Traffic Flow Reconstruction from Limited Collected Data

Nail Baloul
CERMICS, Ecole nationale des ponts et chaussées
Champs-sur-Marne, France
nail.baloul@enpc.fr

Amaury Hayat
CERMICS, Ecole nationale des ponts et chaussées
Champs-sur-Marne, France
amaury.hayat@enpc.fr

Thibault Liard
XLIM, Université de Limoges
Limoges, France
thibault.liard@unilim.fr

Pierre Lissy
CERMICS, Ecole nationale des ponts et chaussées
Champs-sur-Marne, France
pierre.lissy@enpc.fr

***Abstract*—We propose an efficient method for reconstructing traffic density with low penetration rate of probe vehicles. Specifically, we rely on measuring only the initial and final positions of a small number of cars which are generated using microscopic dynamical systems. We then implement a machine learning algorithm from scratch to reconstruct the approximate traffic density. This approach leverages learning techniques to improve the accuracy of density reconstruction despite constraints in available data. For the sake of consistency, we will prove that, if only using data from dynamical systems, the approximate density predicted by our learned-based model converges to a well-known macroscopic traffic flow model when the number of vehicles approaches infinity.**

***Index Terms*—Machine Learning, Neural Networks, Nonlinear Systems Identification, Optimization, Differential Equations, Traffic Management**

I. Introduction

Vehicular traffic is often described at two distinct levels, microscopic and macroscopic. The microscopic approach focuses on individual vehicles and their interactions. It considers the behavior of each driver capturing detailed dynamics such as velocity and acceleration. It enables to account for the impact of individual decisions on overall traffic flow. The macroscopic approach relies on a continuity assumption, describing traffic flow in terms of density, flow rate and mean speed. It treats traffic as a continuous fluid, governed by conservation principles such as the conservation of vehicles. It allows for large-scale traffic management. We refer to the survey papers [1], [3], [10] for a general discussion about traffic models at different scales. Because the state of continuous systems or high-dimensional microscopic systems is impossible to fully monitor, being able to reconstruct it from limited traffic data is a paramount challenge that has attracted a strong attention in the last two decades [2], [6], [9], [11].

We aim to construct a model that closely approximates the original (or intrinsic) traffic model while observing considerably fewer vehicles. This approach is motivated by the fact that monitoring all vehicles on a roadway is very challenging and often impractical due to technical limitations or resource constraints. Our main contribution is to implement a machine learning algorithm that reconstructs traffic flow using only the initial and final positions of probe vehicles. A key advantage of our approach is that it implicitly satisfies conservation laws without having to enforce them with additional constraints. This is achieved by using artificial data created by adjusting well-known traffic flow models and using activation functions in neural networks tailored to underlying physical problems.

The paper is organized as follows. In section 2, we introduce the traffic models used throughout our study. Section 3 paves the way for the formulation of a constrained optimization problem that will be the core of our machine learning algorithm in section 4. Section 5 provides a theoretical guarantee of our model. In section 6 we will illustrate the efficiency of our learning-based method by numerical experiments. Finally, Section 7 concludes the paper and suggests directions for future research.

II. Traffic flow models

A. Microscopic model

We consider a system of $N+1$ vehicles moving on a one dimensional single-lane road. The first vehicle is designated as the leader while the remaining vehicles are referred to as followers. The position of each vehicle i, denoted x_i, is expressed as a function of time t. Each vehicle, regarded as a particle, moves at a speed which is computed via a velocity map v and a distance to the vehicle immediately ahead of it. We assume that the leader has no vehicle in front of it and therefore maintains constant and maximal velocity denoted $v_{\max}$. Additionally, the cumulative length of the vehicles is given by L. In other words, $L = lN$, where l is the average

length of a single car. Then the Follow-the-Leader (FtL) microscopic model consists of the dynamical system

$$\begin{cases} \dot{x}_N^N(t) = v_{\max}, & t > 0, \\ \dot{x}_i^N(t) = v\left(\frac{L}{N\left(x_{i+1}^N(t) - x_i^N(t)\right)}\right), & t > 0, \\ x_i^N(0) = \bar{x}_i, & i = 0, \dots, N, \end{cases} \tag{1}$$

where $\bar{x}_i$ denotes the initial position of vehicle i. In [5], the authors established well-posedness of system (1) under the condition that the initial positions satisfy $\bar{x}_0 < \bar{x}_1 < \cdots < \bar{x}_{N-1} < \bar{x}_N$. This nonlinear model not only becomes high-dimensional and computationally expensive when dealing with a large number of vehicles but also assumes that drivers react solely to the vehicle immediately ahead, disregarding the influence of other vehicles and overall traffic conditions. Thus when we have a large number of vehicles it is often more adapted -at least computationally- to consider a macroscopic model as described in the following.

B. Macroscopic model

The Lighthill-Whitham-Richards (LWR) model describes traffic flow using the one-dimensional conservation law. Indeed, vehicles are treated as a continuous medium similar to particles in fluid which leads us to the initial value problem

$$\begin{cases} \frac{\partial \rho}{\partial t}(t,x) + \frac{\partial f(\rho)}{\partial x}(t,x) = 0, & x \in \mathbb{R}, \quad t > 0, \\ \rho(0,x) = \bar{\rho}(x), & x \in \mathbb{R}, \end{cases} \tag{2}$$

where $\rho(t,x)$ is the traffic density at position x and at time t, $f(\rho)$ is the flux function representing the flow rate of vehicles and $\bar{\rho}$ is the initial vehicular density. Typically, the flux function is expressed as $f(\rho) := \rho v(\rho)$, where $v(\rho)$ is representing the average velocity of vehicles at a given density. Obviously, this model does not capture the behavior of individual vehicles as it focuses on aggregate traffic flow variables such as density and flow rate.

The convergence analysis of the microscopic FtL model (1) towards the macroscopic LWR model (2) has been extensively studied in recent years [4], [7]. The procedure involves considering an initial density $\bar{\rho}$ which is discretized to determine the initial positions of $N+1$ vehicles. We then allow the FtL system (1) to evolve according to its dynamics and compare the resulting solution with the one that satisfies the LWR model (2) as the number of vehicles approaches infinity and the length of each car tends to zero.

C. Data-driven based model

In our scenario, it is crucial to note that, unlike previous works that often rely on a given initial density, we do not have access to this critical information. It represents a significant challenge, as density is a key component for accurately predicting traffic flow. We must compensate for this lack of density data by leveraging other available information, such as the initial and final positions of probe vehicles.

In [2], [9], the authors considered probe vehicles regarded as mobile sensors within the traffic flow. These vehicles are governed by an Ordinary Differential Equation (shortly ODE) system analogous to (1) where traffic density evolves according to a Partial Differential Equation (shortly PDE) related to (2). To reconstruct density, vehicle positions $x_i(t)$, local density $\rho_i(t) = \rho(t, x_i(t))$ at vehicle locations and instantaneous speed $v_i(t) = \dot{x}_i(t)$ were needed in real time. This physics-informed learning method relies on explicit enforcement of conservation laws through PDE-derived Lagrangian terms in their optimization. The main drawback of this method is the lack of theoretical guarantee to prove that the reconstructed density converges to the conservation law added as a constraint in the optimization problem.

In [8] the authors adopted a Physical Informed Neural Network (PINN) strategy using density and flow measurements from synthetic data computed at fixed spatial locations. They illustrated that accuracy of their traffic state estimation improves when the number of locations increases.

In [6], the authors conducted the Mobile Century field experiment where they studied the feasibility of using GPS-enabled mobile phones as a cost-effective traffic monitoring system. The system collected continuous trajectory data, with position and velocity updates recorded every three seconds. This extensive data collection ensured sufficient information for traffic monitoring. Their results suggested that a $2-3\%$ penetration of GPS-equipped phones in the driver population could provide accurate measurements of traffic flow velocity.

In contrast, our approach requires fewer information than [2], [6], [8], [9] as it only assumes the knowledge of initial and final positions of the probe vehicles, thereby simplifying the data collection process and reducing the need for real-time measurements.

III. Mathematical formulation

To formalize our model, we first adapt the FtL microscopic model (1) to a modified framework that involves an optimization variable and a finite time horizon.

A. Parametrized ODE system

We consider the ODE system with finite time T

$$\begin{cases} \dot{x}_n^N(t) = v_{\max}, & t \in (0,T], \\ \dot{x}_i^N(t) = v\left(\frac{\alpha_i^N L}{N\left(x_{i+1}^N(t) - x_i^N(t)\right)}\right), & t \in (0,T], \\ x_i^N(0) = \bar{x}_i, & i = 0, \dots, n, \end{cases} \tag{3}$$

where $\alpha^N = (\alpha_0^N, \alpha_1^N, \dots, \alpha_{n-1}^N)^\top \in \mathbb{R}^n$ is a parameter and L is still representing the total length covered by all of the vehicles in a bumper-to-bumper configuration. In the new system (3), we insist on the fact that considered vehicles represent only a subset of the total involved. For $i = 0, \dots, n-1$, $\bar{x}_i$ and $\bar{y}_i$ denote respectively the given initial and final positions of car i. The goal is to gather

observations from a significantly reduced number n of probe vehicles compared to the total number N.

In system (3), α_i^N stands for the number of vehicles in the segment $[x_i(\cdot), x_{i+1}(\cdot))$ between consecutive probe vehicles i and $i+1$. Therefore, we impose physical conditions

$$\begin{aligned} &\alpha_i^N \in [1, \bar{z}_i], \quad i = 0, \dots, n-1, \\ &\sum_{j=0}^{n-1} \alpha_j^N = N, \end{aligned} \tag{4}$$

where for $i = 0, \dots, n-1$

$$\bar{z}_i \coloneqq \min \left\{ \frac{N(\bar{x}_{i+1} - \bar{x}_i)}{L}, \frac{N(\bar{y}_{i+1} - \bar{y}_i)}{L} \right\}. \tag{5}$$

The lower bound of 1 is trivial, as the segment must contain at least one vehicle. The upper bound ensures that there is enough space between the consecutive probe vehicles for this number of vehicles. Since L/N is the length of a car, there cannot be more than k cars on a segment. This is because kL/N is the minimal possible space required by k cars on the road. Moreover, a key assumption of our model is that overtaking between probe vehicles is not allowed. In other words, similarly to FtL system (1), the number of vehicles between consecutive probe vehicles remains constant over time which translates in the fact that α_i^N is constant with time.

To handle constraints (4) efficiently, we introduce the notation

$$\mathcal{A}_N \coloneqq \left\{ \alpha \in \mathbb{R}^n : \quad \alpha_i^N \in [1, \bar{z}_i], \quad i = 0, \dots, n-1 \right\}. \tag{6}$$

Introducing the parameter α^N enables modelling multiple unobserved vehicles between consecutive probe vehicles. The inclusion of α^N in ODE system (3) preserves the well-posedness of the system while allowing for a parametric adjustment to the dynamics of the vehicles. It will also play a role in establishing convergence to macroscopic model (2) which facilitates the transition from discrete, vehicle-level dynamics to continuous, density-level dynamics.

Although trajectories obtained from solving system (3) depend on N, we choose from now on to drop the superscript when referring to the trajectories for simplicity of notation.

B. Global existence and uniqueness

We assume that the velocity function satisfies the following hypotheses

(v1) $v \in C^1([0, +\infty))$,
(v2) v is decreasing on $[0, +\infty)$,
(v3) $v(0) = v_{\max}$ for some $v_{\max} \in \mathbb{R}$.

Since the parameter α^N is bounded and v is smooth, it is straightforward to show local existence and uniqueness of the solution to ODE system (3) from Picard-Lindelof Theorem. To ensure that the solution exists globally, we have to prove that the distance between two consecutive probe vehicles never vanish. The result has been shown by Di Francesco and Rosini in [5] for the classical FtL model (2) involving all vehicles in the system. We briefly outline it in our context, where the same principle must hold.

Lemma III.1. *Let $(x_0(\cdot), \dots, x_n(\cdot))$ be the solution of ODE system (3) and v satisfy hypotheses (v1)-(v3).*
Then the discrete maximum principle holds: for all $i = 0, \dots, n-1$ and for all $t \in [0, T]$,

$$\frac{\alpha_i^N L}{NM} \leq x_{i+1}(t) - x_i(t) \leq \bar{x}_n - \bar{x}_0 + (v_{\max} - v(M))\, t, \tag{7}$$

where $M \coloneqq \max\limits_{i=0,\dots,n-1} \left(\frac{\alpha_i^N L}{N(\bar{x}_{i+1} - \bar{x}_i)} \right)$ denotes the maximum discrete density at initial time $t = 0$.

Lemma III.1 can be proven using a similar approach to that presented in [5].

C. Discrete density

To clearly articulate the relationship between the state of ODE system (3) and the unknown density in equation (2) we introduce for $i = 0, \dots, n-1$ the discrete density as

$$\rho_i^N(t) \coloneqq \frac{\alpha_i^N L}{N(x_{i+1}(t) - x_i(t))}, \quad t \in [0, T], \tag{8}$$

where the trajectories $x(\cdot)$ of the probe vehicles are solutions to system (3). $\rho_i^N(\cdot)$ represents the local density in the interval between consecutive probe vehicles localised at $x_i(\cdot)$ and $x_{i+1}(\cdot)$. Based on (8), we define for $x \in \mathbb{R}$ the piecewise constant Eulerian discrete density as

$$\rho^N(t, x) \coloneqq \sum_{i=0}^{N-1} \rho_i^N(t) \chi_{[x_i(t), x_{i+1}(t))}(x), \quad t \in [0, T], \tag{9}$$

where χ_A is the characteristic function of the set A.

As proved in [7], ρ^N in (9) can be seen as a discrete approximation of the solution to the initial value problem (2).

IV. Learning-based density estimation

Our main focus is to reconstruct the traffic density only using initial and final positions of probe vehicles. To that end, we will solve an optimization problem that will allow us to construct a traffic discrete density ρ^N in (9) that converges to the solution of the LWR model (2) when N tends to infinity when using only artificial data.

A. ODE constrained optimization problem

We address our optimization problem with an innovation which contrasts with physics-informed methods. It lies in deliberately avoiding direct incorporation of the PDE into the system due to the lack of theoretical guarantee induced by these methods. We instead leverage the established convergence under certain conditions of FtL microscopic model (1) to LWR macroscopic model (2). This convergence implicitly ensures mass conservation without requiring explicit PDE constraints, thus maintaining flexibility in the reconstructed solution space while preserving physical consistency.

Enforcing directly the PDE constraints would not pose a computational challenge but would instead create a data dependency. Indeed, it would require time-varying density measurements at specific points in space, which we exclude from our problem set-up since our methodology assumes no density measurements.

We first compute the leader's trajectory

$$x_n(t) = v_{\max} t + \bar{x}_n, \quad t \in [0, T].$$

We then introduce for $\alpha^N \in \mathcal{A}_N$ the matrix W_{α^N} that accounts for interaction between the probe vehicles

$$\begin{cases} (W_{\alpha^N})_{i,i} \coloneqq -\frac{N}{\alpha_i^N L}, & i = 0, \dots, n-1, \\ (W_{\alpha^N})_{i,i+1} \coloneqq \frac{N}{\alpha_i^N L}, & i = 1, \dots, n-2, \\ (W_{\alpha^N})_{i,j} \coloneqq 0, & \text{otherwise}, \end{cases} \tag{10}$$

and function $b_{\alpha^N}(\cdot)$ defined as

$$b_{\alpha^N}(t) = (0, \dots, 0, b_{n-1}(t))^\top \in \mathbb{R}^n, \tag{11}$$

where $b_{n-1}(t) \coloneqq \frac{N}{\alpha_{n-1}^N L} (v_{\max} t + \bar{x}_n)$ accounts for the influence of the leader towards its followers.
Setting V such that

$$V(z) \coloneqq v\left(\frac{1}{z}\right), \quad z \in \mathbb{R}, \tag{12}$$

we can rewrite system (3) as

$$\begin{cases} \dot{x}(t) = V\left(W_{\alpha^N} x(t) + b_{\alpha^N}(t)\right), & t \in (0, T], \\ x(0) = (\bar{x}_0, \dots, \bar{x}_{n-1})^\top \end{cases} \tag{13}$$

where the state variable $x(\cdot) \coloneqq (x_0(\cdot), \dots, x_{n-1}(\cdot))^\top$ represents the trajectories of the probe followers.

We emphasize that initial density $\bar{\rho}$ of the entire fleet is unknown, as we do not have access to the initial positions of all vehicles. Consequently, the discretized initial density $\bar{\rho}^N$ which would typically depend on initial positions of all vehicles, cannot be determined. Equivalently, the ground truth value of α^N is unknown as it relies on information that is not available to us. This limitation highlights the challenge of reconstructing the system's behavior based on our limited data collection.

Based on (13) and (6) we write our ODE constrained optimization problem as

$$\underset{\alpha^N}{\text{minimize}} \quad \frac{1}{2} \left\| x(T) - \bar{y} \right\|^2 \tag{14a}$$

$$\text{s.t.} \quad \dot{x}(t) = V\left(W_{\alpha^N} x(t) + b_{\alpha^N}(t)\right), \quad t \in [0, T], \tag{14b}$$

$$x(0) = \bar{x}, \tag{14c}$$

$$\alpha^N \in \mathcal{A}_N. \tag{14d}$$

The objective function (14a) is a final position matching term which minimizes the discrepancy between the observed final positions $\bar{y}$ of probe vehicles and the corresponding final positions $x(T)$ predicted by the microscopic ODE model. Moreover, the problem involves constraints related to dynamics of the probe vehicles (14b)-(14c) whose velocities depend on inter-vehicle spacing influenced by α^N that must satisfy (14d). This formulation enables traffic density reconstruction from sparse observations by enforcing realistic vehicle dynamics, without directly incorporating PDE constraints.

Machine learning is effective at finding patterns in data even if the unknown relationships are complex. Our method is based on the use of a neural network designed to understand the dynamics of traffic by breaking down the process into small time steps.

B. Dataset generation

Dataset consist of artificial data based on simulated FtL dynamics. This choice allows us to capture the physics of the model we aim to approximate. Precisely, we let evolve $N+1$ vehicles using the classical FtL model (1) until time T is reached. 10% of the total fleet serve as probe vehicles for training data. These vehicles provide initial and final position data used in the machine learning training process. The probe vehicles represent a balanced sample of the overall traffic dynamics.
An additional 2.5% of the total fleet are selected for testing purposes so that 80% of the artificial data are for training and the remaining 20% are test data. Test data are strictly reserved for testing the model and are not used in any part of the training process. It is crucial to emphasize that the optimization network never "sees" or interacts with the initial and final positions of test vehicles. This separation is fundamental to our methodology, as it allows for an unbiased assessment of the model's performance. This choice enables us to evaluate whether the model has truly learned the underlying physics of traffic flow, rather than merely memorizing training data.

C. Learning architecture

We have designed the network architecture using a residual network (ResNet) approach. Each residual block in our ResNet corresponds to a single time step in the simulation.

The network starts with an input representing the initial positions of probe vehicles. This state is then propagated through time using a single, repeating residual block. The number of times this block is applied corresponds to the number of time steps in the simulation. Starting from initial traffic positions, the network predicts the next traffic state based on the current one. This repeated process allows step-by-step simulation of traffic flow until final time is reached. The structure described mirrors the first-order Euler discretization for ODE system (13).

The residual block incorporates physics-based principles of traffic flow. The weights in (10) and biases in (11) of the network are not independent variables, but are instead functions of α. Nonlinear map (12) acts as a physics-grounded activation function. Optimizing α determine the entire network's behavior.

D. Learning procedure

Our task is to minimize the cost between predicted and observed final positions. We train our machine learning algorithm with the training data. The predictions made by the neural network follow traffic rules defined by dynamics (14b) parameterized by α. Indeed, the final traffic state $x^\alpha(T)$ reconstructed through our ResNet architecture is compared to the observed training data $\bar{y}$. The loss function derived from (14a) writes

$$\frac{1}{n}\sum_{j=0}^{n}|x_j^\alpha(T)-\bar{y}_j|^2 \tag{15}$$

and computes the error between this final predicted state and the final observed state. Optimization parameter α is adjusted using backpropagation to minimize prediction errors (15). This iterative process of prediction, comparison and adjustment enables the model to capture the complex dynamics of traffic flow using just the initial condition and the observed final state.

After training, we recover from α the discrete vehicle densities (8). From these, we construct piecewise constant Euler discrete density ρ^N in the form of (9) and use it to simulate on our test data. In particular, we solve system

$$\begin{cases}\dot{x}_i(t)=v\left(\rho^N(t,x_i(t)^+)\right), & t\in(0,T],\\ x_i(0)=\bar{x}_i^N & i=0,\dots,n_{\text{test}},\end{cases} \tag{16}$$

where $\rho\left(t,x_i(t)^+\right) \coloneqq \lim_{x\to x_i(t);x>x_i(t)} \rho\left(t,x\right)$ which indicates that vehicle i is affected by the density located in front and where the density ρ^N is the approximate density obtained from training. We let evolve system (16) from initial test positions $\bar{x}_i^{n_{\text{test}}}$ and measure the error between simulated final positions and observed test data $\bar{y}_i^{n_{\text{test}}}$.

V. Convergence of the model

In this section we prove that, if only using data from dynamical systems, the approximate density ρ^N constructed by our machine learning model converges to the solution of the LWR macroscopic traffic flow model (2) when the number of vehicles approaches infinity.

Although our setting incorporates probe vehicles and parameter α, we can adapt most of the results established by Di Franco and Rosini in [5] using analogous arguments. We recall that the authors showed the convergence of the solution to microscopic model (1) to the unique entropic solution to macroscopic model (2) when N tends to infinity. To obtain a similar result in our context, the main challenge lies in imposing a condition on the distribution of α which would guarantee convergence.

We demonstrate that the discrete initial density $\rho^N(0,\cdot)$ converges to the initial condition $\bar{\rho}$ in the LWR model (2) under this additional assumption. In addition to the Euler discrete density (9), we consider the empirical discrete density defined by

$$\widehat{\rho}^N(t,\cdot) \coloneqq \frac{L}{N}\sum_{i=0}^{n-1}\alpha_i^N\delta_{x_i(t)}(\cdot), \quad t\in[0,T]. \tag{17}$$

We make the important observation that by construction the initial traffic density must satisfy for $i=0,\dots,n-1$

$$\bar{x}_{i+1}=\sup\left\{x\in\mathbb{R}:\int_{\bar{x}_i}^{x}\bar{\rho}(y)dy\le\frac{\alpha_i L}{N}\right\}. \tag{18}$$

Indeed, although we do not have access to the ground-truth initial car density $\bar{\rho}$, we do know that initial positions $\bar{x}_i$ of our probe vehicles verify (18) which states that the number of unobserved vehicles between $[x_i,x_{i+1})$ is given by α_i.

The following result inspired from [5, Proposition 4] ensures that the discretization aligns consistently with the true initial density when N tends to infinity. To simplify notations, let for $t\in[0,T]$, $\rho(t)\coloneqq\rho(t,\cdot)$ and $\widehat{\rho}(t)\coloneqq\widehat{\rho}(t,\cdot)$. In particular, for $t=0$, $\rho(0)\coloneqq\rho(0,\cdot)$ and $\widehat{\rho}(0)\coloneqq\widehat{\rho}(0,\cdot)$. We will use the Wasserstein distance defined in [5] by

$$\mathrm{W}_{L,1}(f,g)=\|f(]-\infty,\cdot])-g(]-\infty,\cdot]\|_{L^1(\mathbb{R},\mathbb{R})}. \tag{19}$$

We refer readers to [5, Section 2.3] which provides a rigorous introduction to this concept.

Proposition V.1. *Let $\bar{\rho}$ satisfy* (18) *and assume that*

$$\max_{i=0,\dots,n-1}\alpha_i^N=o(N). \tag{20}$$

Then, both sequences $(\rho^N(0))_{n\in\mathbb{N}}$ and $(\widehat{\rho}^N(0))_{n\in\mathbb{N}}$ converge to $\bar{\rho}$ in the sense of the $\mathrm{W}_{L,1}$ -Wasserstein distance in (19)*.*

Remark V.2. *A particular case of assumption* (20) *is when $f(N)=\log N$. This choice often used in practice ensures controlled growth of α_N.*

Proof. Set $l_N\coloneqq L/N$. Using the definition of the $\mathrm{W}_{L,1}$-Wasserstein distance and the expression of discrete density in (8) and mimicking the proof of [5, Proposition 4], we have

$$\begin{aligned}&\mathrm{W}_{L,1}(\rho^N(0),\widehat{\rho}^N(0))\\ &=\sum_{i=0}^{n-1}\int_{\bar{x}_i}^{\bar{x}_{i+1}^N}\left(\alpha_i^N l_N-\rho_i^N(t)\left(x-\bar{x}_i\right)\right)dx\\ &=\sum_{i=0}^{n-1}\alpha_i^N l_N\int_{\bar{x}_i}^{\bar{x}_{i+1}}\left(1-\frac{x-\bar{x}_i}{\bar{x}_{i+1}-\bar{x}_i}\right)dx\\ &\le\max_{i=0,\dots,n}\{\alpha_i^N\}l_N\left(\bar{x}_n-\bar{x}_0\right).\end{aligned} \tag{21}$$

We deduce that it suffices to prove that $(\widehat{\rho}^N(0))_{n\in\mathbb{N}}$ converges to $\bar{\rho}$ with respect to the $\mathrm{W}_{L,1}$-Wasserstein distance.

Using the expression of both Euler (9) and empirical (17) discrete densities, its holds that

$$
\begin{aligned}
&\mathrm{W}_{L,1}(\widehat{\rho}^N(0),\bar{\rho}) \\
&=\sum_{i=0}^{n-2}\int_{\bar{x}_i}^{\bar{x}_{i+1}}\left(\sum_{j=0}^{i}\alpha_j^N l_N-\int_{-\infty}^{x}\bar{\rho}(y)dy\right)dx \\
&\qquad+\int_{\bar{x}_{n-1}}^{\bar{x}_n}\left(L-\int_{-\infty}^{x}\bar{\rho}(y)dy\right)dx \\
&=\sum_{i=0}^{n-2}\int_{\bar{x}_i}^{\bar{x}_{i+1}}\left(\left(\sum_{j=0}^{i-1}\alpha_j^N l_N-\int_{\bar{x}_0}^{\bar{x}_i}\bar{\rho}(y)dy\right)\right. \\
&\qquad\left.+\left(\alpha_i^N l_N-\int_{\bar{x}_i}^{x}\bar{\rho}(y)dy\right)\right)dx \\
&\qquad+\int_{\bar{x}_{n-1}}^{\bar{x}_n}\left(\sum_{j=0}^{n-2}\alpha_j^N l_N-\int_{\bar{x}_0}^{\bar{x}_{n-1}}\bar{\rho}(y)dy\right) \\
&\qquad+\left(\alpha_{n-1}^N l_N-\int_{\bar{x}_{n-1}}^{x}\bar{\rho}(y)dy\right)dx.
\end{aligned}
$$

From (18), we deduce that

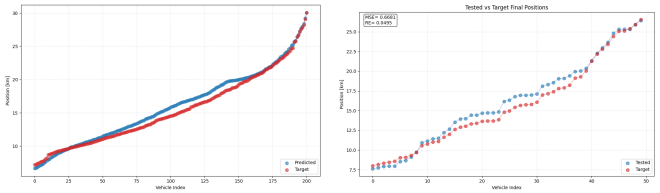
$$
\begin{aligned}
&\mathrm{W}_{L,1}(\widehat{\rho}^N(0),\bar{\rho}) \\
&\leq\sum_{i=0}^{n-2}\int_{\bar{x}_i}^{\bar{x}_{i+1}}\left(\alpha_i^N l_N-\int_{\bar{x}_i}^{x}\bar{\rho}(y)dy\right)dx \\
&\qquad+\int_{\bar{x}_{n-1}}^{\bar{x}_n}\left(\alpha_{n-1}^N l_N-\int_{\bar{x}_{n-1}}^{x}\bar{\rho}(y)dy\right)dx \\
&=\sum_{i=0}^{n-1}\alpha_i^N l_N\int_{\bar{x}_i}^{\bar{x}_{i+1}}\left(1-\frac{1}{\alpha_i^N l_N}\int_{\bar{x}_i}^{x}\bar{\rho}(y)dy\right)dx \\
&\leq\max_{i=0,\dots,n}\left\{\alpha_i^N\right\}l_N(\bar{x}_n-\bar{x}_0).
\end{aligned}
$$

From assumption (20) and estimate (21) we conclude that $(\rho^N(0))_{n\in\mathbb{N}}$ converges to $\bar{\rho}$ in the sense of the $\mathrm{W}_{L,1}$-Wasserstein distance, which achieves the proof. □

Moreover, by leveraging the expression of discrete density (9), we can generalize the convergence to the entropy solution of the conservation law (2), referring to the methodologies from abovementioned work. Specifically, [5, Theorem 3] which asserts the convergence of ρ^N to the unique entropy solution of (2) remains applicable in our case, requiring only minor modifications to the original arguments.

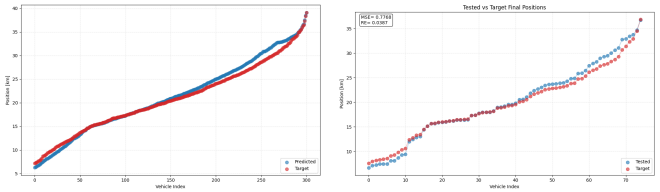
VI. Numerical Experiments

In this section, we present the results of our numerical simulations using the abovementioned training procedure. We consider two distinct traffic scenarios where the maximum allowable speed of traffic is $v_{\max} = 120$ km/h and the maximum traffic density is $\rho_{\max} = 200$ cars/km. We consider



(a) Neural Network learning using 5,000 training epochs (b) Test simulation using reconstructed density

Fig. 1: Comparison between prediction results and observed data with $N = 2000$



(a) Neural Network learning using 5,000 training epochs (b) Test simulation using reconstructed density

Fig. 2: Comparison between prediction results and observed data with $N = 3000$

a special instance of function v, known as the Greenshields velocity

$$
v(\rho) = v_{\max}\max\left\{1-\frac{\rho}{\rho_{max}},0\right\}, \quad \rho\in[0,\rho_{max}].
$$

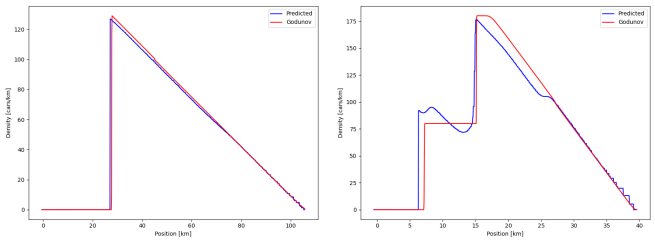
A time period of 0.1 corresponds to a real-world duration of 6 minutes. In the following illustrations, the origin represents the initial position $\bar{x}_0 = 0$ of the last follower at $t = 0$. The right end of the x-axis corresponds to the leader's position at the final time T. The initial domain refers to the portion of the road occupied by all probe vehicles at the start of the simulation.

A. Shock wave simulation

The first traffic scenario represents an abrupt transition in traffic conditions. The shock position where occurs a change from low to high density is set at normalized position 0.5 with respect to the initial domain, where position 0 corresponds to the last vehicle and position 1 corresponds to the leader's initial position.

At initial time, to the right of the shock, traffic is congested with a high density of $0.9\rho_{\max}$, meaning vehicles are packed closely together. To the left of the shock, traffic flows more freely with a lower density of $0.4\rho_{\max}$. Because we increase the number of vehicles while maintaining the same density profile, the actual distance covered by the congested region at $t = 0$ is larger for simulations with higher number of cars N.

The learning procedure outlined in the previous section produced results as shown in Fig.1 and Fig.2. Specifically,



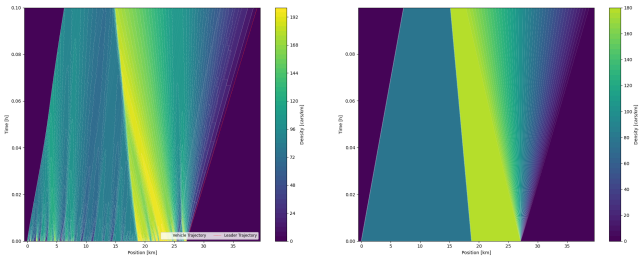
(a) $N = 5000$, $T = 0.5$ (b) $N = 3000$, $T = 0.1$

Fig. 3: Reconstructed final traffic density

during the testing phase where $N = 2000$, Mean Square Error (MSE) was calculated to be 0.6681, while Relative Error (RE) was 0.0495. When the dimension of the simulation was increased to $N = 3000$, MSE increased marginally to 0.7768 while RE slightly decreased to 0.0387. This shows that even with large dimensionality, the method is efficient across different scales.

Fig.3 illustrates that for large final times and a large number of vehicles, such as 5000, the reconstructed final density converges well to the solution obtained from solving the PDE. In shock simulations, longer times reduce system variations, improving reconstruction consistency. More importantly, the right side plot of the same figure presents a scenario with fewer vehicles (3000) and a shorter time frame (6 minutes compared to 30 minutes). Even under these more limited conditions, the reconstruction maintains an obvious accuracy. This suggests that the approach can effectively handle a range of scenarios, from those with ample time and vehicles to those with more limited resources.

Additionally, the method allows us to reconstruct density for all times. Fig. 4 compares the reconstructed discrete density with the solution to the LWR model, which is computed using the Godunov scheme, as well as the ground truth initial density from which the total fleet of vehicles evolved before selecting probe vehicles.

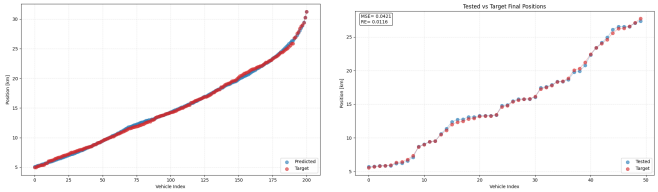


(a) Reconstructed traffic density (b) Godunov traffic density

Fig. 4: Comparison between the predicted density and Godunov scheme density with $N = 3000$ and $T = 0.1$

T	N $(p = 0.1)$	**Training epochs**	**Time cost (hh:mm:ss)**
0.1	1000	5000	00:09:49
0.1	1000	10000	00:17:07
0.1	2000	10000	00:30:53
0.1	3000	5000	00:27:22
0.2	4000	5000	00:50:38
0.2	5000	5000	01:41:36

TABLE I: Learning performance for shock wave



(a) Neural Network learning using 5,000 training epochs (b) Test simulation using reconstructed density

Fig. 5: Comparison between prediction results and observed data with $N = 2000$

B. Stop-and-go wave simulation

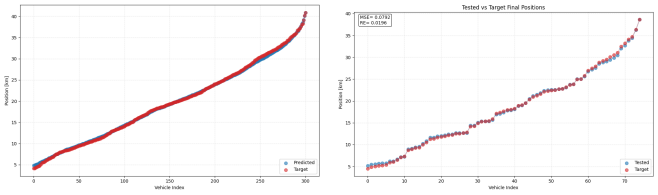
The second scenario captures a more complex traffic pattern characterized by alternating regions of congestion and free flow, represented by multiple waves. The base density is set at $0.6\rho_{\max}$, representing a moderately congested state. The wave amplitude of $0.3\rho_{\max}$ induces oscillations in density between $0.3\rho_{\max}$ and $0.9\rho_{\max}$, mimicking the cyclic nature of stop-and-go traffic.

T	N $(p = 0.1)$	**Training epochs**	**Time cost (hh:mm:ss)**
0.1	1000	5000	00:08:23
0.1	1000	10000	00:15:37
0.1	2000	10000	00:29:10
0.1	3000	5000	01:38:46
0.2	4000	5000	04:17:39
0.2	5000	5000	04:37:44

TABLE II: Learning performance for stop-and-go wave

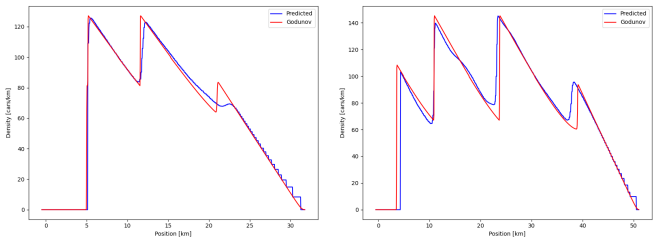
Fig.5 and Fig.6 illustrate the errors obtained by both the training and testing phases of our learning procedure in the case of the stop-and-go simulation. Specifically, during the testing phase where $N = 2000$, MSE was calculated to be 0.0421, while RE was 0.0116. When the number of vehicles was increased to $N = 3000$, both MSE and RE slightly increased to 0.0792 and 0.0196 respectively. This demonstrates that our methods maintains its robustness even when applied to high dimensional data.

Fig. 7 presents the final reconstructed density for two values of N, demonstrating a strong match with the final density obtained by solving the PDE (2) using a Godunov scheme. Notably, when N is increased to 4000, the reconstruction captures the oscillations due to stop-and-go dynamics with even greater precision, nearly perfectly matching the solution.



(a) Neural Network learning using 5,000 training epochs (b) Test simulation using reconstructed density

Fig. 6: Comparison between prediction results and observed data with $N = 3000$



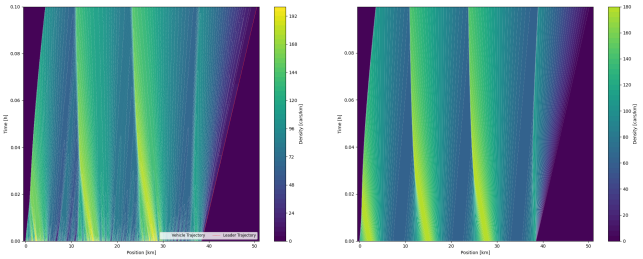
(a) $N = 2000$, $T = 0.1$ (b) $N = 4000$, $T = 0.1$

Fig. 7: Reconstructed final traffic density

Fig. 8 shows that the model can achieve an accurate reconstruction of traffic density over time, even when starting from heterogeneous initial conditions.

VII. Conclusions and Perspectives

In this article, we have developed a model that combines traditional traffic flow models with a data-driven approach. Our method is able to perform effectively with limited data and sparse sensor information (only the starting and ending positions of small number of vehicles) while maintaining low computational complexity. We avoid the need for real-time updates that often require high-frequency data inputs and can be expensive. we opted to use exclusively data generated by traffic FtL model (1) rather than real-world data. The main reason lies on the fact that simulated data allow us to precisely replicate the assumptions and conditions required for the theoretical convergence analysis. Our simulation focused on the use of artificial data generated by microscopic traffic models due to its suitability for theoretical convergence analysis. Indeed, we showed that simulated data allow us to precisely replicate the assumptions and conditions required for demonstrating the convergence of our model to the macroscopic LWR model. The methodology developed remains applicable to real-world data as a key strength of our model is its ability to integrate both artificial and real data. The use of real data would represent an important next step in validating our approach under realistic traffic conditions.



(a) Reconstructed traffic density (b) Godunov traffic density

Fig. 8: Comparison between the predicted density and Godunov scheme density with $N = 3000$ and $T = 0.1$

References

[1] G. Albi, N. Bellomo, L. Fermo, S.-Y. Ha, J. Kim, L. Pareschi, D. Poyato, and J. Soler. Vehicular traffic, crowds, and swarms: From kinetic theory and multiscale methods to applications and research perspectives. *Mathematical Models and Methods in Applied Sciences*, 29(10):1901–2005, 2019.

[2] M. Barreau, M. Aguiar, J. Liu, and K. H. Johansson. Physics-informed learning for identification and state reconstruction of traffic density. *arXiv preprint arXiv:2103.13852*, Sep 2021.

[3] N. Bellomo and C. Dogbe. On the modeling of traffic and crowds: A survey of models, speculations, and perspectives. *SIAM Review*, 53(3):409–463, 2011.

[4] R. M. Colombo and P. Goatin. A well-posed conservation law with a variable unilateral constraint. *Journal of Differential Equations*, 234(2):654–675, Mar 2007.

[5] M. Di Francesco and M. D. Rosini. Rigorous derivation of nonlinear scalar conservation laws from follow-the-leader type models via many particle limit. *arXiv preprint arXiv:1404.7062*, Sep 2015.

[6] J. C. Herrera, D. B. Work, R. Herring, X. J. Ban, Q. Jacobson, and A. M. Bayen. Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4):568–583, 2010.

[7] H. Holden and N. H. Risebro. The continuum limit of follow-the-leader models: A short proof. *arXiv preprint arXiv:1709.07661*, Sep 2017.

[8] D. Inzunza and P. Goatin. A pinn approach for traffic state estimation and model calibration based on loop detector flow data. In *8th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pages 1–6, Jun 2023.

[9] J. Liu, M. Barreau, M. Cicic, and K. H. Johansson. Learning-based traffic state reconstruction using probe vehicles. *arXiv preprint arXiv:2011.05031*, Nov 2020.

[10] B. Piccoli, A. Tosin, et al. Vehicular traffic: A review of continuum mathematical models. *Encyclopedia of Complexity and Systems Science*, 22:9727–9749, 2009.

[11] T. Seo, A. M. Bayen, T. Kusakabe, and Y. Asakura. Traffic state estimation on highway: A comprehensive survey. *Annual Reviews in Control*, 43:128–151, 2017.