Statistical calibration of microscopic and macroscopic traffic flow simulators

A case study on robust model assessment of traffic flow constitutive laws

Ioannis Zachos

Hughes Hall College

Supervisor: Professor Mark Girolami

Advisor: Professor Allan McRobie

Industry supervisor: Dr. Gerard Casey

Department of Engineering

University of Cambridge

October 2021

Contents

Abstract

The proliferation of traffic sensing devices on motorways has spurred interest in data assimilation in macroscopic traffic flow partial differential equations, such as the Lighthill Whitham Richards (LWR) model. The empirical constitutive laws (fundamental diagrams) embedded within these PDEs are uncertain and misspecified. We leverage a Bayesian model assessment framework introduced in (Calderhead and Girolami, 2009) to robustly identify the most plausible FDs accounting for the trade-off between model fit and complexity. This is achieved by computing unbiased estimates of the marginal likelihood for each FD based on ideas from thermodynamic integration while sampling from the resulting power posteriors using a Metropolis Hastings Markov Chain Monte Carlo algorithm. In recognition of the ML's sensitivity to prior diffusivity we also perform a sensitivity analysis of our ML estimates. We compare ten constitutive laws and apply our methodology to synthetic data to identify their structural differences. Finally, we validate the choice of FD in (Coullon and Pokern, 2020) by applying our methodology to loop detector data from the M25 motorway.

Keywords: traffic constitutive laws, macroscopic traffic flow modelling, Bayesian inference, marginal likelihood estimation, thermodynamic integration, Metropolis Hastings.

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Economic growth and urbanisation have significantly increased mobility patterns in cities such as London. The increasing travel demand has posed serious challenges in modelling and regulating traffic flows. Intelligent Transportation Systems (ITS) are collections of hardware and software technologies that address these challenges when optimising the flow of vehicles in a network, increasing the network's capacity to accommodate more vehicles, improving road safety and mitigating against air pollution and greenhouse gas emissions (Rudskoy, Ilin, and Prokhorov, 2021). An important consequence of the development of ITS is the proliferation of traffic sensing devices ranging from stationary loop detector sensors installed on road networks (Eulerian data) to probe devices installed on vehicles (Lagrangian data). This breakthrough in sensing technologies has stimulated an area of research focused on reconciling traffic engineering models with data. This coupling aims to deliver faster, safer solutions to the aforementioned traffic management problems.

Traffic flow phenomena are represented at three different scales: *microscopic*,*mesoscopic* and *macroscopic*. Microscopic models characterise individual agents (e.g. vehicles) and their mutual interaction. In contrast, macroscopic models are coarse representations based on differential equations (DEs) that describe averaging properties of vehicle movement. Mesoscopic models link microscopic with macroscopic models by accounting for the stochasticity embedded in agent interactions and averaging effects of macroscopic models. The scope of the literature review is limited to macroscopic and microscopic inference, where components of the former were studied in the first year while the latter shapes the future PhD direction detailed in Chapter 4. We refer the reader to an extensive review of mesoscopic models in (Kessels, 2019,

ch.3, 6).

1.1 Microscopic modelling

1.1.1 Existing models

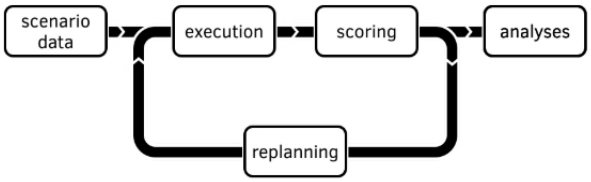


Figure 1.1: Schematic of an agent-based model called MATSim (Axhausen, 2016). Scenario data corresponds to agent population and plan synthesis from input supply and demand data, execution refers to running the mobility simulator, and scoring involves evaluating utility functions. In re-planning agent plans are adjusted according to their utility function and analyses data are derived from the emergent structure of agent dynamics.

The class of microscopic models contains fine-grained representations of individual (agent) dynamics and includes car-following models (Nagatani, 2000; Zhao and Gao, 2005; Panwai and Dia, 2005) and agent-based models (ABMs) (Axhausen, 2016; Smith, Beckman, and Baggerly, 1995). Car-following models describe drivers following each other in a traffic stream using a system of ordinary differential equations (ODEs) based on equations of motion. Schedule-following models (Kieu, Ngoduy, et al., 2019) extend car-following models to account for the adherence of transport modes (e.g. buses) to their proposed schedule. Therefore, bus drivers aim to optimise their flow in a network while meeting their proposed schedule at each bus stop. However, these models neglect the effect of multiple transportation modes in traffic flow.

ABMs deal with multiple interacting modes in an urban transportation network. People are treated as agents interacting dynamically and stochastically within an environment. The environment comprises of a transportation network with land-use (supply) data. Agents and their proposed activities are synthesised from household survey and travel diary (demand)

data. Plans are derived from each agent's proposed activities that shape their movement in the network. A mobility model simulates agent dynamics using queue-based models (vehicles form queues depending on demand instead of following vehicles in front of them) and agent utility functions are computed. Agents adjust their plans to maximise their utility before the mobility simulator is run again. This procedure is repeated iteratively as shown in Figure 1.1 and relevant quantities such as traffic volumes are calculated in the analysis stage. Examples of state-of-the-art ABMs are TRANSIMS (Smith, Beckman, and Baggerly, 1995) and MATSim (Axhausen, 2016) where the latter is open-source whereas the former is not. A complete review of the latest extensions on these models is provided in (Kagho, Balac, and Axhausen, 2020).[1]

1.1.2 Statistical inference

ABMs are better-suited in capturing the dynamics of complex urban environments compared to their macroscopic counterparts reviewed in later sections. Yet, ABMs pose significant statistical challenges during calibration (Banks and Mevin B. Hooten, 2021). We identify these challenges in the review below and shape the PhD research objectives accordingly in Chapter 4. The scope is limited to calibrating mobility simulators used in the execution stage of MATSim (see Figure 1.1). We refer the reader to (Fournier et al., 2018; Garrido et al., 2019) for a complete review of statistical approaches to population synthesis feeding into scenario and data generation. We hereby refer to mobility simulators as ABMs. Calibration approaches are broadly divided in agent-driven and black-box-driven.

Agent-driven approaches model the time evolution of the state-space of individual agents within a Markovian ABM that can be efficiently simulated. A suite of filtering algorithms have been applied to assimilate data into ABMs in an online fashion (M. Wang and Hu, 2015a; Ward, Andrew J. Evans, and N. S. Malleson, 2016; Lux, 2018; Lueck et al., 2019; Clay, Kieu, et al., 2020; Nick Malleson et al., 2020; Kieu, Nicolas Malleson, and A. Heppenstall, 2020; Clay, Ward, et al., 2021; Ju, Heng, and Jacob, 2021). The non-linear nature of agent dynamics

[1]The industry sponsor company of this PhD (Arup) is working extensively with MATSim and therefore the PhD will be focused on calibrating MATSim components.

calls for non-Gaussian likelihoods which is why non-linear approximations of the Kalman filter (KF) and particle filters (PFs) have been employed. The fundamental problem in agent-driven approaches is scalability to high-dimensional agent state-spaces (Nick Malleson et al., 2020). The problem is compounded when state-spaces are partially discrete and continuous. Assuming N agents admitting M discrete values yields M^N state cardinality (M. Wang and Hu, 2015a) which in turn demands exponentially more particles. Mixed state-spaces are modelled by comparing two candidate filter realisations and updating the one achieving the highest acceptance ratio (Clay, Ward, et al., 2021) which further exacerbates the computational load of density estimation. Moreover, global ABM parameters are also augmenting the state-space (Kieu, Nicolas Malleson, and A. Heppenstall, 2020). Ignoring the computational overhead, high-dimensional particle methods suffer from sample *degeneracy* and therefore *impoverishment* (loss of diversity) as well as *deprivation* (deviation from true state) (Snyder et al., 2008). (Kieu, Nicolas Malleson, and A. Heppenstall, 2020; Nick Malleson et al., 2020) proposed roughening, jittering and diffusing particles to rejuvenate their diversity by adding Gaussian noise to the particles' state or parameters vectors. Yet, this requires noise level tuning as too high noise can nudge particles away from true state whereas too low can have no effect on their diversity (Nick Malleson et al., 2020). PF resampling strategies have also been revised. (M. Wang and Hu, 2015b) employ a mixture of standard and component set resampling which leverages particle information across different agents. Regardless, there is no guarantee that good particles survive. (Lueck et al., 2019) proposed a Metropolis Hastings (MH) step that preserves particle and hence agent correlation. The acceptance ratio requires likelihood evaluations which may be expensive to obtain for large particle sizes and mix slowly in high dimensions. Alternatively, Gaussian state density approximations can circumvent particle degeneracy and deprivation. The ensemble and unscented Kalman Filters (EnKF, UKF) (Ward, Andrew J. Evans, and N. S. Malleson, 2016; Clay, Ward, et al., 2021) maximise posterior expressiveness with limited particle sizes. The former applies the standard Kalman updates to an ensemble of randomly initialised particles while the latter applies a predetermined unscented transform to a fixed number of sigma points. UKF and EnKF with small ensemble size lead to posterior variance shrinkage and call upon covariance inflation methods to quantify uncertainty, thus suffering

from similar limitations to PFs. (Ju, Heng, and Jacob, 2021) have introduced Sequence Monte Carlo (SMC) filters that exploit approximate proposals to efficiently explore the state-space at linear in agent size costs and have illustrated their appealing achieved bias-variance trade-off. This contribution assumes negligible ABM simulation cost does not address particle issues in high-dimensions (e.g. more than 100 agents) where approximations for dimensionality reduction utterly fail. State-of-the-art SMC methods could potentially be more robust to particle problems in large dimensions[2].

Agent-driven approaches suffer from another drawback. ABMs are highly granular whereas real data are coarse and cannot be attributed to individual agents making any agent-level model inherently unidentifiable (Lueck et al., 2019; M. Hooten, C. Wikle, and Schwob, 2020). In transportation few vehicle trajectory data are accessible while aggregate traffic volume datasets are richer (Dong, 2016). (Fintzi et al., 2016) dealt with a similar problem in epidemiology by augmenting the latent space with agent-specific dynamics driven by coarse data to account for model misspecification. Augmentation introduces poor MCMC mixing problems in high dimensions and is therefore rendered infeasible for practical applications. With the exception of (Fintzi et al., 2016), online filtering in Markovian ABMs has only dealt with controlled model misspecification in the context of synthetic datasets only. Moreover, existing approaches have not dealt with multi-resolution sensor fusion data which are prevalent in transportation.

Black-box approaches treat the ABM as a simulator of unknown dynamics with limited access to parameter-output simulator pairs determined by a computational budget. Early approaches in ecology (Hazelbag et al., 2020) and economics (Grazzini, M. G. Richiardi, and Tsionas, 2017) employed optimisation routines (Hazelbag et al., 2020) such as genetic algorithms (Rogers and Tessin, 2004; A. J. Heppenstall, Andrew J Evans, and Birkin, 2007) and simulated annealing (Hara et al., 2013) which greedily explore "true" parameters without maintaining a probability density over plausible parameters. Therefore, they may not converge to global extrema if their optimisation objective is not smooth enough. (Grazzini

[2]This remains to be investigated.

and M. Richiardi, 2015; Grazzini, M. G. Richiardi, and Tsionas, 2017) obtained parameters minimising a distance between data moments and moments simulated from the ABM under different parameter regimes known as simulated minimum distance (SMD). *Consistency* of the estimators is proven when the Markovian ABM is in an absorbing equilibrium and the derived estimators are stationary and ergodic in real and simulated data. Such conditions cannot be guaranteed when the ABM lacks analytical structure and thus have to be tested empirically, limiting the applicability of SMD.

The ABM simulator has been emulated/approximated explicitly or implicitly via likelihood or posterior approximations. Implicit approximations include likelihood/posterior approximations (Xu, Dong, and Srihari, 2016; Dong, 2016; Grazzini, M. G. Richiardi, and Tsionas, 2017; Radev et al., 2020; Zheng, Q. Han, and Lin, 2020; Shiono, 2021). (Grazzini, M. G. Richiardi, and Tsionas, 2017) suggested kernel density estimation (KDE) of the likelihood of stochastic ABMs with time series data. KDE can achieve the desired level of expressiveness and smoothness of multi-modal parameter distributions by appropriately choosing kernels and their bandwidth. Nevertheless, KDE computations scale quadratically with the number of ABM simulations and is therefore unsuitable for large-scale ABMs. (S. N. Wood, 2010) recommended a *synthetic likelihood* (SL), where extracted summary statistics are assumed to be normally distributed with unknown parameter-dependent mean and covariance. (Frazier et al., 2019) later extended synthetic likelihoods to a Bayesian setting. Although the Gaussianity assumption is theoretically well founded, there is no guarantee of statistic sufficiency without an explicit data likelihood and therefore SL suffers from the same problems as Approximate Bayesian Computation (ABC) which is reviewed later. Ratio estimation (RE) formulated posterior inference as estimation of the likelihood to marginal likelihood ratio $r(x, \theta) = p(x|\theta)/p(x)$ or equivalently as a logistic classification of data as samples from the either of the two terms. An optimal choice of a likelihood $L(x|\theta) \propto r(x, \theta)$ is the exponential family of distributions. SL and RE have not yet been applied to ABMs with formidable parameter spaces and their posterior inference efficient has not been compared to other methods.

Another principled way of learning a likelihood constrained to a family of distributions

is variational inference (VI). (Xu, Dong, and Srihari, 2016; Dong, 2016) obtained a variational approximation of the latent posterior of discretised stochastic kinetic model (SKM). The choice of approximating distribution is a critical aspect in VI and one which facilitates computational efficiency. Therefore, scalability to large number of agents is not impeded. VI applied on SKM was shown to be competitive to sampling-based methods which mix slowly and require a large number of ABM simulations. (Radev et al., 2020; Shiono, 2021) used VI to approximate posteriors of economic ABM parameters. This approach differs from VI-SKM in that the latter optimises the non-convex Bethe free energy subject to marginalisation and normalisation constraints whereas the latter minimises the convex KL divergence between true and approximated posteriors. In the latter variational parameter learning is facilitated using normalising flows constructed on invertible neural network (NN) outputs. Despite computationally efficient, this approach relies on affine compositions of NNs to approximate an inverse mapping from a Gaussian latent variable to the ABM parameters. Moreover, this approach relies on aggregate data about individuals and does not propagate inference back to the individual agent states.

Implicit likelihood representations are also leveraged in ABC by defining a discrepancy measure between observed data and summary statistics constructed from ABM outputs (Vaart et al., 2015; Lambert et al., 2018). The choice of *sufficient* (information-maximising) summary statistics is far from trivial (Fearnhead and Prangle, 2010). Poor choice of statistics leads to many rejections in rejection ABC sampling especially for small tolerance values (see Chapter 2). (Radev et al., 2020; Shiono, 2021) used a bidirectional Long Short Term Memory neural network (LSTM-NN) to extract summary statistics from time series data. Although the network has necessary properties (e.g. permutation invariance) its construction is prone to over-fitting due to large number of network weights needed to be learned. Moreover, the quality of the summary network is not assessed independently of other model components and not directly used in ABC sampling. (Pacchiardi and Dutta, 2020) applied VI to an exponential family of likelihoods of the form $\frac{\exp\big(\kappa(\theta)^T s(x)\big)h(x)}{Z(\theta)}$. If the Kullback-Leibler (KL) divergence between the true and approximated posteriors is zero then the summary statistics $s(x)$ are *sufficient*. These statistics were used in ABC sampling and achieved smaller KL divergences

compared to ABC using NN learned statistics. However, the likelihood approximation is likely to fail in stochastic, misspecified ABMs with non-identifiable parameters that induce multi-modality in the resulting posterior.

In addition to parameters, ABMs embed constitutive laws (explained in later sections) and complex agent interaction rules that can be learned (Kieu, Ngoduy, et al., 2019; Lu, Maggioni, and Tang, 2021). The schedule-following bus simulator in (Kieu, Ngoduy, et al., 2019) contains an empirical constitutive law between mean bus speed and schedule adherence. To our knowledge, the effect of this law in the emergent structure has not been studied nor learned directly from data except for the work of (Fuhg, Marino, and Bouklas, 2021) which employed local approximate GPs in data-driven constitutive law learning embedded in elastomechanics PDEs. Likewise, agent interactions have been ignored or assumed to be known *a priori* (Ju, Heng, and Jacob, 2021). The novel approach of (Lu, Maggioni, and Tang, 2021) attempted to learn heterogeneous (multi-type) agent interactions ingrained in discretised ODEs defining an ABM as well as establish conditions for learnability. This approach has not yet been scaled to large-scale ABMs with complex interactions.

A third approach to data assimilation involves extracting the macro-scale dynamics from microscopic principles and performing inference on the former. In very niche cases DEs can be derived from microscopic principles (Mevin B Hooten and C. K. Wikle, 2010).(Dong, 2016) leveraged a *mean-field* approximation that assumes agent state conditional independence given the data. (Ju, Heng, and Jacob, 2021) posed a similar assumption in an epidemic ABM by replacing agent infection and recovery rates with their population averages, inducing homogeneity in disease transmission probability. An increasing agent size translates into exponentially more convoluted interactions between them (Nick Malleson et al., 2020) and thus mean-field assumptions break down asymptotically. (Nardini et al., 2021) addressed this limitation and proposed a novel *equation learning* framework for deriving differential equations directly from time-evolving ABMs. The authors collect ABM samples for fixed parameters, pre-define a pool of candidate basis functions (e.g. polynomials) to describe the time-evolving dynamics and learn the basis functions' coefficients using penalised regression methods. This new research

area that has not been extensively developed. For example, derivations of PDEs/stochastic DEs from space-evolving/stochastic ABMs have not been discovered yet.

Finally, we identify understudied open problems in ABM calibration. Incorporating network of agent interactions has only been examined in (Heard et al., 2014; C. K. Wikle and Mevin B Hooten, 2016). Dealing with ABM discontinuities and local phase transitions has not been formally addressed yet (Bijak, Bullock, and Hilton, 2017). Also, AB model comparison has not been performed and marginal likelihoods have been computed in small-scale ABMs (Ju, Heng, and Jacob, 2021) where the likelihood is cheaply available. The computational cost of simulating ABMs has left questions on which ABM components create certain emergent structure and how a complex ABM can be maximally reduced without becoming vastly misspecified unanswered. Finally, simplistic and misspecified ABMs have mostly been considered in the past with notable the exception of the epidemic model in (Andrianakis et al., 2017). Therefore there is great potential in testing the developed statistical machinery to large-scale ABMs such as MATSim (Axhausen, 2016).

1.1.3 Example: Epidemic ABM

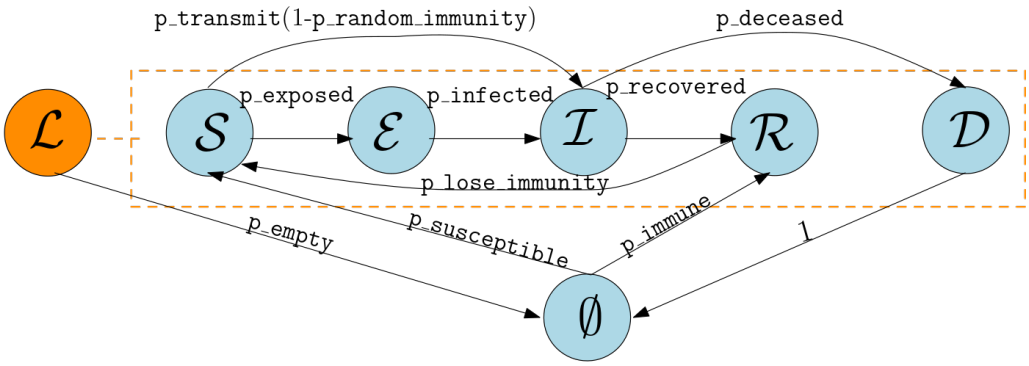


Figure 1.2: Plate diagram of status transitions emerging from the SEIRD model's agent interaction rules for an arbitrary cell.

We illustrate the previously identified statistical challenges in a stochastic contagion ABM called Susceptible-Exposed-Infected-Recovered-Deceased (SEIRD) adapted to a 2D spa-

tial grid (Lekone and Finkenstädt, 2006). Agent interaction rules are depicted in Figure 1.2.

Each cell can be empty ($\emptyset$), occupied by an agent (individual) or possess special properties. The possible agent states are:

- susceptible ($\mathcal{S}$): agent is healthy but susceptible to catching and/or getting exposed to the disease.
- exposed ($\mathcal{E}$): agent is exposed to the disease but showing no symptoms, meaning that it can already infect others when in contact and gets noticed as infected only after the incubation period.
- infected ($\mathcal{I}$): agent is infected and can infect neighbouring agents.
- recovered ($\mathcal{R}$): agent is immune to the disease but can potentially lose it and become susceptible again.
- deceased ($\mathcal{D}$): agent is deceased and will replaced by an empty cell in the next time step.

States $\mathcal{S}$,$\mathcal{E}$,$\mathcal{I}$,$\mathcal{R}$ are summarised as living ($\mathcal{L}$). The special cell states cannot be changed once initialised and are the following:

- source ($\mathcal{S}$): cell is an infection source and can transition neighbouring agents from $\mathcal{S}$ to $\mathcal{E}$.
- inert ($\mathcal{I}$): cell that does not partake in any of the dynamics and can be used to model spatial heterogeneities.

Agent interaction rules are assumed to be known *a priori* (see Figure 1.2) and are denoted by $\mathcal{A} \xrightarrow{\texttt{prob}} \mathcal{B}$ meaning a cell (or its corresponding agent) transitions from state $\mathcal{A}$ to state $\mathcal{B}$ with probability (w.p.) `prob`. Thereby we have:

- $\mathcal{L} \xrightarrow{\texttt{p\_empty}} \emptyset$,
- $\emptyset \xrightarrow{\texttt{p\_susceptible}} \mathcal{S}$,

- $\emptyset \xrightarrow{\texttt{p\_immune}} \mathcal{R}$,
- $\mathcal{S} \xrightarrow{\texttt{p\_exposed}} \mathcal{E}$ or $\mathcal{S} \xrightarrow{\texttt{p\_transmit}(1-\texttt{p\_random\_immunity})} \mathcal{I}$ by virtue of neighbouring cells being in states $\mathcal{E}, \mathcal{I}$,
- $\mathcal{E} \xrightarrow{\texttt{p\_infected}} \mathcal{I}$,
- $\mathcal{I} \xrightarrow{\texttt{p\_recovered}} \mathcal{R}$,
- $\mathcal{I} \xrightarrow{\texttt{p\_deceased}} \mathcal{D}$,
- $\mathcal{R} \xrightarrow{\texttt{p\_lose\_immunity}} \mathcal{S}$,
- $\mathcal{D} \xrightarrow{1} \emptyset$,

where $\texttt{p\_transmit}(1 - \texttt{p\_random\_immunity})$ is the transmission probability from non-immune neighbouring cells. Finally, a living agent moves:

1. w.p. `p_move_randomly` to a randomly chosen empty neighbouring cell, if there are any.
2. away from an infected neighbour to a randomly selected neighbouring empty cell, if there are any.

SEIRD's parameter vector is equal to the ten-dimensional vector of probabilities shown in Figure 1.2. The cell state-space at time t is denoted by $\mathbf{X}_t = [\mathbf{X}_t^{(1)}, \ldots, \mathbf{X}_t^{(N)}]$, where N is the number of cells. The joint cell states are four-dimensional discrete (numerical and categorical) and are listed below for each cell:

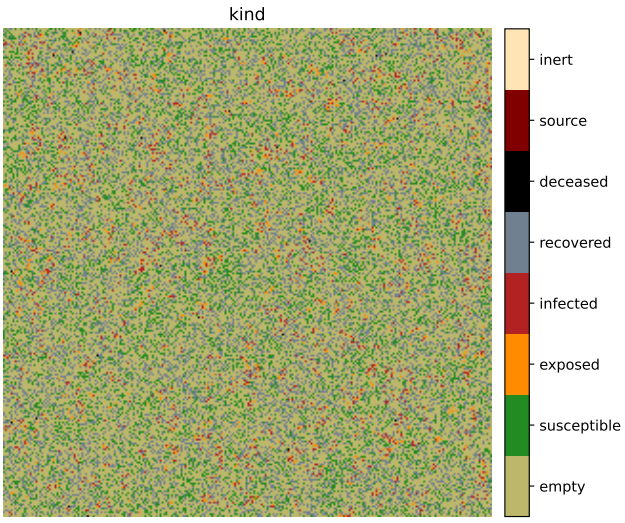
- status: $\emptyset, \mathcal{S}, \mathcal{E}, \mathcal{I}, \mathcal{R}, \mathcal{D}, \mathbb{S}, \mathbb{I}$,
- age: cell age reset after it turns empty,
- exposed_time: time steps a living cell has been exposed to the disease,
- immunity: whether or not cell is immune.

Model output space $\mathcal{X}$ includes cell densities time series by type and cumulative event (status transitions) counters listed above. Figure 1.3 depicts the emergent structure and stochasticity of SEIRD on two simulations under the same parameter regime.

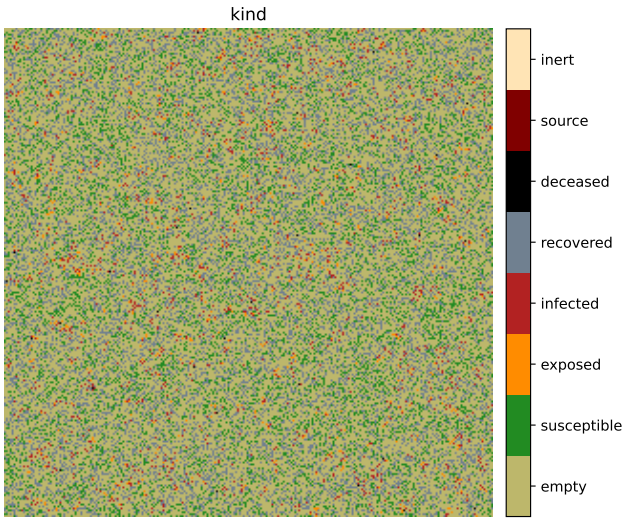
While SEIRD captures key dynamics in epidemics it still omits immunity, exposure, and transmission control measures (e.g. vaccination, quarantine effects). Additional model expressiveness can be achieved by encoding a network structure with supply data (e.g. hospitals) but will render simulation computationally cumbersome as in (Andrianakis et al., 2017). Increased model granularity may reduce model misspecification but will inevitably augment the model's parameter and agent state-space.

An agent-driven approach to inferring agent states and parameters would require exploration of 6^N state-space configurations (5 living statuses and 1 empty status for each agent), which is infeasible for large N. Moreover, infection and deceased data are coarsely available, susceptible or infected rates are not precisely known, the event space is sparsely observed and few agent disease histories are available. The latent space defined by unobserved quantities is enormous to be integrated over and therefore this approach is rendered impractical.

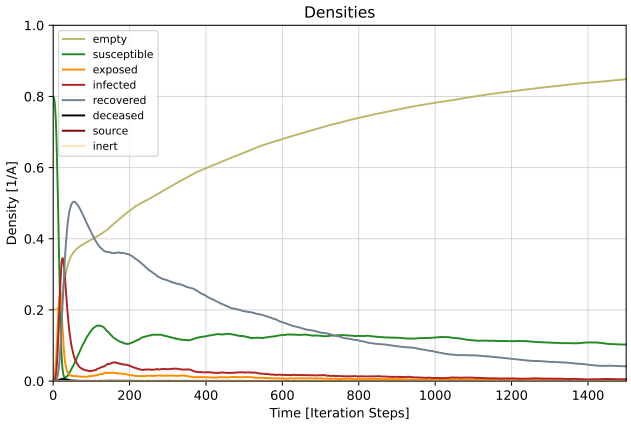
Black-box-driven approaches are better suited in the case of SEIRD. However, as the cardinality of $\boldsymbol{\Theta}$ increases (in complex extensions of SEIRD) it is not guaranteed that a unique set of 'true' parameters will be recovered from a scarce dataset (potentially even in the case of SEIRD). The choice of likelihood approximations becomes non-trivial as exponential family distributions fail to capture multi-modality. Emulators may also encounter challenges if the simulator variance is θ-dependent or if phase transitions occur in the simulator (e.g. agent access to a neighbourhood of the grid is suddenly blocked to reflect quarantine effects). Efficiently obtaining sparse non-stationary GP emulators is an open research problem. Finally, summary statistics used in ABC need to be robust to the non-stationary nature of the output time series shown in Figure 1.3, which is non-trivial if their *sufficiency* needs to be guaranteed for parameter posterior inference.



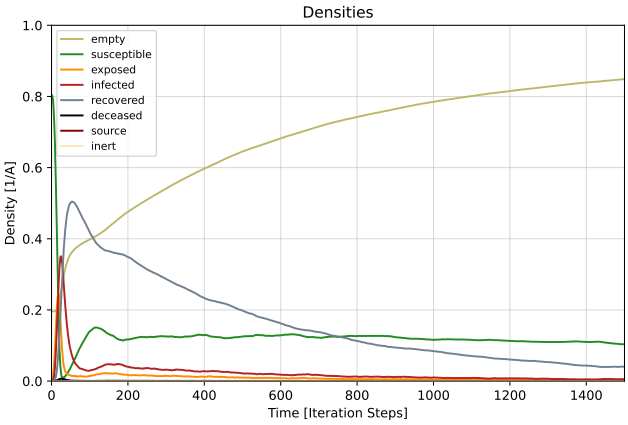
(a) Random seed is fixed to 50.



(b) Random seed is fixed to 100.



(c) Random seed is fixed to 50.



(d) Random seed is fixed to 100.

Figure 1.3: Cellular automata snapshots at time $t = 500$ on a 256x256 grid (top) and time series of cell status counts/densities (bottom) based on two SEIRD realisations coloured by cell status. Figures were generated using the Utopia library (Riedel et al., 2020).

1.1.4 Relevance to MATSim calibration

MATSim faces more calibration challenges than SEIRD since the former is a large-scale unidentifiable ABM with high-dimensional parameters and outputs (traffic analyses). Observational data are multi-resolution (coarse Eulerian traffic volumes and fewer Lagrangian vehicle trajectories), sparsely and partially available as in the SEIRD case. The agent environment is a transportation network while the emergent agent dynamics are stochastic and potentially contain phase transitions from agent re-planning (see Figure 1.1). The agent state-space is mixed (discrete and continuous) and the simulator embeds a constitutive law. Similarly, we argue that black-box-driven approaches are more promising for MATSim calibration and outline two of them in Chapter 2.

1.2 Macroscopic modelling

Macroscopic traffic flow models define relationships between three traffic quantities: *flow* q (number of vehicles per unit time), *density* k (number of vehicles per unit length) and average speed u (average speed of vehicles on a road segment). *Occupancy* $o \in [0, 1]$ of vehicles is sometimes measured instead of density and is equivalent to the density normalised by the road length. These quantities are measured by stationary sensors synchronised with traffic control software such as Motorway Incident Detection and Automatic Signalling (MIDAS) (Morris, 1997) and Split Cycle Offset Optimisation Technique (SCOOT) (Bretherton, K. Wood, and Raha, 1998). Depending on the emerging relationships among these three properties traffic regimes are classified into *free-flow* and *congested*. The former regime is characterised by vehicles moving at their desired speeds while the latter occurs when vehicles react to traffic ahead, decelerating to avoid collisions.

Assimilating loop-detector data in a macroscopic model is an emerging challenge. One such model is the Lighthill Whitham Richards (LWR) model (Lighthill and Whitham, 1955; Richards, 1956) inspired by fluid dynamics models. LWR is a first-order hyperbolic partial

differential equation (PDE) that expresses a conservation law relating the spatial concentration of vehicles (i.e. traffic density k), the rate at which vehicles pass through a cross section of a given road (i.e. traffic flow or flux q) and the average vehicle speed u. This deterministic relationship is quantified in a continuity equation:

$$k_t + u(k)k_x = 0 \quad \forall \ t > 0, \ x \in \mathbb{R}, \tag{1.1}$$

where k_t, k_x are the partial derivatives of density with respect to time and space while $q'(k) = u(k)$ is an *equilibrium constitutive* law linking the derivative of flow with respect to density (speed). Traffic density $k(x,t)$ varies across space-time $\mathbb{R} \times [0, \infty)$. Space is assumed to be one-dimensional to reflect the model's assumption of a single-lane road with one entry/exit.

If $q'(k)$ is a non-linear function of k then equation (1.1) is called non-linear. When $q'(k)$ is constant or a linear function of $k(x,t)$ (1.1) becomes linear or quasi-linear, respectively. The linear case reflects the assumption that vehicles move at constant speed u and traffic flow is increasing with traffic density which is only suitable in a free-flow traffic regime. The non-linear case corresponds to concave flow-density relationships explored in later sections.

LWR model's conservative form captures abrupt flow changes e.g. traffic jams. Despite its widespread use, LWR has been criticised for assuming that drivers respond to local changes in traffic density instantaneously according to $q'(k)$. This is unrealistic given that drivers respond to changing traffic conditions with a delay implying that $u(x,t)$ also depends on speed limits and road geometry traffic conditions. A mitigation against the LWR model's flaw was introduced in (Payne, 1971). Payne proposed coupling the continuity equation in (1.1) with a "momentum conservation"-type equation:

$$v_t + \overbrace{vv_x}^{C} = \frac{\overbrace{u(k) - v}^{R}}{\tau} - \frac{\overbrace{c_0^2}^{A}}{k} k_x, \tag{1.2}$$

where $v(x,t)$ is average vehicle speed and is part of the state vector $(k(x,t), v(x,t))^T$, τ is the relaxation time, and $c_0^2 = -\frac{\mathrm{d}v(u)}{\mathrm{d}k}\frac{1}{\tau} > 0$ is an anticipation constant describing the decreasing rate of equilibrium speed with increasing density. In traffic dynamics momentum is not conserved as vehicles collide, accelerate and decelerate. Equation (1.2) quantifies the delay in

driver reaction to changing traffic regimes. Three terms are introduced: anticipation $A(k,u)$, convection $C(k,u)$ and relaxation $R(k,u)$. Convection describes the changes in average speed due to changes in vehicle inflows/outflows. Anticipation characterises drivers' anticipation of downstream traffic density changes. More general anticipation terms include a diffusion term u_{xx} quantifying the change of vehicle acceleration over the space domain (drivers adjust their acceleration depending on the acceleration of nearby vehicles) (S. P. Hoogendoorn and Bovy, 2001). Relaxation expresses the tendency of traffic to revert to equilibrium speed specified by $q'(k)$ after some delay τ. Payne-type models with diffusion terms violate a fundamental traffic principle: vehicle *anisotropy*. Anisotropy refers to a vehicle's ability to only respond to downstream stimuli i.e. to vehicles in front of it. As a result of anisotropy, characteristic waves (traffic information) travel faster than vehicles which is physically impossible (Coullon and Pokern, 2020).

(Carlos F. Daganzo, 1995) critically reviewed Payne-type models showing how certain initial value problems lead to physically implausible solutions such as vehicle queues "compressing" from their rear end. The anisotropy problem was rectified in (Aw and Rascle, 2000) where (1.1) was coupled with the following equation:

$$v_t + p'(k)k_t + v\left(v_x + p'(k)k_x\right) = 0, \tag{1.3}$$

where $p(k)$ is the traffic pressure constitutive law inspired from gas dynamics and is assumed to be an increasing function of the density. Pressure can be roughly interpreted as driver anticipation of traffic as its space derivative appears in anticipation terms of generalised Payne-type models (S. P. Hoogendoorn and Bovy, 2001). A similar model is Zhang's model (H. Zhang, 2002) which proposes the following acceleration equation:

$$v_t + (v + kq''(k))v_x = \frac{u(k) - v}{\tau}, \tag{1.4}$$

Equation (1.4) yields characteristic speeds smaller than average vehicle speeds satisfying anisotropy.

On a parallel research front, single-lane scalar conservation laws have been generalised to multiple lanes and vehicles classes (e.g. trucks, bikes). Notable mentions are the works of (S. Hoogendoorn, 1999; Michalopoulos, Beskos, and Yamauchi, 1984; Carlos F. Daganzo, 1997).

Nevertheless, it can be argued that ABMs are better suited for modelling the dynamics of different vehicle classes among multiple lanes due to their fine-grained nature.

Although the state-of-the-art traffic dynamic models may resolve issues such as anisotropy and delayed driver reaction to changing traffic density, the resulting models may still be misspecified and therefore additional equations need to be empirically validated under different traffic scenarios (e.g. stop and go waves created by changing traffic lights). The need for empirical validation of model components does not invalidate the epistemic law of vehicle conservation encoded in all traffic PDEs. This need extends to empirical constitutive laws known as *fundamental diagrams* (FDs) which are also misspecified and need to be scrutinised accordingly. We introduce FDs, their desired properties and proposed functional forms in the next section.

Another limitation of traffic flow models is their deterministic nature. Yet, they are employed to describe spatio-temporal traffic waves generated from stochastic traffic phenomena such as accidents on motorways. The attempt by (Polson and Sokolov, 2015) to discretise the LWR model and add a Gaussian error term to account for stochasticity in the data is ad-hoc for reasons that (Coullon and Pokern, 2020) pointed out. First, there is no guarantee that the proposed numerical scheme converges to a specific stochastic PDE (SPDE) under refinement of the discretisation grid. Second, the Gaussian error term does not necessarily preserve positivity of the traffic density and violates the vehicle conservation law in (1.1). A "relaxation" of the conservation law has been proposed in (Müller and Bock, 2020) who added a stochastic forcing diffusion term (Brownian motion) to the LWR PDE. However, the proposed solution has only been illustrated in contrived examples with stochastic forcing and does not generalise to nonlinear drift terms. To our knowledge, a traffic SPDE inverse problem has not been formally introduced in generality yet.

1.3 Traffic constitutive laws

The emergent relationship between traffic flow, density and average speed is called the fundamental diagram relationship (Greenshields, Thompson, et al., 1934). FDs are constitutive laws derived from microscopic principles and/or empirical observations of traffic phenomena and are embedded in microscopic and macroscopic models. There are three equivalent formulations of FDs: flow-density ($q(k)$), speed-density ($u(k)$) and flow-speed ($q(u)$) relationships shown in Figure 1.4. Equivalence follows by virtue of *equilibrium* (stationary) assumptions

$$q = ku. \tag{1.5}$$

The nature of the FD relationship depends on road properties (e.g. number of lanes), vehicle and driver composition (e.g. fraction of trucks, buses), external conditions (e.g. weather and lighting conditions), and traffic laws (e.g. speed limits) (S. P. Hoogendoorn, Botma, and Minderhoud, 2007). In a free-flow regime as vehicles concentrate on a road the number of them exiting the road (or crossing a section of the road) increases until the road's *capacity* is reached (see maximum of top graph in Figure 1.4). A road's capacity or maximal intensity is the largest possible flow of vehicles achieved at the *critical density* (see maximum x-location of top graph in Figure 1.4). The critical density is the density after which any additional vehicle on the road contributes to congestion formation. Therefore, when "too many" vehicles accumulate on the road the flow is reduced and eventually fades to zero in the case of a traffic jam (vehicles are bumper to bumper). Traffic jams occur at *jam density* (see second roots of top and bottom left graphs of Figure 1.4). The average vehicle speed is a decreasing function of the density since a larger spatial concentration of vehicles leads to deceleration and lower speeds. Average vehicle speed is equal to the *free-flow* speed under mild traffic congestion (see y-intercept of bottom right graph of Figure 1.4). As more vehicles travel per unit of time the average speed of vehicles is reduced to avoid collisions and reaches zero in the case of a traffic jam.

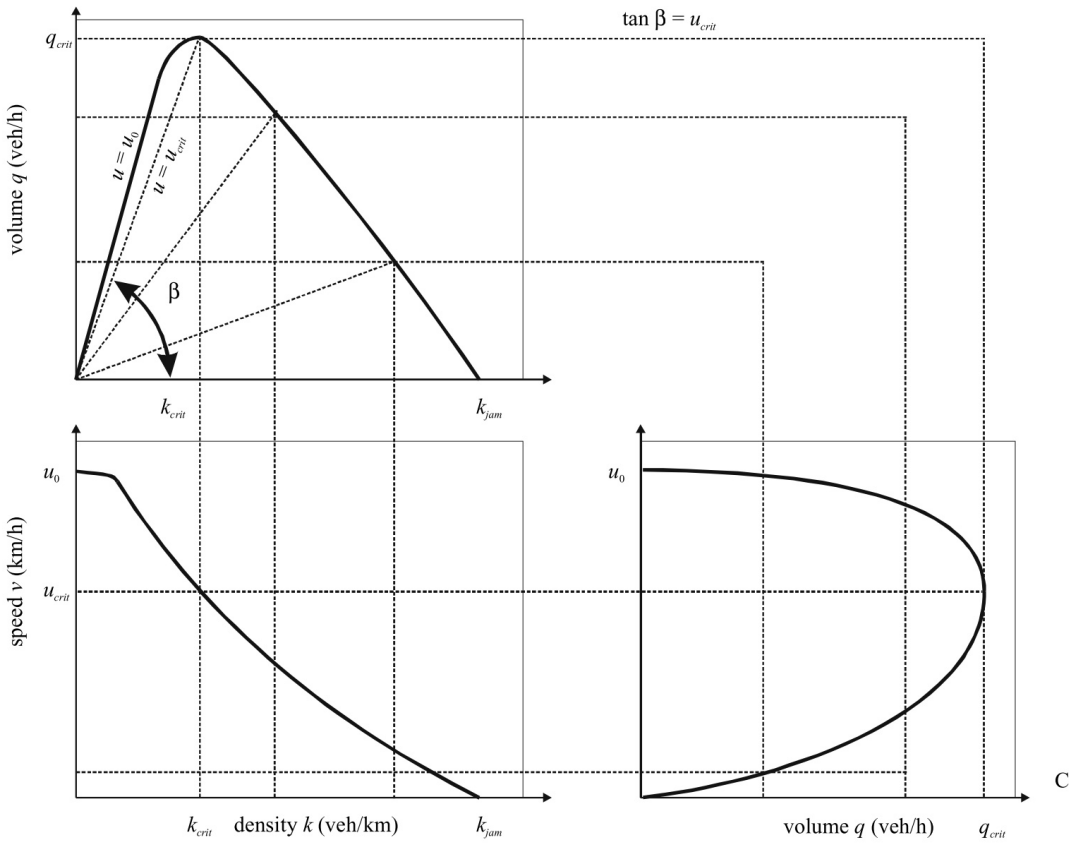


Figure 1.4: Idealised fundamental diagram relationships of flow-density (top), speed-density (bottom left) and speed-flow (bottom right) obtained from (S. P. Hoogendoorn, Botma, and Minderhoud, 2007)[p.82]. Volume is equivalent to traffic flow.

1.3.1 Key properties and functional forms

Empirical studies of FDs have elicited assumptions that a FD should satisfy in theory. (Castillo, 2012a) provides an account of the most common FD properties:

- Strict concavity [**P1**]: $q''(k) < 0 \; \forall \;\; 0 < k < k_{jam}$.
- Bounded speed [**P2**]: $0 < u < u_{max}$, where u_{max} is the maximum free-flow vehicle speed.
- Bounded density [**P3**]: $0 < k < k_{jam}$ where k_{jam} is the traffic jam density (vehicles are bumper to bumper).

- Boundary speed [**P4**]: $u(k = k_{jam}) = 0, u(k = 0) = u_f$, where u_f is the free-flow vehicle speed.
- Boundary flow [**P5**]: $q(k = 0) = 0, q(k = k_{jam}) = 0$ by virtue of (1.5).
- Boundary kinematic wave speed [**P6**]: $\frac{\mathrm{d}q}{\mathrm{d}k}\Big|_{k=0} = u_f$, $\frac{\mathrm{d}q}{\mathrm{d}k}\Big|_{k=k_{jam}} = c_f$, where c_f is the kinematic wave speed at jam density.

Condition **P1** is of high theoretical importance. In variational traffic flow theory (Carlos F. Daganzo, 2005a; Carlos F. Daganzo, 2005b) the LWR model's solution with a strictly concave FD was shown to be equivalent to a set of all shortest wave paths in space-time (x, t) starting from the boundary conditions. The set of wave paths are equivalent to the set of kinematic waves defined by the characteristic curves Γ defined in Section 1.4. Theoretical implications of a strictly concave FD have also been considered by (Ansorge, 1990) who showed that **P1** acts as an additional entropy condition for deriving physically relevant solutions to the LWR model. Therefore, empirically evaluating the assumptions listed above is crucial for determining LWR model misspecification.

Traffic engineers have proposed a number of FD models partially satisfying properties **P1** - **P6**. We list the ten most commonly used FDs in Table 1.1. In terms of their structural similarities, DelCastillo's negative power model ($Q_7(k)$) is a close approximation to Daganzo's in the limit of $\omega \to \infty$ as shown in Figure 1.5. DelCastillo's u and ω scaling parameters are related to the critical density k_{crit} via

$$k_{crit} = (1 + u^{\omega/(\omega+1)})^{-1} \tag{1.6}$$

Also, DeRomph's model reduces to Smulder's for $\alpha = k_{jam}$ and $\beta = 1$. Relaxations of Smuder's and DeRomph's models are considered that allow for a jump discontinuity to account for a phenomenon called *capacity drop*. This phenomenon encapsulates the observation that a transition from a free-flow traffic regime to a congested one achieves a higher maximum vehicle flow (capacity) than its reverse transition (S. P. Hoogendoorn, Botma, and Minderhoud, 2007)[p.88].

Name	Model	Parameters	Reference
Greenshield's	$Q_1(k) = u_f k \left(1 - \frac{k}{k_{jam}}\right)$	u_f, k_{jam}	(Greenshields, Bibbins, et al., 1935)
Greenberg's	$Q_2(k) = u_f k \log\left(\frac{k_{jam}}{k}\right)$	u_f, k_{jam}	(Greenberg, 1959)
Underwood's	$Q_3(k) = u_f k \exp\left(-\frac{k}{k_0}\right)$	u_f, k_0	(Underwood, 1961)
Northwestern's	$Q_4(k) = u_f k \left(\exp\left(-\frac{1}{2}\left(\frac{k}{k_0}\right)^2\right)\right)$	u_f, k_0	(Drake, Schofer, and May, 1965)
Newell's	$Q_5(k) = u_f k \left(1 - \exp\left(-\frac{\lambda}{u_f}\left(\frac{1}{k} - \frac{1}{k_{jam}}\right)\right)\right)$	u_f, k_{jam}, λ	(Newell, 1961)
Wang's	$Q_6(k) = \left(u_f k\right) * 1/\left(1 + \exp\left(\frac{k - k_{crit}}{s}\right)\right)$	u_f, k_{crit}, s	(H. Wang, Li, et al., 2011)
Daganzo's	$Q_7(k) = \begin{cases} \frac{q_{crit}}{k_{crit}} k & k < k_{crit} \\ q_{crit} \frac{k_{jam} - k}{k_{jam} - k_{crit}} & k \geq k_{crit} \end{cases}$	q_{crit}, k_{crit}, k_{jam}	(Carlos F Daganzo, 1994)
DelCastillo's	$Q_8(k) = Z\left[\left(u_f \frac{k}{k_{jam}}\right)^{-\omega} + \left(1 - \frac{k}{k_{jam}}\right)^{-\omega}\right]^{-1/\omega}$	Z, u_f, k_{jam}, ω	(Castillo, 2012b)
Smulder's	$Q_9(k) = \begin{cases} u_f k \left(1 - k/k_{jam}\right) & k < k_{crit} \\ \gamma k \left(1/k - 1/k_{jam}\right) & k \geq k_{crit} \end{cases}$	u_f, k_{crit}, k_{jam}, γ	(Smulders, 1989)
DeRomph's	$Q_{10}(k) = \begin{cases} u_f k \left(1 - k/\alpha\right) & k < k_{crit} \\ \gamma k \left(1/k - 1/k_{jam}\right)^{\beta} & k \geq k_{crit} \end{cases}$	u_f, k_{crit}, k_{jam}, γ, α, β	(De Romph, 1996)

Table 1.1: Fundamental diagram flow (q) - density (k) relationships with our naming conventions and their associated parameters.

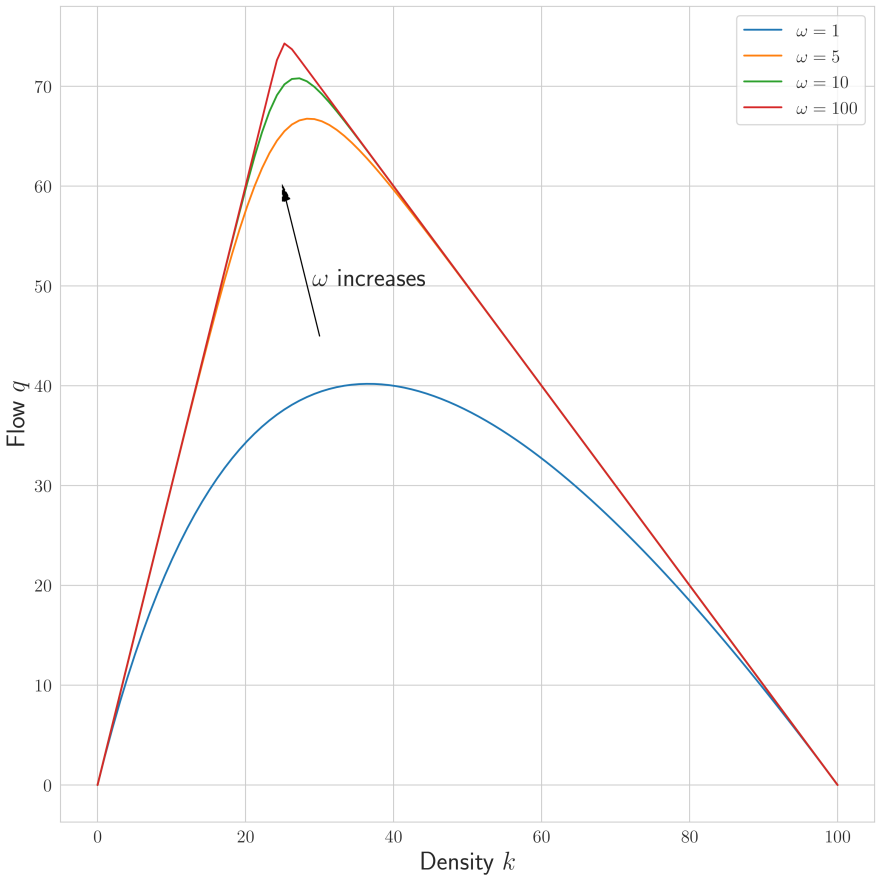


Figure 1.5: DelCastillo's four parameter fundamental diagram plotted for different values of the ω scaling parameter and $Z = 100$,$u = 3$,$k_{jam} = 100$.

1.4 Importance of fundamental diagrams

Moreover, the FD closes the vehicle conservation equation in (1.1). The FD also controls the speed and direction of traffic wave propagation (Mynt-U and Debnath, 2007). Traffic waves can be thought of as traffic information that is passed either upstream (away from a driver) or downstream (towards a driver). Waves can originate from either traffic disturbances such as traffic accidents or smooth traffic operation. Using the *method of characteristics* we can identify paths in (x, t) space along which the solution $k(x, t)$ to (1.1) is constant. We seek

continuous solutions by computing the total differential

$$\mathrm{d}k = \frac{\partial k}{\partial t}\,\mathrm{d}t + \frac{\partial k}{\partial x}\,\mathrm{d}x. \tag{1.7}$$

Dividing by $\mathrm{d}t$ gives

$$\frac{\mathrm{d}k}{\mathrm{d}t} = \frac{\partial k}{\partial t} + \frac{\partial k}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} = 0 \tag{1.8}$$

along the *characteristic curves* Γ defined by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = q'(k). \tag{1.9}$$

The characteristic curves Γ are curves along which the solution to the PDE is constant and are always tangent to the *solution surface* $k(x,t) - k = 0$ in (x,t,k) space. According to (1.9), an increasing $q(k)$ corresponds to upstream wave propagation in a free-flow traffic regime while a decreasing $q(k)$ describes a congested regime of downstream wave propagation.

Solutions to non-linear traffic PDEs can "break" which yields discontinuous solutions that violate uniqueness after some time $t > T_0$ and renders the PDE ill-posed. The "breaking" is evidenced by the appearance of *shock waves* that propagate with FD-governed speeds when characteristic curves intersect. The shock wave shape is also affected by the choice of FD. Empirically observations indicate that downstream shocks maintain their shape in propagation (Whitham, 1999). However, this is only captured by FDs that have a linear congested form whereas FDs with non-linear congested regions smoothen the downstream shock's shape (Carlos F Daganzo, 1994). This motivates the need for evaluating the proposed FD functional forms in 1.1 against data.

The importance of empirically validated fundamental diagrams extends to traffic planning and control (S. P. Hoogendoorn, Botma, and Minderhoud, 2007)[p.83-84]. Determining a road's capacity at equilibrium is critical for evaluating the impact of features such as signs, traffic lights and traffic policies such as diversion strategies on vehicle flow and density. Moreover, traffic flow, speed and density are related to travel time estimates. Such estimates play a pivotal role in traffic assignment which consists of assigning travel routes to vehicles based on an optimality condition such as travel-time minimisation. As a result, well-calibrated FDs are integral to the performance of traffic forecasting systems.

1.5 Problem statement

So far we have introduced a conservation law (traffic PDE) to model traffic wave propagation in space-time and discussed its variants/extensions. While it can be argued that the LWR model in (1.1) misspecifies traffic reality, the conservation of vehicles constraint is valid across any single entry/exit road segment i.e. the conservation law is epistemic and uncertainty-free. Conversely, constitutive laws such as FD relationships are empirical and uncertain as they rely on assumptions **P1** - **P6** which may not necessarily hold in generality. Therefore, misspecified FDs incur errors and uncertainty to PDE solutions which are amplified by the complexity and dynamic nature of traffic phenomena. As a result, it is vital to remove any ad-hoc decisions from FD model selection and quantitatively assess different model hypotheses based on their complexity and ability to match data.

In the following section we discuss the state-of-the-art methodologies for calibrating and comparing different FD models (constitutive laws) on data.

1.6 Related work

Empirical calibration and validation of constitutive laws has attracted significant attention in the transportation literature. Despite the existence of significant contributions in assimilating probe vehicle data (trajectory data from GPS devices) (Seo et al., 2019), we restrict our focus on loop detector data due to their large spatio-temporal coverage of UK motorways. However, it can argued that approaches that fuse Eulerian with Lagrangian data are better suited for areas where sensor coverage or resolution is poor (Ambühl and Menendez, 2016; Neumann, Böhnke, and Tcheumadjeu, 2013).

Constitutive law estimation approaches are split into non-parametric and parametric. Non-parametric approaches make no assumptions of the form **P1**-**P5**. Instead, they allow data to drive the shape of the FD. (Chen et al., 2004) leveraged a generalisation of principal component regression from a family of lines to a family of curves called *principal curves*. Principal

curve regression extracts the latent vector space "generating" the observed occupancy-speed data along which the explained variance of the data is maximised. Although appealing this approach makes limiting assumptions about the data generating mechanism and suffers from three deficiencies. Namely, the inability to predict flow responses for different values of the density covariate, the absence of a complexity penalty term in the regression model, and the manual filtering of outliers limit this approach's ability to generalise to noisy loop detector datasets.

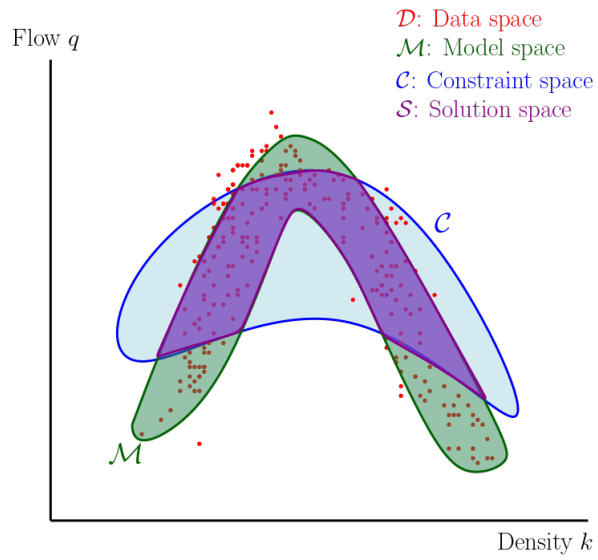


Figure 1.6: Schematic for parametric model paradigm of constitutive law learning. The model space is defined by the set of all plausible parameters consistent with the data space. The solution space is the intersection of the model and constraint spaces where the latter is induced by the conservation law in (1.1).

Despite their flexibility, non-parametric methods often fail at discriminating between flows in the free-flow and congested regimes (S. H. Huang and Ran, 2003). In recognition of this challenge (Einbeck and Tutz, 2006) proposed extracting the *local modes* of the conditional distribution of speed given flow for different flow values and fitting a regression curve over the collection of modes. This is advantageous in that probabilistic estimates of (un)congested branch membership of data points are made which enhances model interpretability in non-parametric constitutive law estimation. Nevertheless, the fitted curves are smooth relations in $C^1([0,\infty))$ instead of functions and therefore cannot be embedded in conservation laws. Another limitation arises from the absence of a regulariser in regression or a thorough discussion of the trade-off between model complexity and data fit.

A Bayesian non-parametric model suitable for constitutive laws embedded within conservation laws is Gaussian processes which quantifies uncertainty about the data-generating process. In the recent work of (Fuhg, Marino, and Bouklas, 2021) GPs were employed for

surrogate modelling of a three-dimensional constitutive law in solid mechanics. Due to their Bayesian nature, a GP's marginal likelihood reward sufficient data fits while penalising model complexity. The computational burden of fitting GPs was eased by leveraging a space-filling experimental design technique to obtain representative subsets of stress-strain data. Although constitutive laws have bounded constraints (in our case flow, density and speed are non-negative), the function space of GPs is loosely constrained (e.g. space of smooth functions). Recent GP model developments induce physical constraints on the GP posterior by applying compositions of transformations (Maroñas et al., 2020). However, to our knowledge this methodology has not yet been applied to surrogate modelling of constitutive laws in traffic.

In the spirit of model-free approaches, a data-driven paradigm was introduced by (Kirchdoerfer and Ortiz, 2016) in computational mechanics and conceptualised in Figure 1.7. The authors examined the elasticity problem and inferred the material constitutive law purely by identifying the stress-strain data points closest to satisfying the governing localisation and conservation laws or a relaxation thereof. The resulting collection of data points is compared against commonly accepted material laws such as Hooke's law. The data-driven solver's greediness was later alleviated in (Kirchdoerfer and Ortiz, 2017) to account for the noise in material datasets. Although a probabilistic data-driven solver was developed, the solver has yet to formally account for the uncertainty in material data in a statistical treatment of the problem. We argue that a Bayesian treatment instill material law robustness to outliers. Also, uncertainty-aware material laws can guide future experiments for material data collection by

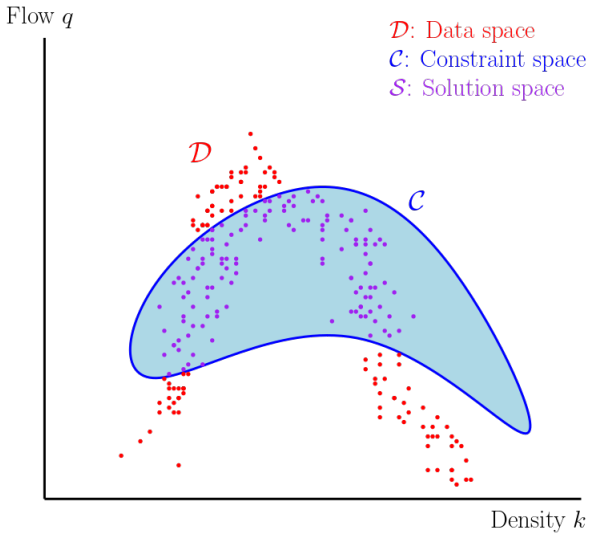


Figure 1.7: Schematic for model-free paradigm of constitutive law learning. The solution space is the intersection of the solution and constraints spaces where the latter is induced by the conservation law in (1.1).

emphasising on uncertain areas of the stress-strain *phase space*.

The limited proliferation of non-parametric model-free methods in traffic constitutive law estimation may be partly attributed to the lack of interpretable parameters such as the ones appearing in their parametric counterparts. As a result, both GPs and model-free paradigms lack the desired level of model interpretability that transportation engineers are interested in. We conclude that the precise motivation for model-free methods in transportation literature is lacking and make the case for a holistic evaluation of parametric approaches to FD estimation.

In contrast, parametric FDs have interpretable and operationally useful parameters. The parametric paradigm is conceptualised in Figure 1.6. FD parameters are inferred based on unconstrained minimisation of an L_2 norm of the model error (Castillo, 2012b; Qu, S. Wang, and J. Zhang, 2015). Least squares regression's propensity to *over-fitting* (accepting more complex model hypotheses than necessary) makes it inferior to any regression method with a regularisation term such as a Tiknhonov regulariser (Golub, Hansen, and O'Leary, 1999). The same limitation applies to a weighted regression model proposed in (Jiang and Y.-X. Huang, 2009) and (Knoop and Daamen, 2017) to learn FD parameters. The regression model assigns weights to speed-density data in an ad-hoc manner or requires a large number of weights to be learned without effectively penalising model complexity. An attempt to infer the different mixtures of speed-density data via k-means clustering prior to applying locally weighted regression is also made in (Jiang and Y.-X. Huang, 2009). The resulting hierarchical model achieved no significant reduction in root mean square percent error (RMSPE) while adding complexity to the modelling framework.

The deterministic error-minimising parameter estimation schemes have been replaced by approaches that formally account for the stochasticity emerging from imperfect loop detector devices and complex traffic dynamics. The work of (H. Wang, Ni, et al., 2013) suggested the use of infinite Karhunen-Loéve (KL) basis expansions which define a stochastic process in square-integrable space $L_2(\Omega)$ and is used to represent a stochastic FD. The stochastic FD is decomposed into a deterministic part of the form found in Table 1.1 and a stochastic component. The latter component is represented by an additive heteroscedastic error structure which

is justified by empirical evidence that traffic flow variation varies with density. However, enforcing positivity via a multiplicative non-heteroscedastic structure is likely to capture enough of that variation while also ensuring that the resulting FD is non-negative and therefore physically plausible. Also, least squares regression is used to relate FD parameters to circumvent *identifiability* issues.

A similar probabilistic treatment of FDs is found in the works of (Polson and Sokolov, 2015) and (Coullon and Pokern, 2020), where they focus on inferring the latent space of traffic densities "driving" traffic density observations. (Polson and Sokolov, 2015) learn that space at each discretisation cell jointly with the parameters of Daganzo's FD (see $Q_8(k)$ in Table 1.1) in an online (real-time) setting using a particle filter. Their choice of FD is solely justified by its computational savings achieved in numerical solvers of the traffic PDE. The dynamic learning of parameters yields a non-equilibrium (non-static) FD whose flexibility is limited by its triangular shape. (Coullon and Pokern, 2020) solve the Bayesian inverse problem in an offline setting instead. They employ a Poisson error model to account for vehicles appearing in each of the four lanes of the road of interest, enforce positivity on the flow values and exploit that the sum of independent and identically distributed (i.i.d.) Poisson random variables is also Poisson distributed. Despite the benefits of this error structure, the model fails to account for the apparent heteroskedasticity in the data. A direct fit of Daganzo's and DelCastillo's models using a *Random Walk Metropolis Hastings* (RWMH) show a misfit to the data which can be partly attributed to the Poisson error structure. The resulting congested flow waves do not cross the (x, t) domain. Whereas there is still significant FD model misfit, the congested flow waves appear in the domain when the FD is fitted to data by solving the inverse problem.

In terms of model validation and comparison of parametric FD forms, comparative studies discuss the difference in the achieved RMSE, RMSPE or similar error metrics (Jiang and Y.-X. Huang, 2009; Castillo, 2012a). Although these type of metrics are adequate for evaluating goodness-of-fit, they fail to account for model complexity and can therefore score complex models highly. This limitation was acknowledged in (Castillo, 2012a) where the authors motivated the need to examine the plausibility each parameter. Granted that a probabilistic

parameter treatment (Polson and Sokolov, 2015; Coullon and Pokern, 2020) allows for such considerations, to our knowledge a probabilistic framework for constitutive model selection has not been introduced in the literature yet.

On this end, we identify two main gaps in FD calibration and validation literature. First, there is no systematic way of assessing traffic constitutive laws with regards to their achieved trade-off between model complexity (dimension of parameter space) and data fit. Secondly, model-free and non-parametric approaches outlined in (Kirchdoerfer and Ortiz, 2016) and (Fuhg, Marino, and Bouklas, 2021), respectively, have not yet been applied to traffic constitutive and conservation laws and their performance has not been appraised.

1.7 Current work and research objectives

In this report we address the first gap by following a Bayesian approach adopted in (Polson and Sokolov, 2015; Coullon and Pokern, 2020), leveraging the robustness of the *marginal likelihood* (ML) term to model complexity and therefore introducing a probabilistic framework for systematically comparing parametric constitutive laws. We acknowledge the importance of tackling both problems and argue that the former is a natural first step that can holistically evaluate parametric FD estimation and clearly identify the need for non-parametric or model-free approaches.

We improve the state-of-the-art traffic constitutive law validation by extending the works of (Polson and Sokolov, 2015; Coullon and Pokern, 2020) and computing unbiased estimates of the marginal likelihood leveraging concepts from thermodynamic integration and estimating the resulting tempered posteriors using a Metropolis Hastings algorithm (Calderhead and Girolami, 2009). We highlight that the marginal likelihood inherently penalises complex model hypotheses and rewards good data fits contrary to any other error-based metrics proposed in the traffic literature. Our approach is to be contrasted with regularised regression in that we rely on the prior specification to penalise model complexity while penalised regression hinges on the choice of regularisation parameter(s). Moreover, our approach is probabilistic

and accounts for the uncertainty in the data whereas penalised regression is deterministic. The work of (Friel and Pettitt, 2008) discussed the marginal likelihood estimate's sensitivity to prior variance. In recognition of this limitation we compute marginal likelihood estimates for three different prior specifications corresponding to diffuse, concentrated and "regular" priors. In doing so, we propose an adaptive MH proposal to efficiently sample from the tempered posteriors. This methodology is applied to model calibration and comparison of the ten functional FDs listed in Table 1.1 using both synthetic and real data from the UK's M25 motorway. Finally, we release our reproducible codebase in this GitHub repository.

This report aims to answer the following research question:

Which parametric constitutive laws embedded within traffic PDEs achieve the best trade-off between model complexity and data fit?

The research question can be decomposed into the following research objectives:

1. Examine the structural differences between the proposed parametric FDs in Table 1.1.
2. Investigate the effect of prior diffusivity on the model evidence of the aforementioned FDs.
3. Discuss the suitability of each model in modelling synthetic and real-world flow-density data.

1.8 Synopsis

Chapter 1 introduced the societal need for assimilating data into transportation models and laid out the foundations of microscopic and macroscopic traffic flow modelling. Statistical inference in microscopic agent-based models is discussed and open problems for future research are identified. Macroscopic traffic conservation and constitutive laws are introduced and critically evaluated. The chapter ends with a review of the existing approaches to constitutive law

estimation, an identification of two key literature gaps and a proposal to improve the state-of-the-art parametric FD model assessment. In Chapter 2 we describe the statistical machinery employed to obtain unbiased estimates of the marginal likelihood of constitutive laws in Table 1.1. We also introduce two black-box-driven approaches to ABM calibration that will be leveraged in the rest of the PhD. In Chapter 3 we discuss our experimental results in response to the aforementioned research objectives. Chapter 4 outlines a detailed plan and research objectives for next the two years of research derived from the statistical ABM calibration review.

Chapter 2

Statistical methodology

The purpose of this Chapter is twofold: to describe the statistical methodology used in Chapter 3 to robustly assess constitutive laws and to introduce two black-box-driven approaches to ABM calibration that will be used in the PhD.

2.1 Robust model assessment methodology

2.1.1 Bayesian inference

We define the vectors of FD flow and density data by $\mathbf{q}$,$\mathbf{k}$ respectively where q_i,k_i correspond to the i-th flow-density pair. We further denote the j-th fundamental diagram model of consideration in Table 1.1 by $M_j = Q_j(k)$ and its associated parameters by $\boldsymbol{\theta}_j$. We begin by assuming the following multiplicative error generative model about the noisy flow-density relationship

$$\log(q_i) = \log(q_i(\boldsymbol{\theta}_j)) + \log(\epsilon_i), \quad \forall i = 1, \dots, N_d, \tag{2.1}$$

where $\xi_i \triangleq \log(\epsilon_i)$ are assumed to be *independent and identically distributed* (i.i.d.) Gaussian random variables $\xi_i \sim N(0, \sigma^2)$, and $\log(q_i(\boldsymbol{\theta}_j))$ is the j-th FD ($Q_j(k)$) evaluated at the i-th density point using parameters $\boldsymbol{\theta}_j$. Therefore, the error distribution induces a distribution $p(\mathbf{q}|M_j, \boldsymbol{\theta}_j)$ over the *likelihood* of the observed flow data conditional on the FD model output. We can express the log-likelihood in log-data scale by applying the Jacobian transformation of

log denoted by $\mathbf{J}(\mathbf{q})$

$$p(\mathbf{q}|\boldsymbol{\theta}_j) = p(\log(\mathbf{q})|\boldsymbol{\theta}_j) \det \left| \frac{\partial \log(\mathbf{q})}{\partial \mathbf{q}} \right| = p(\log(\mathbf{q})|\boldsymbol{\theta}_j) \prod_{l=1}^{N_d} \frac{1}{q_l} \tag{2.2}$$

where $\log(\mathbf{q}|\boldsymbol{\theta}_j) \sim \mathcal{N}(\log(\mathbf{q}(\boldsymbol{\theta}_j)), \sigma^2\mathbf{I})$.

We follow a Bayesian approach and define a set of *prior* beliefs over the j-th FD model parameters $\boldsymbol{\theta}_j$ in the form of a probability distribution denoted by $\pi_0(\boldsymbol{\theta}_j)$. By Bayes rule it follows that the *posterior* distribution over the FD parameters is equal to

$$\pi(\boldsymbol{\theta}_j|\mathbf{q}) = \frac{p(\mathbf{q}|\boldsymbol{\theta}_j)\pi_0(\boldsymbol{\theta}_j)}{\int p(\mathbf{q}|\boldsymbol{\theta}_j)\pi_0(\boldsymbol{\theta}_j)\,\mathrm{d}\boldsymbol{\theta}_j}, \tag{2.3}$$

We have removed the dependence on M_j in the above equation for notational convenience. We also note that the denominator term

$$p(\mathbf{q}|M_j) = \int p(\mathbf{q}|\boldsymbol{\theta}_j, M_j)\pi(\boldsymbol{\theta}_j|M_j)\,\mathrm{d}\boldsymbol{\theta}_j \tag{2.4}$$

is called *model evidence* or *marginal likelihood* (ML). The marginal likelihood encodes the evidence provided by the underlying model hypothesis postulated by the constitutive law. The plausibility of two model hypotheses can be objectively evaluated against data by using Bayes factors:

$$\underbrace{\frac{p(M_j|\mathbf{q})}{p(M_l|\mathbf{q})}}_{\text{posterior odds}} = \underbrace{\frac{p(M_j)}{p(M_l)}}_{\text{prior odds}} \underbrace{\frac{p(\mathbf{q}|M_j)}{p(\mathbf{q}|M_l)}}_{\text{Bayes factor}}, \tag{2.5}$$

where the Bayes factor (BF) is the ratio of marginal likelihoods for models $j, l \in \{1, \ldots, 10\}$. We assume a uniform prior over the model universe $\mathcal{M} = \{Q_j(k) : j = 1, \ldots, 10\}$ of constitutive laws and therefore the posterior odds equal the Bayes factor. This assumption reflects our desire to remain agnostic about the most plausible constitutive law. Table 2.1 gives a scale for interpreting BFs suggested in (Jeffreys, 1998, app. B).

2.1.1.1 Robustness of marginal likelihood

Furthermore, an intuitive explanation of the marginal likelihood can be obtained by considering an approximation to the integral in (2.4) (Svensén and Bishop, 2007, ch. 3.4). Consider the

$\log_{10}(BF)$	BF	**Strength of evidence**
0 to 1/2	1 to 3.2	Not worth more than a bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	> 100	Decisive

Table 2.1: Scale for interpreting strength of model evidence favouring M_j over M_l based on log Bayes factors where M_l is the null hypothesis while M_j is the alternative one.

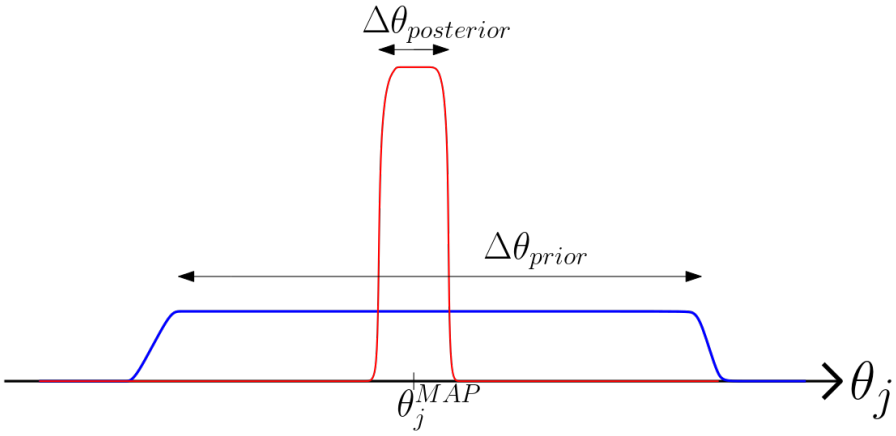


Figure 2.1: Schematic of idealised concentrated posterior around model θ_j^{MAP} and flat prior distributions over parameter θ_j.

case when $dim(\boldsymbol{\theta_j}) = 1$ and $\boldsymbol{\theta_j} = (\theta_j)$. We assume a flat prior and a parameter posterior that is concentrated around the posterior mode θ_j^{MAP} with width $\Delta\theta_{posterior}$ as shown in Figure 2.1. Therefore, we can approximate the integral as follows:

$$p(\mathbf{q}|M_j) = \int p(\mathbf{q}|\theta_j, M_j)p(\theta_j|M_j)\,\mathrm{d}\theta_j \approx p(\mathbf{q}|\theta_j^{MAP})\frac{\Delta\theta_{posterior}}{\Delta\theta_{prior}}, \tag{2.6}$$

where by prior flatness $p(\theta_j) = \frac{1}{\Delta\theta_{prior}}$. Taking logs in the expression above yields

$$\log(p(\mathbf{q}|M_j)) \approx \underbrace{\log\left(p(\mathbf{q}|\theta_j^{MAP})\right)}_{\text{data fit}} + \underbrace{\log\left(\frac{\Delta\theta_{posterior}}{\Delta\theta_{prior}}\right)}_{\text{complexity penalty}}. \tag{2.7}$$

The first term in the above equation quantifies the quality of data fit given by the most plausible parameter configuration. The second term is negative when the ratio $\frac{\Delta\theta_{posterior}}{\Delta\theta_{prior}} < 1$ and increases in magnitude when that ratio becomes smaller. A large complexity penalty in incurred when there is a small set of plausible parameters in the posterior distribution (narrow posterior). In other words, for a given choice of data the model has a limited number of parameter configurations that can explain that data well and is therefore more complex than necessary. Equation (2.7) is extended to a multidimensional parameter vector as follows:

$$\log(p(\mathbf{q}|M_j)) \approx \underbrace{\log\left(p(\mathbf{q}|\boldsymbol{\theta}_j^{MAP})\right)}_{\text{data fit}} + \underbrace{\left|\boldsymbol{\theta_j}\right| \log\left(\frac{\Delta\boldsymbol{\theta}_{posterior}}{\Delta\boldsymbol{\theta}_{prior}}\right)}_{\text{complexity penalty}}, \tag{2.8}$$

where $\left|\boldsymbol{\theta_j}\right| \triangleq dim(\boldsymbol{\theta_j})$ is the number of parameters in model M_j. As the dimensionality of the parameter vector increases so does the model complexity term in magnitude. However, a complex model is likely to achieve an increase in the data fit term due to its added flexibility in explaining the data. Therefore, a trade-off between between model complexity and data fit emerges. Although the approximation in (2.8) is crude it provides sufficient intuition about this trade-off when interpreting marginal likelihoods.

A further insight can be gained by examining how the marginal likelihood favours models of intermediate complexity. Consider a distribution over possible datasets $\mathcal{D}$ and three models M_1, M_2, M_3 of increasing complexity. The datasets are generated by sampling from the parameter prior $\boldsymbol{\theta}^{(i)} \sim \pi_0(\boldsymbol{\theta})$ and then simulating $p(\mathcal{D}|\boldsymbol{\theta}^{(i)}) \quad \forall \quad i$. Figure 2.2 illustrates theoretical distributions over datasets for each of the three models. Simple models generate datasets of limited variability that can be easily reproduced by other more complex models. On the other hand, complex models can generate vastly dissimilar datasets that may not be adequately represented by simple models. Given that the distributions in Figure 2.2 are normalised, dataset $\mathcal{D}_0$ achieves the highest evidence for the model M_2 of intermediate complexity.

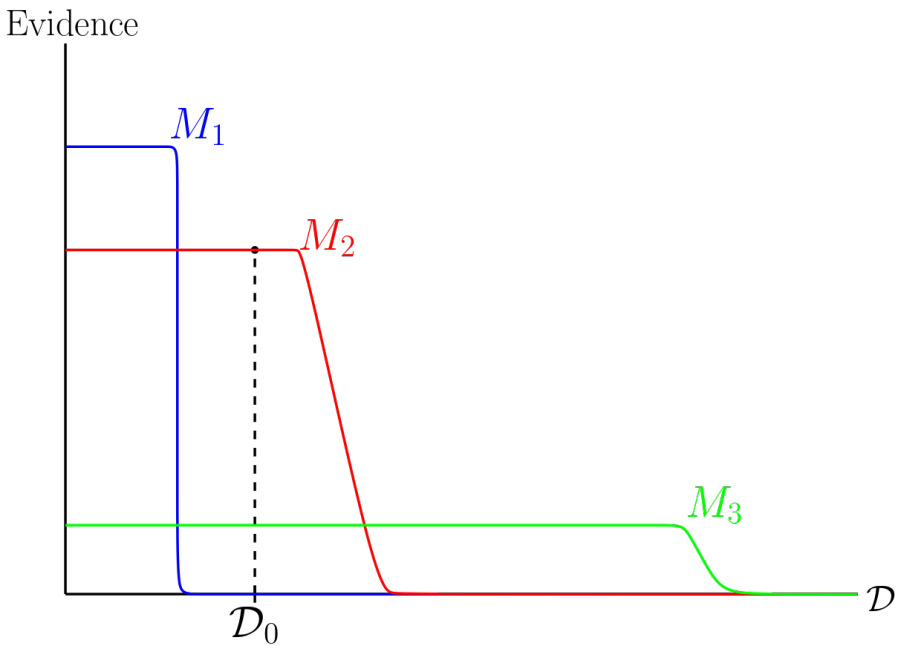


Figure 2.2: Schematic illustration of normalised distributions over datasets $\mathcal{D}$ for a simple, an intermediate and a complex model M_1,M_2,M_3, respectively. Dataset $\mathcal{D}_0$ achieves the highest evidence under M_2.

This is because the simpler model poorly fits the data while the complex model's predictive distribution has high variance and so assigns small probability to each dataset. We note that it is possible for the Bayes factor to misidentify the correct model in a finite dataset. That is a why an expectation over the distribution of datasets has to be taken when computing BFs to ensure that the correct model is identified *on expectation* (assuming the "true" model exists in the model universe).

2.1.1.2 Comparison with R^2

On the contrary, the goodness-of-fit (R^2) criterion rewards complex models when used in model selection. First, we derive the posterior predictive for model M_j

$$p\left(\hat{\mathbf{q}}|\mathbf{q},\mathbf{k},\mathbf{k}^*\right) = \int \underbrace{p\left(\hat{\mathbf{q}}|\mathbf{q},\mathbf{k},\mathbf{k}^*,\boldsymbol{\theta}_j\right)}_{\text{predictive likelihood}} \underbrace{p\left(\boldsymbol{\theta}_j|\mathbf{q},\mathbf{k},\mathbf{k}^*\right)}_{\text{posterior}} \mathrm{d}\boldsymbol{\theta}_j \tag{2.9}$$

where $\mathbf{k}^*$ are the given traffic densities at which the posterior predictive is evaluated at, N_p are the number of MC samples, and $\boldsymbol{\theta}^{(n)} \sim p\left(\boldsymbol{\theta}_j|\mathbf{q},\mathbf{k},\mathbf{k}^*\right)$. We then compute the expectation

$$\begin{aligned}\mu(\hat{\mathbf{q}}) \triangleq \mathbb{E}_{\hat{\mathbf{q}}|\mathbf{q},\mathbf{k},\mathbf{k}^*}\left[\hat{\mathbf{q}}\right] &= \int \hat{\mathbf{q}}\, \underbrace{p\left(\hat{\mathbf{q}}|\mathbf{q},\mathbf{k},\mathbf{k}^*\right)}_{\text{posterior predictive}} \mathrm{d}\hat{\mathbf{q}} \\ &\approx \frac{1}{N_p}\sum_{n=1}^{N_p} \hat{\mathbf{q}}^{(n)},\end{aligned} \tag{2.10}$$

where $\hat{\mathbf{q}}^{(n)} \sim p\left(\hat{\mathbf{q}}|\boldsymbol{\theta}_j^{(n)},\mathbf{k},\mathbf{k}^*,\mathbf{q}\right)$ and $\boldsymbol{\theta}_j^{(n)} \sim p\left(\boldsymbol{\theta}_j|\mathbf{q},\mathbf{k},\mathbf{k}^*\right)$. Subsequently, R^2 is computed as follows:

$$R^2 = 1 - \frac{\sum_{l=1}^{N_d}\left(q_l - \mu(\hat{q}_l)^2\right)}{\sum\left(q_l - \bar{q}\right)^2}, \tag{2.11}$$

where $\bar{q} \triangleq \frac{1}{N_d}\sum_{l=1}^{N_d} q_l$.

2.1.2 Marginal likelihood estimation

In all but simple cases the marginal likelihood in (2.4) is intractable and numerical methods have been employed to compute it (C. Han and Carlin, 2001). Traditional estimators such as

the posterior harmonic mean (Raftery et al., 2006) and the prior arithmetic mean (McCULLOCH and Rossi, 1992) estimators have been rendered unsuitable for ML estimation as they introduce bias for finite samples.

We resort to a new way of estimating the ML based on ideas of path sampling or thermodynamic integration (Gelman and Meng, 1998). In the remainder of this section we assume that $\boldsymbol{\theta}_j = \boldsymbol{\theta}$. Let $t \in [0, 1]$ be an auxiliary variable called *temperature* and define the *power posterior* to be

$$p_t(\boldsymbol{\theta}|\mathbf{q}) \propto p(\mathbf{q}|\boldsymbol{\theta})^t \pi_0(\boldsymbol{\theta}) \tag{2.12}$$

with a normalising constant equal to

$$z(\mathbf{q}|t) \triangleq \int p(\mathbf{q}|\boldsymbol{\theta})^t \pi_0(\boldsymbol{\theta}) \, \mathrm{d}\theta. \tag{2.13}$$

We note that by construction $z(\mathbf{q}|t = 0) = 1$ since the prior normalised to one. Moreover, $z(\mathbf{q}|t = 1)$ is the marginal likelihood in (2.4). It follows that

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}|\mathbf{q},t}\left[\log p(\mathbf{q}|\boldsymbol{\theta})\right] &= \int \log(p(\mathbf{q}|\boldsymbol{\theta}) p_t(\boldsymbol{\theta}|\mathbf{q}) \, \mathrm{d}\boldsymbol{\theta} \\
&= \frac{1}{z(\mathbf{q}|t)} \int \log(p(\mathbf{q}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{q})^t \pi_0(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} \\
&= \frac{1}{z(\mathbf{q}|t)} \frac{\mathrm{d}}{\mathrm{d}t} \left(\int p(\boldsymbol{\theta}|\mathbf{q})^t \pi_0(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} \right) \\
&= \frac{1}{z(\mathbf{q}|t)} \frac{\mathrm{d}}{\mathrm{d}t} \left(z(\mathbf{q}|t) \right) \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \left(\log \left(z(\mathbf{q}|t) \right) \right),
\end{aligned} \tag{2.14}$$

where line 3 follows from $\frac{\mathrm{d}a^x}{\mathrm{d}x} = \log(a)a^x$ and by integral-derivative interchangeability. By integrating (2.14) in the interval $[0, 1]$ we get that

$$\int_0^1 \mathbb{E}_{\boldsymbol{\theta}|\mathbf{q},t}\left[\log p(\mathbf{q}|\boldsymbol{\theta})\right] = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} \left(\log \left(z(\mathbf{q}|t) \right) \right) \mathrm{d}t = \log(p(\mathbf{q})). \tag{2.15}$$

By computing the ML on the log scale we ensure numerical stability of our estimates. Treating t as a random variable with prior distribution $p(t)$ allows us to sample jointly from $p(\boldsymbol{\theta}, t)$:

$$\log(p(\mathbf{q})) = \int_0^1 \frac{\mathbb{E}_{\boldsymbol{\theta}|\mathbf{q},t}\left[\log p(\mathbf{q}|\boldsymbol{\theta})\right] p(t)}{p(t)} \, \mathrm{d}t = \mathbb{E}_{\boldsymbol{\theta},t|\mathbf{q}}\left[\frac{\log p(\mathbf{q}|\boldsymbol{\theta})}{p(t)} \right]. \tag{2.16}$$

2.1.2.1 Thermodynamic integration

We discretise the integral over $[0, 1]$. We solve (2.12),(2.13) for $\pi_0(\boldsymbol{\theta})$ and express and rearrange the following equality

$$\frac{z(\mathbf{q}|t_n)p(\boldsymbol{\theta}|\mathbf{q}, t_n)}{p(\mathbf{q}|\boldsymbol{\theta})^{t_n}} = \frac{z(\mathbf{q}|t_{n-1})p(\boldsymbol{\theta}|\mathbf{q}, t_{n-1})}{p(\mathbf{q}|\boldsymbol{\theta})^{t_{n-1}}}$$
$$\therefore \frac{z(\mathbf{q}|t_n)}{z(\mathbf{q}|t_{n-1})}p(\boldsymbol{\theta}|\mathbf{q}, t_n) = p(\mathbf{q}|\boldsymbol{\theta})^{\Delta t_n}p(\boldsymbol{\theta}|\mathbf{q}, t_{n-1}) \tag{2.17}$$

which we assume to hold for $\Delta t_n \triangleq t_n - t_{n-1} \to 0$ and $0 = t_1 < \cdots < t_N = 1$. We proceed by taking logarithms, multiplying both sides by $p(\boldsymbol{\theta}|\mathbf{q}, t_n)$ and integrating with respect to $\boldsymbol{\theta}$ in (2.17):

$$\int \log\left(\frac{z(\mathbf{q}|t_n)}{z(\mathbf{q}|t_{n-1})}\right) p(\boldsymbol{\theta}|\mathbf{q}, t_n)\,\mathrm{d}\boldsymbol{\theta} + \int \log\left(p(\boldsymbol{\theta}|\mathbf{q}, t_n)\right) p(\boldsymbol{\theta}|\mathbf{q}, t_n)\,\mathrm{d}\boldsymbol{\theta} =$$
$$\Delta t_n \int \log(p(\mathbf{q}|\boldsymbol{\theta}))p(\boldsymbol{\theta}|\mathbf{q}, t_n)\,\mathrm{d}\boldsymbol{\theta} + \int \log\left(p(\boldsymbol{\theta}|\mathbf{q}, t_{n-1})\right) p(\boldsymbol{\theta}|\mathbf{q}, t_n)\,\mathrm{d}\boldsymbol{\theta}$$
$$\therefore \log\left(\frac{z(\mathbf{q}|t_n)}{z(\mathbf{q}|t_{n-1})}\right) = \mathbb{E}_{\boldsymbol{\theta}|\mathbf{q},t_n}\left[\Delta t_n \log(p(\mathbf{q}|\boldsymbol{\theta}))\right] + \int \log\left(\frac{p(\boldsymbol{\theta}|\mathbf{q}, t_{n-1})}{p(\boldsymbol{\theta}|\mathbf{q}, t_n)}\right) p(\boldsymbol{\theta}|\mathbf{q}, t_n)\,\mathrm{d}\boldsymbol{\theta}, \tag{2.18}$$

where the second term on the RHS is equal to the *Kullback Leibler* divergence $KL(p_n||p_{n-1}) \geq 0$. By discretisation of the integral we can derive an upper bound on the integral estimate

$$\log(p(\mathbf{q})) = \sum_n \log\left(\frac{z(\mathbf{q}|t_n)}{z(\mathbf{q}|t_{n-1})}\right)$$
$$= \sum_n \left[\underbrace{\mathbb{E}_{\boldsymbol{\theta}|\mathbf{q},t_n}\left[\log(p(\mathbf{q}|\boldsymbol{\theta}))\right]\Delta t_n}_{\text{upper bound}} - \underbrace{KL(p_n||p_{n-1})}_{\text{bias}}\right]. \tag{2.19}$$

We repeat this procedure to derive a lower bound on the integral estimate by multiplying (2.17) by $p(\boldsymbol{\theta}|\mathbf{q}, t_{n-1})$ which yields a lower bound for the integral estimate

$$\log(p(\mathbf{q})) = \sum_n \log\left(\frac{z(\mathbf{q}|t_n)}{z(\mathbf{q}|t_{n-1})}\right)$$
$$= \sum_n \left[\underbrace{\mathbb{E}_{\boldsymbol{\theta}|\mathbf{q},t_{n-1}}\left[\log(p(\mathbf{q}|\boldsymbol{\theta}))\right]\Delta t_n}_{\text{lower bound}} + \underbrace{KL(p_{n-1}||p_n)}_{\text{bias}}\right]. \tag{2.20}$$

We note that in the limit of $\Delta t_n \to 0$ the power posteriors p_n,p_{n-1} will become indistinguishable causing the bias $KL(p_n||p_{n-1}) \to 0$ and therefore

$$\lim_{\Delta t_n \to 0} \sum_n \mathbb{E}_{\boldsymbol{\theta}|\mathbf{q},t_n} \left[\log(p(\mathbf{q}|\boldsymbol{\theta}))\right] \Delta t_n \to \int_0^1 \mathbb{E}_{\boldsymbol{\theta}|\mathbf{q},\boldsymbol{t}}[\log(p(\mathbf{q}|\boldsymbol{\theta})]dt$$

By taking the average over the lower and upper bounds we derive the trapezium rule for numerical integration of the marginal likelihood integral

$$\log\left(p(\mathbf{q})\right) = \underbrace{\frac{1}{2}\sum_n \Delta t_n \left(\mathbb{E}_{\boldsymbol{\theta}|\mathbf{q},t_n}\left[\log\left(p(\mathbf{q}|\boldsymbol{\theta})\right)\right] + \mathbb{E}_{\boldsymbol{\theta}|\mathbf{q},t_{n-1}}\left[\log\left(p(\mathbf{q}|\boldsymbol{\theta})\right)\right]\right)}_{\text{approximation}} + \underbrace{\frac{1}{2}\sum_n \left[KL(p_{n-1}||p_n) - KL(p_n||p_{n-1})\right]}_{\text{bias}}. \tag{2.21}$$

It is evident that there are two sources of error in the estimation of the marginal likelihood. First, the expectations in the approximation term entail a Monte Carlo sampling error which is reduced based on the number of samples and the efficiency of the sampling scheme. Second, the bias term can be diminished subject to a suitable dense temperature schedule of t_n's.

2.1.2.2 Choice of temperature schedule

There are three popular choices of temperature schedule suggested in (Calderhead and Girolami, 2009) and listed in Table 2.2. We select a prior temperature schedule with $N = 30$, $p = 5$ which achieves one of the lowest biases when computing MLs for linear models in (Calderhead and Girolami, 2009). We note that successive temperature spacing Δt_n increases as the number of temperatures increases. This means that the most dominant terms in (2.21) are traced in high temperatures where the power posteriors are mostly dominated by the likelihood (assuming a flat or non-informative prior) compared to posterior schedules which are prior-dominated in high temperatures. An in-depth consideration of the choice of temperature is outside the scope of this report.

Name	Schedule
Uniform	$t_i = \frac{i}{N}$
Prior	$t_i = \left(\frac{i}{N}\right)^p$
Posterior	$t_i = 1 - \left(\frac{i}{N}\right)^p$

Table 2.2: Geometric-based temperature schedules suggested in (Calderhead and Girolami, 2009).

2.1.3 Posterior sampling

Metropolis Hastings Markov Chain Monte Carlo (MH-MCMC) sampling is employed to sample from the target density $p(\boldsymbol{\theta}|\mathbf{q}, t_n) \ \ \forall\ n = 1, \ldots, N$. Equivalently, by conditional independence between temperature conditionals it follows that

$$p(\boldsymbol{\theta}|\mathbf{q}, \mathbf{t}) \propto \prod_{n=1}^{N} \pi_0(\boldsymbol{\theta}) p(\mathbf{q}|\boldsymbol{\theta})^{t_n}. \tag{2.22}$$

Although it can be supported that more sophisticated MCMC schemes such as population MCMC (Friel and Pettitt, 2008; Calderhead and Girolami, 2009) could be applied in our case, we argue that a simple MH sampler with a carefully constructed proposal mechanism could efficiently sample from the power posteriors. We note that the conditional independence between $p(\boldsymbol{\theta}|\mathbf{q}, t_n)$ facilitates sampling parallelisation across temperature MCMC chains. Also, MH-MCMC is a special case of population MCMC with no global moves in the parameter space. The MH-MCMC algorithm is outlined in Algorithm 1.

There is a number of design choices that have to be made. First, we need to select appropriate priors that preserve the positivity of the constitutive law parameters in Table 1.1. By choosing priors that only provide support for $\mathbb{R}_{>0}$ does not solve the problem as the transition kernel may propose negative samples. Therefore, we use Gaussian priors for all parameters of the ten FDs of interest and transform them in log-space to ensure their positivity following the approach of (Thawornwattana, Dalquen, and Yang, 2018). Thus, we avoid reflecting proposals and reduce the effective samples generated and transform the prior

Algorithm 1 Metropolis Hastings MCMC algorithm for sampling from joint target density $p(\boldsymbol{\theta}|\mathbf{q}, \mathbf{t})$.

Function call: `MH-MCMC`($\boldsymbol{\theta}_0$,$T(\cdot,\cdot)$,$\mathbf{t}$,$p(t)$,N)
Input: Initial parameter vector $\boldsymbol{\theta}_0$, transition kernel $T(\cdot,\cdot)$, temperature schedule $\mathbf{t}$, temperature distribution $p(t)$, number of chains N.
Output: Samples $(\boldsymbol{\theta}^{(i)}, t_i) \ \ \forall\, i = 1, \ldots, N$.
Initialise $\boldsymbol{\theta}_{prev} = \boldsymbol{\theta}_0$ for all temperatures $\mathbf{t}$.
for $i \in \{1, \ldots, N\}$ **do**
    Sample from temperature distribution $t_{new} \sim p(t)$.
    Propose new sample $\boldsymbol{\theta}_{new} \sim T(\cdot, \boldsymbol{\theta}_{new})$ for power posterior corresponding to t_{new}.
    Evaluate acceptance ratio $\alpha' = \alpha(\boldsymbol{\theta}_{new}, \boldsymbol{\theta}_{prev})$
    Sample $u \sim Uniform(0, 1)$.
    if $\alpha' > u$ **then**
        Accept new sample $\boldsymbol{\theta}^{(i)} \leftarrow \boldsymbol{\theta}_{new}$.
        Update last accepted sample $\boldsymbol{\theta}_{prev} \leftarrow \boldsymbol{\theta}_{new}$.
    else
        Reject new sample $\boldsymbol{\theta}^{(i)} \leftarrow \boldsymbol{\theta}_{prev}$.
    end if
end for

distribution by applying the Jacobian of the log transformation denoted by $\mathbf{J}(\boldsymbol{\theta})$. Proposals are generated in log-space and can therefore be kept positive by applying the exp transformation, thus increasing our sampler's efficiency. The general form of the log-transformed priors is therefore

$$\pi_0(\log(\boldsymbol{\theta})|\boldsymbol{\phi})) = \pi_0(\exp(\boldsymbol{\theta})|\boldsymbol{\phi})) \det \left| \frac{\partial \exp(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|, \tag{2.23}$$

where $\boldsymbol{\phi}$ are the fixed prior hyper-parameters which are detailed in Appendix C.

Moreover, it is important to ensure that the transition kernel $T(\cdot, \cdot)$ has the target density $p(\boldsymbol{\theta}|\mathbf{q}, \mathbf{t})$ as its unique stationary distribution. We outline three proposal mechanisms used in our experiments to ensure efficient parameter posterior exploration and derive the corresponding acceptance ratios.

1. **Gaussian Random Walk (GRW)**

$$T_1(\boldsymbol{\theta}_{new}, \boldsymbol{\theta}_{prev}) = \mathcal{N}(\boldsymbol{\theta}_{prev}, \boldsymbol{\Sigma}_{prop}), \tag{2.24}$$

where $\boldsymbol{\Sigma}_{prop}$ is a diagonal matrix of step sizes for parameter space exploration. By symmetry of the kernel we get that the acceptance ratio is equal to

$$\alpha(\boldsymbol{\theta}_{new}, \boldsymbol{\theta}_{prev}) = \frac{\overbrace{\pi_0(\log(\boldsymbol{\theta}_{new}))|\mathbf{J}(\boldsymbol{\theta}_{new})|}^{\text{transformed prior}} \overbrace{p(\mathbf{q}|\boldsymbol{\theta}_{new})^{t_n}}^{\text{likelihood}}}{\pi_0(\log(\boldsymbol{\theta}_{prev}))|\mathbf{J}(\boldsymbol{\theta}_{prev})|p(\mathbf{q}|\boldsymbol{\theta}_{prev})^{t_n}}. \tag{2.25}$$

2. **Truncated Random Walk (TRW)**

$$T_2(\boldsymbol{\theta}_{new}, \boldsymbol{\theta}_{prev}) = \delta_{\mathcal{T}}(t_n)\tilde{\pi}(\boldsymbol{\theta}_{prev}) + \delta_{\mathcal{T}^c}(t_n)T_1(\boldsymbol{\theta}_{new}, \boldsymbol{\theta}_{prev}), \tag{2.26}$$

where $\delta_{\mathcal{T}}(t_n)$ is the delta function, $\tilde{\pi}(\boldsymbol{\theta}_{prev}) = \pi_0(\log(\boldsymbol{\theta}_{new}))|\mathbf{J}(\boldsymbol{\theta}_{new})|$ is the transformed prior density and $\mathcal{T} = \{t_n \in [0, 1] : t_n \leq t'\}$ is the set of all temperatures less than a threshold temperature t'. Since we cannot sample from the transformed prior directly we use a Monte Carlo approximation to sample from the untransformed prior and then apply the log transformation to our sample. This will lead to sample rejections because the log of a negative sample cannot be taken. However, by carefully tuning the temperature threshold t' we manage to limit the rejection rate.

Therefore, the acceptance ratio is equal to

$$\alpha(\boldsymbol{\theta}_{new}, \boldsymbol{\theta}_{prev}) = \frac{\overbrace{\tilde{\pi}(\boldsymbol{\theta}_{new})}^{\text{transformed prior}} \overbrace{p(\mathbf{q}|\boldsymbol{\theta}_{new})^{t_n}}^{\text{likelihood}} \overbrace{\left(\delta_{\mathcal{T}}(t_n)\tilde{\pi}(\boldsymbol{\theta}_{prev}) + \delta_{\mathcal{T}^c}(t_n)T_1(\boldsymbol{\theta}_{new}, \boldsymbol{\theta}_{prev})\right)}^{\text{proposal}}}{\tilde{\pi}(\boldsymbol{\theta}_{prev})p(\mathbf{q}|\boldsymbol{\theta}_{prev})^{t_n}\left(\delta_{\mathcal{T}}(t_n)\tilde{\pi}(\boldsymbol{\theta}_{new}) + \delta_{\mathcal{T}^c}(t_n)T_1(\boldsymbol{\theta}_{prev}, \boldsymbol{\theta}_{new})\right)}$$
$$= \begin{cases} \frac{p(\mathbf{q}|\boldsymbol{\theta}_{new})^{t_n}}{p(\mathbf{q}|\boldsymbol{\theta}_{prev})^{t_n}} & t_n \leq t'; \\ \frac{\tilde{\pi}(\boldsymbol{\theta}_{new})p(\mathbf{q}|\boldsymbol{\theta}_{new})^{t_n}}{\tilde{\pi}(\boldsymbol{\theta}_{prev})p(\mathbf{q}|\boldsymbol{\theta}_{prev})^{t_n}} & t_n > t'. \end{cases} \tag{2.27}$$

This proposal is beneficial in that for low temperatures sample acceptance tends to one and thus it is very efficient.

3. **Mixed Random Walk (MRW)**

$$T_3(\boldsymbol{\theta}_{new}, \boldsymbol{\theta}_{prev}) = (1 - p(\beta \leq t_n))\tilde{\pi}(\boldsymbol{\theta}_{prev}) + p(\beta \leq t_n)q(\boldsymbol{\theta}_{prev}|\boldsymbol{\theta}_{new}), \tag{2.28}$$

where $\beta \sim Beta(\alpha_0, \beta_0) \in [0, 1]$ and α_0, β_0 need to be tuned. The above proposal mechanisms yields

$$\alpha(\boldsymbol{\theta}_{new}, \boldsymbol{\theta}_{prev}) = \frac{\overbrace{\tilde{\pi}(\boldsymbol{\theta}_{new})}^{\text{transformed prior}} \overbrace{p(\mathbf{q}|\boldsymbol{\theta}_{new})^{t_n}}^{\text{likelihood}} \overbrace{\left(1 - p(\beta \leq t_n))\tilde{\pi}(\boldsymbol{\theta}_{prev}) + p(\beta \leq t_n)T_1(\boldsymbol{\theta}_{new}, \boldsymbol{\theta}_{prev})\right)}^{\text{proposal}}}{\tilde{\pi}(\boldsymbol{\theta}_{prev})p(\mathbf{q}|\boldsymbol{\theta}_{prev})^{t_n}\left(1 - p(\beta \leq t_n))\tilde{\pi}(\boldsymbol{\theta}_{new}) + p(\beta \leq t_n)T_1(\boldsymbol{\theta}_{prev}, \boldsymbol{\theta}_{new})\right)}. \tag{2.29}$$

The aim is to control the proportion of samples drawn from the transformed prior and the GRW proposal to achieve an "optimal" exploitation-exploration trade-off subject to priors being appropriately set. T_2, T_3 tend to be more efficient and are used to sample high dimensional parameter posteriors in the case of DeRomph's FD for example. The tuned proposal steps and parameters can be found on our GitHub repository.

2.1.3.1 Convergence diagnostic

The convergence of the MCMC chains is judged by inspection of the MCMC chains to ensure proper mixing and computation of the Gelman and Rubin convergence diagnostic (Gelman and Rubin, 1992) on ten parallel MCMC chains for every temperature. Assuming C parallel MCMC

chains with N samples each and T temperatures we compute for all $c = 1, \ldots, C, t = 1, \ldots, T$ the chain mean after D burnin samples is

$$\bar{\boldsymbol{\theta}}_t^{(c)} = \frac{1}{N-D} \sum_{i=1}^{N-D} \boldsymbol{\theta}_{i,t}^{(c)}, \tag{2.30}$$

and the total mean is

$$\bar{\boldsymbol{\theta}}_t = \frac{1}{C} \sum_{i=1}^{C} \bar{\boldsymbol{\theta}}_t^{(c)}. \tag{2.31}$$

Next we compute the between-chain variance

$$B_t = \frac{N-D}{C-1} \sum_{c=1}^{C} (\bar{\boldsymbol{\theta}}_t^{(c)} - \bar{\boldsymbol{\theta}}_t)^T (\bar{\boldsymbol{\theta}}_t^{(c)} - \bar{\boldsymbol{\theta}}_t) \tag{2.32}$$

and the within-chain variance

$$\begin{aligned} s_{c,t}^2 &= \frac{1}{N-D-1} \sum_{i=1}^{N-D} (\bar{\boldsymbol{\theta}}_{i,t}^{(c)} - \bar{\boldsymbol{\theta}}_t^{(c)})^T (\bar{\boldsymbol{\theta}}_{i,t}^{(c)} - \bar{\boldsymbol{\theta}}_t^{(c)}), \\ W_t &= \frac{1}{C} \sum_{c=1}^{C} s_{c,t}^2. \end{aligned} \tag{2.33}$$

Finally, we get the Gelman and Rubin criterion

$$R_t = \frac{\frac{C-1}{C} W_t + \frac{1}{C} B_t}{W_t} \quad \forall\, t = 1, \ldots, T. \tag{2.34}$$

A convergent chain would exhibit behaviour that results in $B_t \downarrow 0$ and $C \to \infty$, meaning $R_t \to 1$. Therefore, we claim that our sampler has converged if all temperature chains achieved $R_t \approx 1.1$. In practise, C need not be very large and we therefore fix it to $L = 10$.

2.1.4 Sensitivity analysis

Despite the fact that MLs are a robust way of assessing model quality in light of data they are prone to prior specification (Friel and Pettitt, 2008). The authors showed that in the conjugate case of a Gaussian prior and likelihood we get that

$$\mathbb{E}_{\boldsymbol{\theta}|\mathbf{q}, t_n=0} \left[\log \left(p(\mathbf{q}|\boldsymbol{\theta}) \right) \right] \to \infty \tag{2.35}$$

at the same speed as the prior's variance. This effect is larger in low temperatures where the power posterior is prior-dominated. Our choice of temperature schedule yields a significant number of power posteriors where this effect is present. Our dense discretisation in low temperatures mitigates this effect as shown in (2.21) due to the Δt_n term. However, we still need to account for the prior diffusivity's effect on the high temperature power posteriors. Therefore, we vary our priors' variance and get ML estimates for a diffuse (high variance), a "regular" (medium variance) and an informative (low variance) prior as detailed in Appendix C.

2.2 Black-box-driven ABM calibration

2.2.1 Gaussian process emulation

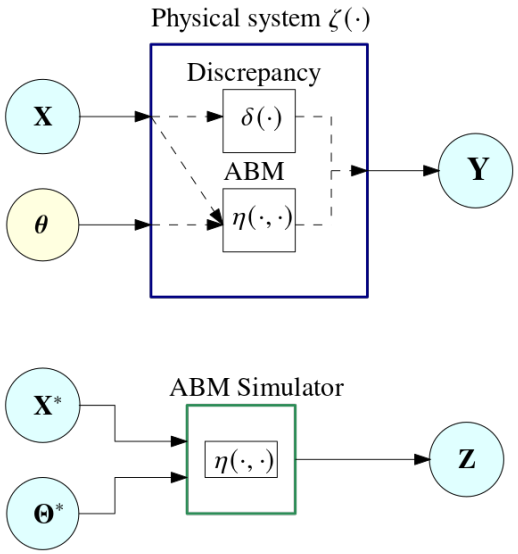


Figure 2.3: Plate diagram of ABM emulation in line with the approach of (Kennedy and O'hagan, 2001). Known and inferred quantities are coloured in light blue and yellow, respectively.

The seminal work of (Kennedy and O'hagan, 2001) introduced a framework for learning a cheap surrogate/emulator of the computer model (in our case ABM) which we depict in Figure 2.3. Let $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_n]^T \in \mathbb{R}^{n \times n_y}$ be observational data about a system under study

e.g. disease epidemics. Under a generative viewpoint, observations are noisy realisations of the true system dynamics $\zeta(\mathbf{X})$ where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]^T \in \mathbb{R}^{n \times n_x}$ is an input to the system. We thus have $\forall\, i = 1, \ldots, n$

$$\mathbf{y}_i = \zeta(\mathbf{x}_i) + \boldsymbol{\epsilon}, \tag{2.36}$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^{n \times n_y}$ is an observation error model. In the SEIRD example of Chapter 1, $\zeta(\mathbf{X})$ may capture number of infected individuals as a function of say space-time input $\mathbf{X}$. The unknown true dynamics ζ are replaced by a misspecified stochastic ABM simulator $\eta(\mathbf{X}, \boldsymbol{\theta}) \in \mathbb{R}^{n_\eta}$ where $\boldsymbol{\theta}$ are *a priori* unknown ABM parameters (e.g. `p_immune` in the SEIRD example). To correct for ABM misspecification a stochastic model bias/discrepancy term $\delta(\mathbf{X})$ is introduced $\forall\, i = 1, \ldots, n$:

$$\mathbf{y}_i = \underbrace{\eta(\mathbf{X}_i, \hat{\boldsymbol{\theta}}) + \delta(\mathbf{X}_i)}_{\zeta(\mathbf{X}_i)} + \boldsymbol{\epsilon}_i, \tag{2.37}$$

where $\hat{\boldsymbol{\theta}} \in \mathbb{R}^{n_\theta}$ is the (unknown) most plausible parameter configuration of the ABM.

We assume that $\eta(\cdot, \cdot)$ is computationally expensive and only N evaluations $\mathbf{Z} \in \mathbb{R}^{N \times n_\eta}$ of $\eta(\cdot, \cdot)$ can be obtained from an experimental design $\mathcal{D} = \{(\boldsymbol{\theta}_i^*, \mathbf{X}_i^*) : i = 1, \cdots, N\}$, where $\mathbf{X}^*$ need not be equal to input data $\mathbf{X}$. A Gaussian process prior is placed on η, i.e.

$$\eta(\mathbf{X}^*, \boldsymbol{\theta}^*) \sim \mathcal{GP}\left(\boldsymbol{\mu}_\eta(\mathbf{X}^*, \boldsymbol{\theta}^*), \boldsymbol{\Sigma}_\eta\left((\mathbf{X}^*, \boldsymbol{\theta}^*), (\mathbf{X}^{**}, \boldsymbol{\theta}^{**})\right)\right),$$

with mean and covariance functions $\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\eta$, respectively. Each covariance kernel has additional hyper-parameters $\boldsymbol{\psi}_\eta$ that need to be learned. Similarly, a GP prior is placed on the stochastic discrepancy term $\delta(\mathbf{X})$:

$$\delta(\mathbf{X}, \mathbf{X}') \sim \mathcal{GP}\left(\boldsymbol{\mu}_\delta(\mathbf{X}), \boldsymbol{\Sigma}_\delta(\mathbf{X}, \mathbf{X}')\right),$$

with hyper-parameters $\boldsymbol{\psi}_\delta$. We denote input locations $\mathbf{X}, \mathbf{X}^*, \boldsymbol{\theta}^*$ by $\boldsymbol{\Xi}$ and kernel parameters $\boldsymbol{\psi} = [\boldsymbol{\psi}_\eta, \boldsymbol{\psi}_\delta]^T$. The marginal likelihood of the observed and simulated data is

$$p(\mathbf{Y}, \mathbf{Z}|\boldsymbol{\Xi}, \boldsymbol{\psi}) = \int p(\mathbf{Y}|\boldsymbol{H}_{\boldsymbol{\theta}}, \boldsymbol{\Delta}) p(\mathbf{Z}|\boldsymbol{H}^*) p(\boldsymbol{\Delta}|\mathbf{X}, \boldsymbol{\psi}) p(\boldsymbol{H}_\theta, \boldsymbol{H}^*|\boldsymbol{\theta}, \boldsymbol{\Xi}, \boldsymbol{\psi}) p(\boldsymbol{\theta}) d\boldsymbol{H}^* \, \mathrm{d}\boldsymbol{\Delta} \, \mathrm{d}\boldsymbol{H}_\theta \, \mathrm{d}\boldsymbol{\theta}, \tag{2.38}$$

where $\boldsymbol{H}_\theta = [\eta(\mathbf{X}_i, \boldsymbol{\theta}), \ldots, \eta(\mathbf{X}_n, \boldsymbol{\theta})]^T$, $\boldsymbol{\Delta} = [\delta(\mathbf{X}_i), \cdots, \delta(\mathbf{X}_n)]^T$, and $\boldsymbol{H}^* = [\eta(\mathbf{X}_1^*, \boldsymbol{\Theta}_1^*), \ldots, \eta(\mathbf{X}_N^*, \boldsymbol{\Theta}_N^*)]^T$. The simulator's stochasticity is assumed to be embedded in η and therefore a noise model links

$\mathbf{Z}$ to $\boldsymbol{H}^*$ in (2.38). The integral in (2.38) is intractable and so we will resort to approximations such as random feature expansions (Marmin and Filippone, 2018).

Further development of this framework will be undertaken to account for sparse discrete spatio-temporal and multi-resolution $\mathbf{Y}$,$\mathbf{Z}$ (e.g. traffic volumes on network links and vehicle trajectories) while refining the model discrepancy $\zeta(\cdot)$. Parameter identifiability will also be addressed since MATSim has large n_θ.

2.2.2 Approximate Bayesian Computation

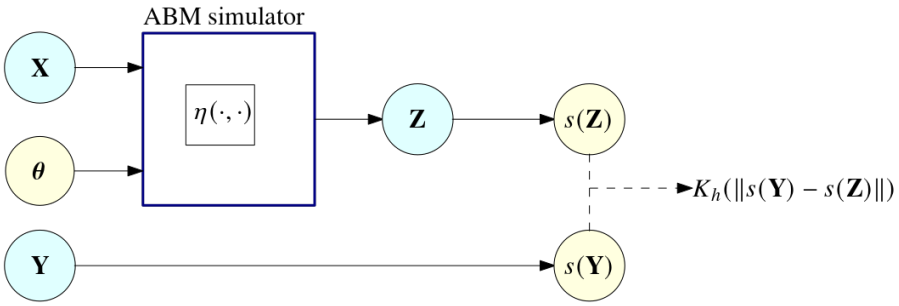


Figure 2.4: Plate diagram of the ABC framework. Known and inferred quantities are coloured in light blue and yellow, respectively.

An approach that circumvents likelihood intractability issues is Approximate Bayesian Computation (Beaumont, W. Zhang, and Balding, 2002) exhibited in Figure 2.4. Likelihood evaluations are replaced by ABM simulations $\mathbf{Z}$ which are obtained by running $\eta(\mathbf{X}, \boldsymbol{\theta})$ for N parameters $\tilde{\boldsymbol{\theta}} = [\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(N)}]^T$ sampled from a prior $\pi_0(\boldsymbol{\theta})$. We implicitly assume that the data and ABM output model spaces are the same even though the latter is typically larger [1]. Summary statistics $s : \mathcal{X} \to \mathcal{D}$ are constructed to reduce data and output model spaces $\mathcal{X} \in \mathbb{R}^{n_\eta}$ to a lower-dimensional space $\mathcal{D}$ while minimising the information lost in the reduction. The likelihood is replaced by an indicator function that ensures summary statistics are in the vicinity of the observed data:

$$p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{X}) \approx \int \mathbb{1}\left(\rho\left(s(\eta(\mathbf{X}, \boldsymbol{\theta})), s(\mathbf{Y})\right) \leq \epsilon_0\right) p(\mathbf{X}|\boldsymbol{\theta})\, \mathrm{d}\mathbf{X}, \tag{2.39}$$

[1] This does not affect the validity of the arguments that follow.

where ϵ_0 is a tolerance value controlling ABC sample acceptances, ρ defines a discrepancy measure and samples from $p(\eta(\mathbf{X}, \boldsymbol{\theta})|\boldsymbol{\theta})$ are obtained via ABM simulation. A rejection sampling scheme from such a hard-threshold likelihood is very inefficient. Smooth generalisations of the likelihood include

$$p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{X}) \approx \int K_h(\|s(\eta(\mathbf{X}, \boldsymbol{\theta})) - s(\mathbf{Y})\|)p(\mathbf{X}|\boldsymbol{\theta})\,\mathrm{d}\mathbf{X}, \tag{2.40}$$

where K_h are suitable chosen kernels of bandwidth h. However, even likelihood approximations such as the one above rely on summary statistic *sufficiency* which ensures $\pi(\boldsymbol{\theta}|\mathbf{Y}) = \pi(\boldsymbol{\theta}|s(\mathbf{Y}))$. Moreover, efficient sampling schemes have been also proposed (Cranmer, Brehmer, and Louppe, 2019). However, the choice of a sampling scheme is of inferior importance if either the kernel or summary statistics are not robustly specified to ABM and/or error model misspefication.

Following (Schmon, P. W. Cannon, and Knoblauch, 2020) and (Dyer, P. Cannon, and Schmon, 2021), we will examine different choices of summary statistics (e.g. path signatures) and kernels (e.g. Wasserstein difference) to robustify ABM inference against misspecification and the non-stationary nature of observational data (e.g. auto-correlated traffic volume time series).

Chapter 3

Experimental results

3.1 Experimental design

In this Chapter we critically evaluate each fundamental diagram in Table 1.1 with regards to their achieved trade-off between model complexity and data fit. We first obtain unbiased estimates of the log marginal likelihood using the Metropolis-Hastings algorithm to sample from each power posterior in (2.21). The MCMC chains are run between 300,000 and 900,000 iterations with burnin between 30,000 and 100,000 depending on model complexity for all parameters and temperatures. The Monte Carlo error in the log ML estimator in (2.21) is quantified by running ten MCMC chains of length 10,000 in parallel for each temperature and computing the mean and variance of the log ML estimators. We also compute the R^2, posterior predictive mean and standard deviation and corresponding residual plot for the full posterior ($t = 1$) based on a "regular" (medium variance) prior specification. We first examine structural differences between constitutive laws using synthetic data simulated from each FD and later apply our method to real-world data from the UK's M25 motorway, which facilitates model validation against loop detector data. Our work is complemented by a sensitivity analysis of the obtained ML estimates in both synthetic and real-world data applications.

3.2 Synthetic data

3.2.1 Model comparison

The log marginal likelihood estimates and their associated Monte Carlo errors introduced in the previous Chapter are summarised in Table 3.1. Row-wise maxima are highlighted in **bold** to indicate the best model for each dataset. Note that the table has been broken into two subtables for better visibility.

Table 3.1 is partially diagonally-dominated and off-diagonally dominated. This is a clear indication that the proposed FD forms are not entirely dissimilar since if that was the case then each FD would be the only model best able to reproduce data generated from itself. Instead, DelCastillo's and DeRomph's models are able to reproduce or sufficiently emulate traffic features from models such as Daganzo's, Newell's, Greenshield's, and Smulder's. We proceed by examining pairs of similar in structure models.

Greenshield's and Greenberg's models are the two most simplistic models with two parameters that control vertical scaling of the FD and the location of the second FD root (traffic jam density). The former model assumes a symmetric flow-density relationship around the maximum (road capacity) while the latter assumes a positively skewed[1] relationship. Both models achieve significant log MLs when applied to their respective simulated datasets. Greenshield's model yields another plausible fit for the only simulated dataset resembling a concave quadratic FD on the entire density domain; Newell's dataset (see Appendix B). However, DelCastillo's model constitutes an arguably better alternative for Greenberg's, Greenshield's and Newell's data since DelCastillo's achieved data fits compensate for its two extra parameters. DelCastillo's and Daganzo's data are smoothed near-linear and piecewise linear respectively and therefore a quadratic model such as Greenshield's can only partially fit the data, scoring lower in the log ML scale. Despite its low dimensional parameter space, Greenberg's inability to adjust its skewness accounts for the majority of its poor data fits and renders it an

[1] We borrow this term from statistics to describe FDs that have denser left than right tails.

Simulation model	*Inference model*				
	Greenshield's	Greenberg's	Underwood's	Northwestern's	Newell's
Greenshield's	176.09 ± 0.15	-84.91 ± 0.19	39.83 ± 0.04	138.08 ± 0.22	146.49 ± 0.06
Greenberg's	-47.03 ± 0.07	**177.22 ± 0.02**	72.78 ± 0.01	-0.64 ± 0.07	82.17 ± 0.12
Underwood's	-357.67 ± 0.16	-183.73 ± 0.23	**172.89 ± 0.36**	-87.88 ± 0.11	21.27 ± 0.17
Northwestern's	-380.13 ± 0.33	-210.5 ± 0.28	-133.89 ± 0.01	**174.48 ± 0.05**	-139.91 ± 0.08
Newell's	141.88 ± 0.07	-33.61 ± 0.02	-0.04 ± 0.04	148.87 ± 0.14	**176.9 ± 0.07**
Wang's	-197.9 ± 0.29	-175.2 ± 0.24	-178.46 ± 0.02	-32.16 ± 0.06	-78.54 ± 0.06
Daganzo's	46.54 ± 0.04	-136.58 ± 0.25	-78.05 ± 0.05	24.69 ± 0.08	113.82 ± 0.13
DelCastillo's	3.39 ± 0.04	11.79 ± 0.05	-2.65 ± 0.03	159.08 ± 0.15	157.87 ± 0.2
Smulder's	-93.58 ± 0.1	2.4 ± 0.14	9.27 ± 0.1	82.75 ± 0.17	91.09 ± 0.15
DeRomph's	-319.38 ± 0.06	-268.34 ± 0.14	-134.92 ± 0.14	45.69 ± 0.24	-202.71 ± 0.2

Simulation model	*Inference model*				
	Wang's	Daganzo's	DelCastillo's	Smulder's	DeRomph's
Greenshield's	142.65 ± 0.13	80.7 ± 0.04	**178.14 ± 0.04**	173.11 ± 0.03	176.3 ± 0.01
Greenberg's	69.23 ± 0.04	14.13 ± 0.04	173.65 ± 0.04	-50.27 ± 0.01	141.21 ± 0.02
Underwood's	126.27 ± 0.01	45.39 ± 0.03	49.48 ± 0.04	50.08 ± 0.01	155.34 ± 0.03
Northwestern's	104.54 ± 0.28	-117.95 ± 0.08	-116.79 ± 0.04	-76.1 ± 0.06	162.55 ± 0.11
Newell's	160.01 ± 0.10	109.08 ± 0.12	**176.9 ± 0.05**	152.86 ± 0.03	160.92 ± 0.01
Wang's	**175.52 ± 0.14**	-36.87 ± 0.13	-35.68 ± 0.06	32.88 ± 0.06	142.11 ± 0.01
Daganzo's	149.83 ± 0.21	176.25 ± 0.05	**177.3 ± 0.04**	115.03 ± 0.03	172.93 ± 0.12
DelCastillo's	152.48 ± 0.14	137.9 ± 0.17	**176.94 ± 0.03**	159.91 ± 0.05	158.01 ± 0.05
Smulder's	120.95 ± 0.07	138.18 ± 0.08	146.45 ± 0.02	175.57 ± 0.1	**177.04 ± 0.21**
DeRomph's	77.34 ± 0.04	-190.22 ± 0.23	-186.72 ± 0.01	-76.79 ± 0.04	**177.11 ± 0.02**

Table 3.1: Unbiased estimates of the log marginal likelihood estimator in (2.21) and their associated Monte Carlo error for every fundamental diagram model and simulated dataset. Each row and column correspond to the FD model used to simulate flow data and the FD model fitted to that data, respectively.

		Inference model				
		Greenshield's	Greenberg's	Underwood's	Northwestern's	Newell's
Simulation model	Greenshield's	**0.98**	0.69	0.88	0.97	0.97
	Greenberg's	0.82	**0.96**	0.89	0.88	0.91
	Underwood's	0.04	0.53	**0.98**	0.85	0.9
	Northwestern's	-0.2	0.29	0.77	**0.99**	0.71
	Newell's	0.97	0.62	0.79	0.97	**0.98**
	Wang's	0.59	0.62	0.07	0.9	0.82
	Daganzo's	0.92	0.18	0.6	0.9	0.97
	DelCastillo's	0.89	0.8	0.76	**0.98**	0.97
	Smulder's	0.76	0.83	0.81	0.95	0.95
	DeRomph's	-0.96	-0.25	0.86	0.98	0.5

		Inference model				
		Wang's	Daganzo's	DelCastillo's	Smulder's	DeRomph's
Simulation model	Greenshield's	0.97	0.95	**0.98**	**0.98**	**0.98**
	Greenberg's	0.9	0.84	**0.96**	0.33	0.93
	Underwood's	**0.98**	0.93	0.93	0.94	**0.98**
	Northwestern's	**0.99**	0.79	0.79	0.92	**0.99**
	Newell's	**0.98**	0.96	**0.98**	0.97	0.97
	Wang's	**0.99**	0.9	0.9	0.96	**0.99**
	Daganzo's	0.98	**0.99**	**0.99**	0.97	0.98
	DelCastillo's	0.97	0.97	**0.98**	**0.98**	0.97
	Smulder's	0.96	0.97	0.97	**0.98**	**0.98**
	DeRomph's	0.99	0.6	0.59	0.94	**1**

Table 3.2: Fitted R^2 values for every fundamental diagram model and simulated dataset. Each row and column correspond to the FD model used to simulate flow data and the FD model fitted to that data, respectively.

oversimplified model hypothesis.

Another pair of similar in structure models is that of Underwood's and Northwestern's models. Both FDs are two-parameter smooth partially concave models that control vertical scaling and location of the maximum (road capacity). The latter has a faster growth and decay rates in the free-flow and congested regions of the FD (before and after the location of the maximum), respectively. Excluding the tails of the FD, Northwestern's model shape can resemble a symmetric around the maximum model more than Underwood's does. Therefore, it fits DelCastillo's, Greenshield's, and Newell's data sufficiently well. Note that inference is made in log space and therefore the existence of a second root in an FD model is irrelevant as long as the right tail of the FD tends to zero at an appropriate rate. However, Northwestern's model has a faster growth rate but slower decay rate than a quadratic model and so the fit to Greenshield's simulated data is slightly poorer. DelCastillo's data are similar to that of a positively skewed quadratic FD that has near-linear tails (see Appendix B) and therefore Northwestern's model is able to yield a better model hypothesis for this dataset. Daganzo's data is a linear approximation of DelCastillo's data and as a result a smooth approximation by Northwestern's model produces lower but still significantly high log ML (159.08 ± 0.15). The subtle difference between Northwestern's and Underwood's decay rate enables the former to fit Smulder's and DeRomph's data sufficiently well despite the existence of a jump discontinuity in the two datasets. The jump discontinuity (capacity drop) can be partially emulated by Northwestern's model as reflected on the log MLs (82.75 ± 0.17,45.69 ± 0.24). DeRomph's jump in the data is considerably larger than Smulder's which explains the discrepancy in the achieved log MLs. We note that the only other model with a non-negligible model evidence for Smulder's and DeRomph's data is Wang's FD because it shares characteristics with Northwestern's which are elaborated on in a later paragraph. Underwood's model is able to substantially outperform Northwestern's only when fitted to Greenberg's and Underwood's data. The positive skewness (asymmetry) of Greenberg's data can be better learned using Underwood's model than its counterpart. Despite the fact that Northwestern's and Underwood's are similar in functional form they do not explain each other's simulated data well as evidenced by the very low log MLs (-87.88 ± 0.11 and -133.89 ± 0.01). This is because attempting to generate a tight fit of one model

over the other model's simulated dataset breaks the i.i.d noise assumption (the log model is close to the data is some regions but quite far away in other regions). As a result, a higher noise level model seems more plausible at first. However, this parameter setting does not contribute significantly to the marginal MC likelihood sum in (2.21). The most dominant terms in the sum come from parameter configurations that achieve an "averaging" effect where on average the fitted model's error is the same across the density domain. Despite comprising the most dominant terms in (2.21), these types of configurations still generate relatively poor data fits that result in low log MLs. Finally, both Underwood's and Northwestern's models insufficiently explain Wang's data because in the free-flow region Wang's data grows near linearly while in the congested region the data decay faster than either Northwestern's or Underwood's model does.

In Chapter 1 we highlighted the similarity between DelCastillo's and Daganzo's models. The former is smoother, more flexible than the latter at the cost of one extra parameter while the latter is triangular in shape and its parameters control vertical scaling, critical and traffic jam densities. Therefore, DelCastillo's FD systematically provides better data fits than Daganzo's. At the tails of the FD, DelCastillo's model ranges from near-linear to quadratic in shape. Daganzo's is a coarse but not entirely poor approximation of data shaped like a symmetric concave quadratic (e.g. Greenshield's, Newell's). Nevertheless, DelCastillo's variable smoothness makes it a better candidate model. DelCastillo's skewness parameter u also makes it a compelling model alternative for positively skewed data such as Greenberg's. The added model complexity from the extra parameter appears to be compensated for with better data fits from DelCastillo's FD. The only shortcoming in DelCastillo's model is the inability to adjust the slope in the congested part of the FD without tampering with the location of the second root or the vertical scaling of the FD. Although Daganzo's model can correct for different slopes in the congested part, its piecewise linear nature is a serious drawback when fitting to data generated from smooth models. Both Daganzo's and DelCastillo's models explain Smulder's data sufficiently well due to the data's non-convex shape on the entire density domain. The existence of capacity drop in Smulder's data is ultimately penalised by the likelihood which reduces the two models' evidence for Smulder's data. Model evidence for the rest

of the simulated datasets is ranked by the degree to which a near-linear model can approximate the data in the congested parts. The only near-linear approximation that leads to a reasonable log ML is that of Underwood's data where the negative exponential decay is roughly modelled by a near-linear model. Since DelCastillo's model is more flexible in the free-flow region of the FD, model fit displays the crude "averaging" effect described earlier. Hence certain regions of the data are overestimated while others are underestimated leading to an overall poor fit. The lowest log MLs are achieved in the cases of Northwestern's and DeRomph's data due to poor model fit coupled with added model complexity and existence of a large jump discontinuity, respectively.

In addition, Newell's model is another flexible three-parameter FD model. Contrary to the aforementioned models, Newell's FD has a parameter for the traffic jam density and two other parameters which jointly determine the vertical scaling, critical density, and skewness of the graph. The model's variable skewness facilitates fitting of near-symmetric around the maximum data such as Greenshield's, DelCastillo's and Daganzo's. Despite its flexibility, the model's vertical scaling cannot be tuned without affecting skewness. As a result, Newell's model can fit Greenberg's data well but within a multiplicative factor. This leads to a lower signal-to-noise ratio and subsequently lower log ML compared to the aforementioned datasets. The concavity assumption on the entire density domain that is encoded in Newell's model implies that the model performs poorly on data that are generated from only partially concave models. An exception is Smulder's data that have a linear congested region and which Newell's model provides sufficient evidence for (91.09 ± 0.15). The presence of a small jump discontinuity in log-scale has a small impact on model fit even though the region above the data's free-flow region's maximum is not captured at all. However, this model hypothesis is deemed plausible due to the multiplicative noise structure of the models (or additive in log-space). When approximating Underwood's, Wang's, Northwestern's and DeRomph's data Newell's model exhibits the aforementioned "averaging" effect and scores poorly on the log ML scale. This effect is apparent at various degrees depending on the decay rate of the data in the congested region and the presence of a large capacity drop.

Regarding Wang's model, it is a partially concave model with variable skewness and a free-flow region which is upper bounded by $v_f k$. These features are sufficient for Wang's to be a competing model to Newell's when modelling simulated data from concave models such as Greenshield's, Newell's, Greenberg's, DelCastillo's datasets. The differences in model evidence between Newell's and Wang's models are substantial but not conclusive as the ML is known to be susceptible to prior diffusion. We discuss more details about this effect in Section 3.2.2. The case for Daganzo's data is notable since Wang's model is able to fit the data better than Newell's for a larger set of parameter configurations thus contributing more to the MC sum in (2.21). This is mainly attributed to parameter s which controls the model's decay rate in the congested region and impacts the maximum's location. For the same reason, Wang's is able to mimic the behaviour of both Underwood's and Northwestern's. Wang's model evidence for Underwood's data is substantially higher than the evidence for Northwestern's data because Wang's is structurally closer to Underwood's when $k_{crit} \approx 0$. Also, the challenge of explaining data with jump discontinuities is to an extent overcome due to the model's ability to control the decay in the congested region. However, the larger capacity drop (jump discontinuity) coupled with a non-linear congested region admits less plausible parameter configurations and renders Wang's model less suitable for DeRomph's data than for Smulder's.

Finally, DeRomph's and Smulder's models are the only two capacity drop FDs and are of piecewise nature. DeRomph's model is a more general version of Smulder's which can be seen by setting $\alpha = k_{jam}$ and $\beta = 1$. This observation is validated by the log ML of the former model on the latter's data (177.04±0.21). It is easy to verify that Smulder's cannot fit DeRomph's data due to the minor structural discrepancy between the two models which causes Smulder's to fail to explain both free-flow and congested regions of DeRomph's data. Excluding DelCastillo's data, DeRomph's model significantly outperforms Smulder's despite having two additional parameters. We note that in the case of DelCastillo's data, DeRomph's and Smulder's FDs are arguably competitive models. A reason for Smulder's marginally higher log ML might be the effect of the prior diffusivity on the marginal likelihood, which we examine at a later Section. The concave near-symmetric datasets of Greenshield's, Newell's and Daganzo's are adequately modelled by both Smulder's and DeRomph's. The discrepancy in model

evidences for Greenshield's data may also be attributed to prior diffusivity differences. In contrast, DeRomph's model is likely a better candidate model for Newell's and Daganzo's data because of the additional flexibility the model gains from parameters α and β. Both parameters allow DeRomph's to fit the marginally asymmetric Newell's data better than the piecewise quadratic and linear Smulder's model. Parameter α grants DeRomph's model the flexibility to fit Daganzo's piecewise linear data by horizontally stretching the quadratic component of the model and obtain a good local linear approximation. Nevertheless, DelCastillo's model provides a much simpler and plausible explanation of Greenshield's, Newell's and Daganzo's data and is ultimately the superior model in these three cases. Moreover, Greenberg's distinctive asymmetry is challenging to model with a linear congested region. That is why DeRomph's has the advantage of providing a smooth approximation of the congested region and on average a larger number of plausible parameters than Smulder's. However, DeRomph's skeweness can only be partially adjusted by jointly configuring $v_f, \rho_c, \gamma, \alpha, \beta$. Inadvertently, the vertical scaling of the model is changed and Greenberg's data can only be tracked within a multiplicative factor. This renders DelCastillo's a much more suitable model for Greenberg's data. Data that are generated from partially concave and convex models such as Northwestern's, Underwood's and Wang's are intuitively better explained by DeRomph's model whose parameter β enables it to become partially convex and closely fit these three datasets. DeRomph's model evidence is reduced when modelling very positively skewed data, such as Underwood's data. The larger the degree of positive skewness the smaller the region that DeRomph's quadratic component can explain and so the poorer the approximation of the data's concave region from DeRomph's convex component. Naturally, Smulder's linear component constitutes a coarse approximation of convex regions of data and the quality of approximation is inversely proportional to the data decay rate in the congested part of the graph.

Regarding Table 3.2, it is apparent that it cannot be used as a basis for selecting the appropriate constitutive law for the data. As highlighted in previous Chapters, the coefficient of determination R^2 is susceptible to favouring complex model hypotheses such as DeRomph's. Indeed, DeRomph's attains the highest R^2 scores across all synthetic datasets despite its six-dimensional parameter space. Although the row-wise maxima of Table 3.2 track those of Table

3.1, there is significant confusion about the most suitable model in the cases of models such as Wang's, Smulder's and DeRomph's. Specifically, Wang's appears to provide competitive fits for Underwood's, Northwestern's and Newell's data whereas Table 3.1 suggests otherwise. For these reasons, the coefficient of determination and similar error-type metrics can be solely used to assess model quality. A more robust way to assess the validity of the log ML estimates would be to examine their sensitivity to prior specification, which we perform in the next Section.

3.2.2 Sensitivity analysis

Although the model evidence estimates allow us to discern the features of most parametric constitutive laws of traffic, the effect of the prior's diffusivity on the log marginal likelihood can potentially obscure our judgement about the most plausible model for a given dataset. Table 3.3 depicts unbiased estimates of the log marginal likelihood for three prior variance specifications: diffuse (high variance), informative (low variance) and "regular" (medium variance) as listed in Appendix C. Row-wise maxima for each prior specification are highlighted in **bold**. We limit our study to the most competitive models identified in the previous Section; Wang's, DelCastillo's, Smulder's and DeRomph's FDs.

The resulting log ML ranges suggest that Greenshield's data can be replicated equally well by Smulder's and DeRomph's model in addition to DelCastillo's FD. This can be verified by identifying Greenshield's model embedded in a similar form in both Smulder's and DeRomph's FDs. Similarly, Daganzo's data admits both DelCastillo's and DeRomph's model as plausible model hypotheses. Indeed, for large enough α and $\beta \approx 1$ DeRomph's model can effectively approximate Daganzo's data. DelCastillo's FD is a competitive model for Greenberg's and Newell's data while DeRomph's FD captures Smulder's data equally as well as Smulder's model does. However, there are cases where the prior's sensitivity does not have an impact on the most plausible model. For example, Wang's, Underwood's and Northwestern's, DelCastillo's, and DeRomph's data cannot be adequately represented by any other model other than their own.

In light of the Tables 3.1,3.3 there is evidence supporting that the model universe of parametric FDs can be maximally reduced to Wang's, DelCastillo's and DeRomph's constitutive laws. However, DeRomph's is a six parameter model describing a two-dimensional phase space and therefore it can be argued that despite its significantly better data fits it will likely not generalise well to real-world noisy datasets. However, such conclusion needs to validated against real-world data. We devote the next Section to an examination of the ten FDs in the context of data from the UK's M25 motorway.

	Inference model			
	Wang's	Delcastillo's	Smulder's	DeRomph's
Greenshield's	141.49 ± 0.14,	**173.95 ± 0.16**,	173.4 ± 0.08,	172.83 ± 0.09
	142.54 ± 0.05,	**176.58 ± 0.05**,	174.33 ± 0.2,	176.19 ± 0.13,
	144.85 ± 0.08	**179.43 ± 0.01**	178.5 ± 0.03	178.29 ± 0.16
Greenberg's	67.54 ± 0.14,	**169.73 ± 0.18**,	-31.09 ± 0.19,	141.19 ± 0.09,
	68.72 ± 0.03,	**172.71 ± 0.07**,	-46.41 ± 0.21,	143.33 ± 0.1,
	69.93 ± 0.03	**175.13 ± 0.01**	-50.53 ± 0.19	146.4 ± 0.01
Underwood's	120.26 ± 0.09,	47.53 ± 0.21,	45.14 ± 0.21,	**154.04 ± 1.27**,
	122.74 ± 0.09,	49.49 ± 0.04,	47.01 ± 0.06,	**157.24 ± 1.08**,
	126.41 ± 0.01	55.79 ± 0.0	50.13 ± 0.03	**162.37 ± 0.25**
Northwestern's	104.07 ± 0.16,	-121.43 ± 0.25,	-78.8 ± 0.3,	**159.39 ± 0.07**
	105.85 ± 0.13,	-118.61 ± 0.22,	-77.08 ± 0.11,	**161.66 ± 0.39**,
	109.78 ± 0.01	-113.46 ± 0.0	-73.58 ± 0.01	**162.81 ± 0.15**,
Newell's	153.06 ± 0.2,	**172.03 ± 0.09**,	151.74 ± 0.22,	156.33 ± 0.52,
	155.97 ± 0.07,	**175.0 ± 0.06**,	155.41 ± 0.02,	159.85 ± 0.48,
	160.48 ± 0.03	**180.4 ± 0.01**	157.92 ± 0.03	162.98 ± 0.72
Wang's	**170.55 ± 0.37**,	-39.73 ± 0.23,	28.21 ± 0.34,	137.4 ± 0.32,
	175.06 ± 0.1,	-36.91 ± 0.13,	29.7 ± 0.13,	139.39 ± 0.34,
	178.04 ± 0.03	-31.65 ± 0.0	33.25 ± 0.09	142.17 ± 0.18

Table 3.3 – continued in next page

Table 3.3 – continued from previous page

Daganzo's	147.89 ± 0.22,	**171.96 ± 0.21**,	106.01 ± 0.18,	**172.1 ± 0.25**,
	151.06 ± 0.08,	**176.22 ± 0.07**,	108.81 ± 0.17,	174.34 ± 0.14,
	156.97 ± 0.0	**182.05 ± 0.0**	115.91 ± 0.04	175.22 ± 0.04
Delcastillo's	151.93 ± 0.09,	**172.22 ± 0.22**,	158.08 ± 0.18,	152.21 ± 0.5,
	154.04 ± 0.11,	**174.84 ± 0.15**,	160.41 ± 0.17,	155.85 ± 0.11,
	159.21 ± 0.01	**178.46 ± 0.04**	163.07 ± 0.15	159.28 ± 0.0
Smulder's	117.58 ± 0.07,	139.81 ± 0.51,	170.51 ± 0.05,	**170.18 ± 0.32**,
	119.59 ± 0.12,	143.04 ± 0.09,	172.64 ± 0.39,	**173.41 ± 0.23**,
	122.63 ± 0.02	149.65 ± 0.0	176.07 ± 0.05	**177.13 ± 0.09**
DeRomph's	76.5 ± 0.26,	-191.64 ± 0.09,	-77.54 ± 0.14,	**170.92 ± 0.33**,
	78.17 ± 0.14,	-189.46 ± 0.02,	-75.88 ± 0.21,	**172.6 ± 0.13**,
	81.04 ± 0.05	-185.08 ± 0.0	-71.54 ± 0.06	**181.33 ± 0.08**

Table 3.3: Sensitivity analysis of the unbiased estimates of the log marginal likelihood estimator in (2.21) and their associated Monte Carlo error for every fundamental diagram model and simulated dataset. Each cell's values correspond to a diffuse, "regular" and informative prior specification of the same model and data.

3.3 M25 motorway data

In this section we examine loop detector data from Highways England's MIDAS 2007 dataset that was also used in (Coullon and Pokern, 2020). Data were collected for 8th January 2007 between 6:21am and 7:09am (inclusive) from a 5km four-lane section of the M25 motorway. The authors extracted density from occupancy and speed measurements averaged every minute for each of the eight fault-free loop detectors installed on the road and spaced every 500m. However, both methods of deriving density rely on unrealistic assumptions. Density from speed data does not account for different vehicle sizes while density from occupancy depends

on different ways of obtaining average vehicle length. We use density data from occupancy as the authors in (Coullon and Pokern, 2020) derive average vehicle length estimates every minute from vehicle type information included in the MIDAS dataset. All road lanes and loop detectors are leveraged to ensure good coverage of the density domain.

3.3.1 Model comparison

		M25 data		
		Diffuse prior	"Regular" prior	Informative prior
Inference model	Greenshield's	147.5 ± 0.11	149.1 ± 0.01	150.96 ± 0.05
	Greenberg's	319.43 ± 0.07	319.5 ± 0.05	321.29 ± 0.06
	Underwood's	317.14 ± 0.1	319.16 ± 0.13	322.26 ± 0.03
	Northwestern's	251.09 ± 0.12	253.12 ± 0.02	255.24 ± 0.03
	Newell's	335.1 ± 0.03	337.48 ± 0.08	340.35 ± 0.05
	Wang's	334.0 ± 0.1	335.79 ± 0.14	337.84 ± 0.06
	Daganzo's	154.72 ± 0.05	156.07 ± 0.04	157.46 ± 0.04
	DelCastillo's	**337.58 ± 0.15**	**340.43 ± 0.1**	**342.94 ± 0.05**
	Smulder's	293.7 ± 0.03	296.99 ± 0.04	299.96 ± 0.14
	DeRomph's	289.48 ± 0.07	291.42 ± 0.17	293.42 ± 0.08
	DeRomph's continuous	293.74 ± 0.08	295.49 ± 0.07	296.84 ± 0.02

Table 3.4: Unbiased estimates of the log marginal likelihood estimator in (2.21) and their associated Monte Carlo error for every fundamental diagram model applied on the M25 dataset. Each row corresponds to the FD model used to simulate flow data for three different prior specifications (diffuse,"regular",informative) while each column corresponds to the respective FD model fitted to that data.

Table 3.4 summarises the unbiased estimates of the log marginal likelihood for the FDs in Table 1.1 applied on the M25 data. The three most plausible models are DelCastillo's, Newell's and

Wang's FDs. Their corresponding data fits and residual plots are shown in Figure 3.1. The better model evidence achieved by DelCastillo's FD across the three prior specifications may be attributed to its ω parameter which can change the smoothness of fit without affecting other FD features such as road capacity. However, it can be argued that a different choice of priors could weaken the evidence supporting the superiority of DelCastillo's model over the other two FDs. In fact, Table 3.5 and Figure 3.1 indicate that three three best models explain the same amount of variance in the data. Wang's model's goodness of fit can be explained by its ability to capture the early saturation of traffic and adjust the decay rate in the congested region of the graph depicted in Figure 3.1. The analysis of the structural differences amongst the FDs suggests that Newell's features can be adequately replicated by DelCastillo's FD (see relevant columns in Table 3.1). Therefore, a preference of Wang's and DelCastillo's models over Newell's can be sufficiently justified. Nevertheless, the achieved R^2 of 0.65 is not significantly high. The residual plots in Figure 3.1 illustrate a degree of *heteroskedasticity* in the data which implies that the multiplicative error structure may be insufficient to model real-world data. Uncertainty quantification alleviates the posterior predictive mean's inability to capture the flow scatter. However, the estimated credible interval is overestimated and underestimated for low and high density values, respectively. This effect can also be mitigated when using a heteroskedastic error structure.

		M25 data
Inference model	Greenshield's	0.08
	Greenberg's	0.62
	Underwood's	0.62
	Northwestern's	0.46
	Newell's	**0.65**
	Wang's	**0.65**
	Daganzo's	0.11
	DelCastillo's	**0.65**
	Smulder's	0.57
	DeRomph's	0.56
	DeRomph's continuous	0.56

Table 3.5: Fitted R^2 values for every fundamental diagram model applied on the M25 dataset. Each row and column correspond to the FD model used to simulate flow data and the FD model fitted to that data, respectively.

In addition, Greenberg's and Underwood's models achieve marginally poorer data fits of $R^2 = 0.62$. Both FDs have two parameters and are positively skewed, which allows them to track the fast growth of the data in the free-flow region. Indeed, the sensitivity analysis in Table 3.4 renders them hardly distinguishable. We note that despite Northwestern's similarity to Underwood's FD, the former has a faster decay rate and subsequently performs poorer than the latter.

Moreover, Smulder's and DeRomph's are overly complex model hypotheses as their fits do not compensate for the additional parameters. Even though the MCMC chains for DeRomph's tempered posteriors converged, we encountered *identifiability* issues for parameters u_f and α in the likelihood-dominated tempered posteriors. This occurred because the congested part of the model dominated the data fit and obviated the free-flow-dependent parameters. Although a different choice of temperature schedule could potentially mitigate the impact of

parameter identifiability on the marginal likelihood, it is likely that it will increase the Monte Carlo error of the log ML estimates as illustrated in (Calderhead and Girolami, 2009). Hence, we make the model identifiable by enforcing continuity at the critical density k_{crit} and ensure positivity of α by introducing the following constraints:

$$\alpha = \frac{k_{crit} u_f}{u_f - \gamma \left(\frac{k_{jam} - k_{crit}}{k_{jam} k_{crit}} \right)^{\beta}} \tag{3.1}$$

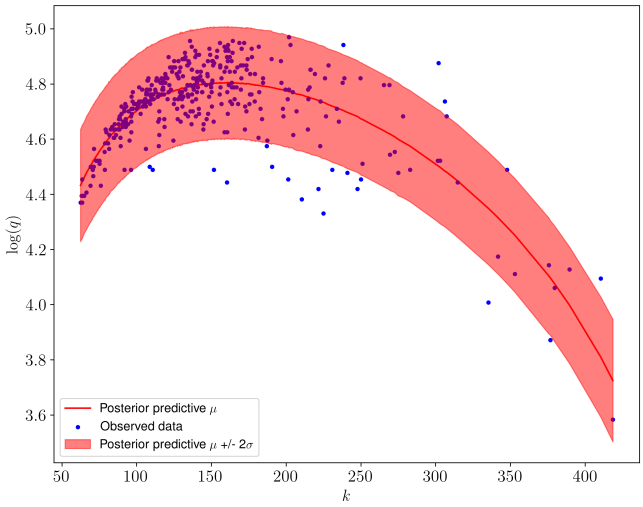
$$u_f \approx \gamma \left(\frac{k_{jam} - k_{crit}}{k_{jam} k_{crit}} \right)^{\beta} \tag{3.2}$$

Constraint (3.1) allows us to reduce the parameter space to $\boldsymbol{\theta} = (k_{crit}, k_{jam}, \gamma, \beta, \sigma^2)$. Constraint (3.2) ensures $\alpha \to \infty$ which corresponds to a near-linear approximation of the free-flow region for small k_{crit} and subsequently guarantees a unique global maximum (capacity). Tables 3.4,3.5 include the original DeRomph's FD and its reduced version for completeness. We note that DeRomhp's model reduction does not yield Smulder's model since the latter is allowed to be discontinuous. The identifiable version of DeRomph's achieves a slightly higher log ML due to parameter reduction but not a better data fit than its non-identifiable counterpart. Given DeRomph's identifiability issues we argue that DeRomph's model may fail to generalise well to other loop detector datasets despite its apparent flexibility highlighted in the synthetic data comparisons.

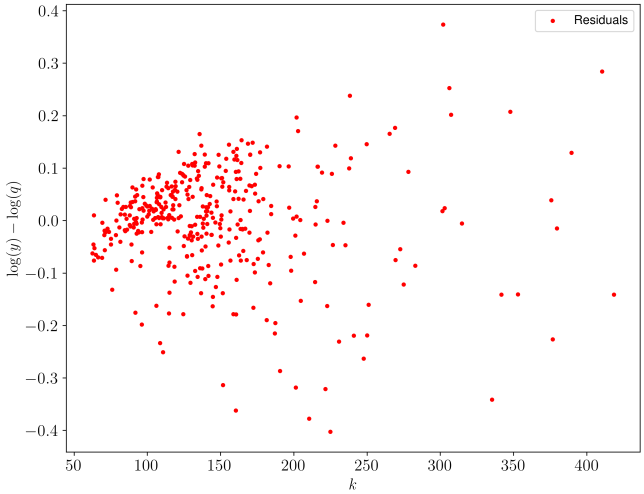
Finally, Greenshield's and Daganzo's FDs yield very poor fits due to their rigid structures (symmetric quadratic and piece-wise linear, respectively) and are therefore excluded from consideration of the most plausible constitutive law. In conclusion, taking into account Tables 3.1,3.4,3.5 and Figure 3.1 a suitable choice of model for M25 data accounting for model complexity and fit would be DelCastillo's or Wang's FDs.

3.4 Discussion

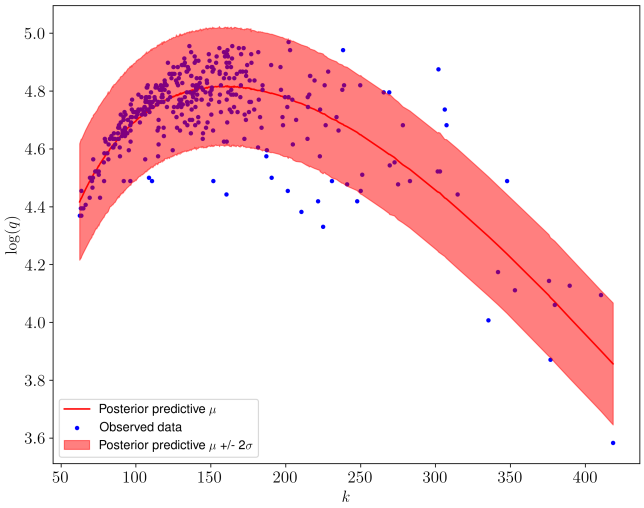
In conclusion, the synthetic data experiments in Section 3.2 allowed us to maximally reduce the model universe of parametric fundamental diagrams to Underwood's, Northwestern's, Wang's,



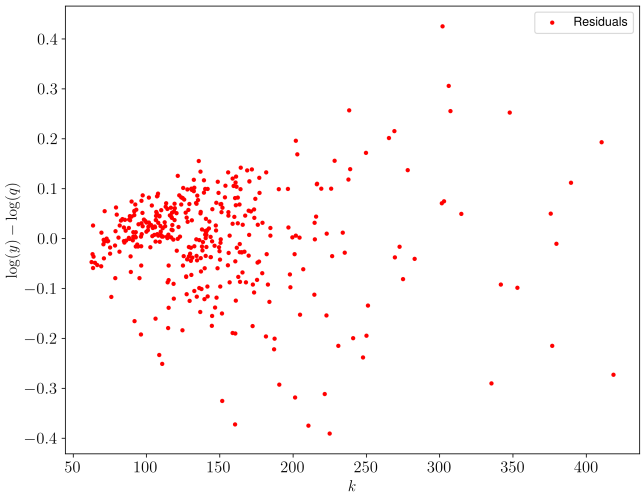
(a) Newell's FD posterior predictive plot.



(b) Newell's FD mean posterior predictive residual plot.

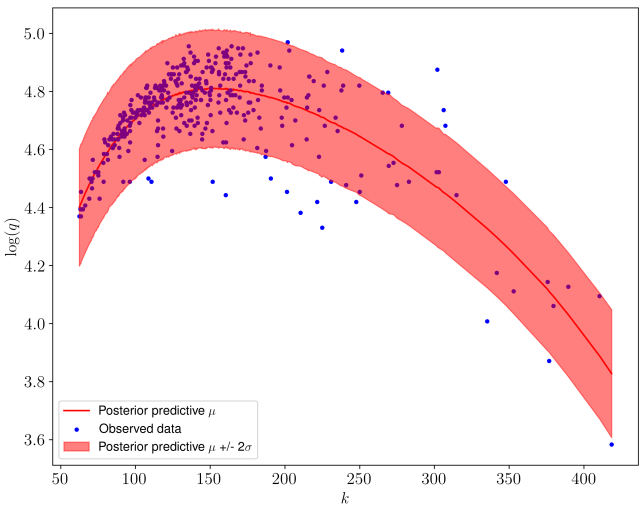


(c) Wang's FD posterior predictive plot.

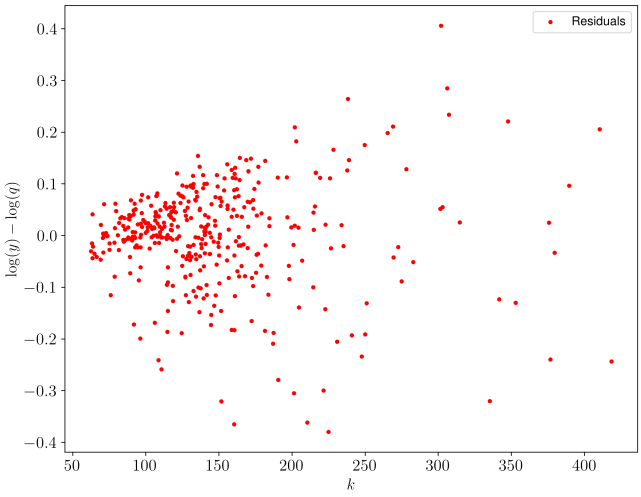


(d) Wang's FD mean posterior predictive residual plot.

Figure 3.1: Log posterior predictive mean, 2-standard deviation credible interval (left) and residual plots for the three fundamental diagrams achieving the highest log marginal likelihood estimates when applied to the M25 data in log scale (right).



(e) DelCastillo's FD posterior predictive plot.



(f) DelCastillo's mean FD posterior predictive residual plot.

Figure 3.1: Log posterior predictive mean, 2-standard deviation credible interval (left) and residual plots for the three fundamental diagrams achieving the highest log marginal likelihood estimates when applied to the M25 data in log scale (right) [continued].

DelCastillo's, DeRomph's models, which are maximally dissimilar in structure. However, the fitting of FDs in M25 data suggested that the trade-off between model complexity and data fit is not the same for all models. We conclude that DeRomph's FD may be an overly complex model hypothesis for two-dimensional phase spaces while Underwood's and Northwestern's FDs may be too simplistic. We also argue that the desired model complexity of fundamental diagrams is between two and four parameter models. The application of our methodology to M25 data allows us to validate the choice of DelCastillo's FD as the appropriate constitutive law in the work of (Coullon and Pokern, 2020). Furthermore, Table 3.6 outlines which of the six common FD properties outlined in Chapter 1 are satisfied by each FD model considered. It is evident that these criteria are not necessary for a model to achieve a sufficient model complexity and data fit trade-off. For example, DelCastillo's FD satisfies all properties while Wang's satisfies only one. However, not all of the examined constitutive laws yield physically plausible solutions to the PDE in (1.1). Fundamental diagrams not satisfying property **P6** may fail to propagate traffic waves with appropriate speeds. Therefore, constitutive laws validated

against data may yield implausible solutions for all or a subset of parameter configurations. Even models such as DelCastillo's need to be carefully treated to ensure that traffic waves cross the domain of interest as in (Coullon and Pokern, 2020). Our proposed methodology for empirical validation of constitutive laws needs to incorporate a data plausibility criterion to ensure physically interpretable solutions to traffic PDEs.

Another potential avenue of future research is to embrace the model-free paradigm of (Kirchdoerfer and Ortiz, 2016) (see Figure 1.7). According to the authors, all models are inherently misspecified and impose redundant assumptions about the true nature of the constitutive law. Therefore, a data-driven (model-free) approach of exploring the fundamental diagram relationship constrained only by the vehicle conservation equation of (1.1) may be better suited to learn the true traffic dynamics.

	Strict concavity (P1)	Bounded speed (P2)	Bounded density (P3)	Boundary speed (P4)	Boundary flow (P5)	Boundary kinematic wave speed (P6)
Greenshield's	✓	✓	✓	✓	✓	✓
Greenberg's	✓		✓			
Underwood's		✓				
Northwestern's		✓				
Newell's	✓	✓	✓	✓	✓	✓
Wang's		✓				
Daganzo's		✓	✓	✓	✓	✓
DelCastillo's	✓	✓	✓	✓	✓	✓
Smulder's		✓	✓	✓	✓	
DeRomph's			✓	✓	✓	

Table 3.6: Six common properties/assumptions about parametric fundamental diagrams outlined in Chapter 1 and their satisfiability by the ten examined FDs in Table 1.1.

Chapter 4

Future research

In Chapter 1 we identified the following gaps in macroscopic traffic flow literature:

1. There is no formal treatment of the well-posedness and solution to the inverse problem of a stochastic traffic conservation law or relaxation thereof.

2. A probabilistic framework for robust constitutive model selection has not been introduced in traffic flow modelling.

3. There is no application of the model-free data-driven paradigm in estimation of constitutive laws of traffic PDEs on either local low-dimensional or global high-dimensional phase spaces corresponding to road links and networks, respectively.

This report addressed the second gap. We argue that the application of traffic flow PDEs has been limited to motorways as they fail to capture the dynamics of urban roads and account for other transportation modes (e.g. rail, underground). Therefore tackling gaps one and three may have limited applications to single entry/exit motorways and not yield tangible benefits to urban transportation modellers.

Instead, we resort to ABM simulations of urban traffic and MATSim (Axhausen, 2016) in particular which is arguably a more realistic representation of traffic reality on a network scale. We aim to use MATSim model extensions developed by the industry partner (Arup) to model transportation in London/Leeds where supply and demand data are available for calibration. Simplified versions of these models may also be considered to ensure that statistical inference is computationally feasible and therefore the choice of model (London or

Leeds) will be rectified accordingly. We prioritise future research objectives based on the open research problems relevant to MATSim calibration identified in Chapter 1:

[**OBJ1**] (High priority): Develop black-box-driven calibration methods to assimilate Eulerian data in a synthetic/real stochastic MATSim city simulator while robustifying them to model miss-specification.

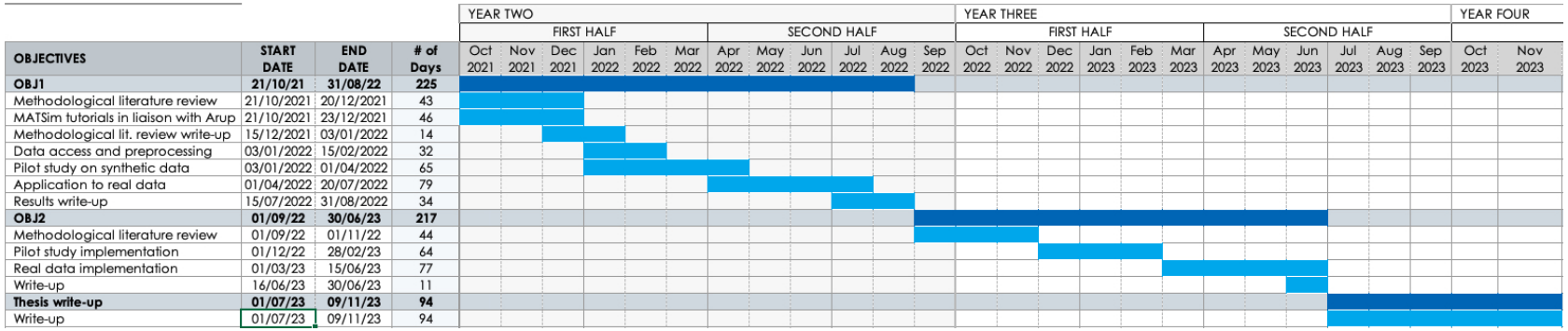
[**OBJ2**] (High priority): Develop existing calibration framework to assimilate sparse multi-resolution Eulerian and Lagrangian data in a synthetic/real stochastic MATSim city simulator.

[**OBJ3**] (Medium priority): Develop existing calibration framework to deal with parameter unidentifiability and induced simulator likelihood multi-modality in the context of a synthetic/real stochastic MATSim city simulator.

[**OBJ4**] (Medium priority): Improve existing calibration framework to account for simulator heterogeneity in input space in the context of a synthetic/real stochastic MATSim city simulator.

[**OBJ5**] (Low priority): Improve existing calibration framework to account for discontinuities and local phase transitions in the context of a synthetic/real stochastic MATSim city simulator.

A Gantt chart of realistic time estimates is depicted in Figure 4.1 addressing the objectives below depending on their priority. We note that objectives scored as high priority will be pursued in the two coming years while medium priority objectives will be revised in the beginning of third year subject to COVID-19 funding extension availability. Low priority objectives will only be undertaken if all other objectives are met ahead of schedule.



				YEAR TWO												YEAR THREE												YEAR FOUR	
				FIRST HALF						SECOND HALF						FIRST HALF						SECOND HALF							
OBJECTIVES	**START DATE**	**END DATE**	**# of Days**	Oct 2021	Nov 2021	Dec 2021	Jan 2022	Feb 2022	Mar 2022	Apr 2022	May 2022	Jun 2022	Jul 2022	Aug 2022	Sep 2022	Oct 2022	Nov 2022	Dec 2022	Jan 2023	Feb 2023	Mar 2023	Apr 2023	May 2023	Jun 2023	Jul 2023	Aug 2023	Sep 2023	Oct 2023	Nov 2023
OBJ1	**21/10/21**	**31/08/22**	**225**																										
Methodological literature review	21/10/2021	20/12/2021	43																										
MATSim tutorials in liaison with Arup	21/10/2021	23/12/2021	46																										
Methodological lit. review write-up	15/12/2021	03/01/2022	14																										
Data access and preprocessing	03/01/2022	15/02/2022	32																										
Pilot study on synthetic data	03/01/2022	01/04/2022	65																										
Application to real data	01/04/2022	20/07/2022	79																										
Results write-up	15/07/2022	31/08/2022	34																										
OBJ2	**01/09/22**	**30/06/23**	**217**																										
Methodological literature review	01/09/22	01/11/22	44																										
Pilot study implementation	01/12/22	28/02/23	64																										
Real data implementation	01/03/23	15/06/23	77																										
Write-up	16/06/23	30/06/23	11																										
Thesis write-up	**01/07/23**	**09/11/23**	**94**																										
Write-up	01/07/23	09/11/23	94																										

Figure 4.1: Gantt chart of proposed research activities for the next two years of the PhD organised by research objectives **OBJ1** and **OBJ2**.

Bibliography

Ambühl, Lukas and Monica Menendez (Oct. 2016). "Data fusion algorithm for macroscopic fundamental diagram estimation". In: *Transportation Research Part C: Emerging Technologies* 71, pp. 184–197. ISSN: 0968090X. DOI: 10.1016/j.trc.2016.07.013. URL: https://linkinghub.elsevier.com/retrieve/pii/S0968090X16301267 (cit. on p. 24).

Andrianakis, Ioannis et al. (Jan. 2017). "Efficient History Matching of a High Dimensional Individual-Based HIV Transmission Model". In: *SIAM/ASA Journal on Uncertainty Quantification* 5.1, pp. 694–719. ISSN: 2166-2525. DOI: 10.1137/16M1093008. URL: https://epubs.siam.org/doi/10.1137/16M1093008 (cit. on pp. 9, 12).

Ansorge, Rainer (Apr. 1990). "What does the entropy condition mean in traffic flow theory?" In: *Transportation Research Part B: Methodological* 24.2, pp. 133–143. ISSN: 01912615. DOI: 10.1016/0191-2615(90)90024-S. URL: https://linkinghub.elsevier.com/retrieve/pii/019126159090024S (cit. on p. 20).

Aw, A. and M. Rascle (Jan. 2000). "Resurrection of "Second Order" Models of Traffic Flow". In: *SIAM Journal on Applied Mathematics* 60.3, pp. 916–938. ISSN: 0036-1399. DOI: 10.1137/S0036139997332099. URL: http://www.siam.org/journals/siap/60-3/33209.html%20http://epubs.siam.org/doi/10.1137/S0036139997332099 (cit. on p. 16).

Axhausen, Kay W. (Aug. 2016). *The Multi-Agent Transport Simulation MATSim*. Ed. by Andreas Horni and Kai Nagel. Ubiquity Press. ISBN: 9781909188754. DOI: 10.5334/baw. URL: http://www.ubiquitypress.com/site/books/10.5334/baw/ (cit. on pp. 2, 3, 9, 69).

Banks, David L. and Mevin B. Hooten (Apr. 2021). "Statistical Challenges in Agent-Based Modeling". In: *The American Statistician*, pp. 1–8. ISSN: 0003-1305. DOI: 10.1080/00031305.2021.1900914. URL: https://www.tandfonline.com/doi/abs/10.1080/00031305.

2021.1900914%20https://www.tandfonline.com/doi/full/10.1080/00031305.2021.1900914 (cit. on p. 3).

Beaumont, Mark A, Wenyang Zhang, and David J Balding (2002). "Approximate Bayesian computation in population genetics". In: *Genetics* 162.4, pp. 2025–2035 (cit. on p. 48).

Bijak, Jakub, Seth Bullock, and Jason Hilton (2017). "Managing Uncertainty in Agent-Based Demographic Models". PhD thesis (cit. on p. 9).

Bretherton, David, Keith Wood, and Neil Raha (Jan. 1998). "Traffic Monitoring and Congestion Management in the SCOOT Urban Traffic Control System". In: *Transportation Research Record: Journal of the Transportation Research Board* 1634.1, pp. 118–122. ISSN: 0361-1981. DOI: 10.3141/1634-15. URL: https://journals.sagepub.com/doi/abs/10.3141/1634-15%20http://journals.sagepub.com/doi/10.3141/1634-15 (cit. on p. 14).

Calderhead, B and M Girolami (2009). "Estimating Bayes factors via thermodynamic integration and population MCMC". In: *Computational Statistics & Data Analysis* 12, pp. 4028–4045. URL: http://eprints.gla.ac.uk/34589/http://eprints.gla.ac.ukURL:http://www.dcs.gla.ac.uk/inference (cit. on pp. iii, 29, 40, 41, 65).

Castillo, Jose M. del (Nov. 2012a). "Three new models for the flow–density relationship: derivation and testing for freeway and urban data". In: *Transportmetrica* 8.6, pp. 443–465. ISSN: 1812-8602. DOI: 10.1080/18128602.2011.556680. URL: https://www.tandfonline.com/doi/abs/10.1080/18128602.2011.556680%20http://www.tandfonline.com/doi/abs/10.1080/18128602.2011.556680 (cit. on pp. 19, 28).

— (Nov. 2012b). "Three new models for the flow–density relationship: derivation and testing for freeway and urban data". In: *Transportmetrica* 8.6, pp. 443–465. ISSN: 1812-8602. DOI: 10.1080/18128602.2011.556680. URL: https://doi.org/10.1080/18128602.2011.556680%20http://www.tandfonline.com/doi/abs/10.1080/18128602.2011.556680 (cit. on pp. 21, 27).

Chen, D. et al. (Dec. 2004). "Freeway Traffic Stream Modeling Based on Principal Curves and Its Analysis". In: *IEEE Transactions on Intelligent Transportation Systems* 5.4, pp. 246–258. ISSN: 1524-9050. DOI: 10.1109/TITS.2004.838226. URL: https://ieeexplore.

ieee.org/stamp/stamp.jsp?tp=&arnumber=1363999%20http://ieeexplore.ieee.org/document/1363999/ (cit. on p. 24).

Clay, Robert, Le-Minh Kieu, et al. (Oct. 2020). "Towards Real-Time Crowd Simulation Under Uncertainty Using an Agent-Based Model and an Unscented Kalman Filter". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 12092 LNAI. Springer, Cham, pp. 68–79. DOI: 10.1007/978-3-030-49778-1{_}6. URL: https://link.springer.com/chapter/10.1007/978-3-030-49778-1_6%20http://link.springer.com/10.1007/978-3-030-49778-1_6 (cit. on p. 3).

Clay, Robert, Jonathan A. Ward, et al. (Dec. 2021). "Real-time agent-based crowd simulation with the Reversible Jump Unscented Kalman Filter". In: *Simulation Modelling Practice and Theory* 113, p. 102386. ISSN: 1569190X. DOI: 10.1016/j.simpat.2021.102386. URL: https://linkinghub.elsevier.com/retrieve/pii/S1569190X21000939 (cit. on pp. 3, 4).

Coullon, Jeremie and Yvo Pokern (Jan. 2020). "MCMC for a hyperbolic Bayesian inverse problem in traffic flow modelling". In: URL: http://arxiv.org/abs/2001.02013 (cit. on pp. iii, 16, 17, 28, 29, 61, 62, 67, 68).

Cranmer, Kyle, Johann Brehmer, and Gilles Louppe (Nov. 2019). "The frontier of simulation-based inference". In: *Proceedings of the National Academy of Sciences of the United States of America* 117.48, pp. 30055–30062. URL: https://arxiv.org/abs/1911.01429v3%20http://arxiv.org/abs/1911.01429 (cit. on p. 49).

Daganzo, Carlos F (1994). "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory". In: *Transportation Research Part B: Methodological* 28.4, pp. 269–287. ISSN: 0191-2615. DOI: https://doi.org/10.1016/0191-2615(94)90002-7. URL: https://www.sciencedirect.com/science/article/pii/0191261594900027 (cit. on pp. 21, 23).

— (Aug. 1995). "Requiem for second-order fluid approximations of traffic flow". In: *Transportation Research Part B: Methodological* 29.4, pp. 277–286. ISSN: 01912615. DOI: 10.

1016/0191-2615(95)00007-Z. URL: https://linkinghub.elsevier.com/retrieve/pii/0191261595000072 (cit. on p. 16).

Daganzo, Carlos F. (Apr. 1997). "A continuum theory of traffic dynamics for freeways with special lanes". In: *Transportation Research Part B: Methodological* 31.2, pp. 83–102. ISSN: 01912615. DOI: 10.1016/S0191-2615(96)00017-3. URL: https://linkinghub.elsevier.com/retrieve/pii/S0191261596000173 (cit. on p. 16).

— (Feb. 2005a). "A variational formulation of kinematic waves: basic theory and complex boundary conditions". In: *Transportation Research Part B: Methodological* 39.2, pp. 187–196. ISSN: 01912615. DOI: 10.1016/j.trb.2004.04.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0191261504000487 (cit. on p. 20).

— (Dec. 2005b). "A variational formulation of kinematic waves: Solution methods". In: *Transportation Research Part B: Methodological* 39.10, pp. 934–950. ISSN: 01912615. DOI: 10.1016/j.trb.2004.05.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0191261505000068 (cit. on p. 20).

De Romph, Erik (1996). "A dynamic traffic assignment model: Theory and applications." In: (cit. on p. 21).

Dong, Wen (May 2016). "Variational Inference with Agent-Based Models". In: URL: www.ifaamas.org%20http://arxiv.org/abs/1605.04360 (cit. on pp. 5–8).

Drake, Jennifer, J Schofer, and A May (1965). "A statistical analysis of speed-density hypotheses". In: *Traffic Flow and Transportation* (cit. on p. 21).

Dyer, Joel, Patrick Cannon, and Sebastian M Schmon (June 2021). "Approximate Bayesian Computation with Path Signatures". In: URL: https://arxiv.org/abs/2106.12555v1%20http://arxiv.org/abs/2106.12555 (cit. on p. 49).

Einbeck, Jochen and Gerhard Tutz (2006). "Modelling beyond Regression Functions: An Application of Multimodal Regression to Speed-Flow Data". In: *Source: Journal of the Royal Statistical Society. Series C (Applied Statistics)* 55.4, pp. 461–475. URL: https://about.jstor.org/terms (cit. on p. 25).

Fearnhead, Paul and Dennis Prangle (Apr. 2010). "Constructing Summary Statistics for Approximate Bayesian Computation: Semi-automatic ABC". In: *Arxiv preprint arxiv:1004.1112*

74.3, pp. 1–28. ISSN: 13697412. DOI: 10.1111/j.1467-9868.2011.01010.x. URL: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1467-9868.2011.01010.x%20https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.01010.x%20https://rss.onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2011.01010.x%20https://onlinelibrary.wiley.com/doi (cit. on p. 7).

Fintzi, Jonathan et al. (June 2016). "Efficient data augmentation for fitting stochastic epidemic models to prevalence data". In: *Journal of Computational and Graphical Statistics* 26.4, pp. 918–929. URL: arxiv:1606.07995%20http://arxiv.org/abs/1606.07995 (cit. on p. 5).

Fournier, Nicholas et al. (2018). "An integration of population synthesis methods for agent-based microsimulation". In: *Proceedings of the Annual Meeting of the Transportation Research Board, Washington, DC, USA*, pp. 13–17 (cit. on p. 3).

Frazier, David T et al. (Feb. 2019). "Bayesian inference using synthetic likelihood: asymptotics and adjustments". In: URL: http://arxiv.org/abs/1902.04827 (cit. on p. 6).

Friel, N. and A. N. Pettitt (July 2008). "Marginal likelihood estimation via power posteriors". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70.3, pp. 589–607. ISSN: 1369-7412. DOI: 10.1111/j.1467-9868.2007.00650.x. URL: http://doi.wiley.com/10.1111/j.1467-9868.2007.00650.x (cit. on pp. 30, 41, 45).

Fuhg, Jan Niklas, Michele Marino, and Nikolaos Bouklas (May 2021). "Local approximate Gaussian process regression for data-driven constitutive laws: Development and comparison with neural networks". In: URL: http://arxiv.org/abs/2105.04554 (cit. on pp. 8, 25, 29).

Garrido, Sergio et al. (Sept. 2019). "Prediction of rare feature combinations in population synthesis: Application of deep generative modelling". In: URL: http://arxiv.org/abs/1909.07689 (cit. on p. 3).

Gelman, Andrew and Xiao-Li Meng (1998). "Simulating normalizing constants: From importance sampling to bridge sampling to path sampling". In: *Statistical science*, pp. 163–185 (cit. on p. 38).

Gelman, Andrew and Donald B. Rubin (Nov. 1992). "Inference from Iterative Simulation Using Multiple Sequences". In: *Statistical Science* 7.4, pp. 457–472. ISSN: 0883-4237. DOI: 10.1214/ss/1177011136. URL: http://projecteuclid.org/euclid.ss/1177011136 (cit. on p. 44).

Golub, Gene H, Per Christian Hansen, and Dianne P O'Leary (1999). "Tikhonov regularization and total least squares". In: *SIAM journal on matrix analysis and applications* 21.1, pp. 185–194 (cit. on p. 27).

Grazzini, Jakob and Matteo Richiardi (Feb. 2015). "Estimation of ergodic agent-based models by simulated minimum distance". In: *Journal of Economic Dynamics and Control* 51, pp. 148–165. ISSN: 01651889. DOI: 10.1016/j.jedc.2014.10.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0165188914002814 (cit. on p. 5).

Grazzini, Jakob, Matteo G. Richiardi, and Mike Tsionas (Apr. 2017). "Bayesian estimation of agent-based models". In: *Journal of Economic Dynamics and Control* 77, pp. 26–47. ISSN: 01651889. DOI: 10.1016/j.jedc.2017.01.014. URL: https://linkinghub.elsevier.com/retrieve/pii/S0165188917300222 (cit. on pp. 5, 6).

Greenberg, Harold (1959). "An analysis of traffic flow". In: *Operations research* 7.1, pp. 79–85 (cit. on p. 21).

Greenshields, B D, J R Bibbins, et al. (1935). "A Study Of Traffic Capacity". In: *Highway Research Board Proceedings* 14 (cit. on p. 21).

Greenshields, B D, J T Thompson, et al. (1934). "The photographic method of studying traffic behavior". In: *Highway Research Board Proceedings*. Vol. 13 (cit. on p. 18).

Han, Cong and Bradley P Carlin (Sept. 2001). "Markov Chain Monte Carlo Methods for Computing Bayes Factors". In: *Journal of the American Statistical Association* 96.455, pp. 1122–1132. ISSN: 0162-1459. DOI: 10.1198/016214501753208780. URL: https://www.tandfonline.com/doi/abs/10.1198/016214501753208780%20http://www.tandfonline.com/doi/abs/10.1198/016214501753208780 (cit. on p. 37).

Hara, Shunsuke et al. (2013). "Configuring Agents' Attributes with Simulated Annealing". In: *Agent-Based Approaches in Economic and Social Complex Systems VII*. Tokyo: Springer

Japan, pp. 45–59. DOI: 10.1007/978-4-431-54279-7{_}4. URL: http://link.springer.com/10.1007/978-4-431-54279-7_4 (cit. on p. 5).

Hazelbag, C. Marijn et al. (May 2020). "Calibration of individual-based models to epidemiological data: A systematic review". In: *PLOS Computational Biology* 16.5. Ed. by Roger Dimitri Kouyos, e1007893. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1007893. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007893%20https://dx.plos.org/10.1371/journal.pcbi.1007893 (cit. on p. 5).

Heard, Daniel et al. (2014). "Statistical Inference Utilizing Agent Based Models". PhD thesis (cit. on p. 9).

Heppenstall, Alison J, Andrew J Evans, and Mark H Birkin (Dec. 2007). "Genetic Algorithm Optimisation of An Agent-Based Model for Simulating a Retail Market". In: *Environment and Planning B: Planning and Design* 34.6, pp. 1051–1070. ISSN: 0265-8135. DOI: 10.1068/b32068. URL: https://journals.sagepub.com/doi/10.1068/b32068%20http://journals.sagepub.com/doi/10.1068/b32068 (cit. on p. 5).

Hoogendoorn, S P, H Botma, and M M Minderhoud (2007). *Traffic flow theory and simulation*. Faculty of Civil Engineering, Geosciences, Transportation, and Traffic (cit. on pp. 18–20, 23).

Hoogendoorn, S P and P H L Bovy (June 2001). "State-of-the-art of vehicular traffic flow modelling". In: *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering* 215.4, pp. 283–303. ISSN: 0959-6518. DOI: 10.1177/095965180121500402. URL: https://journals.sagepub.com/doi/10.1177/09596518012150040 2%20http://journals.sagepub.com/doi/10.1177/095965180121500402 (cit. on p. 16).

Hoogendoorn, S.P. (1999). *Multiclass continuum modelling of multilane traffic flow*. URL: https://repository.tudelft.nl/islandora/object/uuid%3A51cd03aa-01a2-43ce-ac67-9adeec46d024 (cit. on p. 16).

Hooten, Mevin, Christopher Wikle, and Michael Schwob (Aug. 2020). "Statistical Implementations of Agent-Based Demographic Models". In: *International Statistical Review* 88.2, pp. 441–461. ISSN: 0306-7734. DOI: 10.1111/insr.12399. URL: https://onlinelibrary.wiley.com/doi/full/10.1111/insr.12399%20https://onlinelibrary.wiley.com/

doi/abs/10.1111/insr.12399%20https://onlinelibrary.wiley.com/doi/10.1111/insr.12399 (cit. on p. 5).

Hooten, Mevin B and Christopher K Wikle (Mar. 2010). "Statistical Agent-Based Models for Discrete Spatio-Temporal Systems". In: *Journal of the American Statistical Association* 105.489, pp. 236–248. ISSN: 0162-1459. DOI: 10.1198/jasa.2009.tm09036. URL: https://www.tandfonline.com/action/journalInformation?journalCode=uasa20%20http://www.tandfonline.com/doi/abs/10.1198/jasa.2009.tm09036 (cit. on p. 8).

Huang, S H and Bin Ran (2003). *An Application of Neural Network on Traffic Speed Prediction Under Adverse Weather Condition*. The University of Wisconsin-Madison. URL: http://www.wsdot.wa.gov/Rweather/ (cit. on p. 25).

Jeffreys, Harold (1998). *The theory of probability*. OUP Oxford (cit. on p. 33).

Jiang, Zhu and Yong-Xuan Huang (May 2009). "Parametric calibration of speed–density relationships in mesoscopic traffic simulator with data mining". In: *Information Sciences* 179.12, pp. 2002–2013. ISSN: 00200255. DOI: 10.1016/j.ins.2009.02.005. URL: https://dl.acm.org/doi/abs/10.1016/j.ins.2009.02.005%20https://linkinghub.elsevier.com/retrieve/pii/S002002550900070X (cit. on pp. 27, 28).

Ju, Nianqiao, Jeremy Heng, and Pierre E Jacob (Jan. 2021). "Sequential Monte Carlo algorithms for agent-based models of disease transmission". In: URL: http://arxiv.org/abs/2101.12156 (cit. on pp. 3, 5, 8, 9).

Kagho, Grace O., Milos Balac, and Kay W. Axhausen (Jan. 2020). "Agent-Based Models in Transport Planning: Current State, Issues, and Expectations". In: *Procedia Computer Science* 170, pp. 726–732. ISSN: 18770509. DOI: 10.1016/j.procs.2020.03.164. URL: https://linkinghub.elsevier.com/retrieve/pii/S187705092030627X (cit. on p. 3).

Kennedy, Marc C and Anthony O'hagan (2001). *Bayesian calibration of computer models*. Tech. rep., pp. 425–464 (cit. on p. 46).

Kessels, Femke (2019). "Mesoscopic Models". In: *Traffic Flow Modelling*, pp. 99–106. DOI: 10.1007/978-3-319-78695-7{_}6 (cit. on p. 1).

Kieu, Le-Minh, Nicolas Malleson, and Alison Heppenstall (Jan. 2020). "Dealing with uncertainty in agent-based models for short-term predictions". In: *Royal Society Open Science*

7.1, p. 191074. ISSN: 2054-5703. DOI: 10.1098/rsos.191074. URL: https://www.cambridge.org/core/product/identifier/S0266466610000599/type/journal_article%20https://royalsocietypublishing.org/doi/10.1098/rsos.191074%20https://royalsocietypublishing.org/doi/abs/10.1098/rsos.191074 (cit. on pp. 3, 4).

Kieu, Le-Minh, Dong Ngoduy, et al. (Dec. 2019). "A stochastic schedule-following simulation model of bus routes". In: *Transportmetrica B: Transport Dynamics* 7.1, pp. 1588–1610. ISSN: 2168-0566. DOI: 10.1080/21680566.2019.1670118. URL: https://www.tandfonline.com/doi/abs/10.1080/21680566.2019.1670118%20https://www.tandfonline.com/doi/full/10.1080/21680566.2019.1670118 (cit. on pp. 2, 8).

Kirchdoerfer, Trenton and Michael Ortiz (June 2016). "Data-driven computational mechanics". In: *Computer Methods in Applied Mechanics and Engineering* 304, pp. 81–101. ISSN: 00457825. DOI: 10.1016/j.cma.2016.02.001. URL: http://arxiv.org/abs/1510.04232%20http://dx.doi.org/10.1016/j.cma.2016.02.001%20https://linkinghub.elsevier.com/retrieve/pii/S0045782516300238 (cit. on pp. 26, 29, 68).

— (Nov. 2017). "Data Driven Computing with noisy material data sets". In: *Computer Methods in Applied Mechanics and Engineering* 326, pp. 622–641. ISSN: 00457825. DOI: 10.1016/j.cma.2017.07.039. URL: http://arxiv.org/abs/1702.01574%20http://dx.doi.org/10.1016/j.cma.2017.07.039%20https://linkinghub.elsevier.com/retrieve/pii/S0045782517304012 (cit. on p. 26).

Knoop, Victor L and Winnie Daamen (Apr. 2017). "Automatic fitting procedure for the fundamental diagram". In: *Transportmetrica B: Transport Dynamics* 5.2, pp. 129–144. ISSN: 2168-0566. DOI: 10.1080/21680566.2016.1256239. URL: https://www.tandfonline.com/action/journalInformation?journalCode=ttrb20http://dx.doi.org/10.1080/21680566.2016.1256239%20https://www.tandfonline.com/doi/full/10.1080/21680566.2016.1256239 (cit. on p. 27).

Lambert, Ben et al. (June 2018). "Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis". In: *Journal of Mathematical Biology* 76.7, pp. 1673–1697. ISSN: 14321416. DOI: 10.1007/s00285-018-1208-z (cit. on p. 7).

Lekone, Phenyo E. and Bärbel F. Finkenstädt (2006). "Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study". In: *Biometrics* 62.4, pp. 1170–1177. DOI: 10.1111/J.1541-0420.2006.00609.X (cit. on p. 10).

Lighthill, M. J. and G. B. Whitham (May 1955). "On kinematic waves II. A theory of traffic flow on long crowded roads". In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 229.1178, pp. 317–345. ISSN: 0080-4630. DOI: 10.1098/rspa.1955.0089. URL: https://royalsocietypublishing.org/doi/10.1098/rspa.1955.0089 (cit. on p. 14).

Lu, Fei, Mauro Maggioni, and Sui Tang (2021). "Learning interaction kernels in heterogeneous systems of agents from multiple trajectories". In: *Journal of Machine Learning Research* 22, pp. 1–67. URL: http://jmlr.org/papers/v22/19-861.html. (cit. on p. 8).

Lueck, Jordan et al. (Dec. 2019). "Who goes there? Using an agent-based simulation for tracking population movement". In: *2019 Winter Simulation Conference (WSC)*. Vol. 2019-Decem. IEEE, pp. 227–238. ISBN: 978-1-7281-3283-9. DOI: 10.1109/WSC40007.2019.9004861. URL: https://ieeexplore.ieee.org/document/9004861/ (cit. on pp. 3–5).

Lux, Thomas (June 2018). "Estimation of agent-based models using sequential Monte Carlo methods". In: *Journal of Economic Dynamics and Control* 91, pp. 391–408. ISSN: 01651889. DOI: 10.1016/j.jedc.2018.01.021. URL: https://linkinghub.elsevier.com/retrieve/pii/S0165188918300356 (cit. on p. 3).

Malleson, Nick et al. (Sept. 2020). "Simulating Crowds in Real Time with Agent-Based Modelling and a Particle Filter". In: *Journal of Artificial Societies and Social Simulation* 23.3, pp. 1–20. ISSN: 1460-7425. DOI: 10.18564/jasss.4266. URL: https://arxiv.org/abs/1909.09397v1%20http://jasss.soc.surrey.ac.uk/23/3/3.html (cit. on pp. 3, 4, 8).

Marmin, Sébastien and Maurizio Filippone (Oct. 2018). "Variational Calibration of Computer Models". In: URL: https://arxiv.org/abs/1810.12177v1%20http://arxiv.org/abs/1810.12177 (cit. on p. 48).

Maroñas, Juan et al. (Nov. 2020). "Transforming Gaussian Processes With Normalizing Flows". In: URL: https://arxiv.org/abs/2011.01596v2%20http://arxiv.org/abs/2011.01596 (cit. on p. 26).

McCULLOCH, ROBERT E and Peter E Rossi (1992). "Bayes factors for nonlinear hypotheses and likelihood distributions". In: *Biometrika* 79.4, pp. 663–676 (cit. on p. 38).

Michalopoulos, Panos G., Dimitrios E. Beskos, and Yasuji Yamauchi (Aug. 1984). "Multilane traffic flow dynamics: Some macroscopic considerations". In: *Transportation Research Part B: Methodological* 18.4-5, pp. 377–395. ISSN: 01912615. DOI: 10.1016/0191-2615(84)90019-5. URL: https://linkinghub.elsevier.com/retrieve/pii/0191261584900195 (cit. on p. 16).

Morris, Peter (June 1997). "MIDAS system development on the M25". In: *IEE Colloquium on Incident Detection and Management*. Vol. 1997. 123. IEE, pp. 6–6. DOI: 10.1049/ic:19970671. URL: https://digital-library.theiet.org/content/conferences/10.1049/ic_19970671 (cit. on p. 14).

Müller, Nora and Wolfgang Bock (June 2020). "Stochastic Perturbation of the Lighthill-Whitham-Richards Model via the Method of Stochastic Characteristics". In: URL: http://arxiv.org/abs/2006.15301 (cit. on p. 17).

Mynt-U, T and Lokenath Debnath (2007). *Linear Partial Differential Equations for Scientists and Engineers* (cit. on p. 22).

Nagatani, Takashi (Nov. 2000). "Kinetic clustering and jamming transitions in a car-following model for bus route". In: *Physica A: Statistical Mechanics and its Applications* 287.1-2, pp. 302–312. ISSN: 0378-4371. DOI: 10.1016/S0378-4371(00)00455-6 (cit. on p. 2).

Nardini, John T. et al. (Mar. 2021). "Learning differential equation models from stochastic agent-based model simulations". In: *Journal of The Royal Society Interface* 18.176, rsif.2020.0987. ISSN: 1742-5662. DOI: 10.1098/rsif.2020.0987. URL: https://royalsocietypublishi
org/doi/abs/10.1098/rsif.2020.0987%20https://royalsocietypublishing.org/doi/10.1098/rsif.2020.0987 (cit. on p. 8).

Neumann, Thorsten, Philipp L Böhnke, and Louis C Touko Tcheumadjeu (2013). "Dynamic representation of the fundamental diagram via Bayesian networks for estimating traffic flows from probe vehicle data". In: *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. IEEE, pp. 1870–1875. DOI: 10.0/Linux-x86{_}64. URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=06728501 (cit. on p. 24).

Newell, G. F. (Apr. 1961). "Nonlinear Effects in the Dynamics of Car Following". In: *Operations Research* 9.2, pp. 209–229. ISSN: 0030-364X. DOI: 10.1287/opre.9.2.209. URL: http://pubsonline.informs.org/doi/abs/10.1287/opre.9.2.209 (cit. on p. 21).

Pacchiardi, Lorenzo and Ritabrata Dutta (Dec. 2020). "Score Matched Conditional Exponential Families for Likelihood-Free Inference". In: URL: https://arxiv.org/abs/2012.10903v2%20http://arxiv.org/abs/2012.10903 (cit. on p. 7).

Panwai, Sakda and Hussein Dia (Sept. 2005). "Comparative evaluation of microscopic car-following behavior". In: *IEEE Transactions on Intelligent Transportation Systems* 6.3, pp. 314–325. DOI: 10.1109/TITS.2005.853705 (cit. on p. 2).

Payne, Harold J (1971). "Models of freeway traffic and control". In: *Mathematical Models of public systems* (cit. on p. 15).

Polson, Nicholas and Vadim Sokolov (Dec. 2015). "Bayesian analysis of traffic flow on interstate I-55: The LWR model". In: *The Annals of Applied Statistics* 9.4, pp. 1864–1888. ISSN: 1932-6157. DOI: 10.1214/15-AOAS853. URL: http://projecteuclid.org/euclid.aoas/1453993096 (cit. on pp. 17, 28, 29).

Qu, Xiaobo, Shuaian Wang, and Jin Zhang (Mar. 2015). "On the fundamental diagram for freeway traffic: A novel calibration approach for single-regime models". In: *Transportation Research Part B: Methodological* 73, pp. 91–102. ISSN: 01912615. DOI: 10.1016/j.trb.2015.01.001 (cit. on p. 27).

Radev, Stefan T. et al. (Mar. 2020). "BayesFlow: Learning complex stochastic models with invertible neural networks". In: *IEEE Transactions on Neural Networks and Learning Systems*. URL: https://arxiv.org/abs/2003.06281v4%20http://arxiv.org/abs/2003.06281 (cit. on pp. 6, 7).

Raftery, Adrian E et al. (2006). "Estimating the integrated likelihood via posterior simulation using the harmonic mean identity". In: (cit. on p. 38).

Richards, Paul I (1956). "Shock Waves on the Highway". In: *Source: Operations Research* 4.1, pp. 42–51. URL: https://www.jstor.org/stable/167515 (cit. on p. 14).

Riedel, Lukas et al. (Sept. 2020). "Utopia: A Comprehensive and Collaborative Modeling Framework for Complex and Evolving Systems". In: *Journal of Open Source Software* 5.53,

p. 2165. ISSN: 2475-9066. DOI: 10.21105/joss.02165. URL: https://joss.theoj.org/papers/10.21105/joss.02165 (cit. on p. 13).

Rogers, Alex and Peter von Tessin (2004). "Multi-Objective Calibration For Agent-Based Models". In: (cit. on p. 5).

Rudskoy, Andrey, Igor Ilin, and Andrey Prokhorov (Jan. 2021). "Digital Twins in the Intelligent Transport Systems". In: *Transportation Research Procedia* 54, pp. 927–935. ISSN: 23521465. DOI: 10.1016/j.trpro.2021.02.152. URL: https://linkinghub.elsevier.com/retrieve/pii/S235214652100332X (cit. on p. 1).

Schmon, Sebastian M, Patrick W Cannon, and Jeremias Knoblauch (Nov. 2020). "Generalized Posteriors in Approximate Bayesian Computation". In: pp. 2020–2021. URL: https://arxiv.org/abs/2011.08644v2%20http://arxiv.org/abs/2011.08644 (cit. on p. 49).

Seo, Toru et al. (Apr. 2019). "Fundamental diagram estimation by using trajectories of probe vehicles". In: *Transportation Research Part B: Methodological* 122, pp. 40–56. ISSN: 01912615. DOI: 10.1016/j.trb.2019.02.005. URL: http://arxiv.org/abs/1804.05927%20https://linkinghub.elsevier.com/retrieve/pii/S0191261518303527 (cit. on p. 24).

Shiono, Takashi (Apr. 2021). "Estimation of agent-based models using Bayesian deep learning approach of BayesFlow". In: *Journal of Economic Dynamics and Control* 125, p. 104082. ISSN: 01651889. DOI: 10.1016/j.jedc.2021.104082. URL: https://linkinghub.elsevier.com/retrieve/pii/S0165188921000178 (cit. on pp. 6, 7).

Smith, L, R Beckman, and K Baggerly (July 1995). "TRANSIMS: Transportation analysis and simulation system". In: DOI: 10.2172/88648. URL: http://www.osti.gov/servlets/purl/88648-fgWOUT/webviewable/ (cit. on pp. 2, 3).

Smulders, S (1989). "Control of Freeway Traffic Control". In: *Ph. D. thesis* (cit. on p. 21).

Snyder, Chris et al. (Dec. 2008). "Obstacles to High-Dimensional Particle Filtering". In: *Monthly Weather Review* 136.12, pp. 4629–4640. ISSN: 1520-0493. DOI: 10.1175/2008MWR2529.1. URL: https://journals.ametsoc.org/view/journals/mwre/136/12/2008mwr2529.1.xml%20http://journals.ametsoc.org/doi/10.1175/2008MWR2529.1 (cit. on p. 4).

Svensén, Markus and Christopher M Bishop (2007). *Pattern recognition and machine learning* (cit. on p. 33).

Thawornwattana, Yuttapong, Daniel Dalquen, and Ziheng Yang (Dec. 2018). "Designing Simple and Efficient Markov Chain Monte Carlo Proposal Kernels". In: *Bayesian Analysis* 13.4, pp. 1037–1063. ISSN: 1936-0975. DOI: 10.1214/17-BA1084. URL: https://projecteuclid.org/euclid.ba/1510282998 (cit. on p. 41).

Underwood, Robin T (1961). "Speed, volume and density relationships". In: *Quality and theory of traffic flow* (cit. on p. 21).

Vaart, Elske van der et al. (Sept. 2015). "Calibration and evaluation of individual-based models using Approximate Bayesian Computation". In: *Ecological Modelling* 312, pp. 182–190. ISSN: 03043800. DOI: 10.1016/j.ecolmodel.2015.05.020. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304380015002173 (cit. on p. 7).

Wang, Haizhong, Jia Li, et al. (2011). "Logistic modeling of the equilibrium speed–density relationship". In: *Transportation Research Part A: Policy and Practice* 45.6, pp. 554–566. ISSN: 0965-8564. DOI: https://doi.org/10.1016/j.tra.2011.03.010. URL: https://www.sciencedirect.com/science/article/pii/S0965856411000565 (cit. on p. 21).

Wang, Haizhong, Daiheng Ni, et al. (Jan. 2013). "Stochastic modeling of the equilibrium speed-density relationship". In: *Journal of Advanced Transportation* 47.1, pp. 126–150. ISSN: 01976729. DOI: 10.1002/atr.172. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/atr.172%20https://onlinelibrary.wiley.com/doi/abs/10.1002/atr.172%20https://onlinelibrary.wiley.com/doi/10.1002/atr.172 (cit. on p. 27).

Wang, Minghao and Xiaolin Hu (Aug. 2015a). "Data assimilation in agent based simulation of smart environments using particle filters". In: *Simulation Modelling Practice and Theory* 56, pp. 36–54. ISSN: 1569190X. DOI: 10.1016/j.simpat.2015.05.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1569190X15000696 (cit. on pp. 3, 4).

— (Aug. 2015b). "Data assimilation in agent based simulation of smart environments using particle filters". In: *Simulation Modelling Practice and Theory* 56, pp. 36–54. ISSN: 1569190X. DOI: 10.1016/j.simpat.2015.05.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1569190X15000696 (cit. on p. 4).

Ward, Jonathan A., Andrew J. Evans, and Nicolas S. Malleson (Apr. 2016). "Dynamic calibration of agent-based models using data assimilation". In: *Royal Society Open Science* 3.4,

p. 150703. ISSN: 2054-5703. DOI: 10.1098/rsos.150703. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsos.150703%20https://royalsocietypublishing.org/doi/10.1098/rsos.150703 (cit. on pp. 3, 4).

Whitham, G. B. (June 1999). *Linear and Nonlinear Waves*. Hoboken, NJ, USA: John Wiley & Sons, Inc. ISBN: 9781118032954. DOI: 10.1002/9781118032954. URL: http://doi.wiley.com/10.1002/9781118032954 (cit. on p. 23).

Wikle, Christopher K and Mevin B Hooten (Jan. 2016). "Hierarchical Agent-Based Spatio-Temporal Dynamic Models for Discrete-Valued Data". In: *Handbook of Discrete-Valued Time Series*. Chapman and Hall/CRC, pp. 369–386. ISBN: 9780429102189. DOI: 10.1201/b19485-24. URL: https://www.taylorfrancis.com/chapters/edit/10.1201/b19485-24/hierarchical-agent-based-spatio-temporal-dynamic-models-discrete-valued-data-christopher-wikle-mevin-hooten%20https://www.taylorfrancis.com/books/9781466577749/chapters/10.1201/b19485-24 (cit. on p. 9).

Wood, Simon N. (Aug. 2010). "Statistical inference for noisy nonlinear ecological dynamic systems". In: *Nature* 466.7310, pp. 1102–1104. ISSN: 0028-0836. DOI: 10.1038/nature09319. URL: https://www.nature.com/articles/nature09319%20http://www.nature.com/articles/nature09319 (cit. on p. 6).

Xu, Zhen, Wen Dong, and Sargur Srihari (Nov. 2016). "Using Social Dynamics to Make Individual Predictions: Variational Inference with a Stochastic Kinetic Model". In: URL: http://arxiv.org/abs/1611.02181 (cit. on pp. 6, 7).

Zhang, H.M. (Mar. 2002). "A non-equilibrium traffic model devoid of gas-like behavior". In: *Transportation Research Part B: Methodological* 36.3, pp. 275–290. ISSN: 01912615. DOI: 10.1016/S0191-2615(00)00050-3. URL: https://linkinghub.elsevier.com/retrieve/pii/S0191261500000503 (cit. on p. 16).

Zhao, X. and Z. Gao (Sept. 2005). "A new car-following model: full velocity and acceleration difference model". In: *The European Physical Journal B* 47.1, pp. 145–150. ISSN: 1434-6028. DOI: 10.1140/epjb/e2005-00304-3. URL: https://link.springer.com/article/10.1140/epjb/e2005-00304-3%20http://link.springer.com/10.1140/epjb/e2005-00304-3 (cit. on p. 2).

Zheng, Tongjia, Qing Han, and Hai Lin (2020). “PDE-based Dynamic Density Estimation for Large-scale Agent Systems”. In: (cit. on p. 6).

Appendices

Appendix A

Nomenclature

In this report, we adopt standard notation and denote multidimensional objects in bold letters, say $\mathbf{x}$. Lowercase letters are used to denote scalars or vectors (x or $\mathbf{x}$, respectively) while capital letters denote matrices or random variables (the difference between the two is made clear in the context in which the mathematical objects are provided). The i-th element of a vector $\mathbf{x}$ is written as x_i while the (i, j)-th element of a matrix is indexed as X_{ij}. Additional explanations on abbreviations and notation is provided below.

In this report, we adopt standard notation and denote multidimensional objects in bold letters, say $\mathbf{x}$. Lowercase letters are used to denote scalars or vectors (x or $\mathbf{x}$, respectively) while capital letters denote matrices or random variables (the difference between the two is made clear in the context in which the mathematical objects are provided). The i-th element of a vector $\mathbf{x}$ is written as x_i while the (i, j)-th element of a matrix is indexed as X_{ij}. Additional explanations on abbreviations and notation is provided below.

Table A.1: Acronyms and their corresponding descriptions.

Acronym	**Description**
ABM	Agent-based model
EnDTs	Ensemble of decision trees
GRW	Gaussian Random Walk
i.i.d.	Independent and identically distributed
KDE	Kernel density estimation
KF	Kalman filter

KL	Kullback Leibler
LSTM-NN	Long Short-term memory neural network
MAP	Maximum a posteriori
MCMC	Markov Chain Monte Carlo
ML	Marginal likelihood
MLE	Maximum likelihood estimate
MRW	Mixed Random Walk
ODE	Ordinary differential equation
PCE	Polynomial chaos expansions
PDE	Partial differential equation
pdf	Probability density function
PF	Particle filter
RE	Ratio estimation
RMSE	Root mean square error
SEIRD	Susceptible Exposed Infected Recovered Deceased
SKM	Stochastic kinetic model
SL	Synthetic likelihood
SMC	Sequential Monte Carlo
TI	Thermodynamic integration or integral
TRW	Truncated Random Walk
VI	Variation inference
w.p.	with probability

Table A.2: Commonly used notation and its corresponding description.

Symbol	**Description**
σ	Standard deviation
$\boldsymbol{\theta}_i$	Parameter vector of fundamental diagram i
$\mathbf{J}$	Jacobian matrix

$\mathbb{E}[\cdot]$	Expectation
$p(\cdot)$	Probability density function
det	Determinant
$dim(\cdot)$ Dimensionality	
$\mathbb{R}^{\mathbb{M}}$	M-dimensional plane of real numbers
$\mathbb{R}^{\mathbb{M}}{}_{>0}$	positive M-dimensional plane of real numbers
$\sim$	Distributed as
$\forall$	For all
$\exists$	There exists
$\|\|$	'Given'. Used to denote conditionality
∞	Infinity
$\int$	Integral
∂	Partial derivative
d	Ordinary derivative
log	Natural logarithm
lim	Limit

Appendix B

Simulated datasets

Name	**True parameters**
Greenshield's	$u_f = 2.0$, $k_{jam} = 45.0$
Greenberg's	$u_f = 2.2$, $k_{jam} = 45.0$
Underwood's	$u_f = 5.1$, $k_0 = 8.0$
Northwestern's	$u_f = 2.1$, $k_0 = 11.7$
Newell's	$u_f = 1.3$, $k_{jam} = 45.0$, $\lambda = 40.3$
Wang's	$u_f = 1.1$, $k_{crit} = 22.6$, $s = 4.0$
Daganzo's	$q_{crit} = 12.0$, $k_{crit} = 20.0$, $k_{jam} = 45.0$
DelCastillo's	$Z = 26.0$, $u_f = 2.8$, $k_{jam} = 45.0$, $\omega = 2.8$
Smulder's	$u_f = 1.5$, $k_{crit} = 15.0$, $k_{jam} = 45.0$, $\gamma = 16.0$
DeRomph's	$u_f = 2.5$, $k_{crit} = 20.0$, $k_{jam} = 45.0$, $\gamma = 367.0$, $\alpha = 24.0$, $\beta = 2.0$

Table B.1: Fundamental diagram relationships with their respective synthetic data simulation parameters.

Appendix C

Prior specification

	Informative variance	**Regular variance**	**Diffuse variance**
		Greenshields data	
u_f	0.1	0.2	0.3
k_{jam}	10	25	50
σ^2	0.003	0.005	0.01
		Greenbergs data	
u_f	0.05	0.1	0.2
k_{jam}	10	25	50
σ^2	0.001	0.002	0.003
		Underwoods data	
u_f	0.2	1	2
k_c	5	10	20
σ^2	0.001	0.003	0.005
		Northwesterns data	
u_f	0.1	0.25	0.5
k_{crit}	10	20	40
σ^2	0.001	0.003	0.005
		Newells data	
u_f	0.1	0.3	0.5
λ	5	10	20

Table C.1 – continued in next page

Table C.1 – continued from previous page

k_{jam}	10	20	50
σ^2	0.001	0.002	0.003
Daganzos data			
q_c	10	20	30
k_{crit}	0.2	0.5	1
k_{jam}	100	200	300
σ^2	0.003	0.005	0.01
Wangs data			
u_f	0.1	0.3	0.5
k_{crit}	1	3	5
s	5	10	20
σ^2	0.001	0.002	0.003
DelCastillos data			
Z	2	5	10
u	0.2	0.5	1
k_{jam}	10	20	50
ω	0.1	0.2	0.5
σ^2	0.001	0.002	0.003
Smulders data			
u_f	0.1	0.25	0.4
k_{crit}	2	5	10
k_{jam}	50	100	200
γ	5	10	20
σ^2	0.003	0.005	0.01
DeRomphs data			
u_f	25000	50000	100000
k_{crit}	1	2.5	5

Table C.1 – continued in next page

Table C.1 – continued from previous page

k_{jam}	10	25	50
γ	1	2	3
α	100	500	1000
β	0.03	0.05	0.1
σ^2	0.01	0.02	0.03
DeRomphs continuous data			
k_{crit}	1	3	5
k_{jam}	3	5	10
γ	0.3	0.6	1
β	0.02	0.03	0.05
σ^2	0.001	0.002	0.003

Table C.1: Informative, "regular" and diffuse prior variance specification for all ten fundamental diagrams fitted on M25 data.

	Informative variance	**Regular variance**	**Diffuse variance**
Greenshields data			
u_f	0.1	0.25	0.5
k_{crit}	1	2	4
s	0.5	1	2
σ^2	0.001	0.002	0.003
Greenbergs data			
u_f	1	2	3
k_{crit}	0.05	0.1	0.2
s	1	2	3
σ^2	0.001	0.002	0.003
Underwoods data			
u_f	0.1	1	2

Table C.2 – continued in next page

Table C.2 – continued from previous page

k_{crit}	0.001	0.002	0.003
s	0.1	0.5	2
σ^2	0.001	0.002	0.003
Northwesterns data			
u_f	0.1	0.2	0.4
k_{crit}	0.5	1	2
s	0.1	0.5	1
σ^2	0.001	0.002	0.003
Newells data			
u_f	0.005	0.1	0.2
k_{crit}	0.4	1	4
s	0.2	1	2
σ^2	0.001	0.002	0.003
Daganzos data			
u_f	0.01	0.1	0.2
k_{crit}	0.1	1	3
s	0.01	0.5	1.5
σ^2	0.001	0.002	0.003
Wangs data			
u_f	0.01	0.1	0.3
k_{crit}	0.1	1	3
s	0.05	0.1	1
σ^2	0.001	0.002	0.003
DelCastillos data			
u_f	0.01	0.1	0.2
k_{crit}	0.1	1	2
s	0.01	0.5	1

Table C.2 – continued in next page

Table C.2 – continued from previous page

σ^2	0.001	0.002	0.003
Smulders data			
u_f	0.1	0.2	0.4
k_{crit}	0.2	1	2
s	0.2	1	2
σ^2	0.001	0.002	0.003
DeRomphs data			
u_f	0.1	0.3	0.5
k_{crit}	1	3	5
s	5	10	20
σ^2	0.001	0.002	0.003

Table C.2: Informative, "regular" and diffuse prior variance specification for Wangs fundamental diagram applied on synthetic data.

	Informative variance	**Regular variance**	**Diffuse variance**
Greenshields data			
Z	1	2	4
u	0.05	0.1	0.2
k_{jam}	0.1	1	3
ω	0.01	0.1	0.2
σ^2	0.001	0.002	0.003
Greenbergs data			
Z	1	3	6
u	0.005	0.1	0.3
k_{jam}	0.1	1	4
ω	0.001	0.05	0.13

Table C.3 – continued in next page

Table C.3 – continued from previous page

σ^2	0.001	0.002	0.003
Underwoods data			
Z	0.1	1.5	3
u	0.1	1	2.5
k_{jam}	0.1	2	4
ω	0.1	0.5	1
σ^2	0.005	0.0075	0.01
Northwesterns data			
Z	0.1	1	3
u	0.1	0.5	1.5
k_{jam}	0.1	2	4
ω	0.05	0.5	1.5
σ^2	0.01	0.025	0.04
Newells data			
Z	0.1	1	3
u	0.05	0.1	0.3
k_{jam}	0.1	2	4
ω	0.05	0.1	0.3
σ^2	0.001	0.002	0.003
Daganzos data			
Z	0.1	1	4
u	0.005	0.1	0.3
k_{jam}	0.1	1	4
ω	0.1	1	4
σ^2	0.001	0.002	0.003
Wangs data			
Z	0.1	1	4.5

Table C.3 – continued in next page

Table C.3 – continued from previous page

u	0.1	0.25	0.5
k_{jam}	0.1	2	6
ω	0.1	2	6
σ^2	0.01	0.015	0.025
DelCastillos data			
Z	2	3	4
u	0.2	0.3	0.4
k_{jam}	1	3	5
ω	0.2	0.25	0.3
σ^2	0.001	0.002	0.003
Smulders data			
Z	0.1	1	4
u	0.1	0.5	1
k_{jam}	0.1	2	6
ω	0.1	0.5	1
σ^2	0.001	0.002	0.003
DeRomphs data			
Z	0.1	1	2
u	0.1	1	2
k_{jam}	0.1	1	2
ω	0.05	0.5	1.5
σ^2	0.01	0.05	0.1

Table C.3: Informative, "regular" and diffuse prior variance specification for DelCastillos fundamental diagram applied on synthetic data.

	Informative variance	**Regular variance**	**Diffuse variance**
		Greenshields data	
u_f	0.1	0.2	0.4
k_{crit}	1	4	7
k_{jam}	1	3	5
γ	1	4	7
σ^2	0.001	0.002	0.003
		Greenbergs data	
u_f	0.1	1	2.5
k_{crit}	0.1	0.2	0.4
k_{jam}	1	4	9
γ	1	3	5
σ^2	0.001	0.002	0.003
		Underwoods data	
u_f	0.1	0.5	1
k_{crit}	0.1	0.2	0.4
k_{jam}	1	4	9
γ	1	3	5
σ^2	0.001	0.002	0.003
		Northwesterns data	
u_f	0.1	0.2	0.4
k_{crit}	1	2	3
k_{jam}	1	3	5
γ	1	1.5	2
σ^2	0.001	0.002	0.003
		Newells data	
u_f	0.1	0.2	0.4

Table C.4 – continued in next page

Table C.4 – continued from previous page

k_{crit}	1	2	4
k_{jam}	1	3	5
γ	1	2	4
σ^2	0.001	0.002	0.003
Daganzos data			
u_f	0.01	0.05	0.15
k_{crit}	1	2	4
k_{jam}	1	3	5
γ	1	2	4
σ^2	0.001	0.002	0.003
Wangs data			
u_f	0.1	0.15	0.2
k_{crit}	1	2	4
k_{jam}	1	3	5
γ	1	1.5	2
σ^2	0.001	0.002	0.003
DelCastillos data			
u_f	0.1	0.2	0.4
k_{crit}	1	2	3
k_{jam}	1	3	5
γ	1	2	4
σ^2	0.001	0.002	0.003
Smulders data			
u_f	0.1	0.25	0.5
k_{crit}	1	2	3
k_{jam}	1	3	5
γ	1	2	3

Table C.4 – continued in next page

Table C.4 – continued from previous page

σ^2	0.001	0.002	0.003
		DeRomphs data	
u_f	0.1	0.2	0.4
k_{crit}	1	2	3
k_{jam}	1	3	5
γ	0.1	0.5	1
σ^2	0.001	0.002	0.003

Table C.4: Informative, "regular" and diffuse prior variance specification for Smulders fundamental diagram applied on synthetic data.

	Informative variance	**Regular variance**	**Diffuse variance**
		Greenshields data	
u_f	0.1	0.3	0.5
k_{crit}	1	2	3
k_{jam}	1	3	4
γ	0.05	0.1	0.2
α	0.5	1	2
β	0.005	0.01	0.02
σ^2	0.00025	0.0005	0.001
		Greenbergs data	
u_f	0.1	0.3	0.5
k_{crit}	0.1	0.3	0.5
k_{jam}	0.2	0.5	1
γ	0.1	0.2	0.3
α	0.1	0.2	0.3
β	0.005	0.01	0.02

Table C.5 – continued in next page

Table C.5 – continued from previous page

σ^2	0.00025	0.0005	0.001
Underwoods data			
u_f	0.1	0.5	1
k_{crit}	0.1	0.5	1
k_{jam}	1	2	3
γ	1	2	4
α	1	2	3
β	0.02	0.05	0.1
σ^2	0.001	0.002	0.003
Northwesterns data			
u_f	0.1	0.2	0.3
k_{crit}	0.5	1	1.5
k_{jam}	0.5	1	1.5
γ	1	3	6
α	0.1	0.25	0.5
β	0.01	0.05	0.1
σ^2	0.00025	0.0005	0.001
Newells data			
u_f	0.1	0.3	0.5
k_{crit}	1	2	5
k_{jam}	1	2	5
γ	3	5	10
α	5	10	15
β	0.05	0.1	0.2
σ^2	0.001	0.002	0.003
Daganzos data			
u_f	0.02	0.05	0.1

Table C.5 – continued in next page

Table C.5 – continued from previous page

k_{crit}	0.2	0.5	1
k_{jam}	0.5	1	2
γ	0.2	0.5	1
α	1	5	10
β	0.02	0.05	0.1
σ^2	0.001	0.002	0.003
Wangs data			
u_f	0.1	0.2	0.3
k_{crit}	1	2	3
k_{jam}	1	3	5
γ	1	3	5
α	0.5	1	2
β	0.1	0.25	0.5
σ^2	0.001	0.002	0.003
DelCastillos data			
u_f	0.1	0.2	0.3
k_{crit}	1	2	3
k_{jam}	1	2	4
γ	1	3	6
α	1	2	3
β	0.1	0.2	0.3
σ^2	0.001	0.002	0.003
Smulders data			
u_f	0.1	0.3	0.4
k_{crit}	1	2	3
k_{jam}	0.5	1	2
γ	0.1	0.25	0.5

Table C.5 – continued in next page

Table C.5 – continued from previous page

α	0.5	1	2
β	0.01	0.025	0.05
σ^2	0.001	0.002	0.003
DeRomphs data			
u_f	0.05	0.1	0.2
k_{crit}	0.1	0.5	1
k_{jam}	0.5	1	2
γ	1	2	4
α	0.1	0.5	1
β	0.02	0.05	0.1
σ^2	0.001	0.002	0.003

Table C.5: Informative, "regular" and diffuse prior variance specification for DeRomphs fundamental diagram applied on synthetic data.