Learning-based Traffic State Reconstruction using Probe Vehicles [★]

John Liu [*] **Matthieu Barreau** [*,**] **Mladen Čičić** [*]
Karl H. Johansson [*]

[*] *Division of Decision and Control Systems, KTH Royal Institute of Technology Stockholm, Sweden (e-mail: barreau,johnliu,cicic,kallej@kth.se).*
[**] *Corresponding author*

Abstract: This article investigates the use of a model-based neural network for the traffic reconstruction problem using noisy measurements coming from Probe Vehicles (PV). The traffic state is assumed to be the density only, modeled by a partial differential equation. There exist various methods for reconstructing the density in that case. However, none of them perform well with noise and very few deal with lagrangian measurements. This paper introduces a method that can reduce the processes of identification, reconstruction, prediction, and noise rejection into a single optimization problem. Numerical simulations, based either on a macroscopic or a microscopic model, show good performance for a moderate computational burden.

Keywords: Modeling, Control and Optimization of Transportation Systems; Freeway Traffic Control; Connected and Automated Vehicles

1. INTRODUCTION

Reducing traffic emissions while keeping the travel time low is undoubtedly a major concern (Ferrara et al., 2018). Controlling the traffic is one solution that does not imply a modification of the road infrastructure. However, the current control laws (Piacentini et al., 2018; Čičić and Johansson, 2019) require an estimation of the traffic density.

To reconstruct the density, one can derive a backstepping observer (Yu et al., 2019). However, this does not allow the observation of mixed regimes (both free and congested roads) which is the critical phase for traffic control. Dynamic boundary observers as the one by Qi et al. (2018) can correct this issue. However, the state reconstruction is poor and it requires a huge amount of measurements.

One solution to get more measurements without affecting the traffic infrastructure is to use data from PV (Herrera and Bayen, 2010). This is possible thanks to GPS measurements (Amin and al., 2008) or probing vehicles (Seo and Kusakabe, 2015). An empirical reconstruction using PV has been done by Čičić et al. (2020). It is simple and cannot effectively deal with sparse or noisy measurements. Unfortunately, this is often the case in real-world situations when the capacity for data collection may be restrictive.

Instead, one can perform density estimation using a traffic model. The estimation problem then boils down to solving this system of equations. This was solved mathematically by Delle Monache et al. (2019) where the authors de-

veloped a numerical scheme to estimate the bulk traffic density. These results were further extended by Barreau et al. (2020) where the problem was considered from a control perspective. The obtained observer uses a similar model for traffic flow with an exponential or finite-time convergence of the error. These methods indeed solve the traffic flow problem with noiseless data. However, they encounter complications when dealing with inaccurate measurements, even for the simplest traffic flow model.

Raissi et al. (2019) present an alternative model-based approach using a physics informed neural network to solve Partial Differential Equations (PDE). The authors introduce a mechanism into the neural network such that it is capable of solving a PDE. From a control perspective, this can be seen as a model-based neural network and its usefulness for traffic state reconstruction has been discussed by (Huang and Agarwal, 2020) for instance. We intend to further investigate this problem with a special focus on noise rejection, prediction, and concrete application using a high-fidelity microscopic traffic simulator.

In this paper, we use a physics informed neural network to estimate the density. Compared to the literature, we are using measurements from PVs in two traffic scenarios. In the first case, we have full knowledge of the PDE which defines the physics of the simulation while in the second one, we aim at identifying the appropriate model for the system to estimate the traffic density simulated with SUMO. Another contribution of this paper lies in that we use a neural network to reduce the problem of identification, reconstruction, prediction, and noise rejection into one single optimization problem.

In Section 2, we discuss several micro- and macroscopic models for traffic flow and then present a coupled micro-

macro model as well as formulate the general problem that we intend to solve. In Section 3, we describe the fundamentals of learning. We then present the reconstruction results for a finite difference scheme and an external microscopic simulation in Section 4 and, lastly, offer a few concluding remarks in Section 5.

2. TRAFFIC FLOW THEORY

There are two main approaches to model traffic flow, namely microscopic and macroscopic (Ferrara et al., 2018). A microscopic model defines the dynamic of each vehicle in a traffic stream. This enables us to observe specific phenomena that are exclusively due to the behavior of individual drivers. The main drawbacks are the high computational demand and large parameter space.

On the other hand, a macroscopic model defines the dynamics employing aggregate variables in a traffic flow system. This reduces the computational load and model complexity. The compromise is that a macroscopic model can only simulate general behaviors in traffic flow rather than individual interactions between drivers.

In this section, we present first-order models for microscopic and macroscopic simulations. We then combine them into a coupled micro-macro model which we use to define the problem formulation.

2.1 Microscopic Model

A common model for microscopic simulation is the first order Follow-the-Leader (FL1) model which was rigorously formulated by Argall et al. (2002):

$$\begin{cases} \dot{x}_i(t) = V(\rho_i(t)), & i = 1, 2, \ldots, n, \\ \dot{x}_{n+1}(t) = V_L(t), \end{cases} \quad (1)$$

with appropriate initial positions. It consists of one leading vehicle at position x_{n+1} driving with speed $V_L(t)$ and n following vehicles at positions x_1, x_2, …, x_n. The dynamics of the following vehicles are described by the velocity function $V \in C([0,1], [0, V_{max}])$, where

$$\rho_i(t) = \frac{l_n}{x_{i+1}(t) - x_i(t)} \in [0,1] \quad (2)$$

is the normalized density between x_{i+1} and x_i. To avoid collisions $x_{i+1} - x_i > l_n$ must hold. It is satisfied through the inverse dependency between the relative density and the difference in position as long as the inequality is true for the initial positions. Moreover, $l_n = \frac{m}{n}$ is the average length of the vehicles and m is the total mass of the all vehicles. The velocity of the leading vehicle $V_L(t)$ can be chosen freely such that $V_L \geq 0$, to avoid non-physical behaviors. Furthermore, the velocity function $V(\rho)$ needs to be a bounded, non-negative and non-increasing function with $V(1) = 0$ (a fully congested road results in zero velocity). In this paper, we only consider the normalized density meaning that ρ is dimensionless.

Notice that in (1), the number of parameters increases with the number of vehicles, exposing the problem to the curse of dimensionality. In the FL1 model, the dynamics of the system are defined by how the position of vehicles changes over time. Through this assumption, the model is easy to implement in simulations, however, due to how humans perceive traffic in real life, it is natural to consider how the velocity of vehicles evolves. This would imply to derive a second-order microscopic model that is not described in this paper but used in the simulation software.

For high-fidelity simulations, we use SUMO: an open-source software for microscopic traffic simulations. It is the short name for 'Simulation of Urban MObility' using second-order follow-the-leader model. It is used to construct a multimodal road network with various features such as junctions, route assignments, and car communication (Behrisch et al., 2011). In this paper, SUMO is used to build a single-lane road with an attached driveway in the middle and a stochastic influx of vehicles. Moreover, a traffic light is placed on the highway, giving rise to congestion when vehicles from the driveway enter the highway. Nevertheless, for larger traffic systems, the microscopic model is, by design, limited computationally. As an alternative, we can use a macroscopic model that does not have the same restrictions.

2.2 Macroscopic Model

A macroscopic model uses three quantities to describe the temporal development of a traffic system, namely relative density ρ, average velocity v, and flow rate q, which all depend on time t and space x. The relative density ρ is the number of vehicles within a unit of road space normalized over the interval $[0,1]$, where $\rho = 0$ is interpreted as an empty road and $\rho = 1$ is a bumper-to-bumper road. Average velocity v is the mean speed of vehicles on a road segment. Flow rate q is the number of vehicles passing through a cross-section of the road during a time period. These three quantities are related through the hydrodynamic equation:

$$q = \rho v. \quad (3)$$

The conservation of mass leads to the following equation:

$$\rho_t + q_x = 0, \quad (4)$$

Finally, we substitute (3) for the flow rate q in (4) resulting in the continuity equation for traffic flow:

$$\rho_t + (\rho v)_x = 0. \quad (5)$$

The dynamics of Lighthill-Whitham Richards (LWR) model (Lighthill and Whitham, 1955) are defined by

$$\rho_t + (\rho V(\rho))_x = 0, \quad (6)$$

where $V(\rho)$ is the equilibrium velocity analogous to the velocity function in the FL1 model. An important distinction is that $v(t, x)$ is the average velocity at a time t and a position x while $V(\rho)$ is the velocity in an equilibrium state where the density is ρ. In the LWR model, it is assumed that $v(t, x) = V(\rho(t, x))$. Additionally, note that the velocity function in the FL1 model is not necessarily differentiable while the equilibrium velocity in LWR is required to be differentiable for a well-defined macroscopic model.

However, in this paper, we do not consider the system dynamics to be described by the continuity equation in (6) due to the discontinuous nature of its solutions. Instead, as suggested by Nelson (2002), we assume that the traffic flow is characterized by the following parabolic PDE:

$$\rho_t + (\rho V(\rho))_x = \gamma \rho_{xx}, \quad (7)$$

where $0 < \gamma \ll 1$ is a diffusion correction parameter. As $\gamma \to 0$, the solution of (7) converges, in a distributive sense, to the solution of the continuity equation in (6), which has been proven by Hopf (1950). Equation (7) now exhibits a unique smooth solution which is favorable from a modeling and analysis point of view. Admittedly, equation (7) violates the anisotropic property of traffic flow but at the same time enables the possibility to model other traffic flow phenomena for which the LWR model has been proven to be deficient (Nelson, 2002).

The main weakness of a first-order macroscopic model is the assumption that the system is always in an equilibrium state. Thus, the LWR model is unable to predict phenomena such as capacity drop (Kontorinakia et al., 2017). This paper is a preliminary study and therefore, we do not consider higher order traffic models.

2.3 Coupled Micro-Macro Model

Both the microscopic and macroscopic models attempt to describe traffic flow, although, from separate standpoints they are not completely uncorrelated. Colombo and Rossi (2014) showed that the density solution in the FL1 model ρ_i in (2) converges to the density solution in the LWR model ρ, as the number of vehicles n tends toward infinity while the total vehicle mass m in the FL1 model remains constant. Specifically, the solution in the FL1 model describes the trajectory solution for the LWR model.

$$\lim_{n \to 0} \rho_i(t) = \rho(t, x_i) \tag{8}$$

To exploit this relationship, we use a Coupled Micro-Macro (CMM) model, where we consider the parabolic PDE in (7), as a substitute for the LWR model, and the equilibrium velocity given by (1) from the FL1 model:

$$\begin{cases} \rho_t + (\rho V(\rho))_x = \gamma \rho_{xx}, \\ \dot{x}_i = V(\rho(x_i)), \qquad i = 1, 2, \dots, n. \end{cases} \tag{9}$$

In this paper, we study the model in (9). The following subsection states the problem in a general context.

2.4 Problem formulation

The objective of this paper is to reconstruct the true value of the desired quantity from a sparse data set.

General problem   Let $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ be a semi-normed vector space, $z \in \mathcal{H}$ the desired quantity and $\mathcal{H}_c \subseteq \mathcal{H}$. We say that $\hat{z}$ is a **Partial-State Reconstruction** (PSR) of z if

$$\hat{z} \in \mathcal{R}, \quad \mathcal{R} = \operatorname*{argmin}_{\bar{z} \in \mathcal{H}_c} \|z - \bar{z}\|_{\mathcal{H}}^2, \tag{10}$$

where $\mathcal{R}$ is the partial-reconstructed set.

This problem is a special case of the one introduced by Barreau et al. (2021). The authors point out there that if $z \in \mathcal{H}_c$ then $\hat{z} = z$ weakly. In contrast, if $z \notin \mathcal{H}_c$ and H_c is a vector space, a PSR $\hat{z}$ will be the orthogonal projection of z in $\mathcal{H}_c$, which is assumed to be non-empty.

Traffic flow problem   In this paper, we investigate the PSR of the density solution ρ of the LWR model. However, the density ρ is assumed to follow the parabolic equation resulting in the density reconstruction $\hat{\rho}$ being a solution of (7). The work of Hopf (1950) proves the existence of a solution to (7) which enforces the set $\mathcal{H}_c$ to be non-empty.

The reconstruction is performed using density measurements in time and space. They are collected over a time window $[0, T]$ by Probe Vehicles (PV), with position $x_1, x_2, \dots, x_n$, that interact within a traffic stream obeying the ODE in (9). Specifically, we are interested in the PSR in the region between the first and last PV (where the PV trajectories is interpreted as boundary conditions to (7)):

$$\Omega_r = \{(t, x) \in \mathbb{R}^+ \times \mathbb{R} \mid$$
$$x \in [x_1(t), x_n(t)], \ \forall t \in [0, T + \Delta T]\}, \tag{11}$$

which we consider to be the reconstruction domain. Note that we intend to predict the future density as well, where ΔT is the horizon of the future prediction.

In relation to the general problem, $\mathcal{H} = H^1(\Omega_r, [0, 1])$ and $\|\cdot\|_{\mathcal{H}}$ is the traditional $\mathcal{L}_2$-norm. Additionally, $\mathcal{H}_c = C^\infty(\Omega_r, [0, 1])$ is bounded due to the assumption in (7). Note that the model only considers continuous functions while the density solution can exhibit discontinuous tendencies. Furthermore, measurements are only collected along PV trajectories, meaning that we are, effectively, attempting to solve the PDE in (7) without initial conditions. As a result, the reconstruction set $\mathcal{R}$ is most likely not a singleton. To solve this optimization problem, we utilize a neural network as described in the following section.

3. MACHINE LEARNING

Machine learning methods are data-based approaches for pattern recognition and have been applied in numerous fields of research (Goodfellow et al., 2016). Learning is a data-based approach to modeling non-linear relationships. The structure of a Deep Neural Network (DNN) was loosely inspired by the neural processes in the human brain. Although DNN is an outdated model for the human brain it has proven to be useful when trying to predict complex patterns such as solutions to PDEs.

3.1 Fundamentals of Neural Networks

A DNN consists of processing units, referred to as neurons, which form fully connected sequential layers. Each neuron uses the output of all neurons from the previous layer as the input and returns a single value as the output.

Consider the ith layer with input $\mathbf{x}_i \in \mathbb{R}^{n_{i-1}}$, weights $W_i \in \mathbb{R}^{n_i \times n_{i-1}}$ and biases $\mathbf{b}_i \in \mathbb{R}^{n_i}$, where n_i is the number of neurons in the layer. The weights determine the influence of $\mathbf{x}_i$ on each neuron and the biases act as offset values. The output of the layer $\mathbf{y}_i$ is a linear function of $\mathbf{x}_i$ subject to an activation function $\mathcal{A} : \mathbb{R} \to \mathbb{R}$:

$$\mathbf{y}_i = \mathcal{A}(W_i \mathbf{x}_i + \mathbf{b}_i), \tag{12}$$

where $\mathcal{A}$ is operating elementwise (Goodfellow et al., 2016). The input for the following layer is then $\mathbf{x}_{i+1} = \mathbf{y}_i$.

The input to the entire DNN, in our case, is the time t and position x and the output is $\hat{\rho}(t, x)$, the estimated density.

Network Size   The network size of a DNN reflects the complexity of the model for the system and is one of the main trade-offs when constructing a DNN. If the network size is insufficiently large, the DNN will be unable to fully model the system. On the other hand, if the network size

is overestimated, the neural network will model individual data points instead of the underlying structure of the system resulting in overfitting (Goodfellow et al., 2016). To counteract this, we can introduce a regularization agent to restrict the region in parameter space. This is further discussed in Subsection 3.2.

Estimation Cost Function   A cost function is constructed to measure the inaccuracy of the model estimation in relation to the reference value. A common type of cost function for regression networks is a quadratic discriminant function of which the Mean Squared Error (MSE) is the most popular. Suppose we are given a training set $\{(t_i, x_i, \rho_i)\}$ with N_{est} estimation points. The MSE estimation cost function is given by

$$\mathcal{J}_{\text{est}} = \text{MSE}_{(\rho, \hat{\rho})} = \frac{1}{N_{\text{est}}} \sum_{i=1}^{N_{\text{est}}} |\rho_i - \hat{\rho}(t_i, x_i)|^2. \qquad (13)$$

The cost is minimized by determining the optimal weights and biases. There exist many procedures developed to perform parameter optimization.

The most common ones are based on the gradient descent algorithm. They are only capable of finding a global minimum if the cost function is strictly concave. Consequently, for an arbitrary smooth cost function, the gradient descent algorithm lacks mechanics to differentiate a local from a global minimum. Nevertheless, this is not a crucial concern since it has been shown by Choromanska et al. (2015) that local minimums have values close to the global minimum if the complexity of the neural network is sufficiently large. However, the neural network might overfit the dataset. Thus, a regularization agent, as described in the following subsection, is needed.

3.2 Physics Informed Learning

Physics informed learning (Raissi et al., 2019; Sirignano and Spiliopoulos, 2018) is based on the same principle as the traditional learning but extended to enforce the dynamics of the system, similar to model-based learning.

Physics Cost Function   Specifically, we want to ensure that the neural network acts accordingly to a physics property on chosen points in the time-space domain. Let N_{phy} physics points $\{(t_j, x_j)\}$, randomly selected through latin hypercube sampling on Ω_r. Assume that the physical system obeys the relation $f(t, x, \rho) = 0$ on its domain. We can now define a physics cost function in (14) using the physics property f.

$$\mathcal{J}_{\text{phy}} = \text{MSE}_{(0, f)} = \frac{1}{N_{\text{phy}}} \sum_{j=1}^{N_{\text{phy}}} |f(t_j, x_j, \hat{\rho})|^2 \qquad (14)$$

The total cost function $\mathcal{J}$ is constructed as the sum of the estimation cost and physics cost weighted by $\mu \in [0, 1]$:

$$\mathcal{J} = \mu \mathcal{J}_{\text{est}} + (1 - \mu) \mathcal{J}_{\text{phy}}. \qquad (15)$$

The neural network is forced to make a trade-off between minimizing the estimation cost and the physics cost when optimizing the parameters Θ depending on the weight μ. Qualitatively, using $\mu = 0$ means that we are trying to solve a differential equation without initial or boundary conditions. It is unhelpful to consider this case since $\hat{\rho} \equiv 0$ is usually an optimal solution. In contrast, using $\mu = 1$ is

equivalent to a traditional learning, which can result in a poor reconstruction in regions with insufficient measurements. Hence, an appropriate choice of μ is required for a desirable reconstruction. We used $\mu = 0.5$ similar to Raissi et al. (2019).

From a machine learning perspective, the physics cost acts as a regularization agent. The neural network aims at accurately estimating the measurement data under the constraint that the physics cost needs to remain low. Importantly, the physics cost is independent of the measurements and depends only on the parameter tensor Θ. This additional cost can be interpreted as introducing artificial measurements. Hence, it is an efficient solution to prevent overfitting. Moreover, in theory, it is a model-based approach which deters non-physical solutions and, therefore, it is more robust to noise in the measurements.

3.3 Traffic flow problem

For the traffic flow problem, we have the training set $\{(t_i, x_i, \rho_i)\}$, from PV following the ODE in (9). Consequently, the measurements are not well-spread in the reconstructed domain Ω_r, indicating that using physics informed learning could be advantageous compared to traditional learning (Huang and Agarwal, 2020). We assume the vehicle dynamics are described by (9) such that

$$f(t, x, \hat{\rho}) = \hat{\rho}_t(t, x) + (\hat{\rho}V(\hat{\rho}))_x(t, x) - \gamma \hat{\rho}_{xx}(t, x) \qquad (16)$$

is the physics property of the neural network. The total cost function for the traffic flow problem is constructed according to (13), (14) and (15).

Next, we present the reconstruction results from a finite difference scheme and a SUMO simulation using physics informed learning with $\mu = 0,5$ giving the same importance to each cost.

4. RESULTS

In this paper, we have attempted to reconstruct the traffic density from a finite difference scheme and a SUMO simulation. In the finite difference scheme, the vehicle dynamics were characterized by the equation in (7), where the equilibrium velocity was based on the famous Greenshields equation (Greenshields et al., 1935). In the SUMO simulation, vehicles follow a variation of a second-order microscopic model, which is capable of simulating numerous traffic flow phenomena (Krauß, 1998). We have also investigated the reconstruction results for both simulations when noise was added to the measurements.

4.1 Finite Difference Scheme

For the finite difference scheme, the vehicle velocity was determined by Greenshields equation:

$$V(\rho) = v_f(1 - \rho), \qquad (17)$$

where v_f is the vehicle velocity without interference, also called the free flow velocity. The true density solution is shown in Fig. 1 using a spatiotemporal diagram. The colors display the density and the black points indicate the measurement points as well as the PV trajectories.

The reconstruction result when using traditional learning is not shown here since a similar study has been conducted
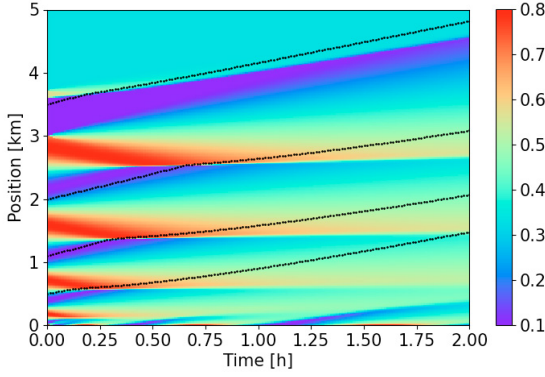
Fig. 1. Spatiotemporal diagram of the true density with a finite-difference scheme ($\gamma = 0.05$). The points display the measurement points and the PV trajectories.
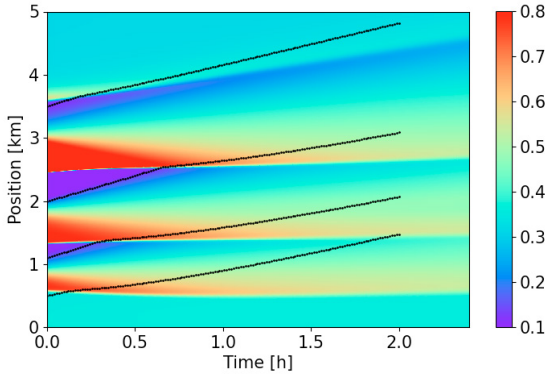

Fig. 2. Spatiotemporal diagrams of the traffic state reconstruction using physics informed deep learning. The points are measurement points along the PV trajectories and noise is added.
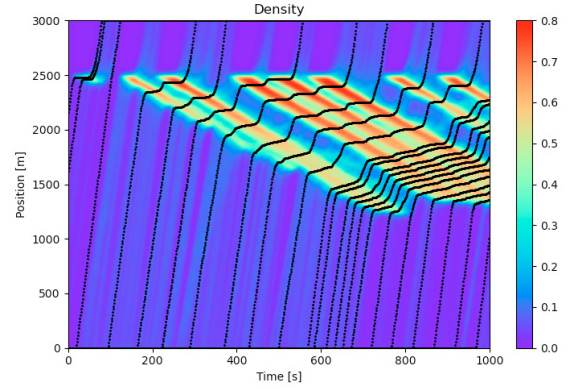

Fig. 3. Spatiotemporal diagram of the true density for the SUMO simulation. The black points display the measurement points and the PV trajectories. No noise is added in this case.
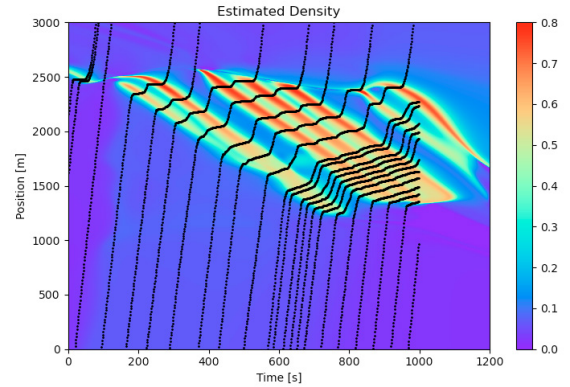

Fig. 4. Spatiotemporal diagrams of the physics informed reconstruction with different models. The black points are measurement points along the PV trajectories.
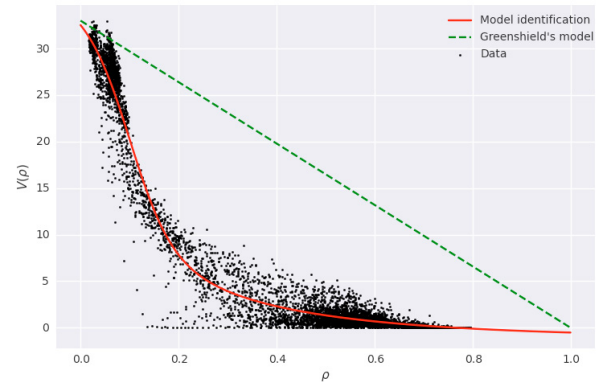

Fig. 5. Equilibrium velocity resulting from Greenshields equation and model identification. The black points are empirical measurements.

by Huang and Agarwal (2020). The conclusion was that there is a perfect reconstruction of the density along the PV trajectories, however, it is far from the true density in between. Specifically, the neural network can predict a positive characteristic speed in congested regions.

If the number of probe vehicles is increasing, the reconstruction becomes more precise, even in the case of $\mu = 1$.

For the physics informed reconstruction, in Fig. 2, we can observe that the results are nearly indistinguishable from the true density for later times t, with an $\mathcal{L}_2$-error of 0.17. This showcases the ability of physics informed learning to compensate for insufficient measurement data. The neural network is also capable of performing accurate reconstruction on future density. The characteristics of the shock waves are preserved at all times in the reconstruction domain. Physics points are sampled throughout the entire domain enforcing the physical property for $t \in [T, T + \Delta T]$.

Admittedly, there are some discrepancies in the noisy reconstruction, specifically at former times t. Nevertheless, this is to be expected since the neural network is unaware of the initial state of the system. From a control point of view, this is not an issue since the reconstruction of the past does not aid our ability to control the traffic flow.

4.2 SUMO Simulation

In the SUMO simulation, vehicles interact on a highway with a connected driveway and a traffic light as a mediator. Vehicles on the highway are stopped by the traffic light as vehicles from the driveway enter which induces congestion. The vehicles in the SUMO simulation follow a second-order microscopic model based on collision prevention. The true density is shown in Fig. 3.

In the SUMO case, we performed model identification using traditional learning beforehand to estimate the equilibrium velocity. Velocity measurements $\{v_i\}$ were constructed through numerical differentiation using the position and time measurements $\{t_i, x_i\}$ and are used only in this step. Fig. 5 shows that the estimated velocity is significantly more accurate compared to the Greenshields equation. The neural network in that case is even capable of making future predictions, where the density characteristics are maintained. The major discrepancy in the reconstruction is around $x = 2500$ m. The reason is most likely due to the unlabelled presence of the traffic light.

The reconstruction with added Gaussian noise using $\sigma_\rho = 0.001$ is not shown here but it has a chaotic density which is considerably less accurate compared to the noiseless reconstruction. This may be due to insufficient model complexity but it could also be model instability. Further investigations are required to fully understand the reconstruction results with added noise.

5. CONCLUSION

In this paper, we have introduced a model-based approach for traffic state reconstruction using a neural network and presented the results for a finite difference scheme and a SUMO simulation. With the leverage of machine learning our method can distill the problem of identification, reconstruction, and prediction into one optimization problem. Additionally, the method is robust when introducing unbiased noise in density given a sufficient understanding of the system dynamics. Future work will be devoted to extend the method using 2nd order models and develop tools to counteract trajectory noise.

REFERENCES

Amin, S. and al. (2008). Mobile century using GPS mobile phones as traffic sensors: A field experiment. In *15th World Congress on Intelligent Transportation Systems.*

Argall, B., Cheleshikin, B., Greenberg, J., Hinde, C., L., and P.-J. (2002). A Rigorous Treatment of a Follow-the-Leader Traffic Model with Traffic Lights Present. *SIAM J. Appl. Math.*, 63(1), 149–168.

Barreau, M., Liu, J., and Johansson, K.H. (2021). Learning-based state reconstruction of a scalar hyperbolic PDE under noisy Lagrangian sensing. In *L4DC.*

Barreau, M., Selivanov, A., and Johansson, K.H. (2020). Dynamic traffic reconstruction using probe vehicles. In *Conference on Decision and Control (CDC).*

Behrisch, M., Bieker, L., Erdmann, J., and Krajzewicz, D. (2011). SUMO – Simulation of Urban MObility; An Overview. *German Aerospace Centre, Institute for Transportation Research.*

Choromanska, A., Henaff, M., Mathieu, M., Arous, G.B., and LeCun, Y. (2015). The Loss Surfaces of Multilayer Networks. *Proceedings of Machine Learning Research.*

Čičić, M., Barreau, M., and Johansson, K.H. (2020). Numerical investigation of traffic state reconstruction and control using connected automated vehicles. In *23rd IEEE International Conference on Intelligent Transportation Systems.* Rhodes, Greece.

Čičić, M. and Johansson, K.H. (2019). Stop-and-go wave dissipation using accumulated controlled moving bottlenecks in multi-class CTM framework. In *Conference on Decision and Control (CDC).* Nice, France.

Colombo, R. and Rossi, E. (2014). On the micro-macro limit in traffic flow. *Rendiconti del Seminario Matematico della Università di Padova*, 131.

Delle Monache, M.L., Liard, T., Piccoli, B., Stern, R., and Work, D. (2019). Traffic reconstruction using autonomous vehicles. *SIAM Journal on Applied Mathematics*, 79(5), 1748–1767.

Ferrara, A., Sacone, S., and Siri, S. (2018). *Freeway Traffic Modelling and Control.* Springer Nature.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning.* MIT Press.

Greenshields, B.D., Bibbins, J.R., Channing, W.S., and Miller, H.H. (1935). A study of traffic capacity. In *Highway research board proceedings*, volume 14, 448–477. National Research Council (USA).

Herrera, J.C. and Bayen, A.M. (2010). Incorporation of Lagrangian measurements in freeway traffic state estimation. *Transportation Research Part B: Methodological*, 44(4), 460–481.

Hopf, E. (1950). The partial differential equation $u_t + uu_x = \mu u_{xx}$. *Communications on Pure and Applied mathematics*, 3(3), 201–230.

Huang, J. and Agarwal, S. (2020). Physics Informed Deep Learning for Traffic State Estimation. *IEEE Intelligent Transportation Systems Conference.*

Kontorinakia, M., Spiliopouloub, A., Roncolic, C., and Papageorgioub, M. (2017). First-order traffic flow models incorporating capacity drop: Overview and real-data validation. *Transportation Research Part B: Methodological*, 106, 52–75.

Krauß, S. (1998). Microscopic Modeling of Traffic Flow: Investigation of Collision Free Vehicle Dynamics. *Deutsches Zentrum für Luft– und Raumfahrt.*

Lighthill, M. and Whitham, G. (1955). On kinematic waves ii. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London*, 229(1178).

Nelson, P. (2002). Traveling-wave solutions of the diffusively corrected kinematic-wave model. *Mathematical and computer modelling*, 35(5-6), 561–579.

Piacentini, G., Goatin, P., and Ferrara, A. (2018). Traffic control via moving bottleneck of coordinated vehicles. *IFAC-PapersOnLine*, 51(9), 13 – 18. 15th IFAC Symposium on Control in Transportation Systems CTS 2018.

Qi, R., Hao, J., and Zhang, L. (2018). Boundary multi-mode observer for Aw-Rascle-Zhang traffic flow model. In *2018 37th Chinese Control Conference (CCC).*

Raissi, M., Perdikaris, P., and Karniadakis, G. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 686–707.

Seo, T. and Kusakabe, T. (2015). Probe vehicle-based traffic state estimation method with spacing information and conservation law. *Transportation Research Part C: Emerging Technologies*, 59, 391–403.

Sirignano, J. and Spiliopoulos, K. (2018). DGM: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics.*

Yu, H., Bayen, A.M., and Krstic, M. (2019). Boundary observer for congested freeway traffic state estimation via Aw-Rascle-Zhang model. *IFAC-PapersOnLine*, 52(2).