Parameter and density estimation from real-world traffic data: A kinetic compartmental approach

Transportation Research Part B 155 (2022) 210–239

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Parameter and density estimation from real-world traffic data: A kinetic compartmental approach

Mike Pereira [a,b,*], Pinar Boyraz Baykas [c], Balázs Kulcsár [b], Annika Lang [a]

[a] Department of Mathematical Sciences, Chalmers University of Technology & University of Gothenburg, S–412 96 Göteborg, Sweden
[b] Department of Electrical Engineering, Chalmers University of Technology, S–412 96 Göteborg, Sweden
[c] Department of Mechanics and Maritime Sciences, Chalmers University of Technology, S–412 96 Göteborg, Sweden

ARTICLE INFO

Keywords:
Traffic reaction model
Macroscopic model
Hyperbolic PDE
Finite volume scheme
Lax–Friedrichs scheme
Parameter estimation
Viscosity solutions
CFL condition
Gradient descent
HighD
Real traffic data

ABSTRACT

The main motivation of this work is to assess the validity of a LWR traffic flow model to model measurements obtained from trajectory data, and propose extensions of this model to improve it. A formulation for a discrete dynamical system is proposed aiming at reproducing the evolution in time of the density of vehicles along a road, as observed in the measurements. This system is formulated as a chemical reaction network where road cells are interpreted as compartments, the transfer of vehicles from one cell to the other is seen as a chemical reaction between adjacent compartment and the density of vehicles is seen as a concentration of reactant. Several degrees of flexibility on the parameters of this system, which basically consist of the reaction rates between the compartments, can be considered: a constant value or a function depending on time and/or space. Density measurements coming from trajectory data are then interpreted as observations of the states of this system at consecutive times. Optimal reaction rates for the system are then obtained by minimizing the discrepancy between the output of the system and the state measurements. This approach was tested both on simulated and real data, proved successful in recreating the complexity of traffic flows despite the assumptions on the flux–density relation.

1. Introduction

Modeling traffic flow to reflect macroscopic vehicular patterns becomes more and more important in the area of connectivity and autonomy (Treiber et al., 2013; Kessel, 2019; Garavello et al., 2016; Piccoli and Rascle, 2013). In this regard, proposing macroscopic traffic flow models capable of reproducing traffic flow patterns observed in real-world setting is a key problem. Such patterns are traditionally observed through data collected from sensors installed on a road (e.g. loop detectors) which collect vehicle counts or occupancy times, which in turn are aggregated and filtered to yield density, flux and speed estimates (Leduc et al., 2008). Ongoing progress in image capturing and processing capabilities have permitted to multiply and democratize the use of vehicle trajectory data, which arguably provide a more complete and faithful picture of traffic behaviors since the evolution of each vehicle can be tracked along the road (Lu and Skabardonis, 2007; Krajewski et al., 2018). We also expect that the penetration of connective vehicular technology and novel sensing and communication systems will further propel the above transition.

Model-wise, macroscopic traffic flow models play a fundamental role (Garavello et al., 2016) to model network level behavior or management solutions. In particular, first order traffic flow models are predominant and reflect fundamental macroscopic properties

* Corresponding author at: Department of Mathematical Sciences, Chalmers University of Technology & University of Gothenburg, S–412 96 Göteborg, Sweden.
E-mail addresses: mike.pereira@chalmers.se (M. Pereira), kulcsar@chalmers.se (B. Kulcsár), annika.lang@chalmers.se (A. Lang).

https://doi.org/10.1016/j.trb.2021.11.006
Received 27 January 2021; Received in revised form 16 September 2021; Accepted 13 November 2021
Available online 10 December 2021

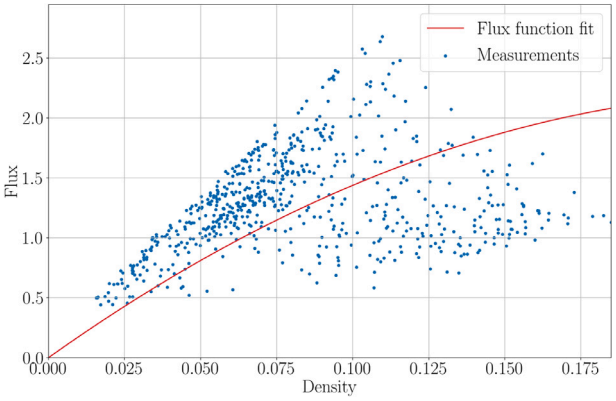


Fig. 1. Fundamental diagram obtained from trajectory data (highD dataset) using Edie's generalized definitions on a space–time grid with time step 2 s and space step ≈ 36 m.

such as conservation and flow property (Kessel, 2019). Such models consider three main fundamental quantities defined over time and along the road (space): the density of vehicles at a given location and time, the flux (or flow) of vehicles passing a given location at a given time and the average speed at a given location and time. These three variables are linked to one another by the fundamental relationship of traffic, and are captured by a first order hyperbolic partial differential equation (Garavello et al., 2016) linking flux and density of vehicles. In the well-studied Lighthill–Witham–Richards (LWR) model, the speed of vehicles is expressed as a function depending on the density only, thus yielding a partial differential equation (PDE) satisfied by this last quantity (Lighthill and Whitham, 1955; Richards, 1956).

The question of assessing the legitimacy of the hypotheses made by the continuous models is natural, especially when working with traffic flow data obtained from real-word measurements (Fan and Seibold, 2013). Take for instance Fig. 1, which shows a fundamental diagram, i.e. a scatter plot of flow vs density measurements done at the same locations and time, obtained from the real-world trajectory data (cf. Section 4.2.1 for more details) of the highD dataset (Krajewski et al., 2018). Modeling the seemingly complex flux–density relationship observed in this diagram by a simple univariate function, as done in the LWR model, then becomes a questionable choice. Hence the following questions motivated this work: How valid is the choice of a univariate (only density dependent) flux function when modeling traffic flows from real-world trajectory data? How can we enhance such traffic flow models for them to better reflect the patterns observed in the available data? Answering these questions and learning to mimick the spatial and temporal change of parameters in fundamental diagrams (flux functions) has two major benefits. First, it contributes to improve the accuracy in describing the traffic dynamics (e.g. propagation of jams). Second, it supports traffic management solution by enabling to reach better performance (i.e. traffic control oriented).

In answering the questions outlined above, most of the focus has been on two (non-exclusive) approaches. The first (Fan and Seibold, 2013) consists of proposing complex functions to fit the flux–density relationship of the fundamental diagram, may they be defined piecewise, discontinuous, or depending on numerous parameters. The second consists of proposing more complex macroscopic models for traffic flow, which takes the form of high order partial differential equations (PDEs) and/or coupled systems of equations involving not only the density of vehicles, but also other quantities such as their free speed. In this work, we advocate for a third way, in which we choose a simple fundamental diagram and a simple modeling PDE, and the model gains its flexibility from its only parameter, by allowing it to become space and/or time dependent.

A first contribution of this work is hence to propose an approach that bridges the gap between a continuous model for traffic flow, namely the LWR model, and measurements of the density. This is done by working on discretized versions of the modeling PDE obtained by finite volume methods, thus extending an approach already used to answer control-related questions linked to these traffic models (Karafyllis and Papageorgiou, 2019; Goatin et al., 2016; Delle Monache et al., 2017). In particular, a discretization scheme called Traffic Reaction Model (TRM) (Lipták et al., 2021), which models flow dynamics along a discretized road as a chemical reaction network, is for the first time used and compared to real data. This leads to a novel physical interpretation of macroscopic traffic flow models using the kinetics of chemical reactions which holds to the test of comparison with real data.

A second contribution of this work is to propose an extension of the LWR model, and subsequently of the TRM, taking the form of space and/or time dependent parameters and aiming at enhancing the modeling capabilities of the model. In particular, this extension allows to consider a setting where the fundamental diagram describing the traffic dynamics is allowed to possibly vary at different points in time and/or space. Our approach then allows to estimate, from density measurements, these variations (through the estimation of the space–time parameters) and therefore the overall dynamics of the traffic flow. The state estimation then becomes a by-product of the parameter estimation, that follows by injecting the estimated parameters back into the (discretized) traffic flow model.

A third contribution comes from providing a multilevel approach to decouple the space–time resolution of the measurements and the space–time resolution on which the PDE is discretized. In particular, the latter can now be arbitrarily small. This decoupling is justified by the fact that, even though the measurements are made on vehicles and therefore suppose some incompressible size of

their support, the quantity actually modeled is, on its side, a macroscopic density theoretically defined on any support. The jump between these two cases is in particular justified by the use of Edie's generalized definition of density and flux.

A final contribution of this work is the interpretation of discretized PDEs as discrete dynamical systems whose parameters are directly linked to those of the original continuous PDE models, thus yielding an approach similar to the cell transmission model of Daganzo (1994). In particular, the smoothness of the numerical flux resulting from the TRM allows to determine the parameters of this system (which also correspond to the parameters of the macroscopic traffic model) from density measurements only using a least-square minimization, which is tackled using gradient-based optimization. An additional validation of the model can then be performed by comparing flux values computed from these parameters (which were obtained using only density measurements) to flux measurements.

This work aims at introducing the TRM, and its kinetic compartmental interpretation of traffic dynamics, as a powerful modeling tool capable of handling the complexity of real datasets. First, this finite volume scheme yields a representation of the traffic dynamics through a system of ordinary differential equations with polynomial nonlinearities (of degree 2), hence embedding the TRM into a well established model class in system theory and opening up the way for a fine understanding of its properties (Lipták et al., 2021). Second, since it is polynomial, the scheme is differentiable and can therefore be readily applied in any traffic estimation method requiring such a property, like for instance data assimilation through the extended Kalman Filter (Wang and Papageorgiou, 2005; Schreiter et al., 2010), without fearing for potential numerical problems or loss of accuracy, as it can be the case with the Godunov scheme (which is nondifferentiable) (Blandin et al., 2012). Third, as our numerical experiments show, the TRM yields better accuracy and robustness than the Lax–Friedrichs scheme, which would be the obvious choice of a differentiable scheme (Wong and Wong, 2002; Göttlich et al., 2013).

The traffic state estimation presented in this work can be seen as a mixture of model-driven (since a physical macroscopic model describing traffic is assumed) and data-driven (since it relies on historical data, here through the least-square minimization step) approaches (see (Seo et al., 2017) for a general introduction and survey). In our setting, we use density data (only) to effectively estimate/extract the mean speed of vehicles (which is in turn considered constant, time dependent, space dependent and finally space–time dependent) thus giving a physical interpretation to the parameters optimized in the minimization approach, which is lacking in some data-driven methods such as those relying on machine learning (Karlaftis and Vlahogianni, 2011; Van Lint and Van Hinsbergen, 2012; Fulari et al., 2017) or time series models (Zhong et al., 2004). But contrary to most model-driven approaches involving the mean speed (Wang and Papageorgiou, 2005; Aw and Rascle, 2000; Zhang, 2002), no assumption on a dynamical model describing the speed is made. As our numerical experiments show, this naive estimation of the density, through the determination of the speed, is robust to missing data (thus opening the way to possible imputation applications) and is capable of great accuracy in density estimation. Besides, it allows to recreate the fundamental diagrams observed in real data (and their distinctive scattering) using the estimated speed and densities, and the fundamental relationship of traffic.

The paper is organized as follows: In Section 2 we recall the derivation of (first order) continuous models for traffic flow, their discretization using finite volume methods and introduce the traffic reaction model, as well as a proposed extension. In Section 3, we present the discrete dynamical system and how its parameters can be tuned with the objective of mimicking observed density measurements. Finally, we present in Section 4 numerical experiments that were conducted on synthetic and real traffic data.

2. Continuous traffic models and their discretization

2.1. First order macroscopic traffic flow model

Let us assume that a traffic flow is studied on a one-directional stretch of road. The density of vehicles ρ and the flux (or flow) of vehicles ϕ are two continuous quantities defined across space (i.e. along the road) and time routinely used to characterize the traffic flow in macroscopic models. Integrating the density function at a time t, across a section Δx of the road, gives the count of vehicles in Δx at t; and integrating the flux function at a location x of the road, across a time interval Δt, gives the count of the vehicles crossing x during Δt.

For any two locations $x_1 < x_2$ on the road and time t, a conservation law can hence be written to express the fact that the variation of the number of vehicles between x_1 and x_2 is equal[1] to the difference between the number of vehicles entering this road section at x_1 and those leaving the section at x_2. This gives the integral representation

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{x_1}^{x_2} \rho(t,x)\,\mathrm{d}x\right) = \phi(t,x_1) - \phi(t,x_2), \quad t \geq 0, \quad x_1 < x_2 \in \mathbb{R}\,, \tag{1}$$

which is equivalent to the following partial differential equation

$$\frac{\partial \rho}{\partial t}(t,x) + \frac{\partial \phi}{\partial x}(t,x) = 0, \quad t \geq 0, \quad x \in \mathbb{R}\,,$$

under suitable regularity conditions on the functions ρ and ϕ.

It is common to assume some additional relationship between the density ρ and the flux ϕ based on some observed links between the two quantities. For instance, when the density is 0 (meaning that the road is empty), so should the flux. Similarly, when the

[1] For sake of data availability and exposition of the method, we assume in this work that no on- or off-ramps are present in the road, meaning that no additional source or sink terms need to be added to the law of conservation of vehicles.

density is at its maximal value (corresponding to a bumper-to-bumper traffic), the flux should be zero as well. The so-called Lighthill–Whitham–Richards (LWR) model (Lighthill and Whitham, 1955; Richards, 1956) in particular stems from these observations by expressing the flux as a function f of the density (only) as follows

$$\phi(t,x) = f(\rho(t,x)), \quad t \geq 0, \quad x \in \mathbb{R},$$

where f is a (univariate) function satisfying $f(0) = 0$ and $f(\rho_m) = 0$, for ρ_m the maximal value the density can take. The simplest form f can take is arguably the Greenshields flux (Greenshields et al., 1935), for which f is a quadratic function defined by

$$f(\rho\,;\, v_m, \rho_m) = \rho \cdot v_m \left(1 - \frac{\rho}{\rho_{\mathrm{m}}}\right), \quad \rho \in [0, \rho_m], \tag{2}$$

where $v_m > 0$ is a parameter that can be interpreted as the maximal speed achievable by vehicles on the road, and the notation $f(\rho\,;\, v_m, \rho_m)$ is used to mark the fact that f is seen as function of the density ρ depending on the two parameters v_m and ρ_m. Then, the conservation law under the LWR model becomes

$$\frac{\partial \rho}{\partial t}(t,x) + \frac{\partial}{\partial x}\left(f(\rho\,;\, v_m, \rho_m)\right)(t,x) = 0, \quad t \geq 0, \quad x \in \mathbb{R}. \tag{3}$$

We will assume that the maximal density $\rho_m > 0$ is a known constant, which characterizes the capacity of the road, i.e., the maximal amount of vehicles that can fit on a road section. As such, it can be directly estimated by considering for instance the ratio between the number of lanes and the typical length of a vehicle. We hence introduce the normalized density function u, which will become from now on our main variable of interest:

$$u = \frac{\rho}{\rho_m} \in [0,1].$$

Then, PDE (3) leads to the following PDE satisfied by u:

$$\frac{\partial u}{\partial t}(t,x) + \frac{\partial}{\partial x}\left(f(u\,;\, v_m)\right)(t,x) = 0, \quad t \geq 0, \quad x \in \mathbb{R}, \tag{4}$$

where f also denotes, with a slight abuse of notation, the normalized flux function defined by

$$f(u\,;\, v_m) = u \cdot v_m (1-u), \quad u \in [0,1]. \tag{5}$$

For any given bounded initial condition $u_0 = u(0,\cdot)$, the existence and uniqueness of (entropy) solutions of PDE (4) on $\mathbb{R}_+ \times \mathbb{R}$, $T > 0$, is guaranteed in the more general case where the maximal speed is taken to be a function $(t,x) \mapsto v_m(t,x)$ that is bounded and Lipschitz continuous (Karlsen and Towers, 2004; Chen and Karlsen, 2005).

2.2. Finite volume approximations of PDEs

In general, no closed-form solution of PDE (4) is available, and therefore, numerical methods must be used to approximate it. In particular, finite volume schemes have been widely used to compute solutions of the (hyperbolic) PDEs of the form (4). As we will see, the quantities computed by such schemes relate to integrals of the solution, which are more appropriate since hyperbolic PDEs often have solutions that develop discontinuities in finite time and therefore for which point evaluations do not make sense everywhere (LeVeque, 2002).

Assume that PDE (4) is approximated on a domain that has been discretized as follows: in time we consider equidistant time steps $t_i = i\Delta t$ for step size $\Delta t > 0$ and $i \in \mathbb{N}$, in space we consider equispaced cells of size $\Delta x > 0$ with centroids $x_j = j\Delta x$ for $j \in \mathbb{Z}$. We then introduce the cell average functions U_j defined by

$$U_j(t) = \frac{1}{\Delta x}\int_{x_j - \Delta x/2}^{x_j + \Delta x/2} u(t,x)\,\mathrm{d}x, \quad t \geq 0, \quad j \in \mathbb{Z}. \tag{6}$$

Hence, U_j is the cell average of the solution u of the PDE on the jth cell. Then, the conservation law (1) applied on the jth cell yields the following differential equation satisfied by U_j

$$\frac{\mathrm{d}U_j}{\mathrm{d}t}(t) = \frac{1}{\Delta x}\left[f(u(t,x_{j-1/2})\,;\, v_m) - f(u(t,x_{j+1/2})\,;\, v_m)\right], \quad t \geq 0, \quad j \in \mathbb{Z}, \tag{7}$$

where $x_{j\pm 1/2} = x_j \pm \Delta x/2$ are the boundary points of the jth cell, and $f(u(t,x_{j\pm 1/2})\,;\, v_m) = \phi(t,x_{j\pm 1/2})/\rho_m$ denotes the (normalized) flux of vehicles at each boundary of the cell.

Finite volume methods propose to turn this set of equations into a system of ordinary differential equations by replacing the right-hand side of (7) by a function of the cell averages $\{U_j : j \in \mathbb{Z}\}$. In the particular case of (3-point) conservative schemes, the flux at a boundary point $x_{j-1/2}$ is replaced by a so-called numerical flux $F(U_{j-1}, U_j\,;\, v_m)$ depending on the cell averages of the two cells $j-1$ and j sharing that boundary and on the parameter v_m defining the flux function.

Hence, Eq. (7) yields a system of ODEs for the finite volume approximations $\widehat{U}_j$ of the cell averages U_j:

$$\frac{\mathrm{d}\widehat{U}_j}{\mathrm{d}t}(t) = \frac{1}{\Delta x}\left[F(\widehat{U}_{j-1}, \widehat{U}_j\,;\, v_m) - F(\widehat{U}_j, \widehat{U}_{j+1}\,;\, v_m)\right], \quad t \geq 0, \quad j \in \mathbb{Z}. \tag{8}$$

We assume, for all j, that the initial condition $\{U_j(0) : j \in \mathbb{Z}\}$ is known, and use it to set the initial values of the finite volume approximation $\widehat{U}_j(t_0) = U_j(t_0)$. Then the numerical solution of the PDE using such schemes can be obtained at times $t_1, t_2, \dots$ by considering an Euler time discretization (of step size Δt) of the system (8), which yields the recurrence relation

$$\widehat{U}_j^{i+1} = \widehat{U}_j^i + \frac{\Delta t}{\Delta x}\left[F(\widehat{U}_{j-1}^i, \widehat{U}_j^i\,;\, v_m) - F(\widehat{U}_j^i, \widehat{U}_{j+1}^i\,;\, v_m)\right], \quad j \in \mathbb{Z}, \quad i \in \mathbb{N}, \tag{9}$$

where $\widehat{U}_j^i$ is the quantity defined by $\widehat{U}_j^i \approx \widehat{U}_j(t_i)$. Different choices of numerical flux F yield different schemes. Among the choices most encountered in the literature, we can cite the Lax–Friedrichs (LxF) scheme and the Godunov scheme which are presented in Appendix A (LeVeque, 2002; Eymard et al., 2000; Barth et al., 2018).

2.3. Traffic reaction model

In this paper, we focus on a particular finite volume scheme (Lipták et al., 2021) for solving PDE (4) under the assumption that f is given by (2): the so-called "Traffic Reaction Model" (TRM). This particular scheme is defined by taking the numerical flux F in (9) as

$$F(u, v\,;\, v_m) = v_m u(1 - v), \quad u, v \in [0, \rho_m].$$

Consequently, for the TRM, the recurrence relation (9) can be rewritten as

$$\widehat{U}_j^{i+1} = \widehat{U}_j^i + \frac{\Delta t}{\Delta x} v_m \left[\tilde{F}_T(\widehat{U}_{j-1}^i, \widehat{U}_j^i) - \tilde{F}_T(\widehat{U}_j^i, \widehat{U}_{j+1}^i)\right], \quad j \in \mathbb{Z}, \quad i \in \mathbb{N}, \tag{10}$$

where F_T is a normalized numerical flux $\tilde{F}_T$ given by

$$\tilde{F}_T(u, v) = u(1 - v), \quad u, v \in [0, \rho_m].$$

As a finite volume scheme, the TRM has several desirable properties shared with other schemes such as the Lax–Friedrichs and the Godunov schemes. First, it is a consistent with the Greenshields flux (5), meaning that numerical flux F_T satisfies $F_T(u, u\,;\, v_m) = f(u\,;\, v_m)$ for any $u \in [0, 1]$. Then, under the assumption that the discretization steps $\Delta t, \Delta x$ satisfy a so-called Courant–Friedrichs–Lewy (CFL) condition given by

$$\frac{\Delta t}{\Delta x} \leq \frac{1}{2v_m}, \tag{11}$$

the TRM is (cf. Lipták et al., 2021):

- monotone: if the recurrence is initialized with two initial conditions $\{\widehat{U}_j^0 : j \in \mathbb{Z}\}$ and $\{\widehat{V}_j^0 : j \in \mathbb{Z}\}$ such that for any j, $\widehat{U}_j^0 \leq \widehat{V}_j^0$ then for any i and any j, $\widehat{U}_j^i \leq \widehat{V}_j^i$.
- L^∞-stable: if there exist two bounds $A, B \geq 0$ such that the initial condition satisfies, for any cell j, $U_j^0 \in [A, B]$, then for any later time i and any cell j, we also have $\widehat{U}_j^i \in [A, B]$. This means in particular that the solution on each cell stays at all time nonnegative and bounded by the capacity of the cell (provided the initial condition has these properties). In other words, the numerical flux of the TRM will never remove vehicles from an empty cell, or add vehicles to a full cell.
- convergent: the L^1 norm between the discrete cell-defined solutions and the true solution[2] converges to 0 as $\Delta x \to 0$ (with $\Delta t/\Delta x$ kept constant). This convergence result is a consequence of the consistency and monotonicity of the scheme (LeVeque, 2002).
- shock-capturing: the scheme is capable of approximating the shockwaves and rarefaction waves that arise from scalar conservation laws, propagates them with the right speed and direction. This is a consequence of the fact that the scheme is in conservative form, and more generally of the convergence result (LeVeque, 2002; Lipták et al., 2021).

The TRM has a physical interpretation that follows from modeling the road traffic dynamic (on the discretized road) as a chemical reaction network. More precisely, the process of vehicles passing from a cell j to the next one is interpreted as a chemical reaction that "transforms" a unit of occupied space O_j in j and a unit of free space Φ_{j+1} in $j+1$ into a unit of free space Φ_j in j and a unit of occupied space O_{j+1} in $j+1$ (cf. Fig. 2). Hence, the road cells are interpreted as compartments containing two homogeneously distributed chemical reactants (Free space Φ_j and Occupied space O_j) and interacting with each other (through the "transfer" reaction).

In this context, the law of mass action then allows to study the kinetics of this network of reactions (Feinberg, 2019). The rate at which a particular reaction happens is then modeled as being proportional to the product of the concentration of each one of its reactants (elevated to the power of the stoichiometric coefficient of the reactant, i.e., the number of "units" of this reactant consumed by a single reaction). Denote by $o_j(t)$ (resp $\varphi_j(t)$) the concentration of Occupied space O_j (resp. Free space Φ_j) in the jth

[2] The term true solution refers here to the notion of entropy solution of the PDE, which is the unique physically-relevant (weak) solution of the PDE (LeVeque, 2002).

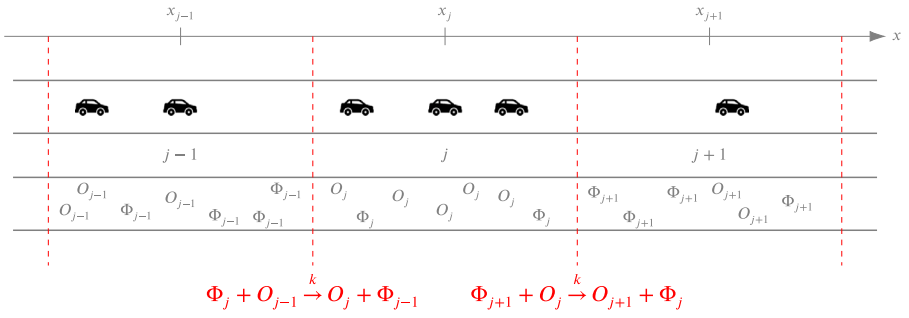


Fig. 2. Representation of the Traffic Reaction Model interpretation of traffic flow.

compartment at time $t \geq 0$. The evolution of these concentrations can be expressed as the difference between the rate at which they are produced by the rate at which they are consumed by the reactions happening in the compartment, thus giving

$$\begin{cases} \dfrac{\mathrm{d}o_j}{\mathrm{d}t} = k_{j-1\to j} o_{j-1}\varphi_j - k_{j\to j+1} o_j \varphi_{j+1} \\ \dfrac{\mathrm{d}\varphi_j}{\mathrm{d}t} = -k_{j-1\to j} o_{j-1}\varphi_j + k_{j\to j+1} o_j \varphi_{j+1} \end{cases}, \quad j \in \mathbb{Z},$$

where, for any j, $k_{j-1\to j}$ is the reaction rate (proportionality) constant of the reaction between the $(j-1)$th compartment and the jth compartment. Adding these two differential equations gives in particular that the quantity $\varphi_j + o_j$ is conserved through time.

Let us assume that all the reaction rates $k_{j-1\to j}$ are constant and equal to k. Identifying then the concentration of occupied space o_j with the density of vehicles ρ_j in the jth road cell/compartment, and the identifying the concentration of free space φ_j with a density of free space, we note that the conserved quantity $o_j + \varphi_j$ can be interpreted as the maximal capacity of the cell, which gives $o_j + \varphi_j = \rho_m$. Hence, we have

$$\frac{\mathrm{d}\rho_j}{\mathrm{d}t} = k\rho_{j-1}(\rho_m - \rho_j) - k\rho_j(\rho_m - \rho_{j+1}), \quad j \in \mathbb{Z}.$$

Dividing this last expression by ρ_m to get back to normalized densities and using an Euler time discretization then gives

$$\hat{U}_j^{i+1} = \hat{U}_j^i + C\left[\hat{U}_{j-1}^i(1-\hat{U}_j^i) - \hat{U}_j^i(1-\hat{U}_{j+1}^i)\right], \quad j \in \mathbb{Z}, \quad i \in \mathbb{N}, \tag{12}$$

where C is the quantity defined by

$$C = \Delta t \rho_m k. \tag{13}$$

Note in particular that by definition of the reaction rate k, the quantity $C\rho_m\Delta x = (k\rho_m\Delta t)(\rho_m\Delta x)$ corresponds to the maximal number of vehicles that can be transferred from the $(j-1)$th to the jth cell during a period Δt. Indeed in the ideal case where the jth cell is empty (i.e., $\varphi_j = \rho_m$), the transfer reaction between $(j-1)$ and j happens at rate $k\varphi_j o_{j-1} = k\rho_m o_{j-1}$. Hence during Δt, the concentration of reactants decreases by $k\rho_m o_{j-1}\Delta t$ times. Hence, if the $(j-1)$th cell is full ($o_{j-1} = \rho_m$), $(k\rho_m\Delta t)(\rho_m\Delta x)$ vehicles will be transferred.

This quantity can also be expressed in terms of the maximal speed v_m of the vehicles. Indeed, during Δt, vehicles can travel a distance of at most $v_m\Delta t$ meaning that at most $\rho_m(v_m\Delta t)$ can cross a cell interface during this period. By equating both expressions we get

$$k = \frac{1}{\Delta x}\frac{v_m}{\rho_m}, \quad C = \frac{\Delta t}{\Delta x} v_m,$$

and therefore, the kinetic equation (12) corresponds exactly to the TRM as defined in (10). In conclusion, the TRM is nothing more than the discretization of the kinetics of a chemical reaction network defined on the road cells, now seen as compartments capable of exchanging reactants with their neighbors.

Remark 2.1. The CFL condition (11) can also be interpreted in the context of reaction kinetics. Indeed, starting from the quantity C defined above, we have that the CFL condition is equivalent to imposing that

$$C = \Delta t \rho_m k = \frac{\Delta t}{\Delta x} v_m < \frac{1}{2}. \tag{14}$$

Let us then consider for instance the interface between the compartment $(j-1)$ and j. Recall that, according to the law of mass action, the rate at which the transfer reaction of this interface happens is $ko_{j-1}\varphi_j$, meaning that during Δt, $N(\Delta t) = ko_{j-1}\varphi_j\Delta t\Delta x$ vehicles are transferred. Note that, since o_{j-1} is upper-bounded by the maximal capacity ρ_m, we can upper-bound $N(\Delta t)$ by $N(\Delta t) \leq k\rho_m\varphi_j\Delta t\Delta x$. Then, the CFL condition yields through (14) an upper-bound for k, which in turn gives

$$N(\Delta t) < \varphi_j \frac{\Delta x}{2}.$$

This means in particular that, during Δt, the number of vehicles transferred from cell $(j-1)$ to cell j is lower than the number of free slots in the left-half of cell j. Similarly, we can prove that this same number is lower than the number of vehicles slots in the right-half of cell $j-1$ (by starting by upper-bounding φ_j).

Hence, the CFL condition allows to decouple what is happening at the different interfaces during a time-lapse Δt: indeed, at each interface between two cells $(j-1)$ and j, the reaction dynamics come down to a transfer of vehicles from the right-half of cell $(j-1)$ to the left-half of cell j, and in this sense are independent of what is happening at other interfaces (or within other half cells).

2.4. Extension of the TRM to spatial and temporal variation in parameter

Reverting to the traffic interpretation of traffic flow dynamics used by the TRM and presented in Section 2.3, working with a constant parameter v_m in the traffic model implies that the traffic flows without perturbation along a homogeneous road: indeed all the reactions between consecutive road compartments happen with the same rate. A direct generalization of this model consists in considering that these reaction rates can now vary in time or across space (i.e., two pairs of consecutive compartments can have different reaction rates).

Let us first assume that the road is infinite and discretized into cells of same size. The evolution of the normalized density in a given compartment would then take the form

$$\frac{\mathrm{d}\widehat{U}_j}{\mathrm{d}t}(t) = \rho_m k_{j-1\to j}(t)\widehat{U}_{j-1}(t)\left(1-\widehat{U}_j(t)\right) - \rho_m k_{j\to j+1}(t)\widehat{U}_j(t)\left(1-\widehat{U}_{j+1}(t)\right), \quad t\geq 0, \quad j\in\mathbb{Z}\,,$$

where, for any j, $k_{j-1\to j}(t)$ is the reaction rate of the reaction between the $(j-1)$th compartment and the jth compartment, at time t. An explicit Euler discretization of this expression then gives the recurrence

$$\widehat{U}_j^{n+1} = \widehat{U}_j^n + C_j^n\widehat{U}_{j-1}^n\left(1-\widehat{U}_j^n\right) - C_{j+1}^n\widehat{U}_j^n\left(1-\widehat{U}_{j+1}^n\right), \quad j\in\mathbb{Z}, \quad n\in\mathbb{N}\,, \tag{15}$$

where for any j, n, the quantity C_j^n is defined, at the nth time step t_n, by

$$C_j^n = \rho_m k_{j-1\to j}(t_n)\Delta t\,.$$

Note that, following the same approach as in Section 2.3, the quantity $C_j^n\rho_m\Delta x$ amounts to the maximal amount of vehicle transfers that can happen between the $(j-1)$th cell and the jth cell during a period Δt starting at time t_n. Assuming that, around the interface $(j-1)/j$, the maximal speed of the vehicles is now a time dependent function $v_m(\cdot, x_{j-1/2}) : t\mapsto v_m(t, x_{j-1/2})$, this quantity can be equated to

$$C_j^n\rho_m\Delta x = \rho_m\int_{t_n}^{t_n+\Delta t} v_m(t, x_{j-1/2})\,\mathrm{d}t\,,$$

by seeing the integral on the right-hand side as the limit of a Riemann sum, and using the similar result derived in Section 2.3 for the constant case. Hence, we have

$$k_{j-1\to j}(t_n) = \frac{1}{\rho_m\Delta x}\left(\frac{1}{\Delta t}\int_{t_n}^{t_n+\Delta t} v_m(t, x_{j-1/2})\,\mathrm{d}t\right), \quad C_j^n = \frac{\Delta t}{\Delta x}\left(\frac{1}{\Delta t}\int_{t_n}^{t_n+\Delta t} v_m(t, x_{j-1/2})\,\mathrm{d}t\right).$$

In particular, following the reasoning of Remark 2.1, we will assume that the quantities C_j^n satisfy the same condition as in the constant case, namely $C_j^n\in(0,1/2)$.

The scheme defined by (15) can be seen as finite volume scheme with a numerical flux consistent with the space–time dependent flux function f defined by

$$f(t,x,\rho(t,x)) = v_m(t,x)\rho\left(1-\frac{\rho}{\rho_m}\right). \tag{16}$$

This type of finite volume scheme was studied in the context of approximation of non-homogeneous scalar conservation laws (Chainais-Hillairet and Champier, 2001). Under some regularity assumption on the speed parameter v_m (that are not stricter than those described earlier for the existence and uniqueness of an entropy solution), this finite volume scheme converges to the entropy solution of PDE (4), hence corresponding to a LWR model with space–time varying parameter v_m (Chainais-Hillairet and Champier, 2001, Theorem 1).

3. A discrete dynamical system to bridge the gap between models and measurements

3.1. Constant parameter case

Let us assume that measurements of the density and flux of vehicles along a road are available. In particular, we assume that these measurements are made along a road discretized into N_x cells of size Δx (centered at locations $x_j = j\Delta x$, $j\in[\![0, N_x-1]\!]$) and at N_t time steps spaced by Δt (and denoted by $t_i = i\Delta t$, $i\in[\![0,N_t-1]\!]$). These measurements are collected into a density matrix $\boldsymbol{D} = \{D_j^i : j\in[\![0,N_x-1]\!], i\in[\![0,N_t-1]\!]\}\in\mathbb{R}^{N_t\times N_x}$ and a flux matrix $\boldsymbol{F} = \{F_j^i : j\in[\![0,N_x-1]\!], i\in[\![0,N_t-1]\!]\}\in\mathbb{R}^{N_t\times N_x}$, whose entries D_j^i and F_j^i are the measurements made at time t_i and location x_j. We will also assume that the maximal density ρ_m of the road is known, and that therefore a normalized density matrix $\boldsymbol{U} = \{U_j^i : j\in[\![0,N_x-1]\!], i\in[\![0,N_t-1]\!]\}\in\mathbb{R}^{N_t\times N_x}$ can be obtained

from $\boldsymbol{D}$ by dividing its entries by ρ_m. We aim at assessing whether the continuous LWR model introduced in the previous section adequately represents the traffic flow as observed through $\boldsymbol{D}$ (or $\boldsymbol{U}$) and $\boldsymbol{F}$.

In order to bridge the gap between continuous models of density and the discrete measurements at hand, and therefore to be able to compare them, we think of the density observations in $\boldsymbol{U}$ as arising from a particular solution of the (continuous) PDE (4) for some unknown (but constant) value of the parameter $\bar{v}_m$. We then propose to leverage the fact that finite volume schemes would naturally provide estimates for the entries of $\boldsymbol{U}$ (assuming $\bar{v}_m$ is known). The idea is then to look for the PDE parameter value v_m^* which gives finite volume estimates $\widehat{\boldsymbol{U}}^*$ closest to the observed data $\boldsymbol{U}$. Then, the quality of the continuous model is assessed by comparing $\widehat{\boldsymbol{U}}^*$ to $\boldsymbol{U}$, and comparing the flux matrix $\boldsymbol{F}$ to flux estimates obtained by applying the flux–density relationship (2) of the model to the density estimates $\widehat{\boldsymbol{U}}^*$.

The optimal parameter value v_m^* is obtained as follows: For any choice of parameter v_m, we can compute the finite volume discretization of PDE (4) by applying the recurrence relation (9) $N_t - 1$ times, thus giving a matrix of estimates $\widehat{\boldsymbol{U}} = \{\widehat{U}_j^i : j \in [\![0, N_x - 1]\!], i \in [\![0, N_t - 1]\!]\}$. The initial state of this recurrence is set up using $\boldsymbol{U}$ as

$$\widehat{U}_j^0 = U_j^0, \quad j \in [\![0, N_x - 1]\!] . \tag{17}$$

Observed data is only available on a road of finite length. As has been seen earlier, the scheme only takes values from the neighboring compartments into account. Therefore, we choose boundary conditions using once again $\boldsymbol{U}$ by imposing

$$\widehat{U}_0^i = U_0^i, \quad \widehat{U}_{N_x-1}^i = U_{N_x-1}^i, \quad i \in [\![0, N_t - 1]\!] . \tag{18}$$

This overall process is seen as computing the output of a discrete dynamical system at times $t_1, \dots, t_{N_t-1}$. Indeed, note that for the finite volume schemes (10), (32) and (33) considered in this paper, we can introduce the (unit-free) scaling parameter C by

$$C = \frac{\Delta t}{\Delta x} v_m \tag{19}$$

and then write the finite volume recurrence relation (9) as

$$\widehat{\boldsymbol{U}}^{i+1}(C) = \mathcal{H}^i(\widehat{\boldsymbol{U}}^i(C), C\,; \boldsymbol{U}), \quad i \in [\![0, N_t - 2]\!] , \tag{20}$$

where $\widehat{\boldsymbol{U}}^i \in \mathbb{R}^{N_x}$ is the vector defined by $\widehat{\boldsymbol{U}}^i = [\widehat{U}_0^i, \dots, \widehat{U}_{N_x-1}^i]^T$, and $\mathcal{H}^i = (\mathcal{H}_0^i, \dots, \mathcal{H}_{N_x-1}^i) : \mathbb{R}^{N_x} \times \mathbb{R}^{N_C} \to \mathbb{R}^{N_x}$ is the transformation defined in part by the boundary conditions (18) as

$$\mathcal{H}_j^i\left(\widehat{\boldsymbol{U}}^i, C\,; \boldsymbol{U}\right) = \begin{cases} U_0^{i+1} & \text{if } j = 0 \\ h(\widehat{U}_{j-1}^i, \widehat{U}_j^i, \widehat{U}_{j+1}^i) + C\left[\tilde{F}(\widehat{U}_{j-1}^i, \widehat{U}_j^i) - \tilde{F}(\widehat{U}_j^i, \widehat{U}_{j+1}^i)\right] & \text{if } j \in [\![1, N_x - 2]\!] \\ U_{N_x-1}^{i+1} & \text{if } j = N_x - 1 \end{cases} , \tag{21}$$

with h and $\tilde{F}$ depending on the choice of numerical scheme (see Equations (10), (32) and (33)). The discrete dynamical system is then defined as follows:

- the state vector of the system contains the finite volume approximations of the PDE across the discretized road, at a given time step;
- the initial state of the system is the vector $\widehat{\boldsymbol{U}}^0$, defined by the initial condition (17) of the scheme;
- the recurrence relation (20) defines the successive state updates;
- the scaling parameter C acts like a control parameter of the system.

Remark 3.1. Note that the CFL condition (11) actually imposes a restriction on the domain of definition of the control parameter C: for the schemes considered in this paper to yield approximations of the cell averages of the solution (and ensure that the recurrence does not diverge), we should only consider $C \in (0, 1/2)$.

The optimal PDE parameter v_m^* is then obtained by finding the value of the control parameter C of the discrete dynamical system that minimizes a cost function measuring the discrepancy between the output of the system and the data $\boldsymbol{U}$. In particular, we consider a least-square approach, meaning that the optimal control parameter C^* will be the solution of the problem

$$C^* = \underset{C \in (0,1/2)}{\operatorname{argmin}} \sum_{i=0}^{N_t-1} \sum_{j=0}^{N_x-1} \left(\widehat{U}_j^i(C) - U_j^i\right)^2 . \tag{22}$$

Following (19), this gives in turn an optimal PDE parameter v_m^* given by

$$v_m^* = \frac{\Delta x}{\Delta t} C^* ,$$

and optimal finite volume estimates given by $\widehat{\boldsymbol{U}}^* = \widehat{\boldsymbol{U}}(C^*)$.

Finally, we turn the minimization problem (22) into an unconstrained minimization problem by introducing the parameter θ defined by

$$C(\theta) = \frac{1}{2}\operatorname{lgt}(\theta), \quad \theta \in \mathbb{R} , \tag{23}$$

where $\mathrm{lgt} : \mathbb{R} \to (0,1)$ is the logit function[3] which defines a strictly increasing bijection between $\mathbb{R}$ and $(0,1)$. The strict monotonicity and smoothness of lgt then allow to cast the minimization problem (22) into the following equivalent minimization problem

$$\theta^* = \underset{\theta\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=0}^{N_t-1}\sum_{j=0}^{N_x-1}\left(\widehat{U}_j^i(C(\theta)) - U_j^i\right)^2 , \tag{24}$$

where, for any $\theta \in \mathbb{R}$, $C(\theta) \in (0,1/2)$ is defined by (23). Then, the optimal control parameter C^* of Eq. (22) is simply obtained by taking $C^* = C(\theta^*)$.

Following from the boundary conditions (18) and initial conditions (17), the minimization problem (24) then boils down to the unconstrained minimization of the cost function L defined by

$$L(\theta) = \frac{1}{2}\sum_{i=1}^{N_t-1}\sum_{j=1}^{N_x-2}\left(\widehat{U}_j^i(C(\theta)) - U_j^i\right)^2, \quad \theta \in \mathbb{R} . \tag{25}$$

This minimization task can be in particular tackled using gradient-based optimization problems since the structure of the recurrence relation (20) can be leveraged to derive analytic expressions for the gradient of L (cf. Appendix C). That is of course if we assume that we can take the derivative of the maps $\mathcal{H}^i$ in (20), which in our case forces us to work with either the LxF scheme or the TRM scheme. In particular, in the applications presented in this paper, only these two schemes are considered and the conjugate gradient algorithm is used to perform the minimization (Nocedal and Wright, 2006).

3.2. Varying parameter case

The approach presented in the previous section naturally extends to the assumption where parameters varying in space or time are considered (as described in Section 2.4). The discrete dynamical system is defined using the recurrence relation (15), and its output is once again compared to the density data $\boldsymbol{U}$ to derive optimal control parameter values through a minimization approach. In particular,

- the boundary and initial conditions are set in the same way as in the constant case;
- the recurrence relation defining the system takes the form

$$\widehat{U}_j^{n+1} = \widehat{U}_j^n + C_j^n\widehat{U}_{j-1}^n\left(1-\widehat{U}_j^n\right) - C_{j+1}^n\widehat{U}_j^n\left(1-\widehat{U}_{j+1}^n\right), \quad j \in [\![1, N_x-2]\!], \quad n \in [\![0, N_t-2]\!] ;$$

- the control parameters of the system are the coefficients $\boldsymbol{C} = \{C_j^n : j \in [\![0,N_x]\!], n \in [\![0,N_t-1]\!]\}$, and are determined by minimizing (without constraints) a cost function L given as the sum of a least-square cost and a regularization term $R(\boldsymbol{C})$ (clarified below):

$$L(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{N_t-1}\sum_{j=1}^{N_x-2}\left(\widehat{U}_j^i(\boldsymbol{C}(\boldsymbol{\theta})) - U_j^i\right)^2 + \lambda R(\boldsymbol{C}(\boldsymbol{\theta})), \quad \boldsymbol{\theta} \in \mathbb{R}^{(N_x+1)N_t} ,$$

  where $\boldsymbol{C}(\boldsymbol{\theta}) \in (0,1/2)^{(N_x+1)N_t}$ is obtained by applying the function (23) to each entry of $\boldsymbol{\theta} \in \mathbb{R}^{(N_x+1)N_t}$ and $\lambda > 0$ is a hyperparameter balancing the importance of the least-square minimization of the regularization;
- the following regularization term $R(\boldsymbol{C})$ is considered:

$$R(\boldsymbol{C}) = \frac{1}{2}\left(\sum_{j=0}^{N_x}\sum_{i=0}^{N_t-2}(C_j^i - C_j^{i+1})^2 + \sum_{i=0}^{N_t-1}\sum_{j=0}^{N_x-1}(C_j^i - C_{j+1}^i)^2\right), \quad \boldsymbol{C} \in (0,1/2)^{(N_x+1)N_t} . \tag{26}$$

Note that the regularization term introduced above plays two roles. On the one hand it allows to reduce the risk of overfitting: indeed the number of parameters now amounts to $N_t(N_x+1)$ which is larger than the number of terms in the least-square term, and thus increases the risk of overfitting. On the other hand it allows to ensure some kind of smoothness in space and time of the parameters, as sharp changes between consecutive coefficients in space or time are penalized. This kind of smoothness assumption of the space–time varying parameter v_m is usually required to prove the existence and uniqueness of solutions of PDE (4) and explains why we try to enforce it in the estimation approach (Karlsen and Towers, 2004; Chen and Karlsen, 2005).

Finally, the minimization of the cost function is performed once again using the Conjugate gradient algorithm, while using the explicit formula of the gradient given in Appendix C.

3.3. Extension to a multilevel approach

So far, the finite volume approximations $\widehat{\boldsymbol{U}}$ were computed on the same discretization grid as the one used to create the density matrix $\boldsymbol{U}$. This has two consequences. First, if the discretization steps are large with respect to the size of the domain, we can fear that the finite volume scheme will not yield satisfactory approximations of the cell averages of the solution. Second, this choice implicitly imposes a restriction on either the range of admissible parameters v_m or the discretization pattern $\Delta t, \Delta x$ since we also

[3] The logit function is defined by $\mathrm{lgt}(\theta) = (1+e^{-\theta})^{-1}$, $\theta \in \mathbb{R}$, and has an inverse defined by $\mathrm{lgt}^{-1}(y) = -\log(y^{-1}-1)$, $y \in (0,1)$.

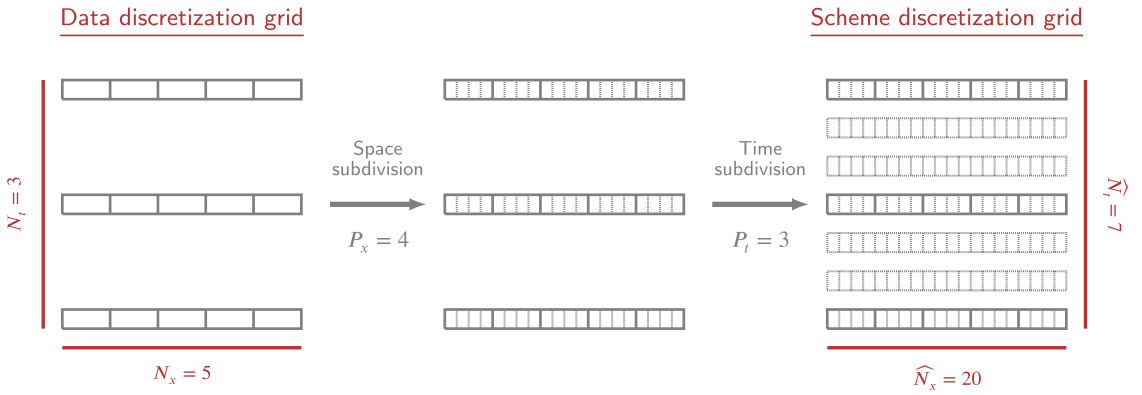


Fig. 3. Space and time subdivisions of the multilevel approach. The discretization grid on which the data are defined (left, in black) is subdivided in space (each cell is subdivided into $P_x = 4$ subcells) and in time ($P_t - 1 = 2$ intermediate time steps are added between each consecutive time steps of the initial discretization). The resulting grid (right, in gray) is used to compute the finite volume density approximations.

impose that for the discretization steps used to compute $\widehat{\boldsymbol{U}}$, the CFL condition (11) should be satisfied. For instance, assuming that we upper-bound the admissible values of the critical speed v_m by a speed of 180 km h^{-1} = 50 m s^{-1}, the CFL condition (11) gives that the discretization steps $\Delta t, \Delta x$ should satisfy $\Delta t/\Delta x < 0.01$ s m^{-1}, or equivalently $\Delta t < 0.01\Delta x$ s. Such a coupling may lead to consider very overly small time step sizes or large space step sizes, which can be limiting if one is not able to change the discretization pattern of the data.

To circumvent these limitations, we propose to use a multilevel approach where a different (and finer) discretization pattern, denoted by $\widehat{\Delta t}$, $\widehat{\Delta x}$, is used for the finite volume computations. In particular, we will take for some $P_t, P_x \in \mathbb{N}^*$,

$$\widehat{\Delta t} = \frac{\Delta t}{P_t}, \quad \widehat{\Delta x} = \frac{\Delta x}{P_x}, \tag{27}$$

meaning that the discretization pattern of the finite volume scheme will be a subdivision of the discretization pattern $\Delta t, \Delta x$ of the data (cf. Fig. 3). Therefore, the CFL condition will become

$$\frac{\widehat{\Delta t}}{\widehat{\Delta x}} = \frac{\Delta t}{\Delta x}\frac{P_x}{P_t} \leq \frac{1}{2v_m}. \tag{28}$$

Thus introducing two additional parameters P_t, P_x can be used to enforce the CFL condition without having to impose anything on the discretization pattern of the data.

The choice (27) yields that each road cell used to compute the density matrix $\boldsymbol{U}$ is subdivided into P_x subcells and each time step is subdivided into P_t steps. Hence, the finite volume matrix $\widehat{\boldsymbol{U}}$ will have size $(P_t(N_t - 1) + 1) \times (P_x N_x)$. We then note that, for $i \in [\![0, N_x - 1]\!], j \in [\![0, N_t - 1]\!]$ and by definition, each entry U_j^i of the density matrix $\boldsymbol{U}$ can be expressed as function of the solution u of the PDE (4) (with parameter $\bar{v}_m$) as

$$U_j^i = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j-1/2}+\Delta x} u(i\Delta t, x)\,\mathrm{d}x = \frac{1}{P_x}\sum_{k=0}^{P_x-1}\frac{1}{\widehat{\Delta x}}\int_{x_{j-1/2}+k\widehat{\Delta x}}^{x_{j-1/2}+(k+1)\widehat{\Delta x}} u(iP_t\widehat{\Delta t}, x)\,\mathrm{d}x,$$

where the integral on the right-hand side corresponds to the cell average of the solution over the kth subcell of the jth road cell, at time $iP_t\widehat{\Delta t}$. Since this last quantity is approximated by the entry $(iP_t, k + jP_x)$ of the finite volume matrix $\widehat{\boldsymbol{U}}$, we deduce that an approximation of U_j^i is obtained by taking the average of the quantities $\{\widehat{U}_{k+jP_x}^{iP_t} : k \in [\![0, P_x - 1]\!]\}$.

In order to use the recurrence relation (9), initial and boundary conditions defined on the finer discretization grid of the finite volume scheme are needed. We deduce them from the data $\boldsymbol{U}$ by imposing an initial condition constant across all subcells of a given road cell, i.e.

$$\widehat{U}_{k+jP_x}^0 = U_j^0, \quad j \in [\![0, N_x - 1]\!], \quad k \in [\![0, P_x - 1]\!], \tag{29}$$

and for the boundary conditions, by considering a linear interpolation of the boundary conditions obtained from the data, i.e.

$$\begin{cases} \widehat{U}_{k+0P_x}^{l+iP_t} = U_0^i + \dfrac{l}{P_t}(U_0^{i+1} - U_0^i) \\ \widehat{U}_{k+(N_x-1)P_x}^{l+iP_t} = U_{N_x-1}^i + \dfrac{l}{P_t}(U_{N_x-1}^{i+1} - U_{N_x-1}^i) \end{cases}, \quad i \in [\![0, N_t - 2]\!], \quad k \in [\![0, P_x - 1]\!], \quad l \in [\![0, P_t]\!]. \tag{30}$$

The minimization problems introduced in the previous sections can then be readily reformulated to account for the difference in discretization steps between the data matrix and the finite volume estimates, as presented in detail in Appendix B.

3.4. Extension to missing data

Up until now, we assumed that the data at hand consisted of density measurements on all of the cells discretizing a given road, and at regular time steps. But our approach can readily be extended to the case where the density at some of these cells is not measured at all. This can be done by simply removing from the least-square term of the cost function all the terms that would involve missing densities, and proceed with the same minimization problem. Following Appendix B and the two previous subsections, this means in particular that the cost function of the minimization problem (in a multilevel setting with space–time varying parameters) now takes the form

$$\tilde{L}_{[P_t,P_x]}(\theta) = \frac{1}{2}\sum_{i=1}^{N_t-1}\sum_{j_c\in I_c}\left(\frac{1}{P_x}\sum_{k=0}^{P_x-1}\hat{U}^{iP_t}_{k+j_cP_x}(\boldsymbol{C}(\theta)) - U^i_{j_c}\right)^2 + \lambda R(\boldsymbol{C}(\theta)), \quad \theta\in\mathbb{R}\,, \tag{31}$$

where $I_c \subset [\![1, N_x - 2]\!]$ is the set of observed road cells, $\lambda > 0$ and $R(\boldsymbol{C}(\theta))$ is the regularization term (26). Following Corollary C.4, the minimization of this cost function can once again be tackled using a gradient-based optimization algorithm and in particular the conjugate gradient algorithm.

In the end, all that is really required to run the estimation method is to know the densities at the boundary cells, and at least one cell in between (so that $I_c \neq \emptyset$). We do not need to know the densities at every cell, and can therefore apply our method to datasets where density measurements are only available at the boundary of the road and at least one point in between (as it is often the case in numerous datasets of density measurements taken from loop detectors on highways).

4. Numerical experiments

The goal of this section is to test the traffic state estimation approach presented in the previous section and assess its robustness and accuracy. Our approach relies on knowing, for a given road segment divided into cells, the evolution of the density on each cell through time in order to estimate reaction rates (which can be assimilated to a mean speed parameter). We first carry out a study on a synthetic dataset, obtained by numerically solving the PDE (4) defining the macroscopic traffic flow model. We then carry out a second study based on real traffic measurements (cf. Section 4.1). To be able to fulfill our requirement for exhaustive data on the evolution of the density along the road, we use in this second study the highD dataset, which consists of trajectory data obtained from processing video recordings of the road activity on a German highway (Krajewski et al., 2018). Edie's generalized definition of density (and flux) is used to create, from these trajectories, density (and flux) estimates at any space–time point of a grid (cf. Section 4.2).

Note that since what is actually needed is the density in each road cell, this second study could be replicated to a setting where occupancy sensors are laid out on each cell, or other types of sensors such as floating car data sensors are used to measure these densities. Note however that in this latter case, the penetration rate of the fleet should be high enough to allow a complete and accurate picture of the density evolution on the road.

However, with the highD dataset, the trajectories of all the vehicles that went through the road segment during the recording period are available: indeed, each vehicle was individually tracked throughout its travel on the segment. Hence, the penetration rate can potentially be 100%. But since the detection of vehicles and trajectories in the video frames was automated, the actual penetration rate is 99%, which is still very high (Krajewski et al., 2018). Regarding the accuracy of the dataset, one should note that the video recordings were made with very high resolution cameras put on drones. In the end, the size of a pixel on a video frame is of only 10 cm (for a 400 m road). Besides, the automated algorithm tasked with locating the vehicles' midpoint in longitudinal and lateral directions was very accurate, yielding mean positional errors below 3 cm in comparison to the manually created labels. To our knowledge, the highD dataset is one of the few datasets available that offer such exhaustive and accurate trajectory data. Hence, contrary to measurements based on probe vehicles, we can basically consider that all the vehicles are tracked through time and therefore and that an accurate picture of the density evolution is available.

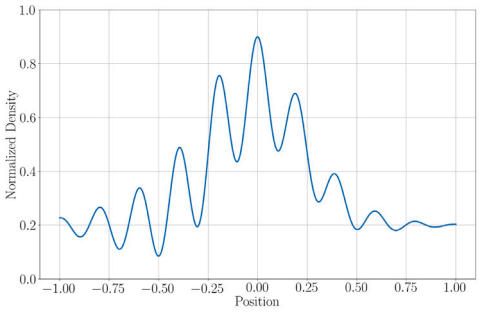
4.1. Application to parameter identification

In this case study, we numerically solve PDE (4) using the Godunov finite volume scheme over a space domain $[-1.5, 1.5]$ and a time frame $[0, 1]$, with a parameter value $\bar{v}_m = 1$ and a maximal density $\rho_m = 1$. In particular, the space discretization step is chosen small compared to the domain extension, namely $\overline{\Delta x} = 10^{-4}$, in order to guarantee that the numerical solution is close to the true solution. As for the time discretization step, it is set to $\overline{\Delta t} = 0.25\overline{\Delta x}/\bar{v}_m$ in accordance with the CFL condition (11). Hence $\bar{N}_x = 30000$ space cells and $\bar{N}_t = 40001$ time steps are considered. The initial condition is taken as

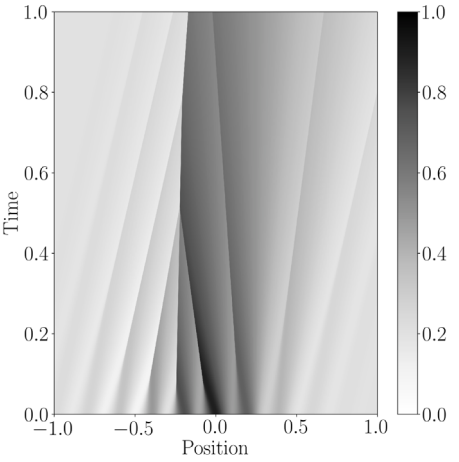
$$u_0(x) = 0.5e^{-10x^2} + 0.2\left(1 + \cos(10\pi x)e^{-(3x^2+x)}\right), \quad x\in[-1.5.1.5]\,,$$

and therefore has a profile that is non-symmetric and has oscillations across space (cf. Fig. 4). Note that reflexive boundary conditions are considered, meaning that we set $U^i_{-1} = U^i_0$ and $U^i_{N_x} = U^i_{N_x-1}$ for any $i \geq 0$ in the recurrence (9). The resulting numerical solution is cropped in space into the section $[-1, 1]$ in order to avoid any possible boundary effect coming from the boundary conditions: this solution, represented in Fig. 4, is from now on considered as the ground truth solution.

Starting from the PDE solution described above, we build density data matrices $\boldsymbol{U}$ corresponding to different choices of discretization steps $\Delta t, \Delta x$. Examples of such density matrices are represented in Fig. 5. We then estimate for each density matrix the

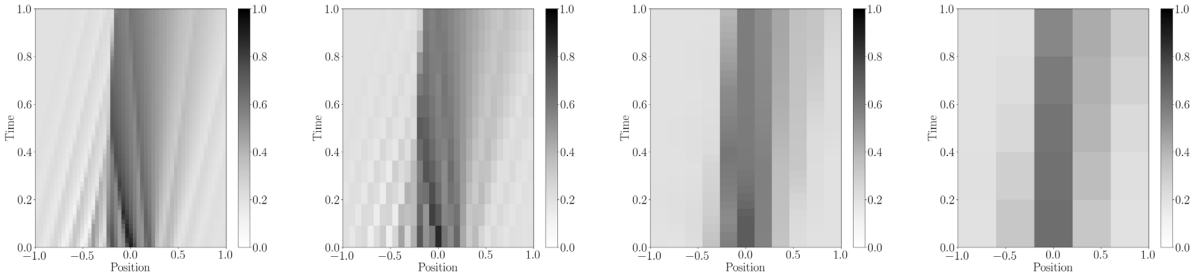


(a) Initial condition.



(b) Solution.

Fig. 4. Initial condition and associated solution of the PDE used in the synthetic case.



(a) $N_t = 51, N_x = 51$. (b) $N_t = 11, N_x = 31$. (c) $N_t = 31, N_x = 11$. (d) $N_t = 5, N_x = 5$.

Fig. 5. Density matrices obtained after various choices of discretizations (into N_x cells in space and N_t steps in time) of the PDE solution.

value of the PDE parameter v_m^* by solving the minimization problem (24), using both the TRM and the LxF scheme. We also compute the Root Mean Square Error (RMSE) between the considered density matrix $\boldsymbol{U}$ and the corresponding finite volume approximations $\widehat{\boldsymbol{U}}^* = \widehat{\boldsymbol{U}}(v_m^*)$. The discretization steps used by these schemes are set according to Section 3.3: the finite volume approximations are computed on a grid obtained by applying $P_x \in \{1, 3, 5\}$ space subdivisions and P_t time subdivisions of the discretization grid of $\boldsymbol{U}$, where for each value P_x, P_t is the smallest integer so that the CFL condition (28) is satisfied.

The relative errors between the estimated parameters and the true value $\bar{v}_m = 1$ are given in Table 1 and the RMSE values between the associated finite volume approximations and the density matrices are given in Table 2. First, one can note that for density matrices with more than 11 space cells, errors on the parameter estimation lower than 15% (and even in some case lower than 5%) can be obtained. This shows that the parameter can indeed be identified from the discretized density matrices. As for the residual error on the parameter, it can be explained by the nature of the density matrices used here: indeed, comparing the PDE solution in Fig. 4 and its discretizations in Fig. 5, suggests that considering too coarse discretizations might smear the solution to a point where identification is no longer possible. In such cases, the features of the original solution which could help to better identify the parameter are no longer visible in the discretized data: for instance, in Fig. 5d, the time and position where sharp changes in the PDE solution occurred (as seen in Fig. 4b) are no longer identifiable.

Then, for both the TRM and the LxF scheme, the errors on the parameter and density estimations seem to only depend on the number of space discretization steps N_x, and not on the number of time discretization steps N_t. Besides, the higher the number of space subdivisions used in the scheme, the better the parameter and density estimates are. A takeaway from these results is that the quality of the parameter and density estimations can be improved independently of the time discretization of the data, by working with fine space discretization steps and by subdividing the cells in space when using the schemes.

Besides, when comparing the schemes, one can note that the TRM systematically and significantly outperforms the LxF scheme in terms of RMSE and generally yields better parameter estimates. To understand why, we represent in Figs. 6 and 7 the finite volume approximations associated with two density matrices (respectively obtained by taking $N_x = 11$ and $N_x = 51$) when using both the TRM and the LxF scheme (with 5 space subdivisions). It can be observed, especially in Fig. 7, that the density estimates

Table 1
Relative error $|\bar{v}_m - v_m^*|/\bar{v}_m$ on the parameter estimation for various choices of discretization steps and schemes, in case where all data is used.

N_t \ N_x	5	11	21	31	51
5	0.84	0.53	0.34	0.23	0.14
11	0.85	0.54	0.38	0.25	0.16
21	0.85	0.55	0.39	0.26	0.16
31	0.85	0.55	0.39	0.26	0.17
51	0.85	0.55	0.39	0.26	0.17

(a) TRM: No Space subdivision

N_t \ N_x	5	11	21	31	51
5	1.00	0.64	0.37	0.16	0.12
11	1.00	0.63	0.34	0.14	0.11
21	1.00	0.61	0.35	0.13	0.11
31	1.00	0.20	0.39	0.15	0.11
51	1.00	0.75	0.41	0.13	0.12

(b) LxF: No Space subdivision

N_t \ N_x	5	11	21	31	51
5	0.70	0.21	0.12	0.09	0.06
11	0.72	0.22	0.14	0.10	0.07
21	0.74	0.22	0.14	0.10	0.07
31	0.74	0.22	0.15	0.10	0.07
51	0.74	0.22	0.15	0.10	0.07

(c) TRM: 3 Space subdivisions

N_t \ N_x	5	11	21	31	51
5	1.00	0.21	0.12	0.10	0.09
11	1.00	0.22	0.11	0.10	0.08
21	1.00	0.21	0.10	0.10	0.08
31	1.00	0.35	0.11	0.10	0.08
51	1.00	0.28	0.11	0.10	0.09

(d) LxF: 3 Space subdivisions

N_t \ N_x	5	11	21	31	51
5	0.46	0.13	0.09	0.06	0.04
11	0.48	0.14	0.10	0.07	0.04
21	0.49	0.14	0.10	0.07	0.04
31	0.49	0.14	0.10	0.07	0.04
51	0.50	0.14	0.10	0.07	0.04

(e) TRM: 5 Space subdivisions

N_t \ N_x	5	11	21	31	51
5	0.82	0.13	0.10	0.08	0.07
11	0.97	0.13	0.10	0.08	0.06
21	0.93	0.12	0.09	0.08	0.06
31	1.00	0.19	0.09	0.08	0.06
51	1.00	0.15	0.09	0.08	0.07

(f) LxF: 5 Space subdivisions

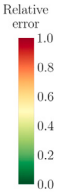


Table 2
RMSE between the density data $\boldsymbol{U}$ and the approximated densities $\hat{\boldsymbol{U}}^*$ for various choices of discretization steps and schemes, in case where all data is used.

N_t \ N_x	5	11	21	31	51
5	0.061	0.044	0.050	0.049	0.045
11	0.057	0.043	0.047	0.048	0.044
21	0.056	0.042	0.047	0.048	0.045
31	0.056	0.041	0.047	0.048	0.045
51	0.055	0.041	0.046	0.048	0.045

(a) TRM: No Space subdivision

N_t \ N_x	5	11	21	31	51
5	0.220	0.157	0.137	0.113	0.096
11	0.222	0.159	0.138	0.111	0.094
21	0.227	0.157	0.135	0.110	0.094
31	0.233	0.175	0.143	0.116	0.098
51	0.238	0.166	0.141	0.109	0.101

(b) LxF: No Space subdivision

N_t \ N_x	5	11	21	31	51
5	0.058	0.023	0.032	0.033	0.030
11	0.055	0.025	0.031	0.033	0.028
21	0.054	0.024	0.031	0.033	0.029
31	0.053	0.024	0.032	0.033	0.029
51	0.053	0.024	0.031	0.033	0.029

(c) TRM: 3 Space subdivisions

N_t \ N_x	5	11	21	31	51
5	0.184	0.104	0.085	0.075	0.062
11	0.191	0.105	0.086	0.074	0.063
21	0.190	0.104	0.089	0.074	0.064
31	0.208	0.122	0.094	0.080	0.067
51	0.223	0.113	0.096	0.074	0.069

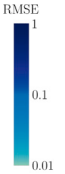
(d) LxF: 3 Space subdivisions

N_t \ N_x	5	11	21	31	51
5	0.050	0.017	0.025	0.026	0.022
11	0.048	0.019	0.025	0.026	0.021
21	0.047	0.018	0.026	0.026	0.021
31	0.047	0.018	0.026	0.027	0.022
51	0.047	0.018	0.026	0.026	0.022

(e) TRM: 5 Space subdivisions

N_t \ N_x	5	11	21	31	51
5	0.157	0.081	0.067	0.060	0.051
11	0.165	0.083	0.068	0.060	0.052
21	0.164	0.082	0.072	0.060	0.053
31	0.187	0.098	0.076	0.065	0.055
51	0.209	0.090	0.079	0.060	0.057

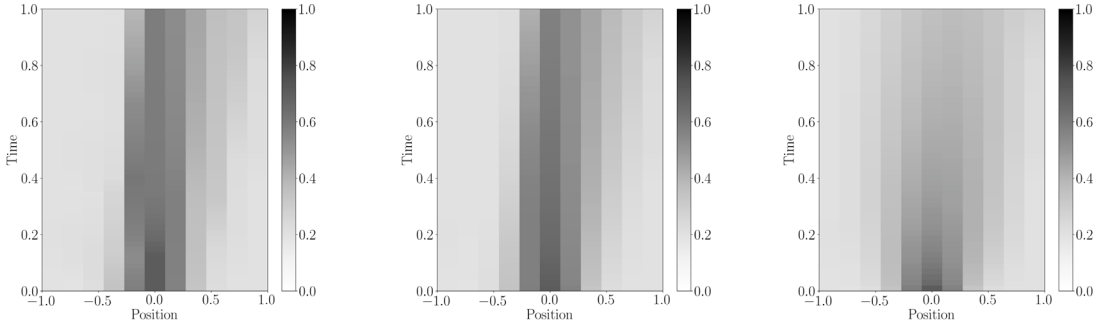
(f) LxF: 5 Space subdivisions



are smoother than the original data. The LxF scheme seems to smear the solution more than the TRM, which explains the higher RMSE on the density estimates.

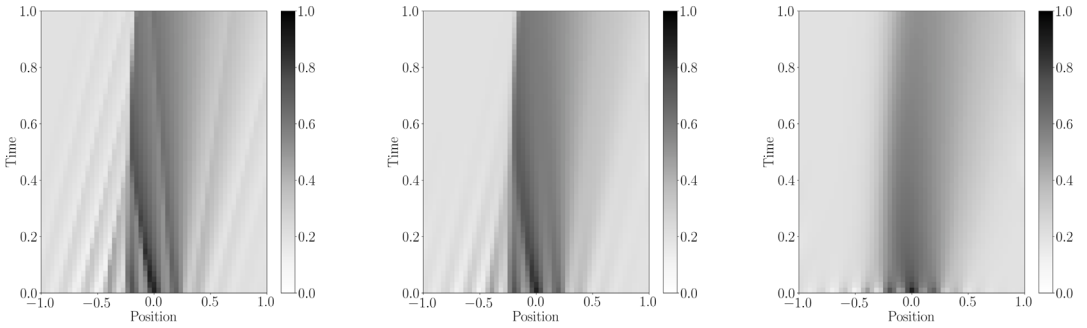
Finally, in order to test the robustness of the approximation, we consider the following approach. Starting from one of the previously formed density matrices, we "hide" some of its columns during the estimation procedure. More precisely, we carry out the parameter estimation (and density approximations) while assuming that only 3 of the density columns are known: the first and the last one (which are used to define boundary conditions) and the center column (i.e., the $((N_x - 1)/2)$th column). However, we will always assume that the 0th row of $\boldsymbol{U}$ is observed (as it is used to define the initial state of the finite volume recurrence). We then use the approach described in Section 3.4 to carry out the estimation.

The relative error between the estimated parameters and the true value $\bar{v}_m = 1$ are then given in Table 3 and the RMSE values between the associated finite volume approximations and the density matrices are given in Table 4. One observes that the TRM is still able to yield good estimates of the parameter and RMSE on the density estimates that are similar to when considering the whole density matrix. However, the LxF scheme now gives poor estimates of the parameter and high-RMSE density estimates. Hence, the TRM proves to be more robust to missing input data than the LxF scheme. From now on, only the TRM will be used as finite volume scheme. In the next section, we apply the same approach as the one used in this case study to real-world measurements of density.



(a) Density data. (b) Estimation from TRM. (c) Estimation from LxF.

Fig. 6. Estimated densities from data (a) discretized into 11 space cells and 51 time steps, using the TRM (b) and LxF (c) schemes with 5 subdivisions.



(a) Density data. (b) Estimation from TRM. (c) Estimation from LxF.

Fig. 7. Estimated densities from data (a) discretized into 51 space cells and 51 time steps, using the TRM (b) and LxF (c) schemes with 5 subdivisions.

Table 3
Relative error $|\bar{v}_m - v_m^*|/\bar{v}_m$ on the parameter estimation for various choices of discretization steps and schemes, in case where only the center road cell is observed through time.

(a) TRM: No Space subdivision

N_t \ N_x	5	11	21	31	51
5	0.89	0.69	0.46	0.29	0.05
11	0.90	0.68	0.44	0.25	0.01
21	0.90	0.67	0.43	0.25	0.00
31	0.91	0.67	0.43	0.24	0.01
51	0.91	0.67	0.42	0.24	0.01

(b) LxF: No Space subdivision

N_t \ N_x	5	11	21	31	51
5	1.00	1.00	0.95	1.00	1.00
11	1.00	1.00	1.00	1.00	1.00
21	1.00	1.00	0.90	1.00	1.00
31	1.00	1.00	1.00	0.99	1.00
51	1.00	1.00	0.91	0.99	1.00

(c) TRM: 3 Space subdivisions

N_t \ N_x	5	11	21	31	51
5	0.87	0.35	0.01	0.17	0.27
11	0.89	0.34	0.03	0.20	0.21
21	0.89	0.32	0.03	0.19	0.20
31	0.30	0.32	0.04	0.20	0.20
51	0.89	0.32	0.04	0.20	0.21

(d) LxF: 3 Space subdivisions

N_t \ N_x	5	11	21	31	51
5	1.00	1.00	1.00	1.00	1.00
11	1.00	1.00	1.00	1.00	1.00
21	1.00	1.00	0.99	1.00	1.00
31	1.00	1.00	1.00	1.00	1.00
51	1.00	1.00	0.99	1.00	1.00

(e) TRM: 5 Space subdivisions

N_t \ N_x	5	11	21	31	51
5	0.85	0.12	0.18	0.28	0.12
11	0.86	0.10	0.18	0.25	0.08
21	0.40	0.08	0.18	0.22	0.07
31	1.00	0.08	0.18	0.22	0.07
51	0.87	0.07	0.19	0.22	0.08

(f) LxF: 5 Space subdivisions

N_t \ N_x	5	11	21	31	51
5	1.00	1.00	1.00	1.00	0.44
11	1.00	1.00	1.00	1.00	0.41
21	1.00	1.00	1.00	0.99	0.36
31	1.00	1.00	1.00	1.00	0.82
51	1.00	1.00	1.00	0.14	0.05

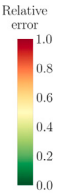


Table 4
RMSE between the density data $\boldsymbol{U}$ and the approximated densities $\hat{\boldsymbol{U}}^*$ for various choices of discretization steps and schemes, in case where only the center road cell is observed through time.

N_t \ N_x	5	11	21	31	51
5	0.062	0.049	0.052	0.049	0.045
11	0.059	0.046	0.047	0.048	0.046
21	0.057	0.044	0.047	0.048	0.047
31	0.057	0.044	0.047	0.048	0.047
51	0.056	0.043	0.047	0.048	0.047

(a) TRM: No Space subdivision

N_t \ N_x	5	11	21	31	51
5	0.220	0.157	0.138	0.118	0.106
11	0.222	0.159	0.139	0.116	0.103
21	0.227	0.157	0.136	0.115	0.103
31	0.233	0.175	0.144	0.119	0.105
51	0.238	0.166	0.142	0.114	0.108

(b) LxF: No Space subdivision

N_t \ N_x	5	11	21	31	51
5	0.061	0.027	0.034	0.044	0.053
11	0.058	0.027	0.034	0.044	0.044
21	0.057	0.026	0.035	0.044	0.043
31	0.063	0.025	0.035	0.044	0.043
51	0.056	0.025	0.035	0.044	0.043

(c) TRM: 3 Space subdivisions

N_t \ N_x	5	11	21	31	51
5	0.184	0.109	0.098	0.092	0.087
11	0.191	0.109	0.096	0.089	0.084
21	0.190	0.108	0.098	0.089	0.084
31	0.208	0.124	0.101	0.092	0.085
51	0.223	0.115	0.103	0.088	0.086

(d) LxF: 3 Space subdivisions

N_t \ N_x	5	11	21	31	51
5	0.060	0.017	0.038	0.050	0.034
11	0.057	0.019	0.037	0.045	0.028
21	0.048	0.019	0.037	0.041	0.027
31	0.067	0.019	0.037	0.041	0.027
51	0.055	0.019	0.037	0.041	0.027

(e) TRM: 5 Space subdivisions

N_t \ N_x	5	11	21	31	51
5	0.157	0.093	0.088	0.086	0.059
11	0.165	0.092	0.085	0.082	0.058
21	0.164	0.091	0.086	0.081	0.057
31	0.187	0.103	0.089	0.083	0.073
51	0.209	0.096	0.091	0.060	0.058

(f) LxF: 5 Space subdivisions

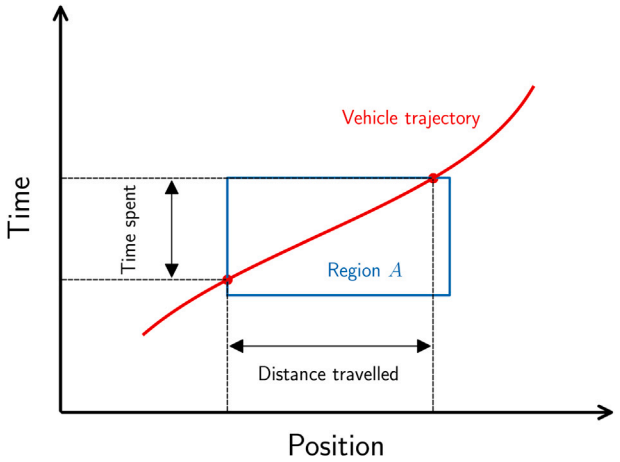


Fig. 8. Quantities used in Edie's generalized definition of density and flow.

4.2. Mimicking real traffic dynamics

4.2.1. Generalized density data

Most tools designed to measure traffic flows and densities are based on counting the number of vehicles passing a given point of the road or present in a given section of the road. The resulting density measurements are then essentially discrete since they depend on these discrete count variables. However, when trajectory data is available, Edie (1963) suggests a generalization of this idea that leverages the continuity of trajectories (in space and time) to yield a continuous estimation of the density. Take a region A of the space–time domain on which the vehicle trajectories lie. The density of vehicles in A is defined as the ratio between the time spent by all the vehicles in A by the area of A. Similarly, the flow of vehicles in A is defined as the ratio between the distance traveled by all the vehicles in A by the area of A. Both quantities can be computed for each vehicle whose trajectory intersects A, using the definition represented in Fig. 8.

Using this definition of a density measurement, it is possible to build a matrix of density measurements $\boldsymbol{D}$ containing the densities associated with a space–time discretization of the domain on which the trajectories lie. Indeed, we discretize this domain into a grid composed of N_x cells in the space dimension and N_t cells in the time dimension (see Fig. 9). Then, $\boldsymbol{D}$ is built as the matrix of size $N_t \times N_x$ for which the (i,j)th entry, denoted by D^i_j, is the estimated density of the (i,j)th grid cell, as obtained by applying Edie's definition on the space–time region defined by the cell.

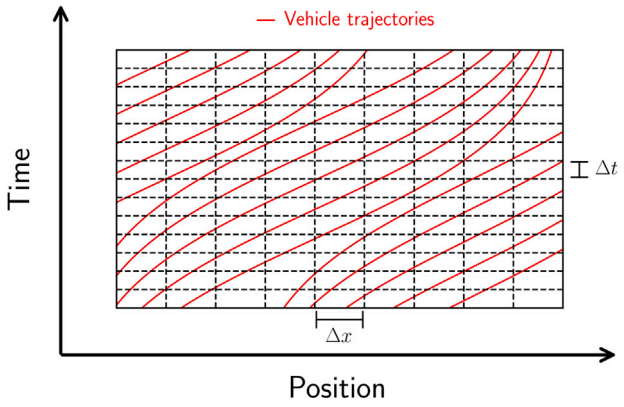
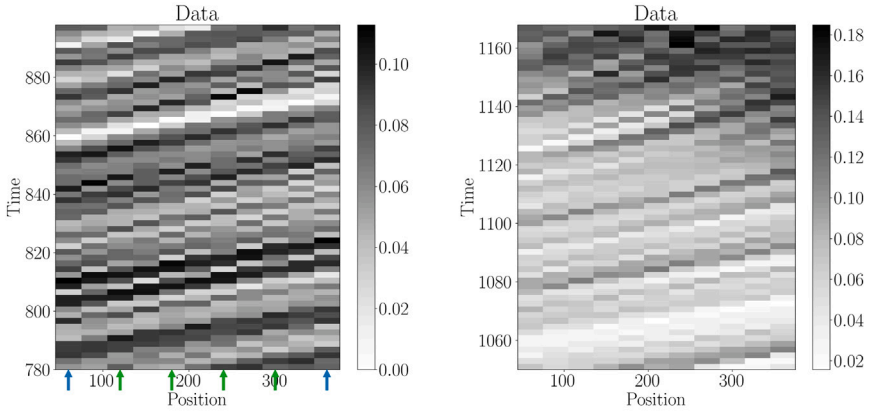


Fig. 9. Space–time discretization grid used to compute the density matrix from Edie's definition, which is applied to each cell of the grid.



(a) Dataset 1. (b) Dataset 2.

Fig. 10. Representation of the two datasets used in the real-data case.

The resulting densities D_j^i are assimilated to a ratio between some kind of continuous count of vehicles (given by the ratio between the total time spent by all vehicles the region A and the time width of A) and the size of the road cells. As such, they can be considered as approximations of the cell average ρ_j^i of the density function ρ, which is expressed as

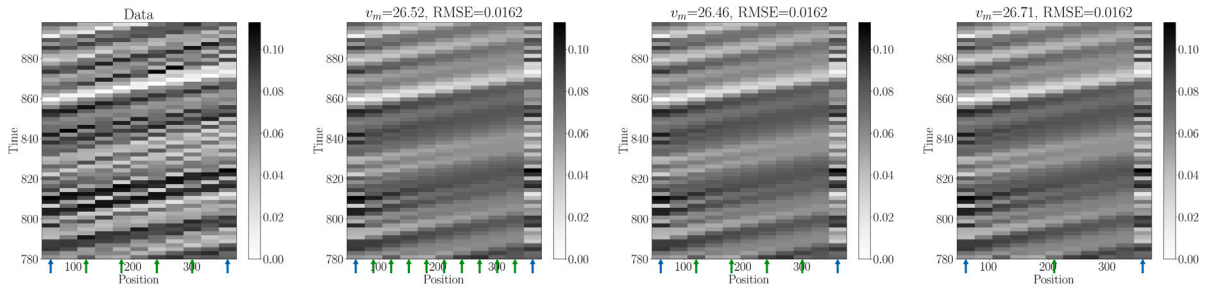
$$D_j^i \approx \rho_j^i = \frac{1}{\Delta x} \int_{x_j - \Delta x/2}^{x_j + \Delta x/2} \rho(t_i, x)\, \mathrm{d}x \ .$$

where (t_i, x_j) are the coordinates of the center the (i, j)th cell of the grid introduced above. The right hand side of this last equality links the cell averages of the normalized density u to the density data D_j^i: indeed, dividing both sides of this equation by ρ_m yields that the cell average U_j^i of the normalized density is approximated by the ratio D_j^i/ρ_m.

The subsequent numerical experiments use trajectory data extracted from the highD dataset. This dataset results from video recordings of sections of German highways made by drones, and from which the trajectory of all the vehicles, or rather of their center of gravity, were extracted using image analysis tools. Hence, we apply Edie's generalized definitions of density (and flux) to the set of trajectory of the center of gravity of the vehicles, which are therefore reduced to points.

We consider a particular section of the studied road with length of about 400 m, which we discretize into 11 road cells (with size $\Delta x \approx 36$ m). Based on this, we build two density matrices, which we call Dataset 1 and Dataset 2, both corresponding to observations of the section over a period of 2 min, so that Dataset 1 reflects free flow conditions only, and Dataset 2 reflects a transition between free flow conditions and a congested state. The time step used to build these density matrices is $\Delta t = 2$ s. The resulting density matrices have $N_t = 60$ rows and $N_x = 11$ columns and can be observed in Fig. 10. Finally we estimate the maximal density of the considered road by dividing the number of lanes by the mean length of the observed vehicles, which gives $\rho_m \approx 0.49\ \mathrm{m}^{-1}$.

Remark 4.1. Note that this macroscopic evaluation of the density, obtained using Edie's generalized definition of density, relies on the trajectories followed by the center of gravity of the vehicles. Hence, in this context, each vehicle is reduced to a point and



(a) Dataset 1. (b) Using all the dataset. (c) Using 4 columns. (d) Using 1 column.

Fig. 11. Estimated densities using different subsets of the Dataset 1 (and the TRM), and associated estimated parameter v_m and RMSE. The blue arrows point to the columns used for the boundary conditions and the green arrows to the ones used in the minimization problem. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tracking these points through space and time allows us to derive a density measurement that is no longer comparable to a mere count of vehicles on a cell, but has a rather macroscopic interpretation.

Besides, this measure of density can still be linked to more usual ones. Indeed, as explained by Cassidy and Coifman (1997), the density over a space–time region $A = \Delta x \times \Delta t$ as defined by Edie is proportional to a measure of occupancy made by a sensor of size Δx during a time interval Δt (the proportionality constant being the average of the length of the vehicles crossing the region A). Hence, the density measurements used in the subsequent applications can naturally be seen as resulting from occupancy sensors paving the roads.

4.2.2. Constant parameter case

We start by testing our approach for parameter identification (and density estimation) on the Datasets 1 and 2 introduced in the previous subsection, under the assumption that the speed parameter v_m is constant.

For each density matrix, we compute the parameter and density estimate using the TRM scheme with the following discretization choice: the space cells are subdivided into $P_x = 3$ subcells and the number of time subdivisions P_t is taken to be the smallest integer so that the CFL condition (28) is satisfied for some rough estimate of the maximal speed $v_m = 130$ km h^{-1}, thus giving $P_t = 15$. Once again, the first and last columns as well as the first row of the density matrix are used as boundary and initial conditions for the finite volume recurrences. Then, three cases are considered: either the whole density matrix is used and hence the cost function (31) with $I_c = \{1, \dots, 9\}$ is minimized, or half of the columns are used and hence the cost function (31) is minimized but with $I_c = \{2, 4, 6, 8\}$, or only one column is used and hence the cost function (31) is minimized with $I_c = \{5\}$.

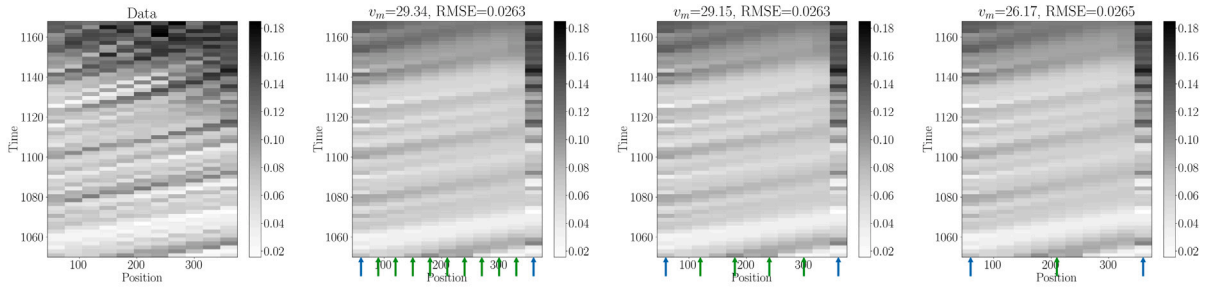
The results of these estimations are represented in Fig. 11 for Dataset 1 and in Fig. 12 for Dataset 2. In both cases we can once again notice the robustness of the estimation since removing some columns from the dataset does not affect significantly the value of the estimated parameter or the RMSE of the estimated densities. Besides, one can note that TRM seems to smooth the true evolution of the densities. When comparing the results obtained for both datasets, the RMSE for Dataset 1 is significantly lower than that of Dataset 2, which can be explained by comparing visually the estimated densities in both cases.

For Dataset 1, the TRM was able to recreate the linear trends of density values appearing in the dataset and that are characteristic of free flow conditions: indeed, in this case, the vehicles are able to travel freely across the road and hence the vehicles can transfer from one cell to the next undisturbed. Therefore, modeling this vehicle transfer with a unique and constant reaction rate, as the TRM does, seems appropriate. For Dataset 2, however, congestion appears in the dataset and hence there is a change in the conditions with which vehicles can transfer from one cell to the other. A single reaction rate becomes now a more controversial choice, which is confirmed by the fact that the estimated densities do not depict the same congestion as in the data.

Another way to understand the difference of approximation quality between both datasets is to compare their fundamental diagrams. A fundamental diagram is a scatter-plot representing density measurements against flux measurements done at the same time and space locations. In our case, flux measurements associated to our datasets can be computed from the trajectory data using once gain the approach in Section 4.2.1. As for the flux "measurements" associated with the estimated densities, we use the Greenshields flux–density relationship (2) assumed by the LWR model, and plug in the estimated parameter v_m^*. The resulting fundamental diagrams are shown in Fig. 13. For Dataset 1, the true fundamental diagram looks quite linear, as expected for free flow conditions, and the quadratic relation of the Greenshields flux of the estimation process then yields an adequate approximation. However, for Dataset 2, the Greenshields flux fails to give a good approximation of the true fundamental diagram, which now shows a mix of linear trend and more diffuse point pattern. In order to improve these estimations, we propose to offer more flexibility to the models by adding new (and physically meaningful) parameters. This is the purpose of the next section.

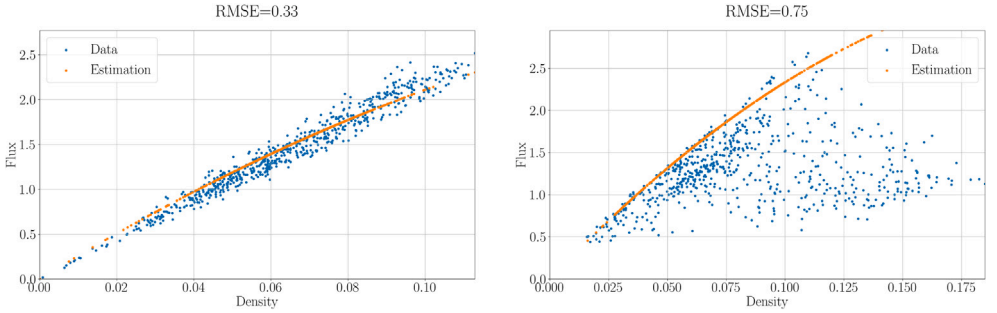
4.2.3. Varying parameter case

Starting from the two datasets introduced in Section 4.2.2, we use the same least-square minimization approach to derive the values of the now varying parameter v_m. We consider three cases:



(a) Dataset 2. (b) Using all the dataset. (c) Using 4 columns. (d) Using 1 column.

Fig. 12. Estimated densities using different subsets of the Dataset 2 (and the TRM), and associated estimated parameter v_m and RMSE. The blue arrows point to the columns used for the boundary conditions and the green arrows to the ones used in the minimization problem. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Dataset 1. (b) Dataset 2.

Fig. 13. Comparison between the fundamental diagram of the two datasets and their estimations through the TRM.

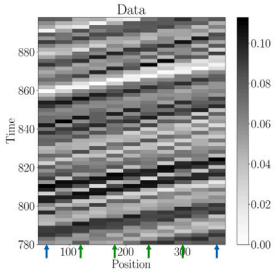
- the parameter v_m is time dependent only, meaning that the scaling parameters of the finite volume recurrence satisfy that for any $n \in [\![0, N_t - 1]\!]$, there exists $C^n \in (0, 1/2)$ such that for any $j \in [\![0, N_x]\!]$, $C^n_j = C^n$. Hence, the actual number of parameters to be estimated in this case is N_t.
- the parameter v_m is space dependent only, meaning that the scaling parameters of the finite volume recurrence satisfy that for any $j \in [\![0, N_x]\!]$, there exists $C^n \in (0, 1/2)$ such that for any $n \in [\![0, N_t - 1]\!]$, $C^n_j = C_j$. Hence, the actual number of parameters to be estimated in this case is N_x.
- the parameter v_m is space–time dependent, and hence, the actual number of parameters to be estimated in this case is $(N_x+1)N_t$.

The estimations are carried out while considering half of the columns of the density matrices (hence $I_c = \{2, 4, 6, 8\}$) and the parameter λ are set so that the overall RMSE between the estimated densities and the whole density matrix is minimized. A TRM with 3 space subdivisions (and 15 time subdivisions) is used as a finite volume scheme, which is the scheme used for the robustness study in the constant case (cf. Section 4.2.2, results in Figs. 11 and 12). The results are presented in Fig. 14 for the time dependent case, Fig. 15 for the space dependent case and Fig. 16 for the space–time dependent case.

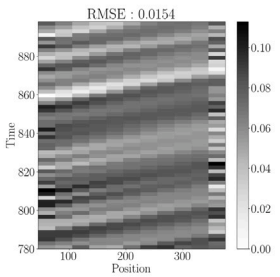
Considering first the results for Dataset 1 (free flow conditions), we can see that the varying parameters estimated in each case stay close and vary around the value estimated under the assumption that the parameter is constant. This is coherent with the conclusions drawn in Section 4.2.2: in free flow conditions, the LWR model with a constant parameter is an adequate choice of model. Comparing now the RMSE of the estimated densities, we see that the time dependent and space dependent parameters yield very similar values and density profiles as in the constant case (cf. Fig. 11). However, a significant decrease of the RMSE is observed when a space–time dependent parameter is considered. Hence, adding small perturbations of the parameters in space and time seems to yield more realistic density profiles and in particular the small scale variations of the density that are not observed in the constant case (due to the smoothness of the estimation).

Considering now the results for Dataset 2 (free flow and congested conditions), we can see that the varying parameters estimated in each case do not stick around the value estimated in the constant case anymore.

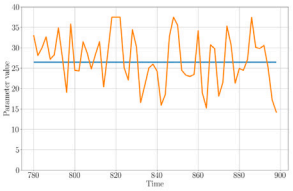
In the time dependent case, we observe two regimes. The first regime spans until $t = 1100$ s, and has the parameter varying around and close to the parameter estimated in the constant case, thus hinting at free flow conditions. The second regime starts at $t = 1100$ s and has the parameter displaying sharp variations between very low and very large values: such behavior can be interpreted as the model trying to accommodate the congested conditions by intermittently stopping or letting all the vehicles go in order to create



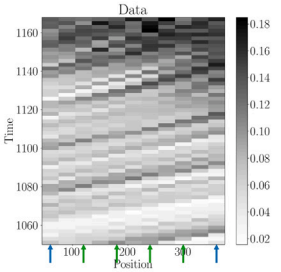
(a) Dataset 1.



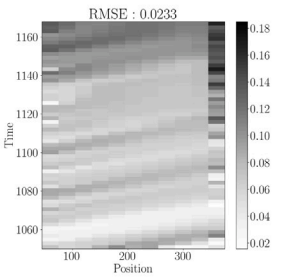
(b) Estimated densities for Dataset 1.



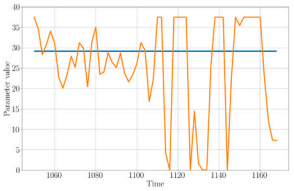
(c) Estimated parameters for Dataset 1.



(d) Dataset 2.

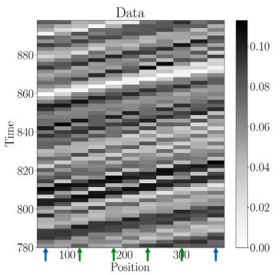


(e) Estimated densities for Dataset 2.

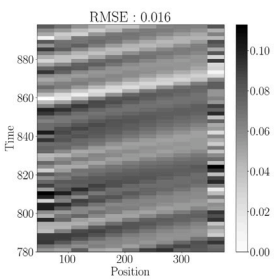


(f) Estimated parameters for Dataset 2.

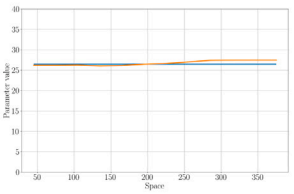
Fig. 14. Estimated densities and parameters for each dataset when time dependent parameters are considered. The plots (c) and (f) represent the evolution through time of the estimated parameter. The blue line represents the estimated parameter in the constant case. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



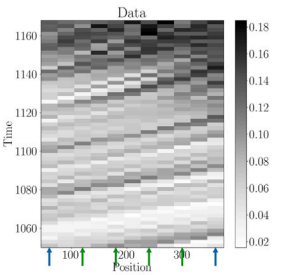
(a) Dataset 1.



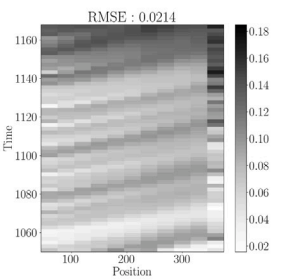
(b) Estimated densities for Dataset 1.



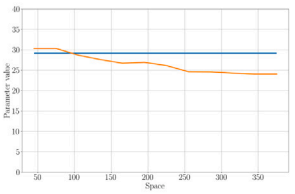
(c) Estimated parameters for Dataset 1.



(d) Dataset 2.



(e) Estimated densities for Dataset 2.



(f) Estimated parameters for Dataset 2.

Fig. 15. Estimated densities and parameters for each dataset when space dependent parameters are considered. The plots (c) and (f) represent the evolution in space of the estimated parameter. The blue line represents the estimated parameter in the constant case. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

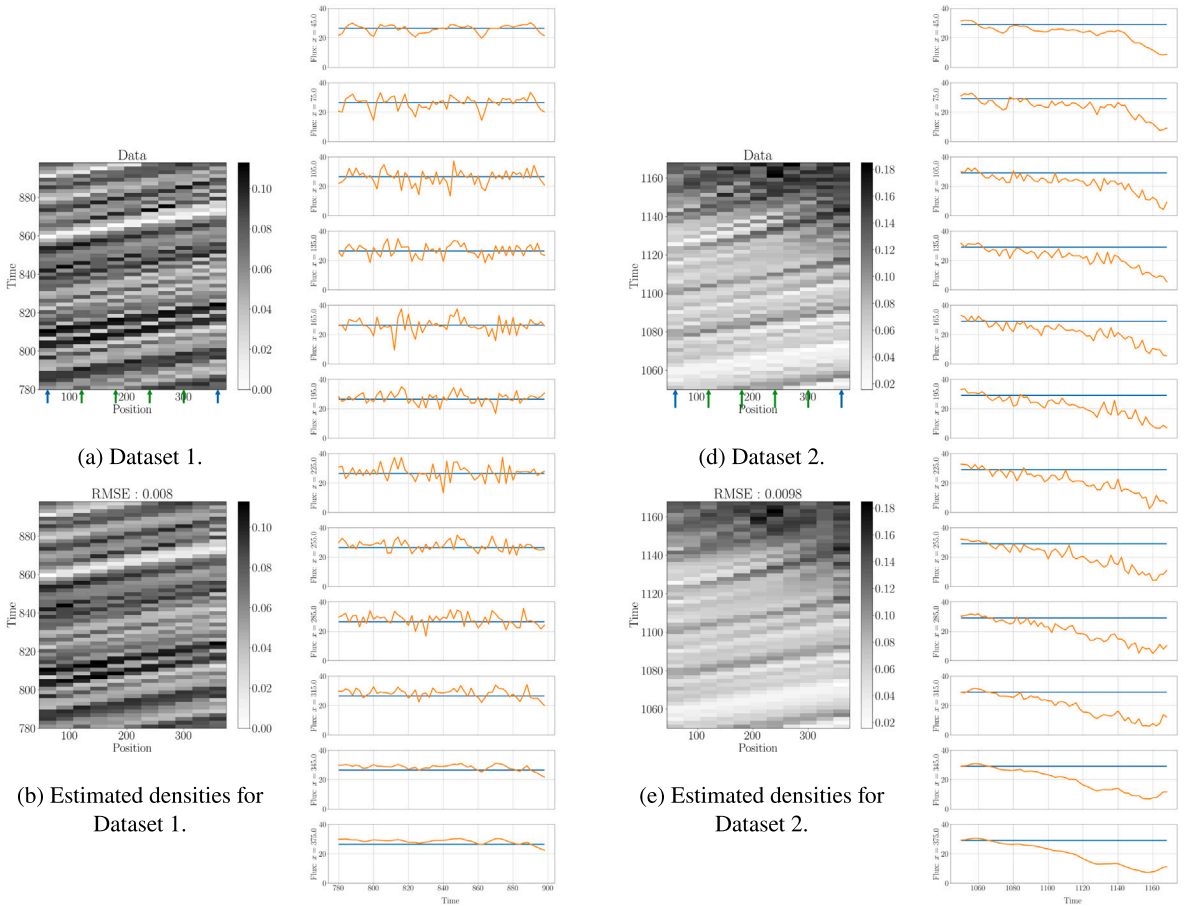


Fig. 16. Estimated densities and parameters for each dataset when space–time dependent parameters are considered. The plots (c) and (f) represent, for each location, the evolution through time of the estimated parameter. The blue line represents the estimated parameter in the constant case. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

congested cells. In the space dependent case, the estimated densities globally decrease across space: this can be seen as an attempt from the model to create congestion by having cells with higher transfer rates upstream (which will tend to let vehicles flow easily), and then gradually decreasing these rates as we go down the road, so that vehicles can accumulate downstream. In both these cases however, the resulting RMSE of the estimated densities is lower but still of the same order as the one from the constant case.

In the space–time dependent case, the estimated parameters at all locations show the same trend: they start close to the value estimated in the constant case and after some time globally decrease with time. Besides, this drop in parameter value occurs at increasing times as we go from the right-most cell to the left-most cell. Hence, the model seems to account for congestion by gradually reducing the transfer rates between the compartments, going from right to left. In this case, the resulting RMSE of the estimated densities is significantly reduced compared to the constant case and the estimated densities display a realistic profile, which also recreates the congestion observed in the data.

Note that, following the link established between the reaction rates and the parameters of the continuous traffic flow models, the gradual decrease of reaction rates observed for Dataset 2 in the time dependent and space–time dependent cases can be interpreted as a gradual drop in road capacity. This observation is corroborated by looking at the actual trajectories corresponding to this dataset and shown in Fig. 17. Indeed, overtaking between vehicles can be observed from trajectory crossings. These overtakings mechanically decrease the overall capacity of the road as less lanes are free. As on can see, these overtaking happen more and more frequently as time passes, and start to appear downhill on the road. The same observations were made when looking at the space–time dependent reaction rates.

Finally, we compare the three choices of parametrizations considered in this section in terms of their ability to recreate a fundamental diagram similar to the one associated with the density data. In particular, flux estimates can be derived from the density estimates by once again applying the Greenshields flux–density relation (2), but using now the varying parameter v_m: to

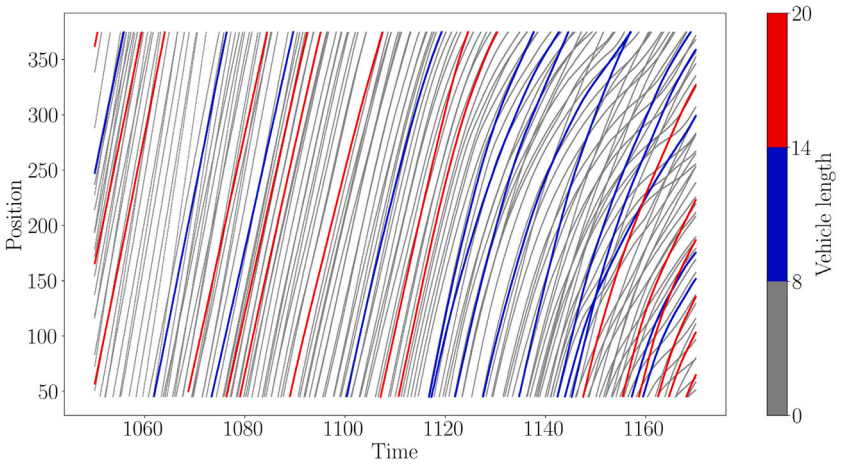


Fig. 17. Trajectories of Dataset 2. Each line corresponds to the trajectory of a given vehicle, the line color is linked to the length of the vehicle. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

compute the flux estimate of the jth cell at time t_n, v_m is taken as the average of the parameter estimates at both boundaries of the jth cell, at time t_n. We obtain the fundamental diagrams shown in Fig. 18. The time-dependent estimates show for both datasets a fundamental diagram which is more scattered than the one from the data, and yield higher RMSE than in the constant case.

On the other hand, the space-dependent estimates produce fundamental diagrams that are similar to the one obtained in the constant case, but with slightly more dispersion. In both cases however, the fundamental diagrams consist in superposition of quadratic functions and seem to fail to reproduce the scattering observed in Dataset 2 (and due to congested conditions). This goal is however achieved with the space–time-dependent estimates which yield fundamental diagrams that nicely overlap the ones from the data, and significantly lower RMSE compared to the constant case.

In conclusion, the recourse to space–time dependent parameters provides more flexibility to the LWR model in a physically sound manner, thus allowing it to recreate real-word density and fundamental diagrams:

- On the one hand, allowing the reaction rates between compartments in the TRM to vary in space and time locally creates conditions that give rise to congestion or sharp changes in the density.
- On the other hand, realistic fundamental diagrams are obtained even though a quadratic relation between flux and density is assumed, by allowing the shape of the relation to change over space and time. Hence, the change of behavior in the fundamental diagram usually interpreted as a capacity drop now becomes a transfer rate drop. Besides, complex point patterns in the diagram can be recreated since in theory each point of the diagram belongs to its own quadratic function.

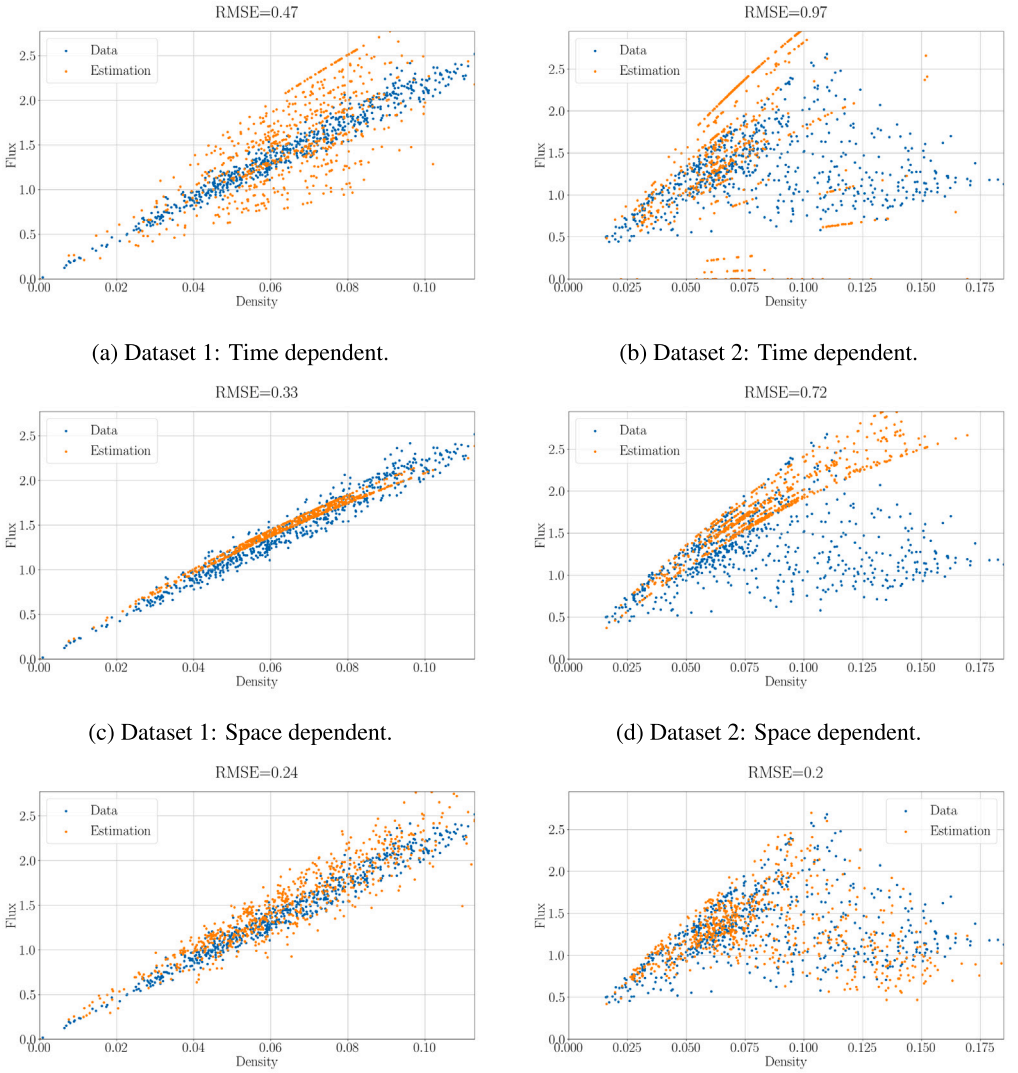
In this work, we restricted our approach to considering (a discretization) of a macroscopic traffic flow model defined from a quadratic fundamental diagram. This might seem like a restriction, especially there seems to be a consensus towards considering piecewise linear flux functions (consisting of a free-flow branch and congested branch) for traffic modeling applications. But it is not. On the one hand, note that the TRM can be generalized to more complex flux–density relationships. Indeed, Lipták et al. (2021) showed that the TRM preserves all its properties (monotonicity, stability, convergence) for any model for which the flux function f can be written as

$$f(\rho) = f_1(\rho) f_2(\rho), \quad \rho \in [0, \rho_m],$$

where ρ_m is the maximal density, f_1 is a non-decreasing Lipschitz-continuous function such that $f_1(0) = 0$, and f_2 is a non-increasing Lipschitz-continuous function[4] such that $f_2(\rho_m) = 0$. Then, by choosing adequately the functions f_1 and f_2 one could retrieve trapezoidal fundamental diagrams (cf. Appendix E).

On the other hand, one should keep in mind that the advantages of considering a quadratic flux are the reduced amount of parametrization needed (only one, the free flow speed) and the fact that the resulting flux function has continuous (and polynomial) derivatives which allows us to readily use gradient-based optimization in our approach. Besides, our numerical experiments prove that even with a simple quadratic flux function, we are able to achieve good accuracy by modifying the underlying macroscopic model itself (by considering space and/or time dependent parameters) rather than trying to find the best choice of flux functions that might fit the fundamental diagram.

[4] In their work, Lipták et al. (2021) actually assume that f_1 and f_2 are continuously derivable, but their results readily extend to the case where these functions are (only) Lipschitz-continuous. Indeed, the only property of continuously derivable functions that they use is the fact that they admit a Lipschitz constant.



(a) Dataset 1: Time dependent.

(b) Dataset 2: Time dependent.

(c) Dataset 1: Space dependent.

(d) Dataset 2: Space dependent.

(e) Dataset 1: Space-time dependent.

(f) Dataset 2: Space-time dependent.

Fig. 18. Comparison between the fundamental diagram of the two datasets and their estimations through the TRM with time dependent, space dependent and space–time dependent parameters.

5. Conclusion

The main motivation of this work is to assess the validity of a LWR traffic flow model to model measurements obtained from trajectory data, and propose extensions of this model to improve it. We answer these questions by comparing continuous models and measurements using a discrete dynamical system defined from a particular discretization of the PDE of the continuous model. This discretization is formulated as a chemical reaction network where road cells are interpreted as compartments, the transfer of vehicles from one cell to the other is seen as a chemical reaction between adjacent compartment and the density of vehicles is seen as a concentration of reactant. Several degrees of flexibility on the parameters of this system, which basically consist of the reaction rates between the compartments, are considered: These rates are taken equal to the same constant value or allowed to depend on time and/or space. We then interpret generalized density measurements coming from trajectory data as observations of the states of the discrete dynamical system at consecutive times, and derive optimal reaction rates for the system by minimizing the discrepancy between the output of the system and the state measurements.

The use of constant reaction rates proves to be enough to reproduce the patterns observed in the density and flux data in free flow conditions but not in mixed conditions where congestion appears. This motivates us to recommend the use of the more flexible models, and in particular the model with space–time dependent reaction rates. This last model proved to perform well both in free

flow and mixed conditions as it mimicked the patterns observed in the density data as well as the fundamental diagrams. Recall that the discrete dynamical system can be seen as a particular finite volume discretization of the LWR model with the flux of vehicles depending quadratically on the density (Greenshields flux). The reaction rates of the system then simply set the shape of this relation (meaning here the maximal value of the flux function). Our numerical experiments hence showed that allowing the shape of this quadratic relation to vary through time, space or even better both, allowed the LWR model to better recreate specific patterns observed in real-world data, such as the appearance of congestion (compared to when a fixed shape is considered).

Direct extensions of the approach presented in this paper are possible. On the one hand, working with situations where on and off ramps are present would be straightforward since the TRM was actually formulated in this setting (Lipták et al., 2021). Similarly, working on networks would be straightforward since the proposed kinetic system can be generalized to this setting by simply dropping the assumptions that the compartments are ordered as chain (which makes sense for a single road) and allowing them to be linked to more than 2 other compartments (thus mimicking the junctions of the network). On the other hand, the use of conventional detector data could be considered, since it would simply come down to the assumption that measurements are only available in some compartments (those where sensors are located), similarly as what was assumed in the robustness tests done in Section 4.1. Finally, a link between the proposed discrete dynamical system and recurrent neural networks was not exploited in this paper but paves the way to exciting outlooks. For instance, ongoing work is undertaken to propose an algorithm for short-term traffic prediction which builds on the ideas presented in this paper. In particular, the fact the actual dynamics of traffic can be uncovered serves as a basis of this upcoming algorithm.

CRediT authorship contribution statement

Mike Pereira: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Pinar Boyraz Baykas:** Conceptualization, Software. **Balázs Kulcsár:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Annika Lang:** Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

The authors thank G. Szederkenyi and Gy. Lipták for the inspiring discussions on the TRM, but also the anonymous reviewers for their helpful comments and their interest in our work. The authors thank for the financial support of the Chalmers AI Research Centre (CHAIR) under the projects STONE, RITE, and SCNN and Transport Area of Advance (Chalmers University of Technology). This work has been partially supported and funded by OPNET (Swedish Energy Agency, 46365-1), the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation, and the Swedish Research Council (project nr. 2020-04170).

Appendix A. Some examples of finite volume schemes

The Lax–Friedrich (LxF) scheme is defined for the choice of numerical flux $F = F_L$ with

$$F_L(u, v\,;\, v_m) = \frac{f(u\,;\, v_m) + f(v\,;\, v_m)}{2} + \frac{\Delta x}{2\Delta t}(u - v)\,.$$

The Godunov scheme is given by the choice $F = F_G$ with

$$F_G(u, v\,;\, v_m) = \begin{cases} \min\limits_{w\in[u,v]} f(w\,;\, v_m) & \text{if } u \le v \\ \max\limits_{w\in[v,u]} f(w\,;\, v_m) & \text{if } v \le u \end{cases}.$$

In the particular case where f is defined by (2), note that the recurrence relation of the Godunov scheme can be rewritten as

$$\widehat{U}_j^{i+1} = \widehat{U}_j^i + \frac{\Delta t}{\Delta x} v_m \left[\tilde{F}_G(\widehat{U}_{j-1}^i, \widehat{U}_j^i) - \tilde{F}_G(\widehat{U}_j^i, \widehat{U}_{j+1}^i)\right], \quad j \in \mathbb{Z}, i \in \mathbb{N}\,, \tag{32}$$

where $\tilde{F}_G$ is a normalized numerical flux (in the sense that it does not depend on the parameter v_m anymore) given by

$$\tilde{F}_G(u, v) = \begin{cases} \min\limits_{w\in[u,v]} w(1-w) & \text{if } u \le v \\ \max\limits_{w\in[v,u]} w(1-w) & \text{if } v \le u \end{cases}.$$

Similarly, for the Lax–Friedrichs scheme, we can write

$$\widehat{U}_j^{i+1} = \frac{\widehat{U}_{j-1}^i + \widehat{U}_{j+1}^i}{2} + \frac{\Delta t}{\Delta x} v_m \left[\tilde{F}_L(\widehat{U}_{j-1}^i, \widehat{U}_j^i) - \tilde{F}_L(\widehat{U}_j^i, \widehat{U}_{j+1}^i)\right], \quad j \in \mathbb{Z}, i \in \mathbb{N} \tag{33}$$

for the normalized numerical flux $\tilde{F}_L$ defined by

$$\tilde{F}_L(u, v) = \frac{u(1-u) + v(1-v)}{2}\,.$$

Appendix B. Minimization problems in the multilevel approach

On the one hand, in the constant parameter case, the minimization problem can be reformulated as

$$\theta^* = \underset{\theta\in\mathbb{R}}{\operatorname{argmin}} L_{[P_t,P_x]}(\theta)\,,$$

where the cost function $L_{[P_t,P_x]}$ is now defined by

$$L_{[P_t,P_x]}(\theta) = \frac{1}{2}\sum_{i=1}^{N_t-1}\sum_{j=1}^{N_x-2}\left(\frac{1}{P_x}\sum_{k=0}^{P_x-1}\hat{U}^{iP_t}_{k+jP_x}(C(\theta)) - U^i_j\right)^2, \quad \theta\in\mathbb{R}\,, \tag{34}$$

with the same mapping $C:\mathbb{R}\to(0,1/2)$ defined by (23). The optimal value υ^*_m of the parameter of PDE (4) is obtained by

$$\upsilon^*_m = \frac{\widehat{\Delta x}}{\widehat{\Delta t}}C(\theta^*) = \frac{P_t}{P_x}\frac{\Delta x}{\Delta t}C(\theta^*)\,.$$

On the other hand, in the varying parameter case, we adopt the following changes:

- the boundary and initial conditions are set in the same way;
- the recurrence relation of the scheme, now defined on the subdivided grid, takes the form

$$\hat{U}^{m+1}_k = \hat{U}^m_k + \hat{C}^m_k\hat{U}^m_{k-1}\left(1-\hat{U}^m_k\right) - \hat{C}^m_{k+1}\hat{U}^m_k\left(1-\hat{U}^m_{k+1}\right), \quad k\in[\![P_x,\,(N_x-1)P_x-1]\!], \quad m\in[\![0,\,P_t(N_t-1)]\!]\,,$$

  where the coefficients $\{\hat{C}^m_k : k\in[\![P_x,(N_x-1)P_x]\!], m\in[\![0,P_t(N_t-1)]\!]\}$ are defined through a bilinear interpolation (in space and time) of the coefficients $\boldsymbol{C}=\{C^n_j : j\in[\![0,N_x]\!], n\in[\![0,N_t-1]\!]\}$ defined in the case where no subdivision is introduced. In particular, for $j\in[\![0,N_x-1]\!]$, $n\in[\![0,N_t-2]\!]$, we have:

$$\hat{C}^{l+nP_t}_{q+jP_x} = \begin{pmatrix}1-l/P_t\\ l/P_t\end{pmatrix}^T\begin{pmatrix}C^n_j & C^n_{j+1}\\ C^{n+1}_j & C^{n+1}_{j+1}\end{pmatrix}\begin{pmatrix}1-q/P_x\\ q/P_x\end{pmatrix}, \quad q\in[\![0,P_x-1]\!], \quad l\in[\![0,P_t]\!]\,;$$

- the coefficients $\boldsymbol{C}=\{C^n_j : j\in[\![0,N_x]\!], n\in[\![0,N_t-1]\!]\}$ are determined by minimizing (without constraints) a cost function $L_{[P_t,P_x]}$ given as the sum of a least-square cost and a regularization term $R(\boldsymbol{C})$:

$$L_{[P_t,P_x]}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{N_t-1}\sum_{j_c\in I_c}\left(\frac{1}{P_x}\sum_{k=0}^{P_x-1}\hat{U}^{iP_t}_{k+j_cP_x}(\boldsymbol{C}(\boldsymbol{\theta})) - U^i_{j_c}\right)^2 + \lambda R(\boldsymbol{C}(\boldsymbol{\theta})), \quad \boldsymbol{\theta}\in\mathbb{R}^{(N_x+1)N_t}\,,$$

  where $\boldsymbol{C}(\boldsymbol{\theta})\in(0,1/2)^{(N_x+1)N_t}$ is obtained by applying (23) to each entry of $\boldsymbol{\theta}\in\mathbb{R}^{(N_x+1)N_t}$, I_c denotes the set of observed columns in the density matrix (excluding the boundary columns) and $\lambda>0$ is a hyperparameter weighting the least-square minimization of the regularization.

In both cases, the minimization can once gain be tackled using the conjugate gradient algorithm, since the same rules can be applied to derive an explicit expression of the gradient of the cost function (cf. Appendix C).

Appendix C. Gradient computation for cost minimization

Let L denote the least-square cost function defined by (25). Starting then from the recurrence relation (20), we see that any $i\in[\![1,N_t]\!]$, $j\in[\![1,N_x-2]\!]$, the finite volume approximation $\hat{U}^i_j(C)$ can be expressed as a composition of the functions $\mathcal{H}^k_l$, for $k<i$ and $l\in[\![1,N_x-2]\!]$. Assuming that the functions h and $\tilde{F}$ are smooth with respect to their arguments, the gradient ∇L of L with respect to θ can actually be computed using the chain rule of derivation, giving the expression given in the next proposition.

Proposition C.1. *Let L be the cost function defined in (25) and assume that the mappings $\mathcal{H}^i$, $i\geq 0$, defined through (21) and (20) are smooth with respect to their arguments. Then, the gradient of L is given by*

$$\frac{\partial L}{\partial\theta}(\theta) = \frac{\partial C}{\partial\theta}(\theta)\cdot\sum_{i=1}^{N_t-1}\operatorname{Jac}_{\hat{U}^i}(C(\theta))^T\left(\hat{\boldsymbol{U}}^i(C(\theta)) - \boldsymbol{U}^i\right), \quad \theta\in\mathbb{R}\,, \tag{35}$$

where

$$\frac{\partial C}{\partial\theta}(\theta) = \frac{1}{2}\operatorname{lgt}(\theta)(1-\operatorname{lgt}(\theta)), \quad \theta\in\mathbb{R}\,,$$

and for $i\in[\![1,N_t-1]\!]$, and $C\in(0,1/2)$, $\operatorname{Jac}_{\hat{U}^i}(C)\in\mathbb{R}^{N_x\times N_C}$ denotes the Jacobian matrix of the mapping $C\mapsto\hat{\boldsymbol{U}}^i(C)$, which can be computed through the recurrence relation

$$\begin{cases}\operatorname{Jac}_{\hat{U}^0}(C) = \mathbf{0}\\ \operatorname{Jac}_{\hat{U}^{i+1}}(C) = \operatorname{Jac}_{\mathcal{H}^i}(\hat{\boldsymbol{U}}^i)\cdot\operatorname{Jac}_{\hat{U}^i}(C) + \operatorname{Jac}_{\mathcal{H}^i}(C), \quad i\geq 0\end{cases} \tag{36}$$

with $\mathrm{Jac}_{\mathcal{H}^i}(\widehat{\boldsymbol{U}}^i) \in \mathbb{R}^{N_x \times N_x}$ *being the Jacobian matrix of the mapping* $\widehat{\boldsymbol{U}}^i \mapsto \mathcal{H}^i(\widehat{\boldsymbol{U}}^i, C\,;\,\boldsymbol{U})$ *and* $\mathrm{Jac}_{\mathcal{H}^i}(C) \in \mathbb{R}^{N_x \times N_C}$ *being the Jacobian matrix of the mapping* $C \mapsto \mathcal{H}^i(\widehat{\boldsymbol{U}}^i, C\,;\,\boldsymbol{U})$.

Proof. Applying the chain rule to (25) yields exactly (35). Then, applying the chain rule to (20) gives the recurrence relation in (36). The fact that $\mathrm{Jac}_{\boldsymbol{U}^0}(C) = \boldsymbol{0}$ follows from the fact that $\boldsymbol{U}^0$ does not depend on C (but is defined in (17) using the data $\boldsymbol{U}$). □

The explicit expression of the Jacobian matrices $\mathrm{Jac}_{\mathcal{H}^i}(\widehat{\boldsymbol{U}}^i)$ and $\mathrm{Jac}_{\mathcal{H}^i}(\boldsymbol{C})$ in Proposition C.1 depends on the choice of the numerical scheme to compute the approximations in $\widehat{\boldsymbol{U}}$. This scheme should only involve smooth functions as assumed at the beginning of this section. This is in particular the case for the TRM and Lax–Friedrichs scheme, and the corresponding Jacobian matrices are given in Appendix D.1.

Algorithm 1: Compute the gradient of the cost function (Forward-Propagation).

Input: Parameter θ, Density matrix $\boldsymbol{U}$.
Output: Gradient $g = \dfrac{\partial L}{\partial \theta}(\theta)$ of the cost function (25).

Set $\boldsymbol{G} = \boldsymbol{0}$ and $g = 0$;
Set $\widehat{\boldsymbol{U}}^0 = \boldsymbol{U}^0$;
for $i = 0, \dots, N_t - 2$ **do**
  $\boldsymbol{G} = \mathrm{Jac}_{\mathcal{H}^i}(\widehat{\boldsymbol{U}}^i) \cdot \boldsymbol{G} + \mathrm{Jac}_{\mathcal{H}^i}(C(\theta))$;
  $\widehat{\boldsymbol{U}}^{i+1} = \mathcal{H}^i(\widehat{\boldsymbol{U}}^i, C(\theta)\,;\,\boldsymbol{U})$;
  $g = g + \boldsymbol{G}^T \left(\widehat{\boldsymbol{U}}^{i+1} - \boldsymbol{U}^{i+1}\right)$;
end
$g = \dfrac{\partial C}{\partial \theta}(\theta) \cdot g$;
return g.

Using these expressions, Algorithm 1 provides a first way to compute the gradient vector (35). This algorithm can be referred to as a Forward-Propagation algorithm: we visit each "time" i sequentially from 0 to the final time to compute the gradient. The finite volume approximations $\widehat{\boldsymbol{U}}^1, \dots, \widehat{\boldsymbol{U}}^{N_t-1}$ are computed on-the-fly, thus saving some storage space. On the other hand, note that each iteration requires matrix–matrix multiplications. Even though the Jacobian matrices involved in these products are sparse, the stored matrix $\boldsymbol{G}$ will fill up as i grows, rendering the computational and storage costs of each iteration more and more expensive. This could become cumbersome in some applications where the size of this matrix, which is $N_t \times N_x$, is large.

Inspired by the theory around the fitting of neural networks we propose a Back-Propagation algorithm which allows us to compute this same gradient while only requiring matrix–vector products, thus keeping the computational and storage costs in check. This algorithm is based on the next result.

Corollary C.2. *The gradient defined in Proposition C.1 satisfies*

$$\frac{\partial L}{\partial \theta}(\theta) = \frac{\partial C}{\partial \theta}(\theta) \cdot \sum_{i=0}^{N_t-2} \mathrm{Jac}_{\mathcal{H}^i}(C(\theta))^T \boldsymbol{\delta}^{i+1}(C(\theta)) \,, \tag{37}$$

where for $C \in (0, 1/2)$*, the sequence* $(\boldsymbol{\delta}^i(C))_{i \in [\![1, N_t-1]\!]}$ *is defined by the recurrence*

$$\begin{cases} \boldsymbol{\delta}^{N_t-1}(C) = \left(\widehat{\boldsymbol{U}}^{N_t-1}(C) - \boldsymbol{U}^{N_t-1}\right) \\ \boldsymbol{\delta}^i(C) = \left(\widehat{\boldsymbol{U}}^i(C) - \boldsymbol{U}^i\right) + \mathrm{Jac}_{\mathcal{H}^i}(\widehat{\boldsymbol{U}}^i(C))^T \boldsymbol{\delta}^{i+1}(C), \quad i \in [\![0, N_t - 2]\!] \,. \end{cases} \tag{38}$$

Proof. See Appendix D.2. □

Corollary C.2 provides an alternative expression for computing the gradient function, which in turn yields Algorithm 2. This last algorithm can be referred to as a Back-Propagation algorithm: we visit each "time" i sequentially from the latest to the initial time to compute the gradient. Consequently, it is no longer possible to compute the density vectors on-the-fly: they must be computed and stored beforehand. Once this is done, each iteration of Algorithm 2 requires the same computational and storage costs, those

associated with products between some sparse matrices and vectors. Hence, these costs are much less influenced by the size of the problem, assuming that there is enough storage space for the density vectors.

Algorithm 2: Compute the gradient of the cost function (Back-Propagation).

Input: Parameter θ, Density matrix $\boldsymbol{U}$.

Output: Gradient $g = \dfrac{\partial L}{\partial \theta}(\theta)$ of the cost function (25).

Set $\widehat{\boldsymbol{U}}^0 = \boldsymbol{U}^0$;

for $i = 0, \dots, N_t - 2$ **do**

  $\widehat{\boldsymbol{U}}^{i+1} = \mathcal{H}^i(\widehat{\boldsymbol{U}}^i, C(\theta)\,;\, \boldsymbol{U})$;

end

Set $\boldsymbol{\delta} = \left(\widehat{\boldsymbol{U}}^{N_t-1} - \boldsymbol{U}^{N_t-1}\right)$, $g = \mathrm{Jac}_{\mathcal{H}^{N_t-2}}(\boldsymbol{C})^T \boldsymbol{\delta}$;

for $l = N_t - 2, \dots, 1$ **do**

  $\boldsymbol{\delta} = \left(\widehat{\boldsymbol{U}}^{l} - \boldsymbol{U}^{l}\right) + \mathrm{Jac}_{\mathcal{H}^l}(\widehat{\boldsymbol{U}}^{l})^T \boldsymbol{\delta}$;

  $g = g + \mathrm{Jac}_{\mathcal{H}^{l-1}}(C(\theta))^T \boldsymbol{\delta}$;

end

$g = \dfrac{\partial C}{\partial \theta}(\theta) \cdot g$;

return g

The minimization problem (22) can then be solved using Algorithm 3. In this algorithm, convergence is understood as the fulfillment of some numerical criterion based on the value of the gradient of the cost function or on the value of the cost function (or both), and chosen by the user. A typical choice is declaring convergence once the norm of the gradient vector is below some predefined threshold. Algorithms allowing to compute descent directions for various gradient descent algorithms can be found in Ruder (2016). We can for instance cite the steepest gradient method for which the descent direction d is computed from the gradient g as

$$d = -\alpha \cdot g$$

for some fixed step size $\alpha > 0$. The (Polak–Ribière) Conjugate gradient algorithm on the other hand uses, at the tth iteration of the process, the descent direction $d^{(t)}$ given by

$$d^{(t)} = -g^{(t)} + \left(\frac{\left(g^{(t)}\right)^T \left(g^{(t)} - g^{(t-1)}\right)}{\left(g^{(t-1)}\right)^T g^{(t-1)}} \right) d^{(t-1)} ,$$

where $g^{(t)}$ denotes the gradient of the cost function at the tth iteration (Nocedal and Wright, 2006). This last algorithm is the one used in the numerical applications of this paper.

Algorithm 3: Find the optimal control parameter of the discrete dynamical system associated with a density matrix (constant case).

Input: Density matrix $\boldsymbol{U}$ defined from a discretization with step sizes $\Delta t, \Delta x$;
Initial value $\theta \in \mathbb{R}$;
A routine grad to compute gradients (Algorithm 1 or 2);
A routine dir to compute descent directions (Steepest gradient method, conjugate gradient algorithm,...).

Output: Parameter C^* of the minimization problem (22).

while *Convergence is not achieved* **do**

  Compute the gradient $g = \dfrac{\partial L}{\partial \theta}(\theta)$ of the cost function (25) with respect to the unrestricted parameters θ:

  $$g = \mathrm{grad}\,(\theta, \boldsymbol{U}) \ ;$$

  Compute the descent direction from the gradient: $d = \mathrm{dir}(g)$;
  Update the parameter:

  $$\theta = \theta + d \ ;$$

end

return $C^* = C(\theta)$.

Proposition C.3. *Let $L_{[P_t,P_x]}$ be the cost function defined in (34) and assume that the mappings $\mathcal{H}^i$, $i \geq 0$, defined through (21) and (20) are smooth with respect to their arguments. Then, following the notations from Proposition C.1, the gradient of L is given by*

$$\frac{\partial L_{[P_t,P_x]}}{\partial \theta}(\theta) = \frac{\partial C}{\partial \theta}(\theta) \cdot \sum_{i=1}^{N_t-1} \mathrm{Jac}_{\widehat{\boldsymbol{U}}^{iP_t}}(C(\theta))^T \boldsymbol{M}_{P_x}^T \left(\boldsymbol{M}_{P_x} \widehat{\boldsymbol{U}}^i(C(\theta)) - \boldsymbol{U}^i \right), \quad \theta \in \mathbb{R}, \tag{39}$$

where $\boldsymbol{M}_{P_x} \in \mathbb{R}^{N_x \times (P_x M_x)}$ is the averaging matrix defined by

$$[\boldsymbol{M}_{P_x}]_{j,m} = \begin{cases} 1/P_x & \text{if } m = l + jP_x \text{ with } l \in [\![0, P_x - 1]\!] \\ 0 & \text{otherwise} \end{cases}$$

and the Jacobian matrices $\widehat{\boldsymbol{U}}^{\iota}(C(\theta))$, $\iota \in [\![0, (N_t-1)P_t]\!]$, are once gain obtained through the recurrence (36).

Algorithm 4: Compute the gradient of the cost function (Forward-Propagation).

Input: *Parameter θ, Density matrix $\boldsymbol{U}$, Subdivision parameters P_t, P_x.*

Output: *Gradient $g = \frac{\partial L}{\partial \theta}(\theta)$ of the cost function (34).*

Set $\boldsymbol{G} = \boldsymbol{0}$;
Set $g = 0$;
Set $\widehat{\boldsymbol{U}}^0 = P_x \boldsymbol{M}_{P_x}^T \boldsymbol{U}^0$;
for $i = 0, \dots, (N_t-1)P_t - 1$ **do**
  $\boldsymbol{G} = \mathrm{Jac}_{\mathcal{H}^i}(\widehat{\boldsymbol{U}}^i) \cdot \boldsymbol{G} + \mathrm{Jac}_{\mathcal{H}^i}(C(\theta))$;
  $\widehat{\boldsymbol{U}}^{i+1} = \mathcal{H}^i(\widehat{\boldsymbol{U}}^i, C(\theta); \boldsymbol{U})$;
  $g = g + \boldsymbol{G}^T \boldsymbol{M}_{P_x}^T \left(\boldsymbol{M}_{P_x} \widehat{\boldsymbol{U}}^{i+1} - \boldsymbol{U}^{i+1} \right)$;
end
$g = \frac{\partial C}{\partial \theta}(\theta) \cdot g$;
return g.

Algorithm 5: Compute the gradient of the cost function (Back-Propagation).

Input: *Parameter θ, Density matrix $\boldsymbol{U}$, Subdivision parameters P_t, P_x.*

Output: *Gradient $g = \frac{\partial L}{\partial \theta}(\theta)$ of the cost function (34).*

Set $\widehat{\boldsymbol{U}}^0 = P_x \boldsymbol{M}_{P_x}^T \boldsymbol{U}^0$;
for $i = 0, \dots, (N_t-1)P_t - 1$ **do**
  $\widehat{\boldsymbol{U}}^{i+1} = \mathcal{H}^i(\widehat{\boldsymbol{U}}^i, C(\theta); \boldsymbol{U})$;
end
Set $\boldsymbol{\delta} = \boldsymbol{M}_{P_x}^T \left(\boldsymbol{M}_{P_x} \widehat{\boldsymbol{U}}^{N_t-1} - \boldsymbol{U}^{N_t-1} \right)$;
Set $g = \mathrm{Jac}_{\mathcal{H}^{N_t-2}}(\boldsymbol{C})^T \boldsymbol{\delta}$;
for $l = (N_t-1)P_t - 1, \dots, 1$ **do**
  $\boldsymbol{\delta} = \mathrm{Jac}_{\mathcal{H}^l}(\widehat{\boldsymbol{U}}^l)^T \boldsymbol{\delta}$;
  if $(l/P_t) \in \{0, \dots, N_t - 2\}$ **then**
    $\boldsymbol{\delta} = \boldsymbol{\delta} + \boldsymbol{M}_{P_x}^T \left(\boldsymbol{M}_{P_x} \widehat{\boldsymbol{U}}^{l/P_t} - \boldsymbol{U}^{l/P_t} \right)$;
  end
  $g = g + \mathrm{Jac}_{\mathcal{H}^{l-1}}(C(\theta))^T \boldsymbol{\delta}$;
end
$g = \frac{\partial C}{\partial \theta}(\theta) \cdot g$;
return g

Equivalently, this gradient can be obtained by

$$\frac{\partial L_{[P_t,P_x]}}{\partial \theta}(\theta) = \frac{\partial C}{\partial \theta}(\theta) \cdot \sum_{\iota=0}^{(N_t-1)P_t-1} \mathrm{Jac}_{\mathcal{H}^\iota}(C(\theta))^T \boldsymbol{\delta}^{\iota+1}(C(\theta)), \tag{40}$$

where for $C \in (0, 1/2)$, the sequence $(\boldsymbol{\delta}^\iota(C))_{\iota \in [\![0,(N_t-1)P_t]\!]}$ is defined by the recurrence

$$\begin{cases} \boldsymbol{\delta}^{(N_t-1)P_t}(C) = \boldsymbol{M}_{P_x}^T \left(\boldsymbol{M}_{P_x} \widehat{\boldsymbol{U}}^{N_t-1}(C) - \boldsymbol{U}^{N_t-1} \right) \\ \boldsymbol{\delta}^\iota(C) = \mathbb{1}_{(\iota/P_t) \in [\![0,N_t-2]\!]} \boldsymbol{M}_{P_x}^T \left(\boldsymbol{M}_{P_x} \widehat{\boldsymbol{U}}^{\iota/P_t}(C) - \boldsymbol{U}^{\iota/P_t} \right) + \mathrm{Jac}_{\mathcal{H}^\iota}(\widehat{\boldsymbol{U}}^\iota(C))^T \boldsymbol{\delta}^{\iota+1}(C), \quad \iota \in [\![0, (N_t-1)P_t - 1]\!] \end{cases} \tag{41}$$

with $\mathbb{1}_A$ denoting the indicator function of a proposition A.

Proof. This result is a direct consequence of Proposition C.1 and Corollary C.2 after noting that replacing, in the expression (25) of the cost function L, the approximation matrix $\hat{\boldsymbol{U}}$ by the averaged approximations $\boldsymbol{M}_{P_x}\hat{\boldsymbol{U}}$ yields the expression (34) of the cost function $L_{[P_t,P_x]}$. The chain rule then yields the result. □

Then, Algorithm 3 can be used to solve the minimization problem, after adjusting the algorithms for computing gradients according to the previous proposition, thus yielding Algorithms 4 and 5.

Corollary C.4. *Let $\tilde{L}_{[P_t,P_x]}$ be the cost function defined by (31) and associated with a set $I_c \subset [\![1, N_x - 2]\!]$ of observed columns. Assume that the mappings $\mathcal{H}^i$, $i \geq 0$, defined through (21) and (20) are smooth with respect to their arguments.*

Then, following the notations from Proposition C.3, the gradient of $\tilde{L}_{[P_t,P_x]}$ is given by Eq. (39) (or equivalently by (40)) after replacing the matrix $\boldsymbol{M}_{P_x}$ by the matrix $\tilde{\boldsymbol{M}}_{P_x}$ defined by

$$[\tilde{\boldsymbol{M}}_{P_x}]_{j,m} = \begin{cases} 1/P_x & \text{if } j = j_c \in I_c \text{ and } m = l + j_c P_x \text{ with } l \in [\![0, P_x - 1]\!] \\ 0 & \text{otherwise} \end{cases}.$$

Proof. This result is a direct consequence of Proposition C.1 Proposition C.3 after noting that replacing, in the expression (34) of the cost function $L_{[P_t,P_x]}$, the matrix $\boldsymbol{M}_{P_x}$ by the matrix $\tilde{\boldsymbol{M}}_{P_x}\hat{\boldsymbol{U}}$ yields an expression equal to the sum of the cost function $\tilde{L}_{[P_t,P_x]}$ (given in (31)) and a term that does not depend on the parameter θ (but only on the entries of the density matrix $\boldsymbol{U}$). Hence, the gradient of this expression (with respect to the parameters) will be the same as the gradient of $\tilde{L}_{[P_t,P_x]}$, which gives the result. □

Consequently, the gradient of the cost function (31) can be computed using either Algorithm 4 or Algorithm 5 and accounting for the modification described in Corollary C.4. Hence, gradient-based optimization can once again be considered to minimize this cost function.

Appendix D. Jacobian matrices and gradient computations

D.1. Jacobian matrices for the TRM and LxF

We derive here, for the TRM and LxF schemes, the expression of the Jacobian matrices needed to compute the gradient of the cost functions considered in this work.

Since the initial condition vector $\boldsymbol{U}^0$, given by (17), does not depend on $\boldsymbol{C}$, $\mathrm{Jac}_{\boldsymbol{U}^0}(\boldsymbol{C}) = 0$.

For the TRM, Eq. (21) is used to derive the expression of the remaining Jacobian matrices involved in the recurrence relation (10): They are sparse matrices, whose non-zero entries are given by

$$\left[\mathrm{Jac}_{\mathcal{H}^i}(\boldsymbol{U}^i)\right]_{j,k} = \begin{cases} C^i_j(1 - U^i_j) & \text{if } k = j - 1 \\ 1 - C^i_j U^i_{j-1} + C^i_{j+1}(1 - U^i_{j+1}) & \text{if } k = j \\ C^i_{j+1} U^i_j & \text{if } k = j + 1, \end{cases} \quad j \in [\![1, N_x - 2]\!], \quad i \in [\![0, N_t - 2]\!], \tag{42}$$

$$\left[\mathrm{Jac}_{\mathcal{H}^i}(\boldsymbol{C})\right]_{j,k} = \begin{cases} U^i_{j-1}(1 - U^i_j) & \text{if } k = j \\ -U^i_j(1 - U^i_{j+1}) & \text{if } k = j + 1 \\ , \end{cases} \quad j \in [\![1, N_x - 2]\!], \quad i \in [\![0, N_t - 2]\!]. \tag{43}$$

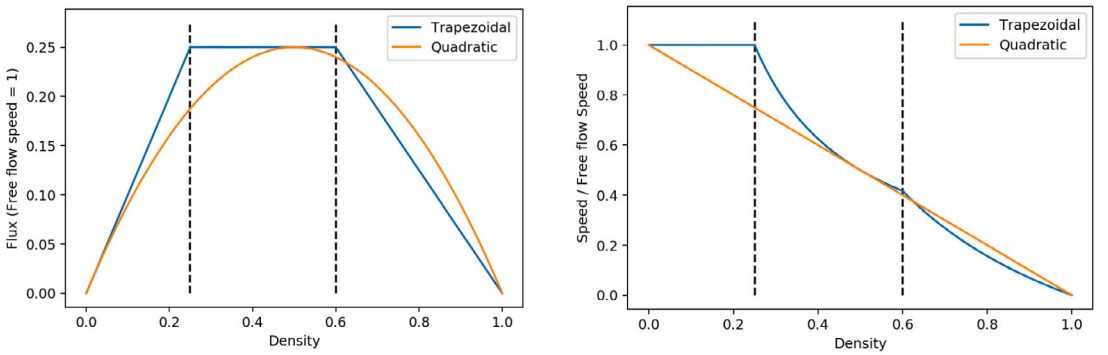
Similarly, we get for the LxF scheme

$$\left[\mathrm{Jac}_{\mathcal{H}^i}(\boldsymbol{U}^i)\right]_{j,k} = \begin{cases} 1/2 + C^i_j(1 - 2U^i_{j-1})/2 & \text{if } k = j - 1 \\ C^i_j(1 - 2U^i_j)/2 - C^i_{j+1}(1 - 2U^i_j)/2 & \text{if } k = j \\ 1/2 - C^i_{j+1}(1 - 2U^i_{j+1})/2 & \text{if } k = j + 1, \end{cases} \quad j \in [\![1, N_x - 2]\!], \quad i \in [\![0, N_t - 2]\!], \tag{44}$$

$$\left[\mathrm{Jac}_{\mathcal{H}^i}(\boldsymbol{C})\right]_{j,k} = \begin{cases} U^i_{j-1}(1 - U^i_{j-1}) + U^i_j(1 - U^i_j) & \text{if } k = j \\ U^i_j(1 - U^i_j) + U^i_{j+1}(1 - U^i_{j+1}) & \text{if } k = j + 1 \\ , \end{cases} \quad j \in [\![1, N_x - 2]\!], \quad i \in [\![0, N_t - 2]\!]. \tag{45}$$

D.2. Back-propagated gradient

We present here the proof of Corollary C.2.



(a) Flux-density relationship

(b) Speed-density relationship (Function f_2)

Fig. 19. Comparison between quadratic and trapezoidal fundamental diagrams. The maximal density ρ_m and the free flow speed v_m are taken equal to 1. The black dotted lines are placed at the critical density values $\rho_1 = 0.25$ and $\rho_2 = 0.6$.

Proof. Using Eq. (38), we have

$$\begin{aligned}\sum_{l=0}^{N_t-2} \mathrm{Jac}_{\mathcal{H}^l}(\boldsymbol{C})^T \boldsymbol{\delta}^{l+1} &= \sum_{l=0}^{N_t-2} \left(\mathrm{Jac}_{\boldsymbol{U}^{l+1}}(\boldsymbol{C}) - \mathrm{Jac}_{\mathcal{H}^l}(\boldsymbol{U}^l)\mathrm{Jac}_{\boldsymbol{U}^l}(\boldsymbol{C})\right)^T \boldsymbol{\delta}^{l+1} \\ &= \sum_{l=0}^{N_t-2} \mathrm{Jac}_{\boldsymbol{U}^{l+1}}(\boldsymbol{C})^T \boldsymbol{\delta}^{l+1} - \sum_{l=0}^{N_t-2} \mathrm{Jac}_{\boldsymbol{U}^l}(\boldsymbol{C})^T \mathrm{Jac}_{\mathcal{H}^l}(\boldsymbol{U}^l)^T \boldsymbol{\delta}^{l+1} \\ &= \sum_{l=0}^{N_t-2} \mathrm{Jac}_{\boldsymbol{U}^{l+1}}(\boldsymbol{C})^T \boldsymbol{\delta}^{l+1} - \sum_{l=1}^{N_t-2} \mathrm{Jac}_{\boldsymbol{U}^l}(\boldsymbol{C})^T \mathrm{Jac}_{\mathcal{H}^l}(\boldsymbol{U}^l)^T \boldsymbol{\delta}^{l+1}\,,\end{aligned}$$

since $\mathrm{Jac}_{\boldsymbol{U}^0}(\boldsymbol{C}) = \boldsymbol{0}$. The definition of the sequence $(\boldsymbol{\delta}^l)_{1\le l\le N_t-1}$ in Eq. (38) then gives

$$\begin{aligned}\sum_{l=0}^{N_t-2} \mathrm{Jac}_{\mathcal{H}^l}(\boldsymbol{C})^T \boldsymbol{\delta}^{l+1} &= \sum_{l=0}^{N_t-2} \mathrm{Jac}_{\boldsymbol{U}^{l+1}}(\boldsymbol{C})^T \boldsymbol{\delta}^{l+1} - \sum_{l=1}^{N_t-2} \mathrm{Jac}_{\boldsymbol{U}^l}(\boldsymbol{C})^T (\boldsymbol{\delta}^l - \boldsymbol{d}^l) \\ &= \mathrm{Jac}_{\boldsymbol{U}^1}(\boldsymbol{C})^T \boldsymbol{\delta}^1 + \sum_{l=1}^{N_t-2} \left(\mathrm{Jac}_{\boldsymbol{U}^{l+1}}(\boldsymbol{C})^T \boldsymbol{\delta}^{l+1} - \mathrm{Jac}_{\boldsymbol{U}^l}(\boldsymbol{C})^T \boldsymbol{\delta}^l\right) + \sum_{l=1}^{N_t-2} \mathrm{Jac}_{\boldsymbol{U}^l}(\boldsymbol{C})^T \boldsymbol{d}^l \\ &= \mathrm{Jac}_{\boldsymbol{U}^1}(\boldsymbol{C})^T \boldsymbol{\delta}^1 + \mathrm{Jac}_{\boldsymbol{U}^{N_t-1}}(\boldsymbol{C})^T \boldsymbol{\delta}^{N_t-1} - \mathrm{Jac}_{\boldsymbol{U}^1}(\boldsymbol{C})^T \boldsymbol{\delta}^1 + \sum_{l=1}^{N_t-2} \mathrm{Jac}_{\boldsymbol{U}^l}(\boldsymbol{C})^T \boldsymbol{d}^l \\ &= \mathrm{Jac}_{\boldsymbol{U}^{N_t-1}}(\boldsymbol{C})^T \boldsymbol{\delta}^{N_t-1} + \sum_{l=1}^{N_t-2} \mathrm{Jac}_{\boldsymbol{U}^l}(\boldsymbol{C})^T \boldsymbol{d}^l\,.\end{aligned}$$

Finally, since from Eq. (38), $\boldsymbol{\delta}^{N_t-1} = \boldsymbol{d}^{N_t-1}$, we get

$$\sum_{l=0}^{N_t-2} \mathrm{Jac}_{\mathcal{H}^l}(\boldsymbol{C})^T \boldsymbol{\delta}^{l+1} = \sum_{l=1}^{N_t-1} \mathrm{Jac}_{\boldsymbol{U}^l}(\boldsymbol{C})^T \boldsymbol{d}^l = \frac{\partial L}{\partial \boldsymbol{C}}(\boldsymbol{C})\,,$$

according to Eq. (35). □

Appendix E. Decomposition of trapezoidal flux functions

Let $\rho_m > 0$ denote some maximal density on a road. Consider the functions f_1 and f_2 defined by

$$f_1(\rho) = \rho, \quad \rho \in [0, \rho_m],$$

and

$$f_2(\rho) = \begin{cases} v_m & \text{if } 0 \le \rho \le \rho_1, \\ v_m \frac{\rho_1}{\rho} & \text{if } \rho_1 \le \rho \le \rho_2, \\ v_m \frac{\rho_1}{\rho_m - \rho_2}\left(\frac{\rho_1}{\rho} - 1\right) & \text{if } \rho_2 \le \rho \le \rho_m, \end{cases}$$

where v_m denotes the free flow speed, and ρ_1, ρ_2 denote some critical densities satisfying $0 < \rho_1 \leq \rho_2 < \rho_m$. Clearly, f_1 is a non-decreasing Lipschitz-continuous function such that $f_1(0) = 0$ and f_2 is a non-increasing Lipschitz-continuous function such that $f_2(\rho_m) = 0$. Then the flux function f defined by

$$f(\rho) = f_1(\rho) f_2(\rho), \quad \rho \in [0, \rho_m],$$

fits into the general TRM framework of Lipták et al. (2021). Also, f has a trapezoidal shape with a plateau in the segment $[\rho_1, \rho_2]$. Fig. 19 gives an example of such a flux function.

References

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

Barth, T., Herbin, R., Ohlberger, M., 2018. Finite volume methods: Foundation and analysis. In: Encyclopedia of Computational Mechanics, second ed. pp. 1–60.

Blandin, S., Couque, A., Bayen, A., Work, D., 2012. On sequential data assimilation for scalar macroscopic traffic flow models. Physica D 241 (17), 1421–1440.

Cassidy, M.J., Coifman, B., 1997. Relation among average speed, flow, and density and analogous relation between density and occupancy. Transp. Res. Rec. 1591 (1), 1–6.

Chainais-Hillairet, C., Champier, S., 2001. Finite volume schemes for nonhomogeneous scalar conservation laws: error estimate. Numer. Math. 88 (4), 607–639.

Chen, G.-Q., Karlsen, K.H., 2005. Quasilinear anisotropic degenerate parabolic equations with time-space dependent diffusion coefficients. Commun. Pure Appl. Anal. 4 (2), 241–266.

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

Delle Monache, M.L., Piccoli, B., Rossi, F., 2017. Traffic regulation via controlled speed limit. SIAM J. Control Optim. 55 (5), 2936–2958.

Edie, L.C., 1963. Discussion of traffic stream measurements and definitions. In: Proceedings of the Second International Symposium on the Theory of Traffic Flow, London. Port of New York Authority, pp. 139–154.

Eymard, R., Gallouët, T., Herbin, R., 2000. Finite volume methods. Handb. Numer. Anal. 7, 713–1018.

Fan, S., Seibold, B., 2013. Data-fitted first-order traffic models and their second-order generalizations: Comparison by trajectory and sensor data. Transp. Res. Rec. 2391 (1), 32–43.

Feinberg, M., 2019. Foundations of Chemical Reaction Network Theory. Springer.

Fulari, S., Vanajakshi, L., Subramanian, S.C., 2017. Artificial neural network–based traffic state estimation using erroneous automated sensor data. J. Transp. Eng. A: Syst. 143 (8), 05017003.

Garavello, M., Han, K., Piccoli, B., 2016. Models for Vehicular Traffic on Networks. American Institute of Mathematical Sciences (AIMS).

Goatin, P., Göttlich, S., Kolb, O., 2016. Speed limit and ramp meter control for traffic flow networks. Eng. Optim. 48 (7), 1121–1144.

Göttlich, S., Ziegler, U., Herty, M., 2013. Numerical discretization of hamilton–jacobi equations on networks. Netw. Heterog. Media 8 (3), 685.

Greenshields, B., Bibbins, J., Channing, W., Miller, H., 1935. A study of traffic capacity. In: Highway Research Board Proceedings, Vol. 1935. National Research Council (USA), Highway Research Board.

Karafyllis, I., Papageorgiou, M., 2019. Feedback control of scalar conservation laws with application to density control in freeways by means of variable speed limits. Automatica 105, 228–236.

Karlaftis, M.G., Vlahogianni, E.I., 2011. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. Transp. Res. C 19 (3), 387–399.

Karlsen, K.H., Towers, J.D., 2004. Convergence of the lax–friedrichs scheme and stability for conservation laws with a discontinuous space–time dependent flux. Chin. Ann. Math. 25 (03), 287–318.

Kessel, F., 2019. Traffic Flow Modeling. Springer.

Krajewski, R., Bock, J., Kloeker, L., Eckstein, L., 2018. The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp. 2118–2125.

Leduc, G., et al., 2008. Road traffic data: Collection methods and applications. Work. Pap. Energy Transp. Clim. Change 1 (55), 1–55.

LeVeque, R.J., 2002. Finite Volume Methods for Hyperbolic Problems, Vol. 31. Cambridge University Press.

Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves II. a theory of traffic flow on long crowded roads. Proc. R. Soc. Lond. Ser. A 229 (1178), 317–345.

Lipták, G., Pereira, M., Kulcsár, B., Kovács, M., Szederkényi, G., 2021. Traffic reaction model. arXiv:2101.10190.

Lu, X.-Y., Skabardonis, A., 2007. Freeway traffic shockwave analysis: exploring the NGSIM trajectory data. In: 86th Annual Meeting of the Transportation Research Board, Washington, DC.

Nocedal, J., Wright, S., 2006. Numerical Optimization. Springer.

Piccoli, B., Rascle, M., 2013. Modeling and Optimization of Flows on Network. In: Lecture notes in Mathematics, vol. 2062, Springer.

Richards, P., 1956. Shock waves on the highway. Oper. Res. 4 (1), 42–51.

Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv:1609.04747.

Schreiter, T., Hinsbergen, C.Van., Zuurbier, F., Lint, J.Van., Hoogendoorn, S., 2010. Data-model synchronization in extended kalman filters for accurate online traffic state estimation. In: TFTC Summer Meeting 2010, Number CONF.

Seo, T., Bayen, A.M., Kusakabe, T., Asakura, Y., 2017. Traffic state estimation on highway: A comprehensive survey. Annu. Rev. Control 43, 128–151.

Treiber, M., Kesting, A., Thiemann, A., 2013. Traffic Flow Dynamics: Data, Models and Simulation. Springer.

Van Lint, J., Van Hinsbergen, C., 2012. Short-term traffic and travel time prediction models. Artif. Intell. Appl. Crit. Transp. Issues 22 (1), 22–41.

Wang, Y., Papageorgiou, M., 2005. Real-time freeway traffic state estimation based on extended kalman filter: a general approach. Transp. Res. B 39 (2), 141–167.

Wong, G., Wong, S., 2002. A multi-class traffic flow model–an extension of lwr model with heterogeneous drivers. Transp. Res. A 36 (9), 827–841.

Zhang, H.M., 2002. A non-equilibrium traffic model devoid of gas-like behavior. Transp. Res. B 36 (3), 275–290.

Zhong, M., Lingras, P., Sharma, S., 2004. Estimation of missing traffic counts using factor, genetic, neural, and regression techniques. Transp. Res. C 12 (2), 139–166.