generally computationally more efficient than other microscopic models. A popular example is the stochastic model due to Nagel & Schreckenberg ([45]).

On the whole, these micro models are used to describe interactions between vehicles on a small scale, and can therefore be used to model small networks or ACC systems. However, it can be seen from (1.10) that one needs n equations to describe a system of n vehicles: modelling a large network can quickly become computationally heavy. In these scenarios, using macroscopic models which consist of a few equations can be preferable.

1.1.2.3 Mesoscopic Models

Mesoscopic traffic flow models describe the dynamics of the probability distributions of vehicle position and velocity, while the behaviour of the model is defined at the scale of the individual vehicle. Two important branches of these models are gas-kinetic models and cluster models.

Gas-kinetic models describe the dynamics of the so-called reduced phase space, which corresponds to the probability distribution of vehicle position and velocity over the length of road considered. This results in a partial differential equation from which one can then derive continuum gas-kinetic models (see [30] for a detailed review).

Cluster models consider traffic flow as a stochastic process, and thus describe the evolution of the dynamics of the so-called master equation (see [41] for a review).

1.1.2.4 Macroscopic Models: First order models

These types of models describe traffic flow using partial differential equations (PDEs), and therefore describe the macroscopic behaviour of traffic flow emerging from the individual interactions between vehicles. These models all consider traffic flow as a continuum without representing the individual behaviour of vehicles. The differences between the models will therefore be in the assumptions made in how the system acts. We point to [7] and [30] for a critical review on the topic.

We will consider macroscopic models that consist of a single equation, which corresponds to a conservation law of the form:

$$\rho_t + (\rho v)_x = 0 \tag{1.15}$$

As before, we use the subscripts ρ_t to denote $\frac{\partial \rho}{\partial t}$ and ρ_x to denote $\frac{\partial \rho}{\partial x}$. In traffic flow, models with a single equation are called "first order models". Although mathematically, the order of a system of differential equations corresponds to the order of the highest derivative rather than the number of equations, we shall follow the convention used in the traffic flow literature. 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 [37] for a derivation). Different assumptions in how the system acts will lead to different ways of closing this conservation equation. Following [7], we can distinguish two main types of procedures for doing this:

- using $v = v[\rho]$, where v may be a functional of ρ. This models the dependency of velocity on density (and possibly density gradients).

- using $v_t = f(v[\rho])$, which explicitly models the time evolution of the velocity.

We will now go over specific ways to close the conservation equation that are special cases of the two procedures above.

(i) We use the following form for the velocity:

$$v = V_e(\rho) \tag{1.16}$$

This assumes that vehicles adapts their speed instantaneously to a change in local density; i.e. traffic flow is always in equilibrium described by the function $V_e(\rho)$. Multiplying this function by the density gives the fundamental diagram $q = \rho V_e(\rho)$. This model is named the LWR model after Lighthill & Whitham [39] and Richards [53] who developed it independently. The velocity function they suggested is the one due to Greenshields and given by $V_e(\rho) = v_{max}(1-\rho)$ with v_{max} the maximum velocity. This gives the fundamental relation:

$$q(\rho) = v_{max}\rho(1-\rho) \tag{1.17}$$

This FD was chosen as it seems to fit the empirical data (for example, compare the parabola (1.17) with figure (1.1 (c))). 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. Indeed, in free flow we have $q'(\rho) > 0$ so characteristics have positive speed, while in the congested regime they have negative speed. We define the characteristic speeds to be the speed of propagation of information in hyperbolic PDEs and are discussed in more detail in section (1.1.2.5). However the model predicts that the downstream shock front becomes smoother over time (see [62]), while empirical data suggests that the downstream front of jams should maintain its shape as it propagates (moving localized clusters in Schönhof and Helbing's classification described in section (1.1.1)). To fix this, one can choose a fundamental diagram that is linear in the congested regime (for

example the so-called triangular diagram suggested by Daganzo in [16]) which would keep the downstream front from losing its shape. However, this is only possible by a specific choice of the FD, and does not explain the scatter in the congested regime observed in data (see 1.1.c). Furthermore, this model cannot describe the spontaneous breakdown of traffic flow (as discussed in section (1.1.1)), and can only describe traffic flow in equilibrium.

(ii) Secondly, one can close the conservation of mass equation by apparent density. We do this by defining the velocity diagram to be:

$$v[\rho] = V_e(\rho^\star) \qquad \text{with } \rho^\star = \rho^\star(\rho, \rho_x) \tag{1.18}$$

The idea behind this procedure is that the driver perceives the density at their location but also anticipates the density around them. This type of model therefore considers vehicles not as particles but as driver-vehicle units, in which the behaviour of the driver is taken into account.

(iii) Another way to close the conservation equation is to consider the delay time between the external stimulus and the response of the driver-vehicle unit. Here we have:

$$v[\rho] = V_e(\rho(t-\tau, x)) \tag{1.19}$$

Bonzani in [7] shows that using this velocity diagram allows the formation of stop and go waves through instabilities, as can be found in real traffic flow.

(iv) Finally, one can use a dynamic velocity to close the conservation equation: this models the adaptation of vehicles to equilibrium velocity:

$$v_t = f(v[\rho]) = \beta(V_e(\rho) - v) \tag{1.20}$$

The difficulty here is choosing a value of β, as it has to be chosen to replicate unsteady features in traffic.

1.1.2.5 Macroscopic models: second order models

We now consider models which consist of the conservation of mass coupled with the conservation of momentum (for some function F of speed and density):

$$\begin{cases} \rho_t + (\rho v)_x = 0 \\ v_t + v v_x = F[\rho, v] \end{cases} \tag{1.21}$$

The second equation is inspired from the conservation of momentum equation in fluid flow. However, in traffic flow, momentum is not conserved (vehicles can break and speed up), and this equation is used to represent the dynamical behaviour of the average velocity. In these second order of models the main difference is in this second equation. Following Hoogendoorn and Bovy's review [30], we consider a class of "Payne type models" (also refered to as PW models), where the general form of the acceleration equation in (1.21) is:

$$v_t + v v_x = \overbrace{\frac{V_e - v}{\tau}}^{R} - \overbrace{\frac{1}{\rho} P_x + \frac{\eta}{\rho} v_{xx}}^{A} \tag{1.22}$$

Here, P is the traffic pressure and η is kinematic viscosity constant. Payne's model (can be found in the review [30]) corresponds to $P = \rho C_o^2$ and $\eta = 0$. The term R corresponds to the relaxation: traffic tends to equilibrium velocity with a certain relaxation time τ. The term A describes the drivers' reaction to a change in traffic density downstream (i.e. an average driver's *anticipation* of downstream conditions). There exist different interpretation of this term. Some authors consider the pressure term P_x to describe the local anticipation of drivers. Indeed, if traffic is denser upstream (modelled by the negative pressure term $-P_x < 0$), vehicles will lower their acceleration (i.e. the total derivative $v_t + v v_x$ will decrease). The term v_{xx} can be thought to describe the spatial rate of change of the acceleration: if the acceleration of vehicles increase, drivers will respond by increasing their own acceleration. These two processes compete to form the anticipation term. However other authors believe that it reflects the change of drivers' states (careful to brisk) (see [30] for a discussion).

However, it is worth stressing that these are interpretations of the various terms.

Ultimately, the ability of the model to reproduce observed features in traffic should decide whether they are appropriate. Only once the terms are shown to contribute to accurate models of traffic flow should one attempt to understand which physical processes they represent. Furthermore, even if a model reproduces features accurately, it is possible that some of the terms are artificial mathematical fixes and do not actually represent a specific process (this is especially relevant for continuum models).
It is also important to note that these Payne type models allow the formation of stop-and-go wave through instabilities (a condition for instability can be in found in [48] for example). This region of instability would allow the spontaneous formation of jams as described earlier in section (1.1.1).

PDE characteristics are an important property of these models; they correspond the wave propagation speed (and direction), or speed of propagation of information. To derive them from a model, consider a system of the form:

$$\mathbf{U}_t + \mathbf{A}\mathbf{U}_x = \mathbf{B} \tag{1.23}$$

Here, $\mathbf{U}$ are the conserved variables, $\mathbf{B}$ the source terms, and $\mathbf{A}$ the jacobian. The characteristics of equation (1.23) correspond to the eigenvalues of $\mathbf{A}$.

In the case of LWR discussed in the previous section, we have $\mathbf{U} = \rho$, $\mathbf{A} = q(\rho)_\rho = V_e(\rho) + \rho V_e'(\rho)$, and $\mathbf{B} = 0$. So the characteristic speed is $V_e(\rho) + \rho V_e'(\rho)$.

For the Payne-Whitham model discussed above (without viscosity), we get the following set of characteristic speeds:

$$\lambda_1 = v + C_o^2, \quad \lambda_2 = v - C_o^2$$

(with $P = \rho C_0^2$)

So this model exhibits waves that propagate in two direction, one faster than traffic and another slower than traffic.

In 1995, Daganzo wrote a paper clearly revealing the flaws in Payne type models (and of two equation macroscopic models in general) [17]. He points out that there are some main differences between traffic flow and fluid flow, namely:

- vehicles are anisotropic particles; they (almost) only respond to downstream

stimuli

- the width of a traffic shock is only of the order of several vehicles, so a macroscopic model cannot represent them accurately
- vehicles-driver units have a personality (timid, aggressive...) which is mostly unchanged by traffic conditions

He shows how behaviour of characteristics in Payne's model is unrealistic: the faster characteristic goes faster than the speed of vehicles, which violates the anisotropy of the vehicles. He also presents initial value problems with unrealistic behaviour, namely vehicles arriving in a queue and compressing it from behind for example. Daganzo argues that this type of behaviour is due to the diffusion term, which causes drivers in the model to be aware of conditions upstream as well as downstream.

Although some of these criticisms were valid at the time of publication, macroscopic models have been developed following this paper which resolve the issue of anisotropic vehicles. For example, the model developed by Aw and Rascle ([2]) does not have characteristics that travel faster than traffic. One can also develop nonlocal models, which include the density and velocity at an anticipated location instead of density and velocity gradients: $\rho_a(x,t) = \rho(x_a,t)$ with $x_a > x$ (similar for v_a) [58]. As for the last objection, continuous models can be used to represent traffic flow resulting from "average" drivers. Thus assuming homogeneous driver-vehicle units can be appropriate, although gas-kinetic models can also be used to derive multiclass models which take into account heterogeneous drivers ([30]).

One of the models which followed Daganzo's paper aiming to rectify these problems is Aw and Rascle's (AR) model [2]. Their model includes the convective derivative of the pressure term which they claim fixes the issues raised by Daganzo. Their acceleration equation reads as follows:

$$v_t + vv_x + P(\rho)_t + vP(\rho)_x = 0 \tag{1.24}$$

with the pressure $P(\rho) = \rho^\gamma$ for $\gamma > 0$. They describe various initial value problems and show that (almost) all of them are realistic and correspond to a set of physical principles

defined in the beginning of the paper. The characteristics of this model are $\lambda_1 = v$ and $\lambda_2 = v - \gamma P$; the faster characteristic moves with the average speed of vehicles, so does not go faster than traffic as recommended by Daganzo.
Another important model resulting from this debate is the one by Zhang ([64]) which was derived from microscopic considerations and which also rectifies the problems of characteristics faster than traffic found in Payne's model; its acceleration equation is:

$$v_t + (v + \rho V_e'(\rho))v_x = 0 \tag{1.25}$$

It has a relaxation term and does not assume the coefficient of v_x to be constant. Its characteristics travel with speed: $\lambda_1 = v$ and $\lambda_2 = v + c(\rho)$; as $c(\rho) \leq 0$, both characteristics travel slower than traffic.

Another model which claimed to improve on the PW models and was derived from car-following models is one due to Berg, Mason, and Woods (BMW) (an overview of the model can be found in J. Morgan's PhD thesis [44]). It has two characteristic speeds ($\lambda_{1,2} = v \pm \sqrt{\frac{-V_e'(\rho)}{2\tau}}$) and its acceleration equation is as follows:

$$v_t + vv_x = \frac{v_e - v}{\tau} + \frac{V_e'(\rho)}{\tau}\left(\frac{\rho_x}{2\rho} + \frac{\rho_{xx}}{6\rho^2} - \frac{\rho_x^2}{2\rho^3}\right) \tag{1.26}$$

Here, higher order terms are included for the density gradients, which are claimed to be significant.

In [44], after having researched several numerical methods for macroscopic traffic models, the author performed a simulation of LWR, PW, AR, BMW, and Zhang's model, and compared the results to traffic flow data from the M25 motorway during July 1999 (a long stretch of road without any on or off-ramps was chosen). The author first simulated the models while removing the relaxation terms and viscosity term (second derivatives of density and velocity) for the models which already included them. She then adds these terms in turn to the higher order models to examine their effect on the models' dynamics. The results of the comparison for the case with no relaxation or viscosity are as follows:

- LWR modelled the speed of the various waves reasonably accurately, but could not capture the interaction between the forward and backward moving waves (due to consisting only of one equation). However once a suitable fundamental diagram was chosen, it provided a robust simplification of how the real traffic behaved.

- PW modelled the speed of low density and high density waves accurately but captured the interactions between the waves poorly. Furthermore, the waves were not as distinct as they were with LWR and the model exhibited negative velocities.

- BMW had similar results to PW, but the waves were better defined and the interactions were closer to real data. Another improvement was that it did not exhibit negative velocities.

- AW exhibited high density waves that propagated in the wrong direction: downstream rather than upstream as in the data (it was the only model to do this). However the low and medium density waves were described adequately. Interestingly, it also exhibited negative velocities (even though it was designed in part to avoid this shortcoming).

- Zhang's model captured the desired features most accurately: it described the movement of waves of all density adequately, along with some of the interaction between waves. However, it predicted at times densities greater than the maximum density.

Adding a relaxation term improved the PW model, but Zhang's model performed better without it. It did not affect the performance of the other two models greatly. Adding viscosity terms generally smoothed out extreme values; for example it removed the higher non-physical density for Zhang's model.

Morgan's conclusions are that Zhang's model performed best with viscosity but without relaxation, and that the AW model was the least accurate. LWR seemed to be a robust model that reproduced adequately the global behaviour of waves.

It is also interesting to note that although the BMW model has a characteristic speed faster than the average velocity, it still did not produce negative velocities unlike

the AW model (which has characteristic speeds lower than vehicle velocity). This reminds us of the need to perform simulations of developed models and to compare them to real data to assess their validity. Indeed, we have seen earlier that there are many ways to interpret the various terms and mathematical features of models (especially when borrowing terms from fluid dynamics such as pressure); therefore complementing theoretical arguments with traffic flow data is essential.

Finally, it is worth noting that this study was performed on a single stretch of road and the traffic variables of the multilane carriage was averaged over all lanes. The assumption made is that the various effects of lane changing cancel each other out when taking the mean. However, Morgan points out in [44] that the drivers' behaviour may change as a result of these lane changes (and any on/off ramps found upstream of the studied stretch of road) which is not taken into account in the processing of the data (nor in the models themselves). Moreover, it is conceivable that the same comparison on a different length of road or point in time would lead to a different "ranking" of the chosen models, due to the realisation of different features in traffic. Another important consideration is the fitting procedure used for the parameters in the models; the author seems to have chosen many of these parameters manually following some pilot runs. The choice of boundary and initial conditions (as we will see in this thesis) can also have a large effect on the performance of the models; the author used density data and interpolated intermediate time points using cubic splines. Finally the numerical method used can also influence the performance of models and cause artifacts in the simulation.

1.1.3 Discussion

In this section we discuss the various issues surrounding macroscopic traffic flow modelling: the misuse of such models, the controversies about their dynamics and the different approaches to deriving them.

Firstly, it is essential when developing and using any model to understand that it is an imperfect description of reality (or more accurately, an imperfect description of the empirical data, which itself is an imperfect description of reality). This means that one must understand what a model attempts to describe, and not use it in an inappropriate context. Macroscopic models describe the macroscopic behaviour of traffic flow, namely the large scale behaviour *emerging* from the microscopic interactions between

individual vehicles. This means that these models are best used to describe long roads with few junctions (such as motorways). If one wishes to describe traffic flow along a short road with many junctions and few vehicles, discrete models (such as car-following or queueing theoretical models) might be more appropriate.

A well known controversy in the traffic flow community is the one initiated by Daganzo in [17]. He clearly exposes the flaws of the second order models present up to that date, and concludes that they are severe limitations intrinsic to all second-order continuous models. His criticism along with the response is discussed in an earlier section (section 1.1.2.5). However in 2008, Helbing and Johansson revisited the discussion in [28] and conclude that although some of the points raised by Daganzo were fully justified (backward moving vehicles or queues compressing from behind), the issue of characteristic speeds greater than vehicle speeds does not constitute a contradiction. They analyse a general class of macroscopic models and discover that the faster characteristic can be within the vehicle velocity variance (described in the pressure term). In these models the vehicle velocity is an *average velocity*, so variation around that mean is allowed. Indeed, Papageorgiou in [50] agrees that macroscopic models describe an average vehicle speed, which allow for variation within that average. The author also points out that instabilities may be caused by microscopic inhomogeneities (some vehicles being faster than others) which supports this view. Other researchers have responded to this debate with a discussion about the correct way to measure the propagation of information in microscopic model which would allow a better comparison with macroscopic models, and so this controversy seems not to have been settled.

Bonzani in [7] raises another issue in traffic flow modelling: she argues that one should include in traffic flow models a parameter relating to the quality of the road and the environmental conditions. This would allow models to describe the change in traffic flow behaviour on different roads and in different conditions. The difficulty here is quantifying these conditions, ”averaging” out their effects, and summarising them into a single parameter.

Macroscopic models are usually developed from microscopic assumptions. Indeed Bonzani argues in [7] that these continuous models *must* be developed from microscopic considerations. She presses the importance that the various properties of models (for example the fundamental diagram describing the equilibrium) should emerge

rather than be artificially plugged in. Indeed in [6], she derives the fundamental diagram by modelling individual driver behaviour. Although this is an excellent way to develop continuous models (as macroscopic behaviour originates from microscopic interactions), one can obtain many different types of macroscopic behaviour by choosing suitable microscopic assumptions. This is due to the fact that human drivers (and hence inhomogeneity and unpredictability) are at the core of traffic flow. Indeed Papageorgiou in [50] argues that even if accurate microscopic models were available, one could not deduce accurate macroscopic models from them (this is also due to the fact that there are relatively few vehicles on the road, which make the use of continuous models difficult). He presses that "any macroscopic traffic flow model that did not undergo a validation procedure against real traffic data must at best be considered a pending suggestion" (p.328). In summary, a main issue in traffic flow modelling is that there are many human and environmental factors affecting its dynamics. As a result, summarising the system in a set of equations is unlikely to ever reproduce satisfactorily the various features observed in traffic. We also noticed that theoretically sound models can perform badly (for example, the AR model did not perform well compared to data in the previous section); this encourages the use of data to check and even drive theory rather than relying only on theoretical justification for certain terms or properties of models (which are of course essential). As there are so many different approaches one can take (which terms to include in the model? Which fundamental diagram? What characteristic speeds?), it is of interest to *constrain* our modelling using data. With this goal in mind, we give an overview in the next sections of an approach to fitting models to data using Bayesian statistics.

1.2 Bayesian inference

1.2.1 Overview of Bayesian inference

Bayesian inference is about fitting a probability model to data and obtaining a probability distribution describing the fitted parameters (see [21] for a reference for doing Bayesian inference in practice). This distribution can then be used to summarise the result using expectations such as the mean or variance of the parameter of interest. In the Bayesian framework (as opposed to the Frequentist one), the parameters θ in the model are random and the data X is fixed. We start with a prior belief for the parameter

of interest described by a *prior distribution* $\pi_0(\theta)$ and wish to update this distribution in light of data X. We assume that data was generated by the model $p(X|\theta)$ and use this to build a *likelihood* $\mathscr{L}(\theta)$; if we assume data is independent and identically distributed (iid), we can build the likelihood as $\mathscr{L}(\theta) = \Pi^N p(X_i|\theta)$ (with N number of observations). We then use Bayes' theorem to update the prior belief with the likelihood to optain a *posterior distribution* $\pi(\theta|X)$:

$$\pi(\theta|X) \propto \mathscr{L}(\theta)\pi_0(\theta) \tag{1.27}$$

We can them summarise the posterior distribution for θ with expectations $E_\pi[g(\theta)]$ with g a function of the parameters, such as the mean $E_\pi[\theta]$ or the variance $E_\pi[(\theta - E_\pi[\theta])^2]$. Furthermore, the posterior can be re-used as a new prior distribution in light of new observations.

The choice of prior is supposed to reflect the practitioner's knowledge of the parameter before seeing data; for example by defining a reasonable range of values the parameter can take, or by specifying the smoothness of a function to infer. However this prior can also be chosen for computational convenience; for example being able to sample from the prior is essential for certain MCMC methods. If the prior is chosen such that the posterior is in the same probability distribution family as the prior, then these distribution are called *conjugate distributions*. If this is possible then inference can be greatly simplified, for example by having the posterior available analytically.

Suppose we have a sample of iid observations $X = (X_1, X_2, ...X_N)$ from a Gaussian distribution $\mathscr{N}(\mu, \tau^{-1})$ with μ the mean to infer and known precision τ, which we summarise in the sample mean $\bar{X} := \frac{1}{N}\sum^N X_i$. If the prior for μ is $\pi_0(\mu) \sim \mathscr{N}(\mu_0, \tau_0^{-1})$, then we have conjugacy and obtain a Gaussian posterior $\pi(\mu|X) \sim \mathscr{N}(\mu_n, \tau_n^{-1})$ (see [21]) with

$$\mu_n = \frac{\mu_0\tau_0 + N\bar{X}\tau}{\tau_0 + \tau N} \tag{1.28}$$

$$\tau_n = \tau_0 + \tau N \tag{1.29}$$

1.2.2 Overview of Markov Chain Monte Carlo

As discussed in the previous section, in Bayesian inference (and indeed in other disciplines such as computational chemistry or physics) we are interested in integrals of the form:

$$I = E_\pi[g(X)] = \int g(x)\pi(x)dx \tag{1.30}$$

If we can obtain N iid samples $X_i \sim \pi$, we can approximate the integral I by:

$$\hat{I}_N = \frac{1}{N}\sum^{N} g(X_i) \tag{1.31}$$

And by the Central Limit Theorem (CLT) (see [40]), we obtain (with $\sigma^2 = \text{var}[g(X)]$):

$$\sqrt{N}(\hat{I}_N - I) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \tag{1.32}$$

However in many applications of interest (for example if the support of π is high dimensional), obtaining independent samples from π is difficult. An alternative approach to overcome this problem is to obtain *correlated* samples from π by defining a Markov chain whose equilibrium distribution is π; this family of algorithms is called Markov Chain Monte Carlo (MCMC).

A general approach introduced in [43] and generalised in [27] is the Metropolis Hastings algorithm, given below.

Metropolis Hastings algorithm: given current state $\mathbf{x}_t \in \mathscr{R}^N$ and proposal distribution $q(x|y)$:

- Sample $\tilde{\mathbf{x}} \sim q(\tilde{\mathbf{x}}|\mathbf{x}_t)$
- Define the acceptance probability: $\alpha = \min\left\{1, \frac{\pi(\tilde{\mathbf{x}})q(\mathbf{x}_t|\tilde{\mathbf{x}})}{\pi(\mathbf{x}_t)q(\tilde{\mathbf{x}}|\mathbf{x}_t)}\right\}$ and sample $u \sim \mathscr{U}(0,1)$. The ratio $\frac{q(\mathbf{x}_t|\tilde{\mathbf{x}})}{q(\tilde{\mathbf{x}}|\mathbf{x}_t)}$ is named the *Hastings correction*.
- Update: $\mathbf{x}_{t+1} = \begin{cases} \tilde{\mathbf{x}} & \text{if } \alpha \geq u \\ \mathbf{x}_t & \text{otherwise} \end{cases}$

A lot of the difficulty in tuning MCMC in practice is choosing the proposal distribution so that the chain *mixes* (an overview of what this means and diagnostics is given later). A common choice for the proposal distribution is a Gaussian:

$q(\mathbf{x}|\mathbf{x}_t) \sim \mathcal{N}(\mathbf{x}_t, \sigma^2)$; the algorithm is then called random walk Metropolis Hastings (RWMH).

If we obtain correlated samples $X_i \sim \pi$ from our distribution, then we can use the same estimator $\hat{I}$ as in equation (1.31), but applying the CLT now results in:

$$\sqrt{\frac{N}{2\tau_{int}}}(\hat{I}_N - I) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \tag{1.33}$$

With τ_{int} the *integrated autocorrelation time* of g given by $\tau_{int}(g) = \frac{1}{2} + \sum_1^\infty \rho_j$ and $\rho_j = \text{corr}[g(X_1), g(X_j)]$ (see [40]). This is similar to the Gaussian in equation (1.32) but with the *effective sample size* $ESS = \frac{N}{2\tau_{int}(g)}$ which corrects for the correlations in the samples. We note that this term depends on the function g and hence must be recalculated for every integral.

1.2.3 MCMC in practice: diagnostics

Running MCMC is in principle a simple task: one defines a Markov chain whose equilibrium distribution is the distribution of interest and one simulates this chain on a computer. If MCMC is successful we say that the chain mixes. A MCMC chain X_t mixes if starting from any initial distribution, its distribution at $t \to \infty$ is the target (see [38]). Namely for some initial distribution f_0 and target distribution π, if $X_0 \sim f_0$ then $X_t \xrightarrow{d} \pi$. An important obstacle to chains mixing is disconnected regions of parameter space, as the chain can fail to explore all these regions.

However when the chain is run there are many questions the practitioner must ask themselves to ensure that the samples are truly representative of the target distribution. Has the chain run for long enough to converge to the target distribution? Does the posterior have multiple modes, namely regions of parameter space separated by low probability regions that all contribute to the total mass of the distribution? If so, does the chain explore these modes in a reasonable amount of time (given the computational budget available)? We can assess the efficiency of the algorithm by estimating the mixing speed which is linked to how correlated the samples are. At the extreme case, if samples are extremely correlated it can take an unreasonable amount of computing time to obtain accurate estimators. This can be seen in the CLT in equation (1.33): long correlations will result in a large value of $\tau_{int}(g)$ which will lower the ESS, resulting in a large variance in the estimator $\hat{I}$. Even more problematic, if the chain does not mix

then samples are unrepresentative of π and the estimator $\hat{I}$ will simply not converge to I.

This difficulty in tuning Monte Carlo methods and their low of efficiency led Alain Sokal to famously warn in [49] that "Monte Carlo is an extremely bad method; it should be used only when all alternative methods are worse". This warning is due in part to the difficulty in running any non-trivial numerical method (there are many ways to make mistakes) and in part due to the slow convergence of Monte Carlo estimators (they have error rate $\mathcal{O}(N^{-1/2})$, as can be seen from equation (1.33)). In contrast, there are deterministic numerical method that have faster rates of convergence in low dimensions. However in high dimensional problems Monte Carlo methods may be the best (and sometimes only) choice. In these cases, one must cautiously check convergence of the method and attempt to lower the correlations in the samples thus improving the proportionality constant in the $\mathcal{O}(N^{-1/2})$ error rate.

We now give an overview of some MCMC diagnostics. These diagnostics are all necessary but not sufficient conditions for mixing, so one can never be certain of convergence. As a result one must check enough of these diagnostics to build confidence in the samples, and if a single one fails the chain must be considered to not have converged. Of course in practice the amount of care taken in the diagnostics should be roughly proportional to the difficulty of the problem: more diagnostics should be used for high dimensional problems resulting from complicated nonlinear models than for simple low dimensional problems. We offer below a non-exhaustive list of diagnostics and good practices for MCMC (see [9] for algorithms and practical considerations related to doing MCMC in practice):

- Run 3 or more chains in parallel and plot the samples as a function of iteration number (called trace plots). The samples from each chain should overlap frequently with one another during the course of the run.

- Compute the variance V of the pooled chains (ie: all chains grouped together) and the average within-chain variance W. Compute the R.hat (or "potential scale reduction factor"): R.hat $= (\frac{V}{W})^{1/2}$. If a chain has perfectly converged then R.hat $= 1$. However, Andrew Gelman and Kenneth Shirly in [9] recommend using a limit of R.hat < 1.1 for MCMC to "pass" this diagnostic.

- Run MCMC on a simpler model, for example one that is a special case of the complicated model. This allows the practitioner to learn about a simpler posterior with a similar structure. For example the simpler model may be lower dimensional, may have certain conditional distributions available analytically, or may simply run faster. As a result one can try a wider range of proposals and settings to see which work best. It maybe be possible that the complicated model "inherits" sampling properties from the simpler one.

- Plot the autocorrelation function (ACF) for many lags and check when the autocorrelation tends to 0. One can also compute the integrated autocorrelation time τ_{int} to summarise the mixing speed. Naively, one could use the estimator $\hat{\tau}_{int}(N) = 1 + \sum_1^N \rho_j$, but Sokal in [49] points out that the variance does not go to zero as N goes to infinity. This is because the *sample* autocorrelation contain more "noise" than "signal" past the number of lags where the autocorrelations are zero. As a result he recommends cutting off the sum at a constant M such that $M \geq c\hat{\tau}_{int}(M)$, with c roughly between 5 and 10. However, this recommendation is given provided $N \geq 1000\tau_{int}$ which is not the case for the samples in this thesis; this estimator is not robust for low numbers of effective samples. As a result we will rather use the estimated *decay time* which we denote by $\hat{\tau}_d$, which is the lag that corresponds to an autocorrelation of e^{-1}. This estimate is more robust as the gradient of the ACF at that point usually still has a high magnitude: estimating this value by eye is therefore practical. We will use this estimator to compare the mixing speed of samplers, though as we will estimate this visually (from the ACF plots) we will consider this to be a fairly approximate way of comparing chains.

- Remove the initial samples from each chain (called the burn-in); the amount of samples removed should be roughly past the point where the autocorrelation goes to zero. As a result the initial conditions are "forgotten". This is especially important if the samples start far away from the mode of the distribution and need time to converge to it. Andrew Gelman and Kenneth Shirly in [9] (Chapter 6) recommend discarding the first half of a run, and if the chain is run again (continuing where the chains left off) discarding half the samples again. So if

one runs a sampler for 100 iterations, one should discard the first 50 samples. If one then runs 100 more iterations, one should discard 50 more samples so that the total number of samples is now 100. However if the model used in the likelihood is computational expensive, discarding half the samples is very wasteful. Using the autocorrelation plots instead to estimate burn-in can be a more efficient method in this case.

- Start chains for random starting points in parameter space; this is called the *multistart heuristic*. One can start either from a sample from the prior, or from an overdispersed distribution (relative to the posterior). For example one could start from the mode (determined from previous runs) and apply a Gaussian error with a large variance. In practice, if the prior was chosen for computational convenience and one has domain knowledge of the posterior mode, starting from a sample from the prior can be wasteful as the sampler might take a long time to reach the posterior mode. In these cases, the second option can be reasonable as well as computationally more efficient. The benefit of the multistart heuristic is that it can help find modes in the posterior that were previously unexplored. If this happens, one must either design a new proposal that can mix between the modes or use MCMC algorithms designed to target multimodal distributions.

- A final diagnostic is simply to run a chain for a long time. Of course what a ”long time” means depends on the autocorrelations in the samples as well as the available computational budget. A long run allows the chain to explore regions of parameter space that contribute a small but non-negligible amount of mass to the total distribution. A long run also allows for detection of *pseudo-convergence* which is when a Markov chain appears to have converged to a distribution when in fact it has not. This can happen when two parts of the state space are poorly connected and the expected time it takes for a chain to move from one to another is higher than the length of the run. Geyer (a strong advocate for long runs) in [9] has a dictum that “the least one can do is to make an overnight run” (Chapter 1, page 19). A second dictum by the same author is that one should start a run when the paper is submitted and keep running until the referees’ reports arrive. This of course must be tempered by practical considerations: if one is using servers that

are shared with other researchers it is difficult to monopolise their use, and if one is renting servers then one must have a considerable (monetary) budget available to put this dictum in practice.

When one is running MCMC on a complicated posterior it is advised to try a wide range of diagnostics. We reiterate that no diagnostic is perfect but that each helps build the confidence in the samples. In contrast, Geyer in [9] argues against the multistart heuristic by claiming that because one cannot start from all parts of the state space (and thus check whether there are modes in the posterior) then this heuristic is "worse than useless: it can give you confidence that all is well when in fact your results are completely erroneous" (Chapter 1, page 19). Of course one cannot start from every corner of state space so one never has guarantees. This can however detect multimodality that no other diagnostic would find. Therefore one must always proceed with caution and use a range of diagnostics.

1.2.4 MCMC in practice: software

A separate but related issue to diagnosing mixing is writing MCMC code. Even if one designs a mathematically correct and well tuned MCMC algorithm, buggy MCMC code can generate samples that are not from the desired distribution. In this section we give an overview of some methods for testing MCMC code (see [25] for a overview of such methods).

There are several difficulties with testing MCMC code:

- The algorithm is stochastic, so there is no single "correct" output. The output is rather a distribution that is unknown before writing the software (if it were known then the software would not be necessary)

- The software may work in simple cases but fail in more complicated cases. So testing the software on a simple distribution may not find a bug that only appears when sampling from more complicated distriutions.

A first thing to keep in mind is to write modular code; namely to structure code into re-usable and loosely coupled blocks. For example in an object oriented paradigm, one might create a `move` class which can propose samples and do the Metropolis Hastings accept reject step, a `backend` class which stores previous samples and regularly saves

them to file, and a `sampler` class which runs the main MCMC loop. The benefits of a modular structure is that is encourages the programmer to think about the structure of the program and to create re-usable components, which means that the components are easier to test.

We distinguish two types of tests: unit tests and integration tests ([25]). Unit tests are about testing small chunks of code such as individual functions and checking that they have the required functionality. Integration tests are about testing the entirety of the software without reference to individual parts. Writing tests simultaneously to writing code also encourages the programmer to explicitly state what the expected output should be as well as consider under what conditions the code could fail. Very similarly to MCMC diagnostics, code with comprehensive tests allow the practitioner (and any user of the software) to trust the outputs.

We now present a non-exhaustive list of tests one could write (besides unit tests) for MCMC code:

- Test the software against analytically known solutions. For example, one can use the MCMC sampler to sample from a Gaussian (both 1 dimensional and higher) and to check the marginal distributions along with the empirical covariance matrix of the samples. This also applies to deterministic software, such as PDE solvers: one can test a solver for a PDE against initial conditions with known solution.

- For a Gibbs sampler (as described in [25]), one can check that the conditional distributions are consistent with the joint distributions. Namely, for any two samples x and x' one can check that the following equality holds: $\frac{p(x'|z)}{p(x|z)} = \frac{p(x',z)}{p(x,z)}$. For most models, if the formula used in the conditional distribution is wrong, then this equality will not hold.

- For any MCMC sampler one can run the following test (sometimes called the prior reproduction test): sample from the prior $\theta_0 \sim \pi_0$, generate data using this prior sample $X \sim p(X|\theta_0)$, and run the to-be-tested sampler long enough to get an independent sample from the posterior $\theta_p \sim \pi(\theta|X)$. The sample from the posterior should be distributed according the prior (derivation in theorem (A.0.1) in the appendix); one can repeat this procedure to obtain many samples θ_p and test

whether they are distributed according to the prior. The authors in [14] suggest a slight variation on this idea: at each replication they sample from the posterior L times (and obtain the samples $(\theta_1, \theta_2, ..., \theta_L)$) and estimate $\hat{q} = \frac{1}{L}\sum_1^L I_{\theta_0 > \theta_1}$. By repeating this test many times (thus obtaining many realisations of $\hat{q}$), they check that $\hat{q}$ is uniformly distributed. For whichever of the two versions one chooses, this type of test has a high coverage (ie: it tests many aspects of the software at once) but can be very computationally expensive to run, as one needs to run the sampler on the posterior for each prior sample θ_0. Furthermore, if the MCMC sampler needs extensive hand tuning then it is not practical to get enough samples θ_p to reliably assess their sampling distribution. One can however mitigate this problem by running the test for blocks of parameters (ie: conditional distributions) rather than for the entire posterior at once.

- In some cases, simply sampling from the prior using the to-be-tested sampler can be a powerful test. This can be very easy to implement (simply remove the likelihood from the posterior) and can run very quickly. If this method is practical for the sampling problem, it can test almost the entire framework (with the exception of the likelihood); for example it can test that the Hastings corrections in the proposals were correctly calculated and implemented. It might however not be applicable in some cases, for example where the prior is not proper or where sampling from the prior is qualitatively different than sampling from the posterior. In the latter case, one might have to implement a different proposal which slightly defeats the purpose of testing the code. However in cases such as the Bayesian inverse problem considered in this thesis, this test is very easy to perform and is almost as powerful as the prior reproduction test (which in this case is impractical to do on the entire posterior).

Similarly to MCMC diagnostics, these tests do not ”prove” that a piece of software works correctly, they rather increase our confidence in it. Furthermore, the amount and variety of automated tests must be roughly proportional to the complexity of the code base.

1.3 Bayesian inverse problems

1.3.1 General framework

Following [54] we give a general overview of inverse problems in the Bayesian framework. Consider data generated by an *observation operator* $\mathcal{G}$:

$$q = \mathcal{G}(\theta) + \eta \tag{1.34}$$

with

- q the observed quantity
- θ the parameter of interest,
- $\mathcal{G}$ the observation operator, which is a mapping from the parameters to the data; this could be the composition of a PDE and an operator that picks out points in the $x-t$ plane corresponding to observations. This is also called the forward problem.
- $\eta \sim \mathcal{N}(0,\Gamma)$ noise that pollutes the observed data with covariance matrix Γ

A first attempt at a solution for this inverse problem might be a least-squares solution (using some appropriate norm $||.||$): $\hat{\theta} = \text{argmin}_u ||\mathcal{G}(\theta) - q||^2$. However, as inverse problems are generally ill-posed, the solution may not be unique or even exist, and may have sensitive dependence on q. A way to overcome this is to impose a prior on the parameter which acts to regularise the problem. If the observational noise is Gaussian and we choose a Gaussian prior $\pi_0(\theta) \sim \mathcal{N}(m_0, \Sigma_0)$ we obtain the posterior:

$$\pi(\theta) \propto \exp\{-\frac{1}{2}||\mathcal{G}(\theta) - q||^2_\Gamma - \frac{1}{2}||\theta - m_0||^2_{\Sigma_0}\} \tag{1.35}$$

A special case of the above is when the observation operator can be written as a linear operator $\mathcal{G}(\theta) = A\theta$ (for some invertible matrix A). Applying theorem 3.1 in [54] the posterior is therefore Gaussian, ie: $\pi(\theta) \sim \mathcal{N}(m, \Sigma)$ with:

$$m = m_0 + \Sigma_0 A^T (\Gamma + A\Sigma_0 A^T)^{-1}(q - Am_0) \tag{1.36}$$

$$\Sigma = \Sigma_0 - \Sigma_0 A^T (\Gamma + A\Sigma_0 A^T)^{-1} A\Sigma_0 \tag{1.37}$$

Furthermore, [54] explores Bayesian inverse problems mainly for elliptic and parabolic PDEs and shows that the existence of the posterior depends on the smoothness of the observation operator. When the observation operator is nonlinear it may not be possible to obtain the posterior analytically. In these cases one may have to resort to Monte Carlo methods to sample from the posterior distribution. In these cases, derivative information about the log posterior (if available) can be used in proposals that are more informative than random walk, as studied in [46].

1.3.2 Hyperbolic Bayesian inverse problems

We now give an overview of some applications of Bayesian inverse problems for hyperbolic PDEs in the literature.

A study close to the topic of this thesis is a Bayesian analysis of traffic flow tested on motorway data by Polson and Sokolov in [51]. The objective of the paper is to develop a methodology (using particle filtering) to estimate in real time traffic density and parameters in the LWR model (introduced in section (1.1.2.4)). This allows real-time estimation of road capacity (maximum possible flow on the road) and critical density (density at which flow is maximised) 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} (the percentage of time a point on the road is occupied by a vehicle), which they use to estimate density. They use this quantity to estimate density with the formula $\rho = \frac{\omega_{occ}}{L}$ (see [29]) 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 they obtain it. See section (3.1) for a discussion of traffic density estimation methods.

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

$$\begin{cases} y_{t+1} = H_{t+1}\theta_{t+1} + \varepsilon^V_{t+1} \quad \text{with} \quad \varepsilon^V_{t+1} \sim \mathcal{N}(0, V_{t+1}) \\ \theta_{t+1} = f_\phi(\theta_t) + \varepsilon^W_{t+1} \quad \text{with} \quad \varepsilon^W_{t+1} \sim \mathcal{N}(0, W_{t+1}) \end{cases} \tag{1.38}$$

Where V_t and W_t are evolution and equation error respectively, y_{t+1} is the vector of measured traffic density, f_ϕ is the LWR evolution equation with Fundamental Diagram parameters ϕ calculated using a Godunov scheme (see [37] for a comprehensive account of these numerical methods and chapter 3 of this thesis for an overview). The observation matrix H_{t+1} picks out the cells with the measurements. The objective of the methodology is to sample from the posteriors $p(\theta_t|y^t)$ and $p(\phi|y^t)$ with $y^t = (y_1, ..., y_t)$ the current history of data.

They consider a length of road on an interstate outside Chicago which seems to have two detectors separated by 845m. They discretise this distance into 4 cells (so each cell corresponds to a distance of 211m) and use a time discretisation of 5 minutes (which seems to be the resolution of data available). They run the analysis for 24 hours 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 boundary conditions. For this simulated data they use a spatial resolution of 300m and a time resolution of 5 seconds over a total road length of 1.5km and a time horizon of 1600s.

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 aim to develop 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 summarise in chapter 3; we will therefore impute the boundary conditions rather than estimate them directly from data.

Another issue with the methodology is the coarse discretisation of the LWR model (211m and 5min 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 (3.2) of chapter 3. Furthermore, each time step in the PDE solver includes a Gaussian error term as seen in equation (1.38); 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 [37]) does not include a random term. Adding a Gaussian error at each time step seems more like the discretisation of a stochastic PDE (the LRW model has no stochastic component), although it is unclear whether this discretisation would indeed converge to a specific SPDE under refinement of the grid. Whether or not this method would converge may depend on the scaling of the variance in the Gaussian error with discretisation step. Moreover, adding a Gaussian error at each time step amounts to adding or removing a random fraction of a vehicle thus violating the conversation of total number of vehicles (ie: conservation of mass). The model used in the paper should therefore rather be considered as an ad-hoc discrete model inspired by the Godunov method for LWR rather than a discretisation of the PDE.

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

Another Bayesian treatment of hyperbolic PDEs in the literature is estimation of parameters in gas networks in [26]. The authors estimate the distribution of some quantities of interest based on a finite number of noisy measurements at the boundary of a 1D pipe using simulated data. As industry regulations ensure that the gas evolves smoothly in gas networks there are no shocks in this system; as a result they consider a simplified version of the Euler equations:

$$p(x,t)_t + c^2 q(x,t)_x = 0 \qquad \text{in } \Omega \times (0,T) \tag{1.39}$$

$$q(x,t)_t + p(x,t)_x = \lambda a(p(x,t),q(x,t)) \qquad \text{in } \Omega \times (0,T) \tag{1.40}$$

With $\Omega := (0,1)$, $T > 0$, $p(x,t)$ the pressure of the gas, $q(x,t)$ the momentum, and $c > 0$ relating to the speed of sound. $a(\cdot,\cdot) : \mathscr{R} \times \mathscr{R} \to \mathscr{R}$ is the friction function and $\lambda \in L^\infty(\Omega)$ the friction coefficient; λa describes the roughness in the interior walls of the pipe. Usually λ is chosen to be a scalar value, but as this coefficient can vary along the length of the pipe, they estimate it nonparametrically. Using some regularity conditions of the forward problem (as the PDE does not allow shocks to form) they

prove well-posedness of the inverse problem. They then generate synthetic data (for different types of functions λ) and estimate it using MCMC (using the pCN algorithm defined in section (1.3.3.6)).

Another interesting study of hyperbolic inverse problems is a paper ([33]) describing how to evaluate gradients of LWR with respect to initial conditions to use in iterative optimisation algorithms even in the presence of shock waves. We first consider a hyperbolic PDE in conservation form:

$$\rho_t + f(\rho)_x = 0 \tag{1.41}$$

If we integrate this against continuous test functions $\phi(x,t)$ we obtain:

$$\int_0^\infty \int_{-\infty}^{+\infty} [\rho_t + f(\rho)_x]\phi(x,t)dx,dt \tag{1.42}$$

Integrating by parts we then obtain:

$$\int_0^\infty \int_{-\infty}^{+\infty} [\rho\phi_t + f(\rho)\phi_x]dxdt = -\int_0^\infty \rho(x,0)\phi(x,t)dx \tag{1.43}$$

As pointed out in [37], the nice feature of equation (1.43) is that the derivatives are now on ϕ rather than on ρ and $f(\rho)$; the equation therefore continues to make sense even if ρ is discontinuous. This then motivates the following definition (also from [37]):

Definition 1.3.1. The function $\rho(x,t)$ is a weak solution of the conservation law (1.41) with given intial data $\rho(x,0)$ if (1.43) holds for all functions ϕ in C_0^1 (all continuous functions with compact support).

The study in [33] uses the weak formulation of LWR to evaluate gradients to use in optimisation problems. This research could perhaps be extended to formulate gradient based proposals in MCMC for hyperbolic PDEs in the presence of shocks; we discuss this further in section (6.2).

1.3.3 Some MCMC methods for inverse problems

We now consider in this section some MCMC methods for inverse problems such as posteriors arising from PDE models. These problems may have the following qualitative properties:

- The forward model might only be available numerically (using a solver) and may be computationally expensive. As a result the sampler needs to efficiently explore the posterior.

- The parameter to infer may be high dimensional and may even result from the discretisation of a function or a field.

- Conditional distributions may be unavailable analytically; this is likely if the forward model involves a nonlinear PDE for example. In this case Gibbs sampling is not possible.

- Gradients may be unavailable analytically or may not even exist (due to discontinuities in the likelihood for example). Furthermore, it may be impractical to rewrite the numerical solver for the forward problem in a language or framework that can obtain gradients automatically (such as the Stan framework [11]).

1.3.3.1 Single site vs multivariate

We consider the inverse problem in Electical Impedance Tomography (EIT) studied in [9] (chapter 16). The task is to infer a 2D field $x(s)$ (with $s \in \Omega$ and $\Omega = [0,1] \times [0,1]$) discretised into a $m = 24 \times 24$ lattice using observations at 16 electrodes evenly located around its edge. The forward model is a PDE solved numerically, and a Markov random field prior π_0 is used for x:

$$\pi_0(x) \propto \exp\{\beta \sum_{i \sim j} u(x_i - x_j)\} \qquad x \in [2.5, 4.5]^m \tag{1.44}$$

With $i \sim j$ denoting horizontal and vertical nearest neighbours, and with $u(\cdot)$ defined by:

$$u(d) = \begin{cases} \frac{1}{s}(1 - |d/s|^3)^3 & \text{if } -s < d < s \\ 0 & \text{if } |d| \geq s \end{cases} \tag{1.45}$$

This prior encourages neighbours to have a similar value, but also occasionally allows large shifts between neighbours.

The authors try a single-site Metropolis sampler, which can also be called a Metropolis within Gibbs sampler. At each iteration, the algorithm scans through each

parameter in the lattice and performs a random walk Metropolis update conditional on the other parameters in the lattice being fixed. They find that the sampler mixes well and that the posterior exhibits bimodality for some of the parameters. They then try a multivariate random walk Metropolis algorithm, which updates the entire lattice at once using a Metropolis sampler. They test different covariance matrices for the Gaussian proposal but find that the performance is noticeably worse than for the single-site update. The lesson from this is that when the space is high dimensional, updating all the parameters at once can be inefficient compared to updating one parameter (or blocks of parameters) at a time.

1.3.3.2 Augmenting the state space with fast approximate simulators

If solving the forward problem is too computationally expensive, one approach can be to augment the state space with faster but approximate solvers. One can do this by introducing an extra parameter $\nu \in [0, 1, .., k]$ denoting a sequence of increasingly coarse approximations of the forward problem (with $\nu = 0$ being the high resolution one). For example in the EIT problem above the simulator can be modified to be computationally cheaper at the cost of a lower level of accuracy.

The sampler is run in two steps: in the first step a single-site Metropolis update is done on each parameter as before (with ν fixed) and in the second step a Metropolis update is done for ν using a discrete random walk as proposal. This scheme allows the sampler to explore the parameter space quickly when the resolution of the solver is lower. Although the final samples are only the ones conditioned on $\nu = 0$, this scheme can result in more efficient sampling as in the case of the EIT study. However more generally, in the case of the forward problem being a PDE it is expected that these efficiency gains will be more apparent for 2 or 3 dimensional cases rather than 1 dimensional cases.

1.3.3.3 Ensemble methods

A popular class of algorithms for black box forward problems is ensemble methods. Some popular versions of these have the property that they do not require gradients, are easy to parallelise, and are simple to understand and implement. One well known example is the Affine Invariant Ensemble Sampler (AIES) from [22] which we introduce here.

Firstly, consider $X \sim \pi_1$ and an affine transformation of X with A an invertible matrix: $AX+b =: Y \sim \pi_2$. A random walk Metropolis Hastings sampler would perform differently on these two transformed distributions (or rather one would need to tune the proposal for each distribution). An affine invariant sampler however would consider sampling from π_1 and π_2 to be equally difficult. For example the performance of an affine invariant sampler on the Gaussian $\pi(x) \propto \exp\{-\frac{1}{2}x^t \Lambda x\}$ would be unaffected by the choice of precision matrix Λ.

The authors in [22] introduce a family of affine invariant ensemble samplers. Define an ensemble $\vec{X} := (X_1, ..., X_L)$ of L walkers $X_k \in \mathscr{R}^N$. The target distribution is the one in which each walker is independent and sampled from π: $\Pi(\vec{x}) := \pi(x_1)\pi(x_2)...\pi(x_L)$, so an ensemble MCMC algorithm is a Markov chain on the state of ensembles. One iteration of the ensemble $\vec{X}(t) \to \vec{X}(t+1)$ involves updating each walker one at a time. Given a walker X_k, define the complementary ensemble to be $\vec{X}_{[k]}(t) := \{X_1(t+1), ..., X_{k-1}(t+1), X_{k+1}(t), ..., X_L(t)\}$. We introduce one of the proposals suggested in the paper: the *stretch move*. To update walker $X_k(t)$, pick $X_j \in \vec{X}_{[k]}(t)$ and propose $Y = X_j + Z(X_k(t) - X_j)$ with $Z \sim g(z)$ and tuning parameter $a > 1$. We define the distribution g to be:

$$g(z) = \begin{cases} \frac{1}{z} & z \in [\frac{1}{a}, a] \\ 0 & \text{otherwise} \end{cases} \tag{1.46}$$

So for $Z = 1$, the proposal is to stay at X_k, and other values of $Z \in [\frac{1}{a}, a]$ move Y on the line joining X_k and X_j. A usual Metropolis Hastings accept-reject step is then performed, and the authors show that this proposal is accepted with probability:

$$\min\{1, Z^{N-1} \frac{\pi(Y)}{\pi(X_k(t))}\} \tag{1.47}$$

Furthermore, to ensure that there is no d-dimensional subspace of $\mathscr{R}^N$ (with $d < N$) that contains all the walkers in the ensemble, the algorithm requires having more walkers than the dimension of the distribution ($L \geq N$).

This algorithm was implemented in the popular Python package *emcee* ([19]), which has been used especially in astrophysics. Compared to other popular MCMC frameworks such as Stan this framework only requires the log posterior and can be

easily parallelised. The fact that one can easily use this sampler on complicated black box forward problems (perhaps written using legacy software) could be one of the factors that contributed to its success.

However a limitation with this sampler (explored experimentally and mathematically in [31]) and possibly ensemble methods in general is that it does not scale well with the dimension of the posterior. An intuitive explanation for this given in [10] is that creating a proposal by interpolating or extrapolating from samples in a distribution in high dimensions is unlikely to fall within the so-called *typical set*. The typical set (as defined for example in [5]) is the region in parameter space of a distribution $\pi(x)$ that contributes significantly to an expectation $\int g(x)\pi(x)dx$ (for some function g). In increasing dimension this set becomes more and more singular, so proposals such as the stretch move described above are unlikely to fall inside of it.

Supporting this claim of poor performance of ensemble methods in high dimensions is the EIT study in [9]. The authors use another ensemble sampler called Differential Evolution MCMC (DE-MCMC) introduced in [56]. In this sampler, to update a walker X_k in the ensemble one picks 2 walkers (without replacement) in the complementary en semble $(X_i, X_j \in \vec{X}_{[k]}(t))$ and proposes $Y = X_k + \gamma(X_i - X_j) + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_N)$. This proposal is symmetric and has tuning parameter $\gamma > 0$. The authors in [9] use this algorithm on the EIT problem and find that it works poorly, which they point out is consistent with the poor performance of the multivariate random walk Metropolis sampler.

1.3.3.4 Tempering methods: Simulated Tempering and Parallel Tempering

If a posterior distribution has well separated modes, then local proposals such as random walk Metropolis Hastings can fail to find all the modes let alone explore them efficiently. Tempering methods are a way to tackle this problem of multimodality.

Simulated Tempering (ST) (as introduced in chapter 10 of [40] and in chapter 11 of [9]) is one type of tempering algorithm. Similarly to the algorithm defined in section (1.3.3.2) above, we augment the state space by introducing an additional discrete parameter $\beta_{temp} \in I_{temp} = \{\beta_1 = 1 < \beta_2 < ... < \beta_L\}$ with $\beta_i \in [0,1]$. Defining the unnormalised target posterior as $\pi(x) \propto \exp\{-k(x)\}$ we can augment the state space in two

ways. In the first case we temper the entire posterior (with $c(\beta_{temp})$ the *pseudo-prior*):

$$\pi_1(x, \beta_{temp}) \propto \exp\{-\beta_{temp} k(x)\} c(\beta_{temp}) \tag{1.48}$$

In the second case (with $-\phi(x)$ the log-likelihood and $-\phi_0(x)$ the log-prior) we temper only the likelihood:

$$\pi_2(x, \beta_{temp}) \propto \exp\{-\beta_{temp} \phi(x) - \phi_0(x)\} c(\beta_{temp}) \tag{1.49}$$

The choice of whether to temper only the likelihood or the entire posterior depends on the sampling problem at hand. Tempering the entire posterior can avoid shrinkage (namely the posterior mean tending towards the prior mean) which can lower the acceptance rate for moves between temperatures. On the other hand, there may be reasons to only wish to temper the likelihood. For example, if one is inferring the dimensionality of the model as well as the parameters in MCMC (as in [47]), one might temper only the likelihood to allow the prior to keep the dimensionality of the model within physically reasonable bounds.

In both cases we have introduced the *pseudo-prior* $c(\beta_{temp})$ in the posterior, named in such a manner because it is unconstrained by the data. These values (one for each inverse-temperature β_{temp}) must be tuned by the practitioner so that the entire chain mixes. Geyer in [9] gives an iterative procedure for tuning these. The intuition for tempering is that by heating up the distribution (β_{temp} is also called the inverse-temperature) the peaks are flattened out and the chain can escape local modes and efficiently mix between them. A related method in optimisation is *Simulated Annealing* (see for example [9] for a brief overview) which can optimise a function with many local minima.

The ST algorithm alternates between "within-temperature" moves and "between-temperature" moves. Let α_0 be the probability of doing a within-temperature move. Given a MCMC chain in state (x, β_i) sampling the posterior π_1 (the algorithm is the same for likelihood tempering), sample $U \sim \mathscr{U}(0,1)$:

- if $\alpha_0 > U$, perform a within-temperature move using any MCMC proposal (it could be a random walk Metropolis proposal or any other proposal)

- if $\alpha_0 \leq U$, propose a new value β_j using a discrete random walk with reflective

boundaries and accept it with probability $\alpha = \min\{1, \frac{\pi_1(x,\beta_j)c(\beta_j)}{\pi_1(x,\beta_i)c(\beta_i)}\}$

When the sampler has converged, the conditional samples $\pi_1(x|\beta_{temp} = 1)$ are samples from the target posterior $\pi(x)$.

An alternative tempering method which avoids the need to hand-tune the pseudo-prior $c(\beta_{temp})$ is *Parallel Tempering* (PT) (also known as Replica Exchange MCMC), which is a type of ensemble method (from section (1.3.3.3) above). Here we augment the state space χ to be the product $\chi \times I_{temp}$ using the same inverse-temperature ladders as in ST ($I_{temp} = \{\beta_1 = 1 < \beta_2 < ... < \beta_L\}$). We therefore run L chains in parallel to target the joint distribution $\Pi_{PT}(x_1,...,x_L) = \pi_1(x_1),...,\pi_L(x_L)$ where $\pi_i(x)$ is the posterior tempered using the inverse-temperature β_i.

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

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

Here there is no need to tune the pseudo-priors as they cancel out in the acceptance ratio. Another benefit is that the chains can be run on parallel cores rather than sequentially (as is the case in ST). However in both PT and ST one needs to choose a temperature schedule I_{temp}. Many extensions of these tempering algorithms have been proposed in the literature, such as ones generalising the between-temperature moves in [55] or ones defining a continuous temperature schedule in [23].

1.3.3.5 Tuning ST and PT

Tuning the temperature schedule is a major difficulty in tuning ST and PT. Atchadé et al. in [1] theoretically investigate the optimal temperature spacings and suggest a practical tuning procedure. They find that too few temperatures results in low acceptance rates for the swaps which results in an inefficient sampler. Similarly, they find that too many temperatures has a high acceptance rate, but also ends up mixing slowly as "good ideas" take too long to propagate from the hot chains to the colder ones. They find under

certain assumptions that the optimal acceptance rate to maximise the expected square jump distance is 23.4%. The tuning procedure they therefore recommend is to find an inverse-temperature low enough for the sampler to mix, and to add another inverse-temperature above it that results in a swap acceptance rate of approximately 23%. One then repeats this procedure until the untempered distribution is reached.

They argue that this result generalises the previous tuning heuristic of geometrically spaced temperatures (namely that consecutive inverse-temperatures follow the relationship $\beta_{i+1} = c\beta_i$ for c a constant to be determined). Firstly, even in cases where the geometric spacing heuristic is appropriate, it gives no indication on how to choose the constant c. Secondly, the authors find cases where the 23% heuristic outperforms the geometric one. For example they tune a PT sampler for the Ising distribution on a $N \times N$ two-dimensional lattice. As the Ising distribution admits a phase transition at the critical temperature, they find that the 23% heuristic forces them to include many inverse-temperature spacings around this critical value. In contrast, a geometrically spaced schedule was found to perform poorly here.

Independently of which heuristic one uses to tune the temperatures, the example of the Ising distribution does point out a limitation in tempering methods: if a distribution admits a phase transition (a qualitative change in the distribution in a narrow range of temperature values or even at a single temperature value) then a large number of inverse-temperatures will be needed for the algorithm to accepts swaps in this region. This is a problem for the performance of MCMC as the time it takes for samples to propagate up the temperature ladder increases exponentially with the number of temperatures.

Another theoretical result in this paper is that under certain assumptions (for example assuming perfectly tuned pseudo-priors) ST is more efficient than PT. However the authors point out that it is unclear what would happen in the case of sub-optimally tuned pseudo-priors (which would often be the case in practice), so that the comparison between these two algorithms is not clear-cut.

We now discuss an applied paper by Sambridge in [47] that aims to introduce PT and ST to the earth sciences and discusses these algorithms both as sampling methods and as optimisation methods. The author investigates the difference in performance between swapping samples between adjacent temperatures only and for random tem-

peratures. They find that for temperatures tuned following the 23% heuristic described above the algorithm performs well, but that adding many more temperatures (in their 1D toy example they use 50 temperatures) and allowing random swaps results in a more efficient sampler. A possible explanation for this could be that with random swaps each chain is coupled with many other chains; therefore a swap move is likely to propose a "novel" region of parameter space (novel compared to previous swap move proposals). In contrast, a chain in the case of adjacent swaps is coupled with either one or two other chains (depending on its location in the temperature ladder). As a result the proposals in a swap move (for a single chain) are likely to be very correlated to each other as a swap proposal corresponds to the position in parameter space of the coupled chain. However the argument for adding many temperatures assumes that increasing them is inexpensive; but as mentioned in section (1.2.3) there is a cost to increase the number of cores in a server. We investigate experimentally in chapter 2 the effect of random swaps as opposed to adjacent swaps in a PT sampler with practical constraints on the number of temperatures.

The author then employs a PT sampler on an inverse problem in the geosciences that exhibits multimodality (using simulated data). They recover the true parameters in the inference and claim that the sampler mixes, but virtually no converge diagnostics were used (or at least reported in the paper). They seem to have treated MCMC more as an optimisation algorithm than as a sampling algorithm (they then explicitly use PT as an optimisation algorithm to show its use for optimisation).

1.3.3.6 MCMC on function space

We now consider the case of sampling from a parameter defined on function space. In practice one will need to discretise this parameter, but the discretisation should converge to the function in the limit of increasing resolution. We first review why a random walk Metropolis Hastings sampler will deteriorate with increasing dimension and then introduce a common method to overcome this limitation.

Consider the problem of sampling from the posterior $\pi(x) \sim L(x)\pi_0(x)$ with $L(x)$ the data likelihood and $\pi_0(x) \sim \mathcal{N}(0,\Sigma)$ the prior, and $x \in \mathscr{R}^N$ is the discretisation of a functional parameter.

If we use the proposal $\tilde{x} = x + \eta$ with $\eta \sim \mathcal{N}(0,\Sigma)$, then the proposal distribution

$q(\tilde{x}|x)$ is Gaussian and we have (by symmetry): $\frac{q(x|\tilde{x})}{q(\tilde{x}|x)} = 1$. We expand the ratio of priors in the acceptance probability:

$$-2\log\frac{\pi_0(\tilde{x})}{\pi_0(x)} = |\Sigma^{-1/2}\tilde{x}|^2 - |\Sigma^{-1/2}x|^2 \tag{1.50}$$

$$= |\Sigma^{-1/2}(x+\eta)|^2 - |\Sigma^{-1/2}x|^2 \tag{1.51}$$

$$= |\Sigma^{-1/2}\eta|^2 + 2 < \Sigma^{-1/2}x, \Sigma^{-1/2}\eta > \tag{1.52}$$

As $\eta \sim \mathcal{N}(0,\Sigma)$ we have $\eta = \Sigma^{1/2}\xi$ with $\xi \sim \mathcal{N}(0,I)$. To see the asymptotic behaviour of this ratio of priors we take the expectation of first term:

$$\mathbb{E}|\Sigma^{-1/2}\eta|^2 = \mathbb{E}|\Sigma^{-1/2}\Sigma^{1/2}\xi|^2 = \mathbb{E}|\xi|^2 = \sum^{N}\mathbb{E}|\xi_k|^2 = \sum^{N}\mathrm{Var}|\xi_k| \tag{1.53}$$

As $\mathrm{Var}|\xi_k = 1|$ for all k, this term grows $\mathcal{O}(N)$ as the grid is refined ($N \to \infty$) ; and so the acceptance probability of the sampler tends to zero. Furthermore, the second term in equation(1.52) has expected value zero.

Consider now a slight modification of the proposal above named the Preconditioned Crank-Nicolson (pCN) algorithm as introduced in [15] (with $\xi \sim \mathcal{N}(0,\Sigma)$ and $\omega \in [0,1]$) :

$$\tilde{x} = (1-\omega^2)^{1/2}x + \omega\xi \tag{1.54}$$

The proposal distribution is now not symmetric and is Gaussian: $q(\tilde{x}|x) \sim \mathcal{N}((1-\omega^2)^{1/2}x, \omega^2\Sigma)$. The acceptance probability is:

$$\alpha = \min\{1, \frac{L(\tilde{x})\pi_0(\tilde{x})q(x|\tilde{x})}{L(x)\pi_0(x)q(\tilde{x}|x)}\}$$

We now define $A(\tilde{x},x) := \pi_0(\tilde{x})q(x|\tilde{x})$ and will show that this term is symmetric.

$$-2\log A(\tilde{x},x) = |\Sigma^{-1/2}\tilde{x}|^2 + \frac{1}{\omega^2}|\Sigma^{-1/2}(x-(1-\omega^2)^{1/2}\tilde{x})|^2 \tag{1.55}$$

$$= |\Sigma^{-1/2}\tilde{x}|^2 + \frac{1}{\omega^2}|\Sigma^{-1/2}x|^2 - 2\frac{(1-\omega^2)^{1/2}}{\omega^2} < \Sigma^{-1/2}x, \Sigma^{-1/2}\tilde{x} > \tag{1.56}$$

$$+ (\frac{1}{\omega^2} - 1)|\Sigma^{-1/2}\tilde{x}|^2 \tag{1.57}$$

$$= \frac{1}{\omega^2}(|\Sigma^{-1/2}\tilde{x}|^2 + |\Sigma^{-1/2}x|^2) - 2\frac{(1-\omega^2)^{1/2}}{\omega^2} < \Sigma^{-1/2}x, \Sigma^{-1/2}\tilde{x} > \tag{1.58}$$

As this term is symmetric, the acceptance probability reduces to (writing $L(x) = \exp\{-\phi(x)\}$):

$$\alpha = \min\{1, \exp\{\phi(x) - \phi(\tilde{x})\}\} \tag{1.59}$$

Therefore the acceptance probability is independent of the dimension of the parameter and only depends on the ratio of likelihoods. As the proposal in (1.54) results from the discretisation of a stochastic partial differential equation, it was designed to work for functional parameters. This follows the general principle advocated by Stuart in [54] that one should avoid discretisation until the last possible moment. Doing so ensures that these methods are principled and will be robust with respect to refining the grid on which the functional parameter is defined.

An obvious limitation with pCN however is that it is simply a random walk proposal. If the likelihood is very informative in a certain subspace and not informative in another, the step size ω has to be lowered so that proposals are accepted (due to the informative likelihood). As a result, proposals in the prior dominated directions are too small and mixing is slow in these directions. There are several extensions in the literature that extend pCN to remedy these limitations. One approach in [36] is to generalise equation (1.54) by considering operators that put weights on different directions of the space of functions.

In the simplest case, consider a Bayesian inverse problem on a Hilbert space H with orthonormal basis $\{\phi_j\}_{j=1}^{\infty}$. Let P project onto a finite basis $\{\phi_j\}_{j=1}^{J}$ and $Q =$

$I-P$ project on the complementary subspace. If observations only inform the subspace projected on by P, then the complement Q is dominated by the prior. One can therefore modify the proposal in equation (1.54) to sample using pCN on PH and to sample independently from the prior on QH:

$$\tilde{x} = (1-\omega^2)^{1/2} Px + (\omega P + Q)\xi \tag{1.60}$$

More sophisticated methods in [36] as well as [4] extend finite dimensional methods to function space by - for example - using derivatives of the likelihood to build proposals.

1.4 Conclusion

We have introduced the field of traffic flow modelling which attempts to model traffic flow on motorways using systems of ODEs or PDEs. Although the field has grown considerably and now includes sophisticated mathematical models, fitting these models rigorously to data remains difficult partly due to the presence of functional parameters such as the initial or boundary conditions. We consider the Bayesian framework to fit and compare models and to quantify the uncertainty in these inferences. We review MCMC methods which are a powerful tool for nontrivial models but require expert knowledge to tune and diagnose.

Chapter 2

Inverse problem for the advection equation

2.1 The advection equation

We start by considering a simple hyperbolic PDE: the advection equation. This models the transport of a *tracer* - a substance with a very small concentration that does not affect the dynamics of the fluid - along a 1D pipe. Setting ρ, and u to be the density and wave speed of the fluid we write the PDE as follows:

$$\rho_t + u\rho_x = 0 \tag{2.1}$$

So the flux function of this equation is linear and is written as $q(\rho) = u\rho$. It is easy to check that the solution to (2.1) can be written as $\rho(x,t) = \rho_0(x-ut)$ with $\rho_0(x)$ the initial condition. This PDE has the property that the solution is constant along rays of the form $X(t) = x_0 + ut$ called *characteristics* (introduced in section (1.1.2.5)). Indeed we check that the solution along these rays does not change with time:

$$\frac{d\rho(X(t),t)}{dt} = \rho(X(t),t)_t + \rho(X(t),t)_x \frac{dx}{dt} \tag{2.2}$$

$$= \rho_t + u\rho_x \tag{2.3}$$

$$= 0 \tag{2.4}$$

To determine the solution at any given point (x,t) along a pipe with endpoints a

and b we must use the initial condition $\rho_0(x)$ as well as the boundary conditions $\rho_{BC}(t)$. Assuming the wave speed to be positive we use the inlet boundary condition at $x = a$. The solution is:

$$\rho(x,t) = \begin{cases} \rho_{BC}(t - \frac{x-a}{u}) & \text{if} \quad a < x < a + ut \\ \rho_0(x - ut) & \text{if} \quad a + ut < x < b \end{cases} \tag{2.5}$$

If the wave speed is negative then the outlet boundary condition (at $x = b$) must be used. From now on we will only consider parameter ranges where the initial condition is needed, as the extension to include the boundary conditions is straightforwards.

Suppose we observe values of flow q at some points in the $x-t$ plane, and consider the problem of reconstructing the initial condition and wave speed. As this problem is ill-posed, there may be no unique solution; we will therefore employ the Bayesian framework as described in section (1.3.1).

2.2 The inverse problem

The ill-posedness of the inverse problem for the advection equation can be seen in several ways. Firstly, the initial condition $\rho_0(x)$ is a function but we observe a noisy solution only at a few detector values. So for a fixed value of u the observed flow will only constrain the initial condition at a few points while the rest of the initial condition can vary freely.

Secondly, consider the case of observing a value of flow at a single detector at location x_1 and time t_1. Let (u, ρ_0) and $(\tilde{u}, \tilde{\rho}_0)$ be two sets of parameters such that (with $\alpha \in \mathscr{R}^+$):

$$\begin{cases} \tilde{u} & = \alpha u \\ \tilde{\rho}_0(x) & = \frac{\rho_0(x + (\tilde{u} - u)t_1)}{\alpha} \end{cases} \tag{2.6}$$

Then we have $\tilde{u}\tilde{\rho}_0(x_1 - \tilde{u}t_1) = u\rho_0(x_1 - ut_1)$ and so the value of flow at the detector for the two sets of parameters is the same. Varying the parameter α will not change the value of flow at the detector, so no amount of data will be able to constrain the solution. Note that if the detector measures the absolute value of flow, then a negative value of α will also result in the same measurement. This problem remains if we have several

detectors measuring flow at exactly the same time t_1.

The two problems described above mean that for some cases the problem is *non-identifiable*, namely that there exists two distinct sets of parameters that result in the same sampling distribution of data.

To overcome this we set a Gaussian prior on both parameters u and ρ_0 which act to regularise the problem:

- $\pi_o(u) \sim \exp\{-\frac{1}{2\sigma_0^2}(u-m_0)^2\}$
- $\pi_{\rho_0}(\rho_0) \sim \mathcal{N}(m_{\rho_0}, C_0)$

With m_0 and m_{ρ_0} the prior means and σ_0^2 and C_0 the variance and covariance function of the parameters. The observation operator $\mathcal{G}$ solves the advection equation for each point in the $x-t$ plane and maps density to flow using the current value of u ($q(x,t) = u\rho(x,t)$). We then build the likelihood using a Gaussian error model with covariance matrix Γ. Multiplying the likelihood by the priors we write the unnormalised posterior as:

$$\pi(u,\rho_0) \propto \exp\{-\frac{1}{2}||\mathcal{G}(u,\rho_0)-q||_\Gamma^2 - \frac{1}{2\sigma_0^2}(u-m_0)^2 - \frac{1}{2}||\rho_0 - m_{\rho_0}||_{C_0}^2\} \tag{2.7}$$

If we consider the conditional distribution $\pi(\rho_0|u)$ the observation operator can be written as a linear operator $\mathcal{G}(\rho_0) = A\rho_0$. Given N observation (x_i,t_i) ($i \in \{1,...,N\}$), one constructs the matrix A by first mapping the observed flow at points (x_i,t_i) back to $t=0$ thus obtaining $x_i^0 := x_i - ut_i$. A then picks out the initial condition times that correspond to these locations (namely x_i^0) and multiplies the density by u to obtain flow. As mentioned in section (1.3.1), we can apply theorem 3.1 in [54] to obtain the conditional posterior analytically, ie: $\pi(\rho_0|u) \sim \mathcal{N}(m,C)$ with:

$$m = m_{\rho_0} + C_0A^T(\Gamma + AC_0A^T)^{-1}(q - Am_{\rho_0}) \tag{2.8}$$

$$C = C_0 - C_0A^T(\Gamma + AC_0A^T)^{-1}AC_0 \tag{2.9}$$

2.3 MCMC

As the posterior defined above cannot be sampled from directly, we will need to use MCMC. We will use a Metropolis Hastings sampler with a proposal that updates both u and ρ_0 simultaneously using a Gaussian proposal for u and the pCN algorithm defined in section (1.3.3.6) for ρ_0. The proposal can be written as:

$$\begin{cases} \tilde{u} = u^{(t)} + \xi \quad \text{with } \xi \sim \mathcal{N}(0, \sigma_u^2) \\ \tilde{\rho_0} = (1-\omega^2)^{1/2}(\rho_0^{(t)} - m_{\rho_0}) + \omega\xi + m_{\rho_0} \quad \text{with } \xi \sim \mathcal{N}(0, C_0) \end{cases} \tag{2.10}$$

With σ_u^2 the variance of the proposal for u and ω the step size for ρ_0.

2.3.1 Results on simulated data

We define a spatial grid $x \in [0,10]$ with 200 cells. We set a uniform prior on the wave speed $u \sim \mathcal{U}(0,1.4)$, and choose 6 detectors at random locations and times such that density at those points in the $x-t$ plane will only be influenced by the initial condition and not the boundary conditions. The location of the detectors in space varies between $[2,10]$ and in time between $[0,2]$.

We define a Gaussian prior on the initial condition with squared exponential covariance function: $\text{cov}(x,y) = \exp\{\frac{1}{2l}(x-y)^2\}$ with $l = 1$ the *characteristic length scale* (see [52]). We set a constant prior mean $m_{\rho_0} = 100$, and show samples in figure (2.1)

We choose a sample from the prior as true initial condition and choose wave speed $u_{true} = 0.5$ to generate data. Independent Gaussian noise is added to the measured flow at each detector with standard deviation 1.

We denote C_0 to be the squared exponential covariance function, and q_i and $\hat{q}_i$ the measured and predicted flow respectively at the i^{th} detector. The posterior can therefore be written as:

$$\pi(u,\rho_0) \propto \exp\{-\frac{1}{2}\sum_{i=1}^{6}(\hat{q}_i - q_i)^2 - \frac{1}{2}|C_0^{-1/2}\rho_0|^2\} \quad \text{if } \ 0 < u < 1.4 \tag{2.11}$$

To sample from this posterior, we run 3 chains for 200K iterations each, with $\sigma_u^2 = 0.0005$ and $\omega = 0.3$. We obtain an acceptance rate of 5.7% for each chain which

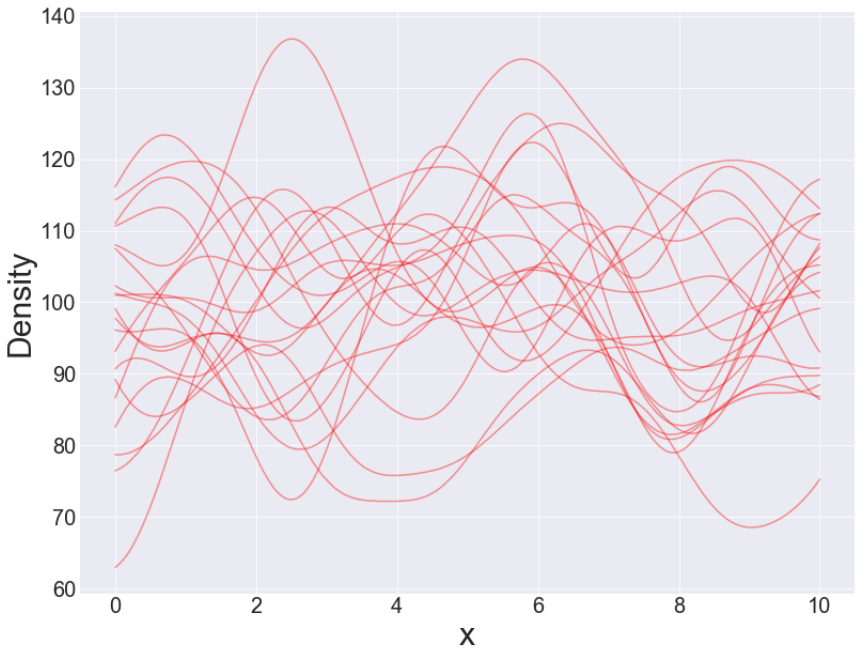

Figure 2.1: Samples from the prior for the initial condition ρ_0

was found to be appropriate following some pilot runs. We show the trace plot for u and a typical time point of ρ_0 thinned by 100 in figure (2.2) which shows good mixing. The 3 colours represent the 3 chains. We show samples from the initial conditions along with the true initial condition in figure (2.3). The ACF for u along with a typical time point for ρ_0 is shown in figure (2.4); we visually estimate the decay time τ_d to be around 500, which we will use as a comparison to later runs.

To analyse the correlations between the two parameters, we plot $\rho_0|u$ samples conditioned on several values of u in figure (2.5). Plotted in red is the true initial condition which approximately aligns with the samples conditioned on $u = 0.5$ (namely, the true value for u). We can clearly see the mean of the conditional posterior being lowered and shifted towards the left as the wave speed increases, which is consistent with the identifiability discussion in section (2.2). To confirm this correlation we plot the mean of these conditional initial condition samples (from the posterior this time) against their corresponding value of u in figure (2.6). Here we see a clear linear correlation: the mean of the initial condition is lowered as the wave speed u increases.

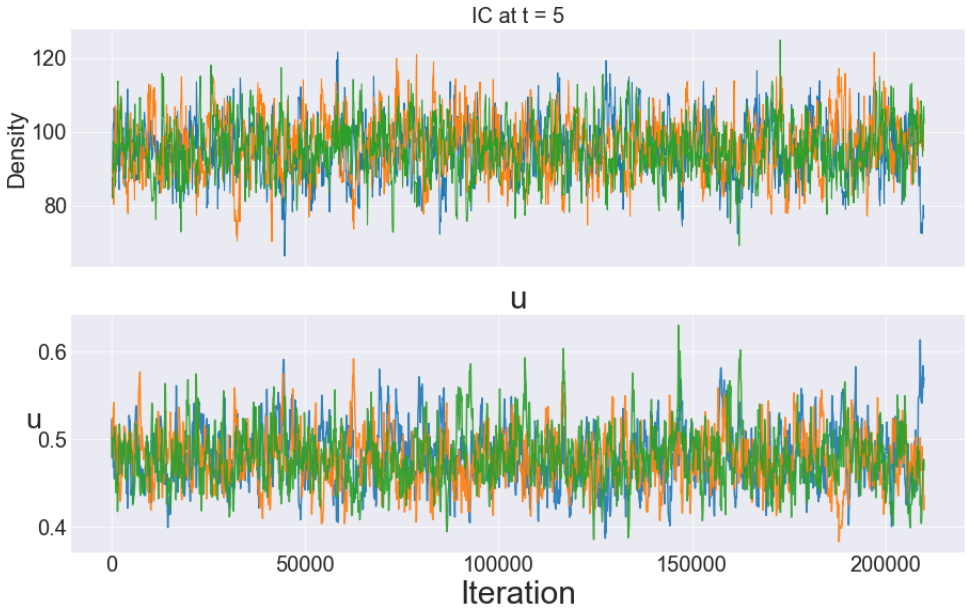

Figure 2.2: Trace plots for u and a typical time point of ρ_0 for a Metropolis Hastings sampler for the advection equation. The 3 colours represent the 3 chains

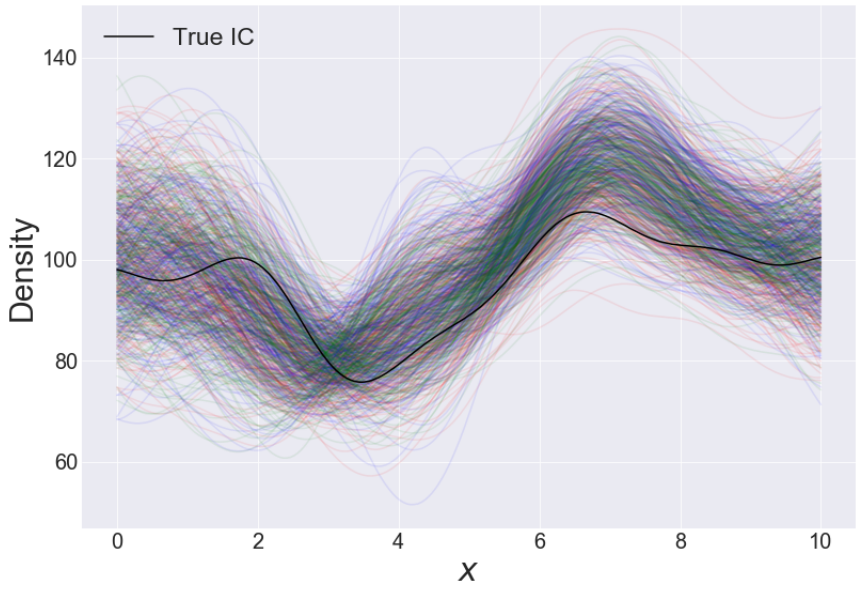

Figure 2.3: Samples for the initial condition ρ_0 for a Metropolis Hastings sampler for the advection equation plotted with the true solution. The 3 colours represent the 3 chains

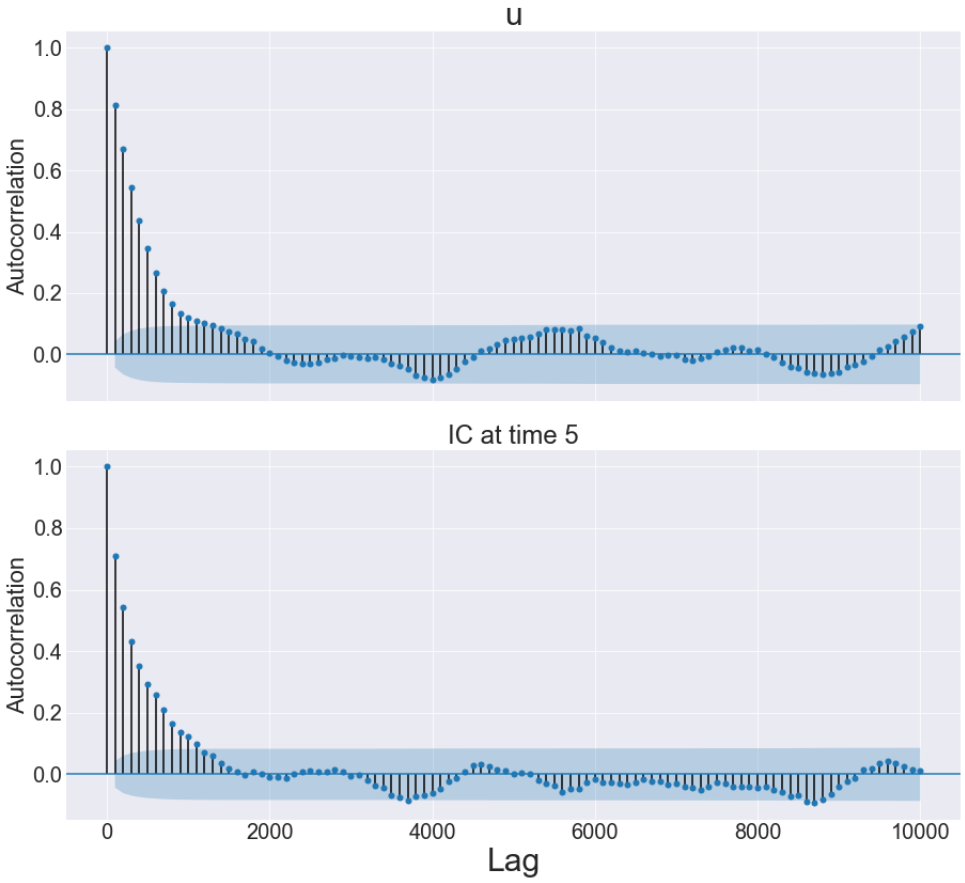

Figure 2.4: Autocorrelation plot with 95% confidence interval for u and a typical time point of ρ_0 for a Metropolis Hastings sampler for the advection equation. We estimate visually the decay time τ_d to be around 500.

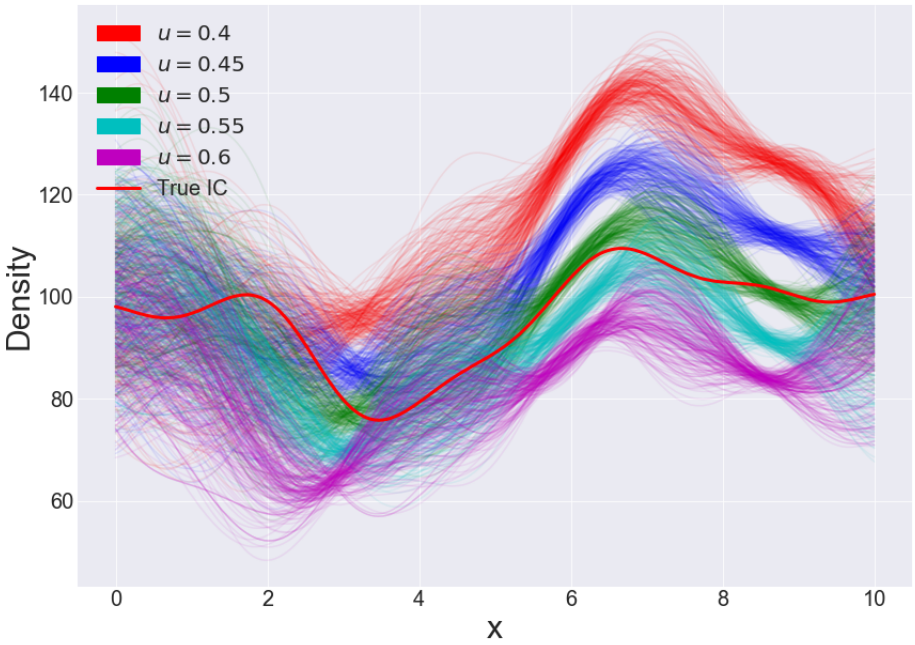

Figure 2.5: $\rho_0|u$ samples conditioned on several values of u for the advection equation. Plotted in red is the true initial condition which approximately aligns with the samples conditioned on $u = 0.5$ (namely, the true value for u). We notice a shift in the initial condition both in the vertical direction as well as in the horizontal direction which is consistent with the identifiabilty discussion in section (2.2)

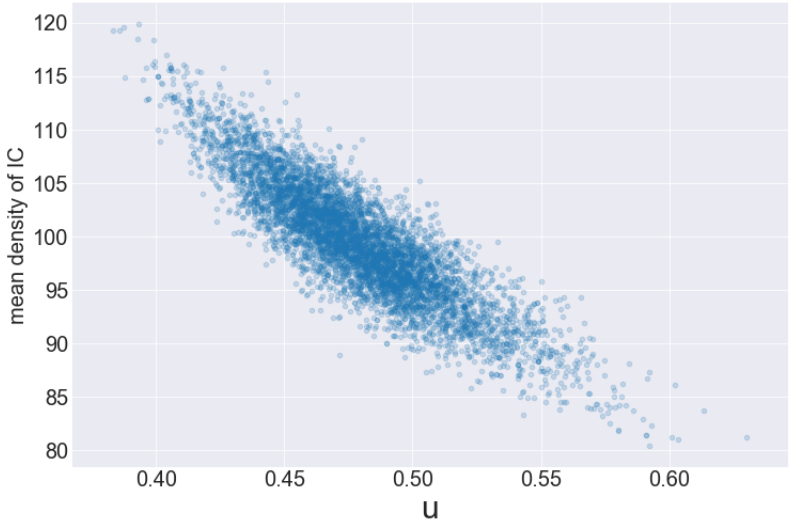

Figure 2.6: Mean of initial condition samples $\rho_0|u$ plotted against their corresponding u samples. We notice a clear linear correlation between these which is consistent with the vertical shift observed in figure (2.5)

2.3.2 The advection equation as a case study for LWR

As discussed in section (1.2.3), when sampling from a difficult posterior it can be helpful to first consider a simpler model. The advection equation is indeed a simpler version of LWR (defined in section (1.1.2.4)): it is a conservation law as in equation (1.15) closed using the relationship $q(\rho) = V_e(\rho)\rho = u\rho$ with u a constant. To sample from the posterior arising from LWR (in chapter 5) we will have to use a Metropolis within Gibbs update for the FD parameters and the boundary conditions (BCs) as updating all the parameters at once results in very slow mixing (similarly to the EIT study discussed in section (1.3.3.1)). As the forward model for the advection equation is computationally inexpensive (it is available analytically), will treat sampling from the advection equation in the same way to be able to easily experiment with the algorithm. Furthermore, the initial condition and boundary conditions play a similar role in these two models, as can be seen from the solution to the advection equation in equation (2.5). We will therefore use the wave speed u and the initial condition ρ_0 as proxies of the FD parameters and BCs in LWR respectively.

We sample from the posterior of the advection equation (arising from the same simulated data as described in the previous section) using a random scan Metropolis within Gibbs sampler for u and ρ_0. We use $\sigma_0^2 = 0.0005$ and $\omega = 0.3$ as before, and update u and ρ_0 with equal probability. We run the sampler for 200K iterations and obtain an acceptance rate of 21.8%, 21.2%, and 21.5% for the u parameter (for the 3 chain respectively) and 11.92%, 12.20%, and 12.06% for the ρ_0 parameter. From figure (2.8) we estimate the decay time for u to be $\tau_d = 1000$ which is more than in the case of the joint proposal in the previous section. As we have seen that the two parameters are highly correlated, this difference is however to be expected. We show in figure (2.7) the trace plot for u and the same time point for ρ_0 as in the previous section.

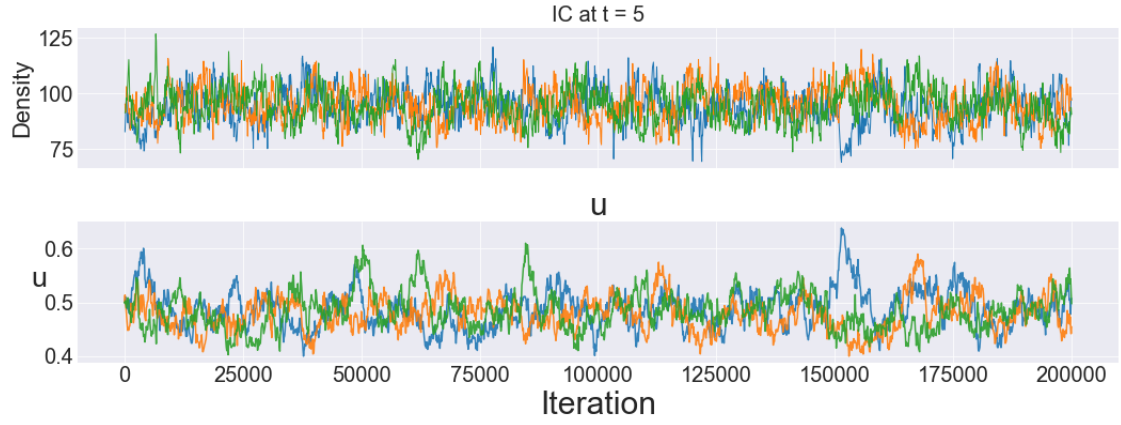

Figure 2.7: Trace plots for u and a typical time point of ρ_0 for a Metropolis with Gibbs sampler for the advection equation. The mixing is noticeably slower than for the sampler with joint proposal in figure (2.2), in particular for the u parameter

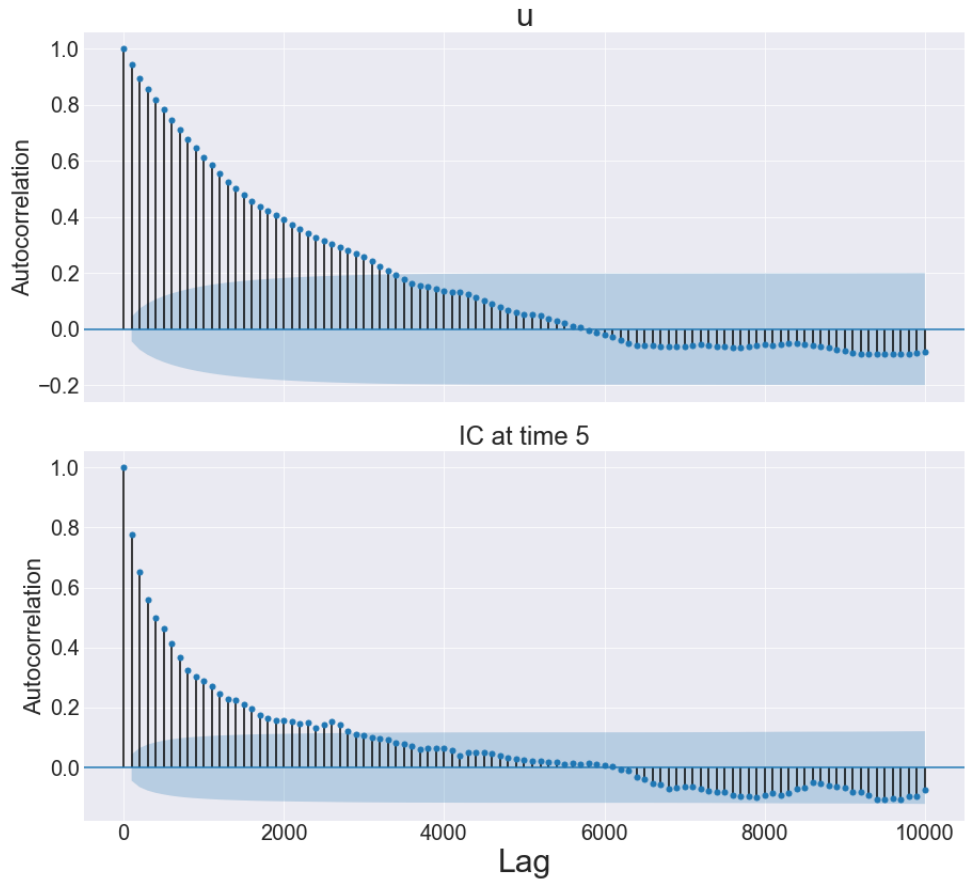

Figure 2.8: Autocorrelation function for u and a typical time point of ρ_0 for a Metropolis with Gibbs sampler for the advection equation. Here the decay time for u is approximately $\tau_d = 1000$, which shows longer autocorrelations than in figure (2.4)

To improve this slow mixing we will now employ a Parallel Tempering (PT) sam-

pler as introduced in section (1.3.3.4). We choose PT rather than Simulated Tempering for the convenience of not having to tune the pseudo-priors. We temper only the likelihood as opposed to the whole posterior as tempering the whole posterior would cause the priors on ρ_0 between temperatures to be incompatible. An intuition for this can be seen by writing the tempered log prior for ρ_0 (with mean 0 for convenience): $\log(\pi_{\rho_0}(\rho_0)) \propto -\frac{1}{2}\beta_{temp}\rho_0^T C_0^{-1}\rho_0$. The inverse-temperature β_{temp} will modify to prior covariance matrix, so samples that get accepted for a given value of β_{temp} will be rejected by another.

We use 4 inverse-temperatures (`[0.027, 0.09, 0.3, 1]`) tuned using the iterative procedure given in [1]: we found a low inverse-temperature that mixes sufficiently quickly and successively added higher inverse-temperatures such that the acceptance rates for adjacent between-temperature swaps was approximately 23%. We could have used less inverse-temperatures, but we chose 4 as this is the number used in the PT sampler for LWR in Chapter 5. We alternate between N_{within} within-temperature moves (for all temperatures in parallel) and a single between-temperature move using a deterministic scan sampler going through each pair of temperatures in turn. So the sampler will update N_{within} within-temperature moves, then update the temperature pairs (β_1, β_2), then update N_{within} within-temperature moves, and then update (β_2, β_3) and so on.

For the within-temperature moves we use the same random scan Metropolis within Gibbs sampler as before (with equal probability for u and ρ_0) but we tune the proposal variance and step size for each temperature. We give these settings in the following table:

Table 2.1: Proposal variances and step sizes for u and ρ_0 for the 4 temperatures in the PT sampler for the advection equation

β_{temp}	σ_u^2	ω
1	0.0005	0.3
0.3	0.002	0.45
0.09	0.008	0.7
0.027	0.02	0.93

However two practical questions remain: should we use random swaps between temperatures as suggested in [47], and how many within-temperature moves should we do between each between-temperature proposal ? We address these two questions empirically.

We start by considering the number of within-temperature moves needed in the sampler. On one side, a swap move can propose very large jumps so one should propose these often. On the other side if two coupled chains swap parameters too often, neither will be able to efficiently explore the state space locally using within-temperature moves. This is expected to be especially a problem for posteriors with highly correlated parameters as in the case of the advection equation. To test this we run several PT samplers with a different number of within-temperature moves (with adjacent swaps and with all other tuning parameters as described above). For each number of within-temperature moves we run a PT sampler for approximately 200K iterations and calculate the mean square euclidean jump distance (MSEJD) of the u parameter ($S^2_{Euc}(u) := \frac{1}{N-1}\sum_{i=1}^{N-1}(u_{i+1}-u_i)^2$) for the untempered chain (the posterior of interest). However it is possible with a PT sampler to have a high MSEJD but a low mixing speed. Indeed consider a PT sampler with 2 temperatures that only proposes swaps (so no within-temperature moves) and with both coupled chains starting at equally likely (but separate) parameters. This sampler can have 100% acceptance rate and a very high MSEJD (especially if the parameters are far apart) but not mix at all. To control for this effect we therefore also estimate the decay time $\hat{\tau}_d$ for the u parameter of the untempered chain. We plot these two quantities as a function of N_{within} in figure (2.9). We notice that the MSEJD mostly decreases as the number of within-temperature moves increases, but that the decay time seems to first decrease and then increase, suggesting an optimal value around $N_{within} = 5$. This gives evidence that too low values of N_{within} can artificially increase the MSEJD by frequently swapping samples between coupled chains. We note that while in principle we would like an uncertainty estimate for $\hat{\tau}_d$, we do not provide one as we only use it as a rough estimate. From this we conclude that $N_{within} \in [1, 10]$ might be an appropriate amount for a PT sampler for the advection equation in this chapter and for LWR in later chapters.

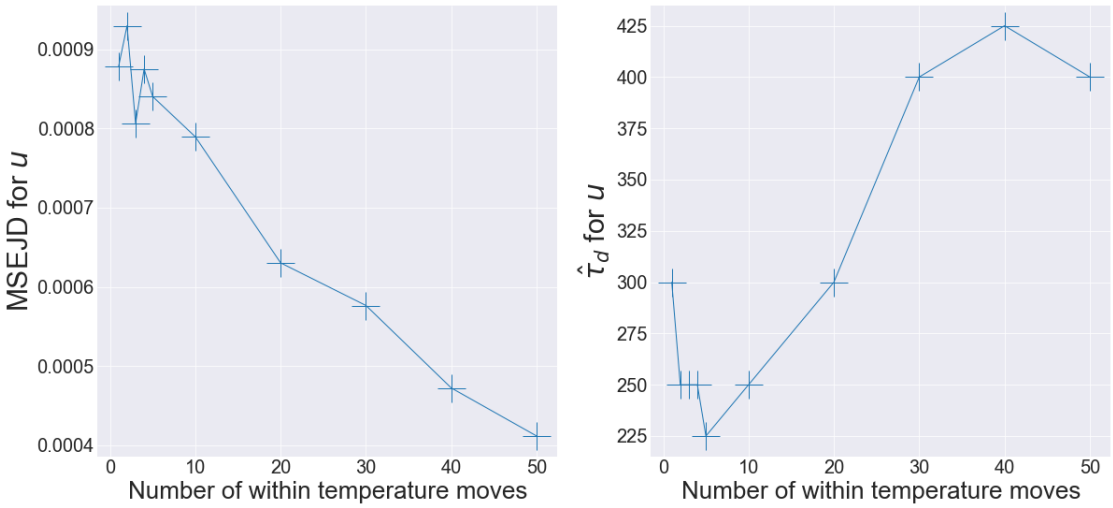

Figure 2.9: MSEJD and $\hat{\tau}_d$ for u for the untempered chain in a PT sampler for different numbers of within-temperature moves. The MSEJD mostly decreases as the number of within-temperature moves increases, but the decay time seems to decrease then increase again, suggesting an optimal value around $N_{within} = 5$. This also gives evidence that too low values of N_{within} artificially increase the MSEJD by frequently swapping samples between coupled chains.

We now investigate the second question which is whether random swaps or adjacent swaps are optimal. We employ a PT sampler for the advection equation with $N_{within} = 2$ and the same settings as above and compare the MSEJD for u in the untempered chain for random swaps and adjacent swaps. We run 100K samples for each case, and report the results in table (2.2) below. We also report the acceptance rates between inverse-temperatures pairs for both runs in table (2.3) below.

Table 2.2: MSEJD for the u parameter in the untempered chain for random temperature swaps and adjacent temperature swaps. We notice that the MSEJD is higher for every temperature in the case of adjacent swaps

β_{temp}	$S^2_{Euc}(u)$ for random swaps	$S^2_{Euc}(u)$ for adjacent swaps
1	0.000027	0.000048
0.3	0.000057	0.000110
0.09	0.000091	0.000162
0.027	0.000145	0.000182

Table 2.3: Acceptance rates for pairs of inverse-temperatures for random temperature swaps and adjacent temperature swaps. We define $(\beta_1, \beta_2, \beta_3, \beta_4) := (1, 0.3, 0.09, 0.027)$

Pairs	Random swaps	Adjacent swaps
β_1, β_2	21.5%	21.0%
β_2, β_3	22.6%	23.2%
β_3, β_4	27.9%	28.2%
β_1, β_3	1.9%	/
β_1, β_4	0.3%	/
β_2, β_4	3.0%	/

We observe in table (2.2) that the MSEJD is higher for every temperature in the case of the adjacent swaps. This is not necessarily inconsistent with the findings in [47] as they seemed to have used many more temperatures than suggested by the 23% heuristic in [1]. However as we pointed out in section (1.3.3.5), increasing the number of cores has a cost, so given a limit of 4 temperatures it seems that adjacent swaps is more efficient (at least in the case of the advection equation). We also note that the acceptance rate decreases for widely spaced temperatures pairs (as expected). The way in which they decrease is also to be expected: jumps between 2 temperature spacings are of the order of magnitude of $0.23^2 \approx 0.05$, and the pair (β_1, β_4) has an acceptance rate (approximately) of the order of magnitude of $0.23^3 \approx 0.01$. Based on this test we will now only use adjacent temperatures for the remainder of this thesis.

We now finally employ a PT sampler with the same settings as in the previous runs but with $N_{within} = 2$ and adjacent swaps. We run 3 chain for 200K iterations each and show in figure (2.10) the traceplots for u and a typical time for ρ_0. We also show the ACF of u and the same time point for ρ_0 in figure (2.11). We estimate the delay time to be $\hat{\tau}_d \approx 500$ which shows faster mixing than the the Metropolis with Gibbs sampler without Parallel tempering and similar to the Metropolis Hastings sampler originally developed in this chapter. We therefore conclude that a PT sampler can help with sampling highly correlated parameters such as u and ρ_0, and we will apply these lessons to the problem of sampling from LWR using a PT sampler in chapters 4 and 5.

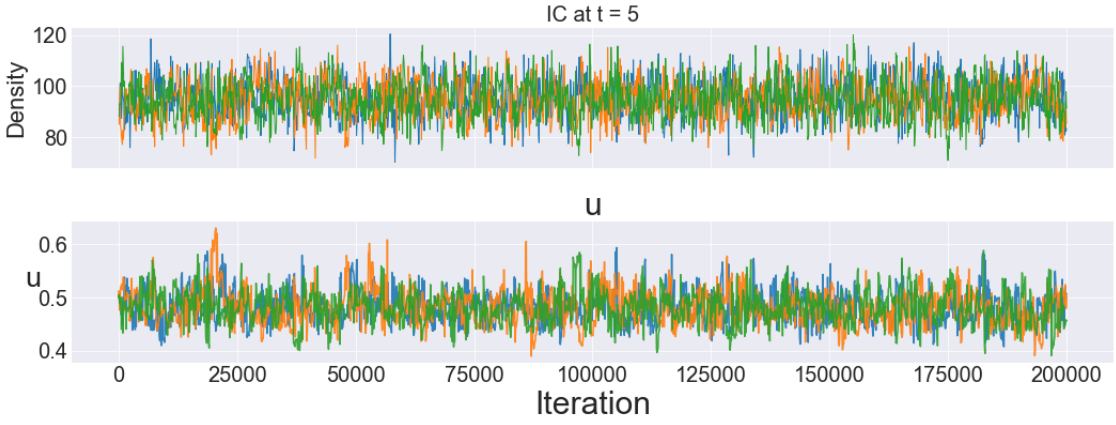

Figure 2.10: Trace plots for u and a typical time point of ρ_0 for a PT sampler for the advection equation with $N_{within} = 2$ and adjacent swaps

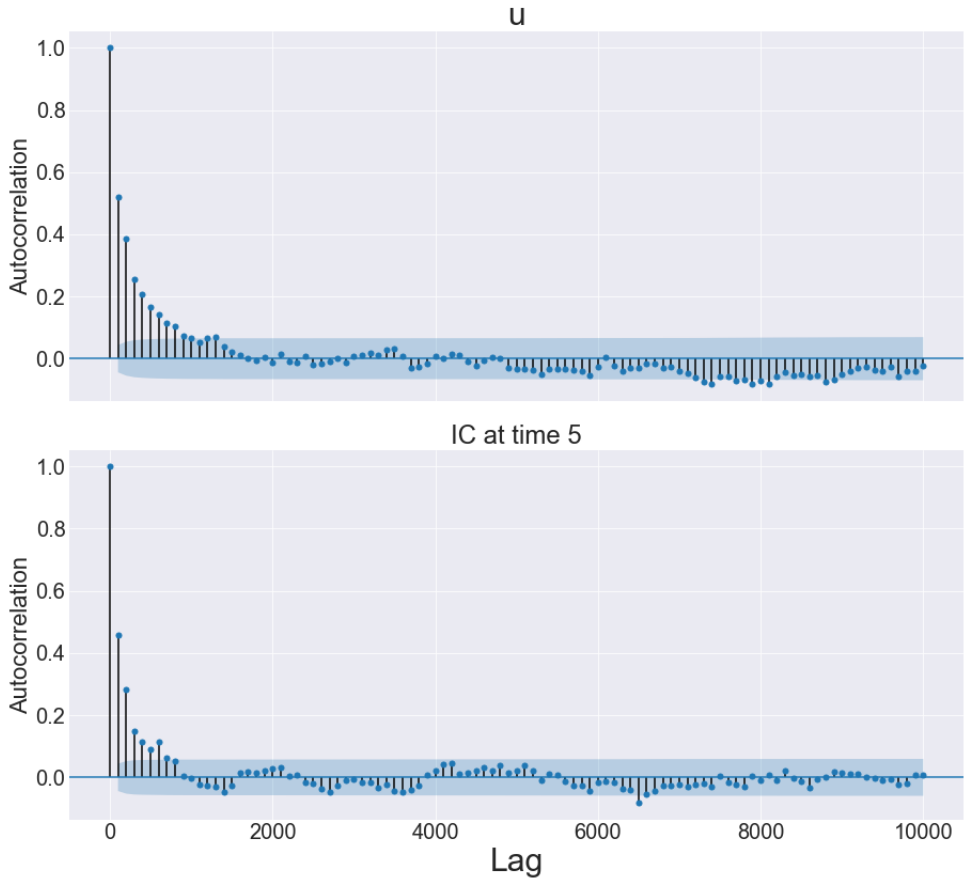

Figure 2.11: Autocorrelation function for u and a typical time point of ρ_0 for a PT sampler for the advection equation with $N_{within} = 2$ and adjacent swaps. We notice that the delay time is approximately 500 which shows faster mixing than the Metropolis with Gibbs sampler without Parallel tempering.

2.4 Conclusions

In this chapter we consider the advection equation which is a simplified version of LWR and inferred the wave speed u and initial conditions ρ_0 using a RWMH and pCN algorithm. As these two parameters are analogous to the FD parameters and boundary conditions in LWR, we use this computationally inexpensive model to tune the Parallel Tempering (PT) algorithm: we considered different numbers of within temperature moves and random vs adjacent swaps. We find that between 1 and 10 within temperature moves was optimal (both for MSEJD and autocorrelations) and that adjacent swaps was best given our computational constraints. As u and ρ_0 are very correlated, a Metropolis within Gibbs sampler mixes very slowly. A PT algorithm is found to significantly help increase mixing speed.

Chapter 3

Estimating FD parameters

Throughout the rest of the thesis we will build the tools to estimate parameters in the LWR model, and test the methodology on motorway traffic data. We start in this chapter by describing the data and the forward model before estimating parameters for two fundamental diagrams (FD) with fixed boundary conditions (BCs). We do this firstly using simulated data, then using the M25 data. In later chapters we will estimate the boundary conditions as well as the FD.

3.1 Data

We use MIDAS data from the Highways Agency on the M25 in 2007. The data is measured on loop detectors spaced every 500m on the road which takes 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, so therefore corresponds to flow (number of vehicles per unit time). Occupancy 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). 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). We plot $\frac{q}{v}$ vs occupancy for lane 3 on the 8th January 2007 between 6:21am and 7:09am

on the chosen section of the M25 (how the road is chosen is discussed later). Note that lane 4 is the fastest lane, and lane 1 is the slowest.

We see a consistent relationship between the two variables under 40% occupancy, but beyond that there is a scatter.

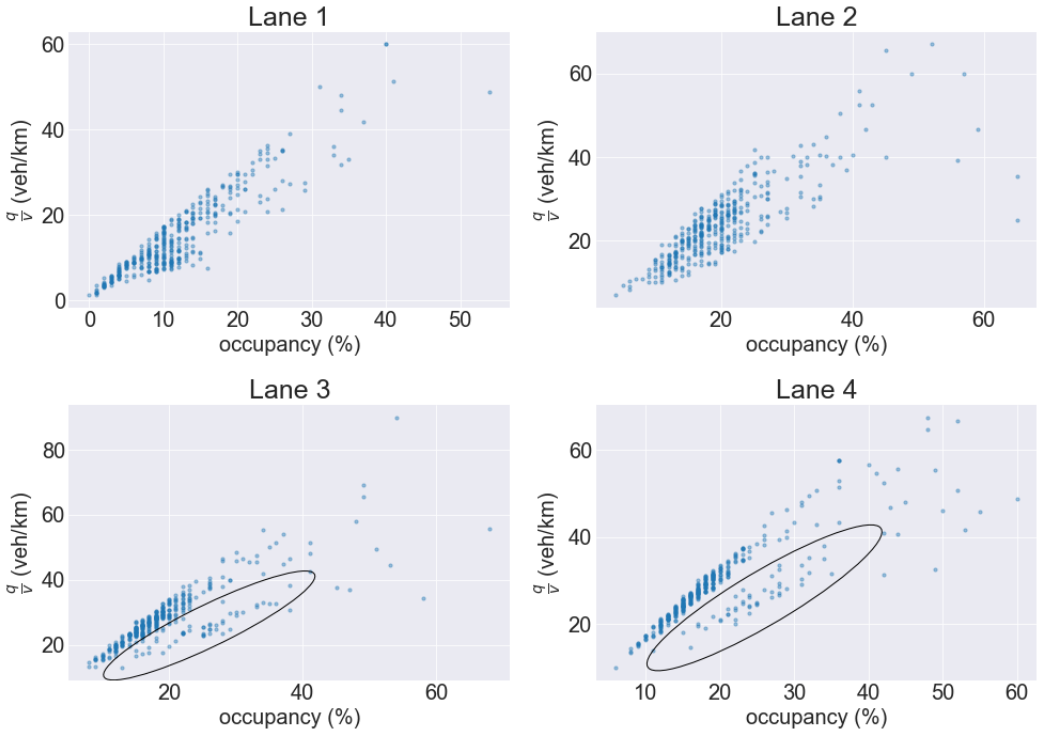

Figure 3.1: Section of M25 on the 8th January 2007 between 6:21am and 7:09am. Density from speed vs occupancy for all four lanes. Apart from the extra 'branch' (circled in black on the plot) which is due to a slightly different flow-density relationship, we observe a linear correlation up until around 30% occupancy where there is scatter

We have the following relationship between occupancy and density ([29]) (with L the average vehicle length, ω_{occ} occupancy):

$$\rho = \frac{\omega_{occ}}{L} \tag{3.1}$$

Using equation (3.1) and the relation $q = v\rho$ we obtain:

$$\frac{q}{v} = \frac{\omega_{occ}}{L} \tag{3.2}$$

Using equation (3.1) and (3.2) we obtain two ways to estimate density: density from occupancy ($\rho_{occ} := \frac{\omega_{occ}}{L}$) and density from speed ($\rho_{speed} := \frac{q}{v}$).

Equation (3.2) means that the slope of figure (3.1) corresponds to the reciprocal of the average vehicle length on that lane. The shows that the vehicle length is fairly uniform up to a certain occupancy, but after that point there is variability due to the variety of vehicle lengths. This variability occurs for high occupancy (which corresponds to high density; ie: congestion). This means that density estimated from speed will have a wider error above this critical occupancy.

There are several ways to estimate the average vehicle length L (or its reciprocal L^{-1}): in [29] 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: 4m vehicles, type 2: 6m, type 3: 9m, type 4: 16m). 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 includes 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 vs density using these two methods in figure (3.2) 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 chapter 5 we will estimate density in the BCs (as well as estimate the FD parameters) and compare its wave speeds to those in 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 macroscopic 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 5km section of the M25, and as the detectors are spaced every 500m there are 12 detectors (the endpoints of the section are included). However some of the detectors have faults, so we use 8 detectors in our inference at the following locations (in

Figure 3.2: Section of M25 on the 8th January 2007 between 6am and 10am. We plot flow vs density for two estimation methods: density from occupancy and density from speed 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.

km): $[0, 1, 2, 2.5, 3, 4, 4.5, 5]$. We use the detector measurements between 6:21am and 7:09pm (including both endpoints) on the 8th January 2007; we therefore use 48min of data which corresponds to 49 time points.

As the two methods to estimate density give very different results, we would like to build our likelihood based on a quantity that has less 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 inter-arrival 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. To assess the fit of this model (and the fit of LWR) we will perform a residual analysis once the parameters in LWR are fit to data in chapter 5. We can see the sum of flows and density in space and time in figures (3.3) and (3.4). In these plots time is denoted on the horizontal axis and space (distance on the road) on the vertical axis. Vehicles move from distance 0km to 5km (namely upwards in the plot) and forwards in time (namely to the right in the plot) so therefore move diagonally (upwards 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 5km section of road. This is consistent with a vehicle speed of approximately 120km/hr (ie 2km/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 (low flow) waves. These backwards 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.

3.2 LWR

We would like to solve LWR for arbitrary FD parameters, initial conditions, and boundary conditions. However we cannot solve it analytically in the general case. We will use the open-source software Clawpack ([13]) which is a package for solving conservation laws using finite volume methods (see [42] 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 [37].

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 (3.3)), which ensures that the method behaves correctly in the presence of shock waves:

$$\rho_t + f(\rho)_x = 0 \tag{3.3}$$

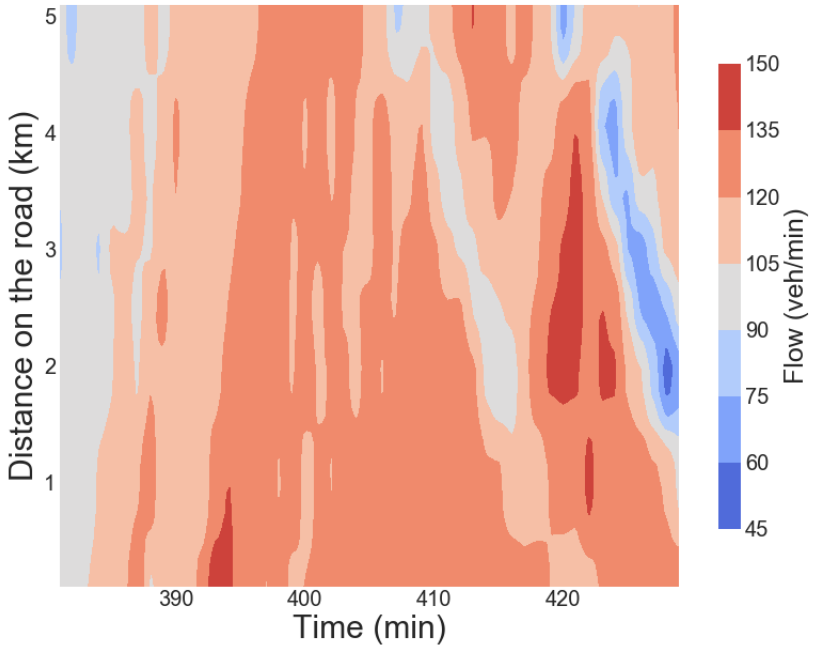

Figure 3.3: Flow for the section of M25 on the 8th January 2007 between 6:21am and 7:09am. We observe forward moving free flow waves between minutes 381 and 405 which correspond to the movement of vehicles. We also observe backwards moving low flow waves in the second half of the X-T plane.

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

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{\Delta x}(F_{i+1/2}^n - F_{i-1/2}^n) \tag{3.4}$$

With:

- ρ_i^n density at cell i, time n
- $F_{i+1/2}^n$ flux (flow) at the the right boundary of cell i at time n
- $F_{i-1/2}^n$ flux (flow) at the the left boundary of cell i at time n
- Δt and Δx the time and space discretisation

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 computed analytically - at each boundary to find $F_{i+1/2}^n$ and $F_{i-1/2}^n$.

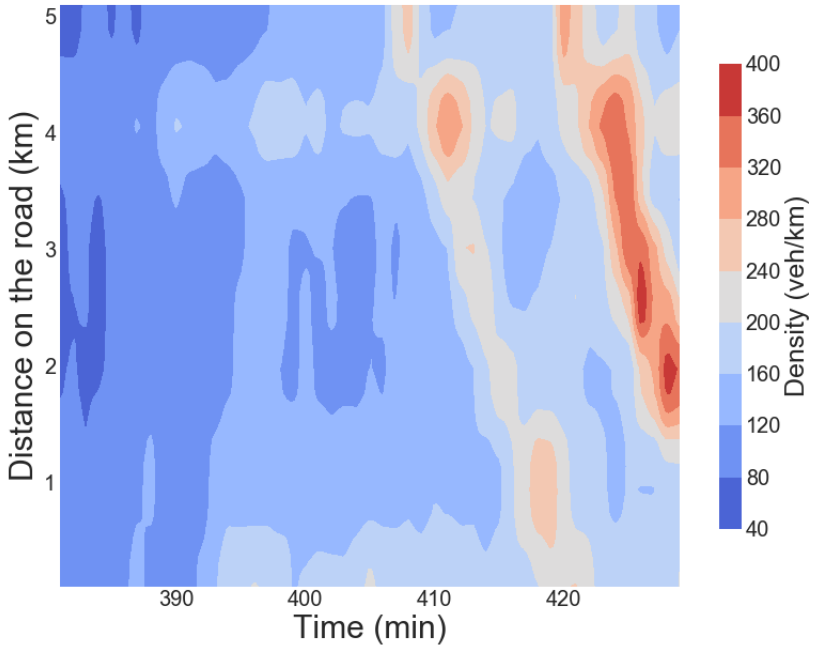

Figure 3.4: Density estimated from occupancy for the section of M25 on the 8th January 2007 between 6:21am and 7:09am. We observe forward moving free flow waves between minutes 381 and 405 which correspond to the movement of vehicles. We also observe backwards moving high density waves in the second half of the X-T plane.

There are two main cases in the Riemann Problem:

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

In this case we will have a shock wave as seen in figure (3.6): the discontinuity will simply be advected with speed: $v_{shock} = \frac{f(\rho_i) - f(\rho_{i-1})}{\rho_i - \rho_{i-1}}$

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

Here we will have a rarefaction wave as seen in figure (3.5).

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 smooth out the solution. One can then extend Godunov's method to create second order numerical methods such as the Lax-Wendroff method (see [37] for a detailed account) that models smooth solutions more accurately than the first order Godunov scheme but fails at discontinuities. To correct for 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

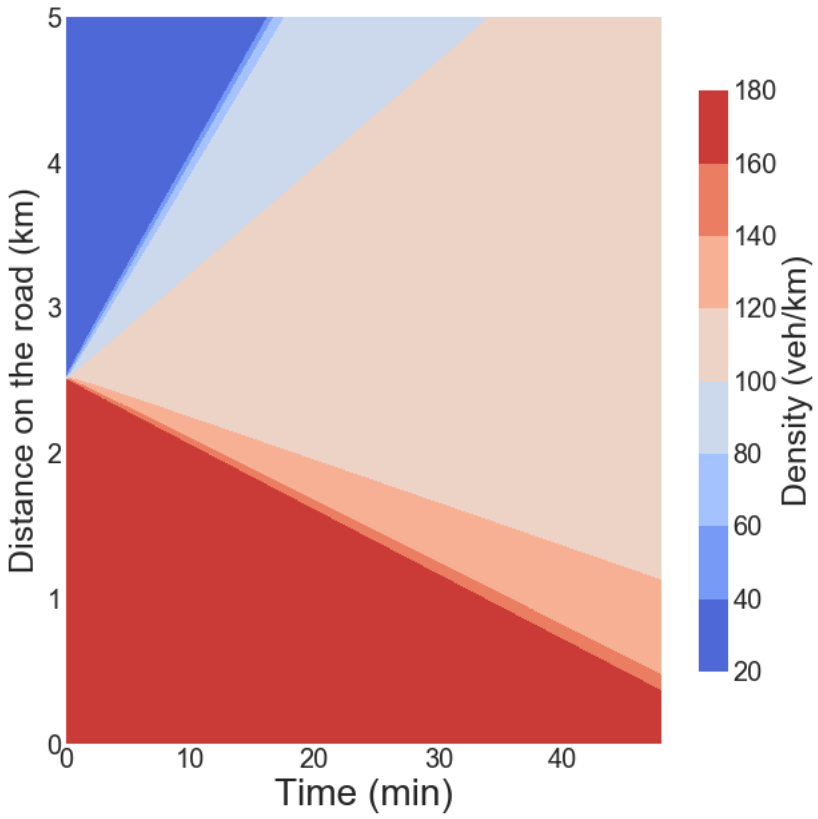

Figure 3.5: The Riemann Problem with $\rho_{i-1} > \rho_i$ causes a rarefaction wave. The initial condition (namely at $t = 0$) consists of a constant value of high density for $x \in [0, 2.5]$ and a constant value of low density for $x \in [2.5, 5]$. As the simulation moves forward in time we observe a rarefaction wave, or a fanning out of density values between the low and high values of the initial condition.

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 fundamental diagram for two different times to the output of Clawpack in figure (3.6).

To understand the relevance of shocks and this numerical viscosity to Bayesian inference, consider the problem of estimating the FD parameters in LWR using a Riemann Problem (case 1 above) as initial condition and a single detector in the x-t plane (for $t \neq 0$). As we vary the speed of the shock wave (by varying the parameters in the FD), the discontinuity will pass over the detector and cause the likelihood to jump. This discontinuity in the likelihood means that we cannot use gradient based meth-

ods in MCMC and that this may cause mixing difficulties. However as the numerical method smooths the discontinuity slightly, mixing might be slightly improved. Thus numerical viscosity is 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 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* - a necessary condition for convergence - is satisfied (see [37]). 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 (for example equation (3.4)), this means that information must only come from the adjacent cells and not from cells further away. We then define λ_{max} to be the fastest wave speed of the PDE and we require $\Delta t \leq \frac{\Delta x}{|\lambda_{max}|}$, ie:

$$\left| \frac{\Delta t \lambda_{max}}{\Delta x} \right| \leq 1 \tag{3.5}$$

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 non-adjacent 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 19m (as we have 259 cells for a 5km length of road), and pass in a new boundary condition density value every 1.5s (which corresponds to the time resolution required by CFL for typical FD parameter values found in data). Clawpack therefore chooses $\Delta t \leq 1.5s$ depending on the FD parameter values. We used this resolution for

the square wave test in figure (3.6). We note that this resolution (both time and space) is much finer than the one used by Polson and Sokolov in [51] when analysing real data (211m and 5min).

3.3 Exponential FD

We now estimate the parameters in LWR with the Exponential fundamental diagram introduced in section (1.1.2.1). We recall the model for this FD which has two parameters $\alpha > 0$ and $\beta > 0$:

$$q(\rho) = \alpha\rho \exp^{-\beta\rho}$$

We start by creating simulated data using LWR as the forward model with FD parameters $(\alpha, \beta) = (5, 0.012)$, and with i.i.d. Poisson error. We choose these parameters as they generate data that is similar to the M25 data; the algorithm developed for this data is therefore more likely to also work on the real dataset. We will use this simulated data throughout chapters 4 and 5 to develop the sampling methodology before testing it on traffic flow data from the M25 (introduced in the previous section). To create the simulated data we use boundary conditions sampled from the prior (a "log Ornstein-Uhlenbeck" process). For a given BC (ie: 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 following 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 standard deviation of the Wiener process respectively (see [32]). The elicitation of this prior is described in the next chapter, and from it we fit the parameters to obtain $\beta = 0.22$ and $\sigma = 0.256$.

We plot the BCs used to simulate data in figure (3.7), and show samples from the inlet prior in figure (4.8) (in the next chapter).

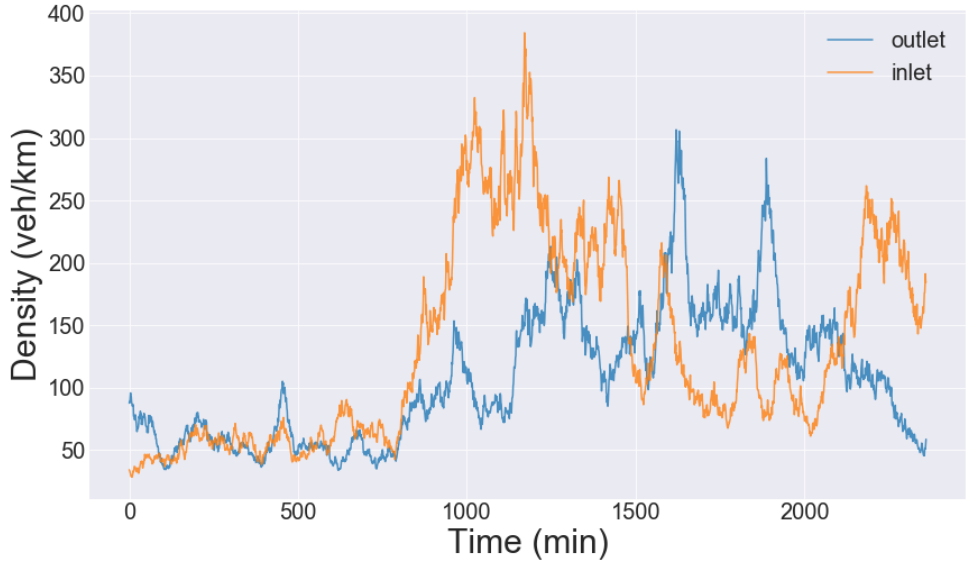

Figure 3.7: Inlet and outlet BCs used to generate simulated data. These are samples from the prior.

LWR requires the initial condition as well as the boundary conditions. This density will be propagated along characteristics (along either free flow or congested flow waves) 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 detectors times to build the likelihood; as a result the likelihood is unaffected by the initial condition and is only affected by the choice of boundary conditions and FD parameters. To be able to only remove a small amount of points in the $x-t$ plane (ie: 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 (3.8)). 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 thesis. Letting θ be the vector of FD parameters in LWR, and let (ρ_i, q_i) (with $i = 1 \rightarrow N$) be observed density and flow. We assume that data is iid with a Poisson error model which leads to the log-likelihood:

$$l_{LWR}(\theta) \propto \sum_{i=1}^{N} (-\hat{q}_i^{LWR}(\theta) + q_i \log(\hat{q}_i^{LWR}(\theta))) \tag{3.6}$$

With $\hat{q}_i^{LWR}(\theta)$ the predicted flow for the i^{th} observation resulting from the observation option $\mathscr{G}(\theta)$ as defined in section (1.3.1). This operator solves the LWR model given some boundary and initial conditions as well as 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. We use a uniform prior for the FD parameters (equation (3.7) below) and obtain the posterior.

$$\begin{cases} \alpha \sim \mathscr{U}(1,50) \\ \beta \sim \mathscr{U}(0.001,10) \end{cases} \tag{3.7}$$

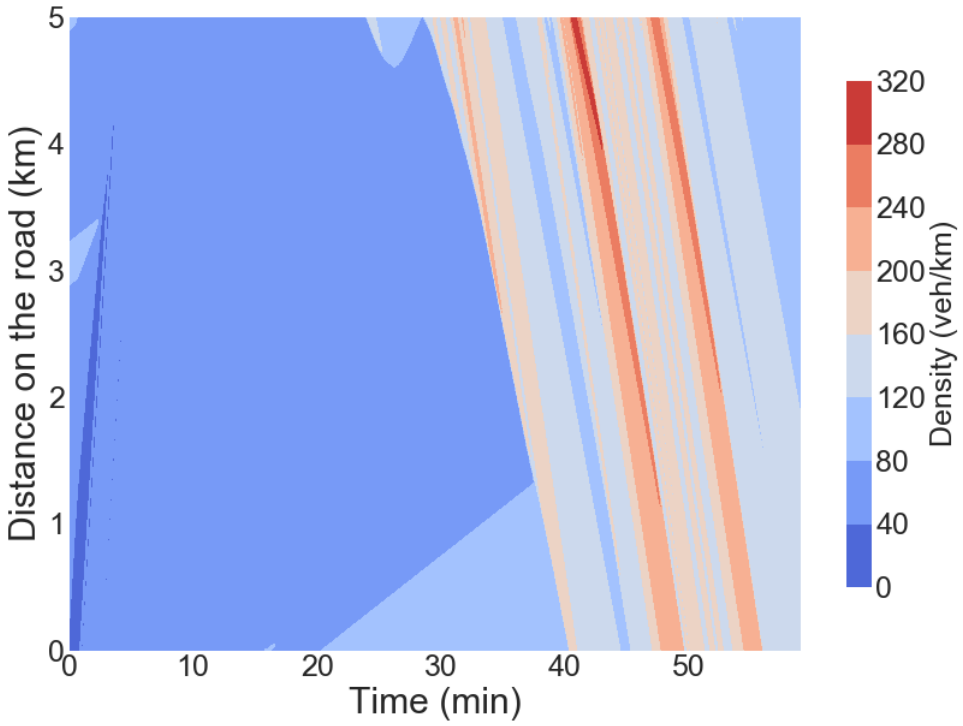

Figure 3.8: Simulated data from LWR with Exponential FD parameters $(\alpha,\beta) = (5, 0.012)$.

3.3.1 Simulated data

We employ a random walk Metropolis sampler on the two parameters $(\alpha,\beta)^T$ with covariance matrix $\begin{pmatrix} 0.0003865 & 0.00000088 \\ 0.00000088 & 3.269e-09 \end{pmatrix}$ and run 3 chains for 5K iterations. We tuned the covariance matrix by sampling both parameters (using an initial guess for the proposal variances) and fitting a covariance matrix to the joint samples. We repeated

this procedure a few times until the acceptance rate was approximately 23% (see [9] for an overview of results of optimal scaling of RWMH). We obtain acceptance rates for the 3 chains of 28.9%, 29.4%, and 29.1%, and show the traceplots (unthinned) in figure (3.9). We note that the true parameters $(\alpha, \beta) = (5, 0.012)$ are recovered in the inference. Visually, α and β seem to mix quickly, and indeed we observe in the autocorrelation plots for one of the chains (in figure 3.10) that the delay time is approximately 5. We also show the sampled FDs plotted against true flow-density values in figure (3.11).

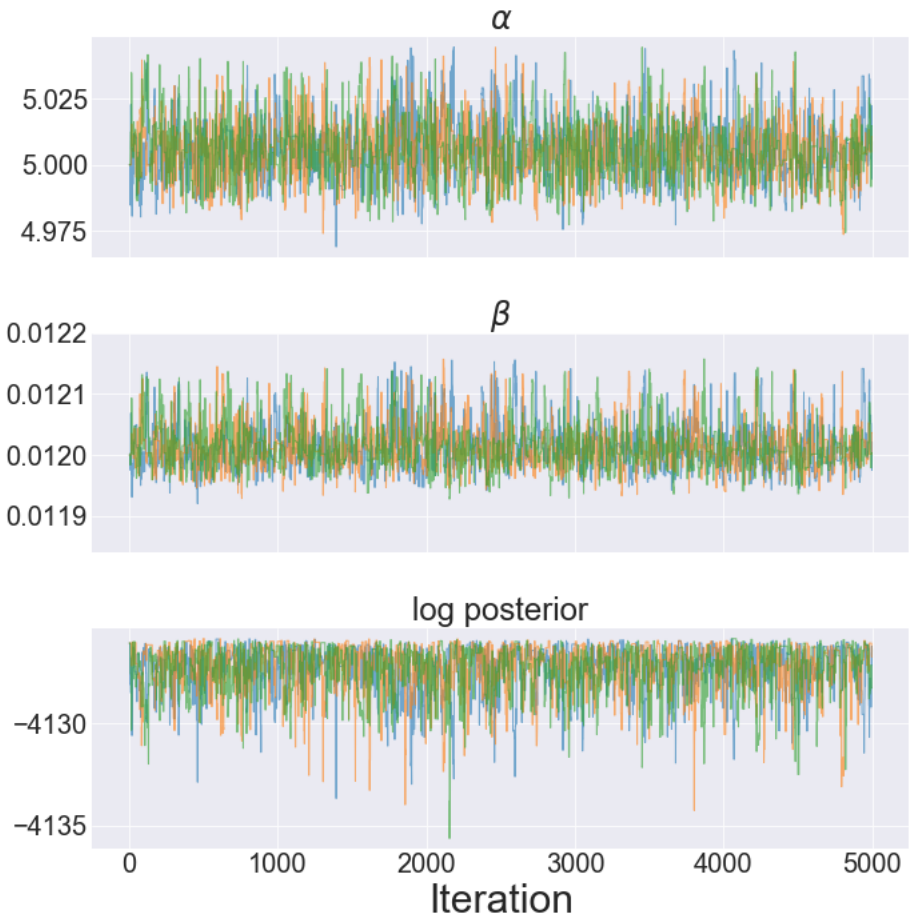

Figure 3.9: Unthinned trace plots from a random walk Metropolis sampler on the two parameters in the Exponential FD within LWR for simulated data. The 3 colours denote the 3 chains

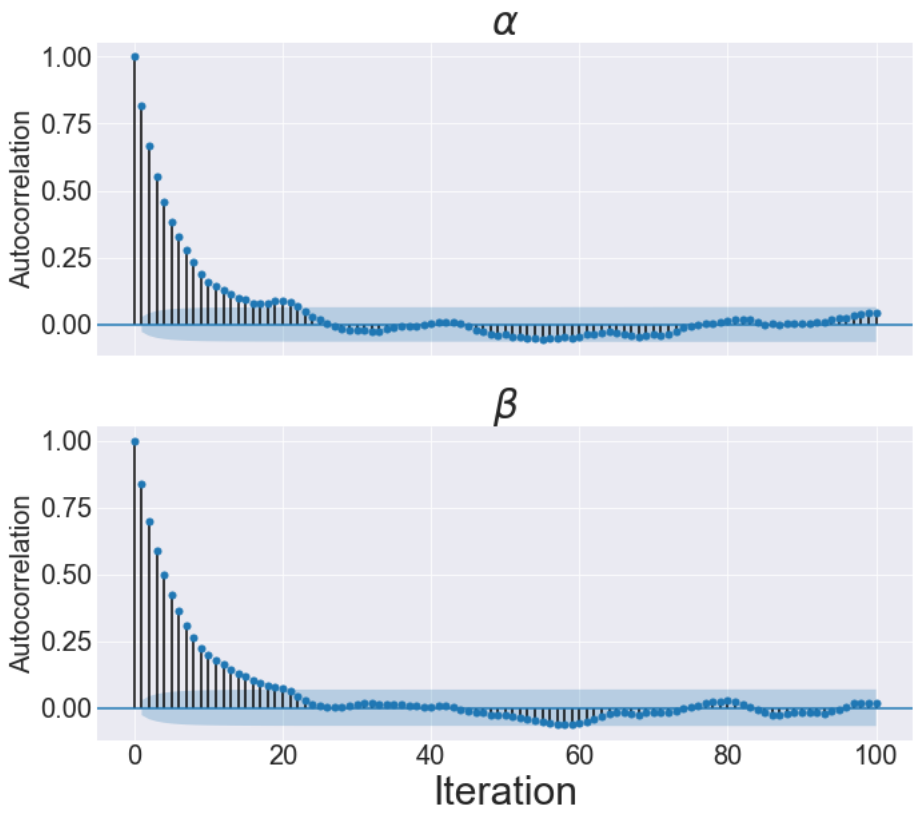

Figure 3.10: Autocorrelation plot for a single chain. The sampler is a random walk Metropolis sampler on the two parameters in the Exponential FD within LWR for simulated data

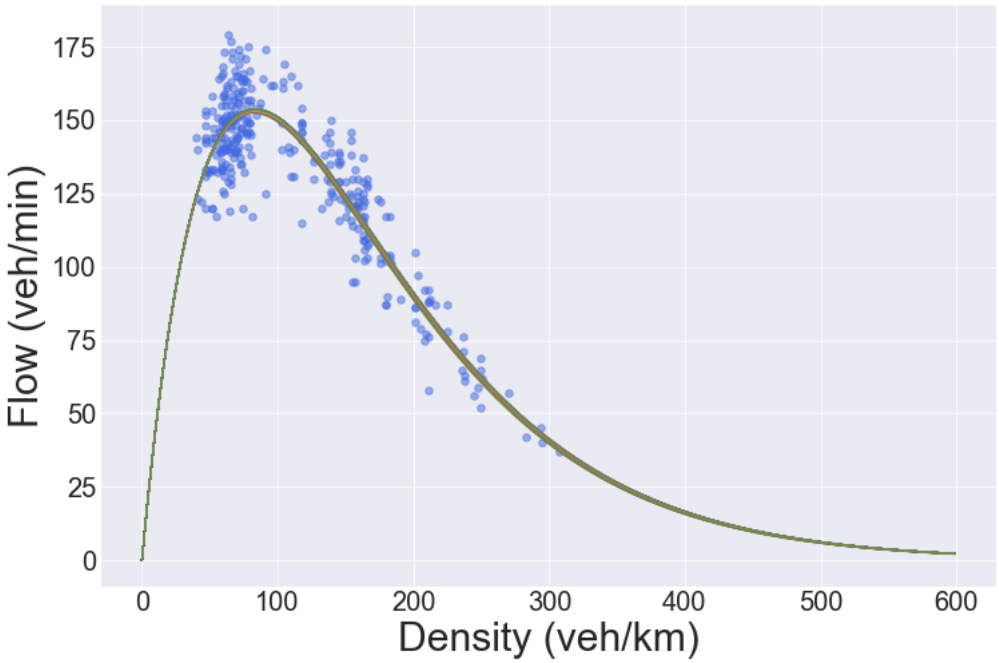

Figure 3.11: Exponential FD from sampled parameters plotted against simulated flow-density data

3.3.2 M25 data

We now estimate the parameters in the exponential FD for motorway data on the M25 described in the previous section. We start by first fitting the FD parameters to flow and density from occupancy data directly (namely without LWR). We then estimate them using LWR as a forward model.

3.3.2.1 Direct fit

We start by fitting the Exponential FD directly to M25 data shown in figure (3.3) and (3.4). As in the case for simulated data, we denote observed density and flow as (ρ_i, q_i) (with $i = 1 \rightarrow N$) and we assume that data is iid with a Poisson error model. The log-likelihood $l(\alpha, \beta)$ is written:

$$l(\alpha,\beta) \propto \sum_{i=1}^{N} (-\hat{q}_i(\alpha,\beta) + q_i \log(\hat{q}_i(\alpha,\beta))) \tag{3.8}$$

But this time $\hat{q}_i(\alpha,\beta)$ is flow predicted from observed density ρ_i using the Exponential FD with parameters (α,β) (so without using the LWR model):

$$\hat{q}_i(\alpha,\beta) = \alpha\rho_i \exp\{-\beta\rho_i\}$$

As this is a simple model and the likelihood is suitable informative, there is no particular need for a proper prior and so we use a flat prior $\pi_0(\alpha,\beta) \propto 1$. We therefore obtain the posterior: $\pi(\alpha,\beta) \propto l(\alpha,\beta)$.

We sample from this posterior using a RWMH algorithm with covariance matrix $\begin{pmatrix} 3.95e-03 & 1.10e-05 \\ 1.10e-05 & 3.56e-08 \end{pmatrix}$. We run 3 chains for 5K iterations and obtain acceptance rates of 26.3%, 24.6%, and 24.6%. We show the trace plots in figure (3.12) which visually seem to mix well. Indeed inspecting the ACF plots for α and β (not shown) shows that the delay time is approximately 5 for both parameters. We show FDs with sampled parameters (thinned by 100) in figure (3.13)

We then show the output from LWR in the $x-t$ plane using the posterior mean samples in figure (3.14). Comparing this plot to the real data in figures (3.3) and (3.4) 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 suggest a misfit of the model.

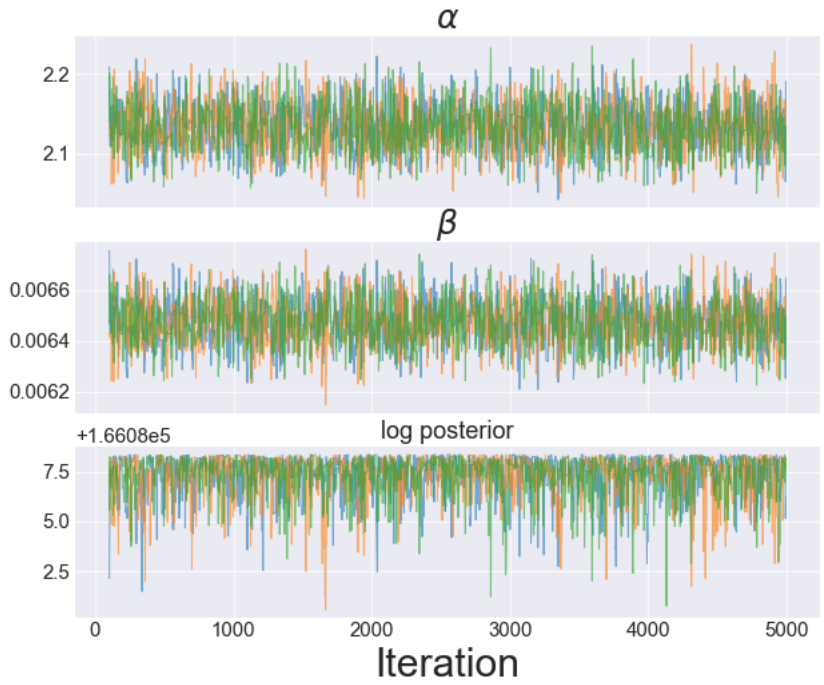

Figure 3.12: Trace plots for the two estimated parameters from a direct fit of exponential FD to M25 data.

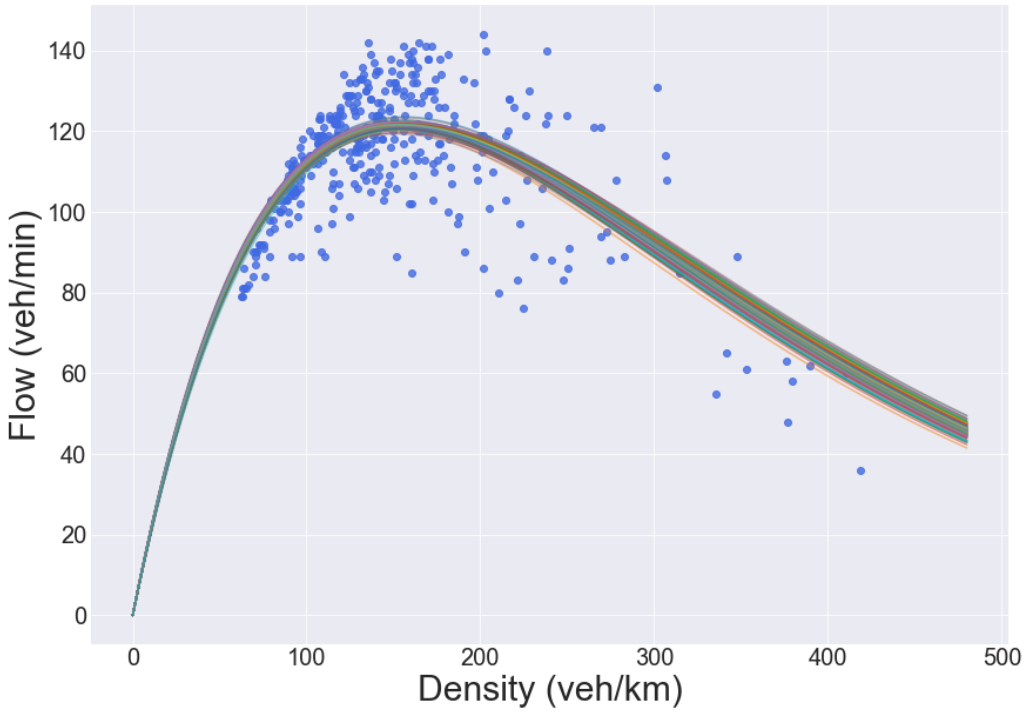

Figure 3.13: Plot of FDs from sampled parameters against flow-density data from a direct fit of exponential FD to M25 data.

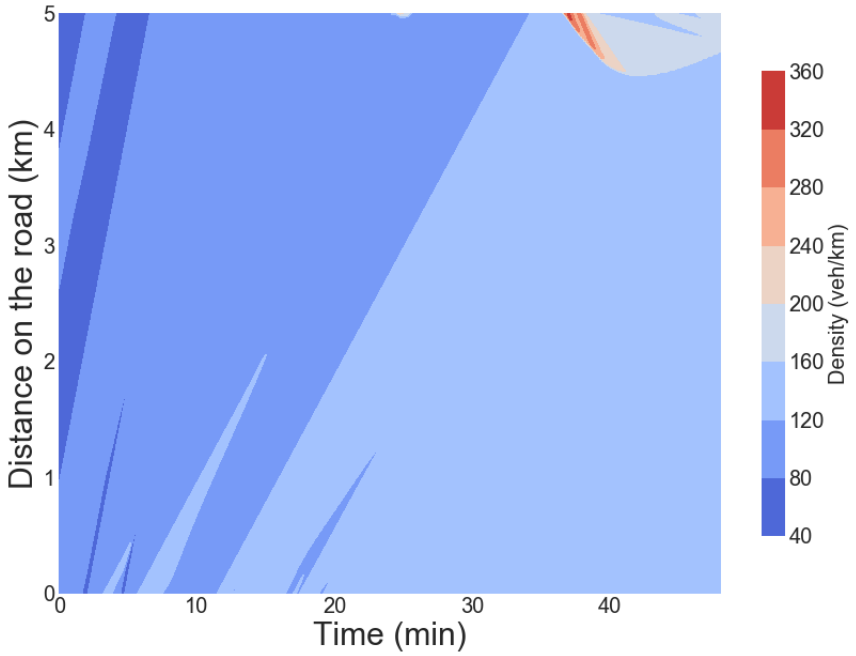

Figure 3.14: Density in the $x-t$ plane from LWR. Parameters used are the posterior mean from samples arising from a direct fit of the Exponential FD to M25 data. We notice that the congested flow waves (seen in the top right hand corner) in this plot do not cross the domain as they do in the data. Indeed in figure (3.4) (X-T plot for the M25 data) these congested flow waves cross the domain.

3.3.2.2 MCMC with LWR

We now fit the exponential FD to M25 data using LWR as the forward model (so using equation (3.6)). We fix the BCs to density from occupancy from M25 data. However, the observations are only available every minute but we need a 1.5s resolution to insert in the solver. We therefore need to interpolate in between each minute to obtain BCs at the correct resolution. We do this by sampling from the prior conditional on the density from occupancy data observed at every minute, and plot both BCs in figure (3.15).

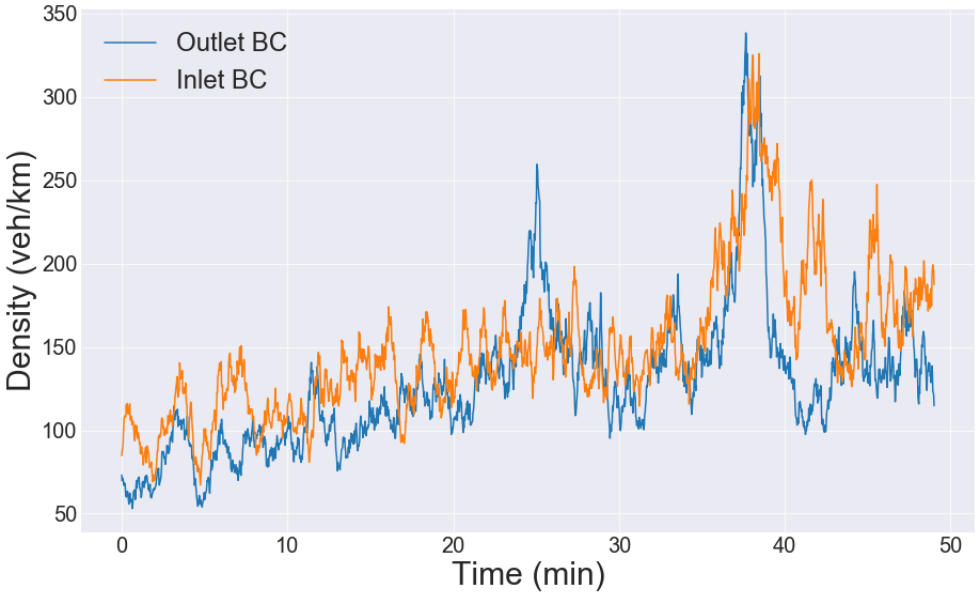

Figure 3.15: Inlet and outlet BCs to use in $FD|BC$ sampling. These are samples from the log-OU prior conditional on density from occupancy (M25 data) at every minute.

We use a random walk Metropolis sampler as before with covariance matrix $\begin{pmatrix} 0.00913695 & 0.00002983 \\ 0.00002983 & 0.0000001 \end{pmatrix}$, run 3 chains for 5K iterations, and obtain acceptance rates of 28.5%, 27.5%, and 29.9%. We show the trace plots in figure (3.17) which seem to show fast mixing. We do not show the ACF plots for the two parameters as they are qualitatively similar to the ones in the simulated data case in figure (3.10); the difference is that here the delay time is approximately 10 for both parameters. We note the bimodality for both parameters (which was not observed in the case of the simulated data) which can be seen more clearly when plotting FD samples in figure (3.16).

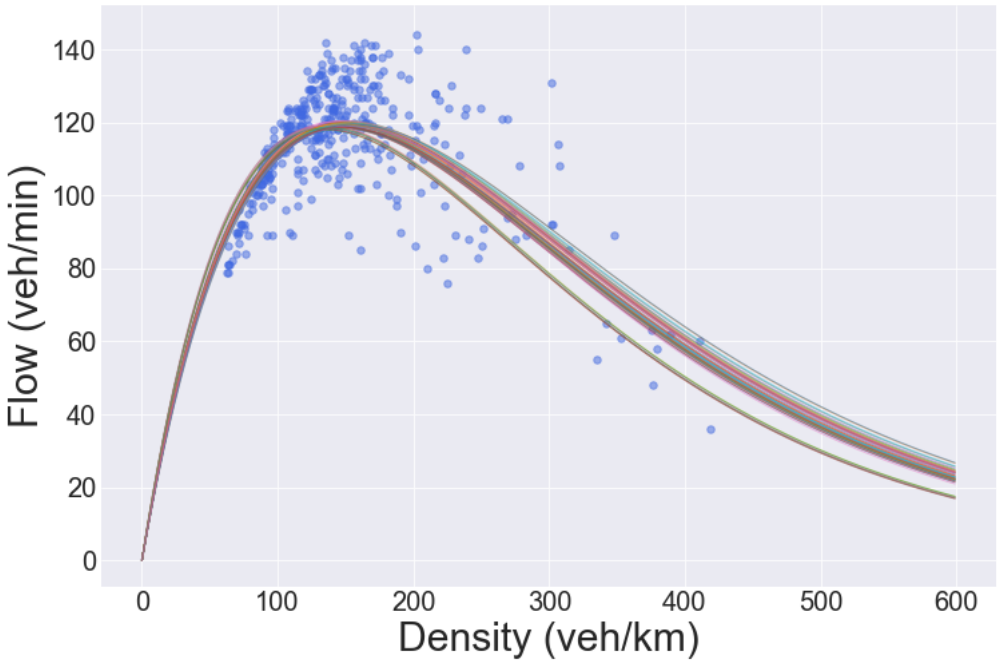

Figure 3.16: Exponential FD for sampled parameters plotted against flow-density data. The samples are from 3 chains fit to M25 data using LWR. We note the bimodality that was not present in the simulated data.

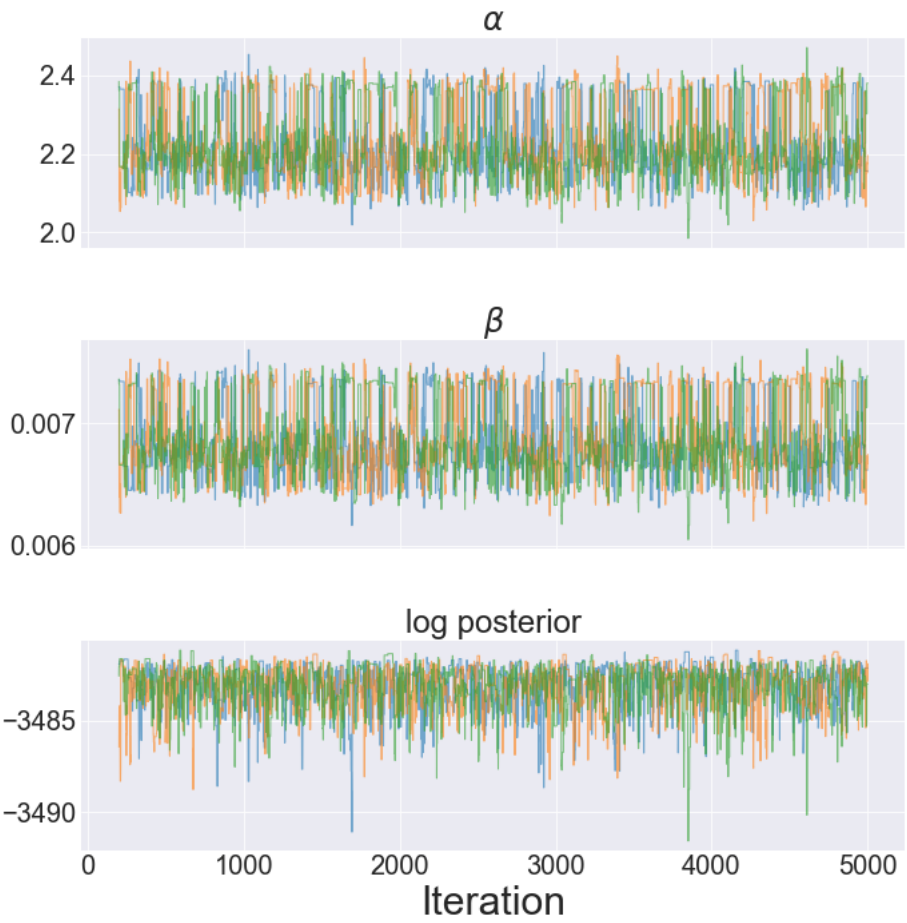

Figure 3.17: Trace plots for the Exponential FD fit to M25 data using LWR. Note the bimodality in both parameter which can be seen in figure (3.16)

As this bimodality was not found in the simulated data, we are led to consider that the model might not fit the data. We define the ”top” mode to be the α and β samples with higher values as seen in figure (3.17), and we define the ”bottom” mode to be the ones with the lower values. We note however that the ”top” mode corresponds to the FD with the lower congested flow branch in figure (3.16) while the ”bottom” mode corresponds to the higher congested flow branch.

We plot in figure (3.18) and (3.19) the outputs of LWR using a sample from the ”top” mode $((\alpha,\beta) = (2.371, 0.0073))$ and ”bottom” mode $((\alpha,\beta) = (2.106, 0.0064))$ respectively. We notice that in both cases, the congested flow waves cross the domain a little more than in the direct fit. This suggests that fitting the FD parameters with LWR as the model (rather than the direct fit) penalises misfitting the congested flow wave speeds. In contrast, the direct fit in figure (3.13) simply goes through the points in the flow-density diagram.

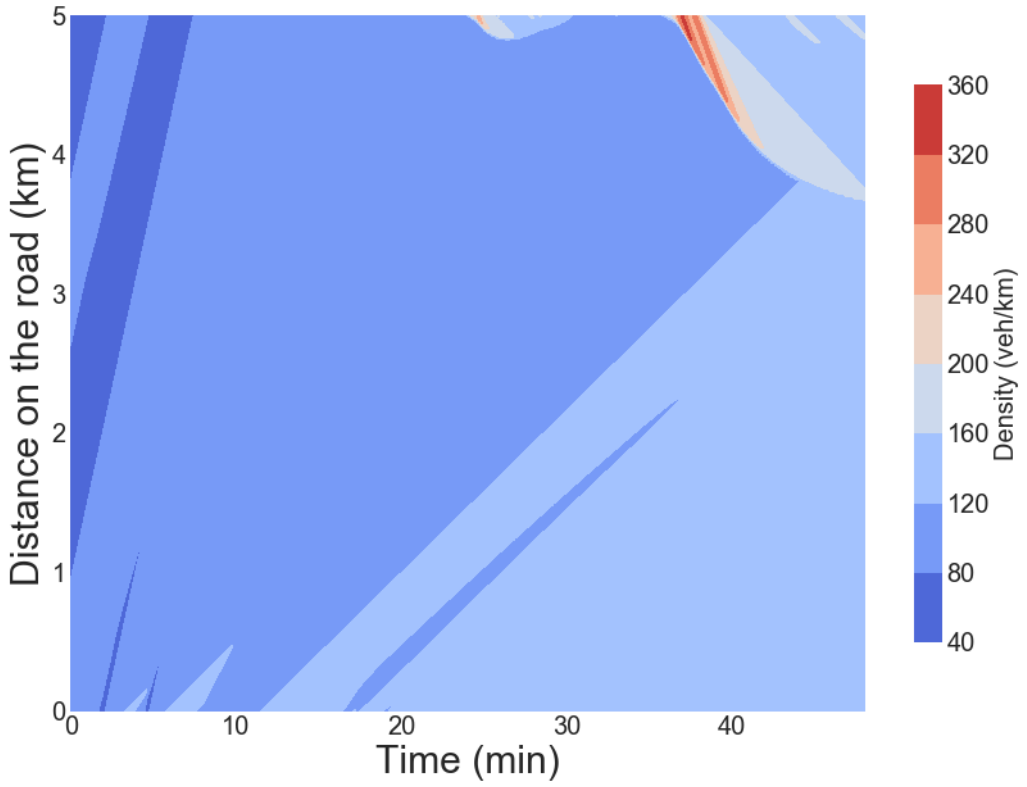

Figure 3.18: LWR output using a sample from the ”top” mode as seen in figure (3.17). The congested flow waves still do not cross the domain, but we see a slight improvement from the direct fit in figure (3.14)

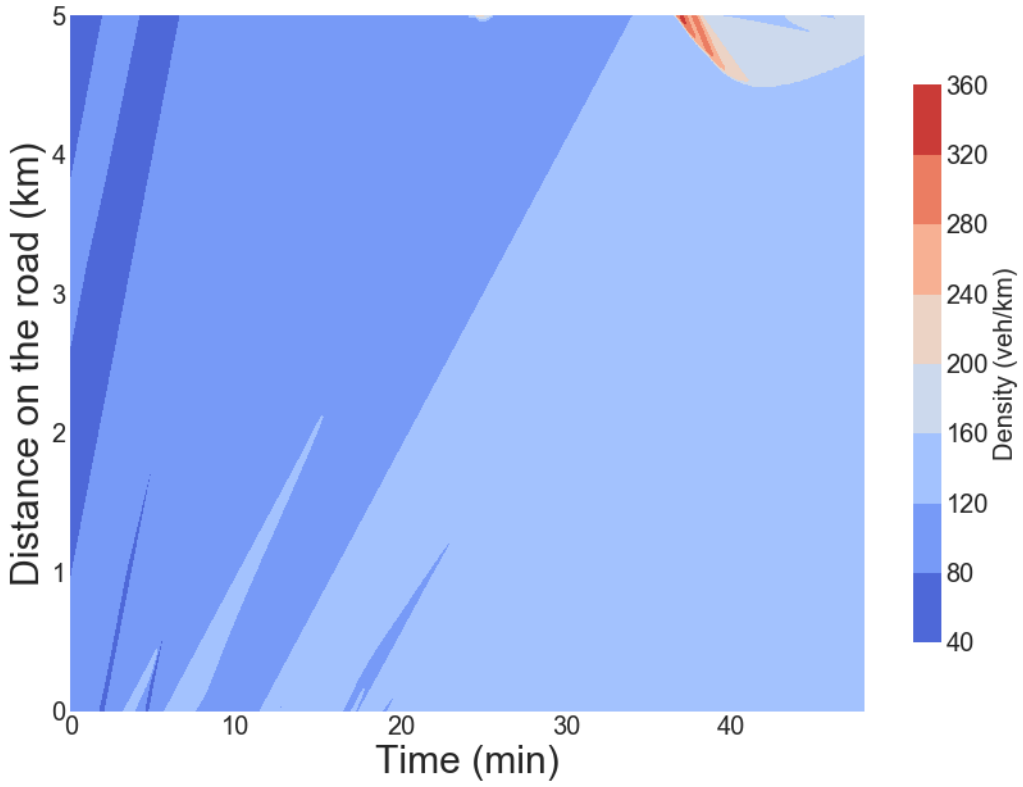

Figure 3.19: LWR output using the samples from the "bottom" mode as seen in figure (3.17). The congested flow waves still do not cross the domain, but we see a slight improvement from the direct fit in figure (3.14)

3.4 Del Castillo FD

We now estimate the parameters in Del Castillo's FD (introduced in section (1.1.2.1) and plotted in figure (1.5)). We recall the model which maps density to flow and is parametrised by (z, ρ_j, u, w):

$$q_{dC}(\rho; z, \rho_j, u, w) = Z\left[(u\frac{\rho}{\rho_j})^{-w} + (1 - \frac{\rho}{\rho_j})^{-w}\right]^{-\frac{1}{w}} \tag{3.9}$$

The w parameter corresponds to the tightness of the peak, so as $w \to \infty$ the FD tends to the Triangular FD (equation (1.2) in section (1.1.2.1). In order to allow this parameter to take high values but while also allowing the Markov chain to converge, we invert this parameter and hence sample $\frac{1}{w}$ (we continue to denote this inverted parameter as w). We use the same likelihood model as in the case of the Exponential FD in equation (3.6) and set uniform priors for the FD parameters shown in equation (3.10) below:

$$\begin{cases} z \sim \mathscr{U}(100,400) \\ \rho_j \sim \mathscr{U}(300,800) \\ u \sim \mathscr{U}(1,10) \\ w \sim \mathscr{U}(0.004,10) \end{cases} \tag{3.10}$$

3.4.1 Simulated data

We simulate data using parameters $(z,\rho_j,u,w) = (180,410,3.2,0.1)$ (which generate similar data to the M25 data) and the same BCs as in the simulated data for the exponential FD. We plot density for this simulated data in the $x-t$ plane in figure (3.20)).

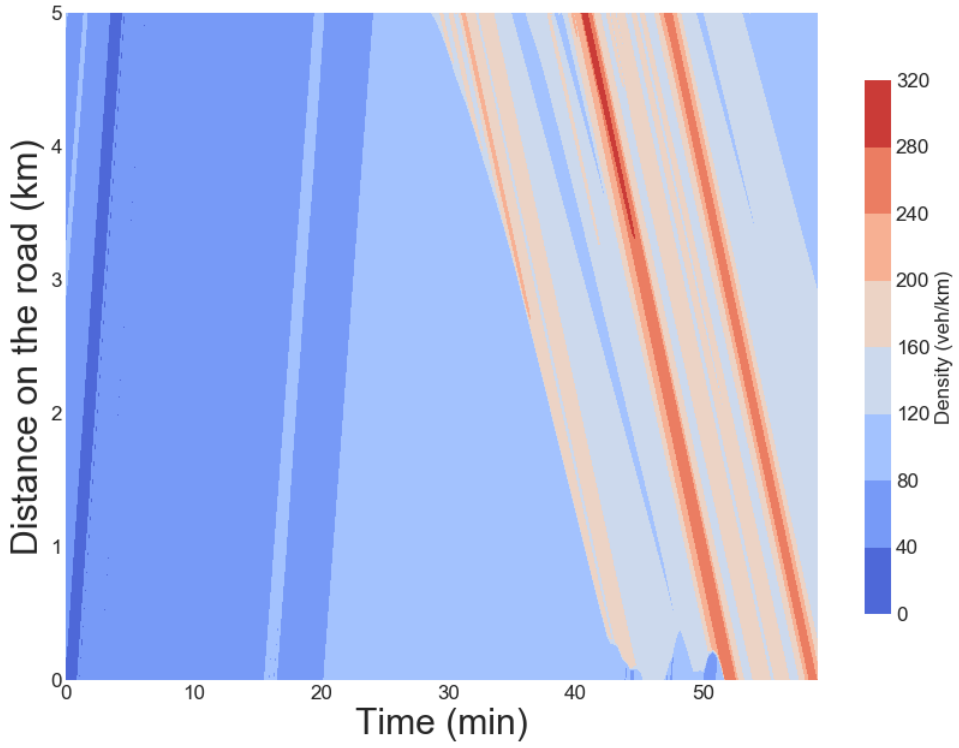

Figure 3.20: Simulated data from LWR with del Castillo's FD with parameters $(z,\rho_j,u,\omega) = (180,410,3.2,0.1)$

We run a RWMH sampler to test whether we can recover the true parameters in the inference. We use a multivariate Gaussian proposal for the four parameters conditional on $w > 0$. The covariance matrix used is given in the Appendix (section (B.1)), and we run 3 chains for 5K iterations.

We obtain acceptance rates of 20.7%, 19.7%, and 18.6% for the 3 chains, and we show the trace plots in figure (3.21) which seem to show good mixing. Indeed

inspecting the ACF plot (not shown) shows that the delay time is approximately 10. We also show the sampled FDs against flow and density data in figure (3.22). As in the case of simulated data from the Exponential FD, we recover the true parameters in the inference.

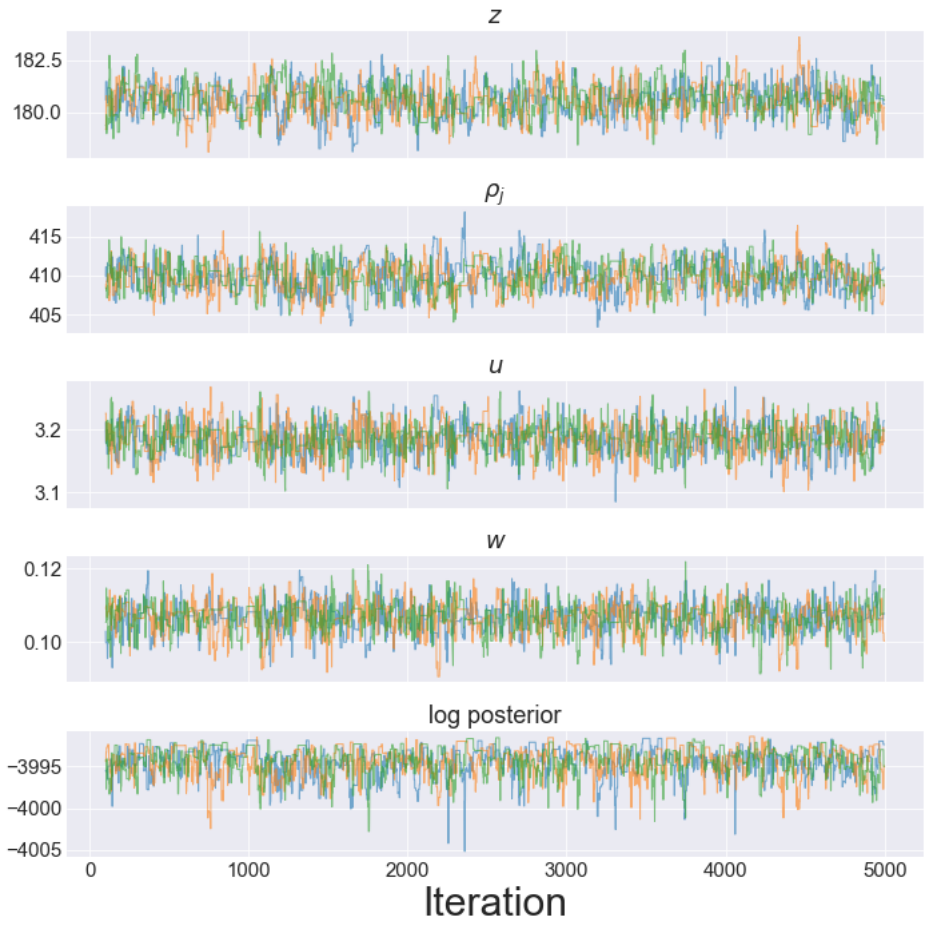

Figure 3.21: Trace plots for the 4 parameters in del Castillo's FD for a RWMH sampler for LWR on simulated data.

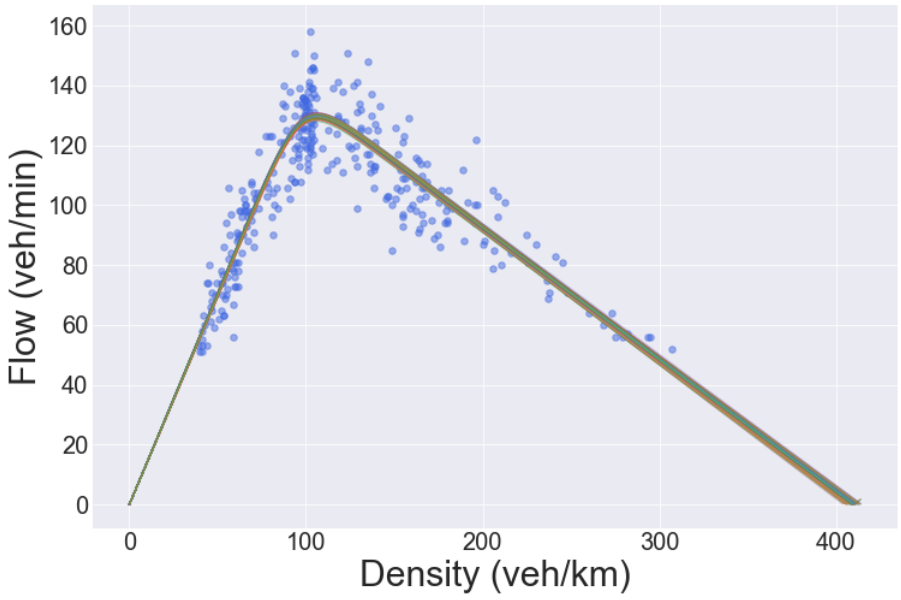

Figure 3.22: Del Castillo's FD plotted using sampled parameters against simulated flow-density data

3.4.2 M25 data

3.4.2.1 Direct fit

As in the previous section, we first fit the FD directly to flow-density data from the M25. We use the same error model (iid Poisson) as for the direct fit of the Exponential FD. As a result, the likelihood is the same as equation (3.8) but with the flow $\hat{q}_i$ predicted using del Castillo's FD $q_{dC}(\rho; z, \rho_j, u, w)$ defined in equation (3.9):

$$\hat{q}_i(z, \rho_j, u, w) = q_{dC}(\rho_i; z, \rho_j, u, w)$$

We use a flat prior as in the direct fit of the Exponential FD, and sample from the posterior using a RWMH algorithm. We run 3 chains for 5K iterations with covariance matrix given in the appendix in section (B.2). We obtain acceptance rates of 24.4%, 23.1%, and 23.8%, and show the trace plots for the parameters in figure (3.23) which suggest good mixing. The ACF (plot not shown) reveals that the delay time is approximately 9, which further reinforces the conclusion of good mixing. We show FD plots with sampled parameters against flow-density data in figure (3.24); as in the case of the Exponential FD the curve simply goes through the points as expected. However as

before this lack of dynamic information could be the cause of the misfit to data seen in figure (3.25) where we see the congested flow waves not crossing the $x-t$ plane.

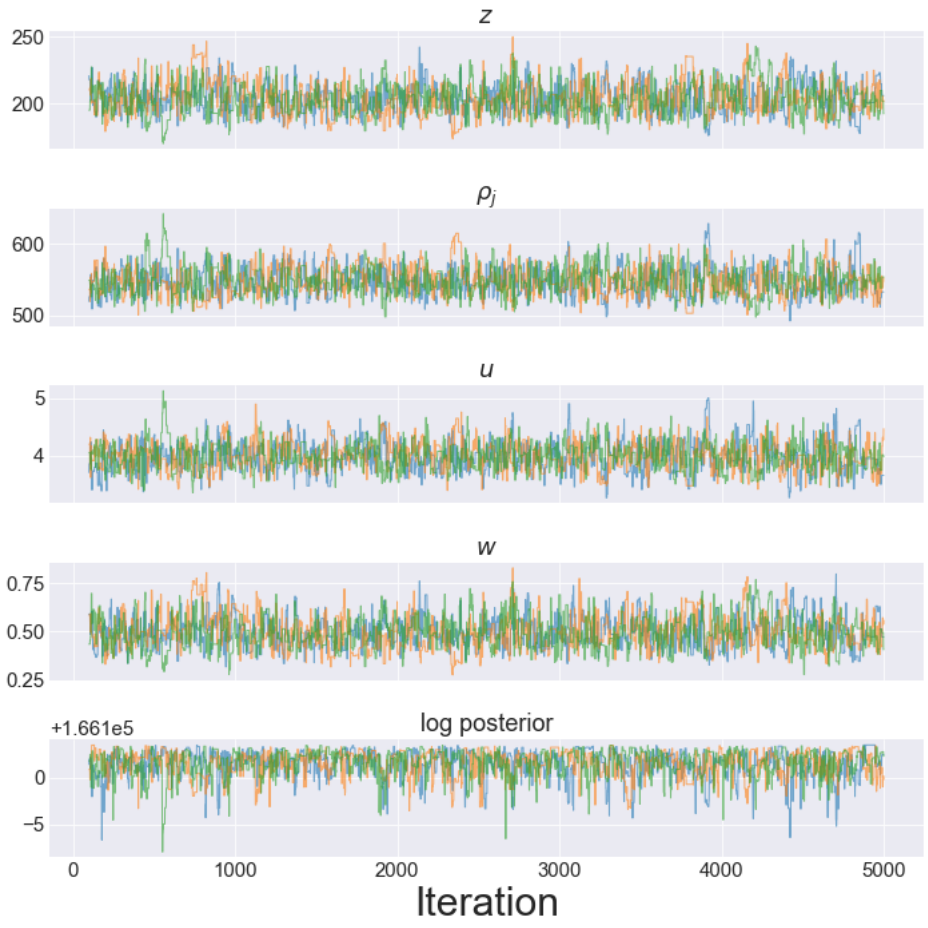

Figure 3.23: Trace plots for del Castillo's parameters from a direct fit of del Castillo's FD to M25 data.

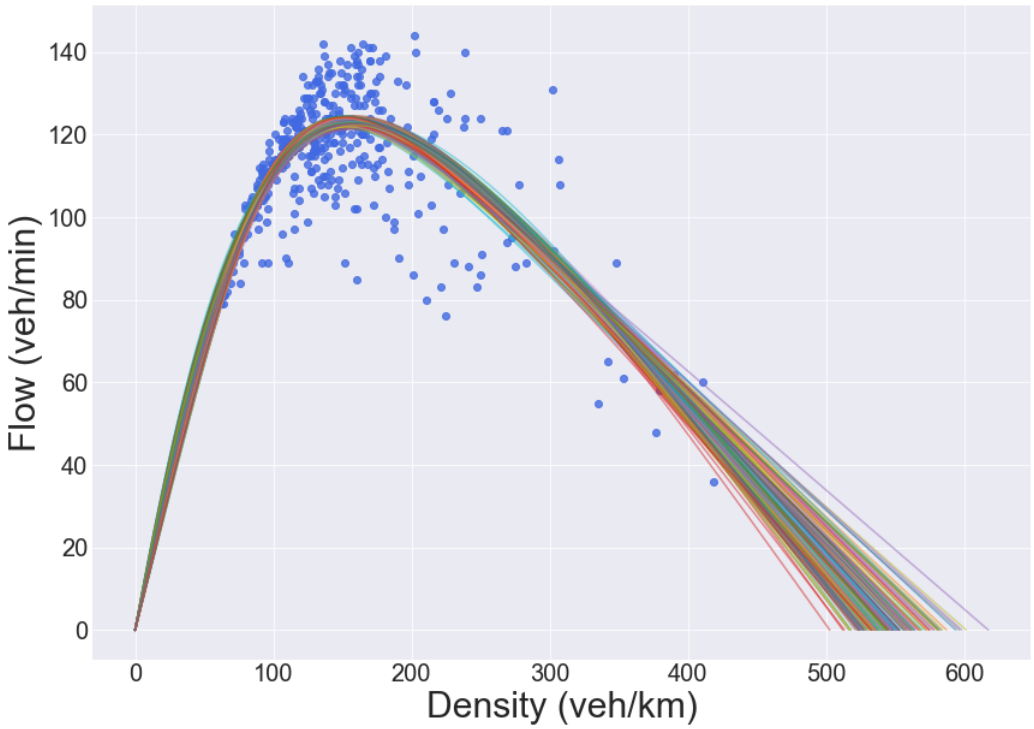

Figure 3.24: Plotted FDs from sampled parameters against flow-density data from a direct fit of del Castillo's FD to M25 data.

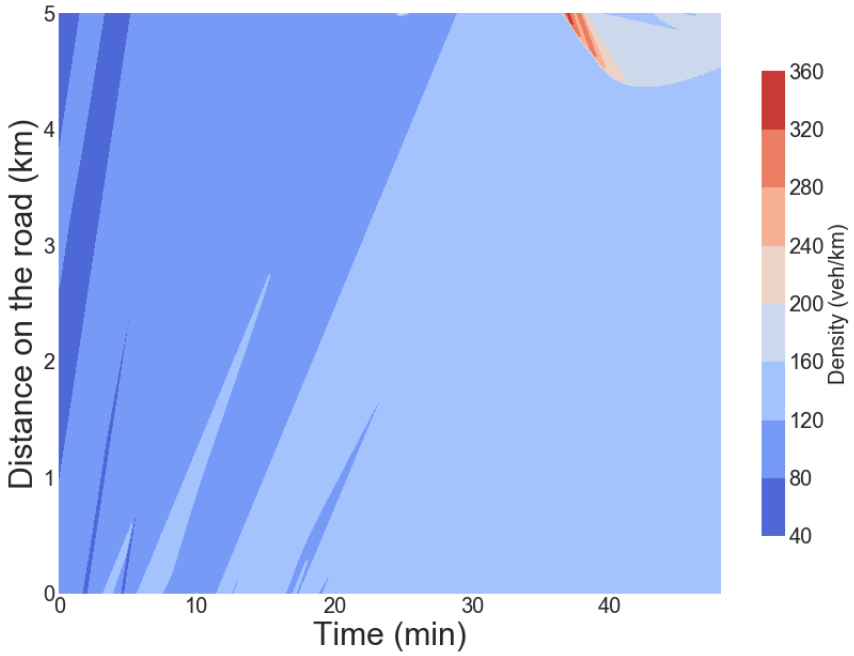

Figure 3.25: Density in the $x-t$ plane from LWR. Parameters used are the posterior mean from samples arising from a direct fit of del Castillo's FD to M25 data. We notice that the congested flow waves do not cross the domain as they do in the data

3.4.2.2 MCMC with LWR

We now employ a RWMH sampler to fit del Castillo's FD to M25 data using LWR as the forward model (so using equation (3.6)). As in the case of the Exponential FD, we fix the BCs to density from occupancy while interpolating in between each minute using the prior. We recall that the BCs are plotted in figure (3.15).

We use the covariance matrix given in the appendix in section (B.2) and run 3 chains for 5K iterations. We obtain the acceptance rates 24.2%, 26.3%, and 25.3% for the 3 chains. We show the trace plots in figure (3.26) which seems to show good mixing and from the ACF plots (not shown) we observe that the delay time is approximately 10. We also show the FD samples in figure (3.27).

Here we see in figure (3.28) that the congested flow waves cross the domain (as in the real data), unlike in the direct fit to data. Similarly to the case of the Exponential FD in section (3.3.2.2), this suggests that the likelihood built with LWR (rather than the direct fit) penalises misfitting the congested flow wave speeds. As the direct fit does not consider time or space, this result confirms the intuition that using a likelihood that includes dynamic information about the system yields a better fit to data.

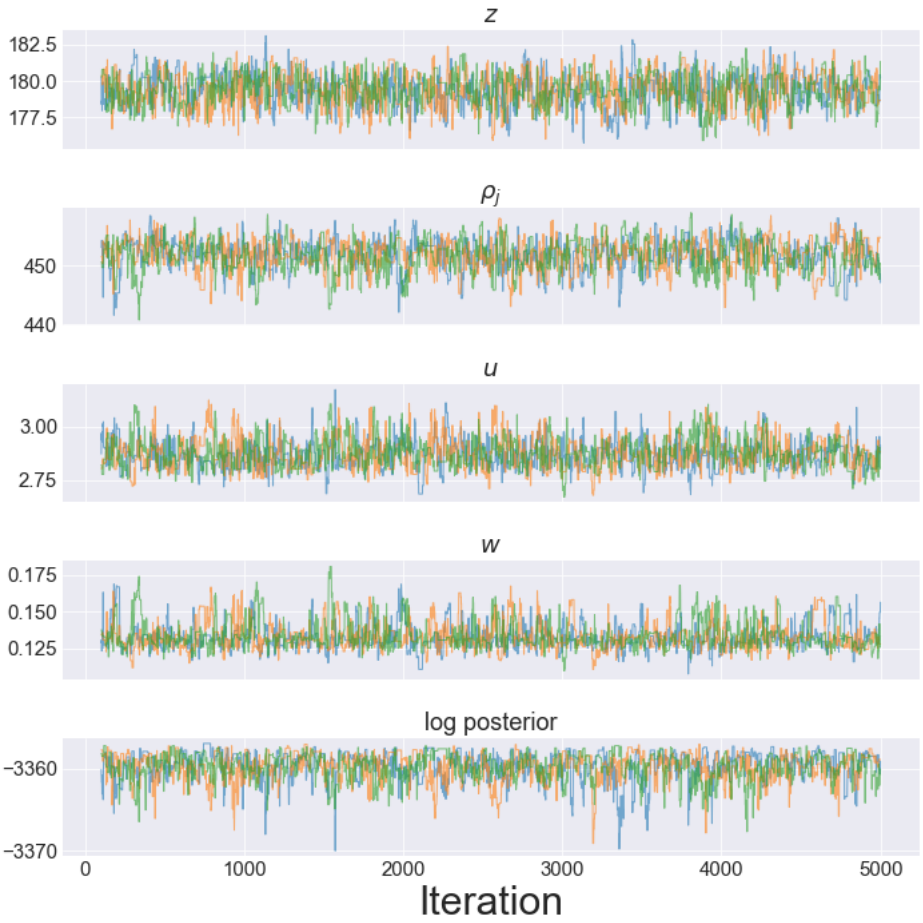

Figure 3.26: Trace plots for del Castillo's parameters when fit to M25 data using LWR.

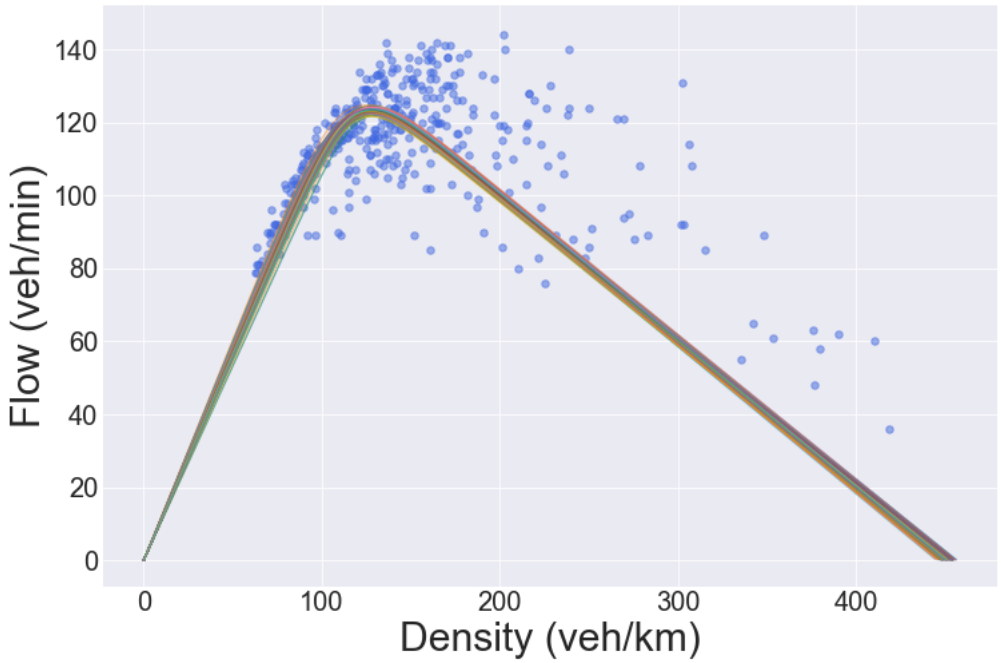

Figure 3.27: Del Castillo's FD for sampled parameters plotted against M25 flow-density data.

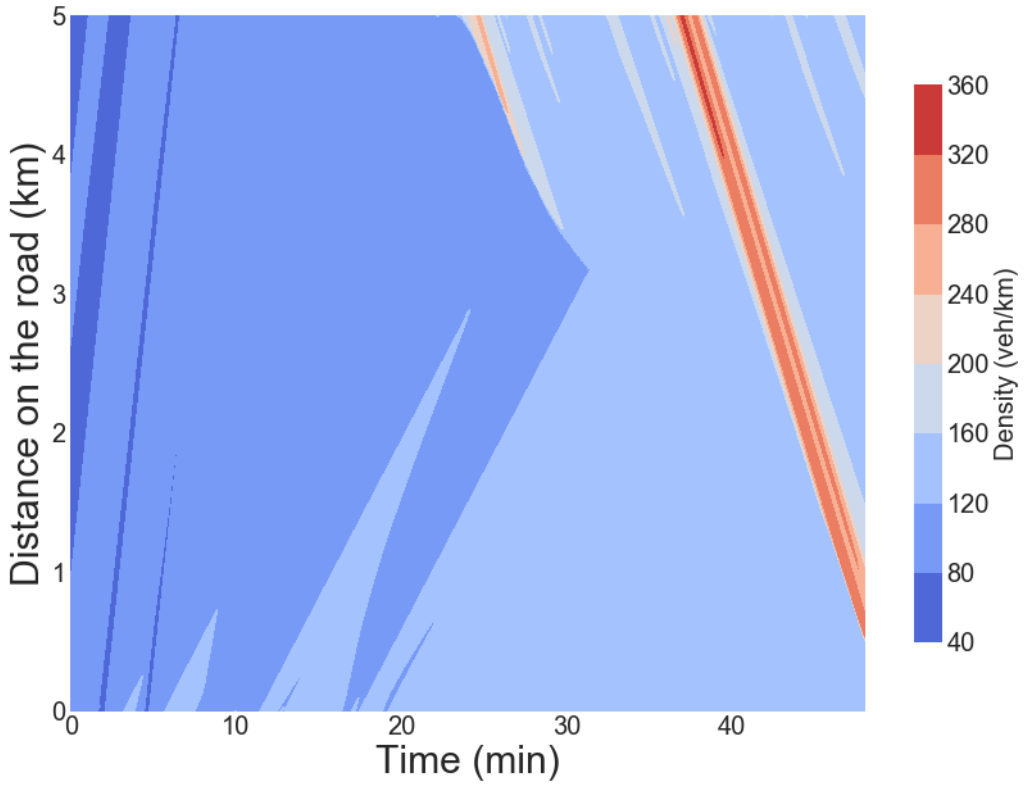

Figure 3.28: Density in the $x-t$ plane from LWR using the posterior mean arising from LWR fit to M25 data with del Castillo's FD. Unlike in the case of the direct fit to data, the congested flow waves cross the domain as in real data show in figures (3.4) and (3.3).

3.5 Conclusion

In this chapter we process M25 motorway data and fit the Exponential and del Castillo FD directly to data (ie without LWR). We find that this fit gives the wrong congested flow wave speed. We then fit these two FDs with LWR as the forward model both using simulated data and the M25 data, with fixed Boundary Conditions (BCs). For the simulated data we sample BCs from a log-OU process, and for the M25 data we use BCs estimated directly from data using a standard approach from the traffic flow literature. Del Castillo's FD is found to give the correct wave speeds which the Exponential FD does not. The fitted del Castillo FD is close to the Triangular FD which confirm practitioners' belief that the latter is a robust model to use in practical applications.

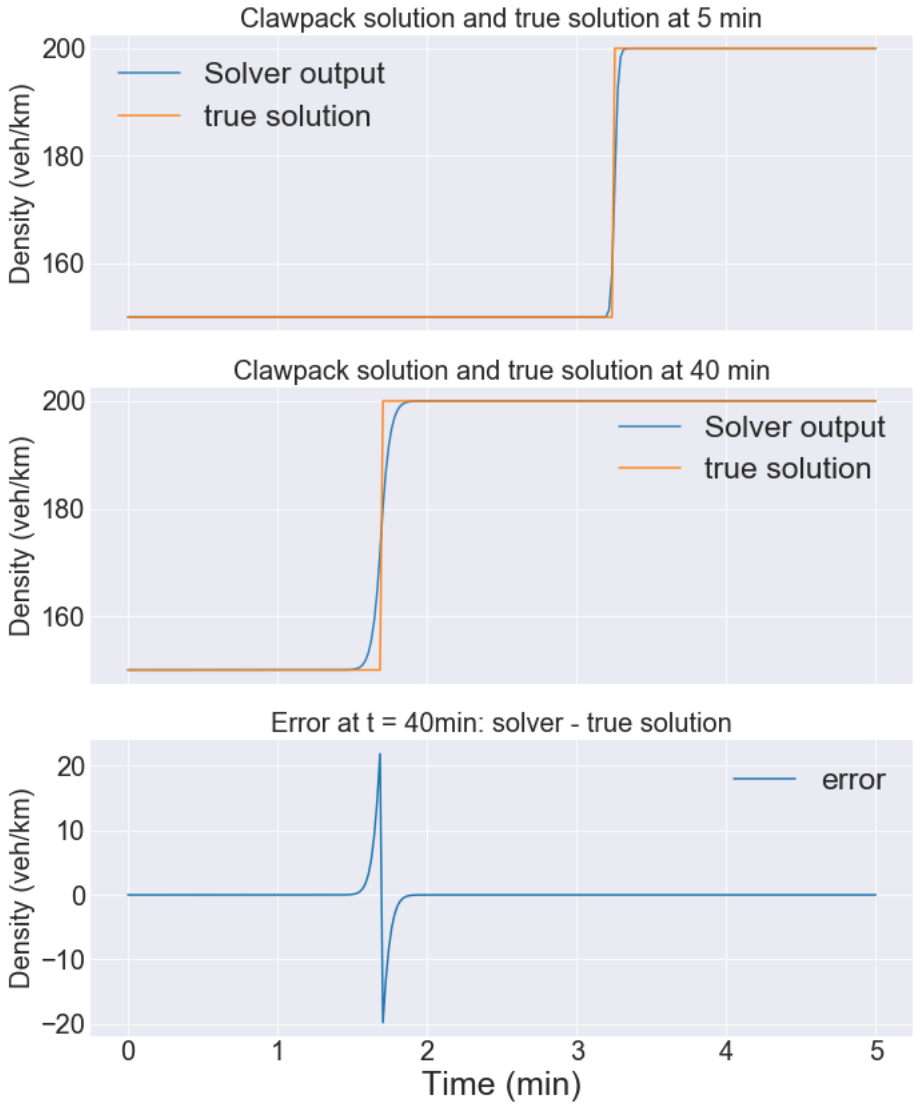

Figure 3.6: We plot the analytic solution to the Riemann Problem along with it's numerical solution using Clawpack. As time progresses we observe that the discontinuity is smoothed slightly. However we notice that the position of the shock wave remains accurate.

Chapter 4

Estimating Boundary Conditions

In the previous chapter we used density estimated from occupancy to define the boundary conditions (BCs) in LWR as this method was determined to be the most appropriate. However density is a continuous quantity and is the result of a limiting procedure while traffic is composed of discrete vehicles. We therefore prefer to not impose it and rather treat it as a parameter to infer.

4.1 Eliciting the prior

As density from occupancy (see equation (3.1)) is considered an appropriate method for estimating density, we will use it to elicit the prior for the BCs. As discussed in the previous chapter, 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 (as described in the previous chapter) to fit a Gaussian process prior for the BCs (of course discarding the day that we use in inference).

4.1.1 OU process prior

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 (so this is the stochastic process version of a log-normal random variable). We choose this prior for 3 reasons:

- We would like density to always be positive; a log-OU prior satisfies this requirement.
- 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 (explained below).

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 (ie: 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 following 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 standard deviation of the Wiener process respectively (see [32])).

Solving the SDE (see section 1.13.1 in [32]) we obtain the density of the distribution of X_t given $X_0 = x_0$:

$$(X_t|X_0 = x_0) \sim \mathcal{N}\left(x_0 e^{-\beta t}, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})\right)$$

We can then discretise the process into N cells using time step Δt, and write the joint distribution of the random variables $(X_0, X_{\Delta t}, ..., X_{(N-1)\Delta t})$:

$$P(X_0, X_{\Delta t}, ..., X_{(N-1)\Delta t}) = P(X_0) \prod_{i=1}^{N-1} P(X_{i\Delta t}|X_{(i-1)\Delta t})$$

We choose $P(X_0)$ to be the invariant distribution $\mathcal{N}(0, \frac{\sigma^2}{2\beta})$, which results in the marginal distribution of the process at every point being the same. We therefore obtain the joint distribution:

$$P(x_0, x_{\Delta t}, ..., x_{(N-1)\Delta t}) \propto \exp\left\{-\frac{\beta}{\sigma^2}\right\} \exp\left\{-\frac{\beta}{\sigma^2(1 - e^{-2\beta\Delta t})} \sum_{i=1}^{N-1} (x_{i\Delta t} - x_{(i-1)\Delta t} e^{-\beta\Delta t})^2\right\}$$

We can rewrite the right-hand side in matrix form (which gives a precision matrix

Λ) to obtain the following:

$$\pi(\underline{x}) \propto \exp\{-\frac{1}{2}\underline{x}^T \Lambda \underline{x}\}$$

This means that we can consider the values that makes up the logarithm of the BCs to follow a multivariate Gaussian distribution with precision matrix Λ found above.

To sample from this Gaussian $\underline{x} \sim \mathcal{N}(0, \Lambda^{-1})$, we first decompose the precision matrix into Cholesky factors: $\Lambda = LL^T$ and then solve the linear system $L^T \underline{x} = \underline{\xi}$ with $\underline{\xi} \sim \mathcal{N}(0, I)$.

4.1.2 Fitting the OU process

We first fit an OU process to centred log-BCs for each inlet and outlet BC separately. The inlet BC is a function of time that returns density, and corresponds to the inlet of the studied stretch of road (ie: $x = 0$) . Similarly, the outlet BC corresponds to the outlet of the road (ie: $x = 5$km). As the inference for the two BCs results in similar OU parameters, we then fit the prior for both log-BCs together.

For the inlet and outlet detector data we keep only weekdays, discard the 1st January, 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 the 8th January as this is the dataset used in the inference.

We recall the notation defined in section (4.1.1): 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. 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 minute). 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 separately using MCMC. Defining X^i to be the i-th measured density curve and Λ the precision matrix for the OU process, we write the likelihood as:

$$l(\Lambda) \propto -\frac{N}{2}\log|\Lambda| - \frac{1}{2}\sum^{N} X^i \Lambda X^i \tag{4.1}$$

We use a flat prior for the parameters and use a random walk Metropolis sampler to sample from the posterior. We show the trace plots for the inlet and outlet BCs in

figures (4.1) and (4.2) respectively, and the kde plots for both BCs in figure (4.3). As the model is a simple Gaussian we can claim from the trace plots alone that this chain is mixing and we do not use further diagnostics.

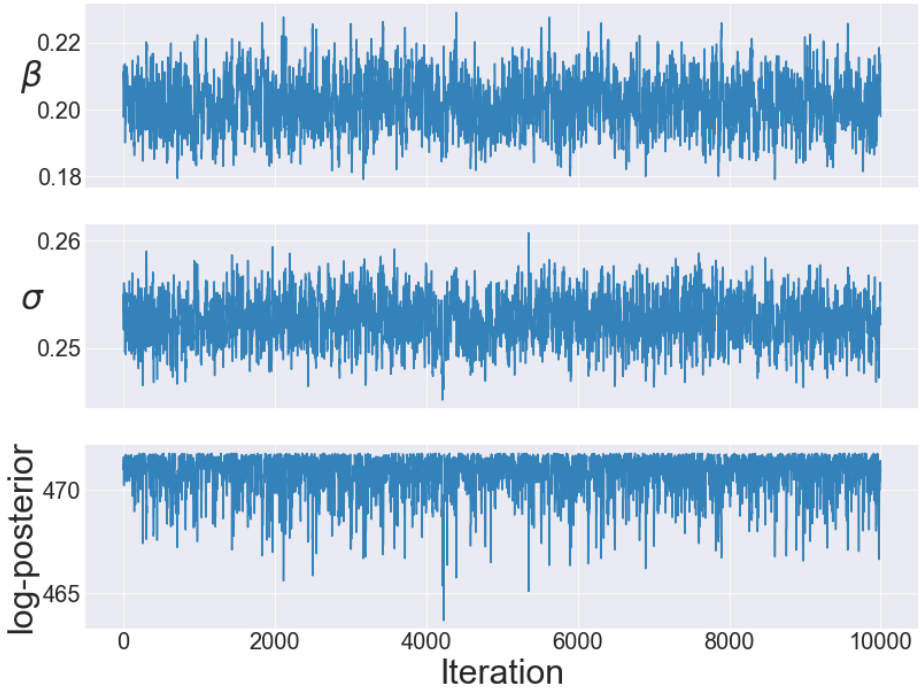

Figure 4.1: Trace plots for β and σ for the inlet BC

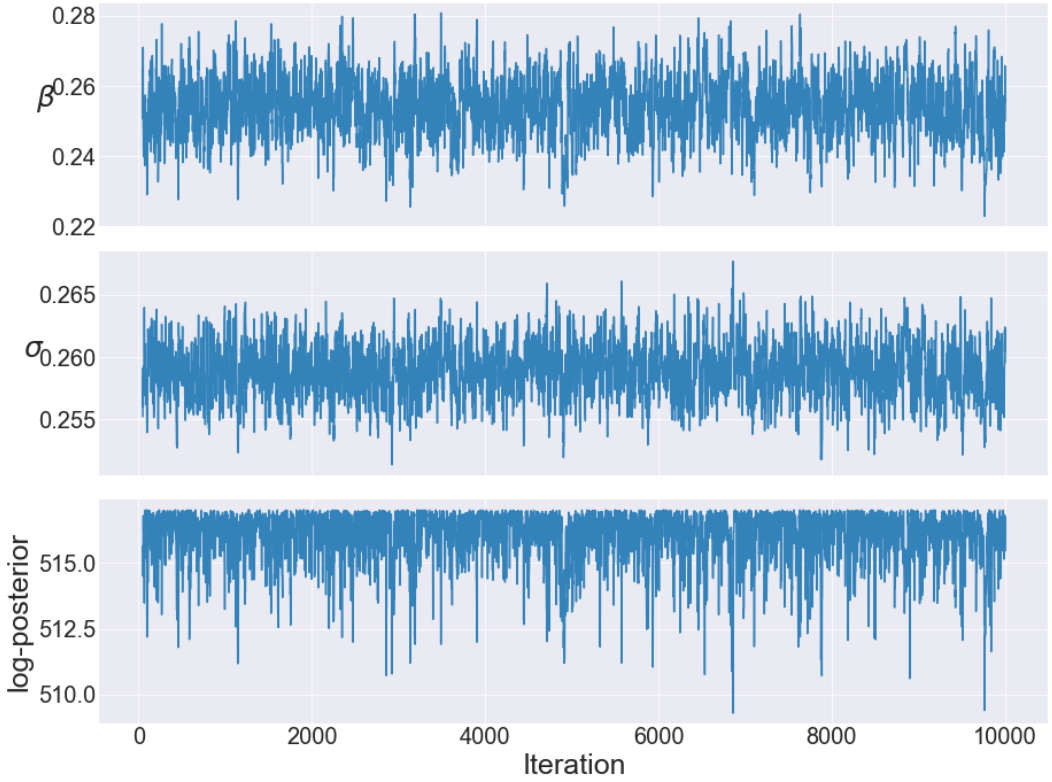

Figure 4.2: Trace plots for β and σ for the outlet BC

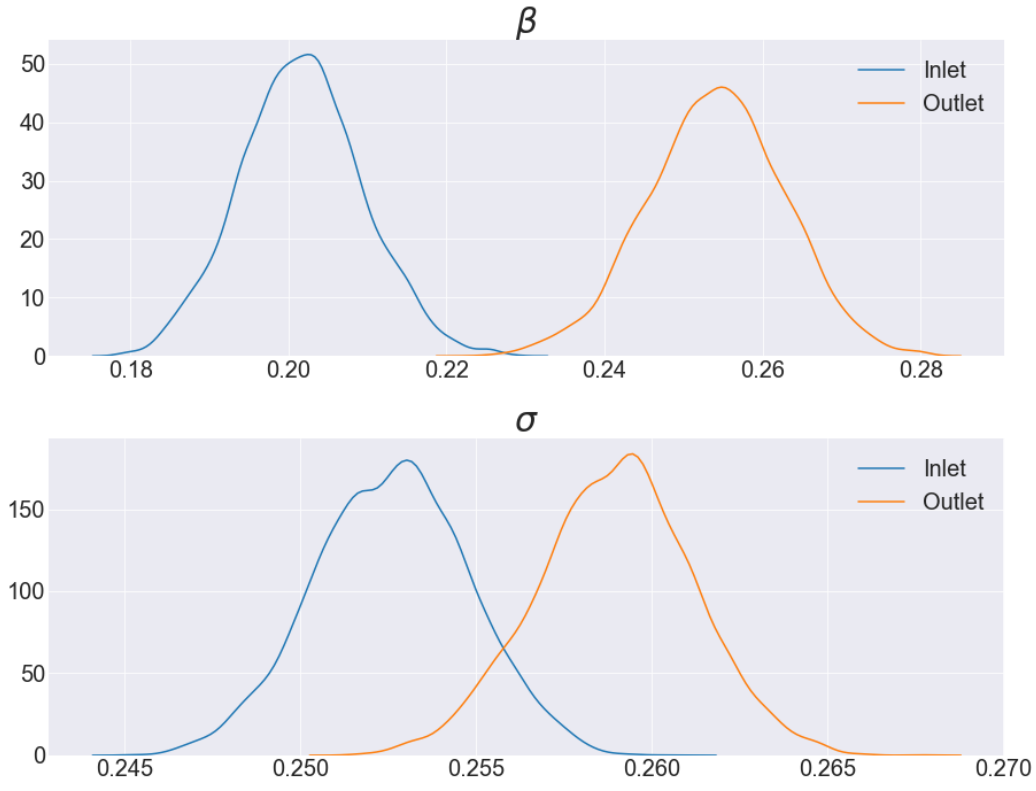

Figure 4.3: Kde plots for β and σ for the inlet and outlet BC

We notice that the variances of the σ parameter overlap, but that the variances of the β do not quite overlap. However as they are close, we estimate the two parameters for both BCs together rather than separately. The traceplots and kdeplots are shown in figures (4.4) and (4.5). The posterior mean is $\beta = 0.22$ and $\sigma = 0.256$.

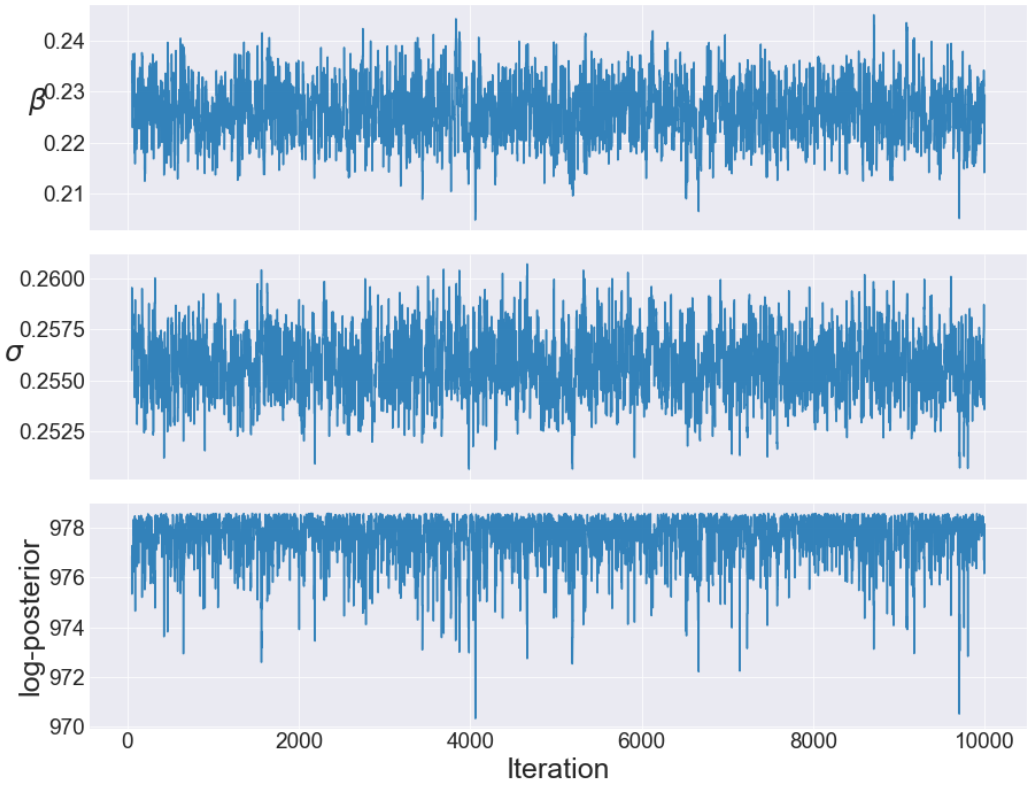

Figure 4.4: Trace plots for β and σ for both BCs

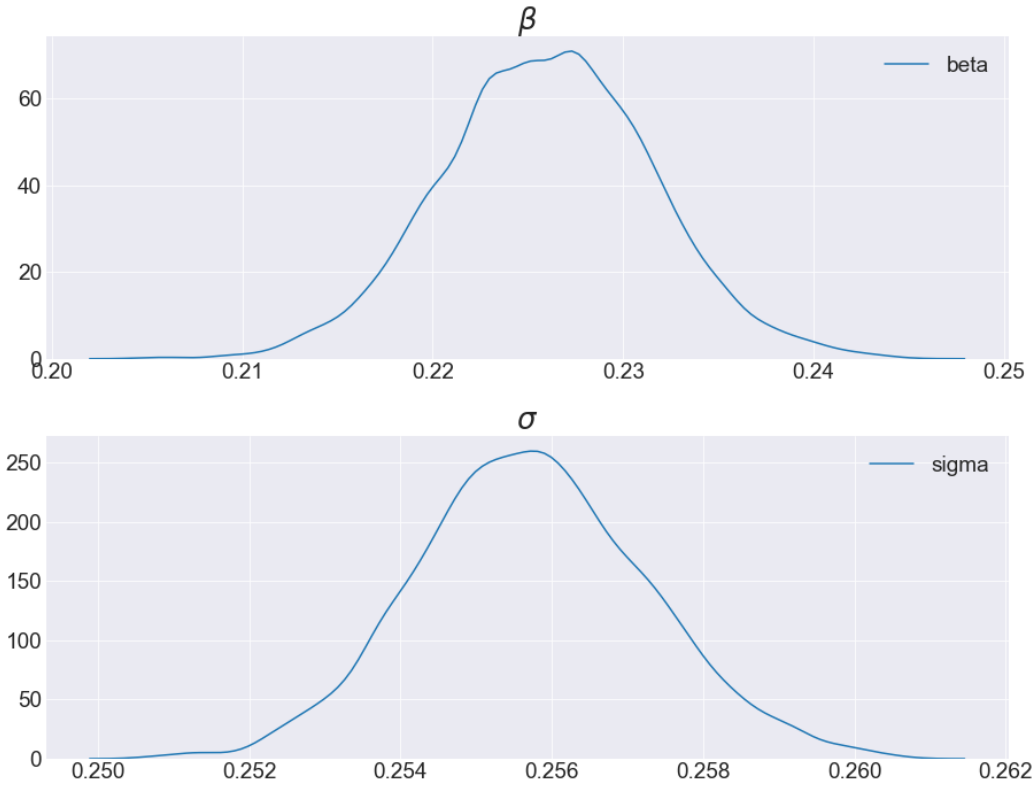

Figure 4.5: Kde plots for β and σ for both BCs

We plot in figure (4.6) and (4.7) inlet and outlet 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 BCs from data.

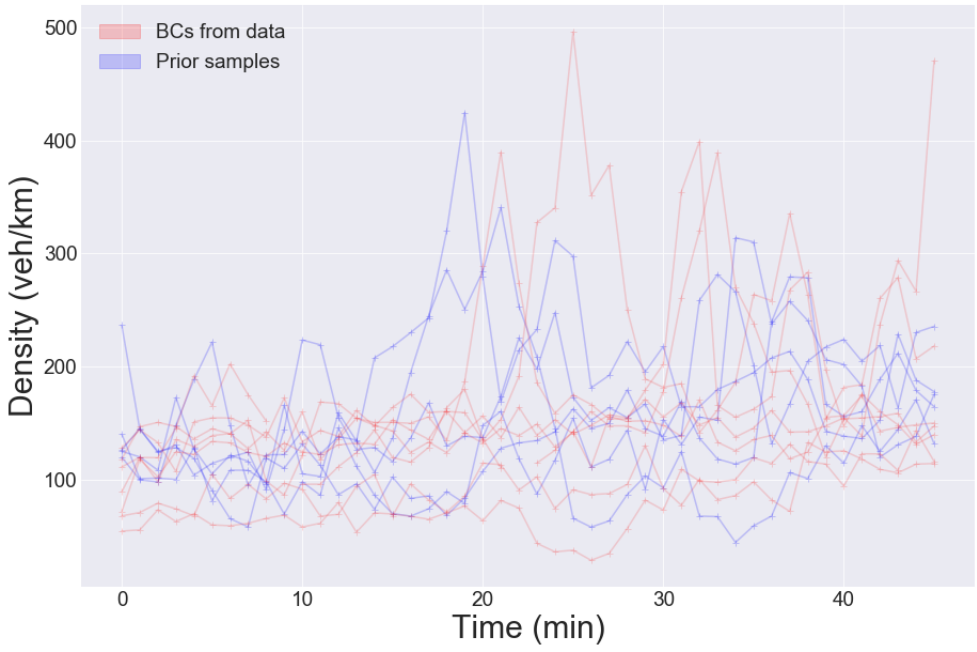

Figure 4.6: Inlet BCs from data (using density from occupancy) along with prior samples at 1 min resolution

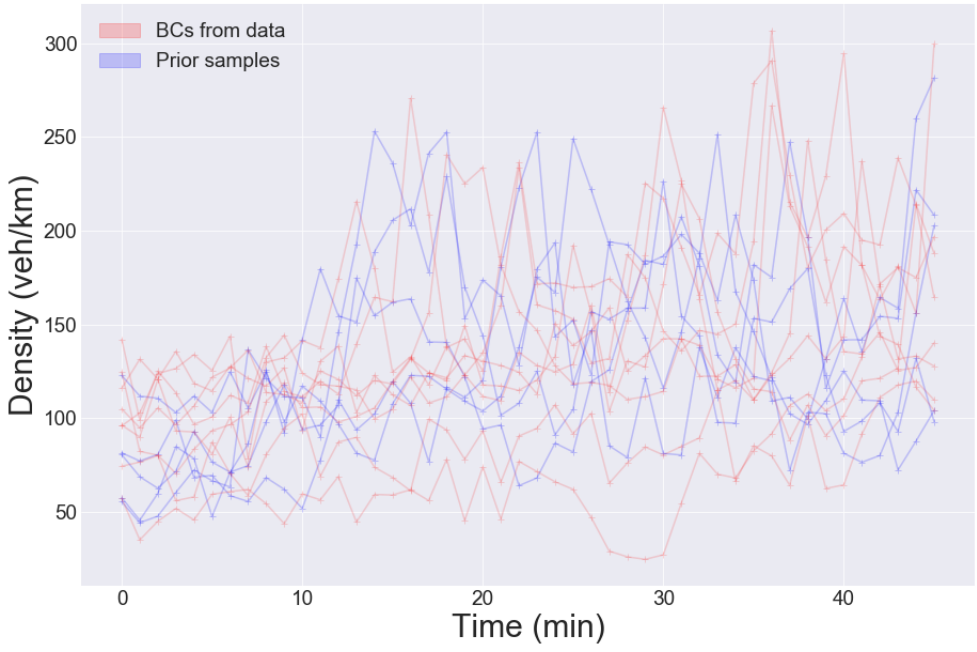

Figure 4.7: Outlet BCs from data (using density from occupancy) along with prior samples at 1 min resolution

We also plot in figure (4.8) samples from the inlet BCs at full resolution: one point every 1.5 seconds, 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 we can see from figures (4.6) and (4.7) that a resolution of 1min does not capture the details of the high density waves.

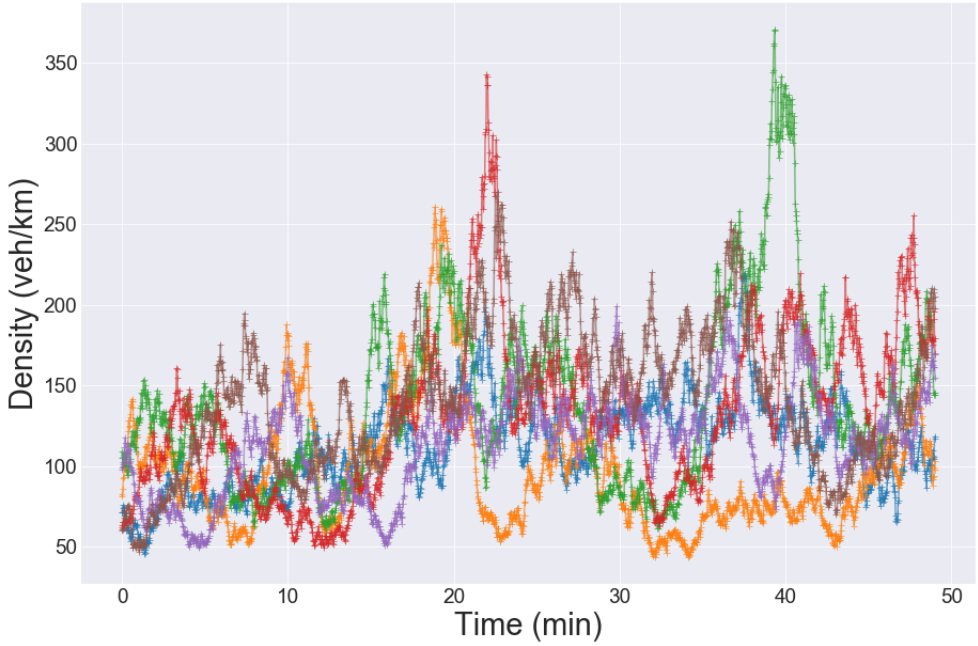

Figure 4.8: Sample from the prior for the inlet at full resolution: 1 point every 1.5 seconds

4.2 Estimating BCs with the Exponential FD: simulated data

In this section we develop the methodology to sample the conditional posterior of $\pi(BC|FD)$ parameters. As in the previous chapter, we start with simulated data to develop the methodology, and then test it on M25 motorway data. For simulated data we fix the FD to the true parameters. For real data we fix the FD (for both FDs) to be the mean of the posterior $\pi(FD|BC)$ estimated in the previous chapter. We recall in figure (4.9) the inlet and outlet sampled from the prior that were used to simulate data.

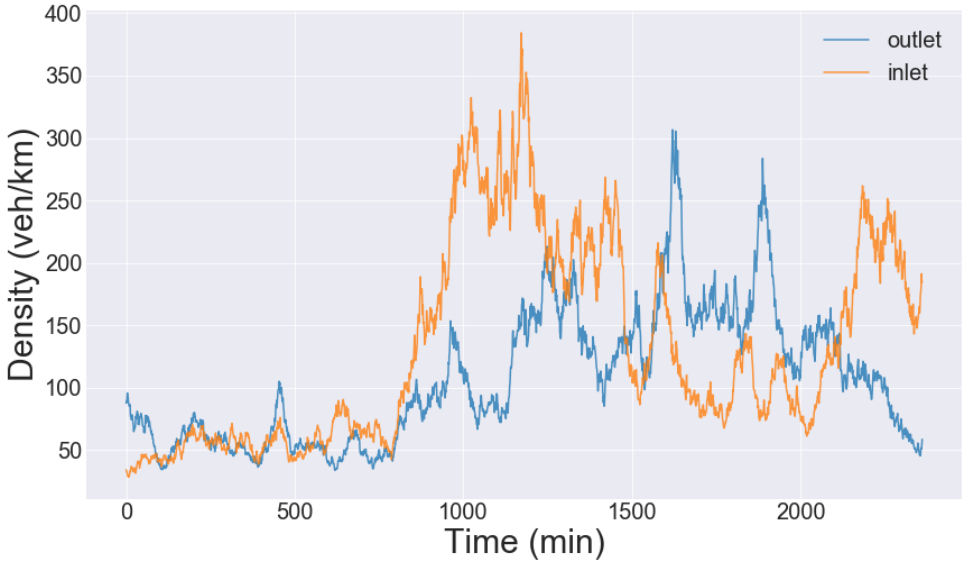

Figure 4.9: Inlet and outlet BCs used to generate simulated data. These are samples from the prior.

4.2.1 A first attempt

The simplest method to try is to run the pCN algorithm (introduced in section (1.3.3.6)) for both the BCs as we did for the initial condition in the advection equation (in chapter 2).

As proposal we use:

$$\tilde{Y}_t = \xi\,\omega + (1-\omega^2)^{1/2}(Y_t - \mu) + \mu \tag{4.2}$$

with:

- Y_t: logarithm of the BC
- $\xi \sim \mathcal{N}(0, \Lambda^{-1})$ a sample from the prior with mean 0 (OU process)
- $\omega \in [0,1]$: step size parameter to tune in the pCN algorithm
- μ: prior mean (of the logarithm of BCs)

We use a step size of $\omega = 0.02$ and run 3 chains for 20K iterations. We obtain acceptance rates of 21%, 24%, and 24% for the 3 chains, and show the samples in figures (4.10) and (4.11). We also show the trace plots for 3 typical outlet BC times in figure

(4.12). The first thing we notice is that the chains are not mixing; it is not unreasonable to suggest that several orders of magnitude more iterations would be needed to get a mixing chain, which exceeds our computational budget (20K samples takes around 12 hours to run). Inspecting the ACF plots (not shown) suggests that the delay time for the BC time points vary between 1K and 5K lags, but 20K iterations seems not enough to estimate these reliably. The trace plots in figure (4.12) show that the sampler is clearly not mixing and that the delay time for BC times such as time 2184 (54min) is likely to be much higher.

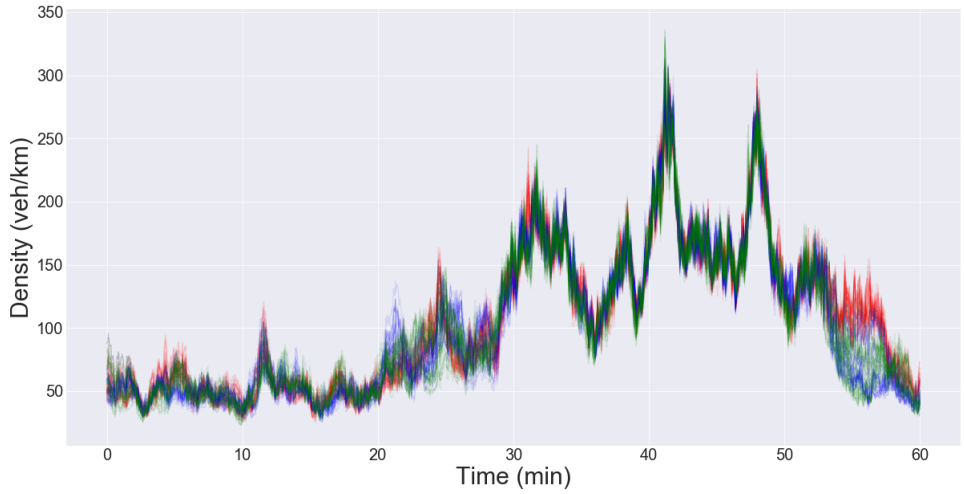

Figure 4.10: A first try of BC sampling using a pCN sampler with $\omega = 0.02$: inlet boundary conditions samples

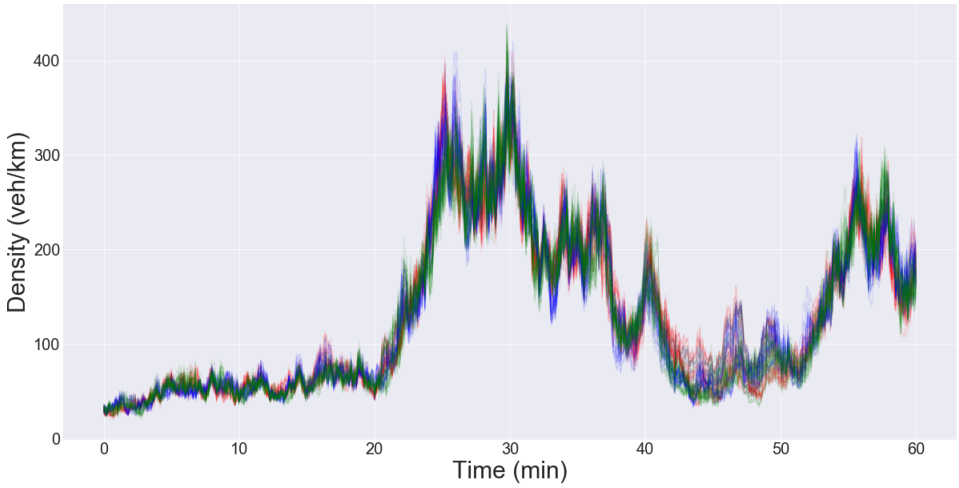

Figure 4.11: A first try of BC sampling using a pCN sampler with $\omega = 0.02$: outlet boundary conditions samples

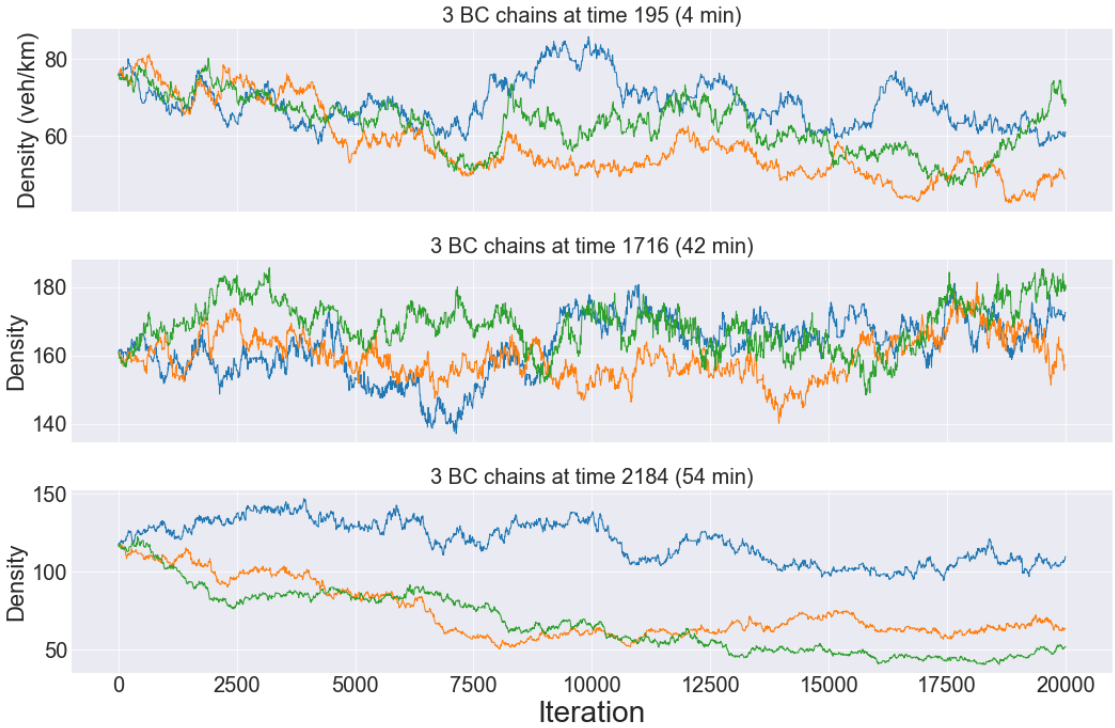

Figure 4.12: A first try of BC sampling using a pCN sampler with $\omega = 0.02$: 3 typical time points for the outlet BC

However another point to note about the chain is that the posterior seems to have different variances for different sections of the BCs. Namely, sections of low and high density seem to have low variance, while sections of medium density seem to have high

variance. If one maps density to flow using the FD, this corresponds to low flow having low variance and high flow having high variance, which is consistent with the Poisson error model.

To obtain a faster sampler, we hypothesise that these different sections in the BCs are independent. Indeed for a fixed FD wave speed, a section in the outlet BC will influence the (x,t) plane mainly within a thin strip with wave speed given by the FD. Another section in that BC will influence another strip parallel to the first section, so these two sections are therefore expected to be fairly independent. Of course the inlet and outlet BCs are correlated, as a wave leaving the outlet BC will hit the inlet BC and vice-versa. However if this assumption is approximately true we can sample these different sections separately using a Metropolis-within-Gibbs sampler. This would allow us to tune the ω parameter in the pCN algorithm for each section. It would also allow us to make bigger jumps as each section would be much shorter than the entire BC. This idea of tuning ω differently for different sections of the BC is similar to operator weighted proposals mentioned in section (1.3.3.6).

4.2.2 Gibbs blocking

4.2.2.1 Building the proposal

We first recall the formula for conditioning in a Gaussian. Let $Z \sim \mathcal{N}(\mu,\Sigma)$ and $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$. If we split the mean and covariance matrix into blocks $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, we have the conditional distribution $Y|(X=x) \sim \mathcal{N}(\mu^c_{Y|X}, \Sigma^c_{Y|X})$ with:

$$\mu^c_{Y|X} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x-\mu_1) \tag{4.3}$$

$$\Sigma^c_{Y|X} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \Sigma|\Sigma_{11} \tag{4.4}$$

With $\Sigma|\Sigma_{11}$ called the Schur complement. We can use the Schur complement to write the inverse of the covariance matrix as:

$$\Sigma^{-1} = \begin{pmatrix} \Sigma^* & -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma|\Sigma_{11})^{-1} \\ -(\Sigma|\Sigma_{11})^{-1}\Sigma_{21}\Sigma_{11}^{-1} & (\Sigma|\Sigma_{11})^{-1} \end{pmatrix} \tag{4.5}$$

With $\Sigma^* := \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma|\Sigma_{11})^{-1}\Sigma_{21}\Sigma_{11}^{-1}$. As the inverse of the covariance matrix is the precision matrix, we have:

$$\Sigma^{-1} = \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \tag{4.6}$$

Identifying terms, we obtain $\Lambda_{22} = (\Sigma|\Sigma_{11})^{-1}$ along with the identity $\Sigma_{21}\Sigma_{11}^{-1} = -\Lambda_{22}^{-1}\Lambda_{21}$.

Using these two identities we can rewrite the conditional distribution above as $Y|(X = x) \sim \mathcal{N}(\mu_{Y|X}^c, {\Lambda_{Y|X}^c}^{-1})$ with:

$$\mu_{Y|X}^c = \mu_2 - \Lambda_{22}^{-1}\Lambda_{21}(x - \mu_1) \tag{4.7}$$

$$\Lambda_{Y|X}^c = \Lambda_{22} \tag{4.8}$$

As we are interested in sampling from a section of $\log(BC)$ conditional on the rest, let $Y_t = \begin{pmatrix} Y_{t,1} & Y_{t,2} & Y_{t,3} \end{pmatrix}^T$ be the vector of the logarithm of a BC. We would like to sample from the conditional distribution $Y_{t,2}|Y_{t,1}, Y_{t,3}$. We rearrange the sections to obtain $\begin{pmatrix} Y_{t,1} & Y_{t,3} & Y_{t,2} \end{pmatrix}^T = \begin{pmatrix} Y_* & Y_{t,2} \end{pmatrix}^T$ with $Y_* := \begin{pmatrix} Y_{t,1} & Y_{t,3} \end{pmatrix}^T$.

We then write the precision matrix for Y_t:

$$\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{13} & \Lambda_{12} \\ \Lambda_{31} & \Lambda_{33} & \Lambda_{32} \\ \Lambda_{21} & \Lambda_{23} & \Lambda_{22} \end{pmatrix} \tag{4.9}$$

$$= \begin{pmatrix} \Lambda_{**} & \Lambda_{*2} \\ \Lambda_{2*} & \Lambda_{22} \end{pmatrix} \tag{4.10}$$

With $\Lambda_{**} = \begin{pmatrix} \Lambda_{11} & \Lambda_{13} \\ \Lambda_{31} & \Lambda_{33} \end{pmatrix}$, $\Lambda_{*2} = \Lambda_{2*}^T = \begin{pmatrix} \Lambda_{12} \\ \Lambda_{32} \end{pmatrix}$

Using equations (4.7) and (4.8) We can now write the conditional distribution $Y_{t,2}|Y_{t,1}, Y_{t,3}$:

$$Y_{t,2} \sim \mathcal{N}(\mu_{2|1,3}, \Lambda_{22}^{-1}) \tag{4.11}$$

With $\mu_{2|1,3} = \mu_2 - \Lambda_{22}^{-1}\Lambda_{21}(y_{t,1} - \mu_1) - \Lambda_{22}^{-1}\Lambda_{23}(y_{t,3} - \mu_3)$

We show proposals (in red) for the inlet BC for a single section between time points 20min and 30min with parameter $\omega = 0.7$ in figure (4.13).

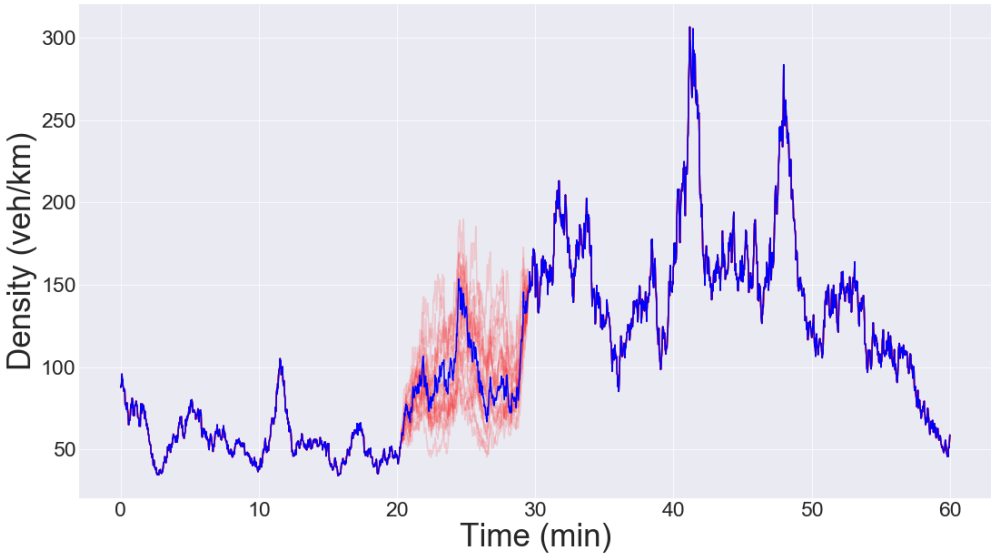

Figure 4.13: Proposals (in red) for a section of the inlet BC with $\omega = 0.7$.

4.2.2.2 MCMC using Gibbs blocking

We now try Gibbs blocking for sampling the BCs. We use the following guidelines to choose the Gibbs blocks (based on experimentation):

1. Choose a block size between 10min and 30min. Choosing a size that is too big will result in a low acceptance rate, while choosing too small results in too many blocks (and hence slow mixing for the chain overall).

2. Choose a section where the density - and hence the variance - is fairly similar. Tune ω based on the variance of the block.

3. Check the acceptance rate and aim for around 20%

We run MCMC for 50K iterations using 16 blocks (given in the appendix in table (B.1) along with the acceptance rates for each block) and obtain BC samples in figures (4.14) and (4.15). We show the trace plots for a portion of the outlet BC in figure (4.16). We note that the chains mix a lot better than in the previous section, but that

we still do not have perfect mixing. Using delay times to compare this chain with the previous chain is not practical as it was not possible to reliable estimate them for the previous section (as discussed). However comparing the trace plots alone shows clear improvement. By inspecting the BC samples and the trace plots for this run we notice :

- The variance does indeed vary between the sections: note for example the high variance in the outlet at minutes $20-30$ compared to the peak of the high density waves. Tuning ω for each section allows more efficient exploration of the posterior.

- We can see that the posterior seems to exhibit bimodality, for example around minutes 10 and 15 in the outlet, and minute 40 in the inlet. By using the FD to map density to flow, we notice that the two values of density in the bimodality seem to correspond to similar values of flow. This would explain why both values of density are similarly likely. This idea will be discussed in more detail in section (4.2.3.2).

- Some of the time points still mix very slowly even after tuning the blocks. In particular the sections which exhibit bimodality have trouble efficiently jumping between the modes. Indeed in figure (4.16) we notice that at time 351 (minute 8) a chain stayed in the top mode while the other ones stayed in the bottom with rarely any mixing. This is reflected in the ACF plots (not shown) where we observe that the delay time for BC times ranges from 200 to over 10K.

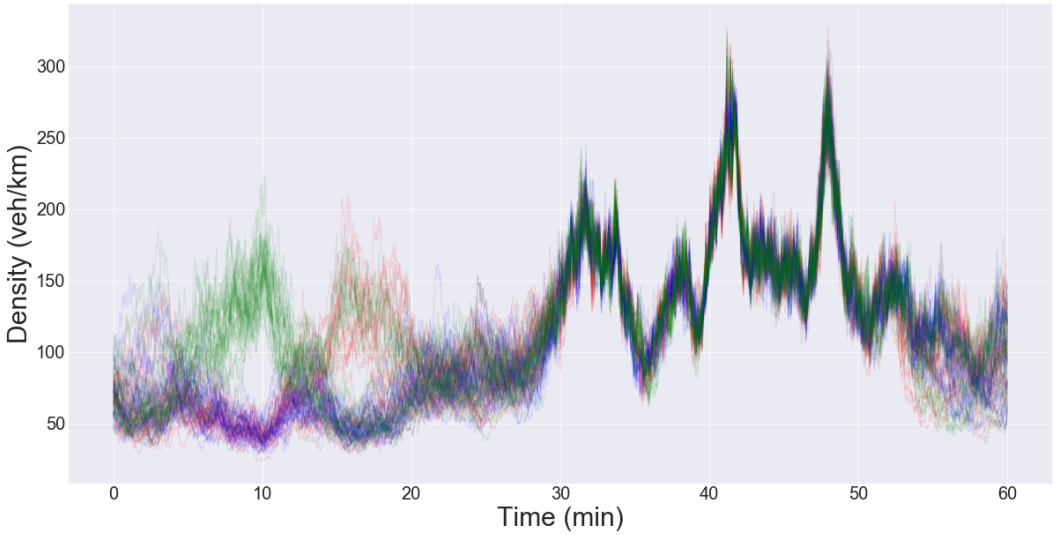

Figure 4.14: Exponential FD for simulated data: outlet BC samples with Gibbs blocks, with the 3 colours denoting the 3 chains. We notice that not all the chains explore the bimodality around minutes $[5-12]$ and $[12-20]$

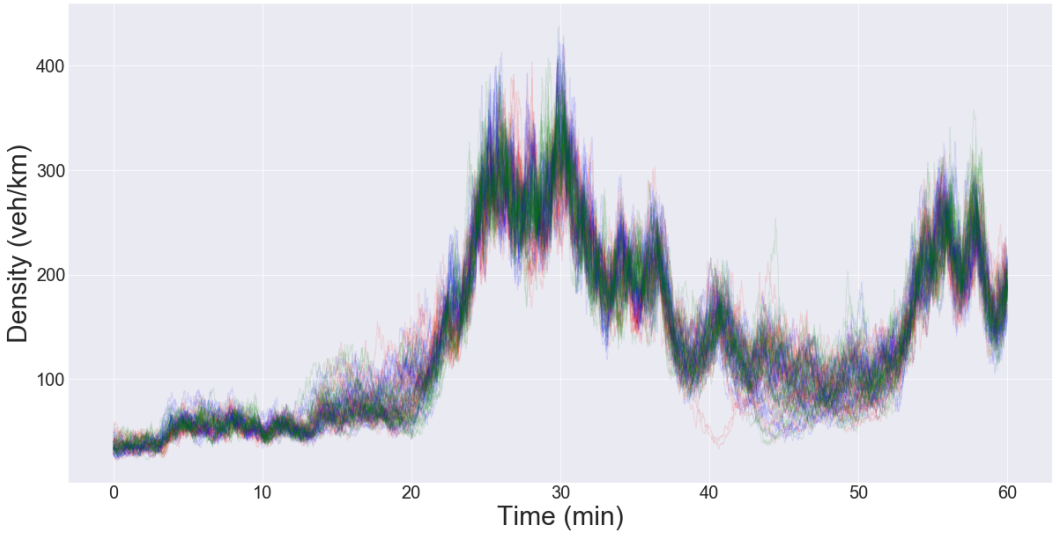

Figure 4.15: Exponential FD for simulated data: inlet BC samples with Gibbs blocks, with the 3 colours denoting the 3 chains. We note the bimodality around minute 40 that seems to only have been explored by a single chain.

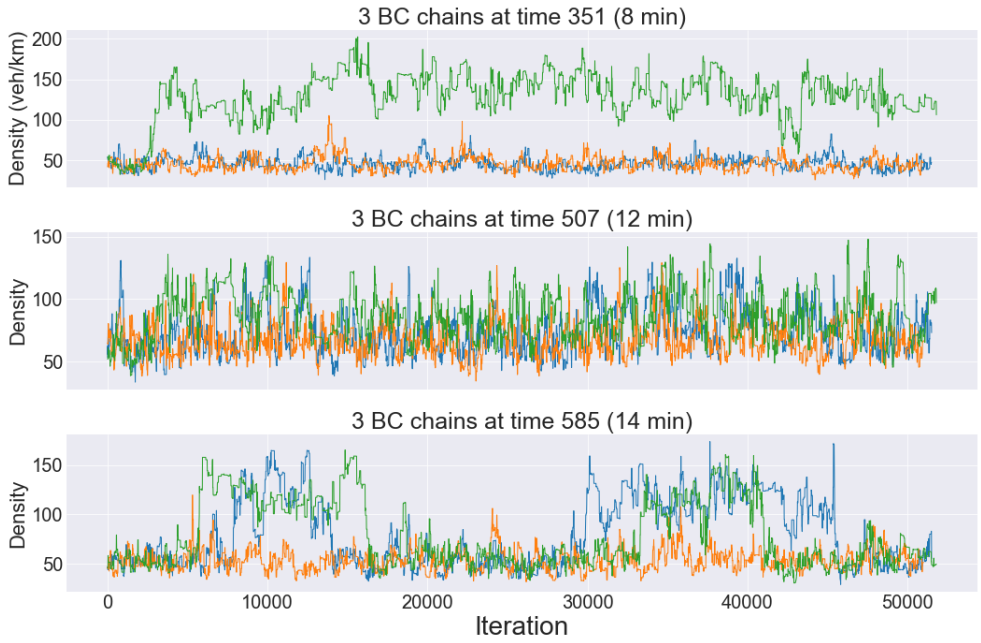

Figure 4.16: Exponential FD for simulated data: outlet BC traceplots for 3 typical times. Some of times seem to mix well (such at minute 12 shown here), while others do not mix very well (such as minutes 8 and 14 shown here). These two non-mixing times (minutes 8 and 14) can be seen in the samples in figure (4.14)

4.2.3 Parallel tempering

The use of Gibbs blocking has significantly helped the chain to explore the posterior but has also revealed the presence of bimodality, which is one of the causes of the poor mixing. Furthermore, there may be some more sections of the BCs that exhibit bimodality that the Gibbs blocks were never able to jump to. We also note that some other time points in the BCs (that do not exhibit bimodality) also mix very slowly; this could perhaps be due to strong correlations between the inlet and outlet.

To remedy the problem of bimodality (and the conjectured problem of strong correlations) we employ parallel tempering (PT), which we introduced in Chapter 2.

4.2.3.1 Tuning

We raise the likelihood to the power β_{temp} with $\beta_{temp} \in [0,1]$, and find that a value of $\beta_{temp} = 0.5$ allows the chain to mix well. Following the iterative tuning procedure outlined in [1], we then find another temperature ($\beta_{temp} = 0.63$) such that the swap acceptance rate is approximately 23%. We then fit a geometric schedule between these two values ($0.5 * 1.26 = 0.63$) and take powers of 1.26 to obtain 4 temperature spacings: `[0.5, 0.63, 0.8, 1]`. We then check that the acceptance rate for temperature

swaps is around 23% for these spacings.

We run a deterministic scan (Metropolis-within-) Gibbs sampler for the within-temperature and between-temperature proposals, and choose 5 within-temperature proposals for every swap proposal. This is based on the empirical tests performed for the posterior arising from the advection equation which has a similar structure (a functional parameter for the initial condition as well as a parameter in the flux function). We choose slightly more within-temperature proposals than for the advection equation as there are more parameters to sample from. Indeed there are a dozen Gibbs blocks, so the sampler need to be able to explore all of these components.

4.2.3.2 MCMC results

We run 3 chains for 200K iterations with blocks given in the appendix in section (B.3.2), and plot thinned samples for both BCs (for the untempered chain) in figures (4.17) and (4.18). We can see the bimodality that we found in ealier runs, but the 3 colours (one for each chain) seem to mix better this time. Indeed we show the trace-plots for a few typical times of the outlet and inlet BCs in figures (4.19) and (4.20) respectively and see that the chains jump between the modes. The higher temperature chains made these jumps more easily, and this information was propagated up to the untempered chain. We also point out the very clear (but rare) bimodality at time point 1638 (ie: 40 minutes), which we can spot in figure (4.18).

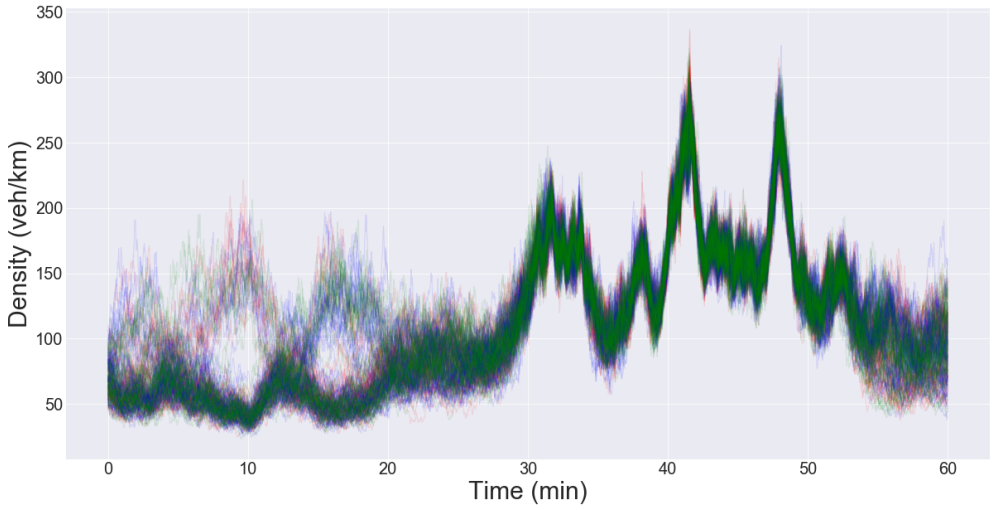

Figure 4.17: Exponential FD for simulated data using parallel tempering: outlet BC samples

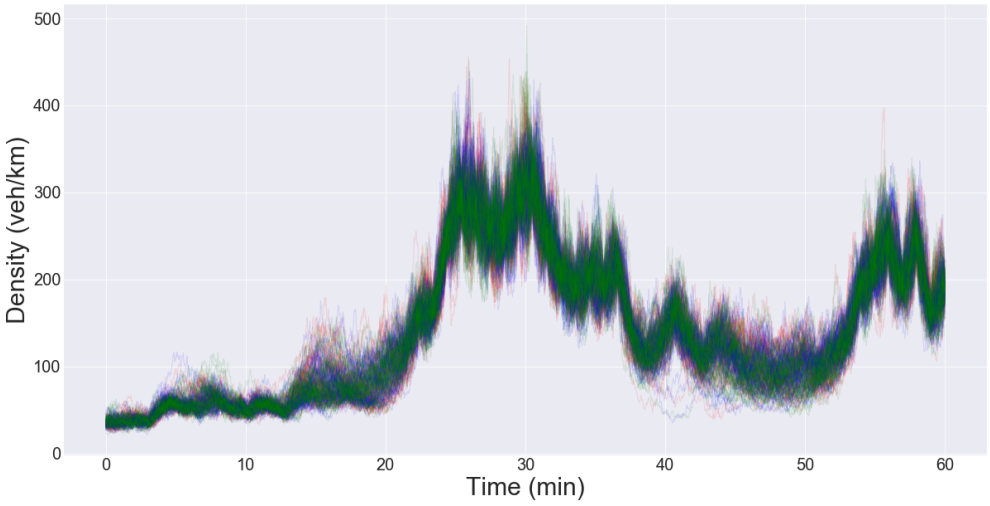

Figure 4.18: Exponential FD for simulated data using parallel tempering: inlet BC samples

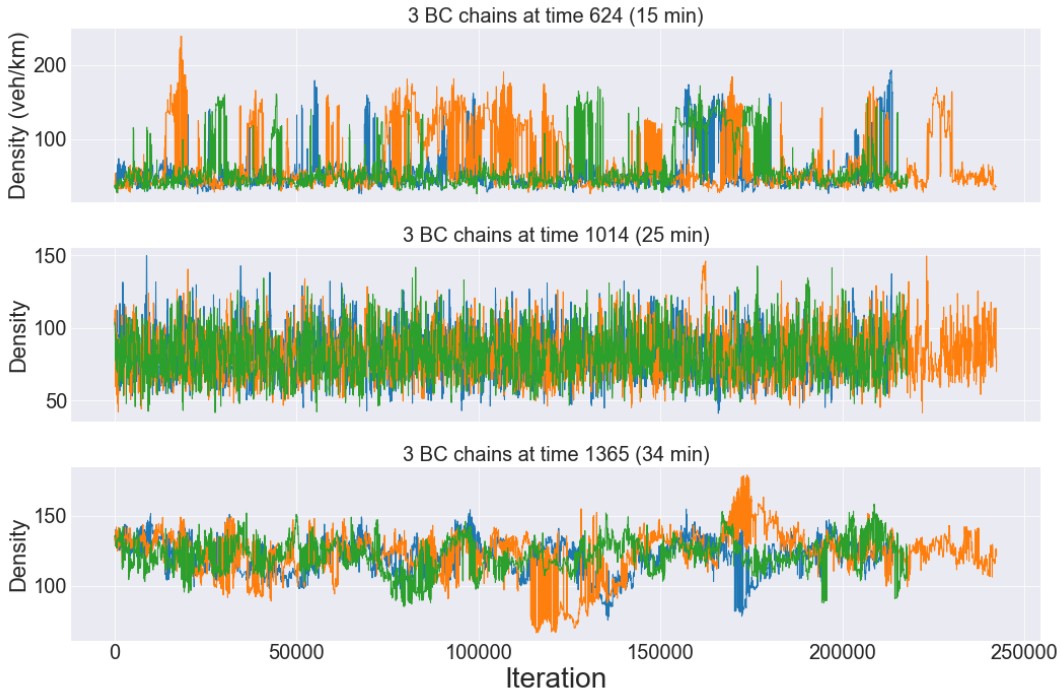

Figure 4.19: Exponential FD for simulated data using parallel tempering: outlet BC trace plots for three typical time points.

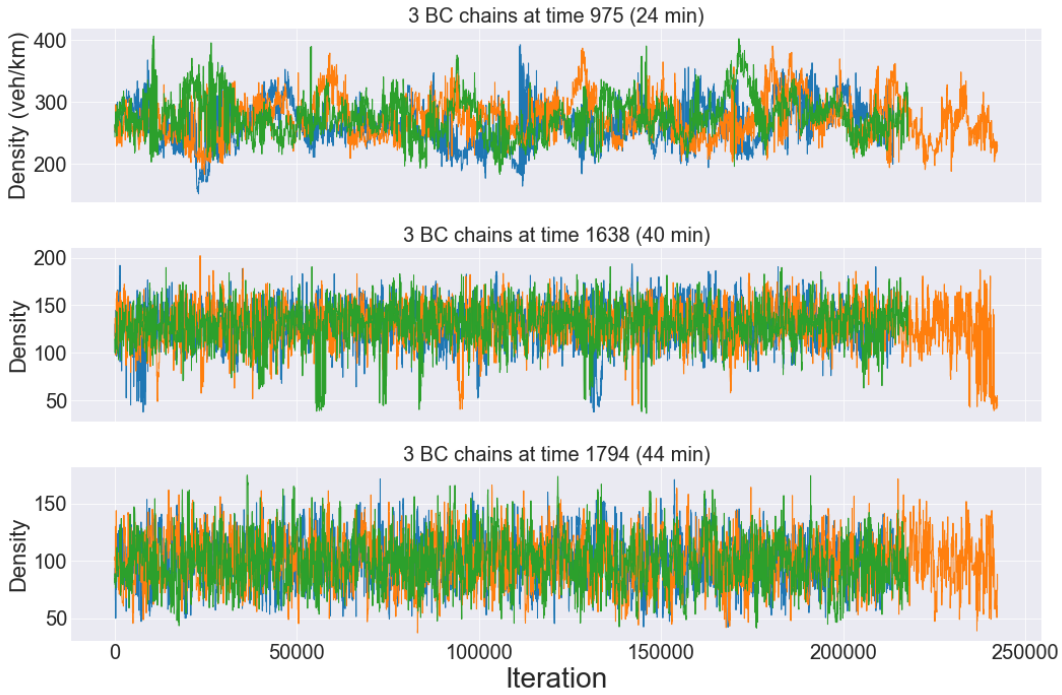

Figure 4.20: Exponential FD for simulated data using parallel tempering: inlet BC trace plots plots for three typical time points.

To explain the bimodality present in figures (4.17) and (4.18), 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. To illustrate this, we plot in figure (4.21) the Exponential FD with the true parameters values from simulated data, and we plot two vertical lines for two values of density ($\rho_1 = 40$ and $\rho_2 = 150$) 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 bimodality, 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. However, the bimodality observed in this posterior and in the later ones studies in this thesis approximately fit this explanation.

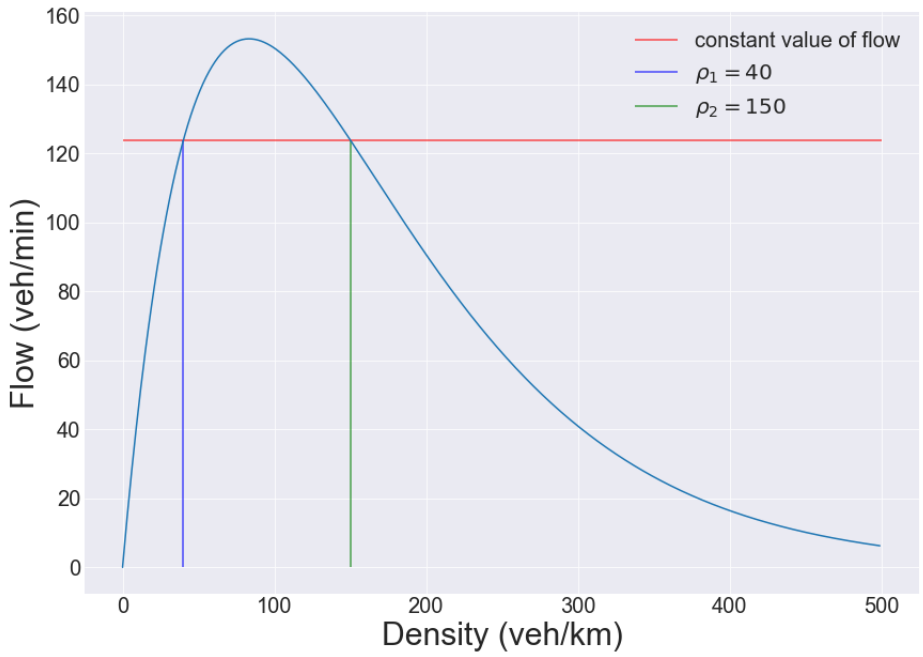

Figure 4.21: Exponential FD with true parameters from simulated data. The two vertical lines correspond to two values of density ($\rho_1 = 40$ and $\rho_2 = 150$) that map to the same value of flow. As the likelihood is built from flow, these two values of density are equally likely and therefore the posterior exhibits bimodality.

We also plot the 'R.hat' diagnostic for the outlet and inlet BCs in figures (4.22) and (4.23) respectively, and find that all the time points have a value below the prescribed limit of 1.1 (as discussed in [9]). From the ACF plots (not shown) we find that the delay time is on average 1K-2K, with some extremes of 5K for the slower mixing BC times. We check in figure (4.24) that the output of LWR with posterior mean BCs in the $(x-t)$ plane is similar to the output with the true parameters (in Chapter 3, figure (3.8)).

Finally, we also run chains with random starting points (centered around the mode) to see if they converge back to the mode (indeed they do). This diagnostic is named the 'multistart heuristic' and can be helpful in finding modes in the posterior that would not be found by running the chain for a long time (see Chapter 1 in [9] for a criticism of this heuristic).

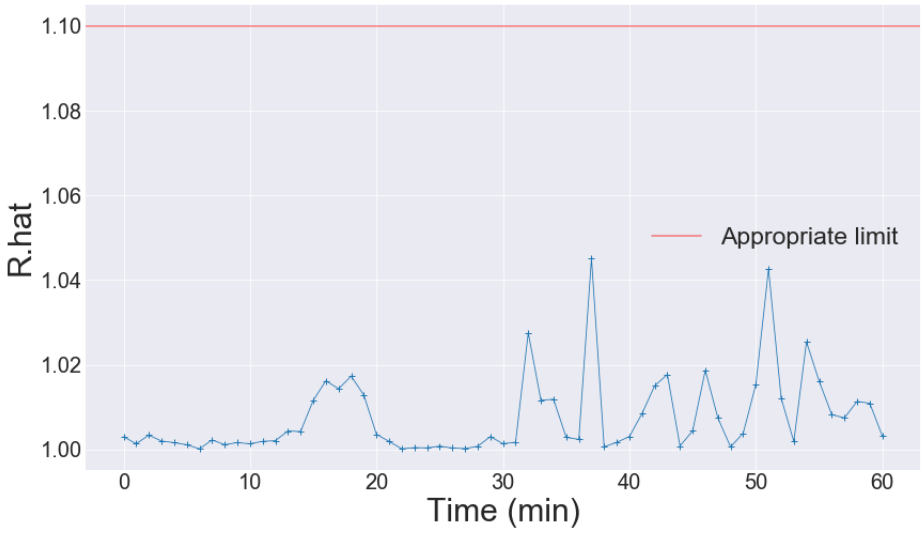

Figure 4.22: Exponential FD for simulated data using parallel tempering: R.hat for the outlet BC

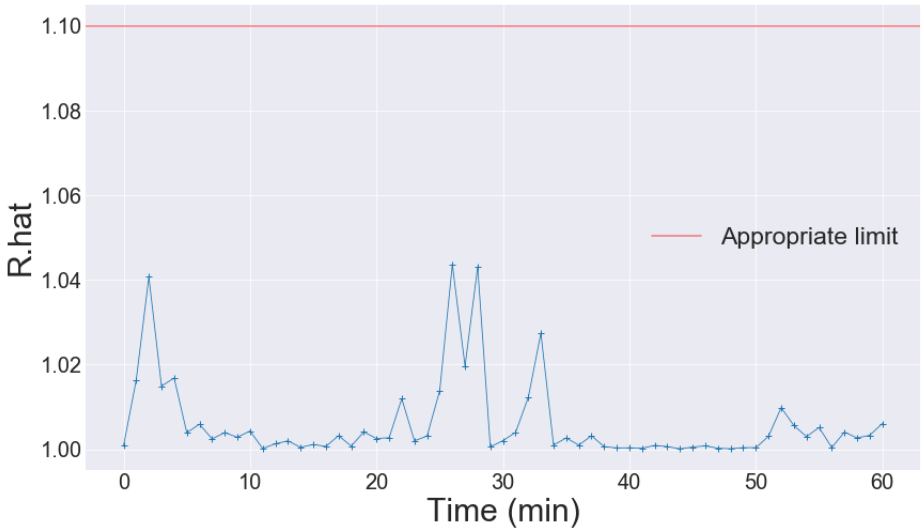

Figure 4.23: Exponential FD for simulated data using parallel tempering: R.hat for the inlet BC

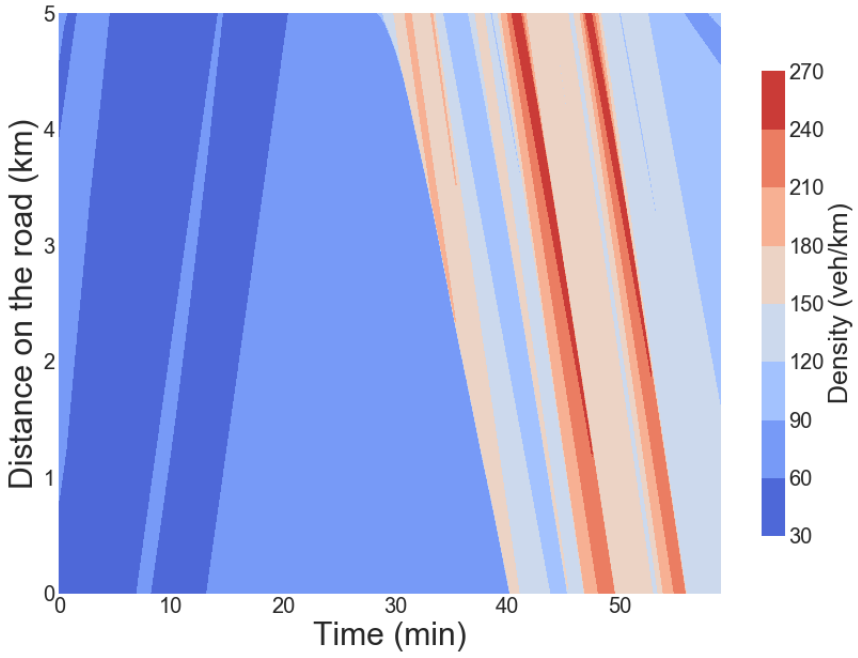

Figure 4.24: Exponential FD for simulated data: $x-t$ plane using the posterior mean BCs

4.3 Estimating BCs with the Exponential FD: M25 data

In this section we use the same methodology as above but this time with M25 data (the same as in the previous chapter). We fix the FD to the mean of the posterior found in the previous chapter (namely $\begin{pmatrix}\alpha & \beta\end{pmatrix}^T = \begin{pmatrix}2.227 & 0.00686\end{pmatrix}^T$) and sample the BCs.

We tune the Gibbs blocks as in the previous section and find similar problems: the chain does not quite mix. Some sections mix very slowly, and other exhibit bimodality where the different chains struggle to jump between the modes. We then tune parallel tempering (PT) in the same way as above and find a schedule with 3 temperatures `[0.58, 0.76, 1]`. We give the tuning of the blocks in section (B.3.3). The M25 data is for 49 minutes rather than 60 minutes of simulated data, so the length of each BCs is shorter; this may be why the sampler can accept bigger temperature jumps more often.

After tuning the PT sampler, we obtain outlet and inlet samples in figures (4.25) and (4.26). We see a portion of trace plots for the outlet and inlet BCs for 3 typical times in figures (4.27) and (4.28) respectively. We show the 'R.hat' diagnostics in figure (4.29) and (4.30) which is always below the prescribed value of 1.1. We show in figures (4.31) and (4.32) the autocorrelation function plotted for two typical times of

a BC chain. The delay times for the BC times are usually around 1K to 2K with some extremes at 500 and at 3K.

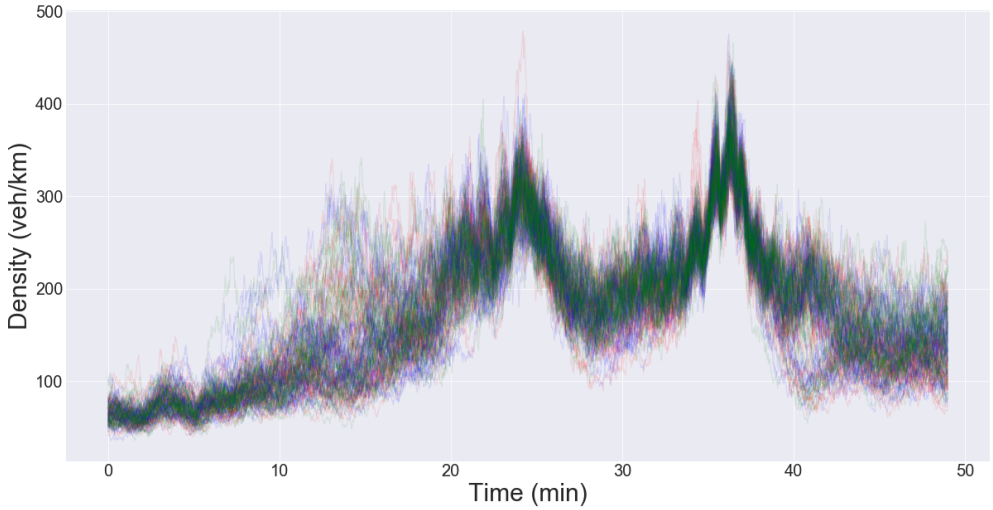

Figure 4.25: Exponential FD for M25 data using parallel tempering: outlet BC samples

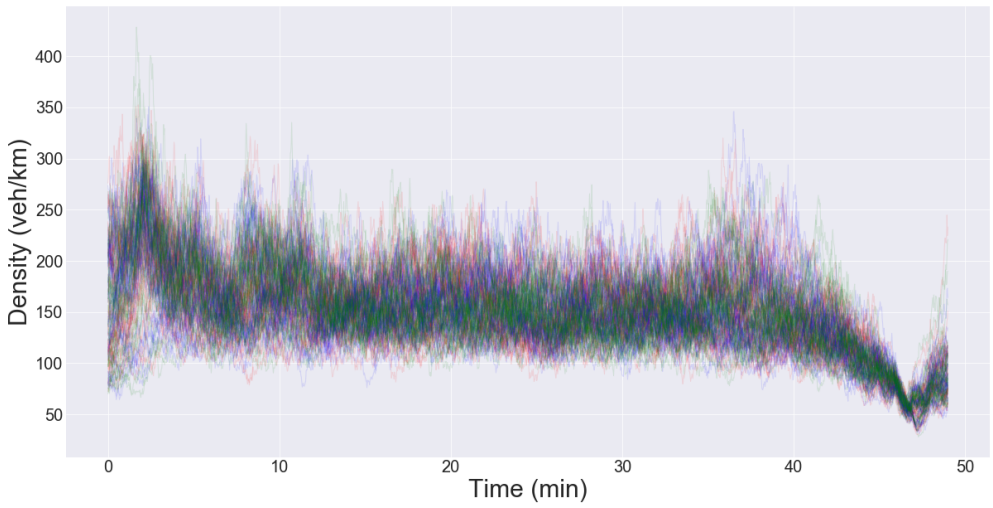

Figure 4.26: Exponential FD for M25 data using parallel tempering: inlet BC samples

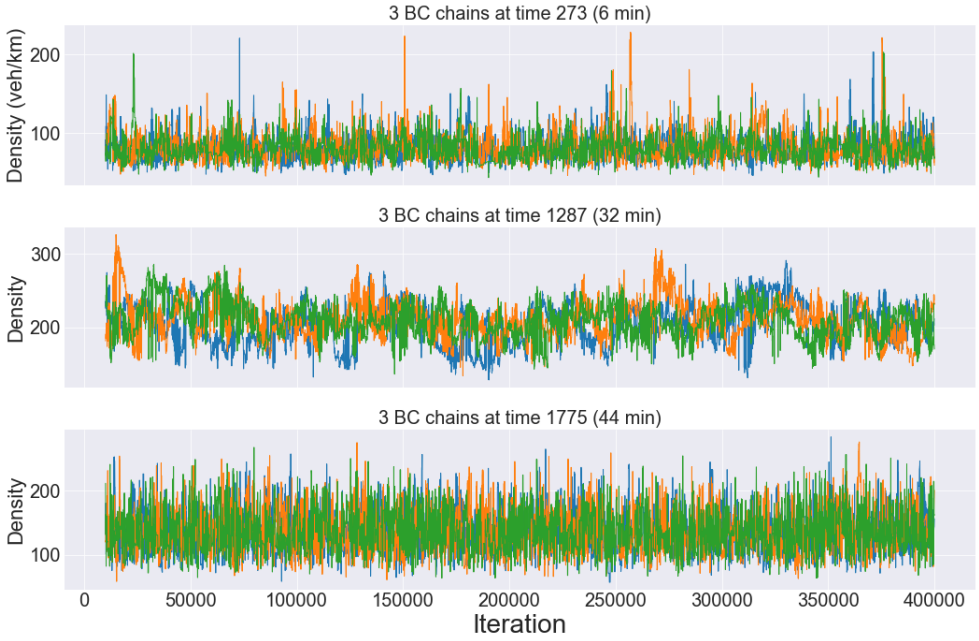

Figure 4.27: Exponential FD for M25 data using parallel tempering: outlet BC traceplots for 3 typical times.

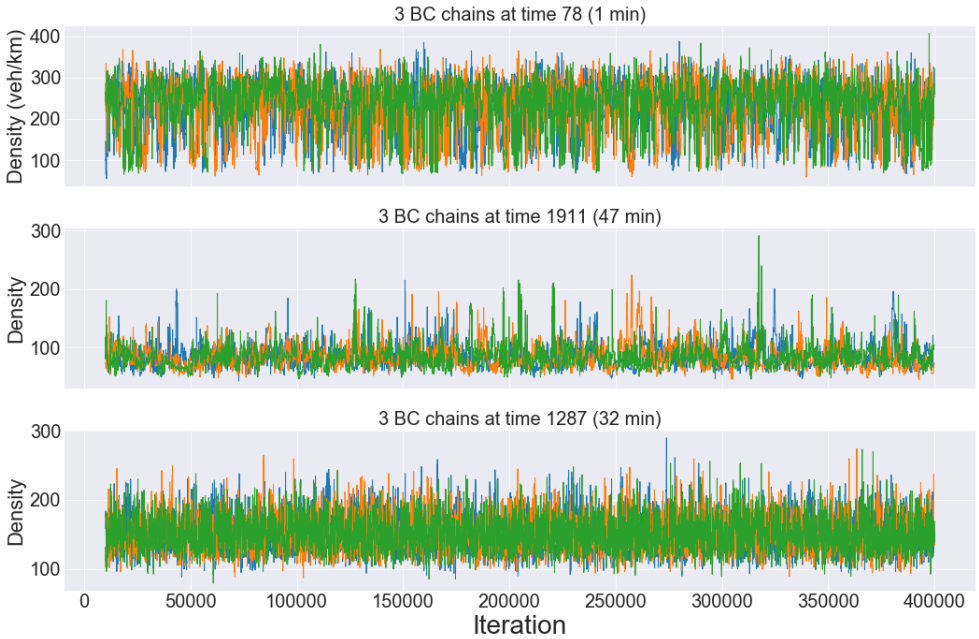

Figure 4.28: Exponential FD for M25 data using parallel tempering: inlet BC traceplots for 3 typical times.

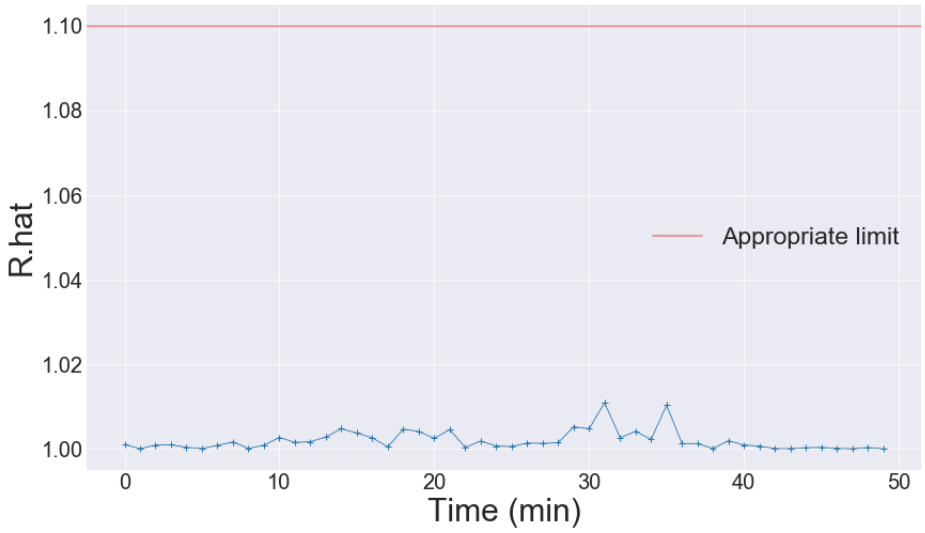

Figure 4.29: R.hat for the outlet BC. We use a PT sampler for Exponential FD for M25 data

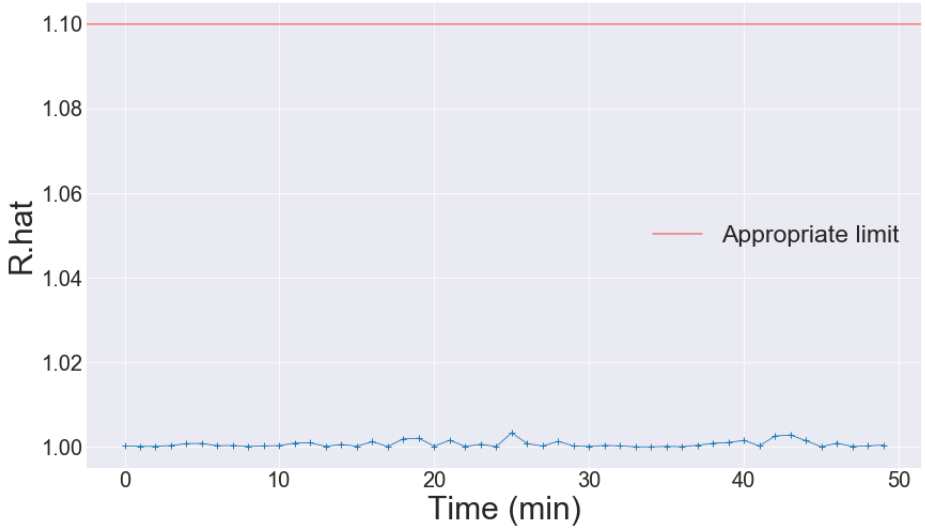

Figure 4.30: R.hat for the inlet BC. We use a PT sampler for Exponential FD for M25 data

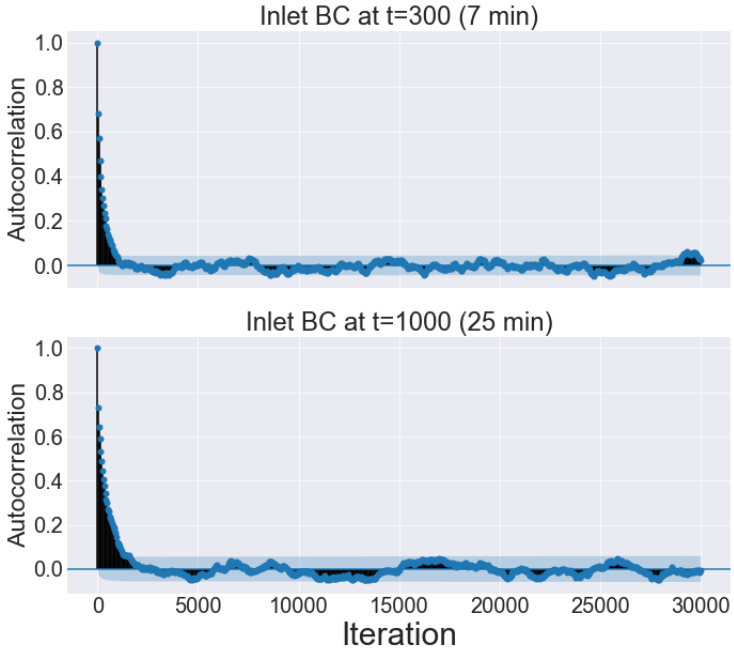

Figure 4.31: ACF plotted for two typical times of the inlet BC. We use a PT sampler for Exponential FD for M25 data

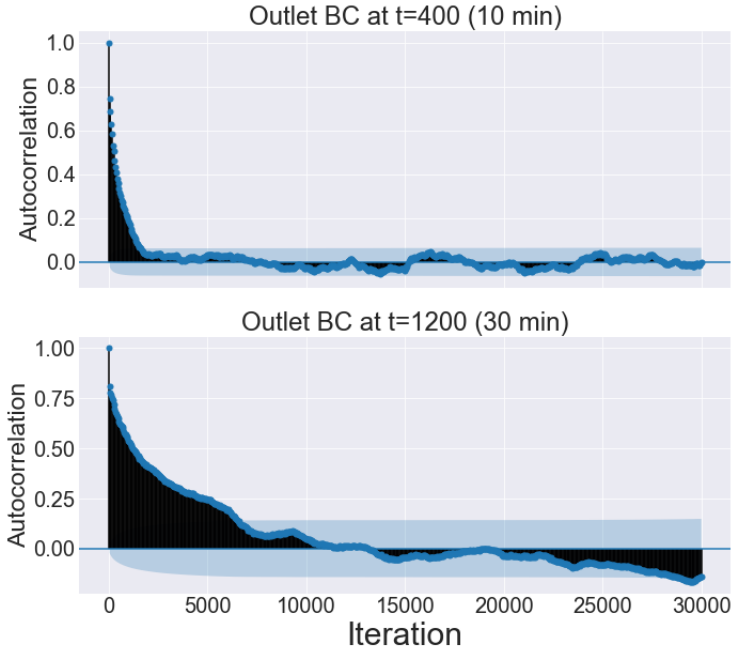

Figure 4.32: ACF plotted for two typical times of the outlet BC. We use a PT sampler for Exponential FD for M25 data

Finally, we plot the output of LWR in the $x-t$ plane using the posterior mean BCs in figure (4.33). We notice that although the waves still do not completely cross the

domain, they have progressed further than in the previous chapter (see figures (3.18) and (3.19)) where we estimated the FD parameters.

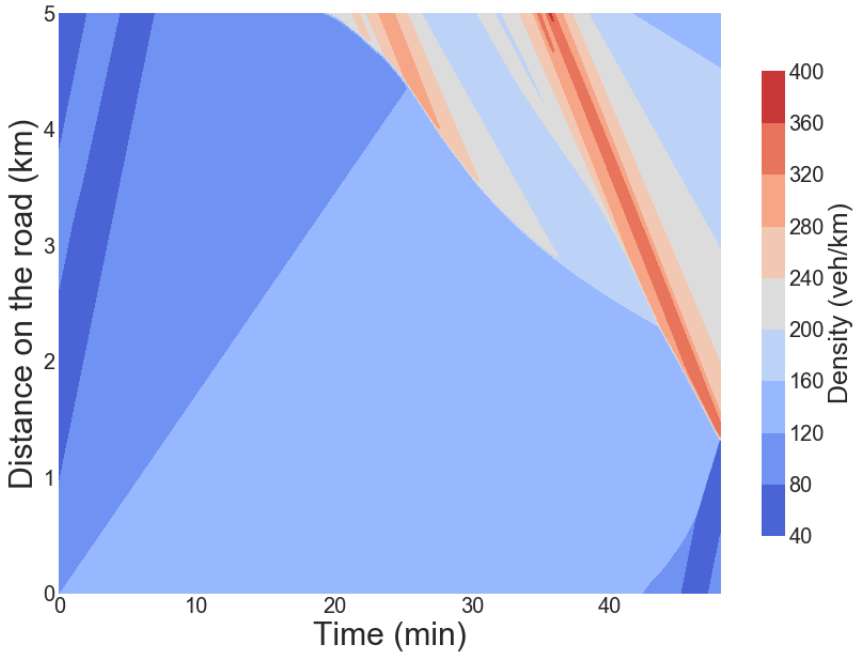

Figure 4.33: Exponential FD for M25 data: XT plane using the posterior mean BCs

4.4 Estimating BCs with del Castillo's FD: simulated data

We now apply the methodology to the BCs in LWR with del Castillo's FD. Here the process is similar to the case of the Exponential FD: Gibbs blocking greatly improves mixing compared to updating the entire BCs using pCN but we still observe slow mixing as well as chains not jumping between modes. We then try the same temperature schedule as for the Exponential FD with simulated data (`[0.5, 0.63, 0.8, 1]`) which has a reasonable acceptance rate for the between-temperature moves (around 14%). We give the Gibbs blocks in section (B.4.1).

We run 200K iterations and plot the outlet and inlet samples in figures (4.34) and (4.35). The trace plots for both BCs seem to suggets good mixing; we show 3 typical times for both BCs in figures (4.36) and (4.37). The R.hat values (not shown) are all below the prescribed limit of 1.1 for both BCs. We show the ACF plots for two typical BC times (one for each BC) in figure (4.38); the delay times for the BC times are

usually around 2K to 3K for this run.

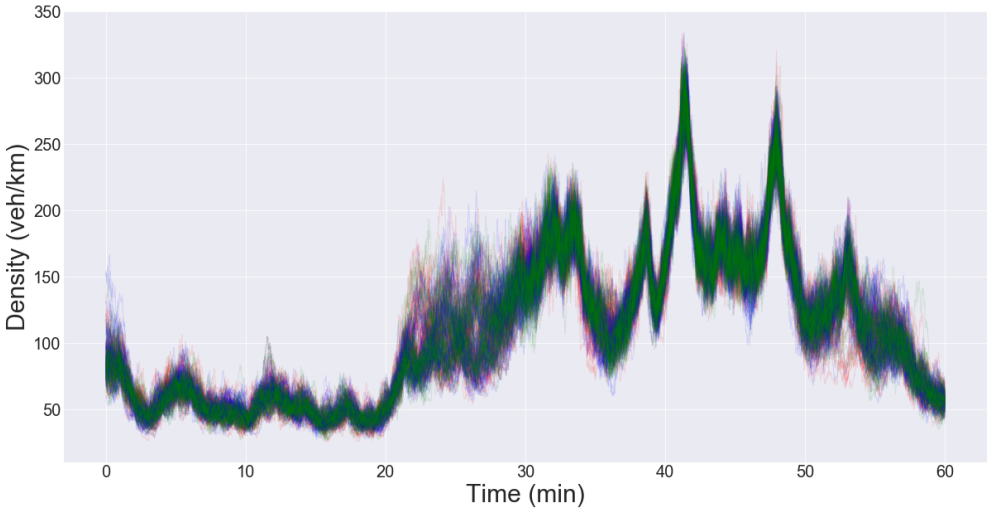

Figure 4.34: Outlet BC samples for del Castillo's FD for simulated data using parallel tempering.

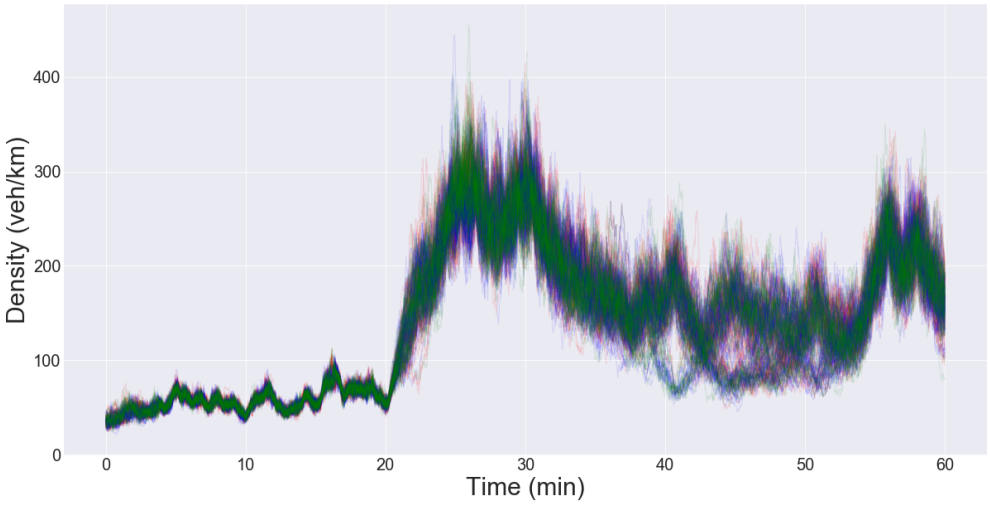

Figure 4.35: Inlet samples for del Castillo's FD for simulated data using parallel tempering.

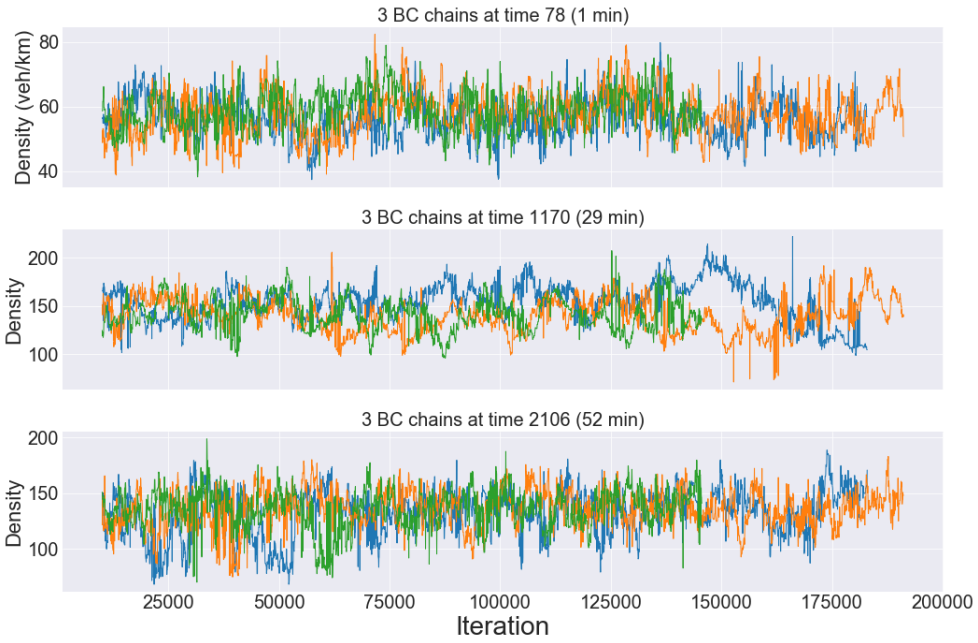

Figure 4.36: Outlet BC trace plots for 3 typical times. The 3 chains are for del Castillo's FD for simulated data using parallel tempering.

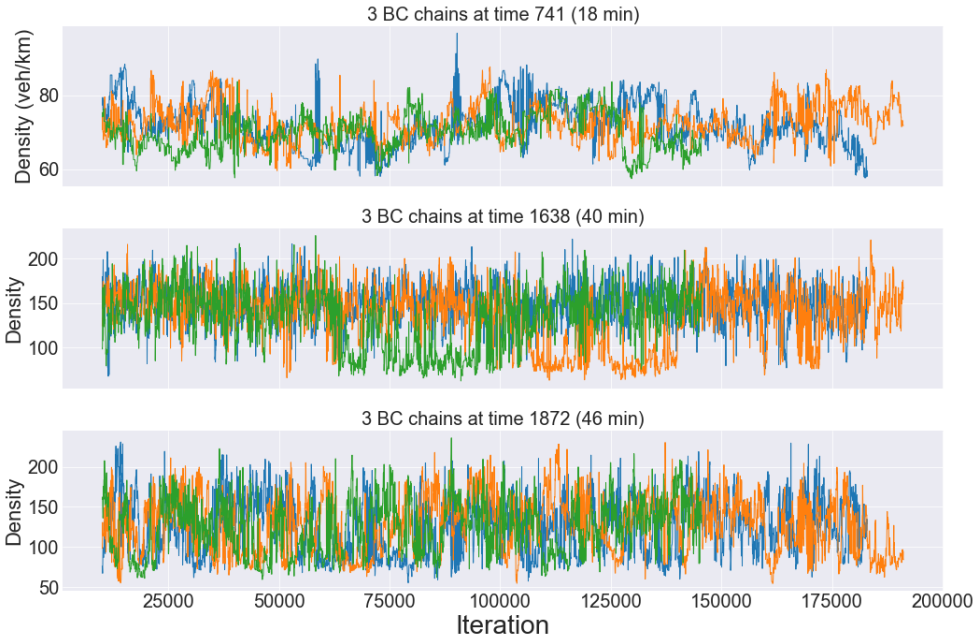

Figure 4.37: Inlet BC trace plots for 3 typical times. The 3 chains are for del Castillo's FD for simulated data using parallel tempering.

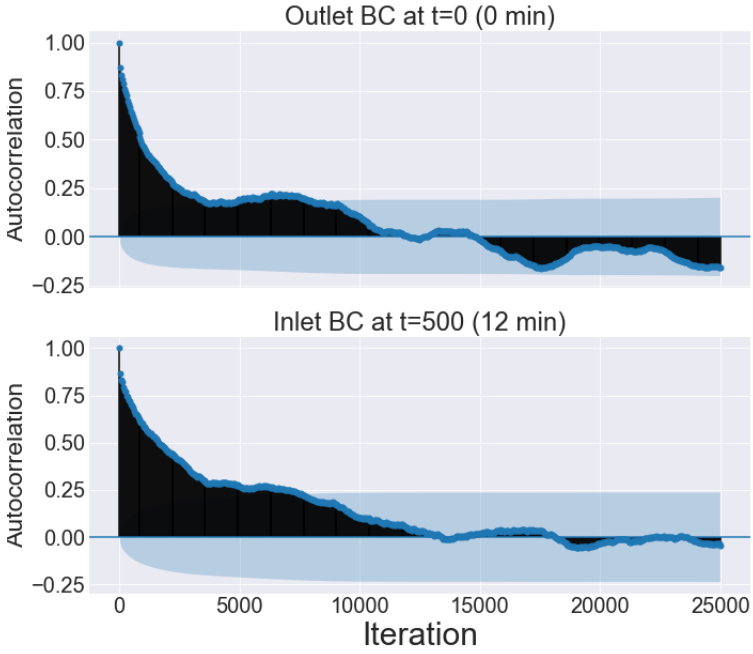

Figure 4.38: ACF for two typical times for the inlet and outet BCs. We use a PT sampler for del Castillo's FD for simulated data.

4.5 Estimating BCs with del Castillo's FD: M25 data

As in the case of the Exponential FD, we fix the FD to the mean of the posterior $\pi(FD|BC)$ found in the previous chapter, namely: $\begin{pmatrix} z & \rho_j & u & w \end{pmatrix}^T = \begin{pmatrix} 179 & 451 & 2.87 & 0.1334 \end{pmatrix}^T$.

We employ a PT sampler with inverse-temperatures `[0.58, 0.76, 1]` and run 3 chains for around 220K iterations. We report the blocks used and their acceptance rates in table (B.4.2). We show the outlet and inlet BC samples in figures (4.39) and (4.40), and notice that the outlet BC exhibits bimodality between times $[0-10]$ for only one chain. Indeed this is confirmed in the trace plots for a few times in the outlet BC in figure (4.41): we can see one of the chains jump to a mode at a density of around 200. This is due to a chain at a higher temperature finding this mode and a swap being accepted. We note that this only occurred after approximately 175K iterations (which takes approximately 72 hours to run); this reinforces the argument made by Geyer (see section (1.2.3)) that one should run MCMC for as long as possible. Running the chains for several days allowed one of them to find this low probability mode. However it is possible that more tuning would allow this mode to be explored more regularly, for

example by adding more Gibbs blocks for those times for all temperatures. The trace plots for 3 typical times for the inlet BC is shown in figure (4.42); this BC appears to not have any mixing problems.

We show the ACF plots for 2 typical times in the outlet and inlet BC in figure (4.43); most of the BC times have delay times between 1K and 3K iterations. Finally, we show the density output of LWR with the posterior mean as BC parameters in figure (4.44): the congested flow waves do cross the domain as they do in data. Of course we are cautious about drawing conclusions based on a non-mixing chain.

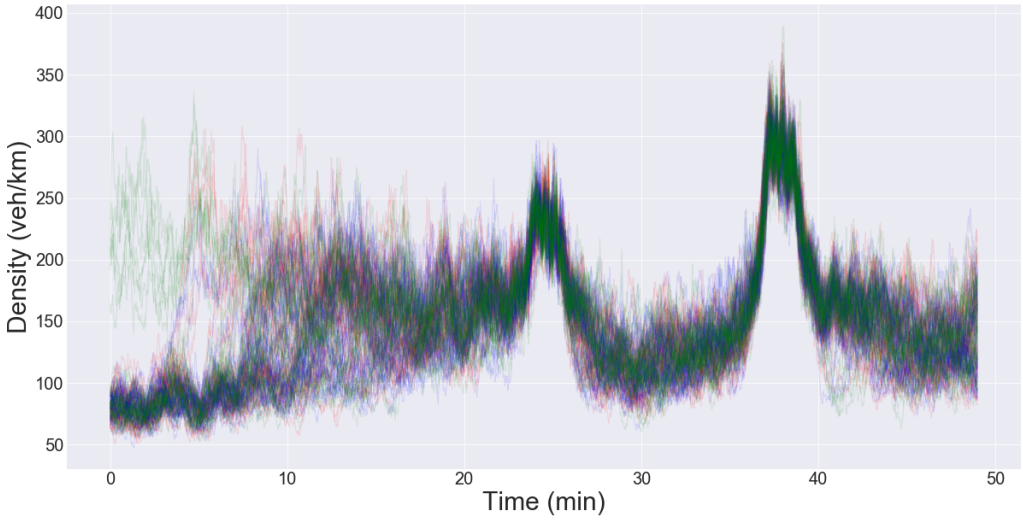

Figure 4.39: Outlet BC samples for a PT sampler with del Castillo's FD on M25 data. We observe that minutes $[0-10]$ exhibit bimodality but that not all the chains explore this mode.

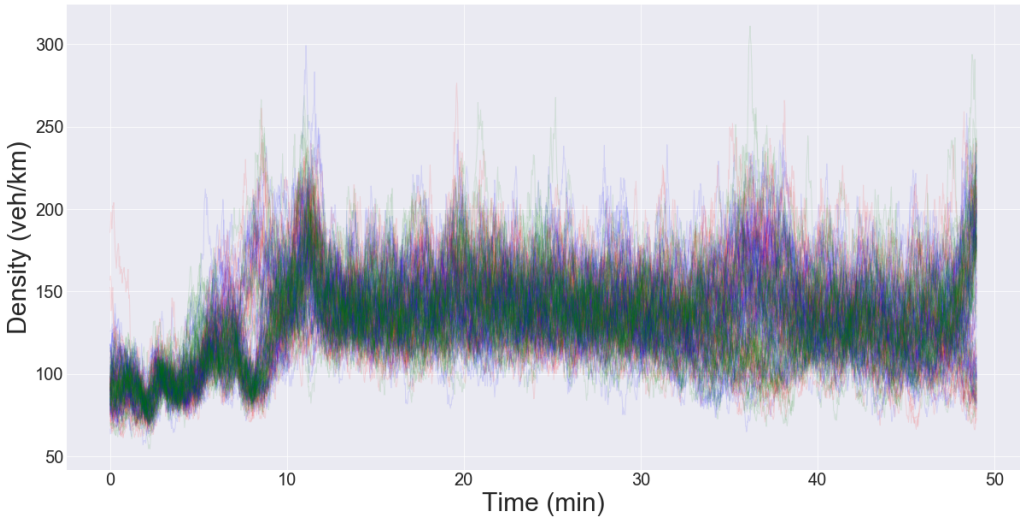

Figure 4.40: Inlet BC samples for a PT sampler with del Castillo's FD on M25 data.

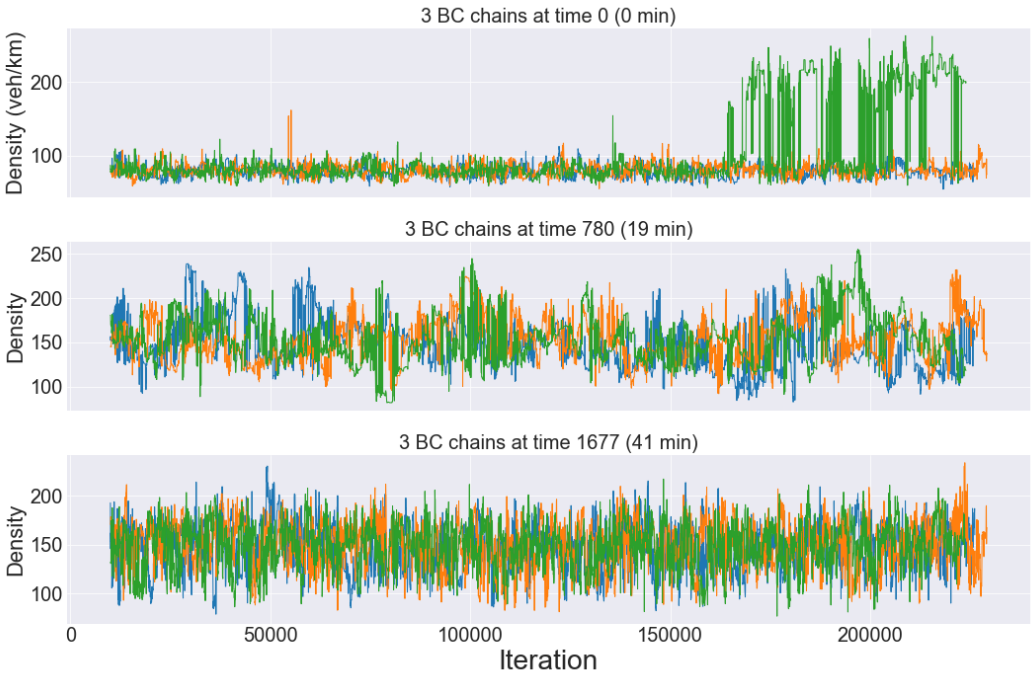

Figure 4.41: Outlet BC samples for a PT sampler with del Castillo's FD on M25 data. We can see at minute 0 the chain that explores the mode found in figure (4.39); however the other chains never find this mode.

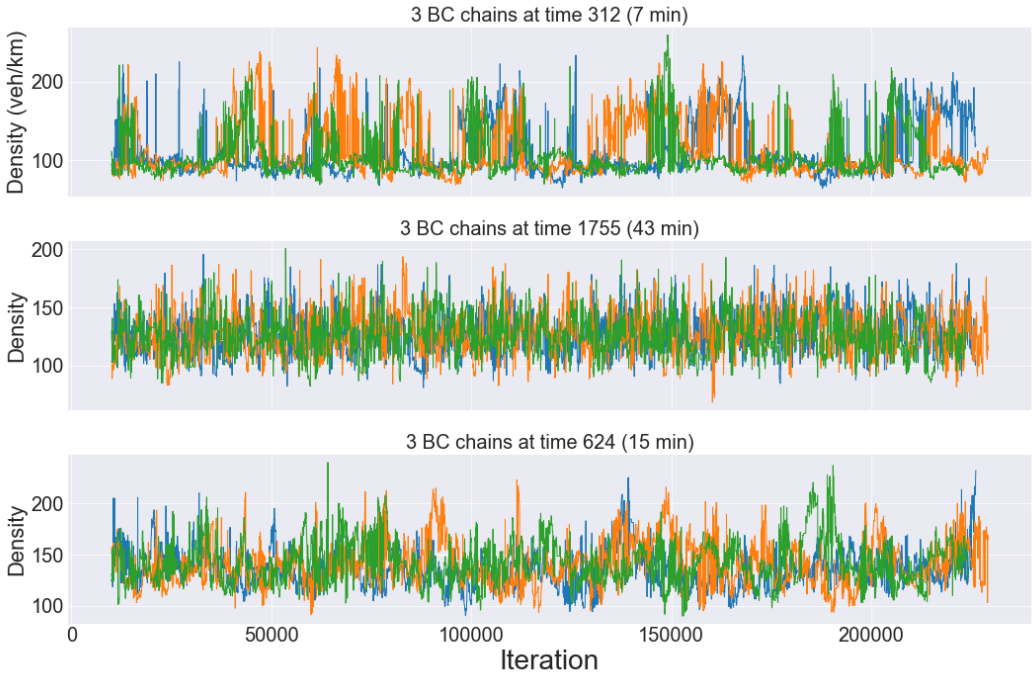

Figure 4.42: Inlet BC samples for a PT sampler with del Castillo's FD on M25 data.

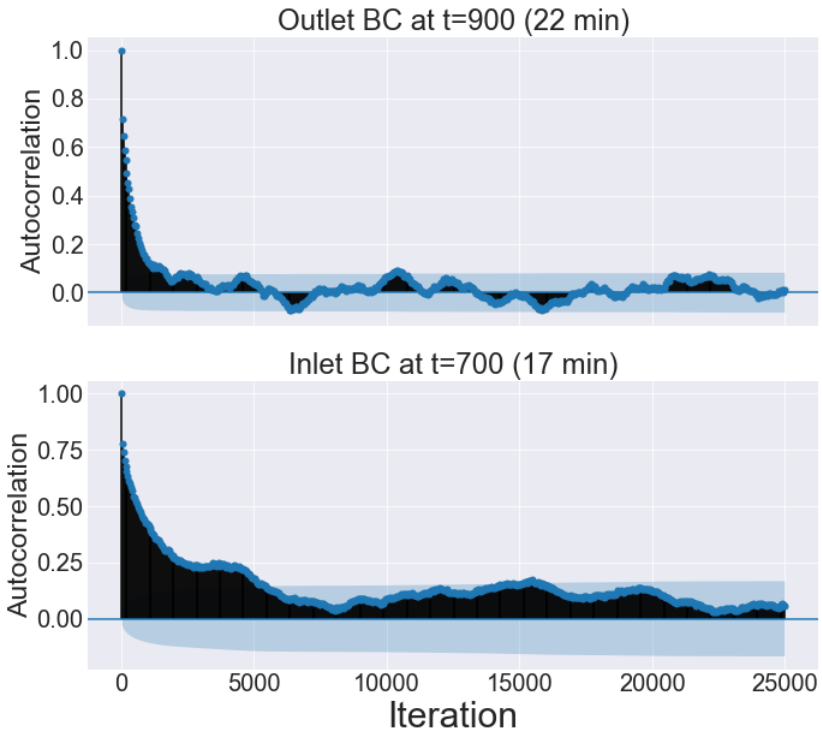

Figure 4.43: ACF for two typical times for the inlet and outlet BCs. We use a PT sampler for del Castillo's FD for M25 data.

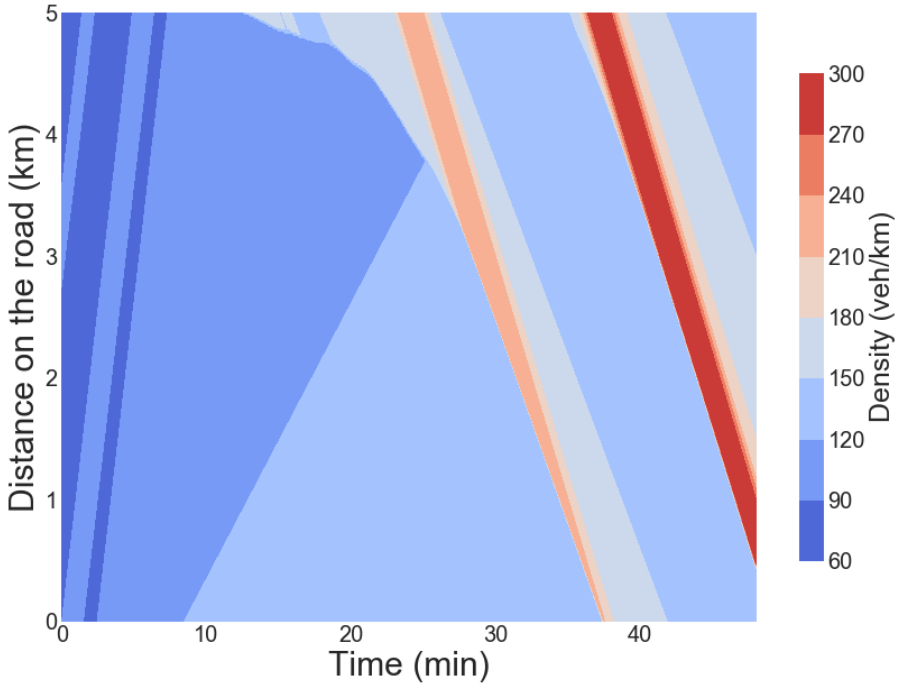

Figure 4.44: We plot the density output of LWR from the posterior mean. We use a PT sampler for del Castillo's FD for M25 data. The congested flow wave speeds cross the domain as they do in real data

4.6 Conclusions

In this chapter we fit the BCs given fixed FD parameters. We elicit a log Ornstein-Uhlenbeck prior with time-dependent mean for the BCs because of computational convenience and qualitative fit to data. We first try a pCN algorithm for the BCs but find that the sampler does not mix. We then try Gibbs blocking, which helps significantly, but has trouble mixing between modes. We then use a Parallel Tempering algorithm (tuned in Chapter 2) and find that the sampler mixes in most cases (Del Castllo's FD on M25 does not quite mix). We give an intuitive explanation for bimodality based on the properties of the FD: different values of density will result in the same value of flow so will result in the same value of the likelihood.

Chapter 5

Estimating FD and BCs in LWR

In this chapter we estimate the FDs and BCs together in LWR. This requires combining the tools developed in the previous chapters and investigating the relationship between these two sets of parameters.

5.1 Exponential FD on simulated data

As in the previous chapters, we start by estimating parameters on simulated data. We employ a parallel temping sampler with the same 4 inverse-temperatures as in the previous chapter (`[0.5, 0.63, 0.8, 1`) and with 5 within-temperature moves. As before, swaps between the temperatures are proposed using a deterministic scan Metropolis-within-Gibbs sampler that goes through each pair in turn; so a swap for each pair is only proposed every 18 iterations. For the within-temperature proposals, we use a random scan Metropolis-within-Gibbs sampler, with probabilities `0.1` and `0.9` for the FD and BC parameters respectively. The proposal for the FD is a Gaussian proposal; the covariance matrices for each temperature are reported in the appendix in section (B.5.1.1) (tuned for fixed BCs). The tuning of the blocks in given in the appendix in table (B.16), (B.17), (B.18), and (B.19) along with their acceptance rates.

We plot the trace plots for the two FD parameters and the log posterior in figure (5.1), which covers the true parameters of $\alpha = 5$ and $\beta = 0.012$ and visually seems to mix well. Indeed R.hat takes values 1.0021 and 1.0001 for α and β respectively. From inspecting the ACF plots the delay time for these two parameters is approximately 3.5K. The FD samples are plotted in figure (5.2) and go through the simulated data appropriately. We show the outlet and inlet samples in figure (5.3) and (5.4) and see the usual bimodality (as described in section 4.2.3.2) (the 3 colours in these plots

correspond to the 3 chains). Here we also seem to have good mixing; the R.hat values in figures (5.5) and (5.6) are below the 1.1 limit for all time points. We show representative trace plots for BC times in figure (5.7) and (5.8) which show a slower mixing time point, a faster mixing one, and a time point exhibiting bimodality. Finally, we plot the ACF for two time points for 30K lags for the inlet and outlet samples in figures (5.9) and (5.10). These show the extremes lags that occur throughout the time points, however most of the them have delay times between 1K and 3K.

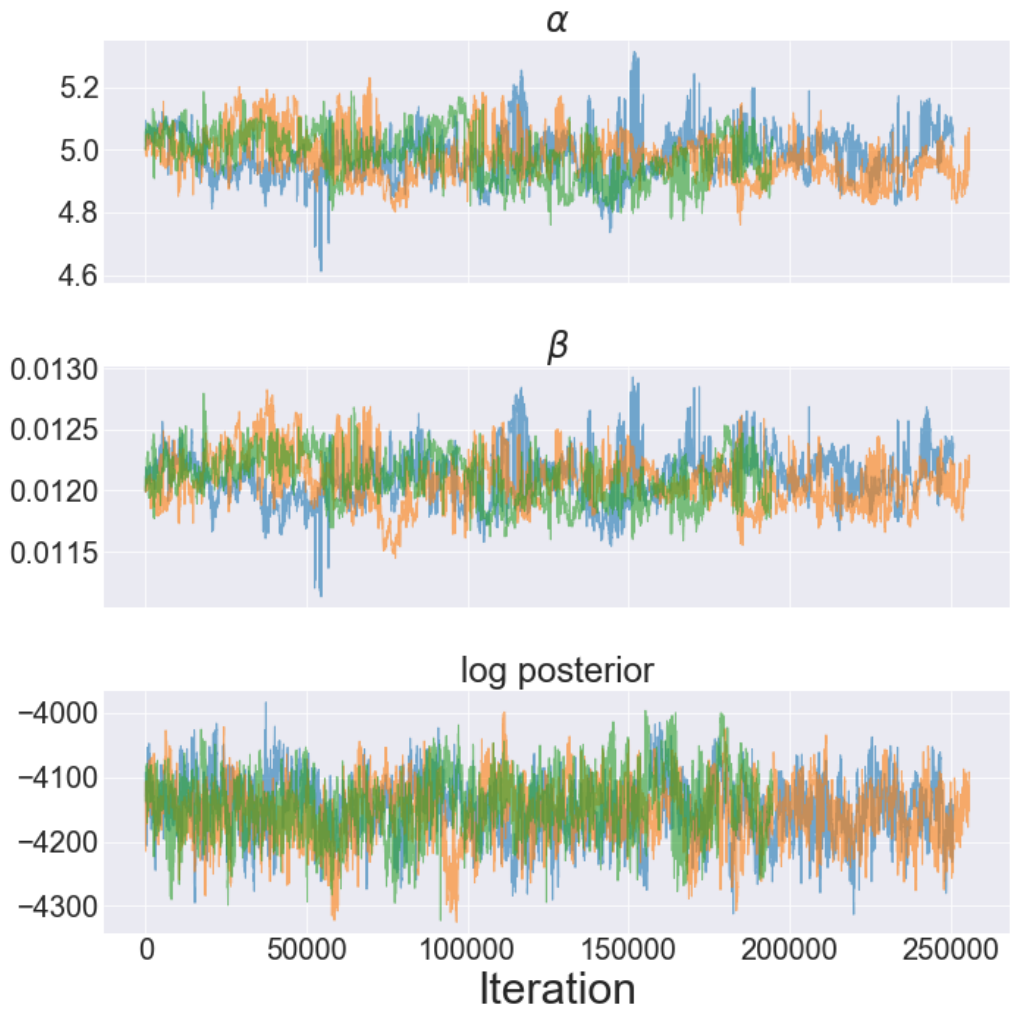

Figure 5.1: Traceplots for the FD parameters α and β along with the log posterior. The samples are from FD and BC sampling for the Exponential FD for simulated data using a parallel tempering sampler. The 3 colours denote the 3 MCMC chains (one chain ran for 200K, and two for 250K)

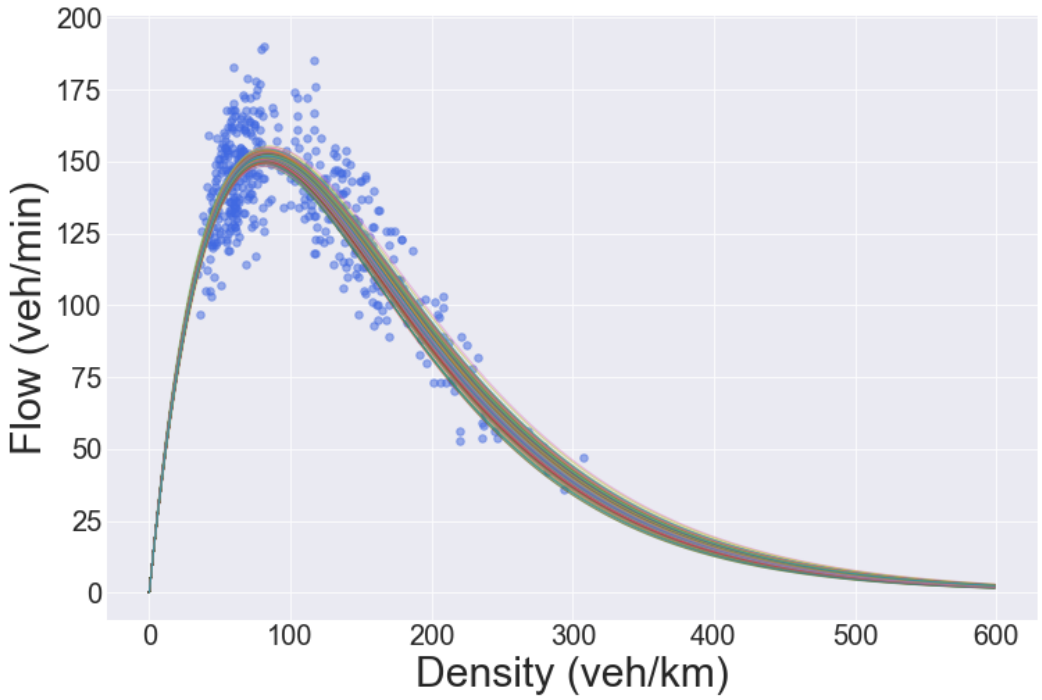

Figure 5.2: FD samples plotted with simulated data. The samples are from FD and BC sampling for the Exponential FD for simulated data using a parallel tempering sampler.

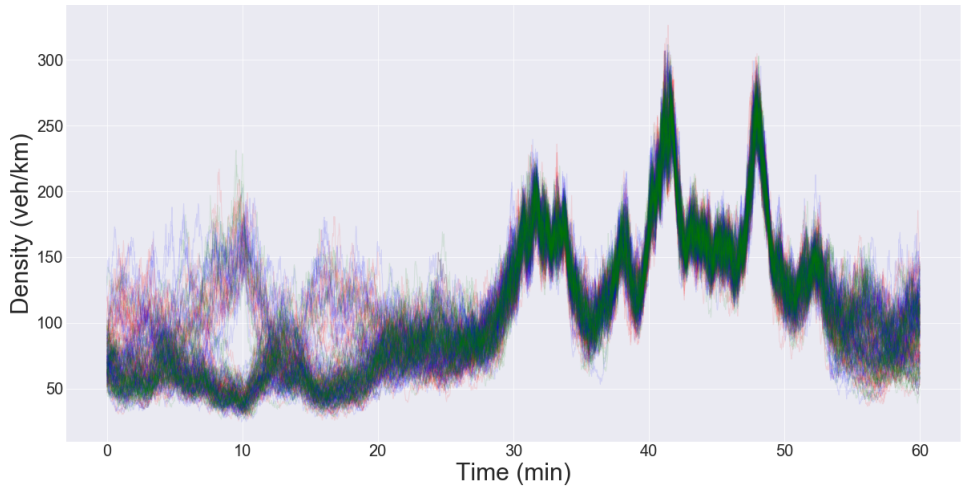

Figure 5.3: Outlet BC samples from FD and BC sampling for the Exponential FD for simulated data using a parallel tempering sampler. The 3 colours denote the 3 MCMC chains.

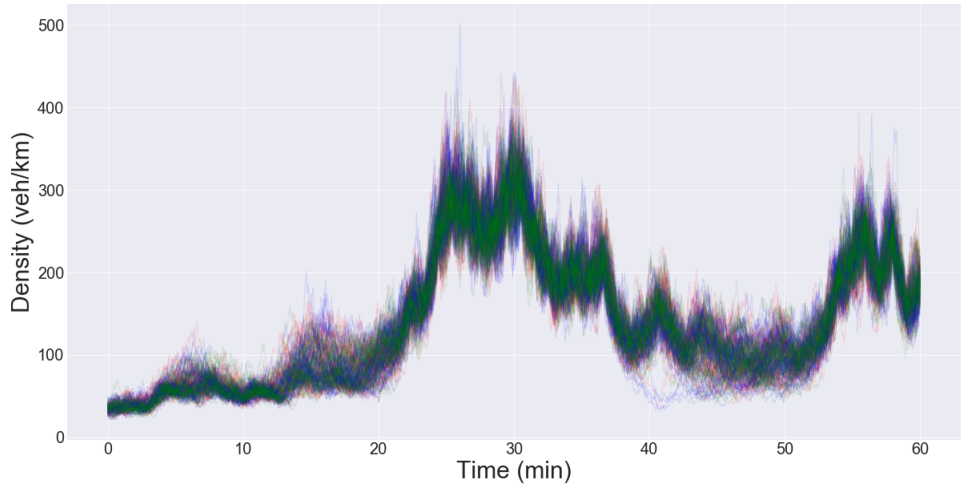

Figure 5.4: Inlet BC samples from FD and BC sampling for the Exponential FD for simulated data using a parallel tempering sampler. The 3 colours denote the 3 MCMC chains.

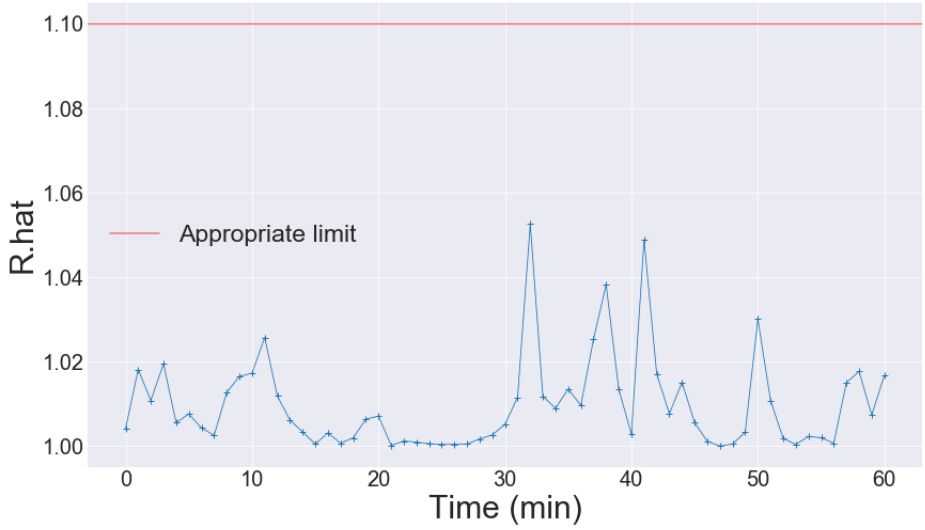

Figure 5.5: R.hat values for the outlet BC from FD and BC sampling for the Exponential FD for simulated data using a parallel tempering sampler. All the time points are below the recommended limit of 1.1

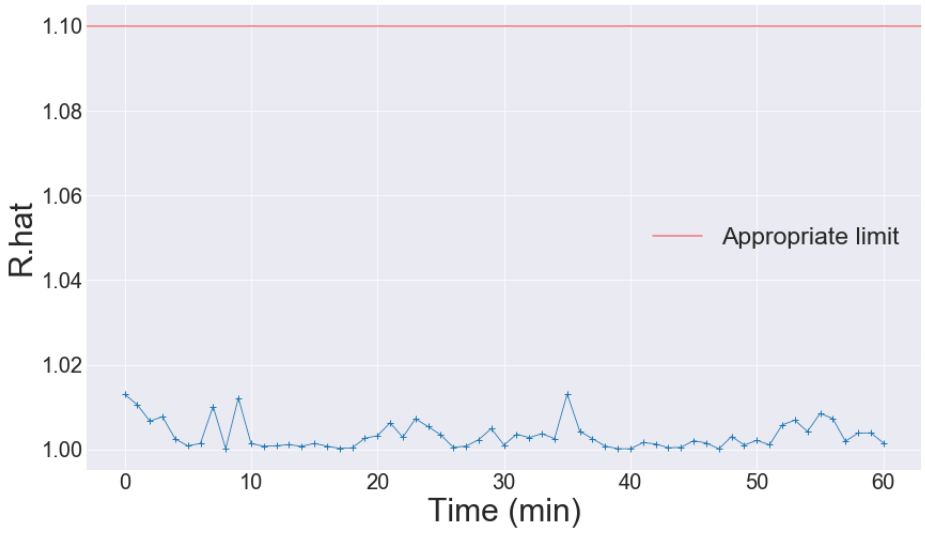

Figure 5.6: R.hat values for the inlet BC from FD and BC sampling for the Exponential FD for simulated data using a parallel tempering sampler. All the time points are below the recommended limit of 1.1

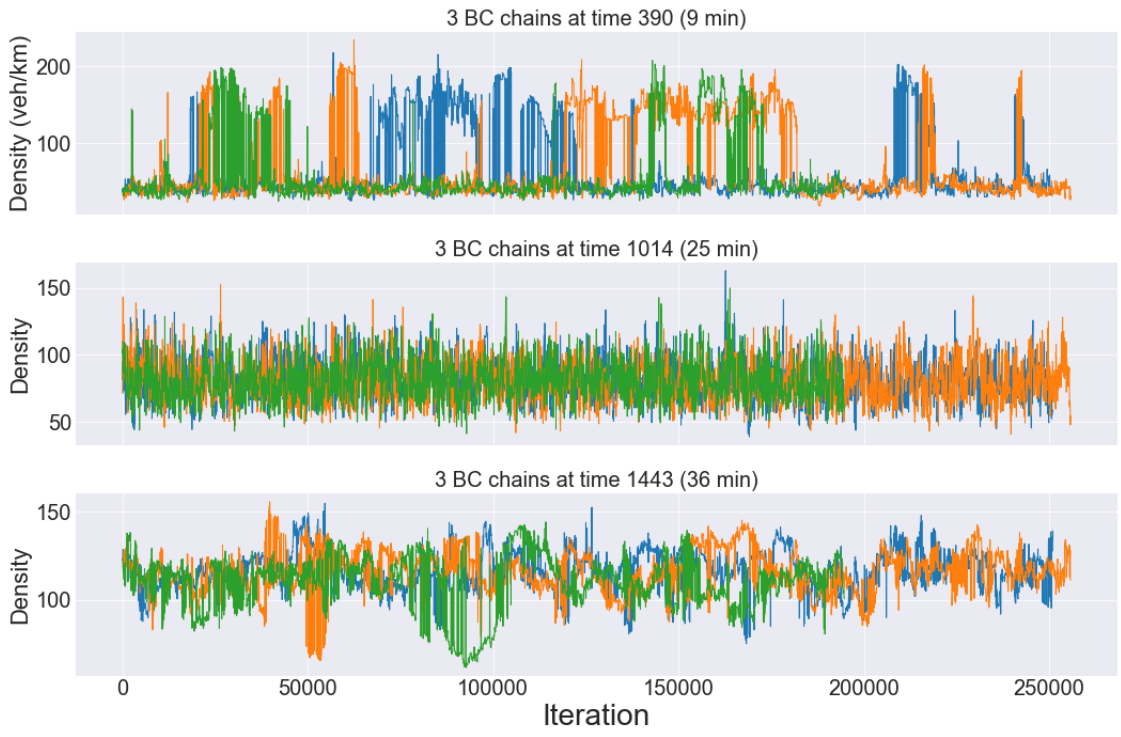

Figure 5.7: Trace plots for 3 outlet BC times for FD and BC sampling for the Exponential FD for simulated data using a parallel tempering sampler. The 3 time points were chosen to be representative of the remaining trace plots of the BCs.

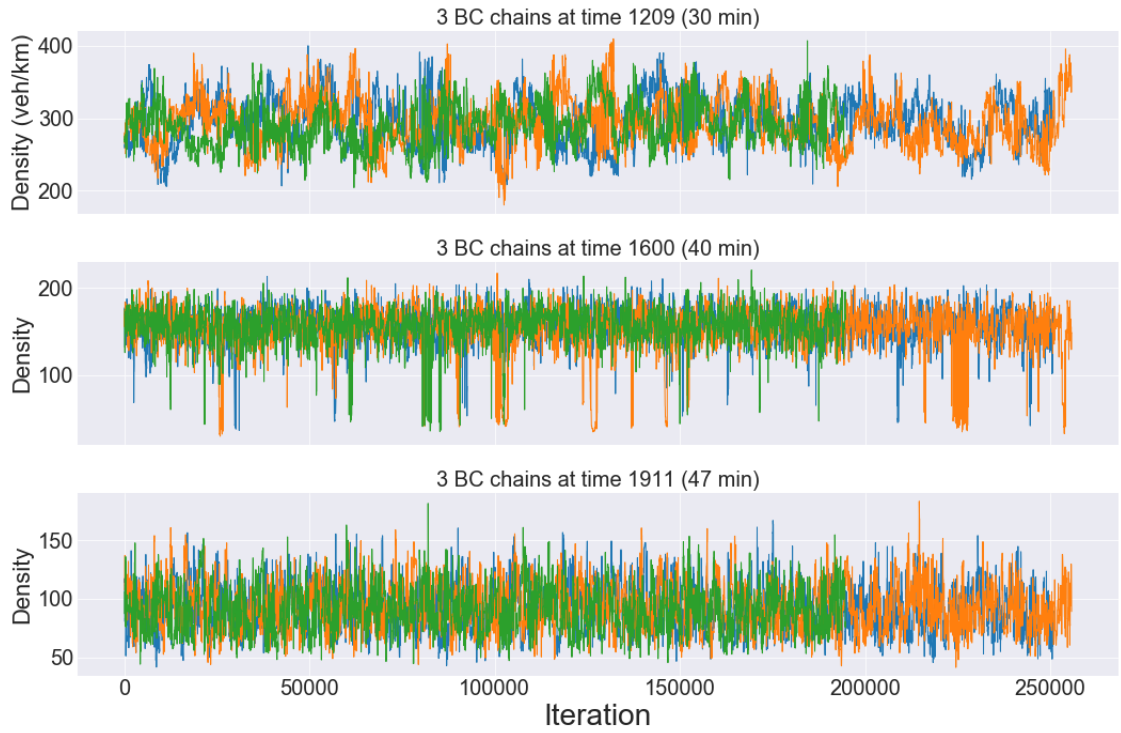

Figure 5.8: Trace plots for 3 inlet BC times for FD and BC sampling for the Exponential FD for simulated data using a parallel tempering sampler. The 3 time points were chosen to be representative of the remaining trace plots of the BCs.

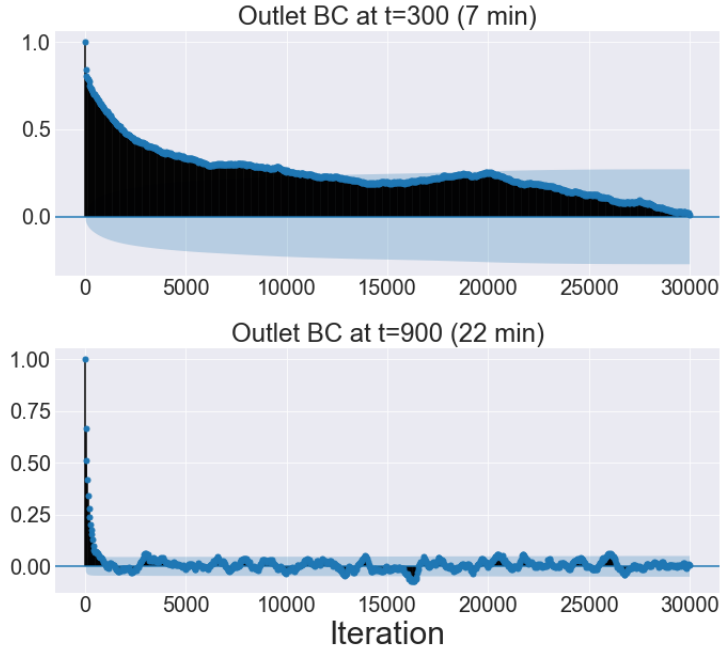

Figure 5.9: ACF for 2 outlet BC times for FD and BC sampling for the Exponential FD for simulated data using a parallel tempering sampler. The 2 time points represent the extremes of the ACF plots of the BCs.

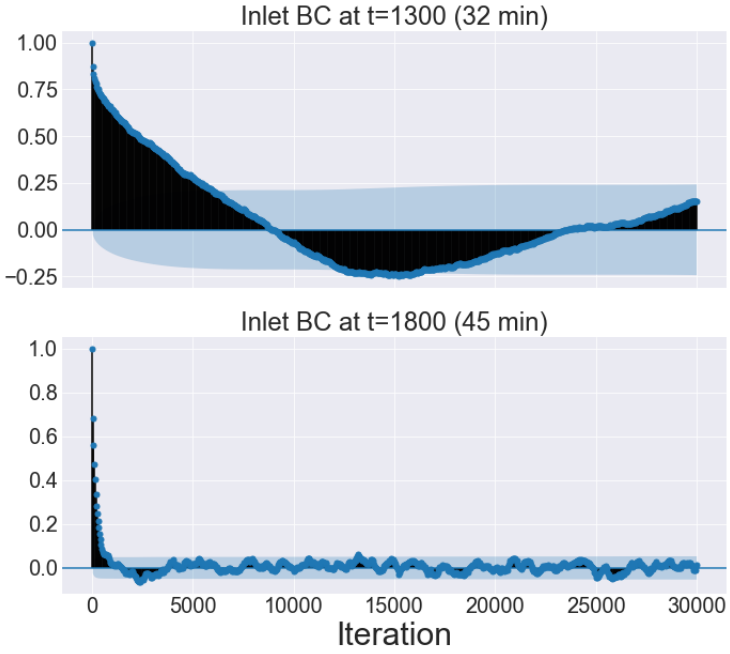

Figure 5.10: ACF for 2 inlet BC times for FD and BC sampling for the Exponential FD for simulated data using a parallel tempering sampler. The 2 time points represent the extremes of the ACF plots of the BCs.

5.2 Exponential FD on M25 Data

We now run the sampler developed in section (5.1) on the M25 dataset to assess its performance. We employ a parallel tempering sampler as in chapter 4 but now with 4 inverse-temperatures (`[0.44, 0.58, 0.76, 1]`) and 4 within-temperature moves. For the within-temperature proposals we propose a new FD and new BCs with probabilities `0.3` and `0.7` using a Gaussian proposal for the FD and Gibbs blocks for the BCs. The covariance matrices and blocks for each temperature are given in the appendix in section (**??**) and tables (B.20), (B.21), (B.22), and (B.23).

We show in figure (5.11) the traces for the FD parameters which seem to mix well. The R.hat for α and β is `1.007`, and the delay time for these two parameters is 3K. We show in figure (5.12) the FD samples plotted with M25 flow data against 3 density estimates: density from speed, from occupancy, and estimated in 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 density estimated in the BCs does not agree with density estimated from occupancy but is similar to density

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 (5.15)) seem to agree with the waves in M25 data (in figures (3.3) and (3.4)), this suggests that estimating the density in LWR rather than estimating it in a preprocessing step yields a better fit of the wave speeds.

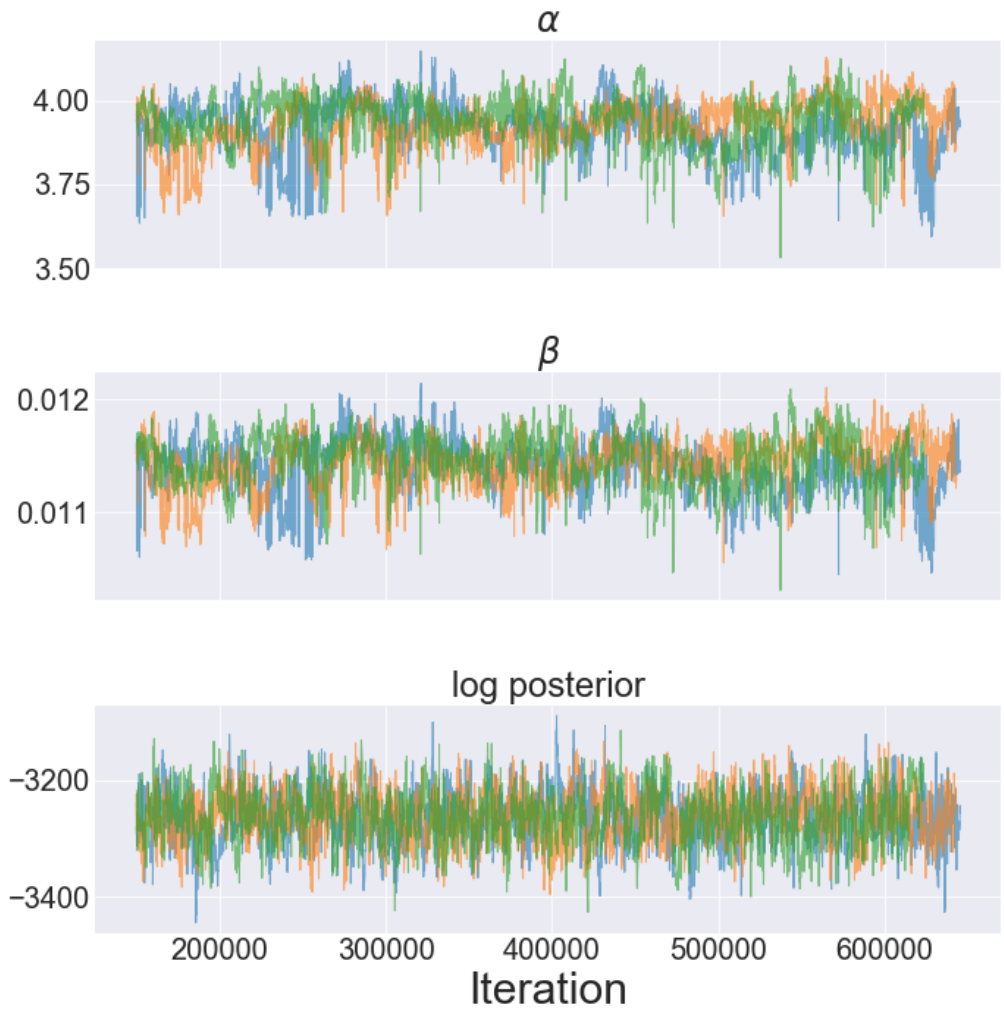

Figure 5.11: Trace plots for the FD parameters α and β along with the log posterior. The samples are from FD and BC sampling for the Exponential FD for M25 data using a parallel tempering sampler. The 3 colours denote the 3 MCMC chains with a burnin of 150K

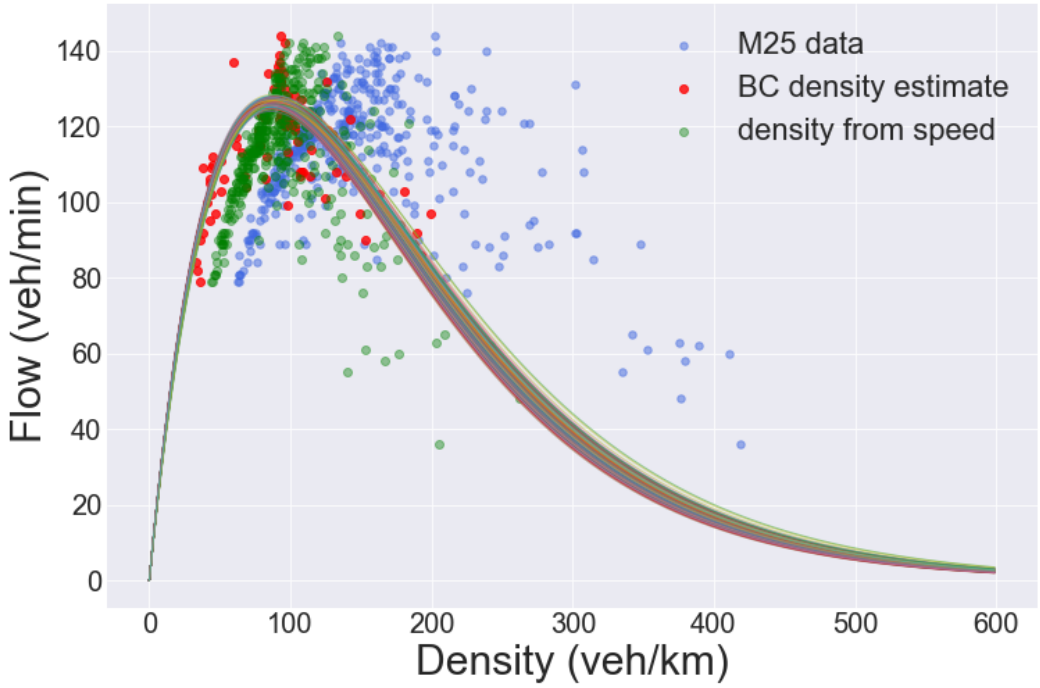

Figure 5.12: FD samples plotted with M25 flow and three density estimates: density from occupancy, density from speed, and density from the mean BCs. The samples are from FD and BC sampling for the Exponential FD for M25 data using a parallel tempering sampler. We note how the density estimated in the BCs seems to agree with density estimated from speed rather than density from occupancy. However the congested flow wave speed in the fitted model seems to be in between the congested flow wave speeds implied by the two other density estimates.

We plot the outlet and inlet BC samples in figures (5.13) and (5.14). The outlet samples look similar to figure (4.25) in chapter 4, but we note how the inlet samples have changed from figure (4.26). If we plot the output of LWR in the $x-t$ plane using the posterior mean parameters in figure (5.15), we can see that the congested flow waves cross the domain and reach the inlet BC (as they do in data). The can explain the change in inlet samples; the high density waves that reach the inlet constrain the likelihood. The BC have excellent R.hat diagnostics (close to 0) as can be seen from figures (5.16) and (5.17), and we show representative traces for the outlet and inlet BCs in figure (5.18) and (5.19). Furthermore, the delay times of the BCs are generally between 500 and 2K which is slightly faster than the simulated data. We also show the extreme ACF plots in figures (5.20) and (5.21).

We inspect the fit of the model by plotting residuals in figure (5.22). Although the

variance does not seem to be increasing with flow as would be expected of a Poisson error model, there is no apparent structure to the residuals; this suggests a good fit of the model.

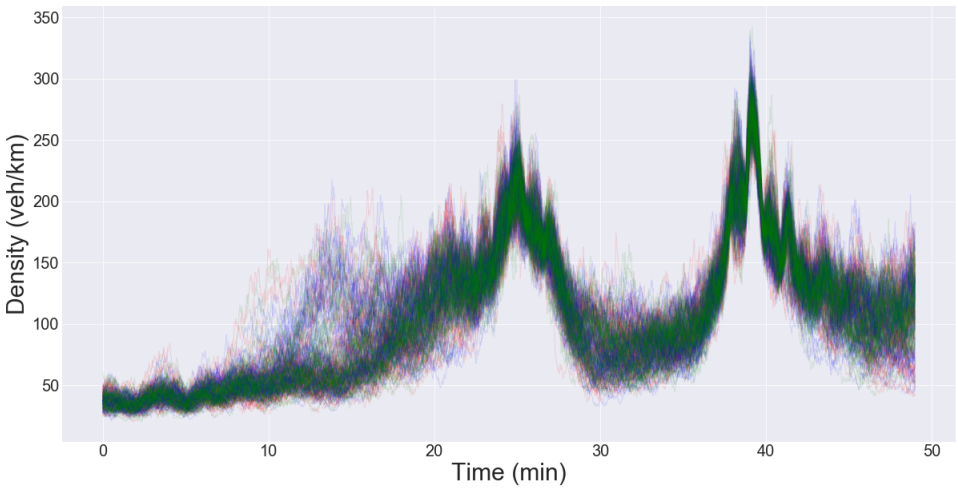

Figure 5.13: Outlet BC samples from FD and BC sampling for the Exponential FD for M25 data using a parallel tempering sampler. The 3 colours denote the 3 MCMC chains.

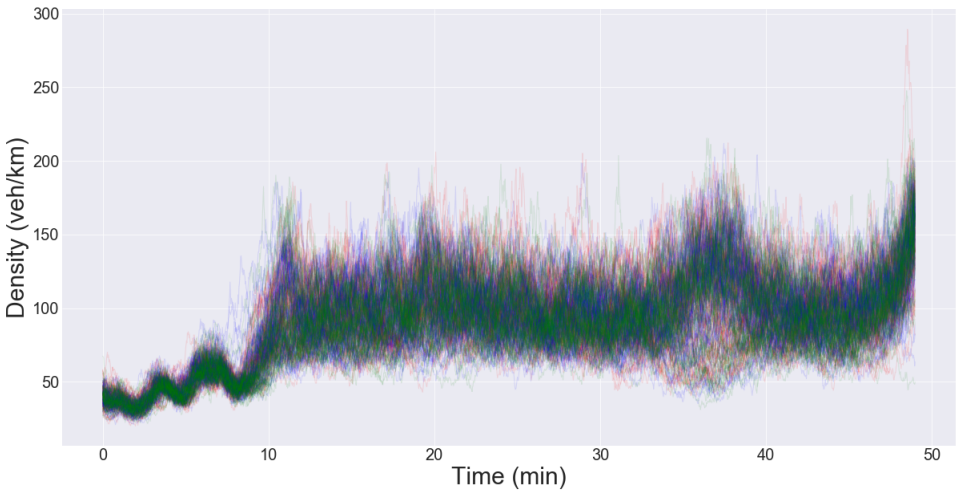

Figure 5.14: Inlet BC samples from FD and BC sampling for the Exponential FD for M25 data using a parallel tempering sampler. The 3 colours denote the 3 MCMC chains.

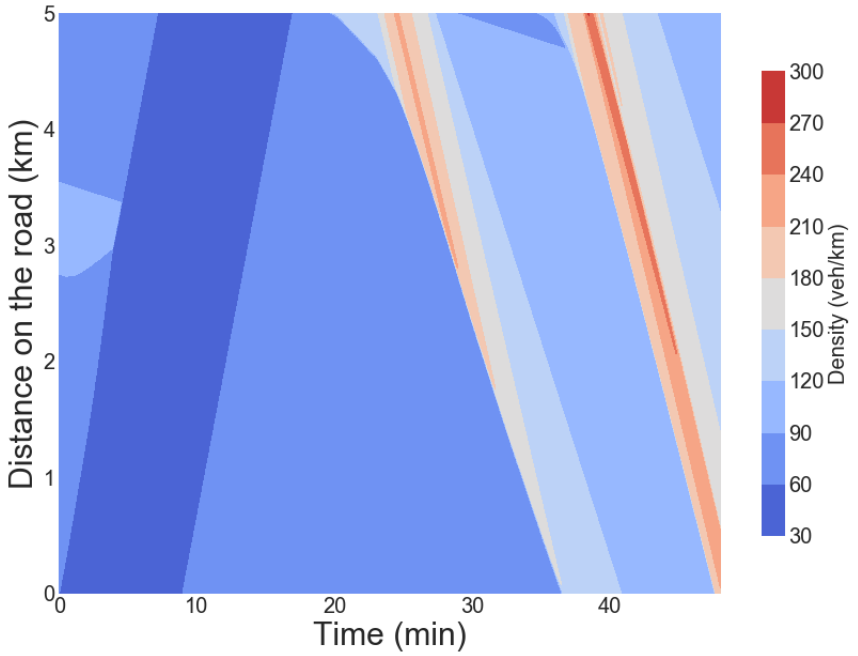

Figure 5.15: Output of LWR in the $x-t$ plane with posterior mean FD and BCs for FD and BC sampling for the Exponential FD for M25 data using a parallel tempering sampler. The congested flow waves cross the domain as they do real data (in figures (3.3) and (3.4)).

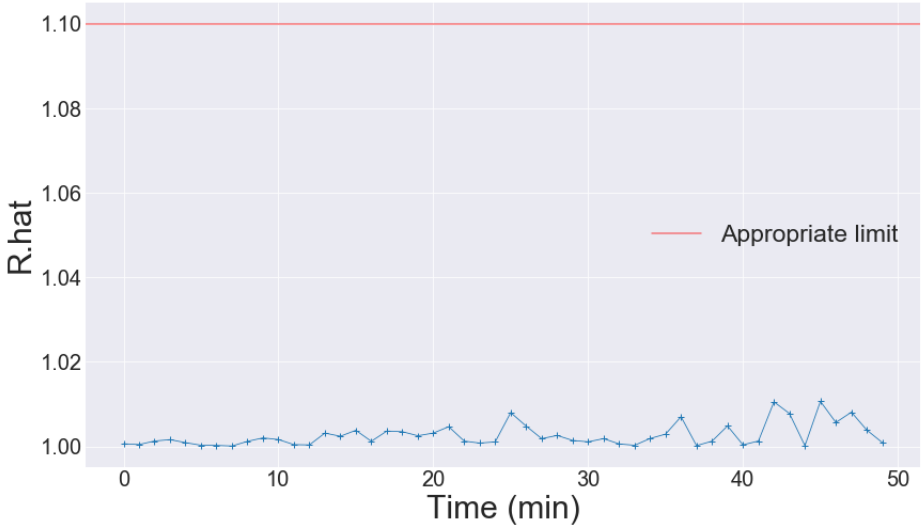

Figure 5.16: R.hat values for the outlet BC from FD and BC sampling for the Exponential FD for M25 data using a parallel tempering sampler. All the time points are well below the recommended limit of 1.1

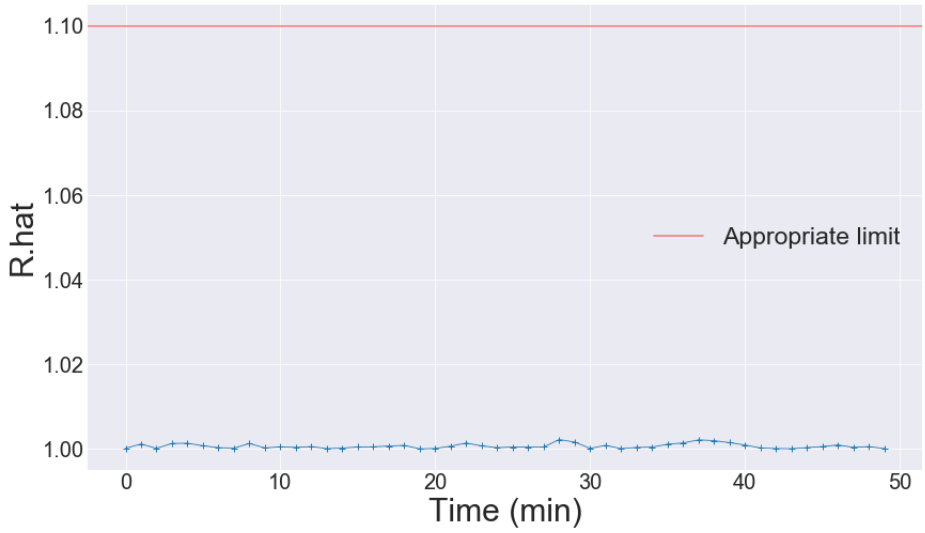

Figure 5.17: R.hat values for the inlet BC from FD and BC sampling for the Exponential FD for M25 data using a parallel tempering sampler. All the time points are well below the recommended limit of 1.1

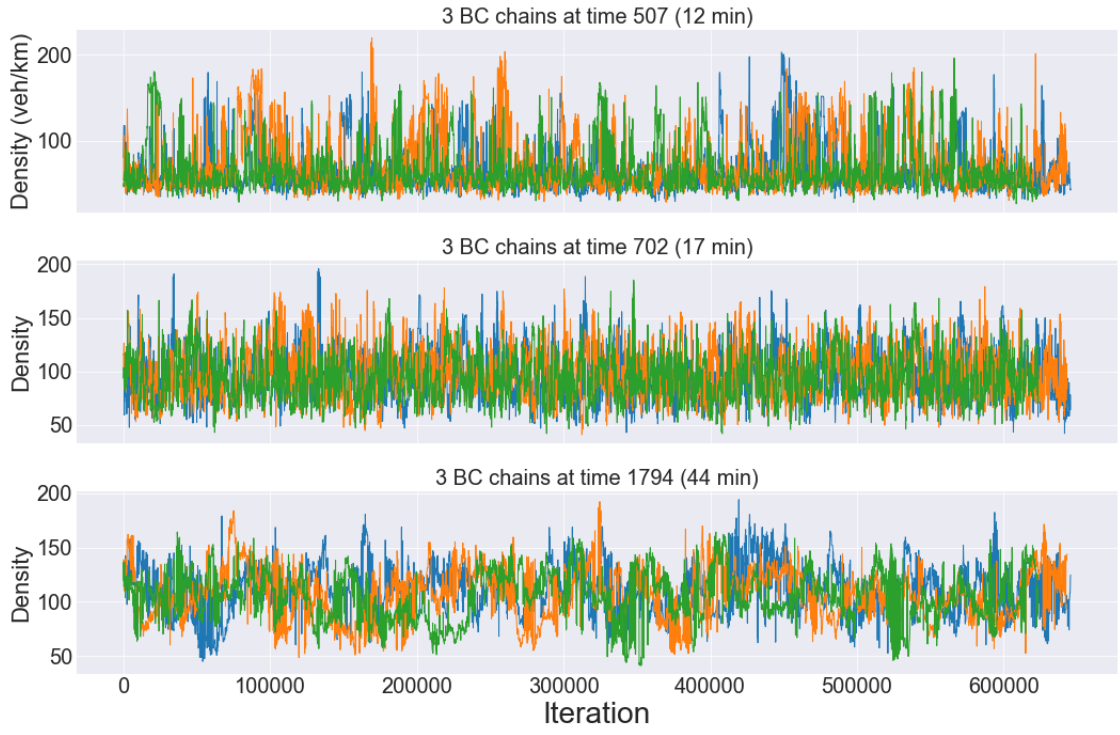

Figure 5.18: Trace plots for 3 outlet BC times for FD and BC sampling for the Exponential FD for M25 data using a parallel tempering sampler. The 3 time points were chosen to be representative of the remaining trace plots of the BCs.

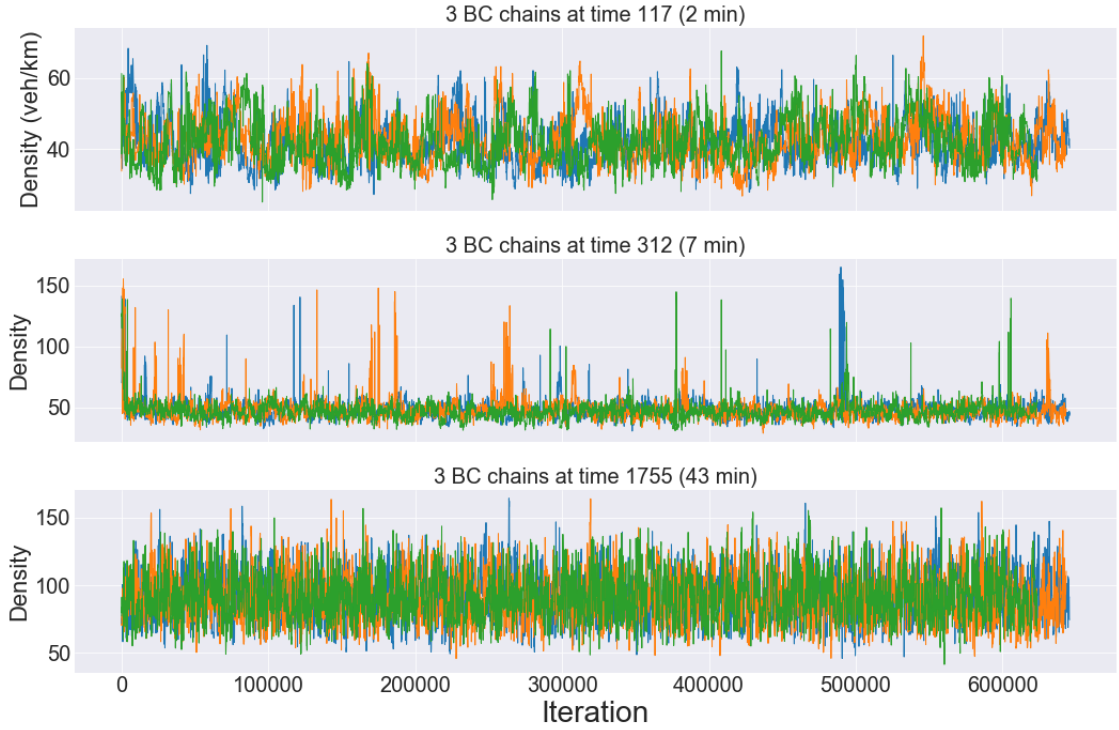

Figure 5.19: Trace plots for 3 inlet BC times for FD and BC sampling for the Exponential FD for M25 data using a parallel tempering sampler. The 3 time points were chosen to be representative of the remaining trace plots of the BCs.

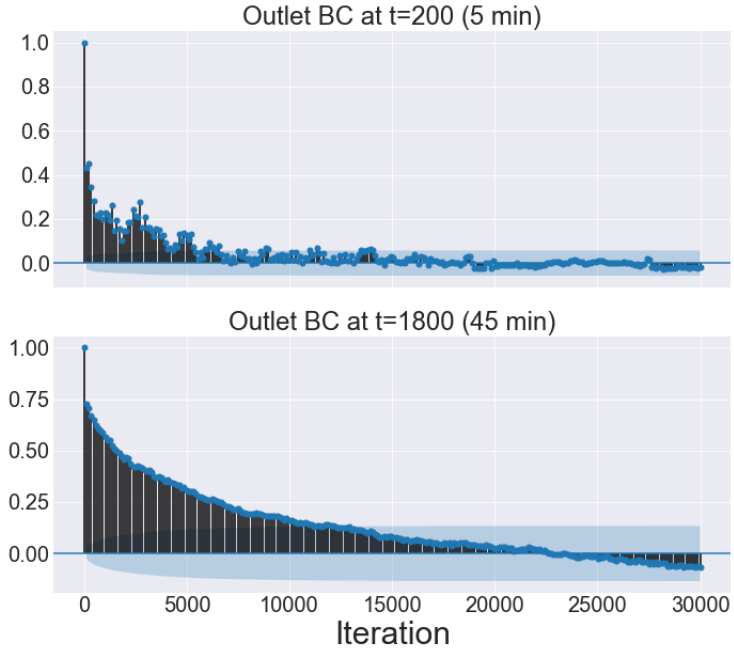

Figure 5.20: ACF for 2 outlet BC times for FD and BC sampling for the Exponential FD for M25 data using a parallel tempering sampler. The 2 time points represent the extremes of the BC ACF plots.

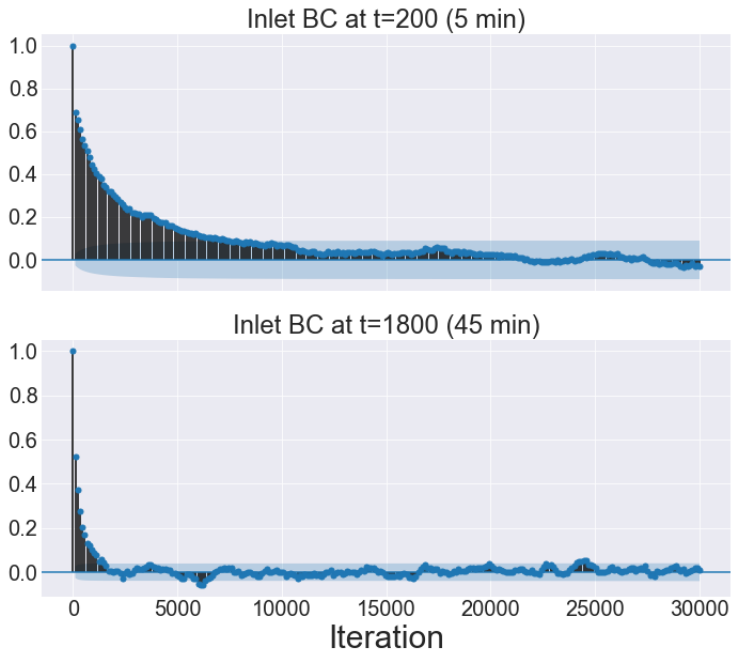

Figure 5.21: ACF for 2 inlet BC times for FD and BC sampling for the Exponential FD for M25 data using a parallel tempering sampler. The 2 time points represent the extremes of the BC ACF plots.

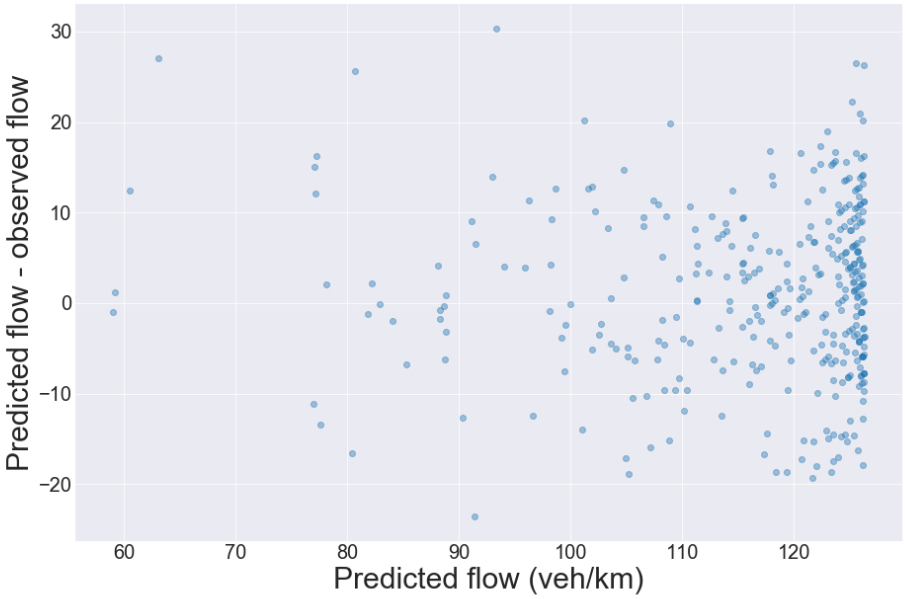

Figure 5.22: Residuals for FD and BC sampling for the Exponential FD for M25 data using a parallel tempering sampler. There is no apparent structure which suggests a good fit to data.

5.3 Del Castillo FD on simulated data

We now estimate FD parameters and BCs using del Castillo's FD (as introduced in section (1.1.2.1)). As the simulated data with 60 minutes was causing mixing problems, we cut the data to only include 40 minutes (cutting 11 minutes from the start and 9 minutes from the end). We use a parallel tempering sampler with inverse temperatures `[0.42, 0.56, 0.75, 1]` and 5 within-temperature moves. We run a random scan Metropolis within Gibbs sampler with probabilities `0.4` and `0.6` for the FD and BC respectively. The covariance matrices for the Gaussian proposal in each temperature and the BC blocks are given in the appendix in section (B.6.1.1) and in tables (B.24), (B.25), (B.26), and (B.27).

We show the trace plots for the FD parameters and log posterior in figure (5.23) and estimate using the ACF plots (not shown) that the delay time for these 3 parameters is approximately 2K. However we notice that one of the chains was interrupted earlier than the other two (around iteration 175K). We also notice that one of the chains is not mixing perfectly: the z, ρ_j, and u parameters jump around iterations 300K and 350K. We can see these excursions into extreme values in the FD samples in figure (5.24): these correspond to the extreme free flow wave speeds and low and high values of ρ_j. We can also see a shift in the BCs for the same range of iteration values in figures (5.29) and (5.30): we can see a jump up and down of the BCs for some of the time points accompanying the jump in FD. Indeed, in section (5.5.2) we will examine the relationship between the BC means and FD parameters (shown in figure (5.44) and (5.45)) and find that these excursions into extreme values loosely follows the same pattern as the rest of the posterior. This (and previous pilot runs) suggests that this excursion is consistent with our understanding of this posterior distribution. More iterations would in principle be desirable to allow the sampler to properly explore these regions of parameter space, but due to the high computational budget involved with this, a better sampler would be preferable.

The BC samples are shown in figures (5.25) and (5.26): we see the usual pattern of bimodality along with low and high variance depending on flow. The R.hat values for the BCs are all below the 1.1 limit as seen in figures (5.27) and (5.28), and the mixing of the BCs is good and can be seen in figure (5.29) and (5.30). Finally, the range of ACF functions is reasonable, with delay times generally between 1K and 2K. We show

the extremes of the ACF plots in figures (5.31) and (5.32).

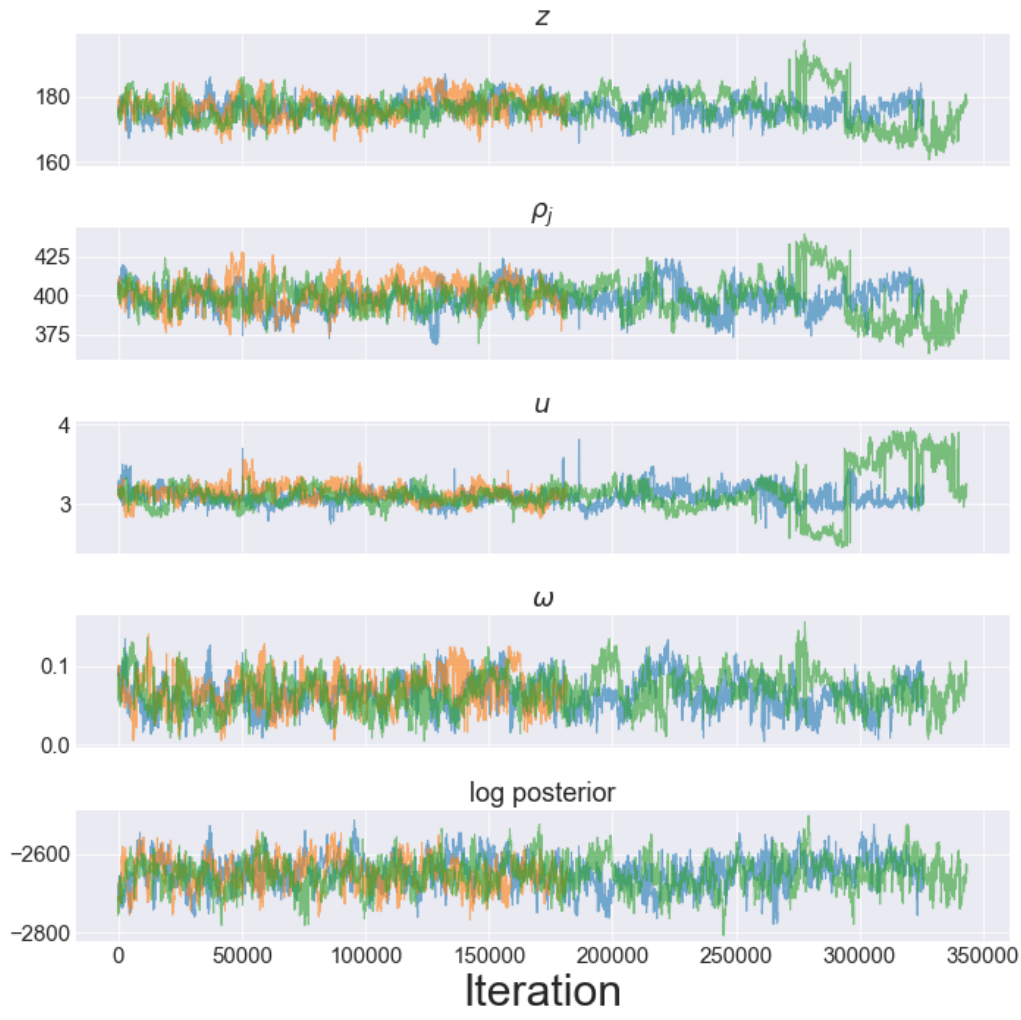

Figure 5.23: Traceplots for the FD parameters along with the log posterior. The samples are from FD and BC sampling for del Castillo's FD for simulated data using a parallel tempering sampler. The 3 colours denote the 3 MCMC chains with a burnin of 150K; we note how one of the chains was interrupted earlier than the others. Another chain exhibits unusual behaviour towards the end of the run in the z, ρ_j, and u parameters.

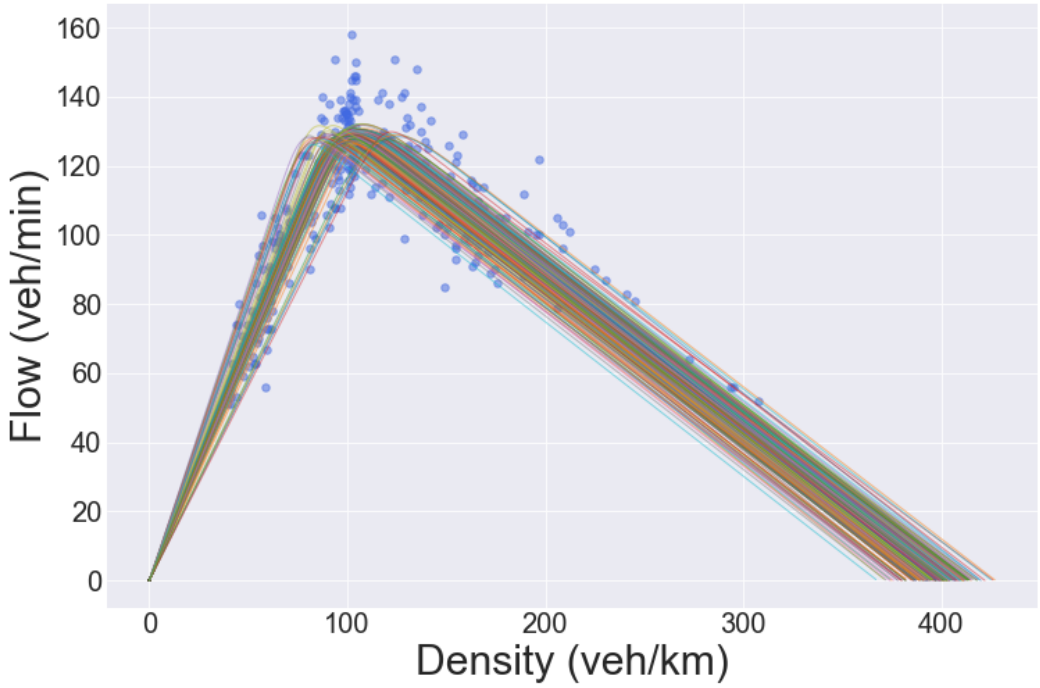

Figure 5.24: FD samples plotted with simulated data. The samples are from FD and BC sampling for del Castillo's FD for simulated data using a parallel tempering sampler. The FDs with extreme values of free flow wave speeds correspond to the excursions of the z, ρ_j, and u parameters seen in figure (5.23).

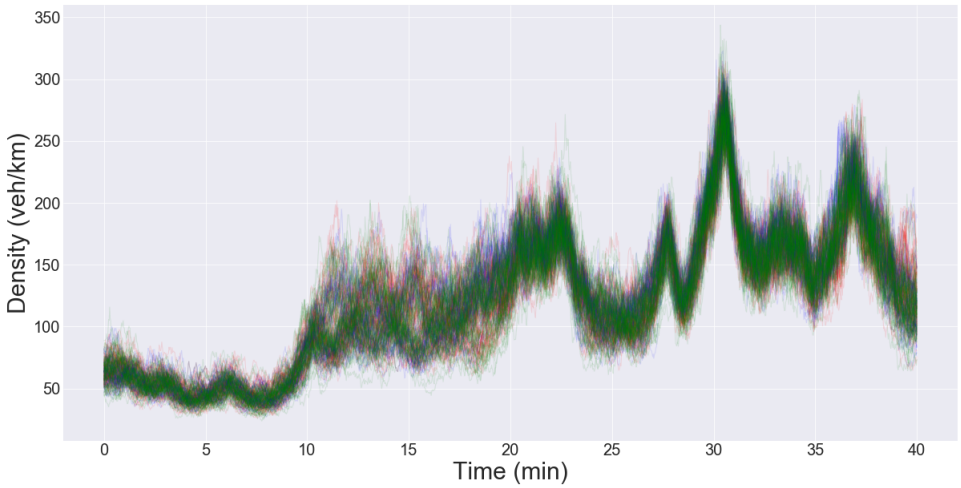

Figure 5.25: Outlet BC samples from FD and BC sampling for del Castillo's FD for simulated data using a parallel tempering sampler. The 3 colours denote the 3 MCMC chains.

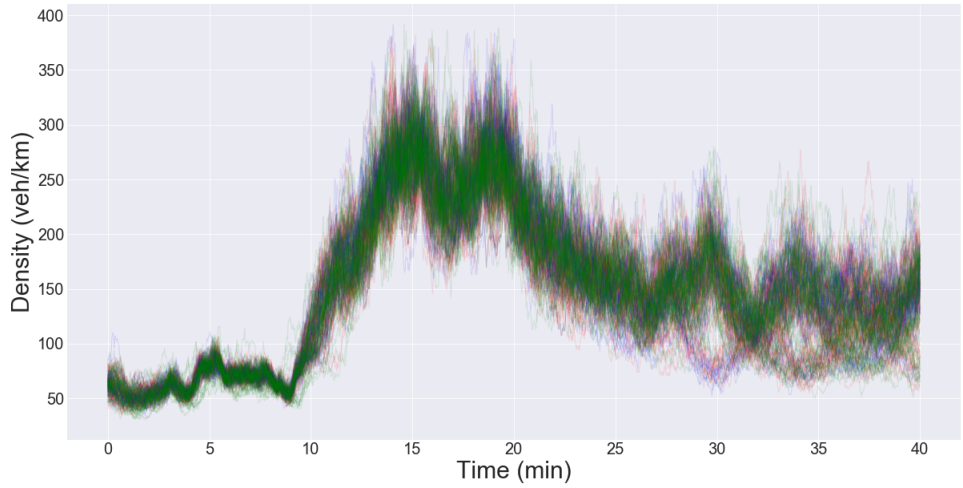

Figure 5.26: Inlet BC samples from FD and BC sampling for del Castillo's FD for simulated data using a parallel tempering sampler. The 3 colours denote the 3 MCMC chains.

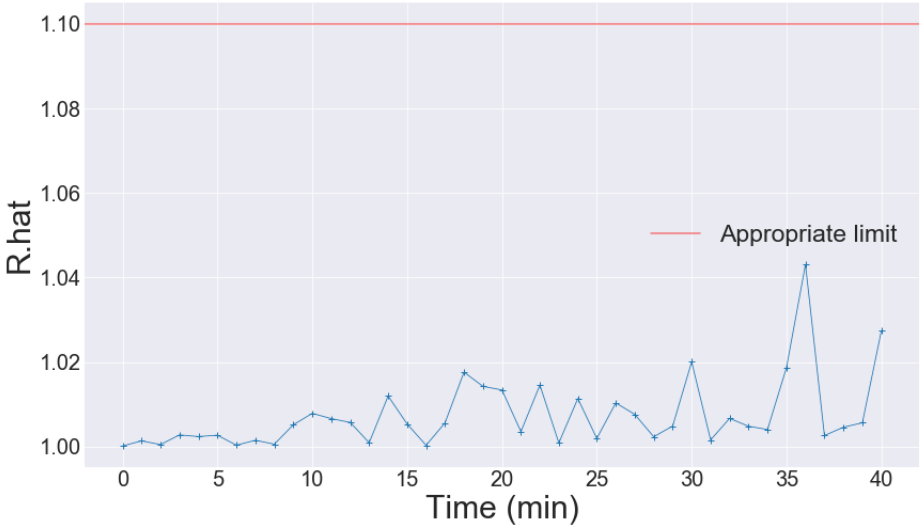

Figure 5.27: R.hat values for the outlet BC from FD and BC sampling for del Castillo's FD for simulated data using a parallel tempering sampler. All the time points are below the recommended limit of 1.1

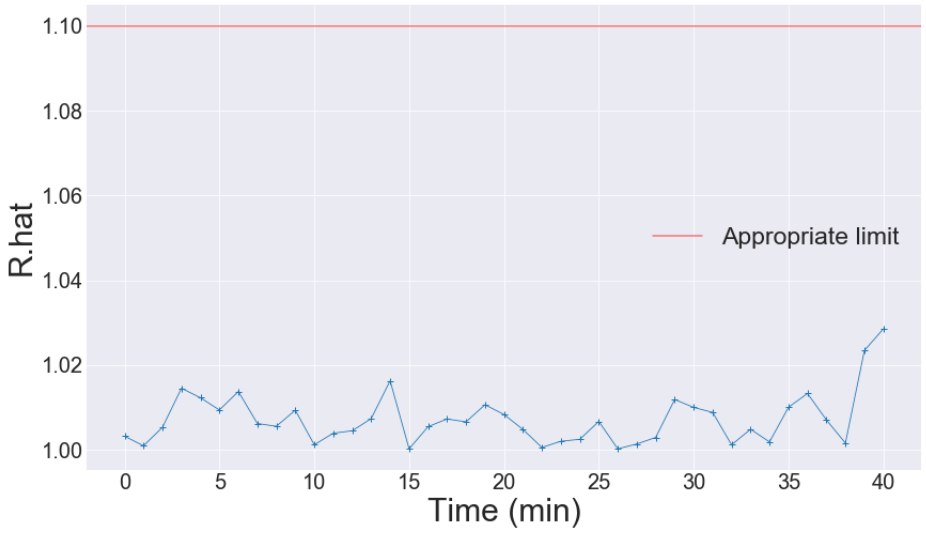

Figure 5.28: R.hat values for the inlet BC from FD and BC sampling for del Castillo's FD for simulated data using a parallel tempering sampler. All the time points are below the recommended limit of 1.1

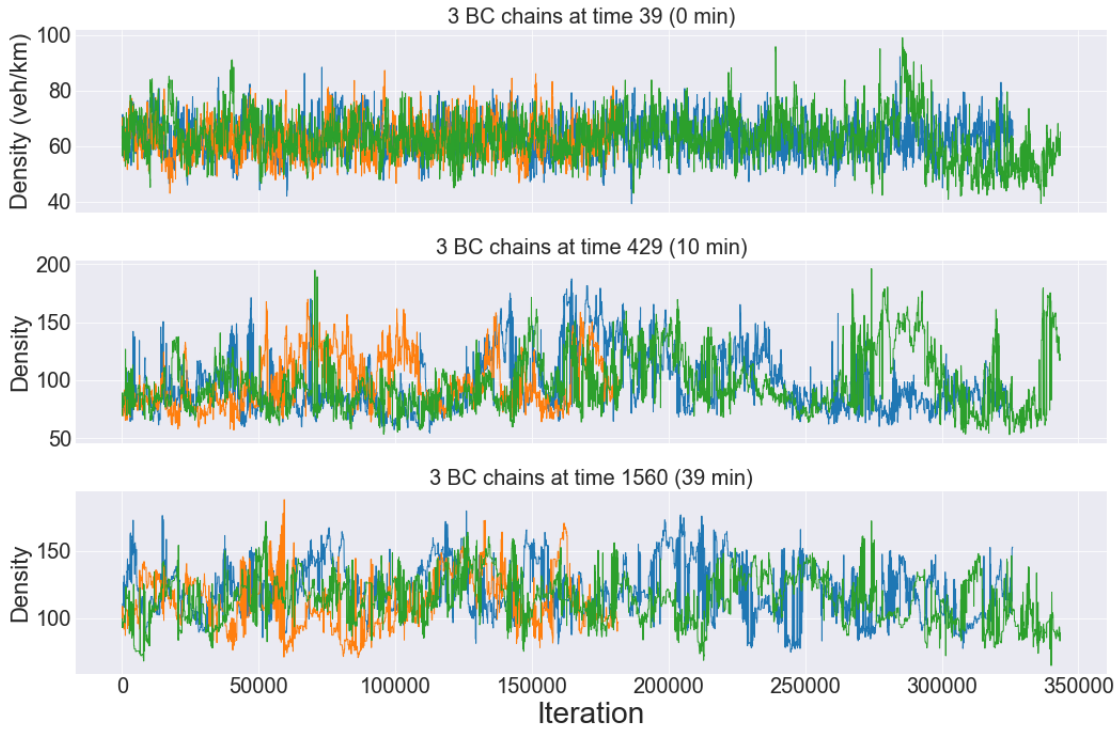

Figure 5.29: Trace plots for 3 outlet BC times for FD and BC sampling for del Castillo's FD for simulated data using a parallel tempering sampler. The 3 time points were chosen to be representative of the remaining trace plots of the BCs. We can see a slight shift in the BC a bit before iteration 300K especially for time point 39 (0min) which corresponds to the excursions in the z, ρ_j, and u parameters seen in figure (5.23).

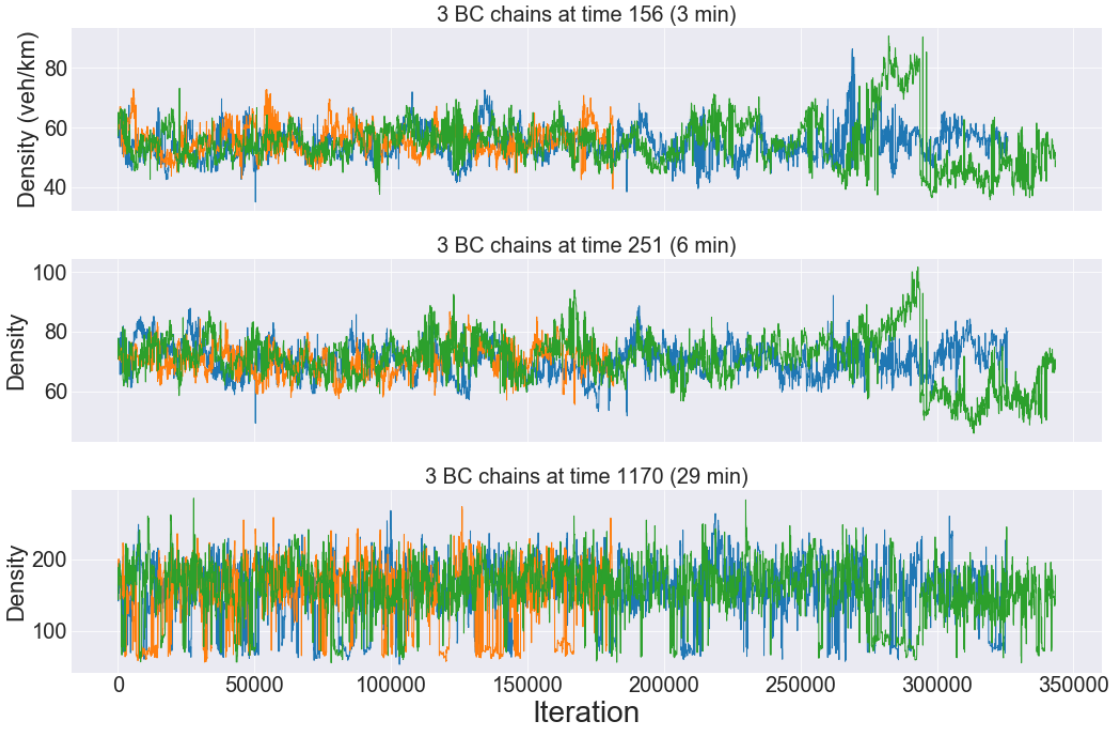

Figure 5.30: Trace plots for 3 inlet BC times for FD and BC sampling for del Castillo's FD for simulated data using a parallel tempering sampler. The 3 time points were chosen to be representative of the remaining trace plots of the BCs. We can see a shift in the BC a bit before iteration 300K especially for time point 156 (3min) and 251 (6min) which corresponds to the excursions in the z, ρ_j, and u parameters seen in figure (5.23).

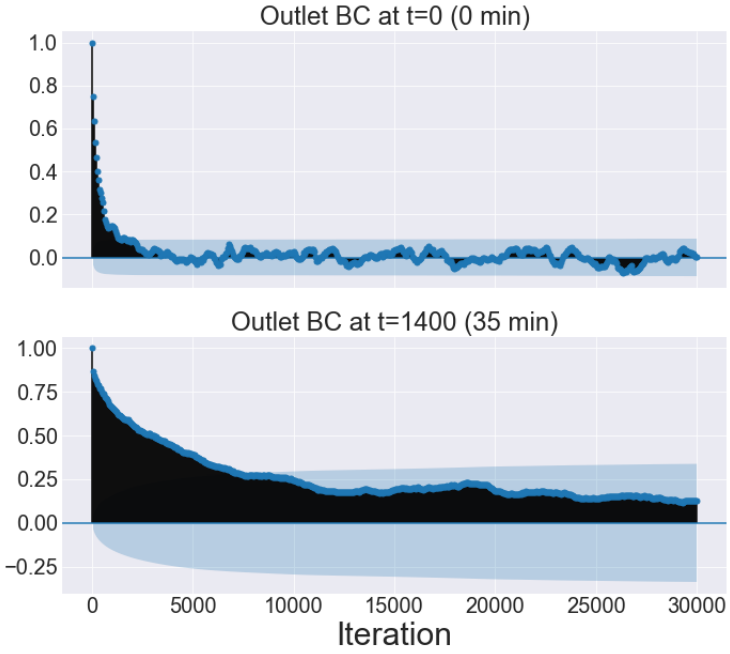

Figure 5.31: ACF for 2 outlet BC times for FD and BC sampling for del Castillo's FD for simulated data using a parallel tempering sampler. The 2 time points represent the extremes of the remaining ACF plots of the BCs.

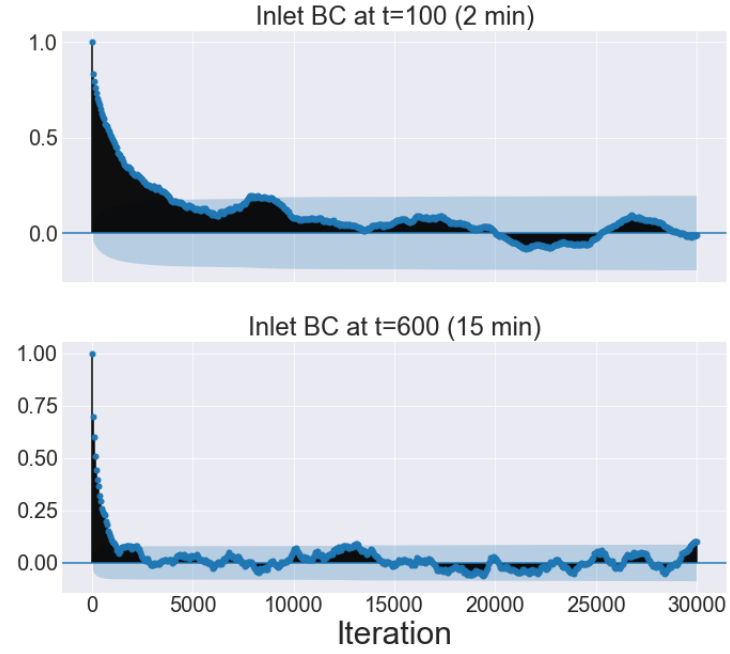

Figure 5.32: ACF for 2 inlet BC times for FD and BC sampling for del Castillo's FD for simulated data using a parallel tempering sampler. The 2 time points represent the extremes of the remaining ACF plots of the BCs.

5.4 Del Castillo's FD on M25 data

We employ a parallel tempering sampler with 4 inverse-temperatures and 3 within-temperature moves. We use the inverse-temperatures `[0.44, 0.58, 0.76, 1]`. For the within-temperature proposals we propose a new FD and new BCs with probabilities `0.4` and `0.6` using a Gaussian proposal for the FD and Gibbs blocks for the BCs. The covariance matrices and blocks for each temperature are given in the appendix in section (B.6.2). With run 3 within-temperature moves for every between-temperature move. We run 3 chains for 230K iterations and we obtain acceptance rates of 15%, 19%, and 18% for the FD proposals for each chain. The R.hat diagnostic for each FD parameter is below 1.1, their delay time is approximately 7.5K, and we show the trace plots in figure (5.33) which suggests good (albeit slow) mixing.

We show in figure (5.34) the FD samples plotted with the three density estimates as we did for the Exponential FD in section (5.2). We observe here a similar result, namely that density estimated in the BCs seems to agree with density estimated from speed rather than density from occupancy. Furthermore, the free flow wave speeds implied by the 3 estimation methods seem to agree, but the congested flow wave speed of the fitted model seems to be in between the wave speeds implied by the other two methods. As in the case of the Exponential FD in section (5.2), wave speeds of the fitted model (in figure (5.39)) seem to agree with those in M25 data (in figures (3.3) and (3.4)). This suggests that the developed MCMC methodology yields a better fit to dynamic information in traffic than estimating density in a pre-processing step.

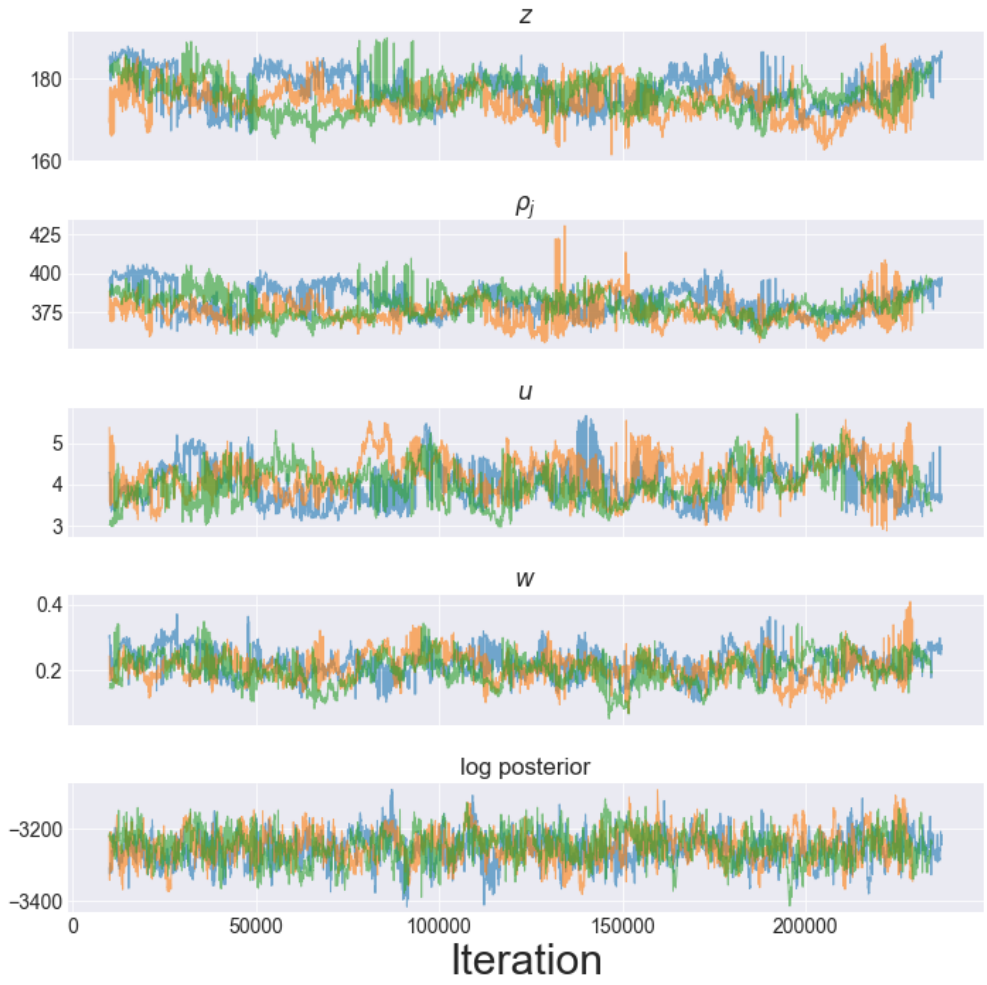

Figure 5.33: Traceplots for the FD parameters along with the log posterior. The samples are from FD and BC sampling for del Castillo's FD for M25 data. The 3 colours denote the 3 MCMC chains with a burnin of 10K

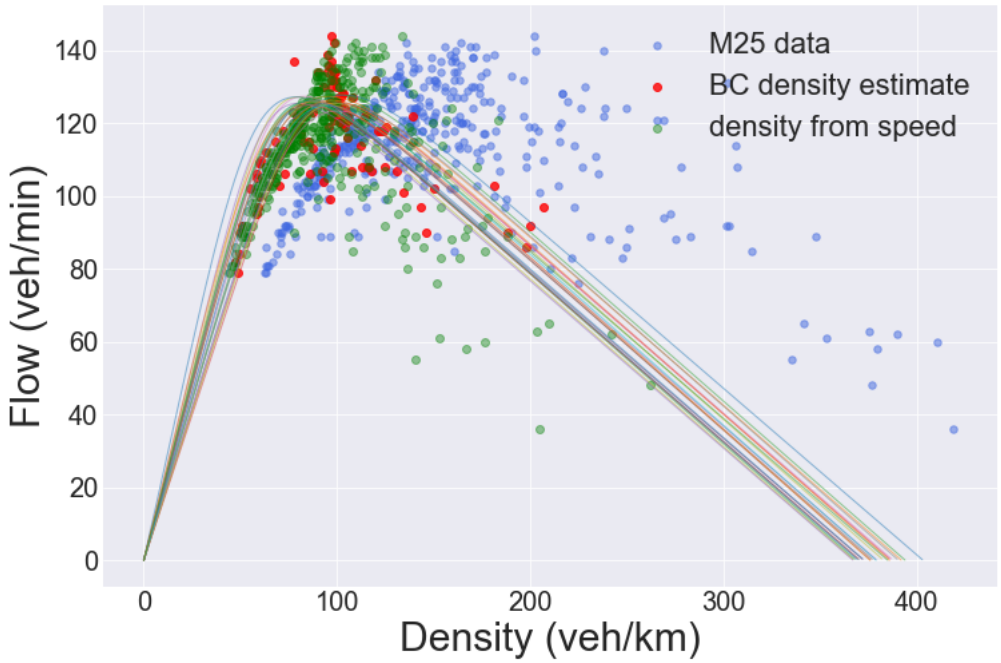

Figure 5.34: FD samples plotted with M25 flow data and three density estimation methods: from occupancy, from speed, and from BCs. The samples are from FD and BC sampling for del Castillo's FD for M25 data. The density estimated in the BCs seems to agree with density from speed, but the congested flow wave speed in the fitted model seems to be different to the wave speeds implied by the other two density estimation methods.

However, we a cautious about this conclusion as this sampler experienced some mixing problems which we now describe.

We show outlet and inlet BC samples in figures (5.35) and (5.36) which visually seems to mix. However we also employed the multistart heuristic and found that one of the chains explored the outlet BC mode around minutes $[0-10]$ which was also found in the $BC|FD$ sampler (see figure (4.41) for the trace plots of that run). The long run (of 230K iterations) however did not explore this mode. We show the trace plots from the multistart chain at time 0 in figure (5.37) as well as 2 typical time points for the outlet and inlet BCs for the long run. All other trace plots (excluding the outlet at times $[0-10]$) minutes in both the multistart run and long run seem to mix well, and the R.hat values are below 1.1 for all BC time points. This result reinforces the prescription in section (1.2.3) that one should use a range of diagnostics rather than rely on a single one. Indeed the multistart heuristic found the mode for this posterior while long runs

found the mode in the case of the $BC|FD$ sampler in section (4.5).

We show the ACF plots for two typical BC times in figure (5.38); most the delay times are between 1K and 3K. We also show the density output of LWR in figure (5.39) which agrees qualitatively with the real data in figures (3.3) and (3.4). Finally we plot the residuals in figure (5.40); as in the case of the Exponential FD, we find no apparent structure in the residuals which suggests a good model fit. However as the chains did not fully mix we are cautious about interpreting the result.

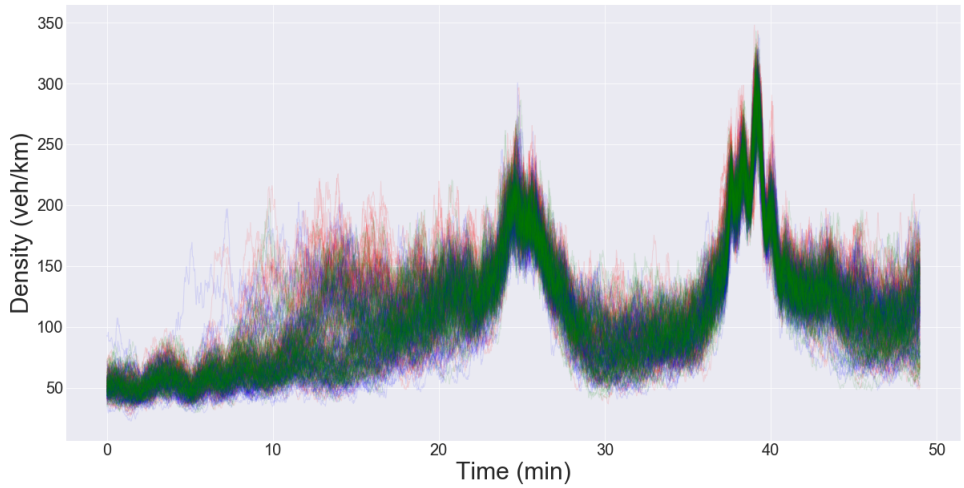

Figure 5.35: Outlet BC samples from FD and BC sampling for del Castillo's FD for M25 data. The 3 colours denote the 3 chains.

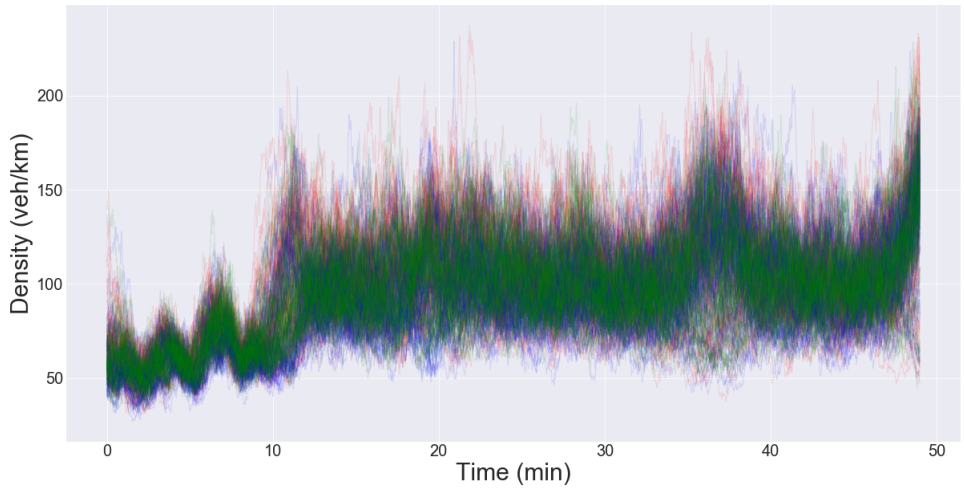

Figure 5.36: Inlet BC samples from FD and BC sampling for del Castillo's FD for M25 data. The 3 colours denote the 3 chains.

Figure 5.37: Trace plots for a typical outlet and inlet BC times (38min and 16min respectively) for FD and BC sampling for del Castillo's FD for M25 data. The third trace plot is from a multistart run: here we observe the same bimodality found in the sampling for $BC|FD$ in figure (4.41)

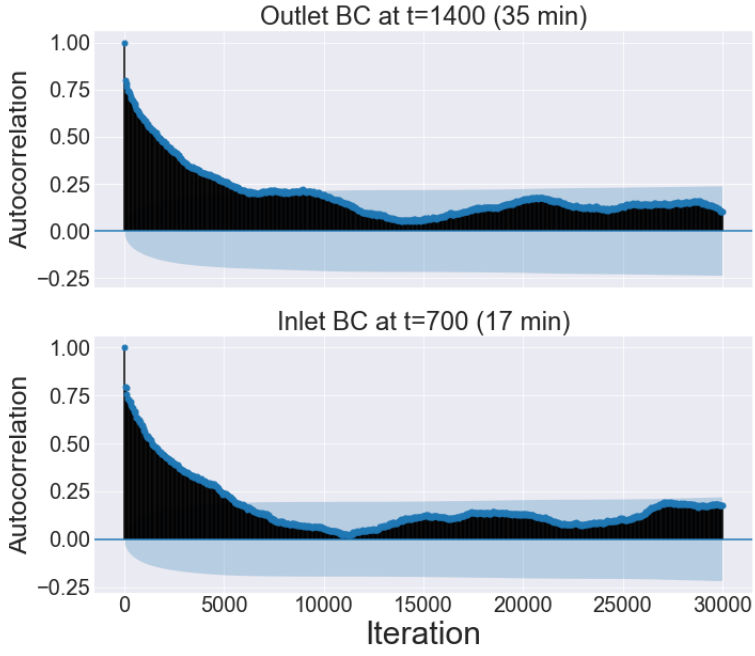

Figure 5.38: ACF for a typical outlet and inlet BC times for FD and BC sampling for del Castillo's FD for M25 data. The delay times for most BC times are between 1K and 3K.

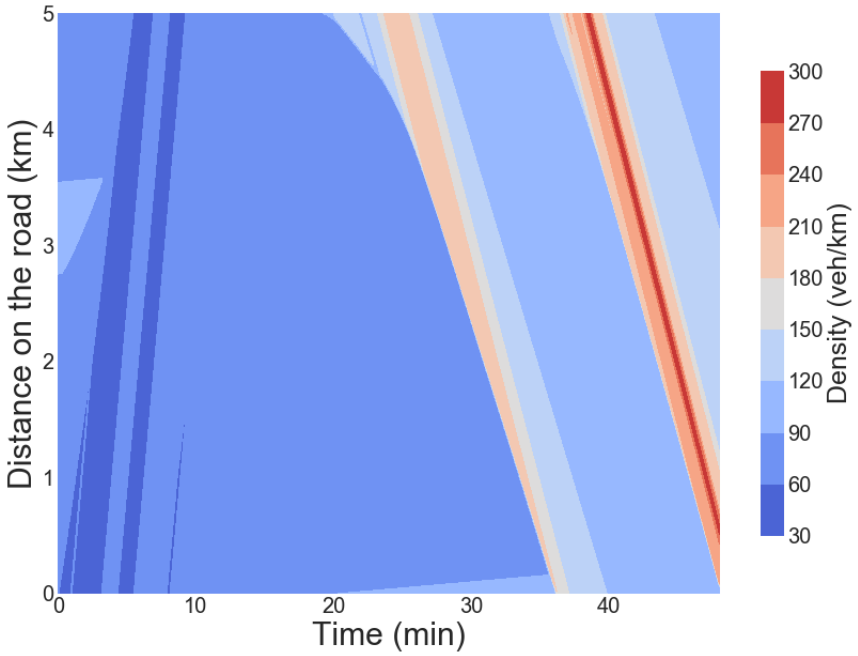

Figure 5.39: Output of LWR in the $x-t$ plane with posterior mean FD and BCs for FD and BC sampling for del Castillo's FD for M25 data. The wave speeds seem to agree with those found in M25 data (in figures (3.3) and (3.4)).

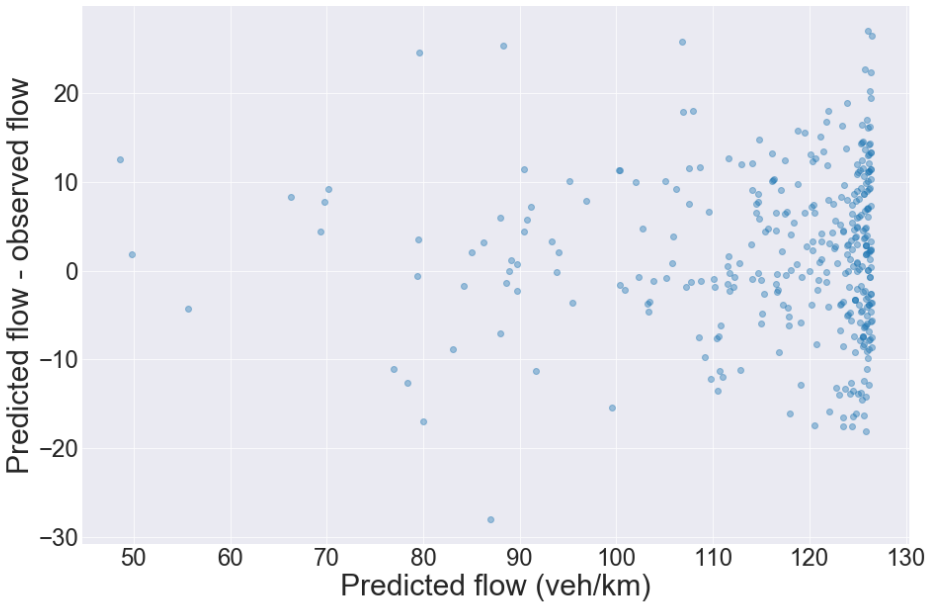

Figure 5.40: The residuals for FD and BC sampling for del Castillo's FD for M25 data show no apparent structure, which suggests a good model fit.

5.5 Joint move

In this section we analyse the correlations between the FD and BC parameters and use these to built a joint proposal. We test the proposal on both Exponential and del Castillo FD but find that it does not improve mixing.

5.5.1 Exponential FD

5.5.1.1 Analysis

To explore the relationship between the FD parameters and the BCs, we first take the mean of the logarithm of each BC. We plot this mean as a function of both FD parameters in figure (5.41). We notice a slight negative correlation for the α parameter.

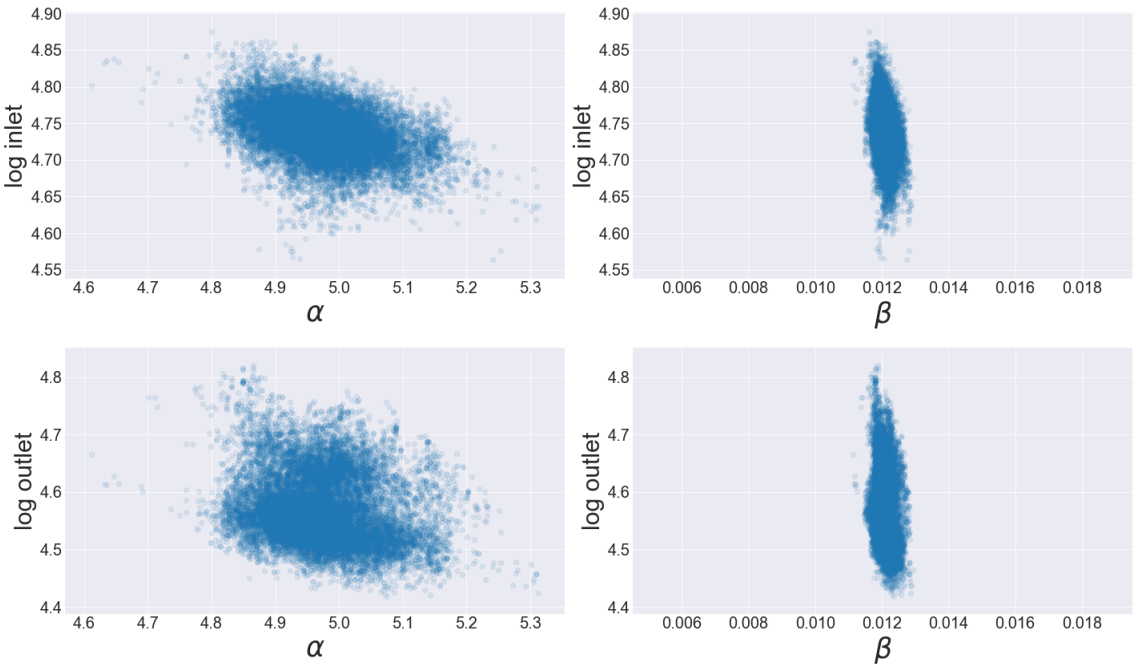

Figure 5.41: Relationship between the mean of log-BCs and the FD parameters. We notice a slight negative correlation between the α parameter and the BC (for both inlet and outlet BCs).

To understand this relationship more closely, we consider 2 sets of samples: samples where $\alpha > 5.15$ and samples where $\alpha < 4.82$. We plot the BCs for these two sets (red for the first set and blue for the second) in figure (5.42), and plot the FDs in figure (5.43).

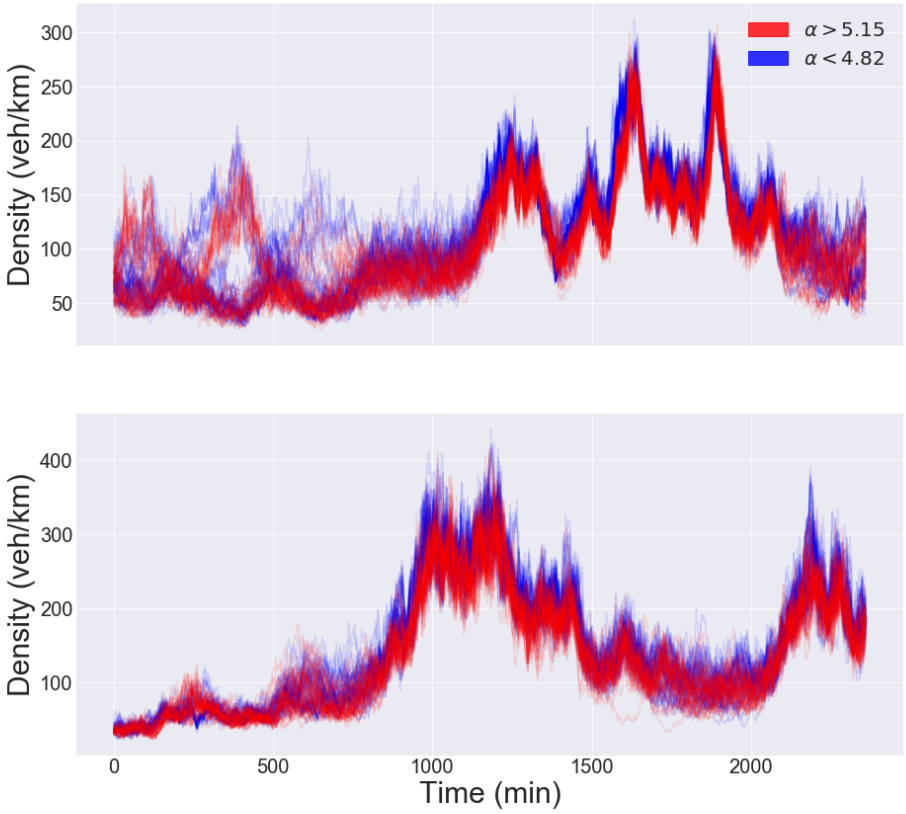

Figure 5.42: Inlet and outlet BCs for two sets of samples: in red are those samples where $\alpha > 5.15$ and in blue are the samples where $\alpha < 4.82$. The sets seem to be shifts vertically, especially for higher values of density

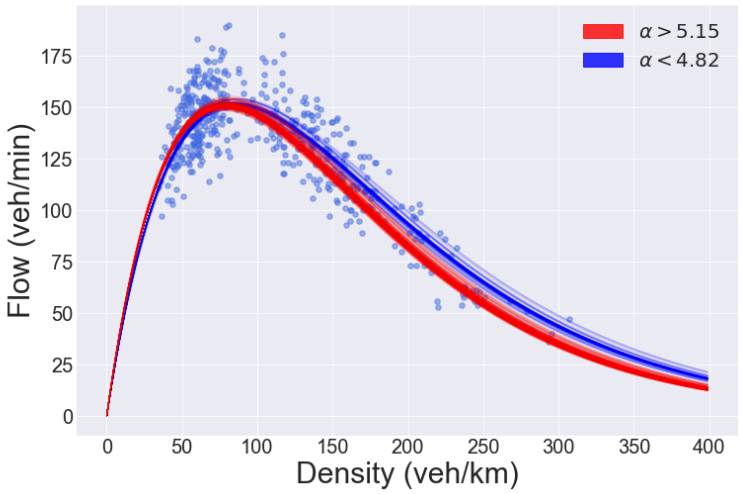

Figure 5.43: FDs samples along with simulated data for two sets of samples: in red are those samples where $\alpha > 5.15$ and in blue are the samples where $\alpha > 4.82$. The main difference in is for higher values of density.

To understand this relationship we recall the discussion about BC bimodality in chapter 4: as the likelihood is built from flow, values of density that map to the same flow under the FD are equally likely. Consider a FD shifting from the red to the blue in figure (5.43) and density in the congested regime; to keep a constant value of flow one needs to increase the density. This is consistent with the BCs in figure (5.42): the blue BCs are shifted upwards relative to the red ones. From figure (5.43) we can also see that this effect is less pronounced for the free flow regime which is consistent with the BC shifts in figure (5.42).

5.5.1.2 Joint move construction

This relationship suggests trying to build a proposal that updates the BCs and FDs together in the following way: propose new FD parameters using a Gaussian update, and simply shift the BC by a constant following a linear model. We assume a linear model for simplicity and because it could model the correlation in figure (5.41). Another way of thinking of this proposal is as a reparametrisation where one of the coordinates corresonds (very approximately) to flow; the proposal would then update the FD parameters while keeping flow constant. This would of course be a very coarse approximation, as we are using a linear model for the shift in BCs and as we are ignoring the more complicated interactions between the FD and the BCs due to LWR.

To define the joint move, we define $Y = \log(BC)$ to be the logarithm of the BCs (vectors of size L) and θ to be the vector of Fundamental Diagram parameters. We then assume a linear model for the mean μ as a function of θ and obtain

$$\psi(\theta) = \gamma_1^T \theta + \gamma_2$$

with $\gamma_1, \gamma_2 \in \mathscr{R}^2$

Using the notation $\overline{1}_N$ to denote a vector of ones of size N, the joint proposal for Y and θ is:

$$\begin{cases} \tilde{\theta} \sim \mathscr{N}(\theta, \Sigma_\theta) \\ \tilde{Y} = \phi_{\theta,\tilde{\theta}}(Y) \end{cases}$$

With

$$\begin{aligned} \phi_{\theta,\tilde{\theta}}(Y) &= Y + \psi(\tilde{\theta})\overline{1}_N - \psi(\theta)\overline{1}_N && (5.1) \\ &= Y + \gamma_1^T(\tilde{\theta} - \theta)\overline{1}_N && (5.2) \end{aligned}$$

Consider the following transformation: $T(\theta, Y) = (Z, R)$ with $Z = \theta$ and $R = Y + \gamma_1^T \theta \overline{1}_N$. In this coordinate system the proposal above is the Metropolis within Gibbs step $\tilde{z} \sim Z|R = r$, namely:

$$\begin{cases} \tilde{Z} \sim \mathscr{N}(Z, \Sigma_\theta) \\ \tilde{R} = R \end{cases}$$

We calculate the Jacobian of the transformation (use $\mathscr{I}_N$ to denote the identity matrix of size N). We define

$$|J| = \begin{vmatrix} \frac{\partial Z}{\partial \theta} & \frac{\partial Z}{\partial Y} \\ \frac{\partial R}{\partial \theta} & \frac{\partial R}{\partial Y} \end{vmatrix} \tag{5.3}$$

$$= \begin{vmatrix} \mathscr{I}_4 & 0 \\ \overline{1}_L \gamma_1^T & \mathscr{I}_L \end{vmatrix} \tag{5.4}$$

$$= 1 \tag{5.5}$$

So the proposal in the (θ, Y) coordinate system is:

$$q_{\theta,Y}(\tilde{\theta}, \tilde{Y}) = q_{Z,R}(\tilde{\theta}, \tilde{Y} + \gamma_1^T \theta \overline{1}_L)|J| \tag{5.6}$$

$$\sim \mathscr{N}(\theta, \Sigma_\theta) \tag{5.7}$$

Which is simply a Gaussian proposal (as the Jacobian is 1), so no Hastings correction is needed.

We then test the effect of the joint move but find no effect. The test is done by running two samplers, one with a BC shift and the other without. Both samplers use a parallel tempering sampler with 5 within-temperature moves and the same four inverse-temperatures as above (`[0.5, 0.63, 0.8, 1]`). A random scan Metropolis-within-Gibbs sampler is used with probabilities `[0.1, 0.2, 0.7]` for the FD move, joint move, and BC move respectively. The FD and BC moves are the same as for the mixing sampler described above. The joint move uses a Gaussian proposal with covariance matrix fit to the marginal posterior of the FD samples (the matrices for each temperature are reported in appendix (B.5.1.1)). The logarithm of the BCs were then shifted based on a linear model for α. They were shifted by a value of $\gamma = -0.243$ and $\gamma = -0.236$ for the outlet and inlet BC respectively (in equation (5.1)). The measure of the effect was MSEJD of the FD and acceptance rate for the joint move; no effect was found for the run with and without BC shift.

5.5.2 del Castillo's FD

5.5.2.1 Analysis

We now analyse the relationship between the BC samples and the FD parameters as we did for the Exponential FD in section (5.5.1.1). If we plot the relationship between the mean of the log BCs and the FD parameters in figure (5.44) and (5.45), we can see a positive linear relationship with the z and ρ_j parameters. This relationship continues for the extreme values of FD parameters as mentioned in the previous section, with the slight exception of the ω parameter in figure (5.45). If we take two sets of parameters (as with the Exponential FD) ($\rho_j > 420$ and $\rho_j < 380$) and plot the BCs we see a clear shift (figures (5.46)). The FDs for the same sets of parameters can be seen in figure (5.47): we see the shift in jam density ρ_j very clearly while keeping the congested flow wave speed constant. However the free flow wave speed does vary. We can also see two free flow branches for $\rho_j > 420$ (in red), which can be explained by seeing the disconnected region of parameter space in figure (5.45) for high values of z and ρ_j parameters. We can see from the plotted FDs that the posterior always keeps the congested flow wave speed fairly constant but allows the free flow wave speed to vary. As del Castillo's FD has more parameters than the Exponential FD and that these parameters allow distinct regions of the FD to change independently of others (such as the jam density ρ_j), it seems that there is more scope than for the Exponential FD to vary the FD and BCs together to keep flow approximately constant. As a result, the FD and BC parameters must move together on this approximately linear subspace (as hinted at by the more pronounced linear relationship in figures (5.44) and (5.45) and by the larger shift in figures (5.46) and (5.47)). This could explain why the sampler for this posterior exhibits some mixing problems.

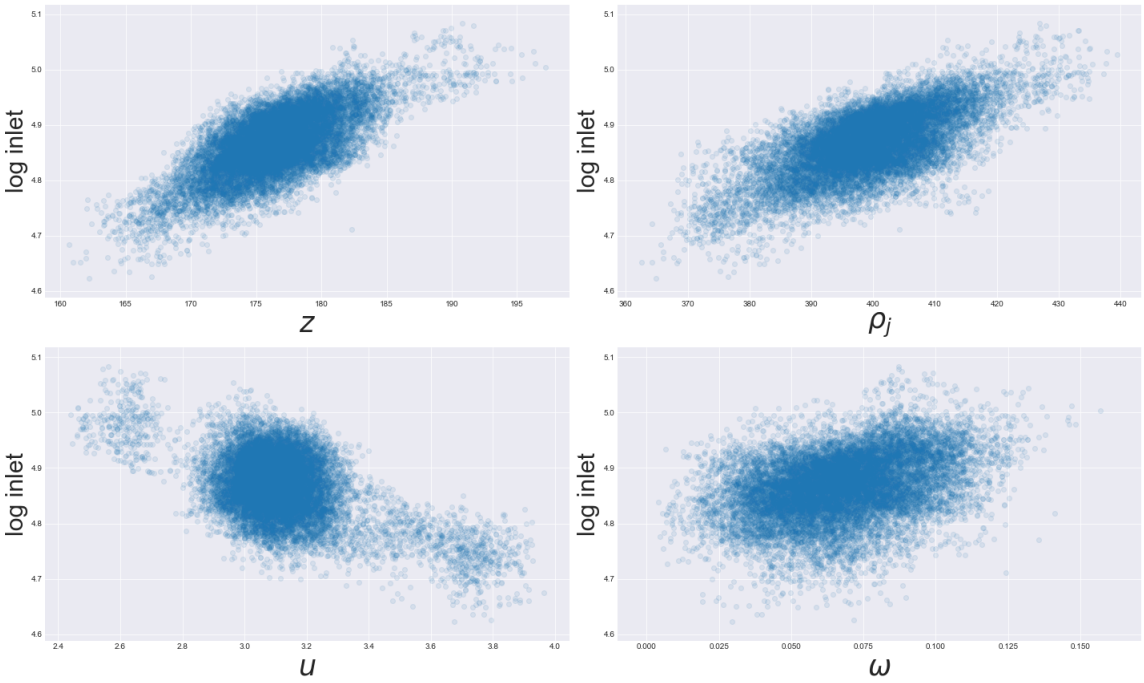

Figure 5.44: Relationship between the mean of log-inlet and del Castillo's parameters. We notice a positive correlation especially between the z and ρ_j parameter and the inlet BCs. We can see the excursions (the extreme value of parameters) in the z, ρ_j, and u parameters seen in figure (5.23). However these extreme still loosely follow the same relationship with the BC mean as the rest of the posterior.

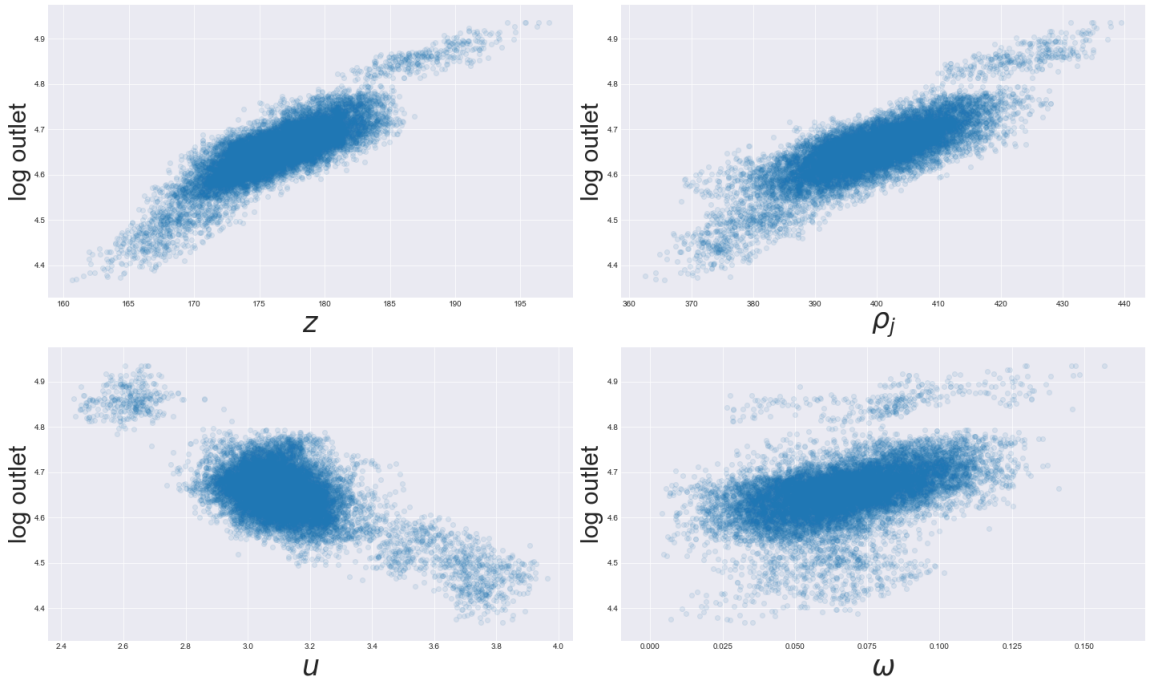

Figure 5.45: Relationship between the mean of log-outlet and del Castillo's parameters. As in figure (5.44), we notice a positive correlation especially between the z and ρ_j parameter and the outlet BCs. We can see the excursions (the extreme value of parameters) in the z, ρ_j, and u parameters seen in figure (5.23). However these extreme still loosely follow the same relationship with the BC mean as the rest of the posterior.

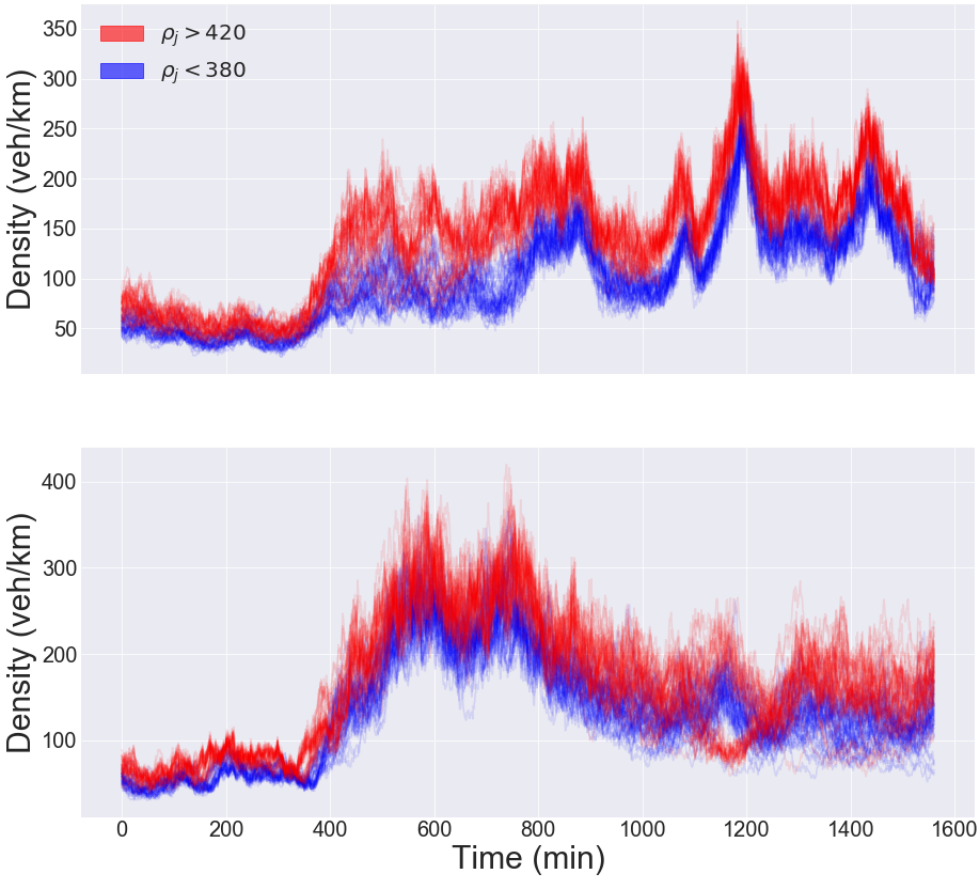

Figure 5.46: Inlet and outlet BCs for two sets of samples: in red are those samples where $\rho_j > 420$ and in blue are the samples where $\rho_j < 380$. The sets seem to be shifts vertically, especially for higher values of density

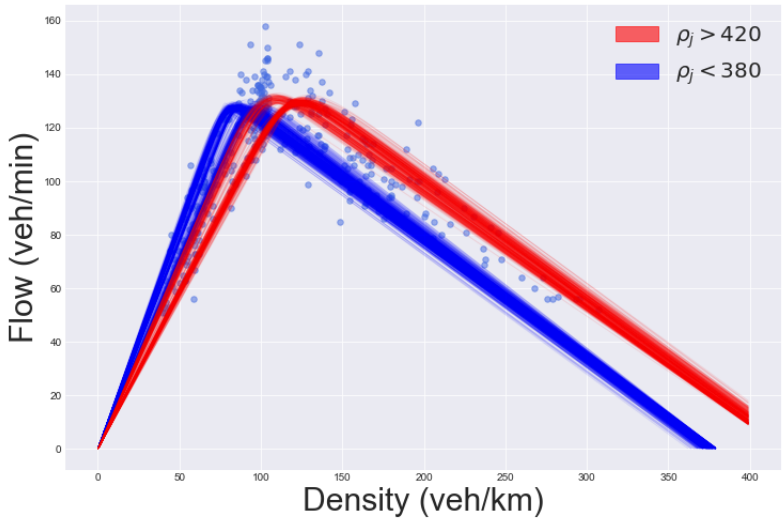

Figure 5.47: FDs samples along with simulated data for two sets of samples: in red are those samples where $\rho_j > 420$ and in blue are the samples where $\rho_j < 380$. The main difference between the two sets is in congested density.

As with the Exponential FD, we test the effect of a joint move (shift in log-BCs). We run a random scan Metropolis within Gibbbs sampler with move probabilities `[0.1, 0.3, 0.6]` for the FD, joint, and BC moves respectively. The FD and BC moves are the same the sampler above without the joint move. The covariance matrices for the joint move for each temperature are given in section (B.6.1.1), and a linear relationship between z, ρ_j and the mean of the log-BCs is fit. Defining $\gamma_1 = \begin{pmatrix} \gamma_z \\ \gamma_{\rho_j} \end{pmatrix}$ to be a vector of parameters in equation (5.1), we obtain a value of $\gamma_{1,outlet} = \begin{pmatrix} 0.00780 \\ 0.00222 \end{pmatrix}$ for the outlet BC and $\gamma_{1,inlet} = \begin{pmatrix} 0.00616 \\ 0.00136 \end{pmatrix}$. We fit this relationship for the chain with $\beta_{temp} = 0.56$. We run 3 chains of this sampler with a BC shift and 3 without a BC shift, but as in the case of the Exponential FD, we find no different in the acceptance rate of the joint move or the MSEJD of the FD parameters.

5.6 Conclusions

In this final chapter we joined together the sampling tools developed in the previous chapters and sample both FD and BC parameters using a Metropolis within Gibbs sam-

pler. We fit the Exponential and del Castillo FD for both simulated and M25 data. The Exponential FD sampler mixes but del Castillo's FD still has mixing problems. In both cases the FD and BC parameters are very correlated and the relationship between these is similar to the relationship in the advection equation between u and ρ_0 in chapter 2. We attempt to use this understanding of the correlations to propose a simple joint update for the FD and BC parameters, but we find that this proposal does not improve mixing in the studied datasets. Finally, we find that the traffic density estimated in the BCs is different to the density estimated using two engineering methods from the traffic flow literature, and that our method results in the correct congested flow wave speeds. When inserted into LWR, the BCs estimated using our new method provides a fit superior to that obtained from BCs using engineering methods.

Chapter 6

Conclusions and future work

6.1 Conclusions

After having given a critical review of microscopic and macroscopic modelling of traffic flow, we gave an overview of MCMC methodology for inverse problems. We constructed a sampler for estimating the wave speed and initial condition for the advection equation and experimented with different ways of tuning parallel tempering (number of within temperature moves along with the type of between-temperature proposals).

We then constructed a sampler for the fundamental diagram parameters as well as the boundary conditions (BCs) in the LWR model of traffic flow. Our sampler alternates between sampling BCs via a pre-conditioned Crank-Nicholson (pCN) method and fundamental diagram parameters using parallel tempering. As the advection equation and the LWR model of traffic flow are both conservation laws, the experimentation mentioned above allowed us to choose tuning parameters in parallel tempering for LWR. Sampling boundary conditions is further helped by blocking which allows the step size ω in pCN to be tuned to the variance of each BC section. The strong correlations between the FD and BCs along with the bimodality present in the BCs causes the sampling problem to be non-trivial, and in some cases the samplers exhibit mixing problems. We then test the developed samplers on real data from the M25 motorway and observe a good fit of LWR to data.

As work on Bayesian inverse problems remains rare in the literature, we have provided a unified statistical model to estimate both boundary conditions and fundamental diagram parameters while respecting the character of LWR as a conservation law. Furthermore, we compare 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 new method provides a fit superior to that obtained from BCs using engineering methods.

6.2 Future work

To further improve the performance of the samplers in LWR, one could use random subsampling of the detector times to build the likelihood. Using less data would increase the posterior variance while keeping a wide range of free flow and congested flow data to constrain the FD parameters (thus allow the different sections of the FD to be fit accurately).

Following the discussion of random and adjacent swaps in parallel tempering (in section (1.3.3.5)), we consider here how to capture the benefits of random swaps while avoiding the limitations. As mentioned in section (1.3.3.5), random swaps allow each chain to be coupled with a large number of other chains (at other temperatures) which might result in low correlations between consecutive swap proposals. However this method suffers from a low acceptance rate for swaps at widely spaced temperatures. An alternative is to still only allow adjacent temperature swaps but to introduce many chains at each inverse-temperature (with the exception of $\beta_{temp} = 1$) which would allow each chain to swap with many other chains. Denoting L the number of temperatures and M the *width* (namely the number of chains per temperature), we obtain $M(L-1)+1$ chains, and $M+M^2(L-2)$ swap combinations between chains. An interesting property of this modification is that increasing M causes the between-temperature moves to be more and more independent, which in the limit would become an independent sampler (with proposals independent from one another) thus causing a significant increase in mixing speed. Such an increase in mixing speed could be very useful in ensuring that the samplers developed in this thesis (such as sampling FD and BC parameters in LWR with del Castillo's FD) are able to efficiently explore the entire distribution including low-volume modes. Initial tests on the advection equation and the Exponential FD (sampling both FD and BC with M25 data) indeed show a substantial increase in mix-

ing speed, so investigating this idea further would be worthwhile. Another approach that can use parallelisation to speed up mixing would be to explore the use of particle filters. These have recently been used for high dimensional inverse problems and can be parallelised relatively easily (for example in [34]).

To introduce another modification on PT, recall that in our implementation we tempered only the likelihood so that BC swaps between temperatures would get accepted. This is because if we tempered the whole posterior the priors on the log-BCs (ie: an OU prior) for each temperature would be incompatible. However to avoid shrinkage it would be of interest to temper the whole posterior rather than just the likelihood. One would therefore need to modify log-BCs proposals between temperatures to be compatible with the prior of new tempered posterior. Some analytic work as well as experimental results would be needed to define this version of parallel tempering on function space.

Furthermore, the work on evaluating gradients in LWR with respect to initial conditions to use in iterative optimisation algorithms in [33] could be used to obtain gradients with respect to the boundary conditions. This could then be used in a gradient-based MCMC method such as a Metropolis adjusted Langevin sampler or a Hamiltonian Monte Carlo sampler. As seen in chapter 5, the mixing speed of the developed sampler for FD and BC parameters is slow (especially considering the computational cost of evaluating the likelihood). Developing gradient based sampling methods might allow more efficient exploration of this problem as well as be a general method for sampling from Bayesian inverse problem in the presence of discontinuities.

Finally, the correlation explored between the flux function parameters and BCs (in both LWR and the advection equation) seemed to be a promising avenue to develop a joint proposal, but ultimately failed to show improved mixing. Let us now view our attempted joint proposal from a slightly different angle from the one taken in chapter 5. The task was to predict a change in log(BCs) given a change in FD parameters (to obtain a sample from the conditional distribution of log(BCs) given FD parameters). We attempted to do this by expanding the log(BCs) in an orthonormal basis $\{\phi_j\}_1^\infty$ but truncating the expansion at a *single* basis element (namely the constant function). For each log(BC), a linear model was then fit between the FD parameters and the coefficient of this single basis element in a pre-processing step and used in a joint proposal (which

was ultimately unsuccessful).

Viewing the attempt in this way makes its failure to show improvement less surprising. It does however also suggest a way to improve it: include more basis elements in the truncation and perform multivariate analysis to predict the change in coefficients of the log(BCs) for a few basis elements given a change in FD parameters.

Another avenue of research following this approach would be to construct an ensemble sampler that extends the AIES or DE-MCMC samplers of section (1.3.3.3) to functional parameters. In this case, proposals such as the stretch move ($Y = X_j + Z(X_k(t) - X_j)$) would interpolate or extrapolate the coefficients of a truncated expansion of the functions (rather than interpolate and extrapolate the samples in a Euclidean basis). Similarly to the DE-MCMC proposal ($Y = X_k + \gamma(X_i - X_j) + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_N)$), one would however need to regularly add a random walk proposal (namely pCN) to ensure that the samples do not reside on a low dimensional subspace of the function space. Constructing such a sampler would leverage the attractive properties of these ensemble methods: straightforwards to tune, embarrassingly parallel, and practical for black-box forwards models such as PDE models. Analytic work would be needed to ensure such proposals can indeed be constructed to yield valid samplers on function space.

The sampling methodology developed in this thesis along with the samplers discussed above could be used to fit more sophisticated traffic flow models to data. Indeed, the second order models discussed and compared in section (1.1.2.5) could be fit rigorously to data. The qualitative comparison of these second order models in [44] (summarised and critically assessed in section (1.1.2.5)) could then be performed quantitatively after fitting the parameters and boundary conditions for each model. This would allow for a more rigorous assessment of the strengths and weaknesses of these models.

Appendix A

Prior reproduction test

Theorem A.0.1 (Prior reproduction test)**.** Define parameter $\theta \in \Theta$ and let X takes values in $\mathscr{R}^N$. The prior reproduction test algorithm is defined as follows:

1. Sample from the prior: $\theta_0 \sim \pi_0$
2. Sample data from the data model: $X \sim f_{X|\theta}(\cdot|\theta_0)$
3. Sample from the posterior $\theta_p \sim f_{\theta|X}(\cdot|X)$

   θ_p is then distributed according to the prior: $\theta_p \sim \pi_0$

Proof. Let $Z(x) := \int_\Theta f_{X|\theta}(x|\theta)\pi_0(\theta)d\theta$

- (θ_0, X) has joint density $f_{\theta_0,X}(\theta, x) = f_{X|\theta}(x|\theta)\pi_0(\theta)$
- (θ_0, X, θ_p) has joint density $f_{\theta_0,X,\theta_p}(\theta_0, x, \theta_p) = f_{\theta|X}(\theta_p|x) f_{X|\theta}(x|\theta_0)\pi_0(\theta_0)$, with $f_{\theta|X} = \frac{f_{X|\theta}(x|\theta)\pi_0(\theta)}{Z(x)}$

  We therefore have

$$
\begin{aligned}
f_{\theta_p}(\theta_p) &= \int_{\mathscr{R}^N} \int_\Theta f_{\theta_0,X,\theta_p}(\theta_0, x, \theta_p) d\theta_0 dx && \text{(A.1)} \\
&= \int_{\mathscr{R}^N} \int_\Theta f_{\theta|X}(\theta_p|x) f_{X|\theta}(x|\theta_0)\pi_0(\theta_0) d\theta_0 dx && \text{(A.2)} \\
&= \int_{\mathscr{R}^N} \frac{f_{X|\theta}(x|\theta_p)\pi_0(\theta_p)}{Z(x)} \int_\Theta f_{X|\theta}(x|\theta_0)\pi_0(\theta_0) d\theta_0 && \text{(A.3)} \\
&= \pi_0(\theta_p) \int_{\mathscr{R}^N} f_{X|\theta}(x|\theta_p) dx && \text{(A.4)} \\
&= \pi_0(\theta_p) && \text{(A.5)}
\end{aligned}
$$

□

Appendix B

MCMC implementation details

B.1 FD sampling: del Castillo's FD on simulated data

Covariance matrix for $\begin{pmatrix} z & \rho_j & u & w \end{pmatrix}^T$:

$$\begin{pmatrix} 0.62772279 & 0.36479498 & -0.01274923 & 0.00033383 \\ 0.36479498 & 4.38784781 & 0.02543674 & -0.0049363 \\ -0.01274923 & 0.02543674 & 0.00167817 & 0.00006093 \\ 0.00033383 & -0.0049363 & 0.00006093 & 0.00003466 \end{pmatrix}$$

B.2 FD sampling: del Castillo's FD on M25 data

Covariance for the direct fit to data for $\begin{pmatrix} z & \rho_j & u & w \end{pmatrix}^T$:

$$\begin{pmatrix} 193 & -299 & -2.480 & 1.288 \\ -299 & 589 & 5.743 & -1.787 \\ -2.480 & 5.743 & 0.092 & -0.008 \\ 1.288 & -1.787 & -0.008 & 0.009 \end{pmatrix}$$

Covariance matrix for sampling from LWR with parameters $\begin{pmatrix} z & \rho_j & u & w \end{pmatrix}^T$:

$$\begin{pmatrix} 1.46565905 & -0.43494473 & 0.04680924 & 0.00179531 \\ -0.43494473 & 5.61913019 & 0.01936478 & 0.0128512 \\ -0.04680924 & 0.01936478 & 0.00575797 & 0.00036424 \\ 0.00179531 & -0.0128512 & 0.00036424 & 0.00008159 \end{pmatrix}$$

B.3 BC sampling: Exponential FD

B.3.1 Exponential FD on Simulated data: Gibbs sampling with no Parallel Tempering

Table B.1: BC Gibbs blocks for sampling $BC|FD$ for the Exponential FD on simulated data with no parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	60	0.07	17.9, 24.0, 22.9
inlet	0	22	0.15	16.5, 16.8, 12.2
inlet	20	37	0.13	29.6, 31.4, 21.6
inlet	35	52	0.37	22.8, 19.4, 20.7
inlet	50	60	0.26	18.0, 14.8, 17.7
outlet	0	60	0.02	30.4, 30.9, 24.9
outlet	0	29	0.33	20.9, 20.4, 11.8
outlet	27	40	0.1	13.3, 17.2, 11.9
outlet	47	60	0.15	22.5, 18.2, 23.4
outlet	19	30	0.6	21.7, 31.5, 26.0
outlet	38	45	0.07	29.7, 17.2, 21.8
outlet	42	50	0.2	6.1, 7.2, 7.6
outlet	0	15	0.45	25.2, 28.2, 18.0
outlet	10	25	0.5	26.5, 33.2, 21.1
inlet	10	22	0.35	18.5, 20.7, 12.5
inlet	37	52	0.7	9.8, 8.3, 8.5

B.3.2 Exponential FD on Simulated data: Parallel Tempering

Table B.2: BC Gibbs blocks for $\beta_{temp} = 1$ for sampling $BC|FD$ for the Exponential FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	60	0.07	17.0, 20.9, 20.0
inlet	0	22	0.15	19.2, 17.6, 14.2
inlet	20	37	0.13	30.3, 29.0, 23.6
inlet	35	52	0.37	18.0, 23.2, 20.5
inlet	50	60	0.26	19.4, 17.3, 21.1
outlet	0	60	0.02	25.5, 27.5, 24.4
outlet	0	29	0.33	22.7, 19.2, 19.5
outlet	27	40	0.08	26.3, 27.5, 25.9
outlet	47	60	0.15	17.1, 18.9, 20.8
outlet	19	30	0.6	28.2, 26.8, 29.6
outlet	38	45	0.07	20.8, 21.1, 20.5
outlet	42	50	0.1	25.6, 27.6, 27.9
inlet	10	22	0.35	20.0, 20.2, 20.7
inlet	37	52	0.58	11.3, 15.7, 16.3
outlet	0	23	0.95	0.0, 0.0, 1.0

Table B.3: BC Gibbs blocks for $\beta_{temp} = 0.8$ for sampling $BC|FD$ for the Exponential FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	60	0.08	23.3, 24.2, 26.7
inlet	0	22	0.18	22.2, 20.4, 17.8
inlet	20	37	0.19	17.2, 14.7, 17.7
inlet	35	52	0.4	22.4, 23.5, 20.8
inlet	50	60	0.29	17.9, 19.2, 17.5
outlet	0	60	0.034	21.0, 21.2, 19.5
outlet	0	29	0.36	22.9, 19.0, 22.4
outlet	27	40	0.09	25.5, 23.5, 25.4
outlet	47	60	0.16	20.6, 19.6, 20.8
outlet	10	32	0.35	12.6, 12.6, 12.5
outlet	38	45	0.07	30.3, 30.5, 32.2
outlet	42	50	0.13	21.3, 21.2, 21.2
inlet	10	22	0.39	19.9, 22.4, 20.6
inlet	37	52	0.7	12.3, 10.8, 13.2
outlet	0	23	0.95	0.3, 0.3, 0.6

Table B.4: BC Gibbs blocks for $\beta_{temp} = 0.63$ for sampling $BC|FD$ for the Exponential FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	60	0.09	22.9, 27.8, 25.3
inlet	0	22	0.2	27.3, 25.3, 19.5
inlet	20	37	0.2	22.1, 20.0, 20.3
inlet	35	52	0.43	25.6, 26.5, 25.6
inlet	50	60	0.31	18.9, 22.6, 19.6
outlet	0	60	0.04	20.2, 20.5, 24.4
outlet	0	29	0.39	27.4, 28.3, 29.8
outlet	27	40	0.1	28.0, 27.7, 26.7
outlet	47	60	0.17	25.0, 22.4, 24.6
outlet	10	32	0.35	15.8, 12.0, 13.7
outlet	38	45	0.07	36.7, 39.4, 43.9
outlet	42	50	0.13	26.3, 26.6, 26.9
inlet	10	22	0.45	27.0, 27.2, 25.8
inlet	37	52	0.7	18.5, 17.5, 15.0
outlet	0	23	0.95	1.6, 1.4, 2.4

Table B.5: BC Gibbs blocks for $\beta_{temp} = 0.5$ for sampling $BC|FD$ for the Exponential FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	60	0.1	28.6, 27.3, 25.8
inlet	0	22	0.25	25.9, 24.3, 25.7
inlet	20	37	0.15	36.8, 39.1, 39.5
inlet	35	52	0.5	24.2, 20.8, 24.6
inlet	50	60	0.33	24.2, 23.3, 20.1
inlet	37	52	0.9	13.9, 13.7, 11.2
outlet	0	60	0.04	27.2, 24.6, 30.8
outlet	0	30	0.5	18.6, 18.8, 17.0
outlet	27	40	0.13	23.3, 22.6, 25.9
outlet	46	60	0.17	18.5, 18.6, 19.4
outlet	38	47	0.13	24.6, 22.9, 24.1
outlet	41	50	0.15	18.1, 18.7, 16.0
outlet	0	23	0.95	4.5, 5.7, 4.9

B.3.3 Exponential FD on M25 data: Parallel Tempering

Table B.6: BC Gibbs blocks for $\beta_{temp} = 1$ for sampling $BC|FD$ for the Exponential FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	49	0.125	23.7, 22.9, 23.1
inlet	0	15	0.48	18.9, 19.3, 19.1
inlet	7	34	0.28	29.0, 30.2, 28.9
inlet	30	40	0.6	47.4, 48.3, 42.7
inlet	40	49	0.4	10.7, 11.6, 14.2
outlet	0	49	0.06	12.2, 11.3, 10.8
outlet	0	20	0.44	16.6, 18.5, 18.2
outlet	12	24	0.38	12.9, 10.9, 10.5
outlet	29	41	0.1	14.0, 16.4, 16.0
outlet	25	35	0.11	26.9, 26.8, 24.1
outlet	33	39	0.15	23.0, 27.7, 20.3
outlet	22	30	0.19	24.4, 23.3, 19.5
inlet	0	10	0.8	5.8, 8.3, 7.8
outlet	18	31	0.9	0.0, 0.0, 0.0
outlet	37	49	0.7	19.1, 21.3, 21.1

Table B.7: BC Gibbs blocks for $\beta_{temp} = 0.76$ for sampling $BC|FD$ for the Exponential FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	49	0.14	20.8, 23.5, 27.0
inlet	0	15	0.5	21.1, 24.4, 20.1
inlet	7	34	0.32	26.6, 27.6, 24.6
inlet	30	44	0.64	28.5, 31.4, 27.1
inlet	40	49	0.44	15.3, 13.7, 15.8
outlet	0	49	0.067	13.7, 11.8, 12.6
outlet	0	20	0.5	17.3, 23.0, 22.5
outlet	12	24	0.45	16.8, 12.3, 13.9
outlet	28	39	0.14	11.0, 12.7, 17.0
outlet	25	35	0.14	23.5, 25.3, 25.2
outlet	33	39	0.2	23.7, 21.3, 24.9
outlet	20	30	0.26	11.5, 8.2, 14.3
inlet	0	10	0.8	14.9, 15.6, 15.6
inlet	42	49	0.7	5.7, 7.3, 6.4
outlet	18	31	0.9	0.2, 0.0, 0.5
outlet	37	49	0.79	24.2, 24.0, 24.0

Table B.8: BC Gibbs blocks for $\beta_{temp} = 0.58$ for sampling $BC|FD$ for the Exponential FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	49	0.18	23.5, 22.9, 23.9
inlet	0	15	0.53	22.5, 28.3, 23.1
inlet	7	34	0.39	23.6, 25.7, 27.4
inlet	30	44	0.73	32.7, 29.7, 31.2
inlet	40	49	0.5	16.9, 18.2, 16.5
outlet	0	49	0.073	15.6, 16.0, 15.6
outlet	0	20	0.59	22.6, 24.3, 21.5
outlet	8	25	0.45	14.8, 11.7, 14.6
outlet	23	39	0.09	22.0, 21.0, 21.8
outlet	25	35	0.18	26.4, 23.6, 25.6
outlet	33	39	0.26	20.8, 16.5, 21.7
outlet	20	30	0.28	11.8, 12.6, 13.1
inlet	0	10	0.8	22.2, 21.4, 20.0
inlet	42	49	0.7	12.9, 12.7, 13.1
outlet	18	31	0.9	1.6, 0.0, 0.5
outlet	37	49	0.85	29.4, 31.4, 28.6

B.4 BC sampling: del Castillo FD

B.4.1 del Castillo's FD on simulated data: Parallel Tempering

Table B.9: BC Gibbs blocks for $\beta_{temp} = 1$ for sampling $BC|FD$ for del Castillo's FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	60	0.07	9.0, 13.1, 10.5
inlet	0	22	0.08	15.7, 12.0, 12.4
inlet	20	37	0.16	27.8, 26.3, 27.3
inlet	35	52	0.3	24.6, 28.2, 24.1
inlet	50	60	0.3	26.8, 24.5, 32.4
outlet	0	60	0.05	22.0, 21.7, 24.5
outlet	0	29	0.2	13.4, 10.9, 9.6
outlet	27	40	0.1	22.0, 23.5, 30.9
outlet	36	49	0.13	15.2, 16.8, 19.4
outlet	47	60	0.2	25.0, 24.2, 16.8
outlet	22	30	0.45	36.5, 26.4, 26.5
inlet	37	49	0.45	27.4, 24.8, 17.7
inlet	19	25	0.7	7.0, 6.4, 5.1

Table B.10: BC Gibbs blocks for $\beta_{temp} = 0.8$ for sampling $BC|FD$ for del Castillo's FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	60	0.075	14.0, 13.4, 17.7
inlet	0	22	0.09	11.5, 18.8, 17.9
inlet	19	37	0.19	18.6, 19.2, 15.9
inlet	35	52	0.36	27.2, 21.7, 29.9
inlet	50	60	0.34	23.6, 29.4, 23.7
outlet	0	60	0.065	14.7, 17.5, 19.1
outlet	0	29	0.2	13.8, 18.2, 19.4
outlet	27	40	0.18	16.7, 16.1, 16.2
outlet	36	49	0.15	18.3, 15.4, 16.9
outlet	47	60	0.24	22.1, 19.7, 21.2
inlet	37	49	0.47	27.6, 25.2, 23.9
inlet	19	25	0.8	4.9, 4.6, 5.1

Table B.11: BC Gibbs blocks for $\beta_{temp} = 0.63$ for sampling $BC|FD$ for del Castillo's FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	60	0.08	13.6, 15.5, 16.8
inlet	0	22	0.09	21.5, 16.9, 18.8
inlet	19	37	0.22	15.3, 19.4, 19.7
inlet	35	52	0.43	27.5, 27.1, 26.8
inlet	50	60	0.39	24.6, 28.1, 20.8
outlet	0	60	0.08	13.9, 15.1, 18.7
outlet	0	29	0.19	24.7, 25.4, 23.5
outlet	27	40	0.2	14.2, 15.0, 15.7
outlet	33	49	0.15	17.4, 20.7, 17.9
outlet	47	60	0.26	23.4, 27.8, 28.8
inlet	37	49	0.57	21.7, 24.9, 24.7
inlet	19	25	0.9	3.3, 3.3, 5.1

Table B.12: BC Gibbs blocks for $\beta_{temp} = 0.5$ for sampling $BC|FD$ for del Castillo's FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	60	0.11	14.8, 10.5, 12.6
inlet	0	22	0.1	21.2, 22.5, 22.3
inlet	19	37	0.25	17.1, 17.4, 17.3
inlet	35	52	0.53	19.6, 28.5, 25.8
inlet	50	60	0.46	21.3, 26.0, 23.2
outlet	0	60	0.1	17.2, 16.4, 14.3
outlet	0	29	0.2	25.1, 21.8, 24.8
outlet	27	40	0.2	24.9, 22.3, 23.0
outlet	34	49	0.17	14.0, 18.0, 20.1
outlet	47	60	0.3	25.3, 26.7, 29.7
inlet	37	49	0.6	31.0, 26.6, 26.1
inlet	19	25	0.95	3.5, 6.0, 3.0

B.4.2 del Castillo's FD on M25 data: Parallel Tempering

Table B.13: BC Gibbs blocks for $\beta_{temp} = 1$ for sampling $BC|FD$ for del Castillo's FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	49	0.07	22.6, 18.9, 18.7
inlet	0	15	0.15	12.1, 10.9, 11.0
inlet	7	35	0.13	23.9, 28.3, 28.5
inlet	33	49	0.4	16.8, 17.4, 21.1
outlet	0	49	0.04	23.7, 24.6, 23.6
outlet	0	15	0.25	16.0, 16.4, 16.6
outlet	12	24	0.21	9.5, 13.1, 10.1
outlet	25	39	0.16	23.6, 23.6, 22.7
outlet	29	49	0.12	22.1, 17.8, 21.0
outlet	6	18	0.4	10.6, 15.0, 18.5
outlet	39	49	0.4	26.8, 23.5, 27.3
inlet	0	12	0.27	3.8, 4.7, 3.2
inlet	20	36	0.25	28.5, 23.9, 27.3
outlet	2	12	0.9	2.7, 2.4, 5.2

Table B.14: BC Gibbs blocks for $\beta_{temp} = 0.76$ for sampling $BC|FD$ for del Castillo's FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	49	0.1	16.0, 19.9, 19.4
inlet	0	15	0.2	15.5, 12.8, 10.4
inlet	7	34	0.18	23.7, 21.8, 21.4
inlet	34	49	0.44	18.3, 24.1, 22.7
outlet	0	49	0.055	19.0, 17.2, 19.4
outlet	0	15	0.3	21.8, 19.7, 18.8
outlet	12	24	0.21	21.3, 22.2, 19.3
outlet	25	39	0.2	20.5, 23.0, 21.1
outlet	29	49	0.14	21.5, 22.0, 22.2
outlet	6	18	0.45	19.8, 17.0, 14.8
outlet	38	49	0.48	23.2, 21.6, 21.1
inlet	0	12	0.3	7.2, 7.5, 3.5
inlet	20	35	0.31	25.3, 25.4, 24.2
inlet	5	15	0.9	2.9, 3.2, 1.5
outlet	2	12	1	4.2, 2.1, 2.5

Table B.15: BC Gibbs blocks for $\beta_{temp} = 0.58$ for sampling $BC|FD$ for del Castillo's FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	49	0.11	28.9, 21.1, 17.2
inlet	0	15	0.21	22.1, 16.3, 16.0
inlet	7	34	0.2	24.4, 26.8, 30.8
inlet	34	49	0.47	23.6, 27.6, 31.1
outlet	0	49	0.06	22.4, 25.8, 23.1
outlet	0	15	0.39	20.8, 21.6, 17.7
outlet	12	24	0.26	23.6, 23.6, 22.2
outlet	25	39	0.2	26.6, 27.7, 27.5
outlet	29	49	0.14	24.2, 25.6, 24.4
outlet	6	18	0.5	30.2, 25.1, 22.1
outlet	38	49	0.51	24.6, 20.7, 22.8
inlet	0	12	0.4	5.4, 7.3, 8.7
inlet	20	35	0.36	25.3, 26.2, 25.7
inlet	5	15	0.9	6.3, 4.3, 3.9
outlet	3	12	1	14.0, 10.4, 11.8

B.5 FD and BC sampling: Exponential FD

B.5.1 Exponential FD on Simulated data

B.5.1.1 Covariance matrices

Covariance matrices for $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ at each temperature:

- For $\beta_{temp} = 1$: $\begin{pmatrix} 0.0003865 & 0.00000088 \\ 0.00000088 & 3.269999e-09 \end{pmatrix}$
- For $\beta_{temp} = 0.8$: $\begin{pmatrix} 0.00036769 & 0.00000095 \\ 0.00000095 & 3.63961760e-09 \end{pmatrix}$

- For $\beta_{temp} = 0.63$: $\begin{pmatrix} 0.00043188 & 0.00000124 \\ 0.00000124 & 4.9803e-09 \end{pmatrix}$

- For $\beta_{temp} = 0.5$: $\begin{pmatrix} 0.00053368 & 0.00000144 \\ 0.00000144 & 0.00000001 \end{pmatrix}$

Covariance matrices for the joint move for each temperature:

- For $\beta_{temp} = 1$: $\begin{pmatrix} 4.2899475e-03 & 1.0665000e-05 \\ 1.0665000e-05 & 3.0000000e-08 \end{pmatrix}$

- For $\beta_{temp} = 0.8$: $\begin{pmatrix} 5.71993e-03 & 1.42200e-05 \\ 1.42200e-05 & 4.00000e-08 \end{pmatrix}$

- For $\beta_{temp} = 0.63$: $\begin{pmatrix} 7.1499125e-03 & 1.7775000e-05 \\ 1.7775000e-05 & 5.0000000e-08 \end{pmatrix}$

- For $\beta_{temp} = 0.5$: $\begin{pmatrix} 8.579895e-03 & 2.133000e-05 \\ 2.133000e-05 & 6.000000e-08 \end{pmatrix}$

B.5.1.2 Boundary condition blocks:

Table B.16: BC Gibbs blocks for $\beta_{temp} = 1$ for sampling FD and BC parameters for the Exponential FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	Acceptance probabilities (%)
inlet	0	60	0.07	22.3, 22.2, 23.4
inlet	0	22	0.15	25.1, 24.2, 19.5
inlet	20	37	0.13	29.1, 26.2, 26.3
inlet	35	52	0.37	23.8, 20.1, 20.1
inlet	50	60	0.26	16.6, 17.0, 18.5
outlet	0	60	0.02	33.3, 35.4, 31.4
outlet	0	29	0.33	19.6, 17.8, 21.1
outlet	27	40	0.08	27.8, 26.2, 22.1
outlet	47	60	0.15	24.7, 21.4, 26.0
outlet	19	30	0.6	31.1, 29.8, 30.3
outlet	38	45	0.07	25.0, 22.1, 27.6
outlet	42	50	0.1	27.3, 25.7, 27.2
inlet	10	22	0.35	22.0, 24.1, 22.0
inlet	37	52	0.58	18.4, 16.2, 18.5
outlet	0	23	0.95	0.9, 0.6, 0.4

Table B.17: BC Gibbs blocks for $\beta_{temp} = 0.8$ for sampling FD and BC parameters for the Exponential FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	Acceptance probabilities (%)
inlet	0	60	0.08	28.1, 20.7, 22.7
inlet	0	22	0.18	25.2, 29.9, 27.2
inlet	20	37	0.19	16.5, 17.6, 18.2
inlet	35	52	0.4	29.7, 27.1, 25.8
inlet	50	60	0.29	20.2, 21.1, 18.2
outlet	0	60	0.034	21.3, 21.6, 20.7
outlet	0	29	0.36	26.4, 27.3, 19.0
outlet	27	40	0.09	31.8, 24.7, 29.9
outlet	47	60	0.16	26.8, 27.8, 26.5
outlet	10	32	0.35	15.4, 9.5, 15.1
outlet	38	45	0.07	35.7, 35.3, 34.1
outlet	42	50	0.13	20.2, 23.9, 21.5
inlet	10	22	0.39	26.5, 31.5, 28.2
inlet	37	52	0.7	14.7, 16.2, 16.7
outlet	0	23	0.95	1.7, 2.1, 1.2

Table B.18: BC Gibbs blocks for $\beta_{temp} = 0.63$ for sampling FD and BC parameters for the Exponential FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	Acceptance probabilities (%)
inlet	0	60	0.09	27.7, 29.8, 27.5
inlet	0	22	0.2	30.7, 24.1, 36.4
inlet	20	37	0.2	20.4, 20.5, 23.1
inlet	35	52	0.43	25.1, 24.5, 23.9
inlet	50	60	0.31	18.8, 19.5, 20.6
outlet	0	60	0.04	26.1, 29.4, 23.8
outlet	0	29	0.39	27.4, 26.3, 25.3
outlet	27	40	0.1	24.4, 27.1, 30.0
outlet	47	60	0.17	28.8, 28.2, 24.3
outlet	10	32	0.35	16.0, 16.0, 14.6
outlet	38	45	0.07	43.1, 38.1, 43.7
outlet	42	50	0.13	26.4, 25.8, 30.2
inlet	10	22	0.45	36.3, 39.2, 30.9
inlet	37	52	0.7	18.1, 17.2, 22.9
outlet	0	23	0.95	3.0, 5.5, 3.6

Table B.19: BC Gibbs blocks for $\beta_{temp} = 0.5$ for sampling FD and BC parameters for the Exponential FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	Acceptance probabilities (%)
inlet	0	60	0.1	33.3, 29.5, 26.8
inlet	0	22	0.25	31.0, 27.4, 34.6
inlet	20	37	0.15	37.4, 40.8, 39.0
inlet	35	52	0.5	32.1, 30.1, 24.6
inlet	50	60	0.33	17.3, 20.4, 20.5
inlet	37	52	0.9	12.6, 13.3, 20.4
outlet	0	60	0.04	28.5, 28.5, 29.4
outlet	0	30	0.5	21.7, 19.5, 18.0
outlet	27	40	0.13	24.6, 20.9, 23.8
outlet	46	60	0.17	23.0, 23.5, 20.8
outlet	38	47	0.13	27.9, 23.4, 23.7
outlet	41	50	0.15	22.0, 23.4, 20.1
outlet	0	23	0.95	6.2, 9.4, 6.8

B.5.2 Exponential FD on M25 data

B.5.2.1 Boundary condition blocks:

Table B.20: BC Gibbs blocks for $\beta_{temp} = 1$ for sampling FD and BC parameters for the Exponential FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	Acceptance probabilities (%)
inlet	0	49	0.129	23.0, 17.2, 18.3
inlet	0	15	0.19	17.8, 17.6, 18.1
inlet	7	34	0.29	24.4, 19.9, 23.4
inlet	30	49	0.52	20.6, 18.6, 21.0
outlet	0	49	0.044	17.3, 13.8, 13.8
outlet	0	20	0.4	16.7, 18.2, 12.2
outlet	8	25	0.24	19.5, 16.9, 21.3
outlet	21	29	0.15	20.2, 23.1, 19.4
outlet	25	39	0.2	23.2, 25.8, 21.7
outlet	35	42	0.1	22.6, 20.3, 18.9
outlet	37	49	0.09	22.9, 19.0, 22.7
outlet	0	17	1	0.0, 0.0, 0.0

Table B.21: BC Gibbs blocks for $\beta_{temp} = 0.76$ for sampling FD and BC parameters for the Exponential FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	Acceptance probabilities (%)
inlet	0	49	0.14	24.3, 21.9, 21.0
inlet	0	15	0.26	15.5, 16.1, 15.5
inlet	7	34	0.32	23.2, 21.6, 28.6
inlet	30	49	0.55	25.8, 23.5, 29.5
outlet	0	49	0.055	14.1, 11.0, 14.3
outlet	0	20	0.45	18.6, 12.7, 20.9
outlet	8	25	0.3	12.2, 13.7, 14.1
outlet	21	29	0.17	19.6, 19.1, 14.6
outlet	25	39	0.24	20.9, 22.4, 17.0
outlet	35	42	0.11	22.9, 22.1, 20.9
outlet	37	49	0.11	26.1, 26.7, 28.5
outlet	0	17	1	0.0, 0.9, 0.4
inlet	1	12	1	0.0, 0.0, 0.0

Table B.22: BC Gibbs blocks for $\beta_{temp} = 0.58$ for sampling FD and BC parameters for the Exponential FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	Acceptance probabilities (%)
inlet	0	49	0.18	24.4, 24.4, 23.2
inlet	0	15	0.3	27.5, 17.6, 24.3
inlet	7	34	0.37	25.2, 22.9, 36.7
inlet	30	49	0.64	31.4, 22.3, 29.7
outlet	0	49	0.069	15.5, 9.2, 14.7
outlet	0	20	0.52	16.3, 13.6, 19.1
outlet	8	25	0.33	23.1, 19.2, 9.6
outlet	21	29	0.18	25.9, 22.1, 18.3
outlet	25	39	0.25	23.3, 28.5, 22.0
outlet	35	42	0.17	22.3, 17.9, 14.7
outlet	37	49	0.15	22.9, 20.9, 16.9
outlet	0	17	1	0.0, 0.8, 3.8
inlet	1	10	1	3.1, 7.1, 4.2
outlet	0	17	1	0.0, 0.8, 3.6

Table B.23: BC Gibbs blocks for $\beta_{temp} = 0.44$ for sampling FD and BC parameters for the Exponential FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	Acceptance probabilities (%)
inlet	0	49	0.21	29.6, 29.7, 25.4
inlet	0	15	0.4	14.5, 22.4, 20.0
inlet	7	34	0.43	30.7, 30.9, 23.5
inlet	30	49	0.78	22.1, 22.4, 24.4
outlet	0	49	0.072	20.1, 15.9, 17.8
outlet	0	20	0.59	20.6, 16.7, 13.5
outlet	8	25	0.45	15.4, 17.5, 14.4
outlet	21	29	0.23	21.5, 25.4, 26.6
outlet	25	39	0.29	24.6, 26.2, 24.5
outlet	35	42	0.2	18.4, 19.1, 22.1
outlet	37	49	0.2	21.3, 25.2, 22.7
outlet	0	17	1	3.8, 2.1, 3.8
inlet	1	10	1	4.7, 6.3, 5.1
outlet	0	17	1	4.1, 3.5, 3.5

B.6 FD and BC sampling: del Castillo's FD

B.6.1 Del Castillo FD on simulated data

B.6.1.1 Covariance matrices

Covariance matrices for $\begin{pmatrix} z & \rho_j & u & w \end{pmatrix}^T$ for each temperature:

- For $\beta_{temp} = 1$:

$$\begin{pmatrix} 0.62772279 & 0.36479498 & -0.01274923 & 0.00033383 \\ 0.36479498 & 4.38784781 & 0.02543674 & -0.0049363 \\ -0.01274923 & 0.02543674 & 0.00167817 & 0.00006093 \\ 0.00033383 & -0.0049363 & 0.00006093 & 0.00003466 \end{pmatrix}$$

- For $\beta_{temp} = 0.75$:
$$\begin{pmatrix} 0.83675448 & 0.48627171 & -0.01699472 & 0.000445 \\ 0.48627171 & 5.84900113 & 0.03390717 & -0.00658009 \\ -0.01699472 & 0.03390717 & 0.002237 & 0.00008122 \\ 0.000445 & -0.00658009 & 0.00008122 & 0.0000462 \end{pmatrix}$$

- For $\beta_{temp} = 0.56$:
$$\begin{pmatrix} 1.75554261 & 0.89897369 & -0.03449611 & 0.00309411 \\ 0.89897369 & 13.55393024 & 0.04391783 & -0.01255817 \\ -0.03449611 & 0.04391783 & 0.00333868 & 0.00009115 \\ 0.00309411 & -0.01255817 & 0.00009115 & 0.00010792 \end{pmatrix}$$

- For $\beta_{temp} = 0.42$:
$$\begin{pmatrix} 2.34072348 & 1.19863159 & -0.04599482 & 0.00412548 \\ 1.19863159 & 18.07190699 & 0.0585571 & -0.01674422 \\ -0.04599482 & 0.0585571 & 0.00445158 & 0.00012154 \\ 0.00412548 & -0.01674422 & 0.00012154 & 0.00014389 \end{pmatrix}$$

Covariance matrices for the joint move for each temperature:

- For $\beta_{temp} = 1$:
$$\begin{pmatrix} 8.20778151 & 10.95302411 & -0.10953903 & 0.03831994 \\ 10.95302411 & 50.85092527 & 0.05612781 & 0.06378059 \\ -0.10953903 & 0.05612781 & 0.00841513 & 0.0001823 \\ 0.03831994 & 0.06378059 & 0.0001823 & 0.00035206 \end{pmatrix}$$

- For $\beta_{temp} = 0.75$:
$$\begin{pmatrix} 10.94370868 & 14.60403215 & -0.14605204 & 0.05109326 \\ 14.60403215 & 67.8012337 & 0.07483708 & 0.08504079 \\ -0.14605204 & 0.07483708 & 0.01122017 & 0.00024307 \\ 0.05109326 & 0.08504079 & 0.00024307 & 0.00046942 \end{pmatrix}$$

- For $\beta_{temp} = 0.56$:
$$\begin{pmatrix} 14.89402706 & 23.56164399 & -0.4327477 & 0.04847274 \\ 23.56164399 & 97.92990939 & -0.47995099 & 0.07121918 \\ -0.4327477 & -0.47995099 & 0.0343587 & 0.00037822 \\ 0.04847274 & 0.07121918 & 0.00037822 & 0.00050598 \end{pmatrix}$$

- For $\beta_{temp} = 0.42$:
$$\begin{pmatrix} 19.85870275 & 31.41552532 & -0.57699693 & 0.06463033 \\ 31.41552532 & 130.57321252 & -0.63993466 & 0.09495891 \\ -0.57699693 & -0.63993466 & 0.04581159 & 0.00050429 \\ 0.06463033 & 0.09495891 & 0.00050429 & 0.00067463 \end{pmatrix}$$

B.6.1.2 Boundary condition blocks:

Table B.24: BC Gibbs blocks for $\beta_{temp} = 1$ for sampling FD and BC parameters for del Castillo's FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	Acceptance probabilities (%)
inlet	0	40	0.07	24.8, 25.6, 30.6
inlet	0	10	0.15	14.2, 11.4, 18.5
inlet	8	18	0.23	26.9, 24.7, 21.9
inlet	15	27	0.22	27.4, 32.9, 35.2
inlet	24	40	0.38	13.6, 19.6, 19.7
outlet	0	40	0.08	21.8, 16.2, 18.0
outlet	0	10	0.23	39.8, 36.9, 39.2
outlet	8	23	0.23	20.7, 19.7, 18.3
outlet	20	30	0.2	22.9, 20.0, 22.3
outlet	27	35	0.19	26.6, 20.8, 20.1
outlet	25	35	0.5	0.8, 0.7, 1.1
inlet	26	33	0.94	11.9, 5.1, 11.0

Table B.25: BC Gibbs blocks for $\beta_{temp} = 0.75$ for sampling FD and BC parameters for del Castillo's FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	Acceptance probabilities (%)
inlet	0	40	0.09	24.7, 22.7, 22.0
inlet	0	10	0.19	16.9, 20.2, 20.2
inlet	8	18	0.28	24.9, 23.0, 22.7
inlet	15	27	0.23	37.4, 42.7, 41.5
inlet	24	40	0.47	16.3, 17.7, 17.5
outlet	0	40	0.1	16.0, 19.8, 13.6
outlet	0	10	0.19	54.7, 56.5, 53.0
outlet	8	23	0.29	17.4, 17.6, 14.0
outlet	20	30	0.15	40.3, 46.6, 36.4
outlet	27	35	0.14	43.5, 42.7, 47.0
outlet	32	40	0.14	56.3, 58.6, 56.3
inlet	8	14	0.95	2.6, 2.1, 0.8
inlet	26	33	0.94	13.9, 17.6, 12.9

Table B.26: BC Gibbs blocks for $\beta_{temp} = 0.56$ for sampling FD and BC parameters for del Castillo's FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	Acceptance probabilities (%)
inlet	0	40	0.1	31.7, 25.8, 28.4
inlet	0	10	0.22	20.8, 20.0, 16.8
inlet	8	18	0.31	25.3, 34.9, 19.3
inlet	15	27	0.27	46.1, 37.2, 37.0
inlet	24	40	0.52	21.6, 21.1, 19.9
outlet	0	40	0.11	19.3, 20.8, 22.6
outlet	0	10	0.22	52.8, 54.7, 50.5
outlet	8	23	0.33	26.6, 24.0, 19.1
outlet	20	30	0.19	41.0, 40.7, 37.9
outlet	27	35	0.17	39.2, 45.8, 38.3
outlet	32	40	0.17	54.7, 54.7, 55.7
inlet	8	14	0.95	2.9, 3.1, 3.5
inlet	26	33	0.94	17.4, 19.7, 21.2

Table B.27: BC Gibbs blocks for $\beta_{temp} = 0.42$ for sampling FD and BC parameters for del Castillo's FD on simulated data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	Acceptance probabilities (%)
inlet	0	40	0.11	33.2, 36.4, 21.6
inlet	0	10	0.25	20.0, 24.0, 23.1
inlet	8	18	0.35	26.3, 27.9, 33.1
inlet	15	27	0.3	39.0, 42.7, 40.1
inlet	24	40	0.59	24.7, 27.3, 30.4
outlet	0	40	0.12	21.8, 25.6, 23.3
outlet	0	10	0.25	49.2, 51.5, 53.7
outlet	8	23	0.38	20.9, 21.7, 21.1
outlet	20	30	0.21	44.4, 47.3, 49.2
outlet	27	35	0.2	44.5, 38.3, 38.3
outlet	32	40	0.2	58.4, 58.0, 58.4
inlet	8	14	0.95	6.9, 8.1, 7.4
inlet	26	33	0.94	30.9, 40.8, 36.9

B.6.2 Del Castillo FD on M25 data

B.6.2.1 Covariance matrices

Covariance matrix for $\begin{pmatrix} z & \rho_j & u & w \end{pmatrix}^T$ for each temperature

- For $\beta_{temp} = 1$: $\begin{pmatrix} 0.66878389 & 0.89393652 & -0.00067186 & 0.00210403 \\ 0.89393652 & 3.33640027 & 0.00612946 & -0.00136162 \\ -0.00067186 & 0.00612946 & 0.00484103 & 0.00047524 \\ 0.00210403 & -0.00136162 & 0.00047524 & 0.00010494 \end{pmatrix}$

- For $\beta_{temp} = 0.75$: $\begin{pmatrix} 0.78746895 & 1.27333014 & 0.00141434 & 0.00308789 \\ 1.27333014 & 5.9597071 & 0.01655901 & -0.00077557 \\ 0.00141434 & 0.01655901 & 0.00573808 & 0.00047403 \\ 0.00308789 & -0.00077557 & 0.00047403 & 0.00011784 \end{pmatrix}$

- For $\beta_{temp} = 0.56$:
$$\begin{pmatrix} 0.91032747 & 0.97317003 & 0.00399942 & 0.00339584 \\ 0.97317003 & 6.24639295 & 0.02674442 & -0.00593384 \\ 0.00399942 & 0.02674442 & 0.0056371 & 0.00051024 \\ 0.00339584 & -0.00593384 & 0.00051024 & 0.00012784 \end{pmatrix}$$

- For $\beta_{temp} = 0.42$:
$$\begin{pmatrix} 1.69715527 & 2.72433505 & -0.03295926 & 0.00081907 \\ 2.72433505 & 10.64697739 & -0.06830345 & -0.01433845 \\ -0.03295926 & -0.06830345 & 0.01229563 & 0.00125811 \\ 0.00081907 & -0.01433845 & 0.00125811 & 0.00026881 \end{pmatrix}$$

B.6.2.2 Boundary condition blocks

Table B.28: BC Gibbs blocks for $\beta_{temp} = 1$ for sampling FD and BCs for del Castillo's FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	49	0.07	28.4, 29.7, 23.4
inlet	0	15	0.15	14.9, 15.3, 18.4
inlet	7	35	0.13	34.2, 26.1, 31.9
inlet	33	49	0.4	29.6, 31.6, 26.0
outlet	0	49	0.04	28.8, 30.9, 31.0
outlet	0	15	0.25	24.1, 26.9, 24.2
outlet	12	24	0.21	20.6, 27.7, 19.6
outlet	25	39	0.16	27.7, 34.4, 24.4
outlet	29	49	0.12	20.1, 24.8, 21.2
outlet	6	18	0.4	35.0, 33.1, 36.4
outlet	39	49	0.4	20.0, 16.9, 17.2
inlet	0	12	0.2	11.8, 11.3, 9.6
inlet	20	36	0.25	44.5, 48.6, 45.4
outlet	2	12	0.9	6.4, 4.3, 2.9
outlet	16	26	0.15	24.4, 20.0, 22.4
inlet	10	18	0.45	34.4, 29.6, 38.9

Table B.29: BC Gibbs blocks for $\beta_{temp} = 0.76$ for sampling FD and BCs for del Castillo's FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	49	0.1	22.7, 15.4, 25.3
inlet	0	15	0.2	13.3, 11.1, 10.1
inlet	7	34	0.18	27.2, 26.8, 28.1
inlet	34	49	0.44	33.3, 27.7, 33.1
outlet	0	49	0.055	21.6, 24.5, 21.3
outlet	0	15	0.3	25.0, 30.6, 34.8
outlet	12	24	0.21	28.6, 21.9, 30.6
outlet	25	39	0.2	18.9, 30.6, 30.6
outlet	29	49	0.14	23.6, 19.5, 15.8
outlet	6	18	0.45	42.4, 43.2, 43.0
outlet	38	49	0.48	5.0, 4.8, 8.7
inlet	0	12	0.24	14.0, 21.1, 13.1
inlet	20	35	0.31	44.2, 46.5, 39.7
inlet	5	15	0.9	2.1, 0.7, 0.7
outlet	2	12	1	4.7, 4.1, 6.6
outlet	16	26	0.21	20.8, 21.3, 16.8
inlet	10	18	0.52	36.7, 37.9, 42.1

Table B.30: BC Gibbs blocks for $\beta_{temp} = 0.58$ for sampling FD and BCs for del Castillo's FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	49	0.11	25.6, 33.1, 24.5
inlet	0	15	0.21	16.5, 18.5, 11.8
inlet	7	34	0.2	32.6, 31.7, 37.1
inlet	34	49	0.47	39.4, 35.5, 42.3
outlet	0	49	0.06	23.2, 29.2, 29.6
outlet	0	15	0.39	26.9, 24.8, 21.4
outlet	12	24	0.26	35.0, 30.8, 28.9
outlet	25	39	0.2	31.8, 36.6, 27.5
outlet	29	49	0.14	24.8, 22.0, 28.1
outlet	6	18	0.5	46.2, 48.9, 35.9
outlet	38	49	0.51	7.9, 12.0, 6.7
inlet	0	12	0.31	12.1, 10.0, 17.0
inlet	20	35	0.36	47.1, 38.5, 46.5
inlet	5	15	0.8	7.7, 7.5, 7.1
outlet	0	13	1	2.2, 1.5, 2.4
outlet	0	13	1	4.1, 3.1, 3.4
outlet	16	26	0.26	25.7, 16.3, 27.4
inlet	10	18	0.55	35.2, 39.4, 37.6

Table B.31: BC Gibbs blocks for $\beta_{temp} = 0.44$ for sampling FD and BCs for del Castillo's FD on M25 data with parallel tempering. We give the start time, end time, and step size ω for the BC Gibbs blocks along with the acceptance probabilities for each of the 3 chains.

BC type	Start (min)	End (min)	ω	acceptance probabilities (%)
inlet	0	49	0.155	26.6, 31.0, 29.0
inlet	0	15	0.25	20.2, 17.4, 28.9
inlet	7	34	0.29	28.8, 28.1, 32.4
inlet	34	49	0.57	34.3, 44.0, 37.5
outlet	0	49	0.078	27.9, 27.9, 27.3
outlet	0	15	0.44	31.0, 37.4, 32.9
outlet	12	24	0.3	36.9, 31.8, 35.3
outlet	25	39	0.24	30.2, 34.8, 26.8
outlet	29	49	0.18	17.3, 23.0, 19.5
outlet	6	18	0.65	45.2, 40.9, 43.6
outlet	38	49	0.5	11.1, 12.5, 14.2
inlet	0	12	0.39	13.7, 16.1, 14.3
inlet	20	35	0.5	39.4, 36.4, 40.1
inlet	5	15	0.9	9.6, 12.1, 8.3
outlet	0	13	1	3.4, 6.1, 6.8
outlet	0	13	1	3.6, 4.8, 5.1
outlet	16	26	0.3	26.6, 27.8, 20.1
inlet	10	18	0.61	51.9, 45.5, 40.7

Bibliography

[1] Atchadé, Y. F., Roberts, G. O., Rosenthal, J. S., 2011. Towards optimal scaling of metropolis-coupled Markov chain Monte Carlo, *Statistics and Computing* 21(4), pp. 555-568.

[2] Aw, A., Rascle, M., 2000. Resurection of "second order" models of traffic flow. *SIAM Journal on Applied Mathematics*, 60(3), pp.916-938.

[3] Bando,M., Hasebe, K., Nakayama, A., Shibata, A., and Sugiyama, 1995. Dynamic Model of traffic congestion and numerical simulation. *Physical Review E* 51, 1035-1042. (doi:10.1103/PhysRevE.51.1035)

[4] Beskos, A., Girolami, M., Lan, S., Farrell, P. E., Stuart A. M., 2017. Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics 335* (15) pp 327-351

[5] Betancourt, M. 2018. A Conceptual Introduction to Hamiltonian Monte Carlo. *arxiv*, stat.ME, DOI: arxiv:1701.02434v2

[6] Bonzani, I., Mussone, L., 2009. On the derivation of the velocity and fundamental traffic flow diagram from the modelling of the vehicle-driver behaviors. *Mathematical and Computer Modelling*, 50, pp. 1107-1112.

[7] Bonzani, I., Gramani Cumin, L. M., 2009. Critical analysis and perspectives on the hydrodynamic approach for the mathematical theory of vehicular traffic. *Mathematical and Computer Modelling*, 50 526-541. doi:10.1016/j.mcm.2009.03.007

[8] Bourrel, E., Lesort, J., 2003. Mixing Micro and Macro Representations of Traffic Flow: A Hybrid Model Based on the LWR Theory. *Transportation Research Record: Journal of the Transportation Research Board*, 1852, pp. 193-200.

[9] Brooks, S., Gelman, A., Jones, G. L., Meng, X, 2011 *Handbook of Markov Chain Monte Carlo*. CRC Press

[10] Carpenter, B., 2017. Ensemble Methods are Doomed to Fail in High Dimensions. [blog] *Statistical Modeling, Causal Inference, and Social Science*. Availalbe at: https://statmodeling.stat.columbia.edu/2017/03/15/ensemble-methods-doomed-fail-high-dimensions/ [Accessed 4th March 2019]

[11] Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A.. 2017. Stan: A probabilistic programming language. Journal of Statistical Software **76**(1). DOI 10.18637/jss.v076.i01

[12] Cassidy MJ, Bertini RL (1999). *Some traffic features at freeway bottlenecks*. Transportation Research Part B Methodology 33(1), pp. 2542

[13] Clawpack Development Team, 2018. Clawpack Version 5.4.0, http://www.clawpack.org, doi:10.5281/zenodo.262111

[14] Cook, S. R., Gelman, A., Rubin, D. B., 2006. Validation of Software for Bayesian Models Using Posterior Quantiles. *Journal of Computational and Graphical Statistics* **15**(3): 675692

[15] Cotter, S. L., Roberts G. O., Stuart, A. M., White, D., 2013. MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. *Statistical Science* 28(3), pp. 424-446 DOI: 10.1214/13-STS421

[16] Daganzo, C., F., 1994. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, 28(4) pp. 269-287.

[17] Daganzo, C., F., 1995. Requiem for second-order fluid approximations of traffic flow. *Transportation Research Part B*, 29(4), pp.277-286.

[18] 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

[19] Foreman-Mackey, D., Hogg, D. W., Lang, D. and Goodman, J. 2013. emcee: The MCMC Hammer. *arxiv* [preprint], astro-ph.IM, Vol 125, pp 306. DOI: 10.1086/670067

[20] Gasser, I., Sirito, G., and Werner, B. 2004. Bifurcation analysis of a class of car following traffic models. *Physica D* 197, p.222-241.

[21] Gelman, A., Carlin, J. B., Stern, H.S., Rubin, D. B., 2004. *Bayesian Data Analysis* (2nd ed.). Chapman and Hall/CRC.

[22] Goodman, J. and Weare, J. 2010. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computer Science* **5**(1), 6580.

[23] Graham, M. M., Storkey, A. J., 2017. Continuously tempered Hamiltonian Monte Carlo. *arxiv* Available at https://arxiv.org/abs/1704.03338v1 [Accessed 5th March 2019]

[24] Greenshields B.D., 1934. The photographic method of studying traffic behaviour. *Proceedings of the 13th Annual Meeting of the Highway Research Board* pp. 382-399.

[25] Grosse, R. B., Duvenaud, D. K. Testing MCMC code. *arxiv* Available from: https://arxiv.org/abs/1412.5218 [Accessed 3rd March 2019]

[26] Hajian, S., Hintermüller, M., Schillings, C., Strogies, N., 2018. A Bayesian approach for parameter identification in gas networks [preprint]. DOI 10.20347/WIAS.PREPRINT.2537, available at https://opus4.kobv.de/opus4-trr154/frontdoor/index/index/docId/231 [Accessed 4th March 2019]

[27] Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**(1): 97-109.

[28] Helbing, D., Johansson, A.F., 2008. On the controversy around Daganzo's requiem for and Aw-Rascle's resurretion of second-order traffic flow models. *The European Physical Journal B*, 69, pp.549-562.

[29] Heydecker, B.G., Addison, J. D., 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*. (pp. 109-120).

[30] Hoogendoorn, S. P. and Bovy, P. H. L., 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. doi: 10.1177/095965180121500402

[31] Huijser, D., Goodman, J., Brewer, B. J., 2015. Properties of the Affine Invariant Ensemble Sampler in high dimensions. *arxiv* [preprint]. Available at https://arxiv.org/abs/1509.02230 [Accessed 4th March 2019].

[32] Iacus, M., Stefano, 2008. *Simulation and Inference for Stochastic Differential Equations*. Springer.

[33] Jacquet, D., Canudas de Wit, C., Koenig, D., 2005. Traffic Control and Monitoring with a Macroscopic Model in the Presence of Strong Congestion Waves. *Proceedings of the 44th IEEE Conference on Decision and Control* DOI: 10.1109/CDC.2005.1582482

[34] Kantas, N., Beskos, A., Jasra, 2014. A. Sequential Monte Carlo Methods for High-Dimensional Inverse Problems: A case study for the Navier-Stokes equation. *SIAM/ASA Journal of Uncertainty Quantification*, 2, 464-489

[35] Kerner, B. S., 2004. *The Physics of Traffic*. Springer-Verlag Berlin Heidelberg.

[36] Law, K. J. H, 2014. Proposals which speed up function-space MCMC. *Journal of Computational and Applied Mathematics* **262**(15) pp 127-138

[37] LeVeque Randall, J. 2004. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press.

[38] Levin, D. A., Peres, Y., Wilmer, E. L., 2009. *Markov Chains and Mixing Times*. American Mathematical Society.

[39] Lighthill, M. H., and Whitham, G.B., 1955. On Kinematic Waves 2: A Theory of Traffic Flow on Long Crowded Roads. Proceedings of the Royal Society of London A, 229, pp.317-345.

[40] Liu, J. S. 2001, *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag New York

[41] Mahnke, R., Kaupuzs, J., Lubashevsky, I., 2005. Probabilistic description of traffic flow. *Physics Reports*, 408(2005), 1-130.

[42] Mandli, K.T., Ahmadia, A.J., Berger, M.J., Calhoun, D.A., George, D.L., Hadjimichael, Y., Ketcheson, D.I., Lemoine, G.I., LeVeque, R.J., 2016. Clawpack: building an open source ecosystem for solving hyperbolic PDEs. PeerJ Computer Science. doi:10.7717/peerj-cs.68

[43] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., 1953. Equations of state calculations by fast computing machines, *Journal of Chemical Physics* **21**(6): 1087-1091

[44] Morgan, J.,V., 2002. *Numerical Methods for Macroscopic Traffic Models*. PhD dissertation. University of Reading.

[45] Nagel, K., and Schreckenberg, M., 1992. A cellular automaton model for freeway traffic. *Journal de Physique, 1 France*, *2*(12), pp. 2221-2229

[46] Roberts, G., and Tweedie, R., 1996. Exponential convergence of Langevin distributions and their discrete approximations, *Bernoulli* **2**(4), 341363.

[47] Sambridge, M., 2014. A Parallel Tempering algorithm for probabilistic sampling and multimodal optimization. *Geophysical Journal International* **196** pp 357-374. DOI: doi: 10.1093/gji/ggt342

[48] Schönhof, M. and Helbing, D., 2007. Empirical Features of Congested Traffic States and their Implications for Traffic Modeling. *Transportation Science* 41(2), pp. 135-166

[49] Sokal, A. D., 1989. *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*, lecture notes (unpublished). In *Cours de Troisieme Cyle de la Physique en Suisse Romande*., Lausanne.

[50] Papageorgiou, M., 1997. Some remarks on macroscopic traffic flow modelling. *Transportation Research Part A*, 32(5), pp. 323-329.

[51] Polson, N., Sokolov, V., 2015. Bayesian analysis of traffic flow on interstate I-55: the LWR model. *The Annals of Applied Statistics* **9**(4), pp 18641888

[52] Rasmussen, C.E. and Williams K. I.. 2002. *Gaussian Processes for Machine Learning*. MIT Press.

[53] Richards, P.I., 1956. Shock Waves on the Highway. *Operations Research*, 4 (1) pp.42-51

[54] Stuart, A. M., 2010. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19, pp 451-559 doi:10.1017/ S0962492910000061

[55] Tawn, N. G., Roberts, G. O., 2018. Accelerating Parallel Tempering: Quantile Tempering Algorithm (QuanTA). *arxiv* [preprint] Stat.ME. Available at: https://arxiv.org/abs/1808.10415v1 [Accessed 5th March 2019]

[56] ter Braak, C. J. F. 2006. A Markov chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces *Statistics and Computing*, **16**(3), pp 239-249.

[57] Treiber, M., Hennecke, A., and Helbing, D., 2000. Congested traffic states in empirical observations and microscopic simulations. *Physical Review E* 62(4) pp.1805-1824

[58] Treiber, M., and Kesting, A., 2013. *Traffic Flow Dynamics*. Germany: Springer.

[59] Underwood, R. T., 1961. Speed, volume and density relationships, Quality and Theory of Traffic Flow, *Yale Bureau of Highway Traffic*, p141-88

[60] Van Wageningen-Kessels, F., van Lint, H., Vuik, K., Hoogendoorn, S., 2014. *Genealogy of traffic flow models*. EURO J Transp Logist. doi: 10.1007/s13676-014-0045-5

[61] Ward, J., 2009. *Heterogeneity, Lane-Changing and Instability in Traffic: a Mathematical Approach*. PhD dissertation. University of Bristol.

[62] Whitham, G., B., 1974. *Linear and Nonlinear Waves*. New York: John Wiley and Sons.

[63] Wilson, R.E., 2008. Mechanisms for spatio-temporal pattern formation in highway traffic models. *Philosophical Transactions of the Royal Society A* 368, pp.4455-4479.

[64] Zhang, H.M., 2000. A non-equilibrium traffic model devoid of gas-like behaviour. *Transportation Research Part B*, 36 (2002) pp.275-290.