A physics-informed machine learning for generalized bathtub model in large-scale urban networks

Eunhan Ka, Jiawei Xue, Ludovic Leclercq, Satish Ukkusuri

A Physics-Informed Machine Learning for Generalized Bathtub Model in Large-Scale Urban Networks

Eunhan Ka[a], Jiawei Xue[a], Ludovic Leclercq[b], Satish V. Ukkusuri[a,*]

[a]*Lyles School of Civil Engineering, Purdue University, West Lafayette, 47907, Indiana, United States*
[b]*Univ Gustave Eiffel, ENTPE, LICIT-ECO7, Lyon, F-69675, France*

Abstract

Traffic management strategies contribute to alleviating urban traffic congestion by improving the efficiency of urban road networks. Network traffic flow models are vital to improving the effectiveness of traffic management strategies by estimating traffic states and describing traffic dynamics. Despite having robust theoretical foundations, existing network traffic flow models struggle to model complex and dynamic real-world traffic data - especially the variance and heterogeneity in large-scale urban networks. These challenges arise from both the inherent dynamics of traffic flows and external factors such as changes in travel demand and traffic control. Many studies used machine learning (ML) methods to estimate traffic states with high accuracy, but ML methods have limited interpretations since the relationship between the variables is not explicitly visible. To ease these limitations, we propose a hybrid physics-informed machine learning model with a generalized bathtub model (PIML-GBM), which leverages the interpretability of physical models and ML methods for their powerful modeling ability. This study tests the proposed PIML-GBM on mobile location data and a large-scale road network in Indianapolis, United States. The experimental results show that the proposed PIML-GBM model has superior accuracy and interpretability in estimating traffic state over existing algorithms.

Keywords: Network Traffic Flow Model, Physics-Informed Machine Learning, Generalized Bathtub Model, Large-Scale Urban Network, Mobile Location Data

1. Introduction

Urban traffic congestion occurs due to limited road network supply (such as capacity and the number of lanes) and spatio-temporal variations of travel demand. Urban congestion causes travel time losses, traffic accidents, and wasted fuel leading to severe air pollution [1]. From 2000 to 2019, the yearly delay time per auto commuter increased by about 42% (from 38 to 54 hours), the wasted fuel per auto commuter increased by about 47% (from 15 to 22 gallons), and the national congestion cost increased by 2.46 times (from $ 77

*Corresponding author. E-mail address: sukkusur@purdue.edu (S.V. Ukkusuri).

to \$ 190 billion) in the United States [2]. In the short term, constructing new infrastructure and extensions to existing infrastructure may not be optimal solutions to reduce congestion due to the tremendous cost and growth in travel demand induced by increasing road capacity. The efficient way to alleviate traffic congestion is to make better use of existing road capacity and improve the efficiency of large urban road networks through various traffic management strategies. The first step for developing these management strategies is to accurately describe the traffic states (i.e., traffic speed, density, and flow) and identify their relationships with travel demand in urban areas.

1.1. Network Traffic Dynamics Models

Network-level traffic models have been studied and advanced significantly after the macroscopic fundamental diagram (MFD) was experimentally proven in urban areas using loop detector data (LDD) and microsimulation [3, 4]. The MFD describes the stable relationship between average travel speed and average vehicle density in homogeneous urban areas [5, 6]. Daganzo (2007) proposed the equilibrium condition of the road network as an ordinary differential equation (ODE) to approximate traffic dynamics [7], as shown in Figure 1. The proposed accumulation-based model described a road network as a single reservoir system of traffic flows based on MFD: $\frac{dn(t)}{dt} = f(t) - G(n(t))$, where $n(t)$ is the number of vehicles in the road network, $f(t)$ is the inflow rate, and $G(n(t))$ is the outflow rate at time t. In the accumulation-based model, the outflow rate is derived from the travel production $P(n(t))$ over the average trip length L (i.e., $G(n(t)) = \frac{P(n(t))}{L}$). For calculating the travel production $P(n(t))$, the average travel speed $v(t)$ is calculated from the stable relationship between the average travel speed and average vehicle density of the MFD. The accumulation-based model has been developed into multiple reservoir systems by various researchers [8, 9, 10, 11].

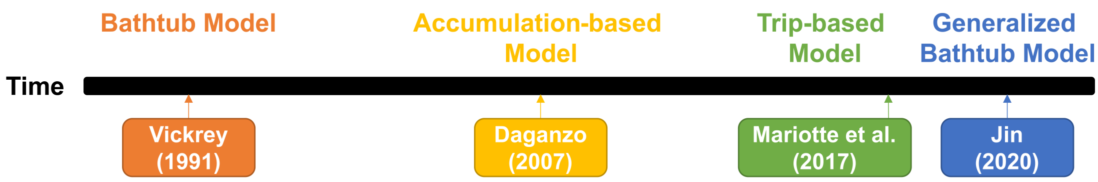

Figure 1: Timeline of Traffic Dynamics Models

Though the accumulation-based model scales down the complexity of traffic models by aggregating individual road segments, it has four drawbacks that undermine its accuracy in expressing heterogeneous traffic dynamics [12]. First, the assumption that all vehicles have constant trip length is not likely to be reasonably practical, curtailing the accuracy of the accumulation-based model in depicting traffic dynamics. In the real world, trip lengths are sensitive to the OD (i.e., origin-destination) matrix and traffic conditions (1st drawback) [13]. Second, there is no sufficient explanation for how far the trip lengths of vehicles have progressed (i.e., there is no space dimension) [14]. Mariotte et al. (2017) found that the absence of space dimension caused inaccuracy of travel time evolution when there was a sudden change in demand (2nd drawback) [15]. Lastly, Geroliminis and Sun (2011) demonstrated the third and fourth limitations of lower accuracy in heterogeneous traffic dynamics: the different degrees of spatial heterogeneity in vehicle density during the onset and offset of the peak period (3rd drawback), and the synchronized occurrence of transient periods and capacity drop in the offset of congestion (4th drawback) [16].

For relaxing the assumption of constant trip length, previous studies proposed the trip-based model, which assumes that each vehicle i has its trip length L_i by accumulating the mean speed obtained from the MFD while it exists in the road network [15, 17], as shown in Figure 1. Lamotte et al. (2018) proposed the M-model, a hybrid model between accumulation-based and trip-based models to incorporate the remaining distance into traffic dynamics models [18]. The M-model described variations in the remaining distance to be traveled into the average remaining distance. These studies relaxed the assumption of constant trip length in accumulation-based models and were tested in either discrete time-based or event-based simulations, not real road networks [15, 17, 18, 19].

The bathtub model, proposed in 1991 but published in 2020 by Vickrey (2020), is another approach to describing traffic dynamics in road networks by extending a single bottleneck model to a network-wide traffic model [20], as shown in Figure 1. In the bathtub model, the process of trips entering (leaving) the traffic network in urban areas is similar to the inflow (outflow) of water into (out of) a bathtub where vehicle density would correspond to the level of water in the bathtub [21]. Since the process of the bathtub model is similar to that of the accumulation-based model, the bathtub model has similar assumptions of the accumulation-based model: conservation of the number of trips, the network-level speed-density relation, and homogeneously distributed congestion. Arnott and Buli (2018) developed a numerical algorithm for the solution of the departure-time user equilibrium, as the equilibrium solution of the bathtub model is generally analytically intractable [22]. This study demonstrated the traffic dynamics of active trips and travel times based on bathtub models under the assumption of constant trip length by delay-differential equations.

For relaxing the strict assumption of constant trip length, Jin (2020) proposed the generalized bathtub model (GBM) in which the trip length distribution is not restricted to a specific distribution function [23], as shown in Figure 1. The GBM differs from the accumulation-based and bathtub models because it assumes that every vehicle has its individual assigned trip length by controlling its speed to the mean speed of the reservoir obtained from the speed-density relation, which is the same as the trip-based model. This assumption enables the GBM to describe flow evolution more accurately than accumulation-based and bathtub models when inflow is abruptly changed. GBM overcomes the limitations of the accumulation-based and bathtub models by modeling the number of vehicles with the information on the remaining trip length at all times under heterogeneous trip lengths.

As bathtub, accumulation-based, and trip-based models can provide the number of vehicles (or traffic density) in road networks, traffic operators can elaborate congestion pricing policies, traffic signal controls, and other traffic management strategies based on GBM, which provides the number of trips with remaining trip lengths. Jin et al. (2021) proposed a modeling framework to describe trip dynamics and traffic congestion in shared mobility systems, as information on remaining distances based on GBM can improve fleet management for shared mobility systems [24]. Since several partial differential equations governed GBM over a time-distance domain (i.e., time is a time instance and distance is a remaining distance), $K(t, x)$ is directly linked to the "unloading process" of the network. GBM formulates a continuous system of traffic dynamics that can be easily discretized for any mesh. As the GBM also uses a stable relationship between traffic density and speed, the estimation of the relationship from MFD has a crucial impact on its performance in modeling traffic dynamics. The bathtub model and GBM have been implemented in hypothetical networks, not real networks. In this study, we use GBM as the traffic flow model to learn physics knowledge since GBM can model heterogeneous trips under a continuous system of traffic dynamics.

1.2. Limitations of Network Traffic Dynamic Models

Although existing models (i.e., GBM, bathtub, accumulation-based, trip-based models) have the advantage of describing traffic dynamics under theoretically ideal conditions, they suffer from two critical challenges related to the variance of real traffic data and calibrations based on real data.

First, existing models cannot capture the variance of traffic data. These variances come from the inaccuracy of data sources (e.g., loop detectors, floating cars, mobile phones), and the stochastic nature of traffic. After Geroliminis and Daganzo (2008) experimentally proved that there are relationships between variables in traffic state in urban areas with the homogeneity assumption [3], Buisson and Ladier (2009) empirically examined the effect of heterogeneity with LDD in Toulouse, France [25]. They showed that heterogeneous traffic flows influence the shape and scatter of MFD. Hence, existing models based on MFD are insufficient to provide rather accurate traffic state estimation results under real-world data. Knoop et al. (2015) proposed a generalized MFD to handle inhomogeneity caused by the variance of real traffic data, which is expressed as the standard deviation of densities in road networks [26]. However, this study was tested on a 4 x 4 toy network, not real road networks.

Second, existing models have difficulty calibrating MFD's parameters with real-world data. The calibration of MFD requires significant observed traffic data, potentially requiring significant infrastructure installations (e.g., loop detectors) and maintenance costs for data collection. Existing simulation-based studies also require that the form and parameters of the speed-density relationship in the MFD to be specified [15, 17, 18, 19, 27, 28, 29].

1.3. Machine Learning Models in Traffic Flow Modeling

To enhance the accuracy of physical models in describing urban traffic, we introduce machine learning (ML) methods. ML methods do not require theoretically ideal assumptions or prior knowledge, and they can directly estimate traffic states from massive traffic data [30]. Some examples of ML methods in transportation studies are k-nearest neighbor [31], convolutional neural networks [32, 33], long short-term memory [34], graph neural networks [35, 36], and reinforcement learning [37]. Compared to physical models, these ML methods can extract nuanced relationships between different traffic state variables considering the variance in traffic data. Moreover, the model training process is becoming more convenient with the advancement of powerful optimizers (e.g., Adam optimizer [38]).

One disadvantage of ML methods is that they are developed as "black boxes" with limited interpretation because the relationship between input features and output labels is not explicitly visible. To overcome these limitations, Raissi et al. (2019) have pioneered the development of physics-informed machine learning (PIML), a methodological advancement that facilitates the integration of physical constraints into deep learning models, thus aligning the computational approach with established physical principles [39]. Recent works proposed PIML models in the transportation area, to harness the advantage of physical traffic models as interpretability and ML methods as their powerful modeling ability [40, 41, 42]. However, these models were applied only to road segments of highways, not to the road network in urban areas. An extension of the network traffic flow model is required to identify urban traffic congestion.

1.4. Data for Modeling Traffic Dynamics

In terms of data sources, most studies estimate the relationship between traffic state variables using LDD. However, these data collection techniques, including loop detector and camera, have some limitations, such as inaccuracy from the detector location [43], the enormous cost of installation, the long computation time for real-time estimation [44], and no trip information (e.g., origin, destination, trip length).

Floating car data (FCD) providing vehicular trajectory data is an alternative to modeling traffic flows with trip information [5, 45]. However, the data sparsity of FCD makes it hard to generate the frequent trips made by travelers (e.g., residents). With big data analytics and advanced data collection techniques, location-based data (LBD) generated from mobile applications is increasingly available and treated as a beneficial data source [44, 46]. LBD has two powerful advantages overwhelming other data sources: comprehensive coverage of the population in the urban network and highly variable sampling spatiotemporal interval [46]. Furthermore, the distinct characteristics of LBD render it particularly apt for implementing GBM, necessitating the distribution of trip length within the road network. Compared to traditional data sources such as LDD and FCD, LBD offers an enriched dataset through its superior coverage and adaptability, traits that are integral for GBM-based models. The capability of LBD to generate comprehensive trip-length distributions, combined with its adaptability to diverse spatiotemporal conditions, not only enhances its applicability but also positions it as a leading data source in the field of contemporary traffic modeling.

1.5. Objectives and Contributions

This study bridges the gap between network traffic flow models and ML methods by overcoming three challenges: (1) the stringent assumptions of physical models, for example, the stable speed-density relationship, (2) the absence of physical relationships between input and output variables in the training process of deep learning models, and (3) the limited applicability of traffic flow models that are tested in hypothetical scenarios instead of real large-scale urban networks.

To achieve this, we propose a PIML model to describe the traffic dynamics for large-scale urban road networks with GBM (PIML-GBM). Specifically, we make the following contributions:

- Estimating and predicting the number of trips with a remaining distance not smaller than x at time t (i.e., $K(t,x)$) using a deep neural network and boundary condition without explicit theoretical assumptions, which is regularized by physics knowledge
- Quantifying the ability of PIML-GBM to capture the randomness and uncertainty of traffic dynamics more accurately from the neural network structure by comparing existing solution methods
- Describing traffic dynamics over a continuous time-distance domain with boundary points of a discretized time-distance domain by using the PIML-GBM
- Implementing GBM to the real road network for estimating traffic states with mobile location-based data.

The remainder of this paper is organized as follows. Section 2 describes the preliminaries of GBM (i.e., definitions of variables and mathematical formulations). Section 3 formulates the framework of PIML-GBM. Section 4 describes details of the collected data, data processing procedure, and experiments of PIML-GBM. Section 5 presents a case study in the Indianapolis network to evaluate the proposed PIML-GBM and discusses the results. Lastly, Section 6 concludes our work and summarizes future research.

2. Preliminaries

We first introduce five types of variables and GBM used in this study, including assumptions and mathematical formulations.

2.1. Definitions of Variables

There are five types of variables: supply variables (network variables), inflow variables, outflow variables, traffic state variables, and active trip variables [23].

2.1.1. Supply variables (network variables)

We consider a road network as a single reservoir, not multiple reservoirs. Supply variables are related to the network structure and network speed-density relationship. The network's total length of road segments is defined as L_{net} and T denotes the total time step. Let $\mathbf{T}$ be a time horizon (i.e., $\mathbf{T} = [0, T]$). The network speed-density relationship follows MFD:

$$v(t) = V(\rho(t)), \tag{1}$$

where $v(t)$ is the average speed of vehicles running on the road network at time $t \in \mathbf{T}$, $V(\cdot)$ is the function of the density based on MFD, and $\rho(t)$ is the average density per unit road length at time t. The corresponding average flow-density relationship can be described by:

$$q(t) = \rho(t)v(t) = \rho(t)V(\rho(t)), \tag{2}$$

where $q(t)$ is the traffic flow at time t.

2.1.2. Inflow variables (demand variables)

Inflow variables are entering trip (in-flow) rates (i.e., $f(t)$, $F(t)$) and distributions of entering trip distance (i.e., $\tilde{\varphi}(t, x)$, $\tilde{\Phi}(t, x)$). These variables represent the demand pattern of entering trips. We denote the entering trip rate at time t by $f(t)$. The cumulative in-flow at t is defined as $F(t)$, which is represented by:

$$F(t) = \int_0^t f(s)ds. \tag{3}$$

Let $\mathbf{X}$ be a spatial-distance domain (i.e., $\mathbf{X} = [0, X_{max}]$), where X_{max} is the maximum of trip distance. We denote the probability density function of the entering trip's distance, $x \in \mathbf{X}$, at time t as $\tilde{\varphi}(t, x)$ and the cumulative distribution function of the entering trips with distances not smaller than x at time t as $\tilde{\Phi}(t, x)$, which satisfy:

$$\int_0^\infty \tilde{\varphi}(t, x)dx = 1, \tag{4}$$

$$\tilde{\Phi}(t, x) = \int_x^\infty \tilde{\varphi}(t, s)ds. \tag{5}$$

Here, $\tilde{\varphi}(t, x) \geq 0$, $\tilde{\varphi}(t, \infty) = 0$, $\tilde{\Phi}(t, 0) = 1$ and $\tilde{\Phi}(t, \infty) = 0$. The $\tilde{\Phi}$ can be interpreted as the ratio of the number of vehicles entering the network at time t and having the travel distance larger than x over the number of vehicles entering the network at time t. We denote the average distance of the entering trips at time t by $\tilde{B}(t)$, which satisfies:

$$\tilde{B}(t) = \int_0^\infty x\tilde{\varphi}(t, x)dx. \tag{6}$$

2.1.3. Traffic state variables

Traffic state variables are the probability density function of the remaining trip distance x and the average distance of remaining trips of vehicles at time t in a road network. We denote the probability density function of the remaining trip distance x at time t by $\varphi(t,x)$, which satisfies:

$$\int_0^\infty \varphi(t,x)dx = 1, \tag{7}$$

where $\varphi(t,x) \geq 0$ and $\varphi(t,\infty) = 0$. Based on the definition of $\varphi(t,x)$, we define the cumulative distribution function of the trips with remaining distances not smaller than x at time t as $\Phi(t,x)$, which satisfies:

$$\Phi(t,x) = \int_x^\infty \varphi(t,s)ds, \tag{8}$$

where $\Phi(t,0) = 1$ and $\Phi(t,\infty) = 0$. Based on Equation 8, we have: $\varphi(t,x) = -\frac{\partial\Phi(t,x)}{\partial x}$. Besides, the average distance of remaining trips of vehicles on the road at time t is denoted by $B(t)$:

$$B(t) = \int_0^\infty x\varphi(t,x)dx. \tag{9}$$

2.1.4. Active trip variables

Active trip variables are the number of active trips (i.e., $\lambda(t)$) and the density of active trips (i.e., $\rho(t)$). We denote the number of active trips (traveling vehicles) at time t by $\lambda(t)$. Consequently, the density of vehicles per unit length of the road at time t equals $\rho(t) = \frac{\lambda(t)}{L_{net}}$. Based on $\lambda(t)$ and $\varphi(t,x)$, we know that the density of active trips with a remaining distance x at time t can be expressed as:

$$k(t,x) = \lambda(t)\varphi(t,x). \tag{10}$$

Since $\int_0^\infty \varphi(t,x)dx = 1$, we have $\int_0^\infty k(t,x)dx = \int_0^\infty \lambda(t)\varphi(t,x)dx = \lambda(t)$. Finally, we define the number of trips with a remaining distance not smaller than x at time t by $K(t,x)$:

$$K(t,x) = \lambda(t)\Phi(t,x). \tag{11}$$

The variable $K(t,x)$ is a major variable in the bathtub model. $K(t,x)$ implies that there are exactly $K(t,x)$ on-road vehicles having the remaining trip distance at least x at time t. Jin (2020) defined the main variable as the number of trips with a remaining distance not smaller than x at time t, $K(t,x)$ [23]. If $K(t,x) = n$, it implies that the n-th longest active trip has a remaining distance of x at time t. $K(t,x)$ can provide trip information on remaining trip distances in road networks. For example, there are 100 vehicles in a road network at time t (i.e., $K(t,0) = 100$) and two cases: 1) half of the vehicles have remaining distances of at least 10 miles (i.e., $K_{case1}(t,10) = 50$) and 2) half of the vehicles have remaining distances of at least 50 miles (i.e., $K_{case2}(t,50) = 50 \leq K_{case2}(t,10)$). If the average speed is the same in two cases, traffic operators can predict that the first case (i.e., $K_{case1}(t,10) = 50$) will alleviate traffic congestion faster than the second case (i.e., $K_{case2}(t,50) = 50$). In summary, the comparison between the two examples illustrates the significant role of $K(t,x)$ in determining future network states. The boundary condition between $\lambda(t)$ and $K(t,0)$ is $\lambda(t) = K(t,0)$ because trips with negative distances are assumed to have exited the network (i.e., $\rho(t) = \frac{\lambda(t)}{L_{net}} = \frac{K(t,0)}{L_{net}}$).

2.1.5. Outflow variables

Outflow variables are the outflow rate of exiting trips (i.e., $g(t)$) and cumulative outflow rate (i.e., $G(t)$). We can obtain the number of exiting trips from Δt to $t + \Delta t$ as follows:

$$g(t)\Delta t = \int_0^{v(t)\Delta t} k(t,x)dx \approx k(t,0)v(t)\Delta t. \tag{12}$$

Hence, the outflow rate of exiting trips and cumulative outflow at time t are defined by $g(t)$ and $G(t)$, respectively:

$$g(t) = k(t,0)v(t) = \lambda(t)\varphi(t,0)v(t), \tag{13}$$

$$G(t) = \int_0^t g(s)ds. \tag{14}$$

The variable $G(t)$ implies the total volume of outflow until time t.

2.2. Generalized Bathtub Model

GBM is a framework for modeling network traffic dynamics with general distributions of trip distances and also can track the evolution of traffic dynamics with remaining trip lengths from three conservation laws. GBM has the following assumptions similar to the accumulation-based model:

- Assumption I: All running vehicles have the same average speed at time t
- Assumption II: The distance distribution for entering trips is an arbitrary distribution
- Assumption III: The speed-density relationship in the network follows the network fundamental diagram.

Since Assumption III is based on MFD, GBM needs to estimate the speed-density relationship in the network.

The governing equations of GBM demonstrate the traffic dynamic in road networks:

$$\mathcal{N}(\mathbf{K}, \mathbf{Q}; \Lambda) = \mathbf{0}, \tag{15}$$

where the operator $\mathcal{N}$ includes three conservation laws of GBM, $\mathbf{K}$ is the set of $K(t,x)$ (i.e., $\mathbf{K} = \{K(t,x) | \, t \in \mathbf{T}$ and $x \in \mathbf{X}\}$), $\mathbf{Q}$ is the set of traffic variables related to entering trips (i.e., $\mathbf{Q} = \{f(t), \tilde{\Phi}(t,x) \mid t \in \mathbf{T}$ and $x \in \mathbf{X}\}$), and Λ contains the parameters of GBM. Since the set of traffic variables related to entering trips $\mathbf{Q}$ is derived from real-world data, we obtain $\mathbf{Q}$ from mobile location data. GBM has three conservation laws [23]: conservation of total trip-miles, conservation of the total number of active trips, and conservation of the number of active trips with remaining distances not smaller than a specific value.

The first conservation law is to conserve the total trip-miles. At the time t, the remaining trip-miles (i.e., $\lambda(t)B(t)$) are derived from the sum of the initial trip-miles (i.e., $\lambda(0)B(0)$), the trip-miles added until time t (i.e., $\int_0^t f(s)\tilde{B}(s)ds$), and the trip-miles traveled until time t (i.e., $\int_0^t \lambda(s)v(s)ds$):

$$\lambda(0)B(0) + \int_0^t f(s)\tilde{B}(s)ds - \int_0^t \lambda(s)v(s)ds = \lambda(t)B(t). \tag{16}$$

Recall that $\lambda(t)$, $B(t)$, $f(t)$, $\tilde{B}(t)$, and $v(t)$ denote the number of active trips (unit: $trip$), the average distance of remaining trips (unit: $mile$), entering trip rate (unit: $trip/hour$), the average distance of entering trips (unit: $mile$), and average speed of vehicles (unit: $mile/hour$), respectively.

The second conservation law describes the relationship between variables of flows. To conserve the total number of active trips, the cumulative outflow until time t (i.e., $G(t)$) is obtained by subtracting the number of running vehicles at time t (i.e., $\lambda(t)$) from the sum of the number of initial running vehicles (i.e., $\lambda(0)$) and the cumulative inflow until time t (i.e., $F(t)$) as follows:

$$G(t) = \lambda(0) + F(t) - \lambda(t). \tag{17}$$

To conserve the number of active trips with remaining distances not smaller than any value, Jin (2020) formulated the third conservation law in two versions [23]:

$$\begin{aligned}\text{(Continuous Version)} \quad \frac{\partial}{\partial t}K(t,x) &= v(t)\frac{\partial}{\partial x}K(t,x) + f(t)\tilde{\Phi}(t,x),\\ &= V(\frac{K(t,0)}{L})\frac{\partial}{\partial x}K(t,x) + f(t)\tilde{\Phi}(t,x),\end{aligned} \tag{18}$$

$$\text{(Discrete Version)} \quad K(t+\Delta t, x) = K(t, x+v(t)\Delta t) + f(t)\tilde{\Phi}(t,x)\Delta t, \tag{19}$$

where Δt is a small time interval. Recall that $K(t,x)$, $v(t)$, $f(t)$, and $\tilde{\Phi}(t,x)$ are respectively the number of trips with remaining distance $\geq x$ at time t (unit: $trip$), the average speed of vehicles (unit: $mile/hour$), the entering trip rate (unit: $trip/hour$), and the ratio of entering trips with distances $\geq x$ (unit: 1). The third conservation law captures traffic dynamics with respect to time t and remaining distance x. Jin (2020) derived analytical solutions by using a continuous version of the third conservation law (i.e., Equation 18) [23]. However, Jin (2020) used a discrete version of the third conservation law (i.e., Equation 19) in a numerical experiment since derivatives of $K(t,x)$ are challenging to be calculated [23]. In this study, we use the (continuous version) third conservation law (i.e., Equation 18) in the training of PIML-GBM since our PIML-GBM model can calculate the derivatives of $K(t,x)$ from an automatic differentiation technique.

3. Methodology

We introduce the proposed PIML-GBM to model the traffic dynamics in large-scale urban networks.

3.1. Problem statement

We consider the estimation of the active trip variable in large-scale urban networks. In this study, the large-scale urban network corresponds to a homogeneous urban reservoir which is represented. We discretize a time-distance domain $\mathbf{S}$ into a discretized time-distance domain $\mathbf{G} = \{(t_g^k, x_g^k) \mid k = 1, ..., N_g\}$ with the number of grid points N_g in time and distance, respectively (i.e., $\mathbf{G} \subset \mathbf{S}$). We can control a resolution level of $\mathbf{G}$ by setting the number of time grid N_t and distance grid N_d (i.e., $N_g = N_t \times N_d$). In a time-distance domain $\mathbf{S}$, there are two types of points: (1) observed boundary points $\mathbf{O} = \{(t_o^i, x_o^i)|i = 1, ..., N_o\} \subset \mathbf{G}$ and (2) auxiliary points $\mathbf{A} = \{(t_a^j, x_a^j)|j = 1, ..., N_a\} \subset \mathbf{S}$, where N_o is the number of observed boundary points and N_a is the number of auxiliary points.

The observed boundary points $\mathbf{O}$, which constitute a subset of $\mathbf{G}$, can be recorded from real-world data and then $K(t, x)$ is known given $\forall(x, t) \in \mathbf{O}$. Let $\mathbf{Y} = \{K(t_o^i, x_o^i)|i = 1, ..., N_o\}$ be the ground-truth (or observed labels). The number of observed boundary points depends on the granularity of a discretized time-distance domain $\mathbf{G}$ (i.e., $N_o = (N_t + N_x) \times 2$). Auxiliary points $\mathbf{A}$ are virtual points in a time-distance domain $\mathbf{S}$ and are unique to the framework of PIML. The auxiliary points represent unobserved data calculated by GBM. The proposed method can learn physics knowledge from the governing equations of the GBM on auxiliary points.

Observed boundary points are directly compared to the estimated boundary points from our method (i.e., data loss) and auxiliary points are used in calculating the physics loss from GBM. The physics loss serves to reinforce the underlying governing principles of the GBM model during training neural networks. The number of auxiliary points depends on how much we want to learn physics knowledge. In setting the number of auxiliary points, there is a trade-off between the amount of knowledge obtained from physics and the learning speed. The more auxiliary points we set, the more information about physics that PIML can learn. In contrast, the fewer auxiliary points we set, the faster the learning algorithm completes (i.e., the lower the computational cost).

This study has the same assumptions except for Assumption III of GBM as follows:

- Assumption I: All running vehicles have the same speed at time t
- Assumption II: The distance distribution for entering trips is a given empirical distribution.

Since the proposed method directly uses the average speed from real-world data, it can relax Assumption III of GBM and thus the estimation of MFD is not required.

Our research question is how to estimate and predict all values of the active trip variable over a continuous time-distance domain with a few boundary points and empirical trip length distribution. Estimating the active trip variable over a continuous time-distance space is the main challenge since calculating the derivatives of the governing equations is strenuous. This paper aims to investigate the performance of a machine learning model regularized by a physics-based model for estimating active trip variables over a continuous time-distance domain in large-scale urban networks. The main idea is to embed traffic dynamics into the machine learning model. We formulate and design the learning algorithm for the machine learning model with physics loss derived from the physics-based model.

3.2. Framework of PIML-GBM

PIML-GBM consists of two main parts: a multi-layer neural network (MNN) and physics-informed learning. MNN is parameterized by the network parameters $\mathbf{w}$. The framework of our proposed PIML-GBM is graphically described in Figure 2.

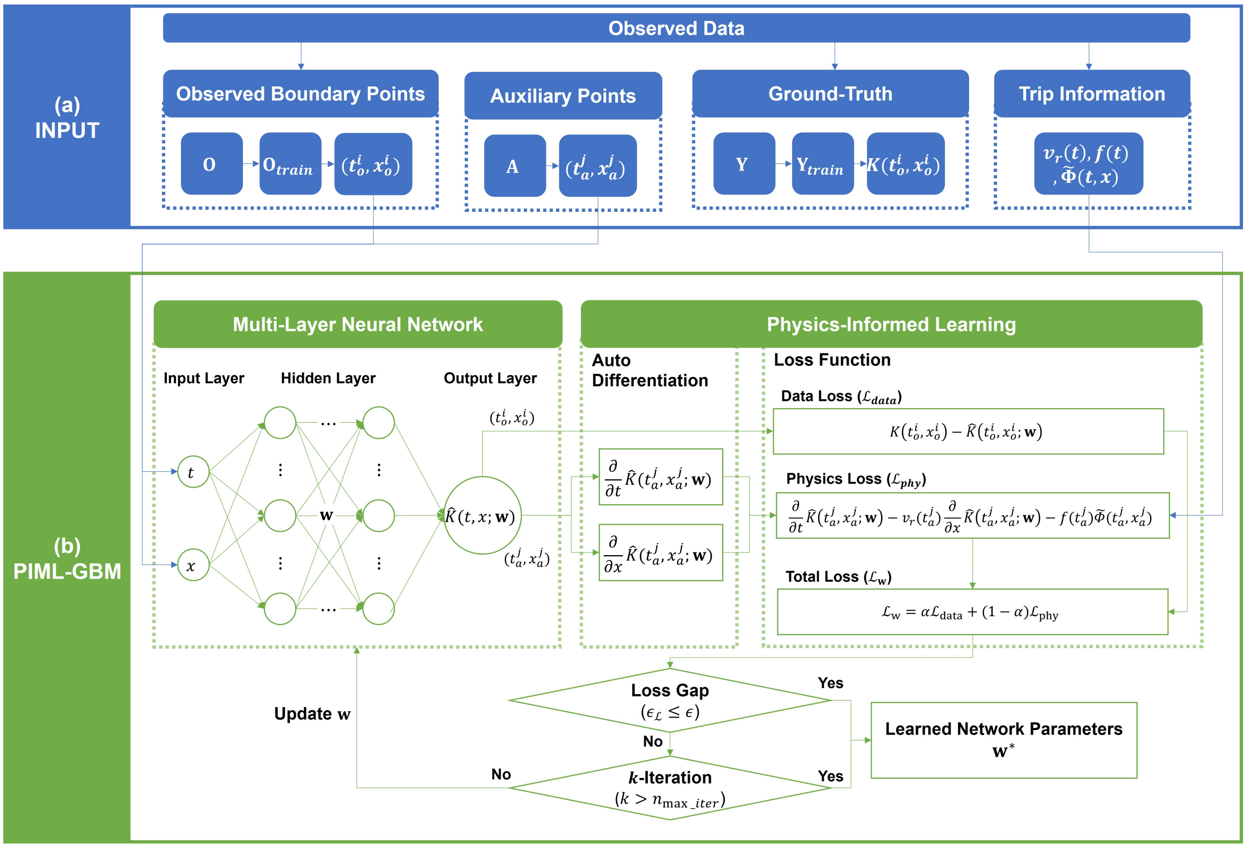

Figure 2: The Framework of Proposed PIML-GBM: (a) Input of the proposed PIML-GBM consists of observed points, auxiliary points, ground-truth, and trip information. All input data except for auxiliary points $\mathbf{A}$ are obtained from real-world data. (b) The part of training PIML-GBM consists of MNN, an automatic differentiation technique (i.e., PDE solving technique), and a loss function. MNN estimates the traffic state variable given $\mathbf{O}_{train}$ by using a fully-connected forward propagation neural network. We use the automatic differentiation with L-BFGS optimizer for calculating the derivative of residuals $\hat{f}(t, x; \mathbf{w}, v_r(t), f(t), \tilde{\Phi}(t, x))$. Losses from data and physics information are calculated by Equations (20, 22, and 23).

The input data of PIML-GBM includes observed boundary points $\mathbf{O}$, auxiliary points $\mathbf{A}$, ground-truth $\mathbf{Y}$, and trip information (i.e., $v_r(t), f(t), \tilde{\Phi}(t, x)$), as shown in Figure 2 (a). Input data except for auxiliary points are obtained from real-world data. In this study, we employ $v_r(t)$, representing the real average speed at time t within the road network, instead of $v(t)$. The use of $v_r(t)$ offers a significant advantage, as it obviates the need for the calibration of the MFD. This simplification streamlines the modeling process and has the potential to improve the reliability and applicability of the PIML-GBM We denote the estimation of $K(t,x)$ from MNN as $\hat{K}(t,x;\mathbf{w})$, as shown in Figure 2 (b). In physics-informed learning, there are two components: the automatic differentiation technique from Tensorflow and the calculation of total loss. Automatic differentiation calculates gradients for minimizing total loss at each learning iteration. The total loss $\mathcal{L}_\alpha$ is calculated from a weighted sum of data loss $\mathcal{L}_{data}$ and physics loss $\mathcal{L}_{phy}$. Data loss $\mathcal{L}_{data}$ is a data discrepancy between the ground-truth $K(t,x)$ and the estimation $\hat{K}(t,x;\mathbf{w})$. Physics loss $\mathcal{L}_{phy}$ to learn physics knowledge is a residual derived from the third conservation law (i.e., Equation 18) to train PIML-GBM. The physics-informed learning part calculates the total loss using the output of MNN $\hat{K}(t,x;\mathbf{w})$ as input. The physics learning part updates network parameters $\mathbf{w}$ to estimate $\hat{K}(t,x;\mathbf{w})$ as close as possible to the ground-truth $K(t,x)$ when a residual would be closer to zero. The output of PIML-GBM is an optimal parameter set $\mathbf{w}^*$ after the training process.

The goal of PIML-GBM is to find the optimal parameter set $\mathbf{w}^*$ given small samples of boundary observations for entering trips, which can estimate the number of active trips with remaining distances $K(t,x)$ at any point in a continuous time-distance domain S. The framework of PIML-GBM consists of the training dataset, model architecture, and training algorithm.

3.2.1. Training dataset for PIML-GBM

The training dataset for PIML-GBM is composed of training points $\mathbf{O}_{train} = \{(t_o^i, x_o^i) | i = 1, ..., N_{train}\}$, training labels $\mathbf{Y}_{train} = \{K(t_o^i, x_o^i) \mid i = 1, ..., N_{train}\}$, auxiliary points $\mathbf{A} = \{(t_a^j, x_a^j) | j = 1, ..., N_a\}$, and trip information (i.e., $v_r(t), f(t), \tilde{\Phi}(t,x)$), as shown in Figure 2 (a). Before sampling the training dataset, we normalize all datasets for a stable training process by scaling between 0 and 1. We sample the observed boundary dataset $(\mathbf{O}, \mathbf{Y})$ into $(\mathbf{O}_{train}, \mathbf{Y}_{train})$ training points with a sample rate r_{train}. Since the labels are values on a domain of training points $\mathbf{O}_{train}$, we denoted the same index of $\mathbf{O}_{train}$ and $\mathbf{Y}_{train}$ as i.

Auxiliary points $\mathbf{A}$ are uniformly selected in a continuous time-distance domain $\mathbf{S}$. Since auxiliary points are not restricted from real-world data, we can set the granularity of the time-distance domain through the number of auxiliary points $N_a = n_a \times N_{train}$ for regularization based on physics information from GBM, where n_a is a multiplier of auxiliary points. For example, the observed and auxiliary points with three multipliers ($n_a = 1, 10, 100$) are described in Figure 3. Increasing the number of auxiliary points in the observed dataset enhances the model's ability to learn from the knowledge embedded in physical traffic flow models. However, this augmentation also increases the training time.

The information for entering trips is obtained from real trip data and $v_r(t)$ is the real average speed at time t in the road network. The average speed $v_r(t)$ used in PIML-GBM is different from the average speed $v(t)$ used in GBM: $v(t)$ is estimated by MFD, but $v_r(t)$ is directly derived from real trip data.

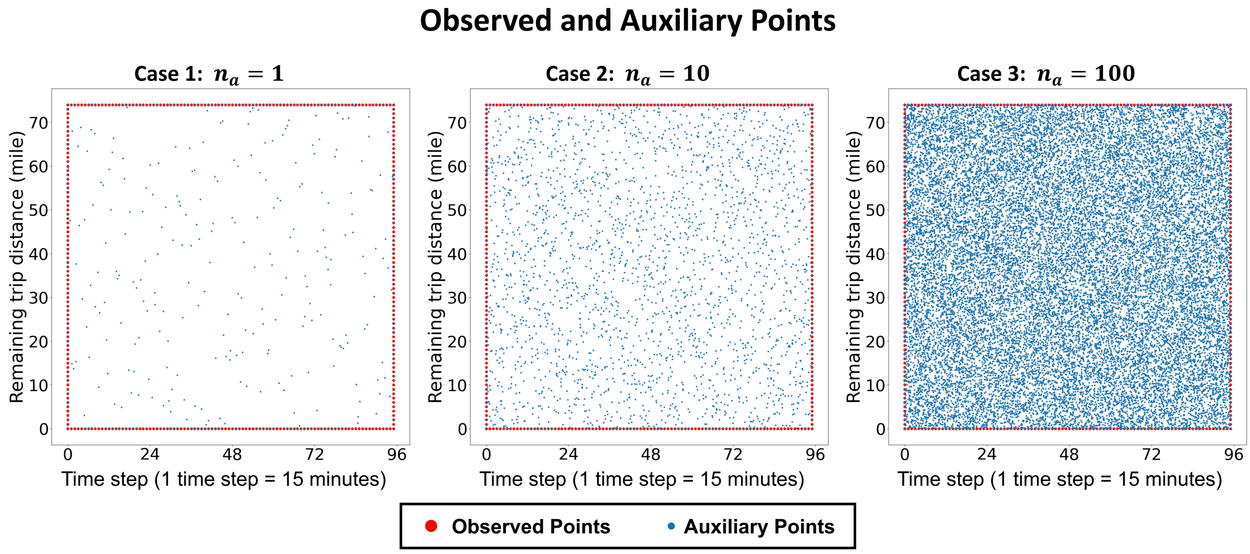

Figure 3: Observed Points and Auxiliary Points over Time-Distance Domain with Various Multipliers ($n_a = 1, 10, 100$)

3.2.2. Model architecture

The PIML-GBM consists of MNN for estimating the traffic state variable $\hat{K}(t, x; \mathbf{w})$, the automatic differentiation technique from Tensorflow, and the calculation of total loss, as shown in Figure 2 (b). The MNN is a fully connected neural network and consists of two input nodes (x, t), hidden layers with network parameters $\mathbf{w}$, and one output node $\hat{K}(t, x; \mathbf{w})$. We initialize $\mathbf{w}$ by using the Xavier uniform initializer and use $Tanh(\cdot)$ as an activation function of each hidden neuron in MNN. The MNN learns the physics knowledge by minimizing the value of loss function with weights $\mathbf{w}$ between layers of MNN. We denote the estimated traffic state variable from MNN as $\hat{K}(t_o^i, x_o^i; \mathbf{w}) \in \hat{\mathbf{K}}_{train}$ for $(t_o^i, x_o^i) \in \mathbf{O}_{train}$. The loss function of MNN in PIML-GBM consists of data loss and physics loss. First, we define the data loss $\mathcal{L}_{data}$ as the gap between observed traffic state variables (i.e., $K(t_o^i, x_o^i)$) and estimated traffic state variables (i.e., $\hat{K}(t_o^i, x_o^i; \mathbf{w})$) given $(t_o^i, x_o^i) \in \mathbf{O}_{train}$:

$$\mathcal{L}_{data} = \frac{1}{N_{train}} \sum_{i=1}^{N_{train}} |K(t_o^i, x_o^i) - \hat{K}(t_o^i, x_o^i; \mathbf{w})|^2. \tag{20}$$

Second, we define the residuals to learn the physics knowledge from the governing equation (i.e., Equation 18):

$$\hat{f}(t, x; \mathbf{w}, v_r(t), f(t), \tilde{\Phi}(t, x)) = \frac{\partial}{\partial t}\hat{K}(t, x; \mathbf{w}) - v_r(t)\frac{\partial}{\partial x}\hat{K}(t, x; \mathbf{w}) - f(t)\tilde{\Phi}(t, x). \tag{21}$$

A difference between Equation 18 and Equation 21 is to use the average speed at time t from real-world data $v_r(t)$ in Equation 21 without any explicit assumptions such as MFD. We obtain $v_r(t)$ from real-world trip data at every time period. The derivatives of the residuals in Equation 21 are calculated by the function of TensorFlow "tf.GradientTape" and parameterized by $\mathbf{w}$. Accordingly, we define the physics loss $\mathcal{L}_{phy}$ from the residuals (i.e., Equation 21):

$$\begin{aligned}\mathcal{L}_{phy} &= \frac{1}{N_a}\sum_{j=1}^{N_a}|\hat{f}(t_a^j, x_a^j; \mathbf{w}, v_r(t_a^j), f(t), \tilde{\Phi}(t,x))|^2 \\ &= \frac{1}{N_a}\sum_{j=1}^{N_a}|\frac{\partial}{\partial t}\hat{K}(t_a^j, x_a^j; \mathbf{w}) - v_r(t_a^j)\frac{\partial}{\partial x}\hat{K}(t_a^j, x_a^j; \mathbf{w}) - f(t_a^j)\tilde{\Phi}(t_a^j, x_a^j)|^2.\end{aligned} \tag{22}$$

We use the total loss of PIML-GBM with a weight of losses α for training the proposed PIML-GBM as follows:

$$\mathcal{L}_\alpha = \alpha\mathcal{L}_{data} + (1-\alpha)\mathcal{L}_{phy}. \tag{23}$$

3.3. Training algorithm

We use the automatic differentiation with the L-BFGS optimizer in Tensorflow for evaluating the derivative of residuals since the L-BFGS optimizer can provide a stable solution with fewer iterations than the Adam optimizer [47]. The training process is terminated if the loss gap $\epsilon_k = |\mathcal{L}_\alpha^{k+1} - \mathcal{L}_\alpha^k|$ between two consecutive total losses at iteration k is less than the termination threshold ϵ or the number of iterations reaches a predetermined maximum number of iterations (i.e., n_{max_iter}). After the training process, we obtain the learned weights of trained PIML-GBM $\mathbf{w}^* = \arg\min_{\mathbf{w}} \mathcal{L}_\alpha$. The training algorithm is shown in Algorithm 1.

Algorithm 1 Training Algorithm of Proposed PIML-GBM

Input: Observed dataset $\{(t_o^i, x_o^i, K(t_o^i, x_o^i))\}_{i=1}^{N_{train}}$; auxiliary points $\{(t_a^j, x_a^j, K(t_a^j, x_a^j))\}_{j=1}^{N_a}$; average speed $v_r(t)$; inflow rate $f(t)$; cumulative distribution function of the entering trips with distances $\tilde{\Phi}(t,x)$

Require: L-BFGS optimizer.

Initialization: Initialized network parameters $\mathbf{w}$; termination threshold ϵ; maximum number of iterations n_{max_iter}; weight of loss functions α; learning rate γ

Procedure:

366

```
k ← 0
J_w^0 = 0
while k < n_max_iter or |L_α^{k+1} − L_α^k| ≥ ε do
    Calculate L_data^k by Equation 20
    Calculate L_phy^k by Equation 22
    Calculate L_α^k by Equation 23
    w^{k+1} ← w^k − γ · LBFGS(w^k, ∇_{w^k} L_α^k) by automatic differentiation
end while
```

Output: Learned network parameters $\mathbf{w}^*$

All notations employed in this study are comprehensively summarized in the Appendix, referenced as Appendix A.

4. Experiments

4.1. Data Description

We obtained LBD data for trip extraction from a mobile phone vendor in Indiana, covering 21 weekdays from March 1st to March 29th, 2019. This dataset comprises 14.4 million unique devices and 4.8 billion records. The vendor collected the LBD directly from first-party, opt-in mobile devices through server-to-server integration. The study area is Marion County, Indiana, United States, as shown in Figure 4 (a). In 2019, there were 964,582 residents living in the study area [48]. We define nodes as intersections and links as road segments. The size of the study area is approximately 396.61 mi^2, consisting of 35,742 nodes and 49,455 links (total length of links = 4851.09 miles).

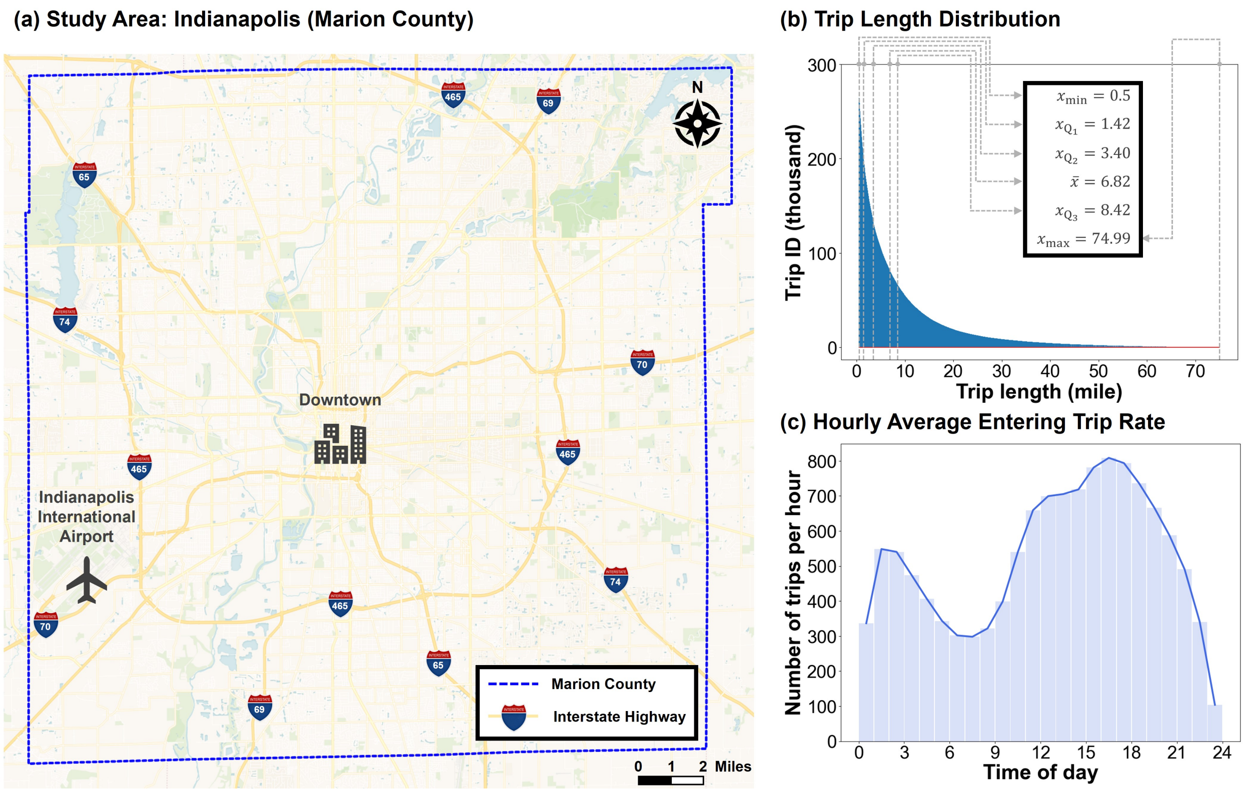

Figure 4: Study Area and Trip Patterns: (a) Study area (Marion County, Indiana, United States; Source: Map data ©OpenStreetMap contributors, ©CartoDB attributions), (b) Distribution of trip lengths, and (c) Hourly average entering trip rate

The weekend records are not included in this study because of their distinct weekday traffic patterns.

Each record consists of the anonymous device ID, location information (latitude and longitude), horizontal accuracy (meters), timestamp, and so on (device type, OS type, ...), as shown in Table 1. The time interval between two consecutive records varies from a few seconds to a few minutes.

Table 1: Description of Mobile Phone Data Used in This Study

Field	Description	Sample data
Device_ID	Device ID of unique anonymous user	5054eb6c-5877-462b
ID_Type	IDFA (iOS) and ADID (Android)	adid
Latitude	Latitude of the record	39.7678718
Longitude	Longitude of the record	-86.1582648
Horizontal_Accuracy	GPS accuracy reported by the device OS	18.0
TimeStamp	Timestamp of the record	2019-03-15 08:31:22
...	...	...

4.2. Data Processing

We extract the trip information (i.e., origin and destination) from the raw mobile data as follows: 1) Detecting Home, 2) Extracting Trips, and 3) Filtering.

The mobile phone data includes location information from a diverse user base, including residents, commercial vehicle drivers, tourists, and others. For this study, we define "valid users" as residents who make regular trips within the study area. Initially, we isolate valid trips based on nighttime location data (collected between 8 PM and 6 AM) and cluster these data points to estimate each user's home location. The centroid of each cluster serves as the estimated home location. Since mobile phone data includes various types of users (such as residents, tourists, commercial vehicle drivers, and drivers on highways), we filter users with few records (less than eight days). In other words, we extract users recorded at least eight days, a threshold selected by balancing the number of filtered users and the noise in the home locations. After detecting the home location, the preprocessed mobile phone data consists of 39,465 valid users.

In the second step, we extract trips from valid users. Users' trips are recorded by several consecutive records with stay regions (origin and destination) and waypoints on the trip within a short time caused by stopping situations (e.g., intersection on the red, traffic jam, and so on). We assume that a new trip starts when the time interval between two consecutive records is more than 30 minutes, which is also used in previous works [46, 49]. For example, a traveler starts a new trip from home to the company at 8:00:10 in Figure 5. Due to congestion caused by traffic signals during trips, the spatial distance between records recorded at t_2 and t_3 is only 15 ft apart, but the time interval is about 4 minutes. In this case, two records are considered as waypoints that constitute one trip. On the other hand, since the time interval exceeds 8 hours, it is treated as a new trip and the records from t_6 constitute a new trip. Based on this approach, we extract 387,496 trips between two consecutive distinct stay regions from 39,465 valid users.

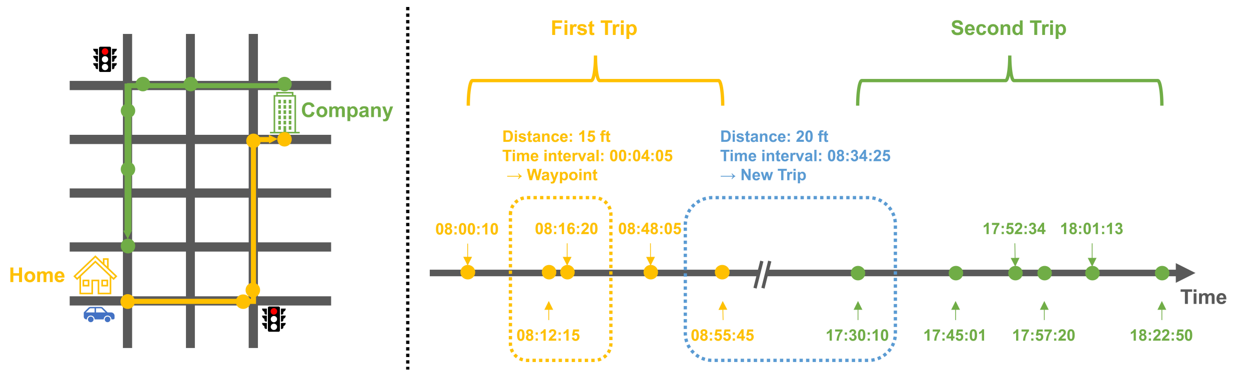

Figure 5: Example of Trip Extraction: There are two trips. Since the time interval between two consecutive records at t_5 and t_6 is larger than 30 minutes, we consider that there are distinct two trips in the example.

In the last step, we remove excessively long and short trips, which are challenging to be considered normal trips with distances less than 0.5 miles and larger than 75 miles (near the 98th percentile). Finally, this study analyzes 34,355 users with 264,620 trips including 4,067,602 waypoints (or data collected points), as shown in Figure 4 (b). We calculate trip lengths from each origin to each destination using the GoogleMaps API. Since the recorded time interval of LDD is not constant, GoogleMaps API was used to estimate a realistic distance considering the geometry of the road rather than the Euclidian distance. In Figure 4 (b), trip length statistics are as follows: x_{Q_1} (25th percentile) $\approx$ 1.42 miles, x_{Q_2} (median) $\approx$ 3.40 miles, and x_{Q_3} (75th percentile) = 8.42 miles. The range spans 0.50 to 74.99 miles and the mean trip length is 6.82 miles (i.e., $x_{\min} = 0.50$, $x_{\max} = 74.99$, $\bar{x} = 6.82$). The percentage of valid users compared to the population of census data is 3.56% (=34,565/964,582). Illustrated in Figure 4 (c), the graph depicts the hourly average entering trip rate, unveiling a discernible pattern. In particular, there is a conspicuous increase in traffic influx around 9 AM, accompanied by a substantial surge from 4 PM to 7 PM, representing the peak hours. The average number of trips for one user is about 7.70 trips per month and about 2.19 trips per day when the user travels.

Figure 6 is composed of two sub-figures that elucidate the spatial distribution of our dataset, which encapsulates a total of 264,620 trips and 4,067,602 waypoints. Figure 6 (a) presents the distribution of the waypoints in the study area, illustrating the complete coverage of the entire road network. This expansive coverage serves as a solid foundation for modeling traffic dynamics in large-scale road networks. Figure 6 (b) focuses on the distribution of waypoints within the downtown region, where a high concentration of commercial offices is located. This downtown-centric view enhances the capability of our data set to accurately represent commuting patterns in areas characterized by a high employment density.

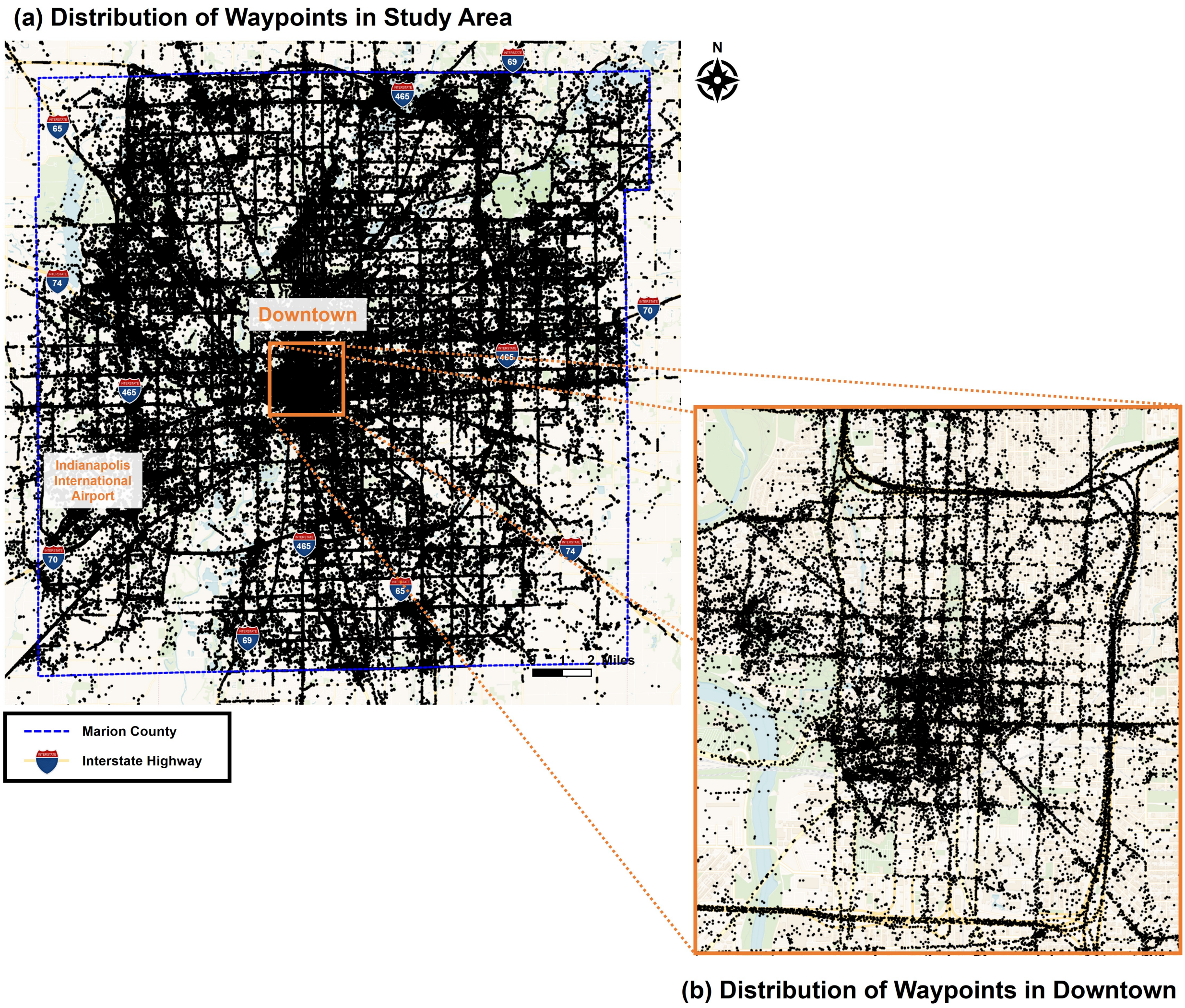

Figure 6: Visualization of Collected Waypoints in Study Area: (a) Distribution of waypoints in Study Area, (b) Distribution of waypoints in Downtown. (Source: Map data ©OpenStreetMap contributors, ©CartoDB attributions)

4.3. Experiment Settings

We set a time-distance domain $\mathbf{S}$ with the maximum trip distance $X_{max} = 75$ miles and the time period $T =$ one day (i.e., 24 hours) on March 1st, 2019. To discretize the time-distance domain, we let a spatial resolution N_x be one mile and a temporal resolution N_t be 15 minutes (i.e., $(t_g, x_g) \in \mathbf{G} = [0, 96] \times [0, 75]$ and $N_g = N_t \times N_x = 96 \times 75 = 7,200$). After discretizing the time-distance domain, we derive observed boundary points ($\mathbf{O}$, $\mathbf{Y}$) from mobile location data with 342 observed boundary points (i.e., $N_o = (N_t + N_x) \times 2 = (75 + 96) \times 2 = 342$). We set a sample rate r_{train} as 70% and randomly select the training dataset ($\mathbf{O}_{train}$, $\mathbf{Y}_{train}$) with 240 training points (i.e., $N_{train} = N_o \times r_{train} = 342 \times 0.7 \approx 240$) from observed boundary points ($\mathbf{O}$, $\mathbf{Y}$). We use a multiplier of auxiliary points as 50 and randomly select 12,000 auxiliary points in $\mathbf{A} \subset \mathbf{G}$ (i.e., $N_a = n_a \times N_{train} = 240 \times 50 = 12,000$).

We derive the entering trip rate $f(t)$, the cumulative distribution function of the entering trips $\tilde{\Phi}(t, x)$, and the average speed of the entering trips $v_r(t)$ from the preprocessed trip data (264,620 trips), as shown in Figure 7 (a-c). In case of the average speed of the entering trips $v_r(t)$, we calculate the speed for each trip by dividing the distance and elapsed time between the recorded points to derive the average speed of the entering trips $v_r(t)$.

Since the speed-density relationship of MFD is needed to obtain solutions of GBM, we use a function (*scipy.optimize.curve_fit*) that automatically adjusts the parameters of a given model function to best match the provided data points, provided by the SciPy library. We estimate the speed-density relationship of MFD from the collected trip data over 21 days, which is established in [50], as follows:

$$V(\rho) = \min(30, \frac{122.685}{\rho} + 9.418). \qquad \text{(unit: mile/hour)} \tag{24}$$

Figure 7 (d) illustrates the estimated speed-density relationship (i.e., free-flow speed = 30 mile/hour), characterized by a coefficient of determination (R^2) of 0.4039 and a root mean square error (RMSE) of 2.5482.

The proposed PIML-GBM consists of 8 hidden layers of MNN and {25, 30, 35, 40, 45} neurons on each hidden layer. Previous studies used 6 - 10 hidden layers and 20 - 60 neurons on each hidden layer of MNN [39, 40, 41, 51, 52]. We adopt an automatic differentiation for calculating derivatives using the L-BFGS algorithm. We set the maximum iteration of training n_{max_iter} as 15,000, the termination threshold ϵ as 1.0×10^{-8} and the learning rate lr as 1.0×10^{-8}. The experiments are conducted on an Intel Core i7-10700K CPU @ 3.80GH and 32 GB of RAM.

4.4. Evaluation Metrics and Baselines

We test the trained PIML-GBM into the discretized time-distance domain $\mathbf{G}$. The testing dataset consists of testing points $\mathbf{O}_{test} = \mathbf{G} = \{(t_g^k, x_g^k) | k = 1, ..., N_g\}$ and testing labels $\mathbf{Y}_{test} = \{K(t_g^k, x_g^k) | k = 1, ..., N_g\}$.

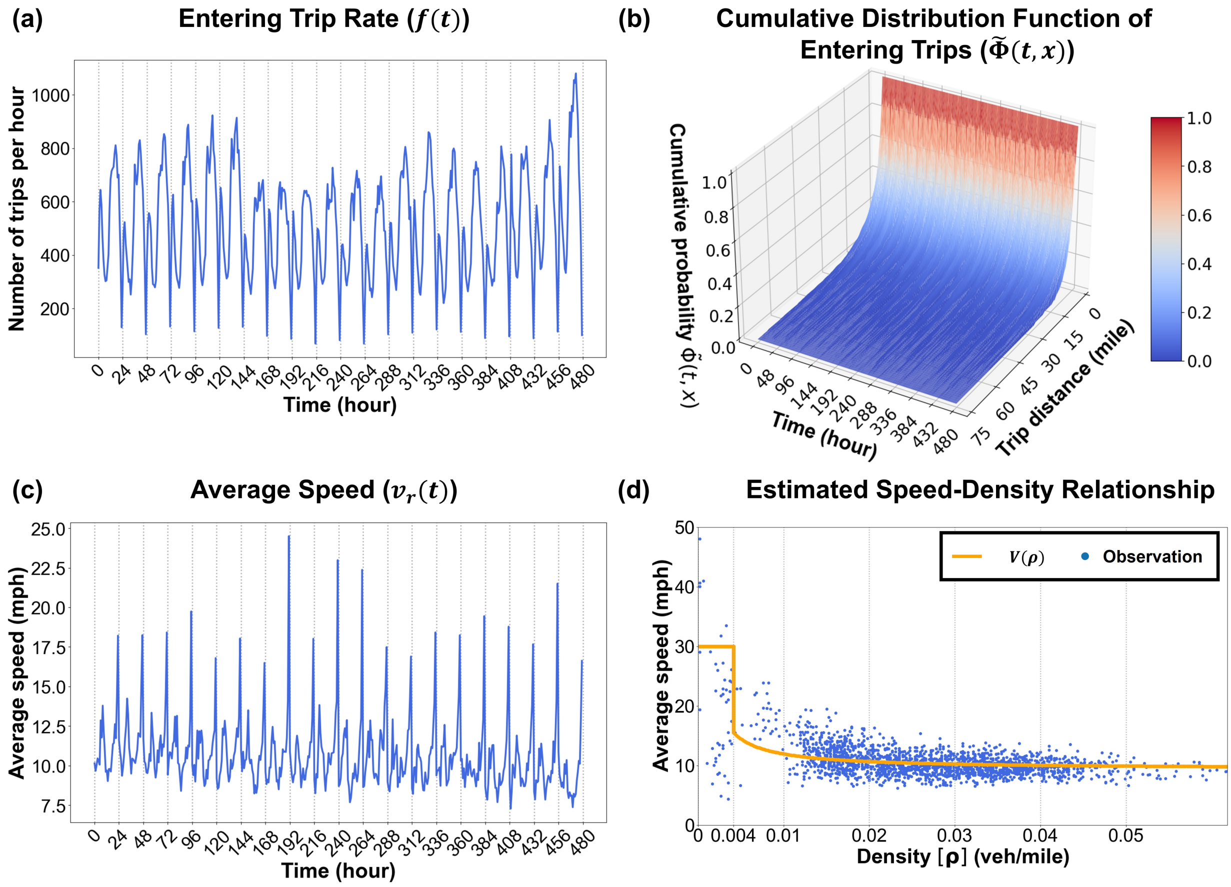

Figure 7: Plots of Trip Characteristics for Training the Proposed PIML-GBM in Marion County for Observed Time Period: (a) Entering trip rate $f(t)$, (b) Cumulative distribution function of entering trips $\tilde{\Phi}(t,x)$, (c) Average speed of entering trips $v_r(t)$, and (d) Estimated speed-density relationship ($V(\rho) = \min(30, \frac{122.685}{\rho} + 9.418)$)

Using PIML-GBM trained with learned network parameters $\mathbf{w}^*$, we can estimate traffic state variables $\hat{K}(t_g^k, x_g^k; \mathbf{w}^*) \in \hat{\mathbf{K}}_{test}$ for testing dataset $k \in \mathbf{O}_{test}$.

We use the Mean Absolute Error(MAE), the Root Mean Square Error ($RMSE$), and the relative L_2 error (Err) to quantify the estimation error between the values of testing labels and estimations of our proposed PIML-GBM over the testing dataset ($\mathbf{O}_{test}, \mathbf{Y}_{test}$) as follows:

$$MAE(\hat{\mathbf{K}}_{test}, \mathbf{Y}_{test}) = \frac{1}{N_g}\sum_{k=1}^{N_g} |\hat{K}(t_g^k, x_g^k; \mathbf{w}^*) - K(t_g^k, x_g^k)|, \tag{25}$$

$$RMSE(\hat{\mathbf{K}}_{test}, \mathbf{Y}_{test}) = \sqrt{\frac{1}{N_g}\sum_{k=1}^{N_g} |\hat{K}(t_g^k, x_g^k; \mathbf{w}^*) - K(t_g^k, x_g^k)|^2}, \tag{26}$$

$$Err(\hat{\mathbf{K}}_{test}, \mathbf{Y}_{test}) = \frac{\sqrt{\sum_{k=1}^{N_g} |\hat{K}(t_g^k, x_g^k; \mathbf{w}^*) - K(t_g^k, x_g^k)|^2}}{\sqrt{\sum_{k=1}^{N_g} |K(t_g^k, x_g^k)|^2}}. \tag{27}$$

MAE measures the average magnitude of the estimation errors in the testing dataset. $RMSE$ is the standard deviation of estimation errors, giving higher weight to large errors. Err normalizes $RMSE$, reducing the impact from the magnitude of ground-truth values. We use MAE to select the optimal trained PIML-GBM because our PIML-GBM aims to estimate a few of large values as well as overall values over the entire domain.

We compare our proposed PIML-GBM to two baselines:

- **Generalized Bathtub Model (GBM)**: GBM numerically estimates $K(t, x)$ based on the solution algorithm proposed by Jin (2020) [23]. The solution algorithm of GBM requires the speed-density relationship of MFD estimated in Equation 24. We adopt GBM to verify the outstanding performance of the proposed PIML-GBM and to quantify the ability of PIML-GBM to capture the randomness and uncertainty of traffic dynamics.

- **Pure Multi-layer Neural Network (PMNN)**: PMNN shares the same architecture with the proposed PIML-GBM except for the physics loss $\mathcal{L}_{phy}$. Since PMNN has no physics loss, auxiliary points $\mathbf{A}$ are not used in PMNN. We use PMNN to confirm the impact of adding physics loss in the training process.

5. Results

This section applies PIML-GBM on real trip data in Marion County, Indiana, United States. First, we show the results of trained PIML-GBM with learned network parameters $\mathbf{w}^*$. Second, we compare the performance of PIML-GBM with baselines (i.e., GBM and PMNN). Lastly, we conduct a sensitivity analysis with respect to the number of neurons on each hidden layer of MNN. We found that the optimal number of neurons is 40 on each hidden layer, which is discussed in Section 5.3.

5.1. Estimation of $K(t, x)$ from PIML-GBM

In this section, we explore the estimation of $K(t, x)$ using our proposed PIML-GBM model. This model is characterized by 40 neurons in each hidden layer, which are chosen from a sensitivity analysis (Section 5.3). We apply the PIML-GBM to calculate the evaluation metrics, defined by Equations (25)-(27). The metrics are visually depicted in Figure 8. In panels (a) and (b) of this figure, we use a logarithm function log to represent the values of $K(t, x)$ and $\hat{K}(t, x; \mathbf{w})$ due to the wide range of these values, leading to the expressions $\log K(t, x)$ and $\log \hat{K}(t, x; \mathbf{w})$.

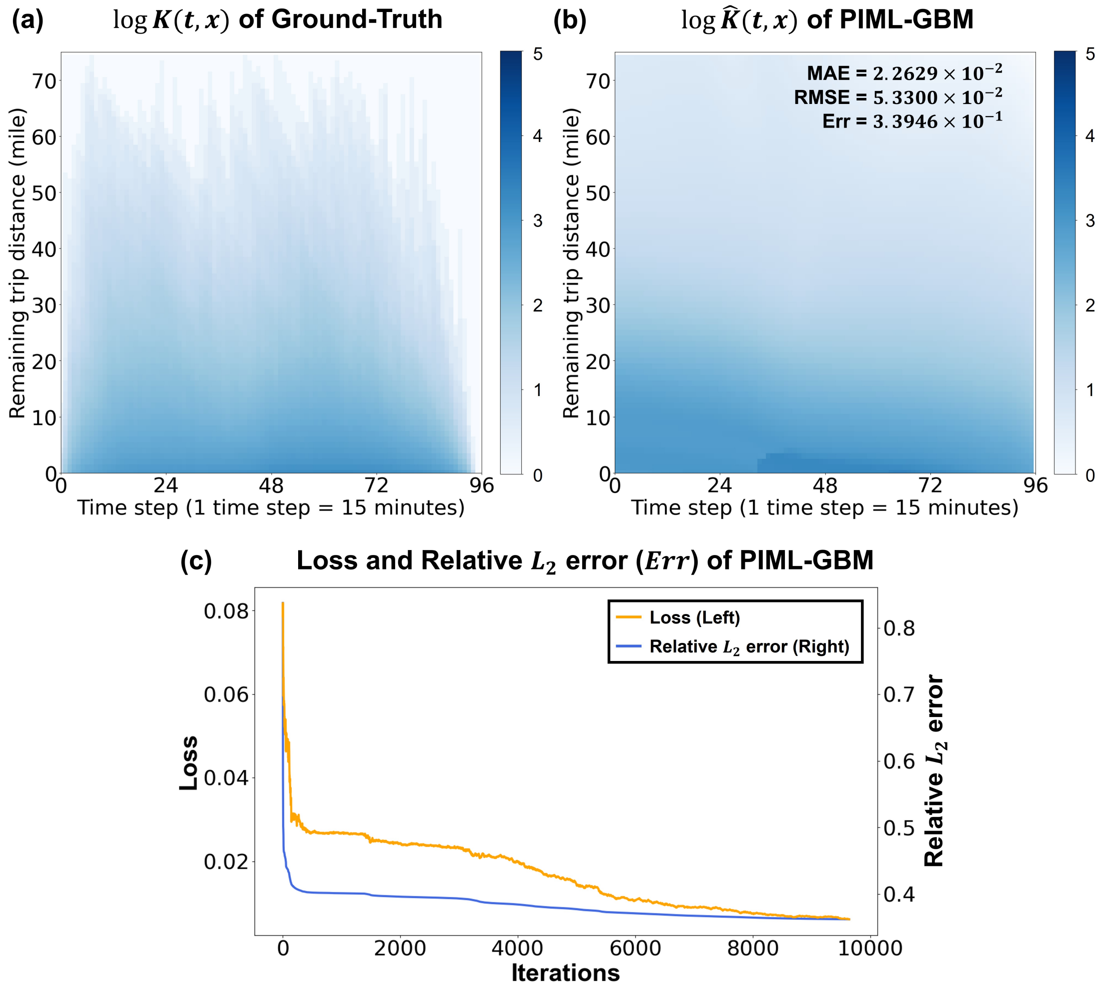

Figure 8: Results of Proposed PIML Model with Learned Network Parameters $\mathbf{w}^*$: (a) Ground-truth values of $\log(K(t,x))$, (b) PIML-GBM estimations of $\hat{K}(t,x;\mathbf{w})$ and evaluation metrics, (c) Loss and Relative L_2 error of PIML-GBM.

The performance of the PIML-GBM model is quantified with the following metrics: MAE $= 0.022629$, RMSE $= 0.053300$, and Err $= 0.33946$. We observe that the values of $\hat{K}(t,x;\mathbf{w})$ are larger than those of $K(t,x)$ for distances $(x > 60)$. However, the estimates align well with the ground-truth values where the remaining distance is shorter than 20 miles $x < 20$, showing the effectiveness of the PIML-GBM model. The loss gap of PIML-GBM converges before reaching the maximum iterations, as depicted in Figure 8 (c). While achieving good accuracy for shorter distances, it exhibits discrepancies for longer distances. Quantitative performance metrics further elucidate the model's capabilities and areas for improvement. The rapid convergence of the loss gap is indicative of the model's ability to learn efficiently.

5.2. Model Comparison

We compare PIML-GBM with two baselines (GBM and PMNN). We can show the outstanding performance of the PIML-GBM by comparing GBM and the impact of physics loss by comparing PMNN. The evaluation metrics in Equations (25) - (27) are calculated in Table 2. First, ML-based models (PIML-GBM and PMNN) significantly outperform the two baselines in Table 2. Since ML-based methods are flexible to theoretical assumptions and capture variance and phenomena from real-world data, PIML-GBM and PMNN can overcome limitations from theoretically ideal assumptions that undermine the model's performance. Second, we can verify the outstanding performance of ML-based methods because ML-based methods particularly outperform GBM. Third, to confirm the impact of physics loss in training, we observe that the evaluation metrics of PIML-GBM are significantly smaller than those of PMNN in Table 2. We can find that the influence of physics knowledge in training MNN improves the accuracy of PIML-GBM.

Table 2: Evaluation Metrics for GBM, PMNN, and PIML-GBM

Model	$MAE(\hat{\mathbf{K}}_{test}, \mathbf{Y}_{test})$ $(\times 10^{-2})$	$RMSE(\hat{\mathbf{K}}_{test}, \mathbf{Y}_{test})$ $(\times 10^{-2})$	$Err(\hat{\mathbf{K}}_{test}, \mathbf{Y}_{test})$ $(\times 10^{-1})$
GBM	37.7495	47.4828	2.7272
PMNN	4.6295	10.7184	7.9794
PIML-GBM	**2.2629**	**5.3300**	**3.3946**

We visualize the estimations of each model $\log \hat{K}(t, x; \mathbf{w})$, values of errors $(\log \hat{K}(t, x; \mathbf{w}) - \log K(t, x))$, and absolute values of errors $|\log \hat{K}(t, x; \mathbf{w}) - \log K(t, x)|$ in Figure 9. The first and second columns of Figure 9 are plots of ground-truth's values and estimations, respectively. The third column of Figure 9 is values of errors $(\log \hat{K}(t, x; \mathbf{w}) - \log K(t, x))$ and the fourth column of Figure 9 is the absolute error values $|\log \hat{K}(t, x; \mathbf{w}) - \log K(t, x)|$. There are two colors in the third column of Figure 9: red indicates underestimations, and blue indicates overestimations. If the colors in the fourth column of Figure 9 are light, it means that the models estimate $\hat{K}(t, x; \mathbf{w})$ close to ground-truth $K(t, x)$.

First, we can observe that GBM has poor accuracy and there are overestimations and underestimations in the third column of Figure 9. In Figure 9, GBM is overestimated on the upper right side of the plot (3) and is underestimated on the lower left side of the plot (3). Since GBM assumes the stable relationship between speed and density based on MFD, the evolution of $K(t, x)$ in a time-distance domain cannot fully capture traffic dynamics from GBM. The plot (4) in Figure 9 shows that significant errors exist over the time-distance domain. Second, we can observe that PMNN cannot estimate $\hat{K}(t, x; \mathbf{w})$ in the time-distance domain except for boundary points from plots (5) and (6) of Figure 9. Since PMNN without physics loss cannot learn the evolution of K(t,x) from auxiliary points, it focuses on estimating only boundary points. Third, PIML-GBM outperforms two baselines in Table 2 and Figure 9. The plot (9) in Figure 9 shows that PIML-GBM has fewer underestimations and overestimations than two baselines and better estimation at any point in the time-distance domain.

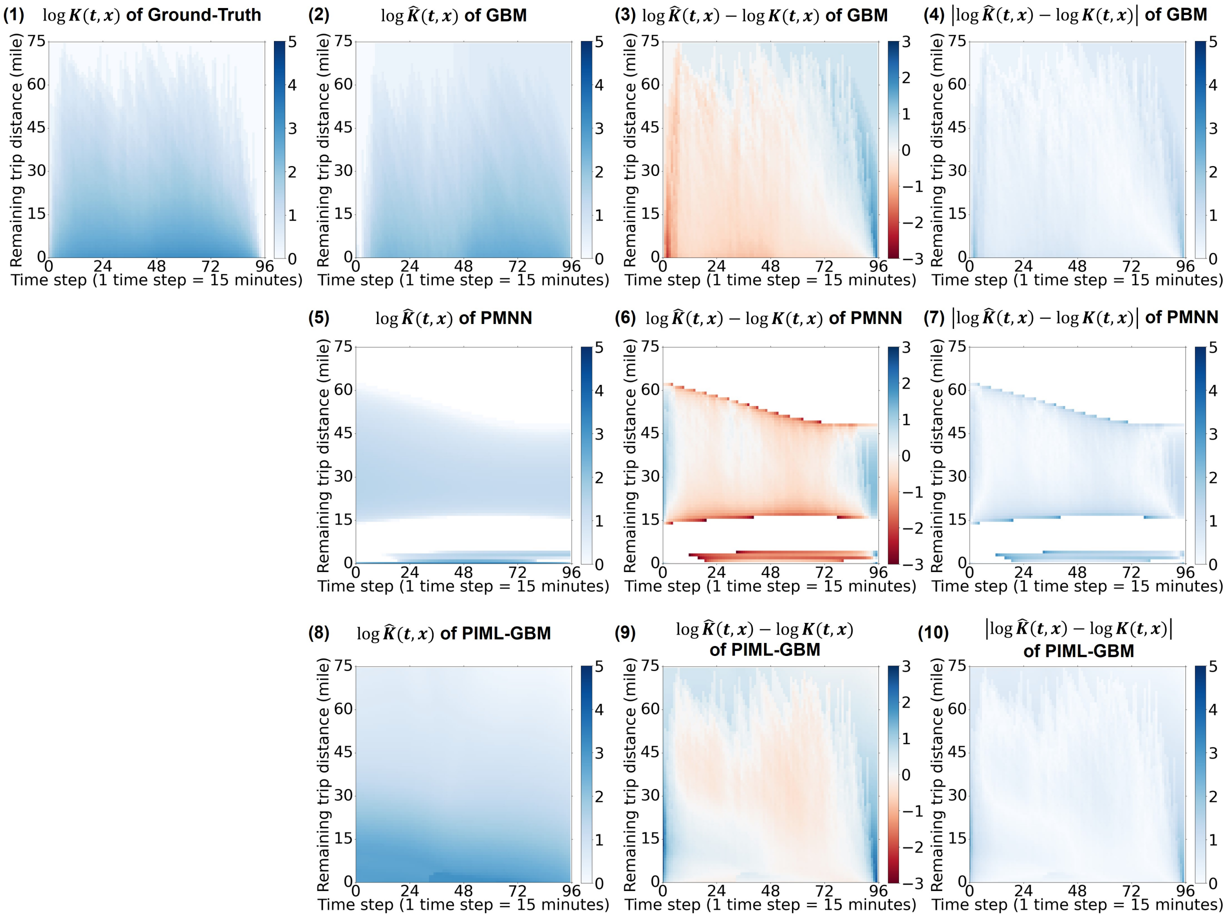

Figure 9: Comparison between PIML-GBM and the two baselines: (First column) Ground-truth values, (Second column) Estimations of each model, (Third column) Values of each model's errors, (Fourth column) Absolute Values of each model's errors. In the third column, the red color means underestimation, and the blue color means overestimation.

5.3. Sensitivity Analysis

This section conducts a sensitivity analysis with respect to the number of neurons in each hidden layer of MNN from 25 to 45 neurons, which are shown in previous studies [40, 41, 39, 51, 52]. There is a trade-off between accuracy and training speed when setting the number of neurons. We can get more accuracy if we set more neurons in each hidden layer, but the training time is longer. Furthermore, we can set more weight between physics loss $\mathcal{L}_{phy}$ and data loss $\mathcal{L}_{data}$ through α. If we set a larger α, we consider data loss $\mathcal{L}_{data}$ is more important than physics loss $\mathcal{L}_{phy}$.

To determine the optimal weight value, denoted by (α^*), we train PIML-GBM models across various (α) values in the set $\{0, 0.1, \ldots, 0.9, 1.0\}$. We then select the best performing (α^*) for each specific number of neurons, balancing between physics loss ($\mathcal{L}_{\text{phy}}$) and data loss ($\mathcal{L}_{\text{data}}$). The evaluation metrics for the PMNN and PIML-GBM models with varying numbers of neurons are detailed in Table 3. The results indicate that as the number of neurons in PMNN models increases, most evaluation metrics show improvement, although there is an exception at 45 neurons. Specifically, PMNN models employing 30 and 35 neurons exhibit nearly identical evaluation metrics. In contrast, the evaluation metrics for PIML-GBM do not reveal a clear trend.

Table 3: Evaluation Metrics of Sensitivity Analysis with PMNN and PIML-GBM

Number of Neurons	PMNN			PIML-GBM			
	MAE ($\times 10^{-2}$)	$RMSE$ ($\times 10^{-2}$)	Err ($\times 10^{-1}$)	α^*	MAE ($\times 10^{-2}$)	$RMSE$ ($\times 10^{-2}$)	Err ($\times 10^{-1}$)
25 neurons	2.9475	5.1829	3.8177	0.5	2.3723	5.6759	3.4710
30 neurons	4.2792	9.4485	6.9407	0.9	2.5370	5.3285	**2.9637**
35 neurons	4.2792	9.4485	6.9407	0.7	2.4788	5.4263	3.1421
40 neurons	4.6295	10.7184	7.9794	0.4	**2.2629**	**5.3300**	3.3946
45 neurons	4.6220	11.2128	8.3103	0.6	2.3028	5.4240	3.1837

The experimental results show that the proposed PIML-GBM overcomes each limitation of the GBM and PMNN by regularizing the model through loss of physics knowledge and using MNN without explicit theoretical assumptions in large-scale road networks.

6. Conclusion

In this research, we have addressed the intricate challenge of estimating traffic states within large-scale urban road networks. The study introduces the PIML-GBM model, an innovative approach designed to leverage the governing equation of the generalized bathtub model to estimate traffic states using mobile location data. This model aims to provide a robust solution, capturing the randomness and dynamics of urban traffic without relying on rigid theoretical assumptions. Our primary contributions include:

- Developing a deep neural network within the PIML-GBM model to estimate traffic state variables without explicit theoretical constraints.
- Demonstrating the ability of the PIML-GBM to accurately capture the randomness of traffic dynamics, offering critical insights for real-world applications.
- Modeling traffic dynamics over a continuous time-distance domain while using boundary points from a discretized time-distance domain.
- Illustrating the promising potential of utilizing mobile location data in large-scale road networks.

The study embarked on a comprehensive examination of the PIML-GBM by contrasting it against established numerical solutions, such as GBM and PMNN. We demonstrated the superiority of PIML-GBM in modeling traffic dynamics across a continuous time-distance domain using location-based data collected from a mobile phone vendor within the Indianapolis road network, United States:

- **Superior Performance of PIML-GBM**: The PIML-GBM has exhibited outstanding accuracy in estimating the values of $K(t, x)$, particularly in comparison to GBM and PMNN. The model's integration of physics loss in training ensures higher accuracy, as reflected in the evaluation metrics.
- **Adaptation to Real-world Data**: Unlike traditional methods that rely on theoretically ideal assumptions, PIML-GBM, along with PMNN, can capture the complex variance and phenomena from real-world data. This adaptability provides a robust foundation for estimating widely ranged values and highlights the advantage of employing ML-based methods.

In summary, the proposed PIML-GBM model marks a significant advancement in the field, providing a more nuanced, accurate, and flexible tool for understanding and predicting complex systems. Its integration of physics knowledge with machine learning techniques not only bridges the gap between theory and practice but also opens new avenues for interdisciplinary research and applications.

The introduction of the PIML-GBM model marks a crucial development in the field, providing both researchers and practitioners with a versatile and precise tool to unravel the complex dynamics of urban mobility. The methodological developments and findings of this work illuminate existing challenges and propose pioneering avenues for ongoing investigation, contributing to the development of intelligent, resilient urban traffic systems. Despite the significant contributions, certain limitations exist that necessitate future works:

- The governing equations may be compromised due to high noise in real-world traffic data, resulting in a loss of physics knowledge. This issue may be mitigated by adding noise to nodes within hidden layers and implementing an attention mechanism to discern non-linear features.
- While the PIML-GBM model effectively estimates current traffic states, predicting future states remains challenging. Future research could extend the PIML-GBM model to forecast traffic states by utilizing a rolling-horizon technique for updating current conditions and employing generative models to sample future data.

In conclusion, this study constitutes a significant step in enhancing our understanding of urban traffic systems. The proposed PIML-GBM model offers an intersection of theoretical advancement and practical utility that can guide the field's future direction. It lays the groundwork for future exploration, pushing the boundaries of our vision for intelligent and resilient urban traffic networks.

References

[1] X. Qian, T. Lei, J. Xue, Z. Lei, S. V. Ukkusuri, Impact of transportation network companies on urban congestion: Evidence from large-scale trajectory data, Sustainable Cities and Society 55 (2020) 102053.

[2] D. Schrank, L. Albert, B. Eisele, T. Lomax, et al., 2021 Urban mobility report, Texas Transportation Institute, 2021.

[3] N. Geroliminis, C. F. Daganzo, Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings, Transportation Research Part B: Methodological 42 (9) (2008) 759–770.

[4] M. Johari, M. Keyvan-Ekbatani, L. Leclercq, D. Ngoduy, H. S. Mahmassani, Macroscopic network-level traffic models: Bridging fifty years of development toward the next era, Transportation Research Part C: Emerging Technologies 131 (2021) 103334.

[5] C. F. Daganzo, N. Geroliminis, An analytical approximation for the macroscopic fundamental diagram of urban traffic, Transportation Research Part B: Methodological 42 (9) (2008) 771–781.

[6] M. Keyvan-Ekbatani, M. Papageorgiou, I. Papamichail, Urban congestion gating control based on reduced operational network fundamental diagrams, Transportation Research Part C: Emerging Technologies 33 (2013) 74–87.

[7] C. F. Daganzo, Urban gridlock: Macroscopic modeling and mitigation approaches, Transportation Research Part B: Methodological 41 (1) (2007) 49–62.

[8] J. Haddad, N. Geroliminis, On the stability of traffic perimeter control in two-region urban cities, Transportation Research Part B: Methodological 46 (9) (2012) 1159–1176.

[9] K. Aboudolas, N. Geroliminis, Perimeter and boundary flow control in multi-reservoir heterogeneous networks, Transportation Research Part B: Methodological 55 (2013) 265–281.

[10] M. Yildirimoglu, N. Geroliminis, Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams, Transportation Research Part B: Methodological 70 (2014) 186–200.

[11] G. Mariotte, L. Leclercq, S. Batista, J. Krug, M. Paipuri, Calibration and validation of multi-reservoir mfd models: A case study in lyon, Transportation Research Part B: Methodological 136 (2020) 62–86.

[12] I. I. Sirmatel, D. Tsitsokas, A. Kouvelas, N. Geroliminis, Modeling, estimation, and control in large-scale urban road networks with remaining travel distance dynamics, Transportation Research Part C: Emerging Technologies 128 (2021) 103157.

[13] L. Leclercq, C. Parzani, V. L. Knoop, J. Amourette, S. P. Hoogendoorn, Macroscopic traffic dynamics with heterogeneous route patterns, Transportation Research Procedia 7 (2015) 631–650.

[14] M. Johari, M. Keyvan-Ekbatani, L. Leclercq, D. Ngoduy, H. S. Mahmassani, Macroscopic network-level traffic models: Bridging fifty years of development toward the next era, Transportation Research Part C: Emerging Technologies 131 (2021) 103334.

[15] G. Mariotte, L. Leclercq, J. A. Laval, Macroscopic urban dynamics: Analytical and numerical comparisons of existing models, Transportation Research Part B: Methodological 101 (2017) 245–267.

[16] N. Geroliminis, J. Sun, Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks, Procedia-Social and Behavioral Sciences 17 (2011) 213–228.

[17] R. Lamotte, N. Geroliminis, The morning commute in urban areas with heterogeneous trip lengths, Transportation Research Part B: Methodological 117 (2018) 794–810.

[18] R. Lamotte, M. Murashkin, A. Kouvelas, N. Geroliminis, Dynamic modeling of trip completion rate in urban areas with mfd representations, in: 2018 TRB Annual Meeting Online, Transportation Research Board, 2018, pp. 18–06192.

[19] L. Leclercq, M. Paipuri, Macroscopic traffic dynamics under fast-varying demand, Transportation science 53 (6) (2019) 1526–1545.

[20] W. Vickrey, Congestion in midtown manhattan in relation to marginal cost pricing, Economics of Transportation 21 (2020) 100152.

[21] Y. Bao, E. T. Verhoef, P. Koster, Leaving the tub: The nature and dynamics of hypercongestion in a bathtub model with a restricted downstream exit, Transportation Research Part E: Logistics and Transportation Review 152 (2021) 102389.

[22] R. Arnott, J. Buli, Solving for equilibrium in the basic bathtub model, Transportation Research Part B: Methodological 109 (2018) 150–175.

[23] W.-L. Jin, Generalized bathtub model of network trip flows, Transportation Research Part B: Methodological 136 (2020) 138–157.

[24] W.-L. Jin, I. Martinez, M. Menendez, Compartmental model and fleet-size management for shared mobility systems with for-hire vehicles, Transportation Research Part C: Emerging Technologies 129 (2021) 103236.

[25] C. Buisson, C. Ladier, Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams, Transportation Research Record 2124 (1) (2009) 127–136.

[26] V. L. Knoop, H. Van Lint, S. P. Hoogendoorn, Traffic dynamics: Its impact on the macroscopic fundamental diagram, Physica A: Statistical Mechanics and its Applications 438 (2015) 236–250.

[27] Q.-J. Gan, W.-L. Jin, V. V. Gayah, Analysis of traffic statics and dynamics in signalized networks: a poincaré map approach, Transportation science 51 (3) (2017) 1009–1029.

[28] M. Ameli, M. S. S. Faradonbeh, J.-P. Lebacque, H. Abouee-Mehrizi, L. Leclercq, Departure time choice models in urban transportation systems based on mean field games, Transportation Science 56 (6) (2022) 1483–1504.

[29] R. Arnott, M. Kilani, Social optimum in the basic bathtub model, Transportation Science 56 (6) (2022) 1505–1529.

[30] Y. Wang, D. Zhang, Y. Liu, B. Dai, L. H. Lee, Enhancing transportation systems via deep learning: A survey, Transportation research part C: emerging technologies 99 (2019) 144–163.

[31] S. Tak, S. Woo, H. Yeo, Data-driven imputation method for traffic data in sectional units of road links, IEEE Transactions on Intelligent Transportation Systems 17 (6) (2016) 1762–1771.

[32] J. Ke, H. Zheng, H. Yang, X. M. Chen, Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach, Transportation research part C: Emerging technologies 85 (2017) 591–608.

[33] Y. Wu, H. Tan, L. Qin, B. Ran, Z. Jiang, A hybrid deep learning based traffic flow prediction method and its understanding, Transportation Research Part C: Emerging Technologies 90 (2018) 166–180.

[34] X. Ma, Z. Tao, Y. Wang, H. Yu, Y. Wang, Long short-term memory neural network for traffic speed prediction using remote microwave sensor data, Transportation Research Part C: Emerging Technologies 54 (2015) 187–197.

[35] Z. Zhang, M. Li, X. Lin, Y. Wang, F. He, Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies, Transportation research part C: emerging technologies 105 (2019) 297–322.

[36] J. Xue, N. Jiang, S. Liang, Q. Pang, T. Yabe, S. V. Ukkusuri, J. Ma, Quantifying the spatial homogeneity of urban road networks via graph neural networks, Nature Machine Intelligence 4 (3) (2022) 246–257.

[37] D. Zhou, V. V. Gayah, Model-free perimeter metering control for two-region urban networks using deep reinforcement learning, Transportation Research Part C: Emerging Technologies 124 (2021) 102949.

[38] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

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

[40] R. Shi, Z. Mo, K. Huang, X. Di, Q. Du, A physics-informed deep learning paradigm for traffic state and fundamental diagram estimation, IEEE Transactions on Intelligent Transportation Systems (2021).

[41] R. Shi, Z. Mo, X. Di, Physics-informed deep learning for traffic state estimation: A hybrid paradigm informed by second-order traffic models, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, 2021, pp. 540–547.

[42] Y. Yuan, Z. Zhang, X. T. Yang, S. Zhe, Macroscopic traffic flow modeling with physics regularized gaussian process: A new insight into machine learning applications in transportation, Transportation Research Part B: Methodological 146 (2021) 88–110.

[43] L. Leclercq, N. Chiabaut, B. Trinquier, Macroscopic fundamental diagrams: A cross-comparison of estimation methods, Transportation Research Part B: Methodological 62 (2014) 1–12.

[44] V. V. Gayah, V. V. Dixit, Using mobile probe data and the macroscopic fundamental diagram to estimate network densities: Tests using microsimulation, Transportation research record 2390 (1) (2013) 76–86.

[45] C. Huang, N. Zheng, J. Zhang, Investigation of bimodal macroscopic fundamental diagrams in large-scale urban networks: empirical study with gps data for shenzhen city, Transportation Research Record 2673 (6) (2019) 114–128.

[46] M. Paipuri, Y. Xu, M. C. González, L. Leclercq, Estimating mfds, trip lengths and path flow distributions in a multi-region setting using mobile phone data, Transportation Research Part C: Emerging Technologies 118 (2020) 102709.

[47] L. Lu, X. Meng, Z. Mao, G. E. Karniadakis, Deepxde: A deep learning library for solving differential equations, SIAM Review 63 (1) (2021) 208–228.

[48] U.S. Census Bureau, Population in Marion County (2019), https://data.census.gov/cedsci/table?q=population%20in%20marion%20county%20in%202019&tid=PEPPOP2019.PEPANNRES, online; accessed 8 July 2022 (2019).

[49] M. Batran, M. G. Mejia, H. Kanasugi, Y. Sekimoto, R. Shibasaki, Inferencing human spatiotemporal mobility in greater maputo via mobile phone big data mining, ISPRS International Journal of Geo-Information 7 (7) (2018) 259.

[50] W.-L. Jin, Y. Yu, Performance analysis and signal design for a stationary signalized ring road, arXiv preprint arXiv:1510.01216 (2015).

[51] C. Rao, H. Sun, Y. Liu, Physics-informed deep learning for incompressible laminar flows, Theoretical and Applied Mechanics Letters 10 (3) (2020) 207–212.

[52] A. J. Huang, S. Agarwal, Physics-informed deep learning for traffic state estimation: illustrations with lwr and ctm models, IEEE Open Journal of Intelligent Transportation Systems 3 (2022) 503–518.

Appendix A. Summary Table of All Notations in This Study

Category	Notation	Definition
Domain		
	$\mathbf{T}$	Time horizon (temporal domain)
	$\mathbf{X}$	Spatial (trip-distance) domain
	$\mathbf{S}$	Time-distance domain
	$\mathbf{G}$	Discretized time-distance domain
Sets		
	$\mathbf{K}$	Set of the number of trips with remaining distance $\geq x$ at time t $K(t,x)$
	$\mathbf{Q}$	Set of traffic variables related to entering trips
	$\mathbf{O}$	Set of observed boundary points
	$\mathbf{A}$	Set of auxiliary points
	$\mathbf{Y}$	Set of ground-truth (or observed) labels
	$\mathbf{w}$	Set of neural network parameters
	$\mathbf{w}^*$	Set of optimal neural network parameter after the training process
	$\mathbf{O}_{train}$	Set of training points from observed boundary points
	$\mathbf{Y}_{train}$	Set of training labels from observed boundary points
	$\mathcal{N}(\mathbf{K}, \mathbf{Q}; \Lambda)$	Governing equations of GBM
Parameters		
	L_{net}	Road network's total length of road segments
	T	Total time step
	$X_{\max}$	Maximum of trip distance
	Λ	Parameters of generalized bathtub model
	N_g	Number of grid points in discretized time-distance domain $\mathbf{G}$
	N_t	Number of time grid points in discretized time-distance domain $\mathbf{G}$
	N_d	Number of distance grid points in discretized time-distance domain $\mathbf{G}$
	N_o	Number of observed boundary points in $\mathbf{O}$
	N_a	Number of auxiliary points in $\mathbf{A}$
	r_{train}	Sample rate for training points
	n_a	Multiplier of auxiliary points
	α	Weight between data and physics losses
	ϵ_k	Allowable loss gap between consecutive total losses
	n_{max_iter}	Allowable maximum number of iterations

Category	Notation	Definition
Variables and Functions		
	$v(t)$	Average speed of vehicles running on the road network at time t
	$\rho(t)$	Average density per unit road length at time t
	$V(\cdot)$	Function of traffic density based on macroscopic fundamental diagram
	$q(t)$	Average traffic flow rate at time t
	$f(t)$	Entering trip (in-flow) rates at time t
	$F(t)$	Cumulative entering trip (in-flow) rates at time t
	$\tilde{\varphi}(t,x)$	Probability density function of the remaining trip distance x at time t
	$\tilde{\Phi}(t,x)$	Cumulative distribution function of the entering trips with distances not smaller than x at time t
	$\tilde{B}(t)$	Average distance of entering trips at time t
	$\varphi(t,x)$	Probability density function of the remaining trip distance x at time t
	$\Phi(t,x)$	Cumulative distribution function of the trips with remaining distances not smaller than x at time t
	$B(t)$	Average distance of remaining trips at time t
	$\lambda(t)$	Number of active trips (traveling vehicles) at time t
	$k(t,x)$	Density of active trips with a remaining distance x at time t
	$K(t,x)$	Number of trips with a remaining distance not smaller than x at time t
	$g(t)$	Outflow rate of exiting trips at time t
	$G(t)$	Cumulative outflow rate of exiting trips until time t
	$\hat{K}(t,x;\mathbf{w})$	Estimation of $K(t,x)$ from MNN in PIML-GBM with neural network parameter $\mathbf{w}$
	$\mathcal{L}_{\alpha}$	Total loss in PIML-GBM
	$\mathcal{L}_{data}$	Data loss in PIML-GBM
	$\mathcal{L}_{phy}$	Physics loss in PIML-GBM
	$v_r(t)$	Average speed in the road network at time t from observed data