A Physics-Informed Spatiotemporal Network for Traffic Prediction

Haoyang Chen

Student ID: 1276594

Supervisors:

Dr. Joyce Zhang

Dr. Mingming Gong

Dr. Jianzhong Qi

Dr. Tingjin Chu

Master of Computer Science

School of Computing and Information Systems

The University of Melbourne

July 11, 2023

Table of Contents

1 Introduction

As part of the smart city project and Intelligent Transportation System (ITS), traffic prediction plays an essential role in urban traffic planning, management and control [3, 40]. The task focuses on predicting future traffic states such as traffic flow or speed for a given time span and region provided with historical traffic information, road condition or other external data (weather, accidents, etc). With the rise of deep learning techniques and the increasing availability of traffic data, studies on traffic prediction have gradually shifted focus from statistical models to neural networks [58]. Enormous research effort devoted to the area in the past two to three decades focuses on capturing the complex spatiotemporal dependencies within traffic data and has pushed the boundary of the research field [40, 52]. Despite the great success of deep learning-based methods for traffic prediction, several intrinsic deficiencies are still present. Firstly, these approaches rely heavily on sufficient data sources and learnable short- and long-term data patterns. The availability of complete traffic data for deep learning is still rare, with such sources mostly found in public benchmark datasets (PeMS, META-LA, etc). Moreover, due to accidents or events, outliers in traffic data are likely to occur, which can worsen the performance of models. Secondly, the effectiveness of deep learning-based methods mainly depends on the carefully-designed model architecture. With the increasing complexity of neural networks, the process is prone to error and requires greater effort in solution ideas. Lastly, neural networks are purely data-driven, which largely ignores the underlying traffic flow physics studied for decades [18, 32, 16].

Compared to research efforts in designing spatiotemporal architectures, embedding physics into deep learning models remains under-explored. For the task of traffic prediction, model-driven approaches and data-driven approaches stay largely separated. In recent years, physics-informed neural network (PINN) has received widespread attention in various research fields [26, 45, 65]. Despite several attempts with this structure in traffic prediction [22, 57], few have provided insights into incorporating observational and inductive biases into the data and neural networks. Furthermore, none of the research has demonstrated a scalable and general model for traffic networks, thus preventing the application of the method to the spatiotemporal domains.

In this research, we propose to investigate physics-informed techniques as observational, inductive and learning biases to be embedded into deep learning architectures for traffic prediction, which is much less studied than pure data-driven or model-driven approaches. We will also study the dynamic division of traffic regions based on data, such that the physics to be incorporated can be adapted for different regions' traffic conditions. These techniques may provide different strengths and weaknesses and therefore complement existing deep learning-based methods. This project may also contribute to integrating traffic flow physics with spatiotemporal neural networks. We plan to first examine existing literature on physics-informed approaches as well as pure deep learning techniques for traffic prediction and then identify suitable ways to embed the physics in. Then, we will compare the physics-informed methods with either model-driven or data-driven models alone to evaluate the effectiveness of the proposed physics-informed architecture. Through these steps, we aim to identify the merits, demerits and complementarity of physics-informed deep learning for traffic prediction, while also providing solution insights of incorporation approaches from multiple perspectives.

2 Related Works

2.1 Overview

Traffic prediction concerns predicting future traffic conditions based on historical traffic data or a statistical model. There is a list of sub-tasks related to different attributes of traffic, such as traffic state prediction, trajectory prediction, estimated time of arrival and map matching [59]. As a core technology of Intelligent Transportation Systems, traffic state prediction has received widespread attention in the past few decades [3, 21, 6, 43, 24, 69, 39]. Two commonly used variables for the task are traffic speed and flow, which is the main focus of our research. To provide a more detailed analysis of traffic state prediction, the following aspects will be covered:

The first section will discuss different categories of traffic prediction models, including statistical models, machine learning models, and hybrid models. For each category, the section will provide an overview of the most commonly used algorithms, their strengths and weaknesses, and their applications in traffic prediction.

The second section will delve into the various data sources used for traffic prediction, such as traffic sensor data, GPS data, and social media data. Additionally, it will examine how these data sources are represented and pre-processed for use in traffic prediction models.

The third section will introduce the fundamental concepts of traffic flow theory, which provides the theoretical foundation for traffic prediction. We will discuss the key principles of traffic flow theory, such as the fundamental diagram and the macroscopic traffic flow models, and how they can be used to improve traffic prediction accuracy.

By covering these three aspects, we aim to provide a comprehensive overview of traffic prediction, highlighting the current state-of-the-art techniques, along with the challenges and opportunities in this field.

2.2 Traffic Prediction

Traffic forecasting approaches can be categorized based on two main factors. The first factor is whether or not data is used. Data-driven models learn the model structure from historical traffic data, while model-driven approaches express the model in a mathematical form. Models can also be categorized as parametric or non-parametric. Parametric models assume a specific form for the model parameters, while non-parametric models do not make any assumptions about the underlying distribution of the data. These categorizations provide a useful framework for understanding the different approaches taken in traffic forecasting, and we will discuss some of the important models for each category.

2.2.1 Model-driven vs Data-driven Approaches

Early research on traffic prediction tends to focus on model-driven approaches due to insufficient data. Many mathematical or statistical models have been proposed and optimised with the purpose of accurately reproducing the dynamics of traffic flow, which can be divided into hydrodynamic models, kinetic models and microscopic models [48]. However, since traffic is stochastic in nature, the internal and external dynamics are hard to model without any assumptions. Thus, great limitations have been posed to model-driven approaches for practical use.

In the recent decade, with the rapid development of sensor technology and increasing availability of traffic databases, data-driven models have received widespread attention in the research field of traffic prediction [40, 52]. Historical Average (HA) and Instantaneous Travel Times are among the earliest models to be applied to the time series of historical traffic data [25, 33]. By estimating the past sequence in a naive way, forecasts can be made in terms of future traffic conditions. However, these models fail to capture the non-stationary and nonlinear patterns in the time series, making the forecasts unreliable and thus impractical to use. Research efforts after the naive models could be divided into parametric and non-parametric models, which will be discussed in the next section.

Either model-driven or data-driven approach alone has its limitations. Model-driven methods typically require complex parameter calibration or mathematical calculations, which pose great challenges to computing resources. On the other hand, data-driven models often require sufficient data to unveil the underlying patterns. Moreover, such models tend to lack interpretability. Hybrid methods aim to combine the advantages of both approaches. A hybrid model typically approximates traffic states with non-linear differential equations and uses traffic data to calibrate the parameters. The traffic flow model was first incorporated into the deep learning models with a hidden physics model [46], which uses nonlinear partial differential equations to extract patterns from high-dimensional data. Some other studies combined the Lighthill-Whisham-Richards (LWR) models and the cell transmission model

(CTM) [38, 28]. Those models tend to focus on general conditions without considering the possible congestion in the road network. The latest study by [41] proposed a hybrid stepwise modelling framework to deal with the congestion situation, and can better capture the dynamic traffic flow states.

2.2.2 Parametric vs Non-parametric Models

Data-driven approaches have been widely explored in traffic state prediction in the past decade and have achieved state-of-the-art results in numerous benchmarks. The methods could mainly be categorised into parametric and non-parametric models.

As early attempts with parametric models, Box-Jenkins Auto-regressive Integrated Moving Average (ARIMA) was applied in traffic prediction [21]. The study demonstrated that an order of (0,1,1) ARIMA is adequate for reproducing the original time series. The model has been widely adopted since and many variants were proposed to improve the model performance [61, 9, 66]. The Kalman filter is another popular parametric model for traffic prediction [17, 29, 11]. The algorithm uses a weighted average of the current and previous measurement of traffic states to estimate the future state, where the weights are determined by the uncertainties associated with each time step. Similarly, Support Vector Regression Machine (SVR) [10] and Vector Auto-Regression (VAR) [6, 34] also considers time-series only. None of the above models can address the spatial correlations between different nodes in a road network. The temporal dimension only considers a single node, which is impractical in real-world traffic situations. Furthermore, traffic is stochastic. Without enough assumptions on the road and traffic conditions, parametric models are unable to reproduce the accurate traffic state.

With the rise of the availability of traffic data, machine learning methods have attracted an increasing amount of attention in traffic prediction. Bayesian Network (BN) and K-Nearest Neighbors (KNN) are among the first to be adopted in the task [51, 50]. Sun et al. [50] attempted to use the adjacent road links to predict the trend of the current link and proved it an effective approach for incomplete data. Pascale et al. [43] proposed an adaptive Bayesian network that adapts the graph structure to the local traffic phase. The study suggests that the optimised graphical model provides a reduced complexity compared to conventional BN while still achieving similar prediction accuracy. Cai et al., [5] introduced an enhanced KNN model based on spatiotemporal correlation and has demonstrated superiority over other existing KNN methods. However, machine-learning-based methods suffer from capturing the non-linearity and the complex spatiotemporal relationships, which makes them less effective in traffic prediction compared to their performance in other areas.

In the past decade, deep learning-based approaches have gained great popularity in multiple disciplines. The neural network-based approach was first adopted in transportation research in 2014 [23]. The model proposed consists of a deep belief network at the bottom and a regression layer at the top. By combining unsupervised feature extraction and supervised learning, the model achieved a 5% improvement in the state-of-the-art in that period. In the same year, an auto-encoder architecture was proposed for traffic flow feature representation as well as prediction [35]. The model stacks multiple autoencoders to process the features and uses a logistic regression layer for the output prediction. Since then, numerous neural network architectures have been proposed [69, 52].

A fundamental challenge for traffic prediction is to model the spatiotemporal dependencies. The early solution uses a convolutional neural network (CNN) to model grid-based node dependencies, and a recurrent neural network (RNN) for learning the temporal patterns [70, 67]. Li et al. proposed a diffusion convolutional recurrent neural network (DCRNN) to model spatial dependencies with bidirectional random walks on the graph and temporal dependencies with encoder-decoder RNN [31]. Zhang et al. proposed Spatio-temporal residual networks (ST-ResNet) to combine and aggregate three residual convolutional units for jointly modelling the pedestrian inflow and outflow [70]. Based on this architecture, research efforts have then been devoted to improving either the temporal or the spatial side of the model.

On the temporal dimension, both Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) have shown promising results towards the problem [54, 14]. Studies also suggest that GRU is a little better than LSTM and usually converges faster [14]. Although the RNN-based methods capture the historical temporal dependencies well, it suffers from capturing long-term dependencies. It is also not parallelisable due to the inherent sequential structure. To solve the problem, transformers and other attention-based models have been experimented with. A traffic transformer was proposed in 2021 to overcome the inherent RNN deficiencies [4]. Four position encoding strategies are introduced in the paper to capture the continuity and periodicity of time series. This was the first attempt to apply transformer architecture to traffic prediction models. After this, many self-attention-based models have been proposed and state-of-the-art results were frequently updated since [72, 64, 20, 24].

Research on capturing the spatial dependencies tends to adopt Graph-Neural-Network(GNN)-based approaches. A study by Yu et al. [68] showed that GNN is better at representing complex road connections since the network is dominated by topological order. Wu et al. [63] proposed Graph WaveNet for deep spatiotemporal graph modelling. In the study, a novel adaptive dependency matrix is learned through node embedding to precisely capture the hidden spatial dependency in the data. In 2020, Multivariate Graph Neural Network (MTGNN) was introduced in [62]. The model incorporates a graph learning layer to optimise the graph structure before feeding into the convolutional modules, which makes it applicable to general graphs. Other attempts seek to combine attention and GNN-based methods and have demonstrated promising results [72, 12, 44].

2.3 Traffic Flow Theory

2.3.1 Macroscopic Variables

The basic variables for traffic can be divided into microscopic level and macroscopic level. Microscopic level refers to individual vehicles/drivers level and macroscopic level refers to flow/stream level. This section describes some macroscopic level variables.

1. Density. Density k reflects the number of vehicles per kilometre of road. For a measurement interval at a certain point in time, such as S_1, k can be calculated over a road section with ΔX length as:
$$k(x_1, t_1, S_1) = \frac{n}{\Delta X}, \tag{1}$$
where the index n indicates the number of vehicles at t_1 on the location interval ΔX. For the location S_1, we take the centre of the measurement interval. The maximal density on a road fluctuates around 100 vehicles per kilometre per traffic lane.

2. Flow Rate (Flow). The flow rate q represents the number of vehicles that passes a certain cross-section per time unit. For the time interval ΔT at any location x_2, such as the measurement interval S_2, the flow rate is calculated as follows:
$$q(x_2, t_2, S_2) = \frac{m}{\Delta T}, \tag{2}$$
where the index m represents the number of vehicles that passes location x_2 during ΔT. This concept is usually exchangeable with volume.

3. Mean Speed. Mean speed u is defined as the quotient of the flow rate and the density. It is also called the ***fundamental relation*** of traffic flow theory:
$$q = k \cdot u. \tag{3}$$

Knowing two of these variables immediately leads to the remaining third variable. Mathematically, it is called the ***continuity equation***. The definition could be further divided into space-mean speeds and time-mean speeds where the mean is calculated across a specific region or period. The precise definitions are also provided:

$$v_t = \frac{\sum_{i=1}^{n} v_i}{n} \tag{4}$$

$$v_s = \frac{n}{\sum_{i=1}^{n} \frac{1}{v_i}} \tag{5}$$

where v_t is the time-mean speed, v_s is the space-mean speed, v_i is the speed of the i^{th} vehicle passing the fixed point on the road, and n is the number of measured vehicle spot speed. Note that the equation can only be used for space-mean speeds but not time-mean speeds [19].

2.3.2 Macroscopic Fundamental Diagram

The fundamental relationship between network-aggregated traffic flow and density (a.k.a occupancy)is commonly known as a ***Macroscopic Fundamental Diagram (MFD)*** usually graphed as a curve. It has been a cornerstone for traffic modelling that associates average density, flow and speed in an urban network [16].

Early research has proved the existence of the fundamental diagram/relationship of traffic variables in urban areas. A study on data collected from streets in central London showed a seemingly linear-decreasing relationship between traffic speed and flow [53]. In the next year, Wardrop [60] introduced a generic relationship between average speed and flow, though still decreasing monotonically. It was not until the recent decade that the research effort started to shift from network-wide analysis to link(individual lane)-level [16, 15]. The studies suggest that the MFD shows a reproducible shape for urban transportation networks, which facilitates its application to traffic planning and control in recent years.

The MFDs were expressed as empirical diagrams or sampled curves in early studies.

Though the forms provided valuable insights for traffic analysis, they lack systematic parameterisation and mathematical applicability. To address this problem, several functional forms of the diagram have been proposed. Lukas et al. introduced a new form of MFD based on the smooth approximation of a time-invariant upper bound of traffic state (uMFD)[2]. A trapezoidal shape along with a new parameter λ are used in the model, which refined existing analytical approaches that usually result in multi-regime functions. Méndez et al[37] introduced a multi-class fundamental diagram where both slow and fast user-class are discussed. The paper provides complete numerical derivation for the theoretical model, which provides new perspectives for modelling the MFD.

2.4 Physics-Informed Neural Networks

Physics-Informed Neural Networks (PINNs) are a class of deep learning models incorporated with mathematically or physically meaningful constraints. Such models are developed with some extent of prior knowledge that can yield more interpretable results, and help improve the robustness of machine learning (ML) algorithms when faced with inconsistent or incomplete data.

Karniadakis et al. identified three ways to incorporate physics into machine learning (ML): observational biases, inductive biases, and learning biases [26]. Observational biases involve augmenting or transforming data using prior knowledge of underlying physics. Inductive biases can be embedded by modifying the ML model architecture such that it satisfies specific physical or mathematical principles. Lastly, learning biases refer to loss constraints or inference algorithms that guide the model's training process. While learning biases provide more flexibility, they may not be fully satisfied, unlike inductive biases, which guarantee compliance with desired principles.

Due to the increasing power of ML technologies and the availability of big data, observational and inductive biases are widely explored for decades. Typical examples include convolutional neural network (CNN) [30] and recurrent neural network (RNN) [1], governing the learning of intrinsic geometry and sequential characteristics respectively. Common learning biases are introduced in some variants of PINNs [45, 42, 27], where the framework was used to solve nonlinear partial differential equations.

The application of PINNs in traffic state estimation and prediction has received an increasing amount of research attention in recent years. A simple loss-constraint approach was proposed by Dahmen et al. [8] that incorporates the continuity equation into a simple neural network. Case studies carried out in the research show that the method outperforms cases where no physical loss constraint is incorporated. The approach demonstrated the feasibility of physics-informed traffic prediction but did not dig deep into the parameters of the fundamental diagram. In the same year, Huang et al. proposed a deep learning framework that combines the LWR and CTM models with neural networks, which further confirms the potential of PINN in traffic prediction [22]. However, neither approach considers the entire traffic network and thus does not apply to real-world road conditions. Later, Esama et al. proposed a Link-Net framework that is specifically designed for a traffic network [57]. The framework adapts to the characteristics of different links with domain decomposition and combines the individual links to find the traffic state over the entire network.

3 Aims and Objectives

For the task of traffic prediction, existing deep-learning-based methods mainly focus on improving the neural network structure, ignoring the traffic flow theory that guides the underlying physics of traffic. On the other side, research introducing physics-informed approaches for traffic prediction put major effort into designing the physics model without paying attention to dedicated data-driven model architecture. Thus, the two fields remain largely separated. Furthermore, due to the "black-box" nature of neural networks, existing methods for traffic prediction lack interpretability, which makes the models less applicable to real traffic conditions. Finally, none of the existing approaches has considered the dynamic division of a city thus the variation in traffic flow dynamics is largely ignored. Therefore, to fill the above gaps between the existing approaches, the overall aim of our project is:

Aim: Introducing physics-informed approaches into deep learning architectures and push the integration of the two research fields

Following this aim, two research questions will lead our investigation:

1. Can physic-informed techniques be applied as learning biases of neural networks for the task of traffic prediction?

2. What are the merits, demerits, and complementarity of these techniques in terms of accuracy and interpretability?

This project further consists of three objectives, in order to achieve the aim and answer the questions:

Objective 1: Demonstrate the viability of applying physic-informed techniques to deep learning architecture for traffic prediction

Although a theoretical basis for the viability has been provided, we will also practically demonstrate it by introducing traffic flow physics through three categories of biases in neural networks.

Hypothesis: Traffic flow physics can be incorporated into existing deep learning architectures for traffic prediction effectively through learning biases in neural networks.

Objective 2: Compare the prediction accuracy of PINNs with either model-driven or data-driven approaches alone

If the physics-informed techniques are applicable to traffic prediction with deep learning, we will then evaluate its prediction accuracy, and compare the new model with baselines that consider either model-driven or data-driven approaches alone.

Hypothesis: Physics-informed neural networks are more accurate than either model-driven or data-driven approaches alone for traffic prediction.

Objective 3: Experiment and compare the effectiveness of learning bias with and without the dynamic division of road network

We will divide the functional regions of a road network based on different traffic flow dynamics. Then, we plan to compare the effectiveness of this method with those that consider only one model for the entire network.

Hypothesis: With the dynamic division of the road network, different characteristics of the subnetworks can be more accurately represented and aggregated into the loss constraints, thus strengthening the effect of embedded learning bias.

For evaluation purposes, we further propose another hypothesis, which guides our overall aim and objectives:

Hypothesis: Different ways to incorporate traffic flow physics into neural networks have different impacts on the new traffic prediction model, which demonstrates its complementarity, interpretability and additional merits or demerits.

4 Preliminaries

4.1 Computational Graph

In deep learning, a computational graph is a systematic approach for depicting the functional divisions and the required mathematical computations of a model [55]. It serves as a visualisation tool for understanding neural network architectures and has been widely adopted in traffic prediction for displaying the overall model architecture [4, 24, 41, 13]. In this research, we will use the computational graph to portray the pure data-driven spatiotemporal network along with the physics-informed architecture.

4.2 Governing equations and Automatic Differentiation

Governing equations are mathematical equations that describe how a system behaves. In the context of traffic flow theory, the governing equations refer to the physical constraints that the traffic flow dynamics should follow. An important tool for solving the governing equations is automatic differentiation. It computes the intermediate numerical values by performing element-wise partial differentiation and has been widely adopted in previous literature for simple physics-informed systems [26, 27, 42]. We plan to adopt this technique to incorporate the physical constraints into loss functions.

4.3 Analytical Solutions to MFDs

There are three main forms of macroscopic fundamental diagrams for traffic flow modelling: Greenshield's, Daganzo's, and the inverse lambda MFD. Previous studies have provided analytical solutions with various physical constraints [7, 36, 22]. In this study, we inherit some of the notations and introduce an analytical solution for each of them.

4.3.1 Greenshield's Fundamental Diagram

Greenshields' fundamental diagram describes the relationship between any two of the traffic speed v, density ρ and flow q. The equation is shown in 6, where ρ_{max} is the maximum density and v_{free} is the free-flow speed.

$$
\begin{aligned}
q(\rho) &= v_{free}(1 - \frac{\rho}{\rho_{max}})\rho \\
v(p) &= v_{free}(1 - \frac{\rho}{\rho_{max}})
\end{aligned}
\tag{6}
$$

The visualised curves for the relationships, along with a spacing (headway) variable are shown in Figure 1.

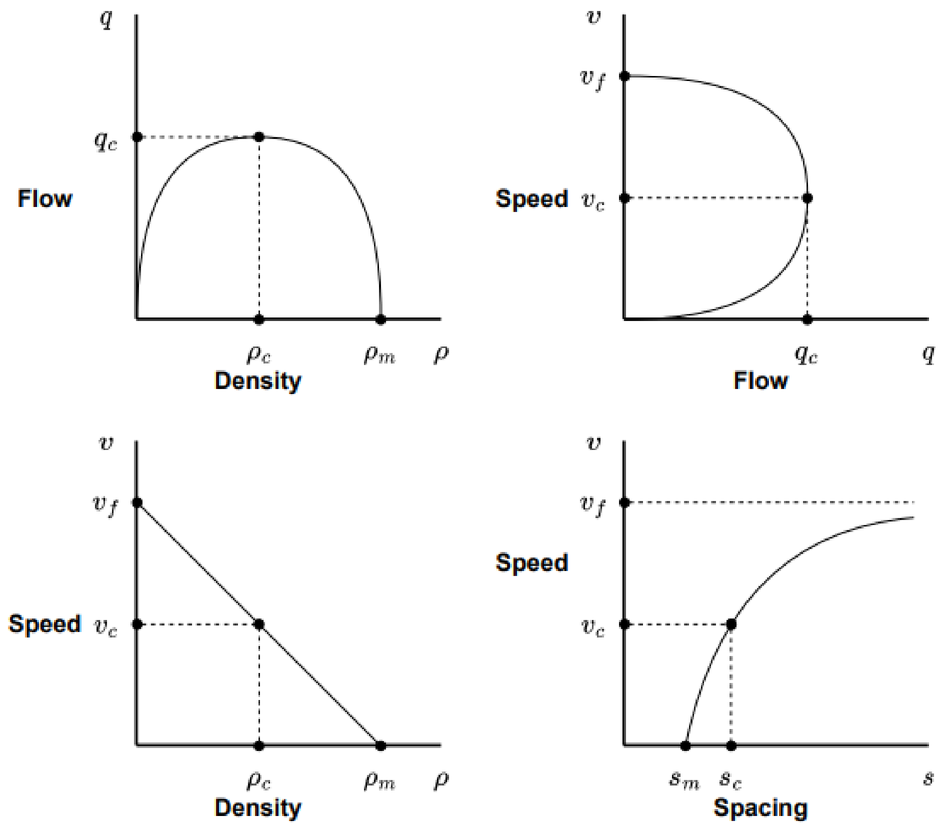


Figure 1: Greenshield's Fundamental Diagrams

4.3.2 Daganzo's Fundamental Diagram

Daganzo's fundamental diagram introduces a critical capacity q_c that indicates the occurrence of a maximum flow. Different from Greenshied's modelling, it uses a triangular shape formed with straight lines to depict the relationship between flow and density. The analytical solution is shown as the piece-wise function in equation (7) and the diagram is shown in Figure 2.

$$q(\rho) = \begin{cases} q_c \frac{\rho}{\rho_c} & \text{if } \rho \leq \rho_c \\ q_c(1 - \frac{\rho - \rho_c}{\rho_m - \rho_c}) & \text{if } \rho > \rho_c, \end{cases} \tag{7}$$

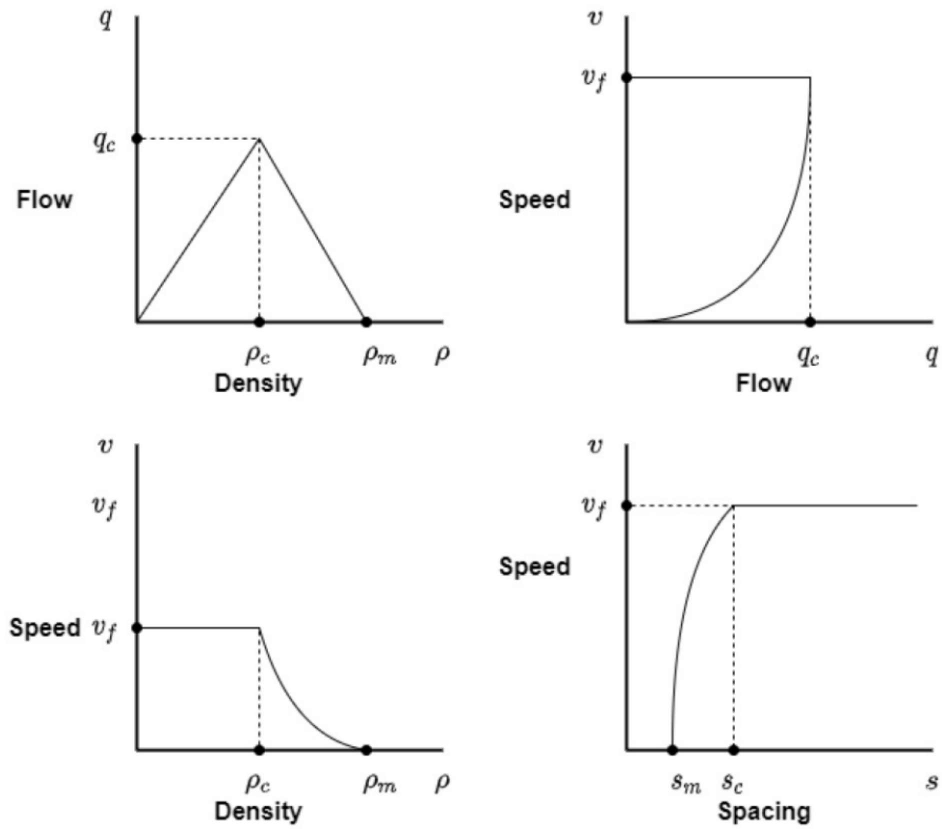


Figure 2: Daganzo's Fundamental Diagrams

4.3.3 Inverse Lambda Fundamental Diagram

Empirical studies suggest that there is often a capacity drop at freeway-ramp merges [49]. A quantified way for this situation leads to the inverse lambda FD. The equation is shown in 8 and the diagram is demonstrated in Figure 3. In the new functional form, the critical capacity is split into q_{c1} and q_{c2} indicating the condition before and after the drop.

$$q(\rho) = \begin{cases} q_{c1}\frac{\rho}{\rho_c} & \text{if } \rho \leq \rho_c \\ q_{c2}(1 - \frac{\rho-\rho_c}{\rho_m-\rho_c}) & \text{if } \rho > \rho_c, \end{cases} \tag{8}$$

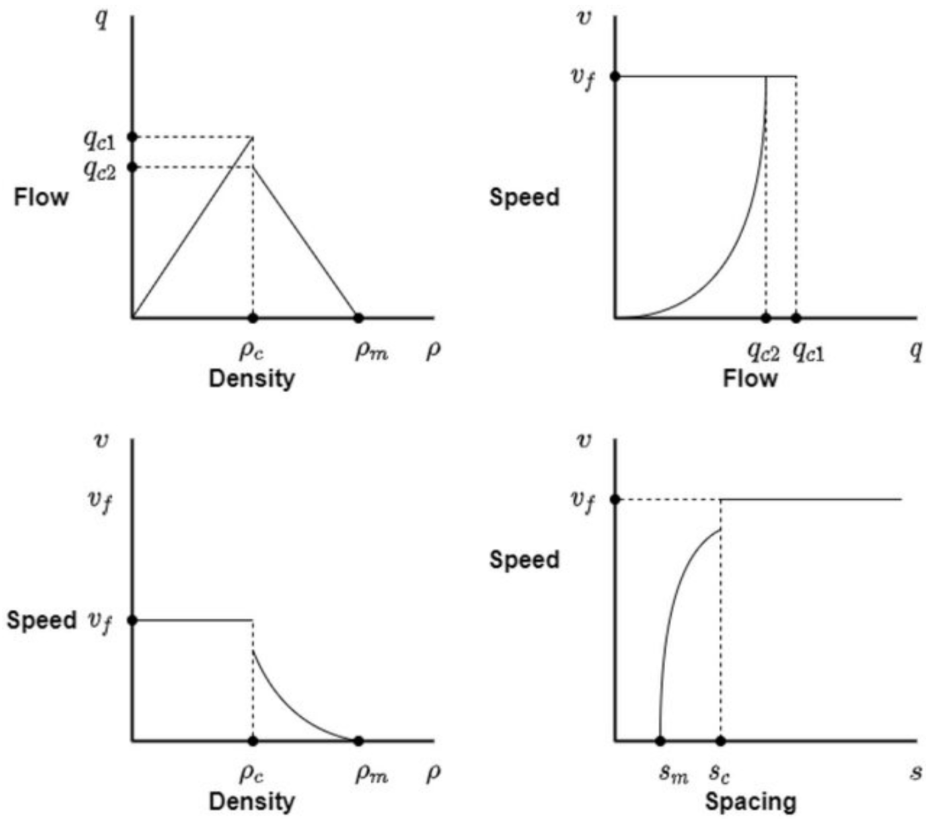


Figure 3: Inverse Lambda Fundamental Diagrams

4.4 Problem Formulation

Definition 1 (Road Network Graph). The road network can be represented as a graph $G_r = (V, E, A)$, where $V = \{v_1, v_2, ..., v_N\}, |V| = N$ denotes a set of nodes, $E_{i,j} = (v_i, v_j), 0 <= i, j <= N, E \subseteq V \times V$ is the set of edges in the graph, and $A \in \mathbb{R}^{N \times N}$ represents the adjacency matrix of graph G_r.

Definition 2 (Adjacency Matrix). The adjacency matrix $A \in \mathbb{R}^{N \times N}$ of a graph $G = (V, E)$ denotes the connections of nodes $||V||$. $A_{i,j} = \epsilon > 0$ if $(v_i, v_j) \in E$ and $A_{i,j} = 0$ otherwise.

Definition 3 (Traffic Feature Matrix). The traffic feature matrix at time t is denoted as $X_t \in \mathbb{R}^{N \times D}$, where N is the number of nodes in the network and D is the dimension of traffic features. For example, if both traffic speed and flow are variables of interest, then

$D = 2$. Based on the above definition, we denote the traffic feature matrix spanning across T units of time as $X = [X_1, X_2, ..., X_T] \in \mathbb{R}^{T\times N\times D}$.

Given historical n_1 time steps of traffic condition along with the road network information, traffic prediction aims to find a functional mapping f that map the given information to the future traffic condition in n_2 time steps. With Section 4.4, the problem can be formulated as follows, with a time span of n before and after time t :

$$f([X_{t-n_1}, ..., X_{t-1}, X_t], G_r) = [X_t, X_{t+1}...X_{t+n_2}]$$

To ensure generality, most literature denotes X as a matrix rather than a vector. However, the variable of concern is mainly monotonic, typically speed or flow only. Thus, $X_t \in \mathbb{R}^{m\times n}$ can be indeed regarded as $X_t \in \mathbb{R}^{m\times 1}$, which is a vector of speed/flow at different nodes in the road network. In our study, we aim to incorporate the fundamental relationship between traffic variables so that they may reinforce each other during model prediction. Thus, the above formulation could also be written as:

$$f([X_{t-n1}, ..., X_{t-1}, X_t], [Z_{t-n1}, ..., Z_{t-1}, Z_t], G) = ([X_t, X_{t+1}...X_{t+n1}], [Z_t, Z_{t+1}...Z_{t+n1}])$$

where both $X_t \in \mathbb{R}^{m\times 1}$ and $Z_t \in \mathbb{R}^{m\times 1}$ are vectors representing a single traffic feature matrix. The two variables could be any two of density, flow or speed. This form is less general than the first formulation. However, it emphasizes the difference between our study and the existing approaches.

5 Research design and Methodologies

This section will describe and provide a rationale for the methods of data collection and analysis, and the materials to be used in research.

5.1 Data Collection

Two datasets will be used for experiments. The first dataset is the traffic signal volume data collected from DATA VIC, which is publicly accessible via [56]. The device used for data collection is a loop detector, which accurately records the number of vehicles passing through the loop wires. We plan to experiment on data collected from January 1 2023 to March 15 2023, and use 2 weeks, 1 month and 2 months of data for training respectively. Meanwhile, we also expect data from the HERE platform and the availability is subject to change in the following months. The second one is the PeMS-BAY dataset. PeMS stands for Caltrans Performance Measu rement System and provides real-time traffic data with various attributes. To ensure consistency with previous studies, we plan to adopt the data used in [31].

5.2 Physics-informed Model Design

To maintain a fair comparison with existing spatiotemporal models and for simplicity, we plan to adopt an encoder-decoder architecture that takes the graph information and time series as input and outputs the future traffic states with desired time steps. The road network will be represented as a graph Laplacian eigenvectors [24]. For the spatiotemporal blocks and the decoder, we will select architectures from the self-attention-based baselines in section 5.3.5. Sample overall frameworks are presented in the deep learning model part of Figure 5, 6 or 7.

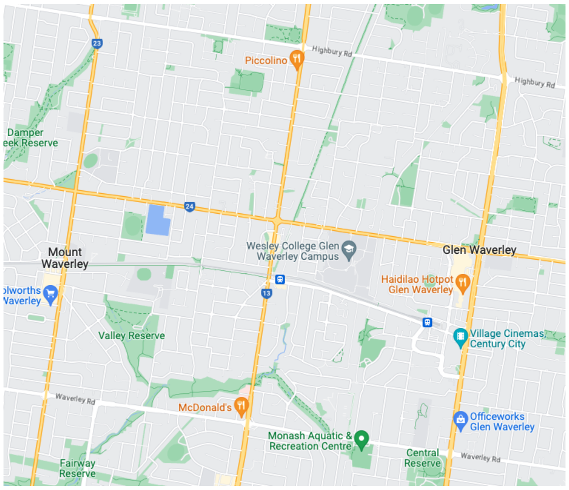
5.2.1 Rationale for Bias Selection

As mentioned in [26], there are three main pathways for embedding physics into an ML model: observational, inductive and learning biases. We will first compare the strengths and weaknesses of the three categories of biases for traffic prediction models, and then provide planned approaches.

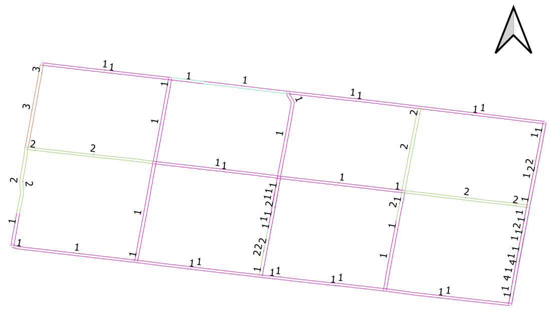
Observational biases require incorporating the underlying physics into the data and learning the physical constraints through the data structure or relationships. This can be considered the easiest way to introduce bias. However, as the relationship between traffic data entries is highly dynamic with respect to space and time, transforming or augmenting the data to a predefined physical form faces a high risk of covering hidden patterns, thus potentially worsening the prediction results. Inductive biases enforce hard constraints on model architectures. Although a customized structure can be effective for tailored application cases, it may also limit the capability of models to capture the variability in traffic data. Learning biases provide only soft penalties to the training process but can serve as guidance towards interpretable results. In the following sections, we will introduce our proposed architecture along with explanations for planned methodologies.

5.2.2 Dynamic Division of Road Network

Due to functional divisions of a city and other external factors such as weather or accidents, different parts of a road network may present different traffic flow dynamics. Previous research mostly considers the network as a whole and ignores the variations of subnetworks. Clustering is an important tool in machine learning to divide the dataset into subsets based on different feature properties, which is suitable for this aim. In this study, we plan to group links (lanes) in a road network into different clusters. For each cluster, the parameters of the traffic flow model will be calibrated through the regional data. By dividing physics into regions, we aim to provide a more accurate representation of the entire network dynamics. We will adopt a combination of clustering methods that considers the correlation coefficient among the traffic flows [47] and a hierarchical clustering [71]. For illustration purposes, we will apply the methods to Melbourne's eastern suburb, with a regional map displayed in Figure 4a and a sample clustering of links in Figure 4b



(a) Melbourne Eastern Suburbs



(b) Cluster Groups Map

5.2.3 Learning Bias

For neural networks, learning biases could be embedded through loss functions. Existing literature mainly demonstrates the viability of three traffic flow models to be incorporated into the loss constraints of deep learning frameworks: the LWR, CTM and ARZ models [22, 57]. Here we derive the analytical solution to LWR so that it could be used as part of loss constraints.

The LWR conservation law could be derived based on the analytical solution to Greenshield's FD (6). The derived equation with is shown in 9, where the traffic speed v is established with respect to location x and time t.

$$\rho_{max}(1 - \frac{2v(x,t)}{v_{free}})\frac{\partial v(x,t)}{\partial x} - \frac{\rho_{max}}{v_{free}}\frac{\partial v(x,t)}{\partial t} = 0 \tag{9}$$

To embed the physical constraints into the loss functions, we could measure the extent of noncompliance of the variables from the physical constraints. For example, the LWR conservation law could be formulated as 10, where N_c stands for the collocation points to be estimated. Assuming that the evaluation metric used for the physics-uninformed part is a mean squared error and that N_0 is the size of observed points, the uninformed loss L_{DL} is shown in 11.

$$L_{PHY} = \frac{1}{N_c}\sum_{i=1}^{N_c} |\rho_{max}(1 - \frac{2\tilde{v}(x_c^i, t_c^i)}{v_{free}})\frac{\partial \tilde{v}(x_c^i, t_c^i)}{\partial x_c^i} - \frac{\rho_{max}}{v_{free}}\frac{\partial \tilde{v}(x_c^i, t_c^i)}{\partial t_c^i}|^2 \tag{10}$$

$$L_{DL} = \frac{1}{N_0}\sum_{i=1}^{N_0} |v(x_c^i, t_c^i) - \tilde{v}(x_c^i, t_c^i)|^2 \tag{11}$$

The total loss could be obtained as the weighted sum of the loss of the deep learning and the physics-informed parts 12.

$$L_{total} = \alpha \times L_{DL} + (1-\alpha) \times L_{PHY} \tag{12}$$

Thus, the overall framework of the learning-bias-based physic-informed neural network is illustrated in Figure 5

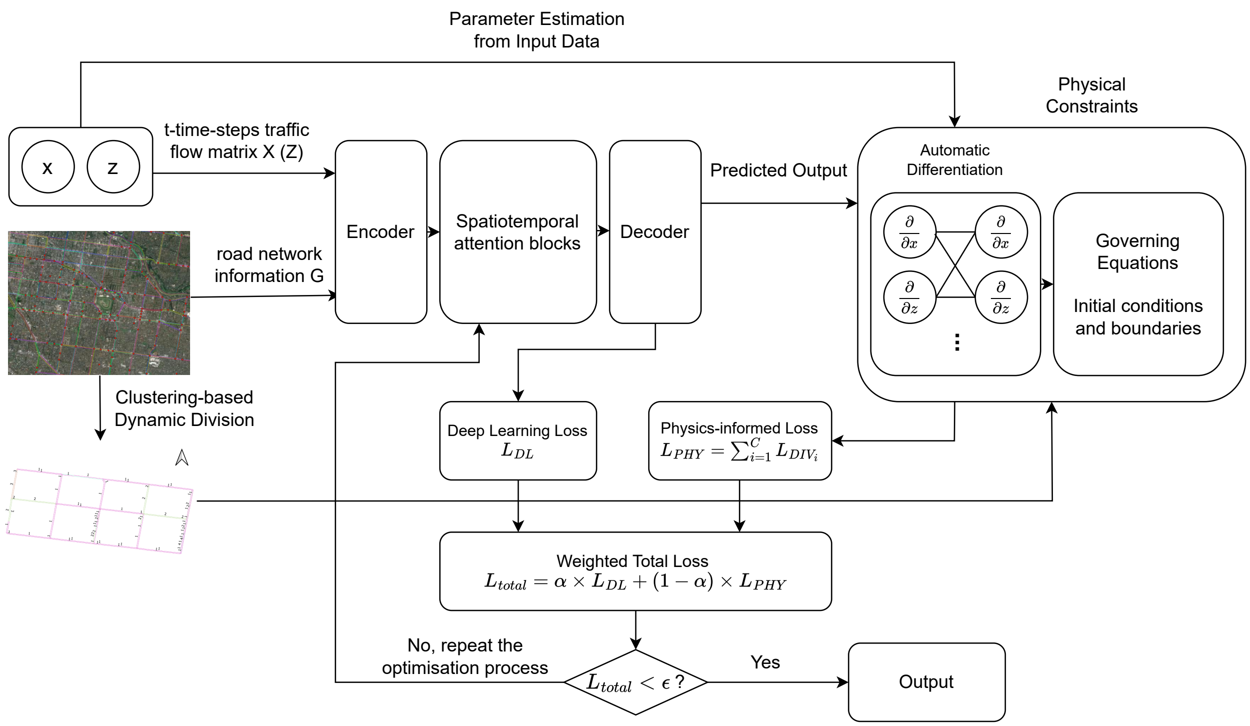


Figure 5: Physics-informed Model Architecture with Learning-Bias Embedded

Name	Number of sensors (nodes)	Time Interval	Time Range
PeMS-BAY	325	5min	Jan 1st 2017 to May 31th 2017
VicRoads Traffic Signal Volume	1291	15min	Jan 1st 2023 to March 15th 2023

Table 1: Basic description of two datasets

5.3 Experimental Design

5.3.1 System and Platforms

The experiments could be divided into 2 stages. In the first stage, trial experiments on a small amount of data as well as model parameter tuning will be carried out. We plan to run the experiments on Google Colab with Python as the implementation language. Pytorch will be used for the deep learning framework construction. In the second stage, we will transfer the fine-tuned model to run on Spartan with the large benchmark datasets. Spartan has 34 nodes, each with 4 80GB Nvidia A100 GPUs, 495000MB RAM and 32 CPU cores. The versions of Python and Pytorch are 3.9.13 and 1.13.1 respectively.

A potential risk associated with this plan lies in the availability of Spartan GPUs. As the demand for the computational resource is high, it may take longer in the queue for our task to be run. Thus, Google Colab can be a suitable alternative. In the worst case, Google Colab might experience a cutoff during the middle of experiments. To avoid this risk, we plan to adopt a combination of Colab and our local laptop for the tasks. The version of the MacOS system is 13.2.1, with 8GB memory and "mps" as the backend accelerated running engine.

5.3.2 Data Processing

The basic information of the two datasets is displayed in Table 1.

To maintain a fair comparison between existing models, we will split the dataset on a ratio of 6:2:2 for train, validation and test set. We plan to use the past 3 steps', 6 steps' and 12 steps' data to predict the next 3, 6 and 12 steps' data respectively. Note that the traffic signal volume dataset is collected on the link level and thus the volume data is directional. Unlike previous research, we will not aggregate the data since averaged or summed volume does not provide meaningful guidance for separate lanes.

5.3.3 Correlations between Input Variables

To investigate the effectiveness of the architecture proposed, we plan to compare the original model with the following variants:

(a) *Variant M/1:* Model with speed v alone as input and output $\tilde{v}$.

(b) *Variant M/2:* Model with flow q alone as input and output $\tilde{q}$.

(c) *Variant M/3:* Both variables are used as input but output is one of them alone.

(d) *Variant M/4:* Both variables are used as input and output.

Through the above experiments, we aim to identify if the variables are reinforcing each other during model prediction with traffic flow models embedded.

5.3.4 Effectiveness of Dynamic Division

To further test the effectiveness of dynamic division of the road network, we will compare the proposed architecture with two other variants: (1) one variant with only one traffic flow model representing the entire system; (2) traffic flow models are calibrated with respect to each link. With these experiments, we aim to demonstrate that the dynamic division can obtain competitive accuracy with respect to both variants without sacrificing much efficiency. The proposed architecture for the two variants is shown in Figure 6 and Figure 7.

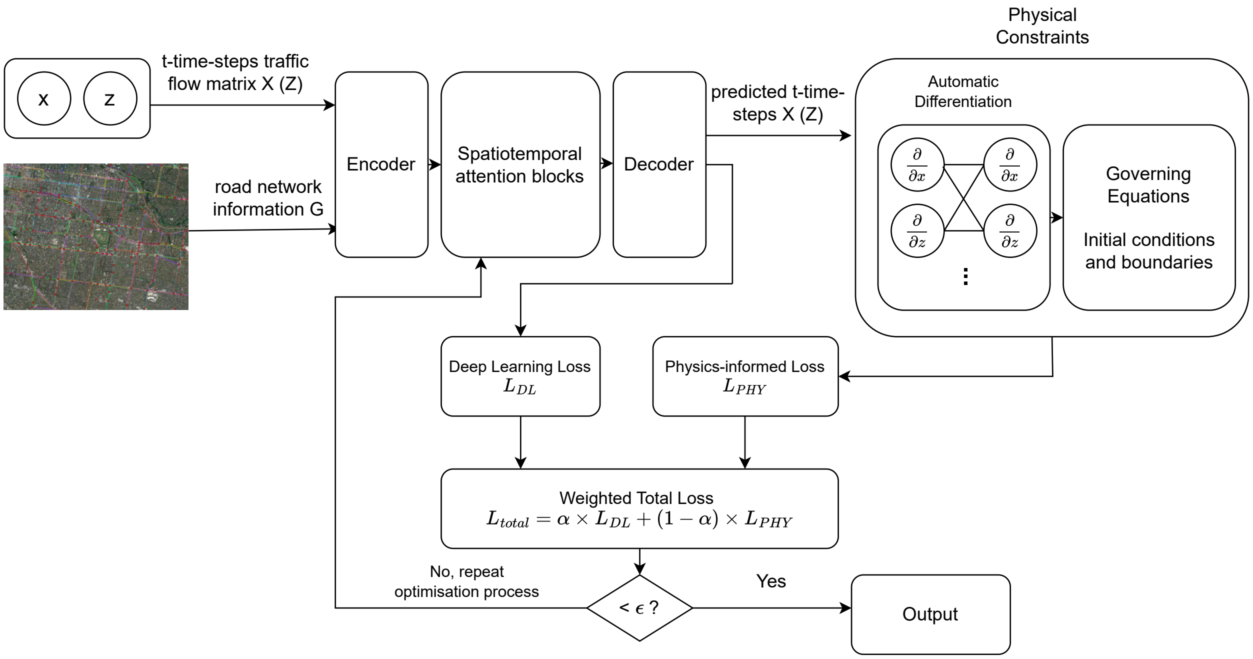


Figure 6: Framework without the dynamic division of the traffic network

5.3.5 Baselines

We plan to compare our proposed model with 11 baselines, falling into the below categories:

1. Time-series based models.

   (a) Seasonal Auto-Regressive Integrated Moving Average (SARIMA) [61]. The model extends the original ARIMA model by including seasonal components, denoted as Seasonal Autoregressive (SAR) and Seasonal Moving Average (SMA), which capture the univariate periodic fluctuations in the data.

   (b) Granger Causality Vector Auto-Regression (GC-VAR) [34]. This is a multivariate time-series model that extends VAR with a Granger causality test, thus identifying causal relationships between variables.

2. GNN-based models.

   (a) DCRNN [31]. The model combines bidirectional-random-walk diffusion convolution with RNN for spatial and temporal dependencies.

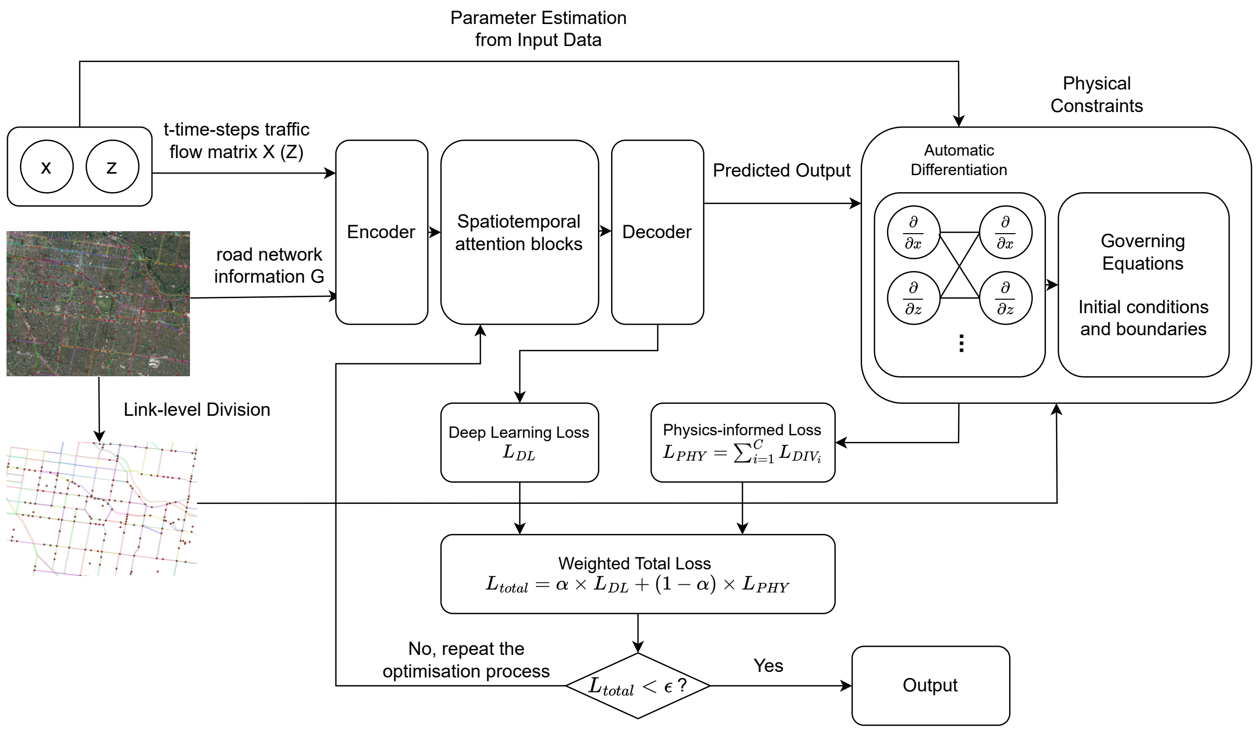


Figure 7: Framework with the link-level division of the traffic network

(b) STGCN [68]. The model integrates graph convolution with gated temporal convolution through spatiotemporal convolution blocks.

(c) Graph WaveNet [63]. An adaptive dependency matrix was developed in the model to better capture hidden spatial dependencies.

(d) MTGNN [62]. The model adapts GNN to multivariate time-series forecasting by introducing graph learning and convolution layers.

3. Self-attention-based models.

(a) Traffic Transformer [4]. The model utilises the transformer architecture with 7 novel positional encoding strategies.

(b) ASTGNN [20]. The model combines self-attention with dynamic graph convolution to capture spatiotemporal correlations in a dynamic manner.

(c) PDFormer [24]. The model introduces a propagation delay-aware mechanism to capture the time delay of traffic streams.

4. PINN-based models.

   (a) Simple neural network with loss constraints [8]. This model incorporates a triangular fundamental diagram into the loss function of a neural network for traffic state estimation.

   (b) Link-Net [57]. This model adapts the PINN-based model to an entire traffic network by solving the LWR traffic flow model.

   In addition, we plan to select one model from each category to embed the traffic flow physics in. Through comparison of these baselines, we aim to demonstrate the applicability of the physics-informed method to state-of-the-art models for traffic prediction.

5.4 Analysis Methods

This section introduces the analysis methods for model efficiency and prediction accuracy. We cover the introduction and justification of metrics we plan to use for the study, and how to interpret the outcomes to answer the research questions.

5.4.1 Evaluation Metrics

To evaluate and compare the performance of different models, and for consistency between existing studies, we plan to adopt three metrics: mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean squared error (RMSE). The objective of traffic prediction is to minimize the difference between the predicted future traffic states and the observed future traffic states, thus pure data-driven losses could be formulated in equations (13) – (15):

$$MAE = \frac{1}{H}\sum_{t=1}^{H}\frac{1}{N}\sum_{i=1}^{N}|X_t^i - \tilde{X}_t^i| \tag{13}$$

$$\text{MAPE} = \frac{1}{H}\sum_{t=1}^{H}\frac{1}{N}\sum_{i=1}^{N}|\frac{X_t^i - \tilde{X}_t^i}{X_t^i}| \tag{14}$$

$$\text{RMSE} = \sqrt{\frac{1}{H}\sum_{t=1}^{H}\frac{1}{N}\sum_{i=1}^{N}|X_t^i - \tilde{X}_t^i|^2} \tag{15}$$

5.4.2 Interpretability

To better understand the learned results from the models, we plan to visualise the change of loss from the PUNN and PINN parts. We will also examine the distribution of predicted traffic states with fundamental diagrams proposed and analyse the trends.

The predicted variables will be scattered with different colours based on the dynamic division of regions. After that, the scatters will be fitted to curves such that they can be compared with the fundamental diagrams. A sample plot is illustrated in Figure 8. Note that the plot does not demonstrate a maximum capacity or where congested conditions are expected. The final plot in the study could vary depending on the dataset characteristics and the algorithms chosen for the dynamic division of regions.

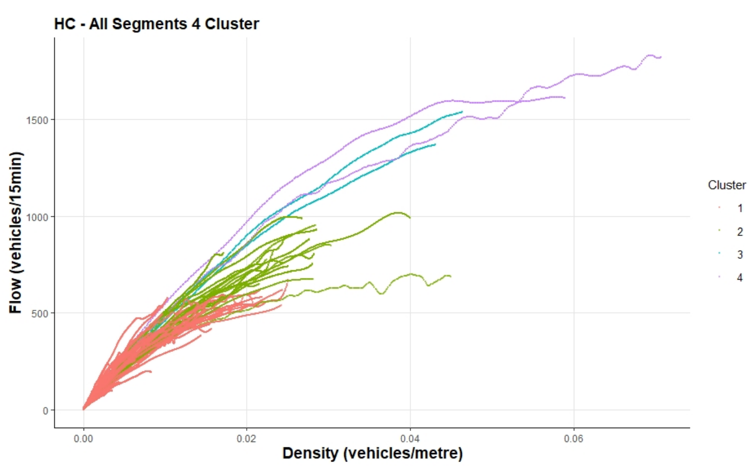


Figure 8: Sample fitted curves for model output in different regions

5.5 Expected Outcomes

The expected outcome is as follows, according to the aim and 3 objectives:

Outcome 1: By introducing traffic flow physics into existing deep learning architectures, we expect to demonstrate the applicability of physic-informed approaches in neural networks and the possibility of integration of the two research fields. Specifically, We expect the model with learning bias incorporated to obtain better prediction performance.

Outcome 2: By evaluating the model variants with respect to input and output variables, we expect to determine if the model with two variables as input and output outperforms other settings. .We would also expect better interpretability of the model if the features are reinforcing each other during the training phase.

Outcome 3: By evaluating the complementarity of the proposed approaches, we expect this research to contribute to the understanding of the strengths and weaknesses of embedding traffic fundamental dynamics into deep learning for traffic prediction.

6 Timeline and Milestones

This section lists the stages of the research project in timeline format and the deadlines for the completion of these stages or tasks.

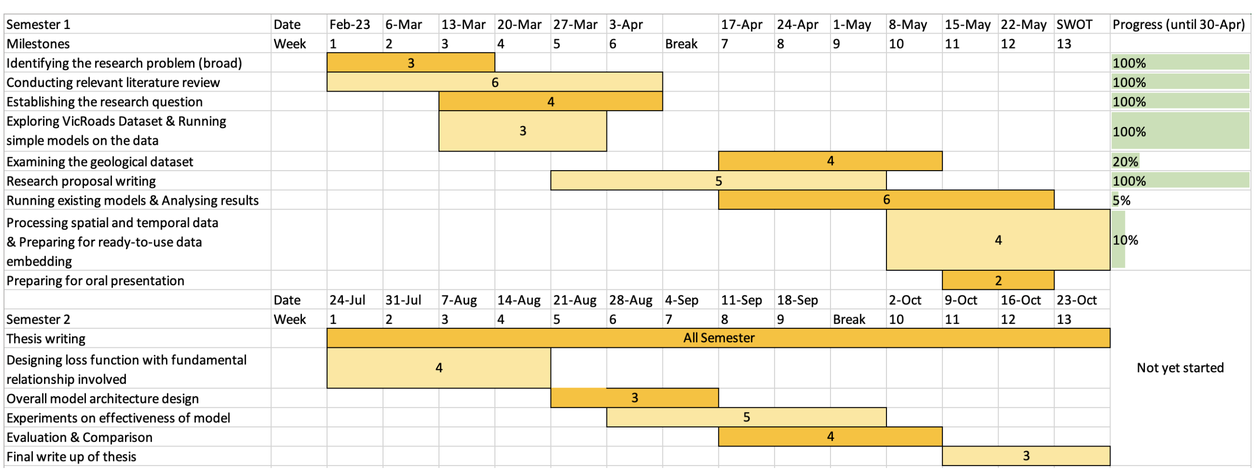


Figure 9: Timeline and milestones

7 Implications of the Study

For the task of traffic prediction, most existing research in the recent decade focus on developing spatiotemporal neural networks. The studies emphasize heavily manually designed model architectures with various assumptions. However, the underlying physics as well as the interpretability of the models are largely ignored. On the other side, existing physics-informed approaches for traffic prediction largely focus on designing functional forms of traffic flow models, and pay much less attention to integrating them into deep learning architectures. Our research aims to fill the gaps between previous studies and existing approaches from the aforementioned perspectives. The significance of this research lies in:

1. We plan to introduce analytical solutions to fundamental diagrams as loss constraints and incorporate them into spatiotemporal neural networks through learning biases. With these forms of physics-informed deep learning, we can utilise both the advantages of model-driven and data-driven approaches. This can potentially contribute to the theory side of the popular data-driven models, while also providing insights for choosing the suitable way to embed physics in deep learning models.

2. We will examine and compare the output distribution with the underlying fundamental diagrams. By evaluating the trends, we expect fewer outliers and that the distribution of prediction will be more reliable. The guided prediction process is also expected to be more interpretable and applicable to real-world scenarios.

3. We incorporate the dynamic division of a road network into the loss function, such that the different traffic flow dynamics in different functional divisions of a city can be more accurately represented. To the best of our knowledge, this is the first attempt for introducing loss functions based on divisional characteristics in a road network, which guide future research to adopt physics-informed approaches for traffic networks.

In conclusion, this study will contribute to the understanding of the theory side of traffic prediction, augmenting existing pure deep-learning-based models with divisional physics-informed approaches, and pushing the integration of the two research fields. We also expect better interpretability and applicability of current spatiotemporal models for the task of traffic prediction.

References

[1] S. -I. Amari. "Learning Patterns and Pattern Sequences by Self-Organizing Nets of Threshold Elements". In: *IEEE Trans. Comput.* 21.11 (Nov. 1972). Place: USA Publisher: IEEE Computer Society, pp. 1197–1206. ISSN: 0018-9340. DOI: 10.1109/T-C.1972.223477. URL: https://doi.org/10.1109/T-C.1972.223477.

[2] Lukas Ambühl et al. "A functional form with a physical meaning for the macroscopic fundamental diagram". In: *Advances in Network Macroscopic Fundamental Diagram (NMFD) Research* 137 (July 2020), pp. 119–132. ISSN: 0191-2615. DOI: 10.1016/j.trb.2018.10.013. URL: https://www.sciencedirect.com/science/article/pii/S0191261517310123.

[3] Sheng-hai An, Byung-Hyug Lee, and Dong-Ryeol Shin. "A Survey of Intelligent Transportation Systems". In: *2011 Third International Conference on Computational Intelligence, Communication Systems and Networks*. 2011, pp. 332–337. DOI: 10.1109/CICSyN.2011.76.

[4] Ling Cai et al. "Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting". en. In: *Transactions in GIS* 24.3 (2020). _eprint: https://onlinelibrary.wiley pp. 736–755. ISSN: 1467-9671. DOI: 10.1111/tgis.12644. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/tgis.12644 (visited on 03/23/2023).

[5] Pinlong Cai et al. "A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting". In: *Transportation Research Part C: Emerging Technologies* 62 (Jan. 2016), pp. 21–34. ISSN: 0968-090X. DOI: 10.1016/j.trc.

2015.11.002. URL: https://www.sciencedirect.com/science/article/pii/S0968090X15003812.

[6] Srinivasa Ravi Chandra and Haitham Al-Deek. "Predictions of Freeway Traffic Speeds and Volumes Using Vector Autoregressive Models". In: *Journal of Intelligent Transportation Systems* 13.2 (May 2009). Publisher: Taylor & Francis, pp. 53–72. ISSN: 1547-2450. DOI: 10.1080/15472450902858368. URL: https://doi.org/10.1080/15472450902858368.

[7] Carlos Daganzo and Nikolas Geroliminis. "An analytical approximation for macroscopic fundamental diagram of urban traffic". In: *Transportation Research Part B: Methodological* 42 (Nov. 2008), pp. 771–781. DOI: 10.1016/j.trb.2008.06.008.

[8] Victoria Dahmen et al. "Traffic State Estimation with Loss Constraint". In: *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. eventplace: Macau, China. IEEE Press, 2022, pp. 1907–1912. DOI: 10.1109/ITSC55140.2022.9921815. URL: https://doi.org/10.1109/ITSC55140.2022.9921815.

[9] Qing Yan Ding et al. "Forecasting Traffic Volume with Space-Time ARIMA Model". en. In: *Advanced Materials Research* 156-157 (2011). Publisher: Trans Tech Publications Ltd, pp. 979–983. ISSN: 1662-8985. DOI: 10.4028/www.scientific.net/AMR.156-157.979. URL: https://www.scientific.net/AMR.156-157.979 (visited on 04/06/2023).

[10] Harris Drucker et al. "Support Vector Regression Machines". In: *Advances in Neural Information Processing Systems*. Ed. by M. C. Mozer, M. Jordan, and T. Petsche. Vol. 9. MIT Press, 1996. URL: https://proceedings.neurips.cc/paper_files/paper/1996/file/d38901788c533e8286cb6400b40b386d-Paper.pdf.

[11] Azadeh Emami, Majid Sarvi, and Saeed Asadi Bagloee. "Using Kalman filter algorithm for short-term traffic flow prediction in a connected vehicle environment". en. In: *Journal of Modern Transportation* 27.3 (Sept. 2019), pp. 222–232. ISSN: 2196-0577.

DOI: 10.1007/s40534-019-0193-2. URL: https://doi.org/10.1007/s40534-019-0193-2 (visited on 04/06/2023).

[12] Mengyuan Fang et al. "FTPG: A fine-grained traffic prediction method with graph attention network using big trace data". In: *IEEE Transactions on Intelligent Transportation Systems* 23.6 (2021). Publisher: IEEE, pp. 5163–5175.

[13] Shen Fang et al. "AutoMSNet: Multi-Source Spatio-Temporal Network via Automatic Neural Architecture Search for Traffic Flow Prediction". In: *IEEE Transactions on Intelligent Transportation Systems* 24.3 (2023), pp. 2827–2841. DOI: 10.1109/TITS.2022.3225553.

[14] Rui Fu, Zuo Zhang, and Li Li. "Using LSTM and GRU neural network methods for traffic flow prediction". In: *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. 2016, pp. 324–328. DOI: 10.1109/YAC.2016.7804912.

[15] Nikolas Geroliminis and Burak Boyacı. "The effect of variability of urban systems characteristics in the network capacity". In: *Transportation Research Part B: Methodological* 46.10 (Dec. 2012), pp. 1607–1623. ISSN: 0191-2615. DOI: 10.1016/j.trb.2012.08.001. URL: https://www.sciencedirect.com/science/article/pii/S0191261512001051.

[16] Nikolas Geroliminis and Carlos F. Daganzo. "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings". In: *Transportation Research Part B: Methodological* 42.9 (Nov. 2008), pp. 759–770. ISSN: 0191-2615. DOI: 10.1016/j.trb.2008.02.002. URL: https://www.sciencedirect.com/science/article/pii/S0191261508000180.

[17] Yi-shan Gong and Yi Zhang. *Research of Short-Term Traffic Volume Prediction Based on Kalman Filtering*. Journal Abbreviation: Proceedings - 2013 6th International Conference on Intelligent Networks and Intelligent Systems, ICINIS 2013 Pages: 102 Pub-

lication Title: Proceedings - 2013 6th International Conference on Intelligent Networks and Intelligent Systems, ICINIS 2013. Nov. 2013. ISBN: 978-1-4799-2809-5. DOI: 10.1109/ICINIS.2013.32.

[18] B. Greenshields et al. "A study of traffic capacity". In: 1935. URL: https://www.semanticscholar.org/paper/A-study-of-traffic-capacity-Greenshields-Bibbins/a3f3c282a7a35b5265ee74757d0edfa163d69bad (visited on 04/24/2023).

[19] *Guide to Traffic Management*. en. Archive Location: Australia and New Zealand Publisher: Austroads. URL: https://austroads.com.au/network-operations/network-management/guide-to-traffic-management (visited on 04/15/2023).

[20] Shengnan Guo et al. "Learning Dynamics and Heterogeneity of Spatial-Temporal Graph Data for Traffic Forecasting". In: *IEEE Transactions on Knowledge and Data Engineering* 34.11 (2022), pp. 5415–5428. DOI: 10.1109/TKDE.2021.3056502.

[21] Hamed Mohammad M., Al-Masaeid Hashem R., and Said Zahi M. Bani. "Short-Term Prediction of Traffic Volume in Urban Arterials". In: *Journal of Transportation Engineering* 121.3 (May 1995). Publisher: American Society of Civil Engineers, pp. 249–254. DOI: 10.1061/(ASCE)0733-947X(1995)121:3(249). URL: https://doi.org/10.1061/(ASCE)0733-947X(1995)121:3(249) (visited on 03/23/2023).

[22] Archie J. Huang and Shaurya Agarwal. "Physics Informed Deep Learning for Traffic State Estimation: Illustrations with LWR and CTM Models". In: *IEEE Open Journal of Intelligent Transportation Systems* 3 (Jan. 2022), pp. 1–1. DOI: 10.1109/OJITS.2022.3182925.

[23] Wenhao Huang et al. "Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning". In: *IEEE Transactions on Intelligent Transportation Systems* 15.5 (Oct. 2014). Conference Name: IEEE Transactions on Intelligent Transportation Systems, pp. 2191–2201. ISSN: 1558-0016. DOI: 10.1109/TITS.2014.2311123.

[24] Jiawei Jiang et al. "PDFormer: Propagation Delay-aware Dynamic Long-range Transformer for Traffic Flow Prediction". In: *AAAI*. AAAI Press, 2023.

[25] Yiannis Kamarianakis and Poulicos Prastacos. "Forecasting Traffic Flow Conditions in an Urban Network: Comparison of Multivariate and Univariate Approaches". In: *Transportation Research Record* 1857.1 (Jan. 2003). Publisher: SAGE Publications Inc, pp. 74–84. ISSN: 0361-1981. DOI: 10.3141/1857-09. URL: https://doi.org/10.3141/1857-09 (visited on 03/29/2023).

[26] George Em Karniadakis et al. "Physics-informed machine learning". In: *Nature Reviews Physics* 3.6 (June 2021), pp. 422–440. ISSN: 2522-5820. DOI: 10.1038/s42254-021-00314-5. URL: https://doi.org/10.1038/s42254-021-00314-5.

[27] Ehsan Kharazmi, Zhongqiang Zhang, and George E.M. Karniadakis. "hp-VPINNs: Variational physics-informed neural networks with domain decomposition". In: *Computer Methods in Applied Mechanics and Engineering* 374 (Feb. 2021), p. 113547. ISSN: 0045-7825. DOI: 10.1016/j.cma.2020.113547. URL: https://www.sciencedirect.com/science/article/pii/S0045782520307325.

[28] Taehooie Kim et al. "A stepwise interpretable machine learning framework using linear regression (LR) and long short-term memory (LSTM): City-wide demand-side prediction of yellow taxi and for-hire vehicle (FHV) service". In: *Transportation Research Part C: Emerging Technologies* 120 (2020). Publisher: Elsevier, p. 102786.

[29] Selvaraj Vasantha Kumar. "Traffic Flow Prediction using Kalman Filtering Technique". In: *TRANSBALTICA 2017: TRANSPORTATION SCIENCE AND TECHNOLOGY: Proceedings of the 10th International Scientific Conference, May 4–5, 2017, Vilnius Gediminas Technical University, Vilnius, Lithuania* 187 (Jan. 2017), pp. 582–587. ISSN: 1877-7058. DOI: 10.1016/j.proeng.2017.04.417. URL: https://www.sciencedirect.com/science/article/pii/S1877705817319471.

[30] Yann LeCun and Yoshua Bengio. "Convolutional Networks for Images, Speech, and Time Series". In: *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, USA: MIT Press, 1998, pp. 255–258. ISBN: 0-262-51102-9.

[31] Yaguang Li et al. "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting". In: *arXiv preprint arXiv:1707.01926* (2017).

[32] Michael James Lighthill and Gerald Beresford Whitham. "On kinematic waves II. A theory of traffic flow on long crowded roads". In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 229.1178 (1955). Publisher: The Royal Society London, pp. 317–345.

[33] Hans van Lint and Marco Schreuder MSc. *Travel time prediction for variable message sign panels: Results and lessons learned from large-scale evaluation study in the Netherlands*. Tech. rep. 2006.

[34] Zheng Lu et al. "Integrating Granger Causality and Vector Auto-Regression for Traffic Prediction of Large-Scale WLANs". In: *KSII Transactions on Internet and Information Systems* 10.1 (Jan. 2016), pp. 136–151. DOI: 10.3837/tiis.2016.01.008.

[35] Yisheng Lv et al. "Traffic Flow Prediction With Big Data: A Deep Learning Approach". In: *IEEE Transactions on Intelligent Transportation Systems* 16.2 (2015), pp. 865–873. DOI: 10.1109/TITS.2014.2345663.

[36] Pierre-Emmanuel Mazaré et al. "Analytical and grid-free solutions to the Lighthill–Whitham–Richard traffic flow model". In: *Transportation Research Part B: Methodological* 45.10 (Dec. 2011), pp. 1727–1748. ISSN: 0191-2615. DOI: 10.1016/j.trb.2011.07.004. URL: https://www.sciencedirect.com/science/article/pii/S0191261511001044.

[37] A R Méndez, W Marques, and R M Velasco. "Multi-class fundamental diagrams from the Prigogine–Herman–Boltzmann equation". In: *Physica Scripta* 94.11 (Aug. 2019). Publisher: IOP Publishing, p. 115008. ISSN: 1402-4896. DOI: 10.1088/1402-4896/ab241e. URL: https://dx.doi.org/10.1088/1402-4896/ab241e.

[38] Lyudmila Mihaylova, René Boel, and Andreas Hegyi. "Freeway traffic estimation within particle filtering framework". In: *Automatica* 43.2 (2007). Publisher: Elsevier, pp. 290–300.

[39] Attila M Nagy and Vilmos Simon. "Survey on traffic prediction in smart cities". In: *Pervasive and Mobile Computing* 50 (2018). Publisher: Elsevier, pp. 148–163.

[40] Attila M. Nagy and Vilmos Simon. "Survey on traffic prediction in smart cities". In: *Pervasive and Mobile Computing* 50 (Oct. 2018), pp. 148–163. ISSN: 1574-1192. DOI: 10.1016/j.pmcj.2018.07.004. URL: https://www.sciencedirect.com/science/article/pii/S1574119217306521.

[41] Yuyan Annie Pan et al. "Incorporating Traffic Flow Model into A Deep Learning Method for Traffic State Estimation: A Hybrid Stepwise Modeling Framework". In: *Journal of Advanced Transportation* 2022 (Nov. 2022). Ed. by Tomio Miwa. Publisher: Hindawi, p. 5926663. ISSN: 0197-6729. DOI: 10.1155/2022/5926663. URL: https://doi.org/10.1155/2022/5926663.

[42] Guofei Pang, Lu Lu, and George Em Karniadakis. "fPINNs: Fractional Physics-Informed Neural Networks". In: *SIAM Journal on Scientific Computing* 41.4 (Jan. 2019). Publisher: Society for Industrial and Applied Mathematics, A2603–A2626. ISSN: 1064-8275. DOI: 10.1137/18M1229845. URL: https://doi.org/10.1137/18M1229845 (visited on 04/16/2023).

[43] A. Pascale and M. Nicoli. "Adaptive Bayesian network for traffic flow prediction". In: *2011 IEEE Statistical Signal Processing Workshop (SSP)*. 2011, pp. 177–180. DOI: 10.1109/SSP.2011.5967651.

[44] Jianzhong Qi et al. "A Graph and Attentive Multi-Path Convolutional Network for Traffic Prediction". In: *IEEE Transactions on Knowledge and Data Engineering* (2022). Conference Name: IEEE Transactions on Knowledge and Data Engineering, pp. 1–1. ISSN: 1558-2191. DOI: 10.1109/TKDE.2022.3179646.

[45] M. Raissi, P. Perdikaris, and G.E. Karniadakis. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations". In: *Journal of Computational Physics* 378 (Feb. 2019), pp. 686–707. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2018.10.045. URL: https://www.sciencedirect.com/science/article/pii/S0021999118307125.

[46] Maziar Raissi. "Deep hidden physics models: Deep learning of nonlinear partial differential equations". In: *The Journal of Machine Learning Research* 19.1 (2018). Publisher: JMLR. org, pp. 932–955.

[47] Unsok Ryu et al. "A clustering based traffic flow prediction method with dynamic spatiotemporal correlation analysis". In: *Transportation* 49.3 (June 2022), pp. 951–988. ISSN: 1572-9435. DOI: 10.1007/s11116-021-10200-9. URL: https://doi.org/10.1007/s11116-021-10200-9.

[48] V. I. Shvetsov. "Mathematical Modeling of Traffic Flows". en. In: *Automation and Remote Control* 64.11 (Nov. 2003), pp. 1651–1689. ISSN: 1608-3032. DOI: 10.1023/A:1027348026919. URL: https://doi.org/10.1023/A:1027348026919 (visited on 04/05/2023).

[49] Anupam Srivastava and Nikolas Geroliminis. "Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model". In: *Transportation Research Part C: Emerging Technologies* 30 (May 2013), pp. 161–177. ISSN: 0968-090X. DOI: 10.1016/j.trc.2013.02.006. URL: https://www.sciencedirect.com/science/article/pii/S0968090X13000363.

[50] Shiliang Sun, Changshui Zhang, and Guoqiang Yu. "A bayesian network approach to traffic flow forecasting". In: *IEEE Transactions on Intelligent Transportation Systems* 7.1 (2006), pp. 124–132. DOI: 10.1109/TITS.2006.869623.

[51] Shiliang Sun, Changshui Zhang, and Yi Zhang. "Traffic Flow Forecasting Using a Spatio-temporal Bayesian Network Predictor". In: *Artificial Neural Networks: Formal*

Models and Their Applications – ICANN 2005. Ed. by Włodzisław Duch et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 273–278. ISBN: 978-3-540-28756-8.

[52] David Alexander Tedjopurnomo et al. "A Survey on Modern Deep Neural Network for Traffic Prediction: Trends, Methods and Challenges". In: *IEEE Transactions on Knowledge and Data Engineering* 34.4 (Apr. 2022). Conference Name: IEEE Transactions on Knowledge and Data Engineering, pp. 1544–1561. ISSN: 1558-2191. DOI: 10.1109/TKDE.2020.3001195.

[53] John Michael Thomson. *Speeds and Flows of Traffic in Central London...* Printerhall, 1967.

[54] Yongxue Tian and Li Pan. "Predicting Short-Term Traffic Flow by Long Short-Term Memory Recurrent Neural Network". In: *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*. 2015, pp. 153–158. DOI: 10.1109/SmartCity.2015.63.

[55] G. Tinhofer et al., eds. *Computational Graph Theory*. en. Vol. 7. Computing Supplementum. Vienna: Springer, 1990. ISBN: 978-3-211-82177-0 978-3-7091-9076-0. DOI: 10.1007/978-3-7091-9076-0. URL: https://link.springer.com/10.1007/978-3-7091-9076-0 (visited on 04/24/2023).

[56] *Traffic Management Training*. en. Archive Location: Australia and New Zealand Publisher: Austroads. URL: https://austroads.com.au/network-operations/traffic-management/traffic-management-training (visited on 03/22/2023).

[57] Muhammad Usama et al. "Physics-Informed Neural Networks (PINNs)-Based Traffic State Estimation: An Application to Traffic Network". en. In: *Algorithms* 15.12 (Dec. 2022). Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, p. 447. ISSN: 1999-4893. DOI: 10.3390/a15120447. URL: https://www.mdpi.com/1999-4893/15/12/447 (visited on 04/19/2023).

[58] Eleni Vlahogianni, Matthew Karlaftis, and John Golias. "Short-term traffic forecasting: Where we are and where we're going". In: *Transportation Research Part C: Emerging Technologies* 43 (June 2014). DOI: 10.1016/j.trc.2014.01.005.

[59] Jingyuan Wang et al. "LibCity: An Open Library for Traffic Prediction". In: *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*. SIGSPATIAL '21. event-place: Beijing, China. New York, NY, USA: Association for Computing Machinery, 2021, pp. 145–148. ISBN: 978-1-4503-8664-7. DOI: 10.1145/3474717.3483923. URL: https://doi.org/10.1145/3474717.3483923.

[60] JG Wardrop. "Journey speed and flow in central urban areas". In: *Traffic Engineering & Control* 8.8 (1968).

[61] Billy Williams and Lester Hoel. "Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results". In: *Journal of Transportation Engineering* 129 (Nov. 2003), pp. 664–672. DOI: 10.1061/(ASCE)0733-947X(2003)129:6(664).

[62] Zonghan Wu et al. "Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks". In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020.

[63] Zonghan Wu et al. "Graph wavenet for deep spatial-temporal graph modeling". In: *arXiv preprint arXiv:1906.00121* (2019).

[64] Haoyang Yan, Xiaolei Ma, and Ziyuan Pu. "Learning Dynamic and Hierarchical Traffic Spatiotemporal Features With Transformer". In: *IEEE Transactions on Intelligent Transportation Systems* 23.11 (2022), pp. 22386–22399. DOI: 10.1109/TITS.2021.3102983.

[65] Liu Yang, Dongkun Zhang, and George Em Karniadakis. "Physics-informed generative adversarial networks for stochastic differential equations". In: *arXiv preprint arXiv:1811.02033* (2018).

[66] Mofeng Yang et al. “Application of the ARIMAX Model on Forecasting Freeway Traffic Flow”. In: July 2017.

[67] Huaxiu Yao et al. “Deep multi-view spatial-temporal network for taxi demand prediction”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 32. Issue: 1. 2018.

[68] Bing Yu, Haoteng Yin, and Zhanxing Zhu. “Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting”. In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. IJCAI’18. event-place: Stockholm, Sweden. AAAI Press, 2018, pp. 3634–3640. ISBN: 978-0-9992411-2-7.

[69] Haitao Yuan and Guoliang Li. “A Survey of Traffic Prediction: from Spatio-Temporal Data to Intelligent Transportation”. In: *Data Science and Engineering* 6.1 (Mar. 2021), pp. 63–85. ISSN: 2364-1541. DOI: `10.1007/s41019-020-00151-z`. URL: `https://doi.org/10.1007/s41019-020-00151-z`.

[70] Junbo Zhang, Yu Zheng, and Dekang Qi. “Deep spatio-temporal residual networks for citywide crowd flows prediction”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 31. Issue: 1. 2017.

[71] L. Zhang et al. *A study of the traffic network in Melbourne eastern suburbs*. Technical Report. Dec. 2021.

[72] Chuanpan Zheng et al. “GMAN: A Graph Multi-Attention Network for Traffic Prediction”. en. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.01 (Apr. 2020). Number: 01, pp. 1234–1241. ISSN: 2374-3468. DOI: `10.1609/aaai.v34i01.5477`. URL: `https://ojs.aaai.org/index.php/AAAI/article/view/5477` (visited on 03/26/2023).