Urban traffic flow predictions using state-of-the-art prediction models for real-time traffic simulations

Von der Fakultät für Ingenieurwissenschaften,

Abteilung Maschinenbau und Verfahrenstechnik der

Universität Duisburg-Essen

zur Erlangung des akademischen Grades

einer

Doktorin der Ingenieurwissenschaften

Dr.-Ing.

genehmigte Dissertation

von

JENITTA PRAGALATHAN
aus
Vellakovil, Indien

Gutachter:
Prof. Dr.-Ing. Dr. h.c. Dieter Schramm
Prof. Dr. Gitakrishnan Ramadurai

Tag der Einreichung: 12.12.2023
Tag der mündlichen Prüfung: 28.05.2024

I can do all things through Christ who strengthens me.

Philippians 4:13

ACKNOWLEDGEMENT

I hereby express my sincere thanks and gratitude to my esteemed guide Prof. Dr.-Ing. Dieter Schramm for rendering valuable guidance and constant support throughout my research work. I take immense pleasure in acknowledging him for giving me an opportunity to carry on my PhD research with all facilities. I always admired his positivity and unflappable nature which also influenced me and gave me hope in difficult situations.

I express my sincere thanks to Prof. Dr. Gitakrishnan Ramadurai for accepting to be my second reviewer and for providing valuable comments for my thesis.

I am extremely thankful to Prof. Dr. Jörg Schönharting for his support and encouragement extended for my research work from the beginning. I am very grateful for his continuous help and guidance and for taking time to provide his expertise.

Special thanks to TRC Transportation Research & Consultancy GmbH as well as Wirtschaftsbetriebe Duisburg (WBD) for providing me with facilities for data collection and software usage. I would like to express sincere recognition to my colleagues at the Chair of Mechatronics who helped me directly or indirectly in the completion of this dissertation.

I am indebted to my husband, Sampaul, for supporting me and encouraging me throughout this PhD journey. I am incredibly thankful to him for managing everything along with our kids and giving me space to pursue my dreams and achieve my goals.

I am always thankful to my beloved parents, my sister, my in-laws, my other family members and friends for having hope in me and encouraging me to complete my thesis in a successful manner.

Jenitta Pragalathan

ABSTRACT

As part of the Digital Single Market Strategy, the EU Commission focuses mainly on Intelligent Transportation Systems (ITS) solutions, for more efficient traffic management strategies (European Commission: Mobility and Transport, 2023). Within the next few years, traffic flow predictions will be used for most of the smart solutions in transportation engineering. Traffic Flow Prediction Models (TFPM) are recognized as a part of the foundations of ITS and many other real-time applications to provide quick and cost-effective solutions for reducing congestion and travel time. TFPM are the models that can estimate traffic flow at a particular location over a given period. These estimations can be carried out based on various techniques (like statistical and deep learning) which have been researched for decades to achieve more accuracy and faster computation. To achieve more flexibility and accuracy, more focus was given towards the deep learning models in the last few years. Even though the accuracy was achieved with such neural network algorithms, the complexity of the model and the processing time were also increasing. Two recent time series prediction models are Fb-Prophet and NeuralProphet. Especially NeuralProphet model which is the successor of the Fb-Prophet model was developed to bridge the traditional and deep learning models.

These models have already started dominating in various other fields for solving current prediction problems. Therefore, being new to traffic flow predictions, this thesis focuses on analysing the model's performances by comparing with a well-known statistical prediction model namely - Seasonal Auto-regressive Integrated Moving Average model (SARIMA) and their applicability for an urban traffic scenario is studied. The property of decomposability of Fb-Prophet and NeuralProphet model enhances them with more flexibility to include the effect of holidays and events. The analyst in the loop concept makes the process of prediction faster and more automated. Both the models can be scaled up to a bigger network because of its automated functioning.

Initially, the performances of these time series models are evaluated in terms of accuracy and computational time. Additionally, the prediction of urban traffic with inclusion of effect of external factors like weather and holidays are carried out by using NeuralProphet model. Thus, the predicted results will be given to a flow generator that can give SUMO readable traffic demand file. Finally, this thesis also outlines the development of a real-time simulation system that can be used for real-time applications, driving simulators, atypical scenario simulations and various other applications. A sample simulation for platoon formation was also run with the developed simulation system. The developed prediction system can be in future used for many research purposes. The predicted results from the model with effect of rainfall can also be simulated in a dynamic driving simulator and used for analysis of vehicle dynamics at various road surface conditions (wet/icy).

KURZFASSUNG

Im Rahmen der Strategie für einen digitalen Binnenmarkt fokussiert die EU-Kommission vor allem auf ITS-Lösungen (Intelligent Transportation Systems) für effizientere Verkehrsmanagementstrategien (European Commission: Mobility and Transport, 2023). In den kommenden Jahren werden Verkehrsflussprognosen für die meisten intelligenten Lösungen in der Verkehrstechnik verwendet werden. Verkehrsflussprognosemodelle (Traffic Flow Prediction Models, TFPM) sind anerkanntermaßen eine Basis für ITS und viele andere Echtzeit-Anwendungen, um schnelle und kosteneffiziente Lösungen zur Verringerung von Verkehrsstaus und Fahrzeiten anzubieten. TFPM sind Modelle, mit denen der Verkehrsfluss an einem bestimmten Ort in einem bestimmten Zeitraum geschätzt werden kann. Diese Prognosen können auf der Basis verschiedener Methoden (z. B. statistisches und Deep Learning) durchgeführt werden, die seit mehreren Jahren erforscht werden, um eine höhere Genauigkeit und schnellere Berechnungen zu ermöglichen. Wegen der geringen Flexibilität und Genauigkeit wurde in den letzten Jahren der Fokus verstärkt auf Deep-Learning-Modelle gelegt. Auch wenn die Genauigkeit mit solchen Algorithmen für neuronale Netze erreicht wurde, wurden die Komplexität des Modells und die Verarbeitungszeit immer höher. Zwei neuere Zeitreihenprognosemodelle sind Fb-Prophet und NeuralProphet. Das NeuralProphet Modell, der Nachfolger des Fb-Prophet Modells, wurde entwickelt, um eine Verknüpfung zwischen den traditionellen und den Deep-Learning-Modellen herzustellen.

Diese Modelle haben sich bereits in verschiedenen anderen Bereichen zur Lösung aktueller Prognoseprobleme durchgesetzt. Da sie für die Verkehrsprognose neu sind, liegt der Schwerpunkt dieser Arbeit auf der Analyse der Performance des Modells, indem es mit einem bekannten statistischen Prognosemodell verglichen wird, nämlich dem Seasonal Auto-regressive Integrated Moving Average Model (SARIMA), und ihre Verwendbarkeit für ein städtisches Verkehrsszenario wird untersucht. Die Eigenschaft der Zerlegbarkeit des Fb-Prophet- und des Neural-Prophet-Modells verleiht ihnen mehr

Flexibilität, um die Auswirkungen von Feiertagen und Ereignissen zu berücksichtigen. Das Analyst-in-the-Loop Konzept macht den Prozess der Prognose schneller und automatisiert. Beide Modelle können aufgrund ihrer automatisierten Funktionsweise auf ein größeres Netzwerk skaliert werden.

Zuerst werden die Leistung der Zeitreihenmodelle in Bezug auf Genauigkeit und Rechenzeit bewertet. Zusätzlich wird die Prognose des städtischen Verkehrs unter Berücksichtigung der Auswirkungen externer Faktoren wie Wetter und Feiertage mit Hilfe des NeuralProphet Modells durchgeführt. Die vorhergesagten Ergebnisse werden dann an einen Flow-Generator weitergegeben, der eine Input-Datei für SUMO erstellen kann. Abschließend wird in dieser Arbeit auch die Entwicklung eines Echtzeit-Simulationssystems skizziert, das für Echtzeitanwendungen, Fahrsimulatoren, Simulationen atypischer Szenarien und verschiedene andere Anwendungen verwendet werden kann. Das entwickelte Simulationssystem hat außerdem eine Beispielsimulation für die Formation von Platoon durchgeführt. Das entwickelte Prognosesystem kann in Zukunft für viele Forschungszwecke eingesetzt werden. Die aus dem Modell prognostizierten Ergebnisse mit der Auswirkung von Regenfällen können auch in einem dynamischen Fahrsimulator simuliert und für die Analyse der Fahrzeugdynamik bei verschiedenen Fahrbahnzuständen (nass/eisig) verwendet werden.

TABLE OF CONTENTS

Page No.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AE	:	Absolute Error
AI	:	Artificial Intelligence
APE	:	Absolute Percentage Error
AIMSUN	:	Advanced Interactive Micro-Simulation for Urban and non-urban Networks
ANN	:	Artificial Neural Network
ARIMA	:	Auto-Regressive Integrated Moving Average
ATIS	:	Advance Traveller Information System
ATMS	:	Advanced Traffic Management System
CNN	:	Convolutional Neural Network
CORSIM	:	CORridor SIMulation
DLR	:	Deutsches Zentrum für Luft- und Raumfahrt
EC	:	European Commission
Fb-Prophet	:	Facebook Prophet
ILD	:	Inductive Loop Detector
ITS	:	Intelligent Transportation System
LSTM	:	Long Short-Term Memory
MAE	:	Mean Absolute Error
MAPE	:	Mean Absolute Percentage Error
MATSIM	:	Multi Agent Transport SIMulation
ML	:	Machine Learning

NETSIM	:	NETwork SIMulation
PARAMICS	:	PARAllel Microscopic Simulation
RMSE	:	Root Mean Squared Error
RNN	:	Recurrent Neural Network
SARIMA	:	Seasonal Auto-Regressive Integrated Moving Average
SUMO	:	Simulation of Urban Mobility
TFPM	:	Traffic Flow Prediction Model
TraCI	:	Traffic Control Interface
TRANSIMS	:	TRansportation Analysis and SIMulation System
VC	:	Vehicle Counter
VISSIM	:	Verkehr In Stadt SIMulation
WBD	:	Wirtschaftsbetriebe Duisburg

CHAPTER 1

INTRODUCTION

Urban road traffic faces numerous problems in day-to-day life, with increasing population and rapid urbanisation. This chapter explains how the urban traffic problems are addressed with recent advancements in technologies and declares the importance of the traffic flow predictions by answering the questions "What is the need for traffic flow predictions?" and "Why is it important now?". It also describes the research problem which will be focused in this thesis and narrates the outline of the chapters.

1.1 MOTIVATION

Congestion is one of the key issues on urban roads which can contribute to other problems like increase in travel delays and fuel consumption. The growing demand for transportation creates greater number of vehicles on the urban roads with limited capacity which eventually leads to congestion and traffic delays. In the past decades, strategies suggested by ITS provided smart solutions focusing on efficient usage of road networks and adaptive traffic control measures (ITS JPO Strategic Plan 2020-2025, 2020; Intelligent Transportation Systems: Joint Program Office, 2023). Hence travel delays and vehicle idling was addressed, leading to reduction in the waiting time and also fuel consumption. With the support of information and communication technologies and availability of huge traffic database, the role of ITS in urban traffic management is increasing.

The two strategies in Intelligent Transportation Systems (ITS) - Advanced Traffic Management Systems (ATMS) and Advanced Traveller Information Systems (ATIS) play a vital role in urban traffic management and to reduce congestion and delays. The implementation of smart ITS solutions is possible with the help of Centralised Traffic Management Centres for urban areas to inform both the network and the traveller about current and future traffic scenarios. Such a well-informed road network is achieved with the help of automated traffic data collection and traffic flow prediction models. Hence Traffic Flow Predictions serves as the foundation of all such ITS strategies to know the real-time or near future or future traffic.

On the other side, the development in automobile industry with self-driving cars, connected vehicles, platooning systems, increases the need to further analyse the influence of such advanced vehicles over the traditional traffic stream, urban road networks and traffic control measures. Such analysis should be done in virtual scenarios with the help of traffic simulations. For advanced research purposes, real-time traffic simulations and digital twinning is possible again with the help of traffic prediction models. Driving simulators can also be used to simulate the traffic scenarios and analyse the vehicle dynamics with the introduction of future vehicles. For all such simulations, the major parameter given as input is traffic demand. Such traffic demand or volume can be predicted and used in the simulations for real-time traffic analysis.

1.2 PROBLEM DEFINITION

For the past few years, the availability of automated traffic data collection and exchange of traffic information particularly created an enormous interest in traffic flow prediction modelling. The efficacy of traffic flow forecasting relies on appropriate data pre-processing and the choice of suitable models. Furthermore, the accuracy and size of traffic data used also play a vital role in the performance of a model. The problem with traffic information forecasting is that the precision of the model alters according to the traffic data used for training and testing the model. However, the everyday pattern of the real-world traffic data depends on a number of external factors, namely temporal dependency i.e., peak/off-peak hours, working/nonworking days, and environmental factors like weather data. Hence, it is important to estimate the traffic flow incorporating external information like weather and holidays and also the temporal dependency of the training and testing datasets. The periodic nature of traffic flow can be analysed by inspecting the trend or pattern of number of vehicles passing in a particular location over a period (short term: every day, medium term: every week, long term: every season).

The recent state-of-the-art time series prediction models are Fb-Prophet and NeuralProphet models. The present work aims to analyse the impact of such temporal dependency of the dataset on the performance of the traffic flow forecasting model. The next decade is likely to witness a considerable use of Traffic Flow Prediction Models (TFPM). These models are acknowledged as part of the foundations of Intelligent Transportation Systems (ITS) to make smart decisions, provide quick and cost-effective solutions and reduce congestion and travel time. Most studies on TFPM have focused on developing models even with hybrid combinations that lead to more complexity and more processing time. There is still some considerable work needed to account for the influence of the periodicity in traffic flow forecasting.

Numerous external factors create this temporal dependency that influences everyday pattern of real-world traffic data. Therefore, it is very much important while forecasting traffic to consider these dependencies and optimise the model performance. Fb-Prophet and NeuralProphet are the two recent time series prediction models which require a study to check the performance of both models, especially for urban traffic forecasting. There were very few work done so far using Fb-Prophet and Neural-Prophet models in traffic flow predictions. Among the TFPM, a well-known statistical model Seasonal Auto-Regressive Integrated Moving-Average (SARIMA) model has proven its accuracy in many research works. Keeping SARIMA as the baseline model, the performance of the Fb-Prophet and NeuralProphet models will be evaluated for urban traffic scenarios. This comparative analysis was carried out in (Pragalathan & Schramm, Apr 2023).

The variations due to trend, seasonality and holidays are included within the model definition in Fb-Prophet (Taylor & Letham, 2017) and NeuralProphet models (Triebe, et al., 2021). Hence the model (Fb-Prophet or NeuralProphet) itself can decide the parameters according to the data provided which makes the process of modelling faster and automatic with lesser human inferences to check the performance of the model and to do some alterations if needed in the small set of intuitive parameters.

Comparative studies of these models were done using data collected for seven days from a National Highway in India (Chikkakrishna, Hardik, Deepika, & Sparsha, 2019; ChikkaKrishna, Rachakonda, & Tallam, 2022). These literature works evaluated the models' performances on a highway and very lesser range (min: 360 veh/hr - max: 608 veh/hr) of traffic data was used for training. Therefore, there is a necessity to investigate the performance of Fb-Prophet and Neural-Prophet models in a busy urban intersection with the traffic flow of a higher range (min: 0 veh/hr - max: 600 veh/hr). The ability of the model to capture strong repetitions of traffic patterns during peak/off-peak hour and weekday/weekend will also be evaluated.

Finding a best possible model with reduced computational cost and increased accuracy will be mainly focused in this research work. Traffic forecasting with inclusion of weather and holiday data will also be carried out and the difference in their performance will be evaluated (Pragalathan & Schramm, 2024). Real-time traffic simulation is another trending research topic in traffic engineering which is developed based on traffic forecasting models. In this research, the development of real-time simulations is also discussed along with a sample application. The various applications of the real-time predictions and simulations will also be briefly explained (Pragalatahan & Schramm, May 2023).

1.3 OBJECTIVES OF THE STUDY

According to classification given by Research Society for Roads and Transport in Germany (Forschungsgesellschaft für Straßen- und Verkehrswesen: FGSV) (Hoyer, et al., April 2012), traffic flow predictions with collected historical data can be carried out based on three methods. They are time series-based, traffic demand-based and traffic pattern recognition-based predictions. Auto Regressive Integrated Moving Average (ARIMA) model was explained as an example of time series prediction. One of the major drawbacks of this classical and precise ARIMA model is that it could not include calendar information. But the newly developed Fb-Prophet and NeuralProphet time series prediction models allow the inclusion of holidays, events and other calendar information. Hence the purpose of this thesis are:

- To analyse the state-of-the-art time series prediction models (Fb-Prophet and NeuralProphet) and study the performance of the models for predicting urban traffic conditions.
- To provide a traffic prediction system that can handle the exogenous factors (holidays/weather) of urban traffic scenarios.
- To simulate real-time traffic scenarios with the support of the final prediction system connecting with SUMO software.
- To illustrate the overall methodology, various stages in the development framework for a real-time urban traffic prediction and simulation.

1.4 ORGANISATION OF THE THESIS

The thesis has been written under six chapters which can be summarized in this section. The outline of the thesis and the corresponding publications are mentioned in Figure 1.1. Part of the research work related to this thesis has been already published in journals and conference proceedings as listed below. The results of the following research works are included in this thesis.

- Data Collection: Quality Assessment (Pragalathan & Schramm, 2019)
- Traffic flow prediction model comparison (Pragalathan & Schramm, Apr 2023)
- Real-time Simulations and Applications (Pragalatahan & Schramm, May 2023)
- Traffic flow Predictions with exogenous variables (Pragalathan & Schramm, 2024)

Chapter 1: Introduction

This chapter gives a short preface about the research. It explains the importance of traffic flow predictions and real-time traffic simulation. It also describes the motivation behind the study and narrates the objectives.

Chapter 2: Theoretical background and state of the art

This chapter elucidates the basic overview of

- data collection methods,
- traffic flow prediction models and
- simulation software- SUMO

This chapter justifies the use of SUMO and also emphasizes various research gaps in each section listed above. It reviews each and every step involved in the development of prediction system and summarizes them at the end of review.

Chapter 3: Methodology

This chapter discusses the detailed steps to attain the formulated research objectives and simulate the predicted results. The details about the data collection and its extraction, the step-by-step procedure for prediction of traffic flow and then modelling of the network and vehicle flow, the various techniques to calibrate the SUMO parameters are given in this chapter.

Chapter 4: Traffic flow predictions

This chapter expounds the common steps followed for forecasting an urban traffic stream. The quality assessment of traffic data, comparison of various prediction models, and the final observations about each model’s performance will be explained in depth in this chapter.

Chapter 5: Real-time traffic simulations and applications

This chapter highlights how the predicted results are given as input into the simulation setup for having real-time traffic scenarios. It also gives an overview about the applications of the developed prediction system and simulation setup. Simulation of platoon formation will be narrated as a sample application. The results of calibration, validation, evaluation of alternate scenarios with different platoon formation ratios are also given in this chapter.

Chapter 6: Summary of the study

The final conclusions of the study, limitations, future scope are discussed in this chapter.

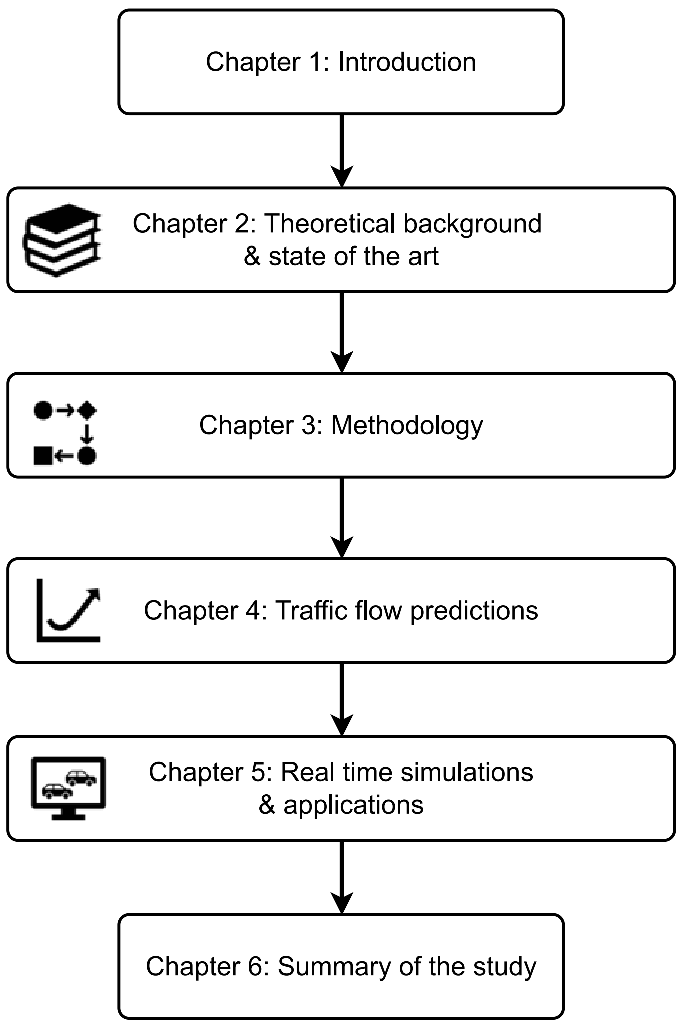

Figure 1.1 Thesis outline

CHAPTER 2

THEORETICAL BACKGROUND AND STATE OF THE ART

This chapter explains the basic concepts about the data collection methods, prediction models and simulations. In addition, it also describes the pros and cons of each data collection methods and the terminologies of time series models. This chapter also outlines the research gaps under each section and narrates the state-of-the-art prediction models.

2.1 DATA COLLECTION

Traffic prediction requires historical data from the field for better prediction. In addition to historical traffic data, the prediction process also requires several other data from various sources. Figure 2.1 shows the list of data generally collected for traffic predictions to incorporate the complex spatiotemporal dependencies and other external dependencies. Traffic data can be collected by various techniques using simple manual counting to complex automated virtual sensors listed by Forschungsgesellschaft fuer Strassen- und Verkehrswesen (FGSV) (Listl, et al., Oktober 2019). The selection of a particular data collection technique is mainly based on the duration of data collection and the type of data required. For real-time applications and traffic prediction processes, long-term traffic data is essential to collect traffic data continuously and should be able to store and transfer data automatically. This section focuses on explaining the fundamentals of the different traffic data collection techniques which are suitable for real-time and continuous data collection. In addition, the section also discusses about the advanced manual counting technique which was used to extract ground truth data from the field for a short duration to assess the quality of the other two methods.

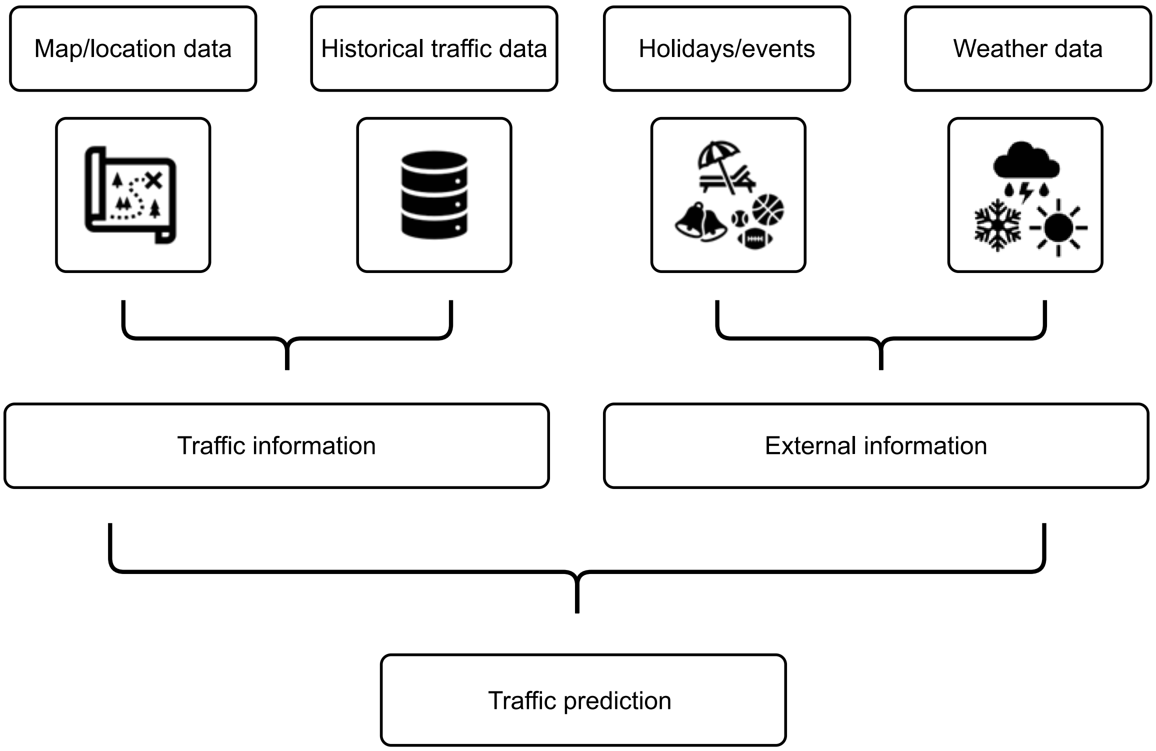

Figure 2.1 Data needed for traffic forecasting

2.1.1 Manual Count Made Easy

Manual counting is the simplest traffic data collection technique which is very apt for short-term data collection with the availability of manpower. Manual Count Made Easy (MCME) (Radhakrishnan & Menon, 2019; MCME V2 User Manual, 2012) is an advanced version of manual counting based on a graphical user interface where the video of the test field for data collection is recorded and the traffic counts are extracted by voice detection technique. It requires human intervention for speech recognition so that whenever a vehicle crosses a particular point in the video, details like the vehicle count, type of vehicle and time of passage can be stored. MCME is very suitable for short-term data collection when there is a possibility to record a video and if microscopic level of traffic data is required. The main advantage of this method is that the MCME method allows to have video as proof to check the vehicle counts and also can be used to extract different vehicle types simultaneously. The major drawback is that the video recordings require huge storage space for longer duration and are very tedious for longer-duration data collection. This method is also better in accuracy without the influence of other external factors like shadow or weather. Figure 2.2 shows the interface where the video file can be loaded, and the traffic data can be extracted as a text file based on voice recognition. This video was recorded from the application of Vehicle counter (Section 2.1.3) where the traffic data was simultaneously recorded by video detection.

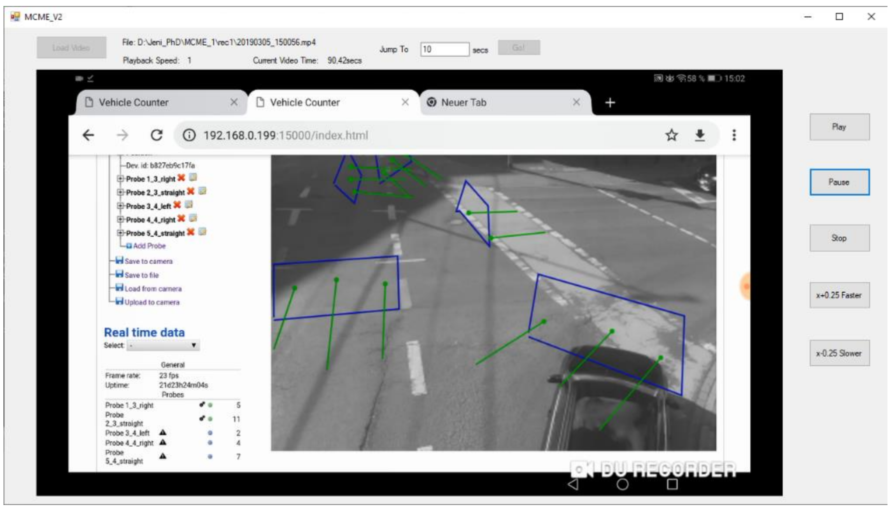

Figure 2.2 Interface of MCME (Manual Count Made Easy)

2.1.2 Inductive Loop Detectors

Inductive loop detectors (ILD) are based on intrusive data collection techniques. The detectors are installed under the roadway and the number of vehicles passing over the detectors is recorded. The accuracy of this method is better than other non-intrusive methods, because external factors like weather and lighting are not influencing the measurements. Moreover, actuated traffic signals at the intersection or any other actuated traffic control measures are getting the traffic data continuously from ILD. ILD is also capable of giving other information like occupancy: how long the vehicle requires to cross the detector, speed of a vehicle, vehicle length, waiting time of the vehicle standing over the detector etc. The major drawback of this method is that it needs proper maintenance which requires cost investment and traffic lane closures. Figure 2.3 shows the basic working of an inductive loop detector in the field.

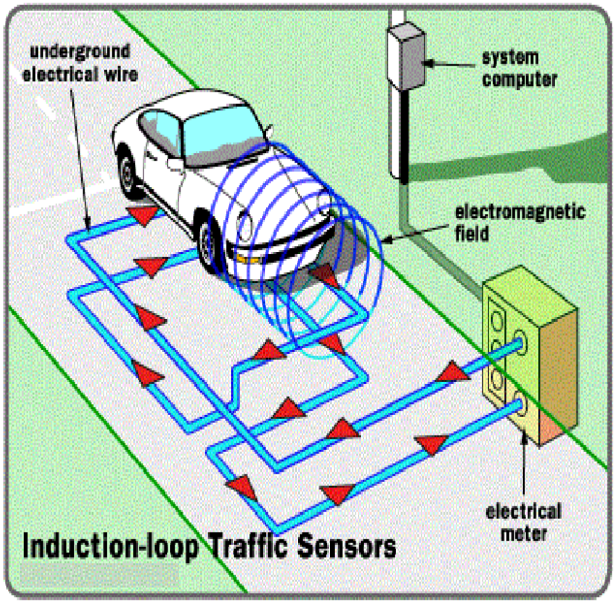

Figure 2.3 Working of Inductive Loop Detectors (ILD)

Source : (Somasekhar, Shirabadagi, & Hegadi, 2014)

2.1.3 Video Detection

Video detection is one of the recently developed methods based on virtual sensors. Without human intervention, this technique can detect the passage of vehicles with the support of cameras and traffic monitoring software. In this research work, a software named Vehicle Counter (Magenta Software Lab: Vehicle Counter, 2019) is used for traffic data collection. Thus, by installing the software in the external surveillance cameras and by fixing the cameras at higher locations like signal posts, traffic data can be extracted continuously and automatically. Since there is no need for video recordings, the storage and exchange of traffic data is also easier and faster. Another advantage is that a single camera can handle traffic in different directions and in different lanes as shown in Figure 2.4. The drawback of this method is that the accuracy of data collection depends on external factors like the location of cameras, lighting, and weather (snowy/cloudy).

Certain events like entering a lane in the wrong direction or high-speed driving can be triggered and the snapshots can be recorded. Thus, it helps with both traffic data collection and traffic monitoring.

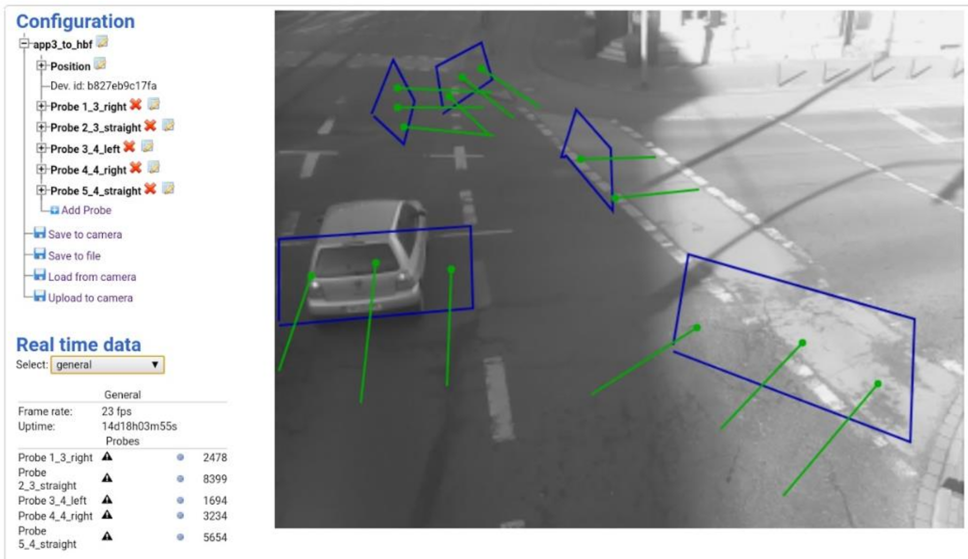

Figure 2.4 Definition of virtual sensors in vehicle counter software

2.1.4 Research gap

Traffic data collection methods evolved from time to time and attained various stages of advancements in many ways. Country-wise traffic data collection methods and the list of manufacturers and limitations of the intrusive and non-intrusive methods were listed in (Leduc, 2008; Skszek, 2001). Pneumatic road tubes, piezoelectric sensors, infra-red detectors, magnetic sensors, ultrasonic acoustic sensors and microwave radar detectors are some of the common methods of traffic data collection. Among various techniques, ILDs are the most commonly used and the most accurate ones (Skszek, 2001). Traffic data collection by video recording or detection is capable of giving traffic information at the microscopic level. The limitations of the video recordings due to area of coverage, mounting of cameras etc. were reported in the study (Hidas & Wagner, 2004). Real-time traffic data collection is one of the basic requirements for various real-time traffic applications and Intelligent Transportation Systems. Video-based detection techniques were preferred in certain situations which require vehicle classification data (Zhang, Avery, & Wang, 2007). One of the main drawbacks of inductive loop detectors is that it is not capable of such microscopic vehicle recognition based on the vehicle length. In this thesis, we need a traffic data collection technique that is suitable for long-term data collection and with more accuracy despite the external factors. Hence a quality assessment study is required to evaluate the various available techniques (Pragalathan & Schramm, 2019). Similar kind of analysis were done in the previous works (Klein, Kelley, & Mills, 1997; Minge, Peterson, & Kotzenmacher, 2011; Mousa & Bonnette, 2016).

2.2 TRAFFIC FLOW PREDICTION

Forecasting is one of the trending topics in many fields because of the availability of huge historical data and recent advances in big data management. Forecasting plays a major role not only in scientific research but also in other areas like business and marketing, weather and so on. Similarly in transportation field, traffic flow forecasting is a largely studied research topic due to the inexorable need for real-time and near-future traffic information for upcoming smart vehicles and smart infrastructures to update the driver and the road network.

Traffic flow forecasting is a technique in which historical data are used to predict future data. As per Forschungsgesellschaft für Strassen- und Verkehrswesen (FGSV) (Hoyer, et al., April 2012), traffic predictions are categorized as short-term, medium-term, and long-term predictions according to the prediction duration. Short-term predictions are the predictions for next few minutes e.g., 5 to 15 minutes. These predictions of traffic conditions help the network to adapt to the current demand and work accordingly. Short-term predictions also help the users in route choice for congestion avoidance. Medium-term predictions are prediction that are carried out for next few hours up to even one day. Such predictions help the users to plan for the next day or in the near future. Long-term predictions are the predictions for the following days or weeks. These predictions are mostly useful at the administrative level where decisions have to be made for future developments, budget allocations and strategic plannings and so on. Atypical or uncertainty predictions are another category which is used for predicting and analysing traffic during special events or holidays or severe weather conditions etc.

In this thesis, time series models are used where the traffic flow with time sequences are given as inputs for training and testing the model. If the past historical and present traffic data is available along with the time sequence, then it is possible to estimate the future traffic at a particular location or a given region. Time series prediction models utilise the available dataset to analyse the trend and changepoints, and then predict the future values up to a specific given period. While comparing time series predictions to algorithm-based predictions using Machine Learning (ML) techniques, time series forecasting is capable of extrapolating patterns outside the training data whereas most ML models cannot do by default. There are numerous techniques from simple linear regression to complex neural networks which can be used for building prediction models. In this thesis, two recent prediction models namely Fb-Prophet (FP) and NeuralProphet (NP) models are considered for traffic flow forecasting at an urban location. The performances of the models are compared with the classical precise statistical model SARIMA as baseline model (Pragalathan & Schramm, Apr 2023).

2.2.1 Terminologies

Stationarity

One of the basic requirements of time series for prediction modelling is that the dataset should be stationary which means the dataset should have constant statistical properties (mean and variance) over the time period. Non-stationarity can be caused by having a trend or seasonal pattern in the dataset. Thus, in a non-stationary dataset, the statistical properties will not be constant and will vary over a time period due to increasing or decreasing trends or repetitive seasonal patterns. By visually inspecting the time series plot in Figure 2.5 for the whole dataset, such trends or seasonal patterns can be found.

Unit root tests such as the Augmented Dicky Fuller test can also be carried out for more precision and the p-value can be verified (<0.05).

Seasonality

Seasonality in a time series refers to repeating patterns that repeat at every certain duration like hourly, daily, weekly, monthly, or yearly. These are obvious patterns that can be included in the prediction models while forecasting time series data. Figure 2.5 shows the time series plot with seasonal patterns that repeats every 24 hours. Differencing is the process of making non-stationary time series into a stationary time series by removing trend or seasonality. Seasonal differencing is carried out by subtracting the current value from the previous season's value.

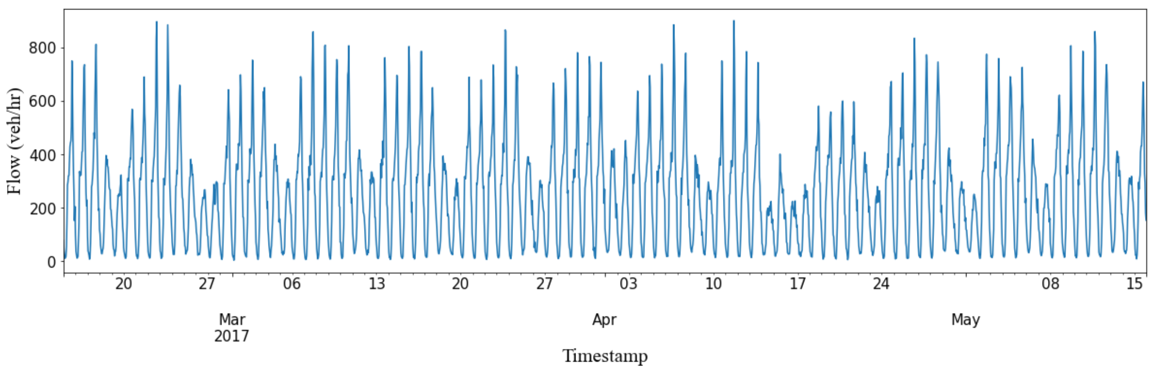

Figure 2.5 Time series plot with stationarity and seasonality

Source: (Pragalathan & Schramm, Apr 2023)

Autocorrelation Function (ACF) Plot

Autocorrelation is the measurement of the linear relationship between the lagged values of a time series (Box & Jenkins, 1976; Hyndman & Athanasopoulos, 2018). The autocorrelation coefficient at a particular lag 'k' quantifies the relationship between the current value at a time 't' and the previous lagged value at '$t - k$'. The autocorrelation coefficients are drawn against the corresponding lags and the plot is named as autocorrelation function (ACF) plot or correlogram as shown in Figure 2.6. When the ACF plots have a gradual decrease of values towards negative, it indicates the time series has a trend in it. Vice versa when the ACF plots have sharp cut-off, it indicates that there is no trend, i.e., stationarity is confirmed. When the coefficients decrease and increase with a repetitive pattern at regular intervals, it depicts the seasonality of the time series. The lag interval at which the repetitive variation is found is the number of datapoints in a single season. The threshold values in ACF plot is calculated by using the formula $\pm 1.96/\sqrt{T}$, where T is the total number of observations. When there are no significant autocorrelation coefficient values for all the lags or more than 95% of values are within the bounds of $\pm 1.96/\sqrt{T}$, then the time series is called white noise. It means that the white noise time series has a random amount of uncorrelated past values.

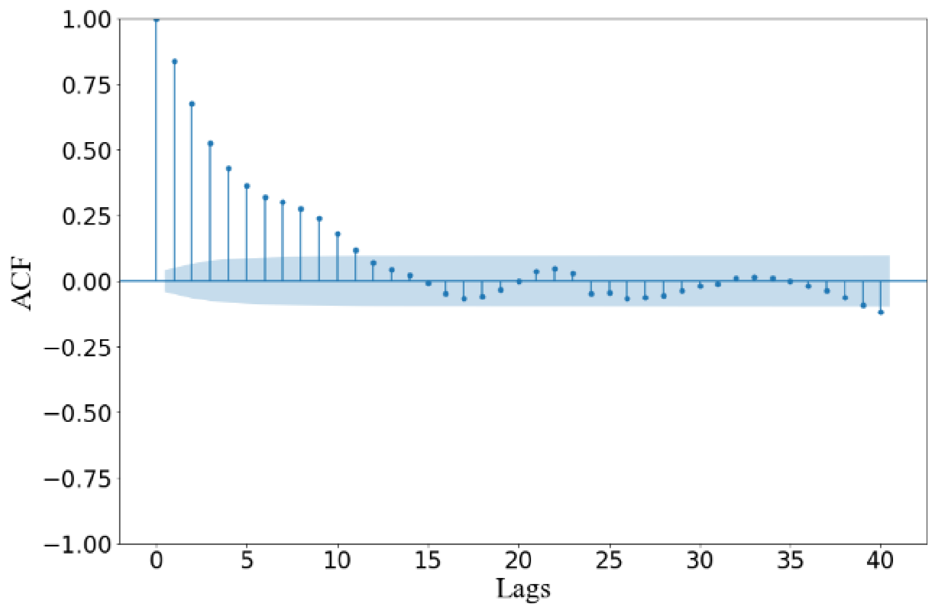

Figure 2.6 Autocorrelation Function (ACF) Plot

Source: (Pragalathan & Schramm, Apr 2023)

Partial Autocorrelation Function (PACF) Plot

Autocorrelation coefficients find the relationship between the current value and the values at specific lags. But the lagged values also correlate with each other. This effect is measured by using partial autocorrelation coefficients after removing the effect of other lags. Here PACF has control over other lags whereas ACF does not have control over other lags. The plots of PACF coefficients are termed as PACF plots as shown in Figure 2.7. Similar to ACF, the threshold values in PACF plot is also calculated by using the formula $\pm 1.96/\sqrt{T}$, where T is the total number of observations. ACF and PACF plots are used for determining the order of the ARIMA model (p, q).

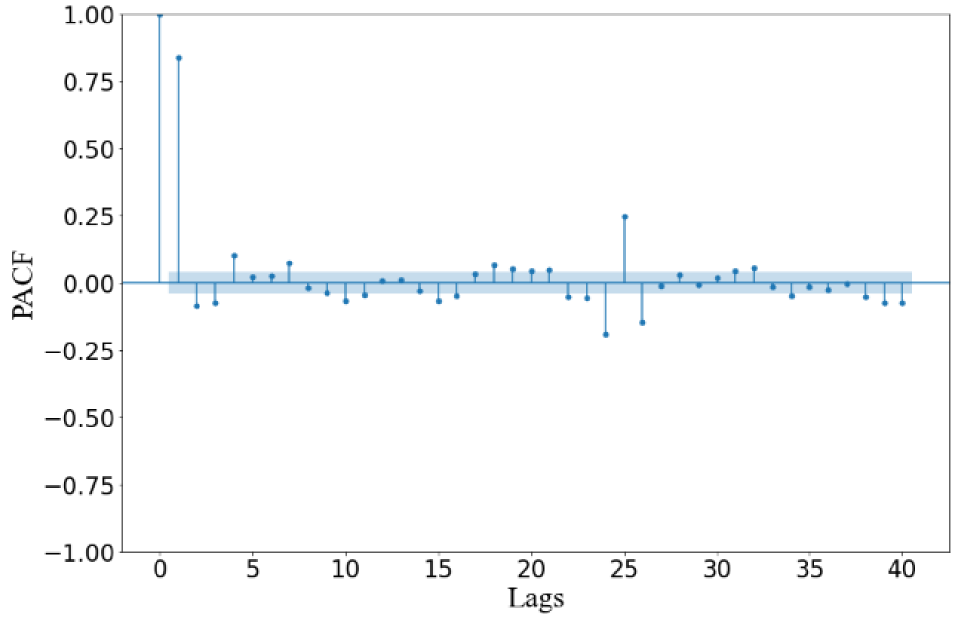

Figure 2.7 Partial Autocorrelation Function (PACF) Plot

Source: (Pragalathan & Schramm, Apr 2023)

2.2.2 State of the art

The complexity of prediction methods differs from model to model. The simplest prediction method is the naïve method, where the latest values are considered as the future estimated values. Another simple method is the average method, where the predictions are found from the average of past values. The complex methods include models with neural networks where various complex hidden layers help to predict future values. The selection of a method is based not only on the quantity and quality of data but also on the prediction requirement and environment.

Some of the time series prediction models are regression models like simple linear regression, multiple linear regression, decomposition models: additive or multiplicative models, exponential smoothing models: simple exponential smoothing, Holt's linear methods, additive damped trend method, additive Holt-Winters' methods, Multiplicative Holt-Winters' methods, Holt-Winters' damped method, ARIMA models: AR, MA, ARMA, ARIMA(X), SARIMA(X) etc. (Hyndman & Athanasopoulos, 2018).

Regression models forecast the time series data with the assumption of having a linear relationship between dependent and independent variables. Decomposition models can include several components each capturing the underlying patterns of separate factors or categories like trend, seasonality, holidays etc. Thus, the decomposition model splits the whole model into required components and predicts the dependent variable for a particular time duration. Exponential smoothing is the technique where prediction is done with the weighted average of the past values and the recent values have higher influence over the predictions and with exponentially decreasing weights for past values. This is the advanced concept of both naïve and average method having more influence of recent value as in naïve and taking the averages of past values like in the average method, but with weights. In the average method, the smoothing parameters are defined within the range 0 and 1, the oldest past value with the least parameter values.

According to previous works (Lana, Ser, Velez, & Vlahogianni, 2018; Barros, Araujo, & Rossetti, 2015; Zhang Y. , 2020), traffic prediction models can be categorised into mathematical, statistical, machine learning - deep learning and hybrid models. Among these models, Auto-Regressive Integrated Moving Average (ARIMA) model is one of the statistical time series prediction models and proven its accuracy for decades (Xu, Li, & Wang, 2016; Ghosh, Basu, & O'Mahony, 2009; Williams & Hoel, 2003). Various approaches (Williams B. , 2001; Van Der Voort, Dougherty, & Watson, 1996; Wang, Li, & Xu, 2017; Chen, Hu, Meng, & Zhang, 2011; Dong, Jia, Sun, Li, & Qin, 2009) have been proposed to attain several variations of ARIMA to improve the model's performance.

Traffic flow time series data exhibits seasonal behaviour with fluctuations that repeat daily due to peak/off-peak hours. Such seasonality can also be included in the ARIMA model by another version of it named SARIMA (Seasonal ARIMA). SARIMA can achieve more accurate predictions when compared to other parametric and non-parametric models (Smith, Williams, & Oswald, 2002). Additionally, SARIMA is also able to overcome the requirement of a huge historical database. It performed satisfactorily even with limited traffic data (Kumar & Vanajakshi, 2015) thus widening its usability when there is also a lack of huge database.

Previous research works also tried to propose several approaches from data handling to finding dependencies to improve the efficacy of the ARIMA model. Some of them are:

- Identification of key features like temporal or external factors that influence the everyday traffic flow (Wang, Li, & Xu, 2017)
- Data clustering the training data as weekday/weekend dataset (Van Der Voort, Dougherty, & Watson, 1996) and as morning/evening peak hour traffic (Dong, Jia, Sun, Li, & Qin, 2009)
- Data aggregation to different time intervals (Van Der Voort, Dougherty, & Watson, 1996; Chen, Hu, Meng, & Zhang, 2011; Dong, Jia, Sun, Li, & Qin, 2009)

Due to the inefficiency of statistical models to consider such dependencies, traffic predictions moved to complex machine learning and deep learning algorithms (Kashyap, et al., 2022). This type of experiment to improve the model efficiency was also carried out in other traffic forecasting models based on machine learning and mathematical techniques (Esugo, Lu, & Haas, 2022; Zhang, Yao, Du, & Ye, 2021).

Numerous models have been developed in the past decades for predicting traffic flow depending on machine learning and deep learning concepts. Based on the literatures (Lv, Duan, Kang, Li, & Wang, 2015; Kashyap, et al., 2022), some of these models are:

- Convolutional Neural Networks (CNN)
- Artificial Neural Networks (ANN)
- Recurrent Neural Networks (RNN)
- Long Short-Term Memory (LSTM), etc.

The complexity, applicability and computational cost of each model vary from one another. Hence every model has both advantages and disadvantages based on these variations. As proof, traffic flow predictions by using Random Subspace learning based on deep CNN (RSCNN) handled heavy traffic conditions successfully but the variation of traffic flow in weekend was not included (Deng, Jia, & Chen, 2019). Likewise, more accurate traffic flow prediction was carried out with Multi-feature Fusion-CNN (MF-CNN), but then the running duration of the model was not ideal and the model's performance was influenced by uncertainties like accidents (Yang, et al., 2019).

ANN based prediction approaches were becoming more familiar in the field of traffic flow predictions. The main drawback of ANN models was that they required a huge database for training and were also computationally more complicated. More experience was needed for optimising the model and its parameters and the performance of the model varied based on proper network structure (Cetiner, Sari, & Borat, 2010). The overfitting issues were another drawback of prediction with ANN (Khaz'ali, Emamjomeh, & Andayesh, 2011). The LSTM-RNN hybrid models were developed to address the problems which standard RNN models could not handle, one of which is requirement of the long-term dependencies (Tian & Pan, 2015). Therefore, the LSTM-RNN based models attained more accuracy while competing with other machine learning models like Support Vector machine (SVM), Random Walk (RW) and Sparse Autoencoder.

Numerous hybrid combinations of models were also developed mainly focusing on improving the accuracy (Song, Guo, Wu, & Ma, 2019). Addressing only the accuracy of the models ended up with more complex models with increased computational requirements (Tselentis, Vlahogianni, & Karlaftis, 2015). The deep learning models have several advantages like capability to capture the non-linearity of traffic flow and to adapt spatial dependencies of traffic flow (Deng, Jia, & Chen, 2019; Sun, Wu, & Xiang, 2020), to adjust to temporal dependencies (Osipov, Nikiforov, & Zhukova, 2020) and to address the issue of missing data (Li, Tan, Wu, Ye, & Ding, 2020). In addition, the disadvantages of deep learning models were also noted in the literatures, which are the requirement of huge historical data for training, overfitting (Feng, Xu, Lin, & Li, 2020) and complex models with development of more hidden layers. The necessity for human expertise for selecting the model's architecture and to fit the data accordingly.

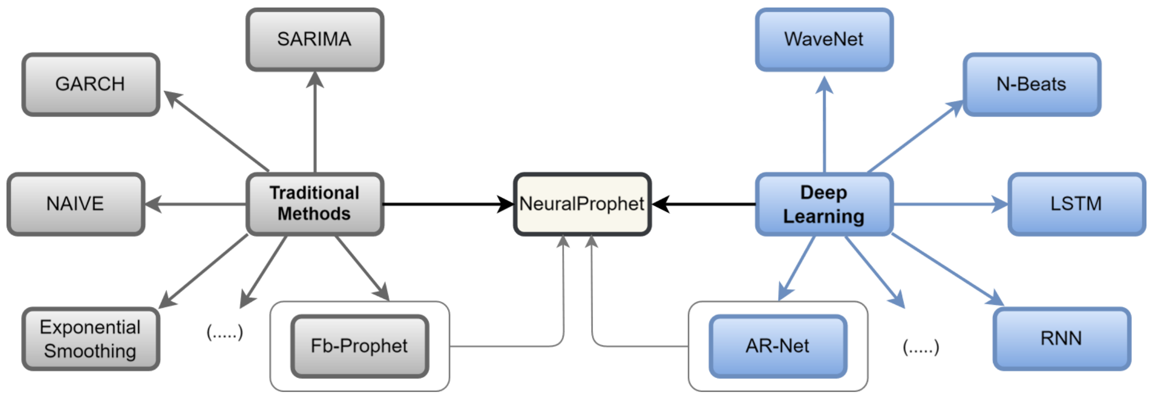

Figure 2.8 NeuralProphet model bridging the gap between statistical and deep learning models

Source: (Pragalathan & Schramm, 2024)

Therefore, for the development of better traffic prediction models, merits of both statistical and neural network models are necessary (Vlahogianni & Karlaftis, 2013). Fb-Prophet model improvises the performance of forecasting based on an additive decomposable model that can include the effect of seasonality and holidays (Taylor & Letham, 2017). Compared to other statistical methods, the Fb-prophet model is fully automated with tuneable parameters. The successor of Fb-Prophet model is NeuralProphet (Triebe, et al., 2021) model by taking the decomposability property from Fb-Prophet model with additional components based on AR-Nets (Triebe, Laptev, & Rajagopal, 2019). Therefore, NeuralProphet fills this gap between statistical and Neural Network algorithms and addresses most of the drawbacks of the algorithms by reducing the requirements of computational cost and human expertise in the process of modelling. Figure 2.8 depicts the various methods used for time series forecasting and shows the bridging of Fb-Prophet and AR-Nets for the development of NeuralProphet model. In this thesis, both of these two models are used for traffic predictions and their performance is compared with the SARIMA model.

2.2.3 SARIMA Model

Auto-regressive Integrated Moving Average (ARIMA) model is a traditional statistical method, based on autocorrelations in data. The Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model is one of the commonly used forecasting methods which works based on historical time series data including the seasonal time series. This model was developed by George Box and Gwilym Jenkins in 1970. SARIMA is a classical model that has proven to perform with higher accuracy in time series prediction problems. It is a combination of three different terms AR, I, and MA. ‘S’ stands for the seasonal part of the model. ‘AR’ stands for the autoregressive model where the data will be regressed on the past values. ‘I’ denotes the integrated part of the model where the differencing step is being done to remove non-stationarity. ‘MA’ indicates the moving average method in which regression error is found from the past error values. It works based on two mathematical models AR and MA with a step of differencing to remove non-stationarity. According to the need, this combination can be written in various forms like AR, MA, ARMA, ARIMA, ARIMAX, SARIMA, and SARIMAX. Figure 2.9 explains the usual procedure followed for forecasting with ARIMA models. The mathematical definitions, assumptions, model selection, and parameter estimation processes will be explained (Pragalathan & Schramm, Apr 2023) in this section.

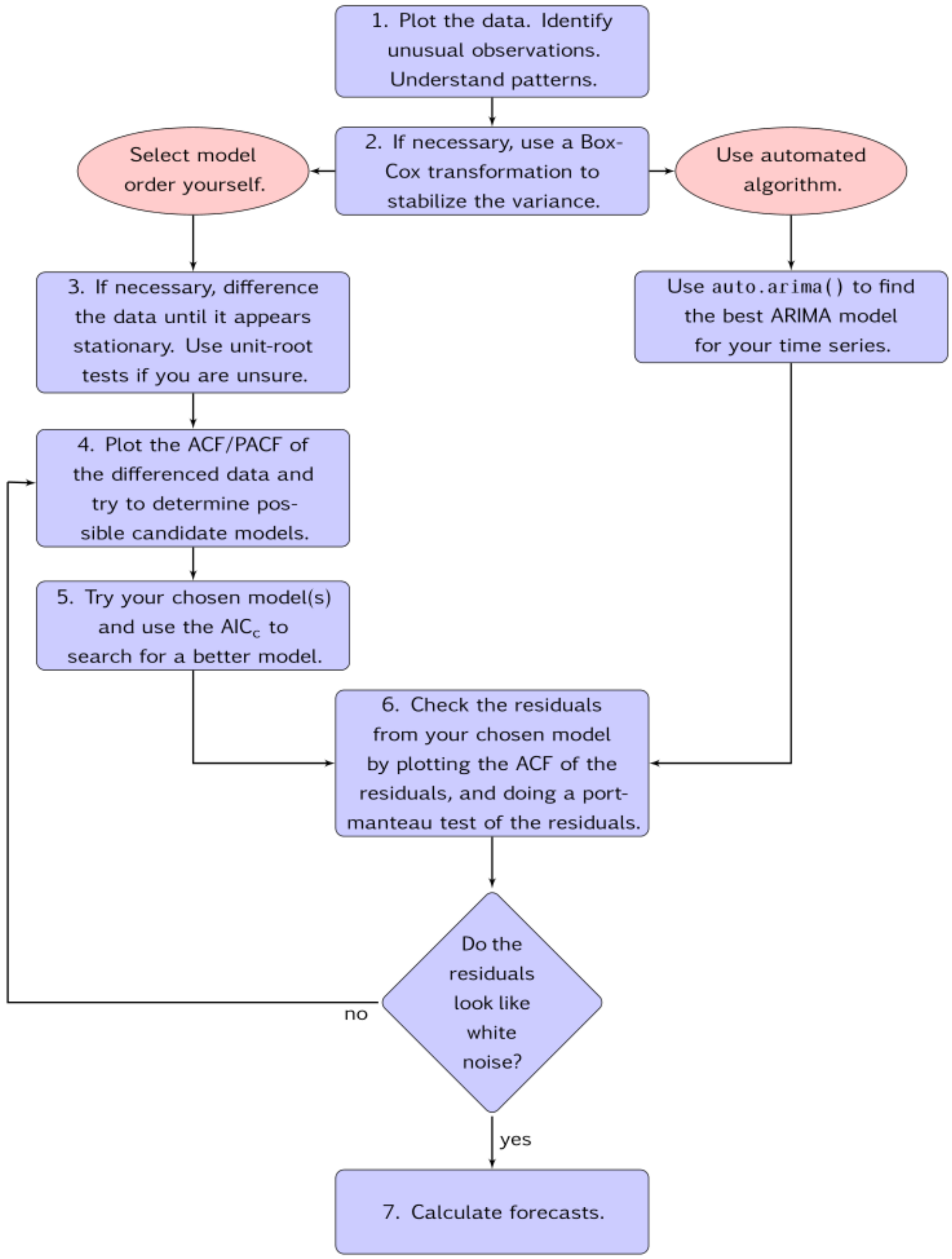

Figure 2.9 General forecasting procedure for ARIMA model

Source: (Hyndman & Athanasopoulos, 2018)

Model Definition

Combining the components of S, AR, I and MA, the mathematical equation can be represented with a summation of all the lags, their multipliers, the constant, and the white noise and it is given in the following equations.

$$y_t = AR(p) + MA(q) + SAR(P) + SMA(Q) \tag{2.1}$$

$$y'_t = c + \sum_{i=1}^{p} \phi_i \, y'_{t-i} + \sum_{i=1}^{q} \theta_i \, \epsilon_{t-i} + \sum_{i=1}^{P} \alpha_i \, y'_{t-si} + \sum_{i=1}^{Q} \beta_i \, \epsilon_{t-si} + \epsilon_t \tag{2.2}$$

Where y'_t represents the differenced data of order d or D, c is a constant, ϕ, θ, α and β are the coefficients of the corresponding models and ϵ_t represents the error value. In the equation, (p, d, q) are the orders of the non-seasonal part of the model, (P, D, Q) are the orders of the seasonal part of the model and s is the number of data points in each season.

Autoregressive Model- $AR(p)$ is a stochastic model where the value at time t can be expressed as a linear aggregate of values at previous equally spaced times $t-1, t-2$, and so on. It can be written as a linear regression model where the dependent variable 'y' at time t (y_t) is regressed on its previous independent values $(y_{t-1}, y_{t-2}, \dots . y_{t-p})$. Moving Average Model- $MA(q)$ is derived based on previous values of forecasting error, whereas the AR model is derived based on regressing its previous values. If the data in the given dataset shows fluctuations or increasing (or decreasing) trend that differ from time to time, this means the dataset does not have constant statistical properties (mean and variance). Such datasets are termed as non-stationary time series. The non-stationary behaviours of datasets can be rectified by doing the process of differencing and the differenced dataset with stationarity behaviour will be having fixed mean and variance. This method of pre-processing is referred to as Integrated and denoted as $I(d)$ where d is the order of differencing (usually $0 < d < 3$) (Box & Jenkins, 1976).

Combining both the AR and MA models, ARMA / ARIMA model will be obtained. If the process of differencing is required and the differenced time series data is used, then the ARMA model will become ARIMA. Suppose the time series data has a repetitive pattern of changes that repeats over a regular number of time periods (S). In that case, it represents seasonal behaviours, where S is the number of times the pattern repeats. For example, in urban traffic flow scenarios, seasonal behaviour is observed every 24 hrs. Such seasonal behaviours will lead to non-stationarity in time series. Because the statistical properties (mean, variance) at specific times within the seasonal span (days/weeks/months) may be different from the statistical properties at other times. Hence seasonal differencing will be required in order to remove non-stationarity.

The ARIMA model is capable of handling non-seasonal datasets. Hence to model the seasonal data set, a step of differencing is being done and such differenced dataset showing non-seasonal structure will be used to model by ARIMA. By effectively using an ARIMA model to a dataset that shows non-seasonal structure by doing differencing, SARIMA models take seasonality into account. Eventually, an additional set of terms will be included in the ARIMA model to include seasonal time series data and the model can be represented as $\text{SARIMA}(p, d, q)(P, D, Q)_S$. D represents the order of seasonal differencing to include seasonality. s represents seasonal periods, and it can also be denoted as h-hourly, m-monthly, and r-yearly seasonality (George & Jenkins, 1976). Time-series data may occasionally exhibit dependence on external variables, which are referred to as exogenous variables. The ARIMAX and SARIMAX models consider such exogenous variables measured at time t that influence the value of the time series data at time t. For this purpose, addition of the terms on the right side of our ARIMA and SARIMA equations will be done.

ACF and PACF Plots in SARIMA Modelling

Plotting ACF and PACF coefficients is one of the important steps in forecasting using ARIMA models. The plots help to identify the order of the model and also to detect patterns and check randomness. ACF plots are also used to identify non-stationarity in time series data. If the ACF drops to zero immediately in the beginning, then it suggests that it is stationary. Whereas if the drop towards zero is gradual and slow, it suggests that it is a non-stationary time series. The application of ACF plots can be explained by the following question “Can MA model be used to model the observed time series? If so, in what order?" Therefore, ACF plots are also used to find the order of MA model (q). Order estimation through ACF and PACF plots can be done only for stationary time series. For non-stationary time series, differenced time series dataset will be used to plot ACF and PACF plots and further order estimation will be done after this step.

The partial autocorrelation at lag k is the autocorrelation between the observed data at time t X_{t_t} and the data at $t-k$ X_{t-k} that is not accounted for by lags 1 through $k-1$. The PACF plot helps to identify whether an AR model can be used or not and if yes, then in which order. When comparing ACF and PACF plots, the following differences can be noticed. The average correlation between observed data points and previous values of the time series that are measured for various lag lengths is plotted in ACF plots. Whereas the only difference between a PACF and an ACF is that each partial correlation takes into account any correlations between observations with shorter lag lengths. Because they both measure the correlation between data points at time t and data points at time t-1, the values of an ACF and a PACF at the first lag are identical. However, the PACF measures the same correlation after controlling for the correlation between data points at time t and those at time t-1, whereas the ACF measures the correlation between data points at time t and data points at time t-2 at the second lag.

2.2.4 Fb-Prophet Model

Fb-Prophet is an open-source forecasting model launched by Facebook's core data science team in 2017 to forecast time series data with a modular regression model with default parameters (Taylor & Letham, 2017). This model can be implemented in Python and R programming languages. One of the main objectives of this model is to provide an easy and fast tool for time series forecasting.

Model Definition

Fb-Prophet model was developed based on an underlying additive model that can be decomposed into various components to include the effect of trend, seasonality, and holidays. The mathematical representation of this nonlinear regression model is:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{2.3}$$

Keeping 't' as the regressor, $y(t)$ represents the value of dependent variable at time 't'. $g(t)$, $s(t)$, $h(t)$ are the functions representing trend, periodic changes due to seasonality and the influence of holidays respectively. ϵ_t denotes the error term due to idiosyncratic changes and it is assumed to follow normal distribution. A saturating growth model and a piecewise linear model are considered to implement the trend component. To include the effect of seasonality, the standard Fourier series is defined with a matrix of vectors seasonality for each value of 't' in the past and future time values. Country-specific holiday lists are used to include the shock created by events or holidays and a matrix of regressors is created in a similar way to seasonality. All these matrixes and changepoint indicators are combined and developed as the Prophet model also named as Fb-Prophet. Because of this property of decomposability of the additive model, it can include the required components as necessary among the trend or seasonality or holidays when it is newly identified and also each component can be visualised separately.

Analyst in the loop

The model uses the "analysts in the loop" approach (Figure 2.10) for the performance analysis where most of the forecasting and evaluation will be done by automated tasks and the modelling and inspection will be carried out by human interferences. Because of this approach, statistical forecasting is fully automated and at the same time, the analysts can intrude on the model with a small set of intuitive parameters. Since the parameters of the underlying model will be automatically extracted and unlike SARIMA, only lesser manual interpretation is needed in Fb-Prophet to define and choose the parameters. The Fb-Prophet model also follows the "forecasting at scale" approach which is a practical way of combining configurable models. With the proposal of a modular regression model with interpretable parameters, the process of forecasting is made automated in a tool that can work with lesser human interference (Taylor & Letham, 2017).

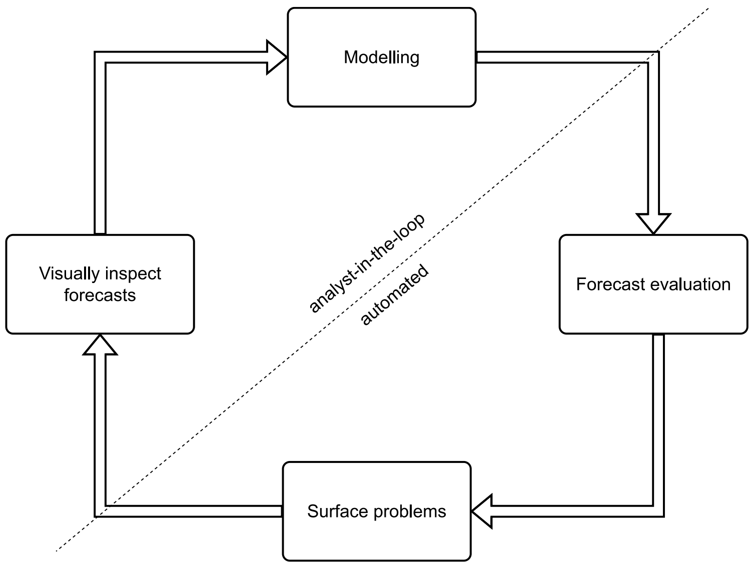

Figure 2.10 Analyst in the loop

Source: (Taylor & Letham, 2017)

Hence, the analysts are allowed to decide the values of smoothing parameters whenever needed even with limited statistical knowledge but based on historical data and domain knowledge. Because of this approach, statistical forecasting is fully automated and at the same time, the analysts can intrude on the model with a small set of intuitive parameters. The parameters and coefficients of the variables of the underlying model will be automatically extracted. Thus unlike SARIMA, manual interpretation is not needed in Fb-Prophet to define and choose the orders and parameters. The Fb-Prophet model follows the "forecasting at scale" approach which is a practical way of combining configurable models. With the proposal of a modular regression model with interpretable parameters, the process of forecasting is made automated in a tool that can work with lesser human interference.

2.2.5 NeuralProphet Model

NeuralProphet is the successor of Fb-Prophet model with some more components added to the model equation. It is also an open-source model developed by a team of researchers along with Facebook in the year 2020. It is a user-friendly tool with more interpretability and configurability. It has more features for automatic capabilities with the support of PyTorch.

Model Definition

The NeuralProphet model bridges the gap between traditional models and deep learning models and performs better for near-term future predictions. This hybrid combination of models (Fb-Prophet & AR-Nets) and the property of modular decomposition makes this NeuralProphet model very accurate and has more additional functionalities like inclusion of non-linear deep layers, autoregression, covariates etc. In addition, seasonality can be included either as additive or multiplicative terms. The mathematical representation of the model can be written as:

$$y(t) = T(t) + S(t) + E(t) + F(t) + A(t) + L(t) \tag{2.4}$$

Where each term represents a specific model at time 't'. $T(t)$ represents the trend function, $S(t)$ represents seasonal effects, $E(t)$ handles the influence of events and holidays, $F(t)$ denotes the effect of regression for future-known exogenous variables, $A(t)$ includes the effect of Auto-regression based on previous values and $L(t)$ signifies the effect of regression for lagged observation of exogenous variables. Both the Fb-Prophet and NeuralProphet models were used for predictions in this paper and the efficacy of the models was evaluated based on accuracy and computational costs.

Comparison of Fb-Prophet and NeuralProphet Models

The original Fb-Prophet model and NeuralProphet models have modular decomposability as the basic common property. Even though, the major difference is the inclusion of the additional components in NeuralProphet model, few other differences giving additional score to the NeuralProphet model are:

- Pytorch as back end helps more extendibility and more optimized methods.
- Hybrid model: includes both linear & Neural Network (Fb-Prophet & AR-Nets).
- NP model supports auto-regression and covariates.

- Automatic Preprocessing features: handling missing data (bi-directional linear interpolation / centred rolling average).
- Feed-Forward Neural Network (FFNN) with non-linear deep layers.

2.3 SUMO SIMULATIONS

SUMO (Simulation of Urban Mobility) (SUMO User Documentation, 2023) is a free and open-source, space-continuous and time-discrete, microscopic simulation software developed for intermodal traffic systems with highly portable models (Lopez, et al., 2018). SUMO can simulate a particular network with given traffic demand and at a level of every single vehicle and thus it is a microscopic simulation software. Every single vehicle can be modelled with explicit routes and other vehicle moving behaviours. It is also capable of handling larger urban networks and their complex traffic movements. SUMO is very much portable since the simulation can be run on different platforms like Windows, Linux etc., It is flexible since simulation can be called not only from GUI (Figure 2.11) but also from other options like command line prompt and python scripts. This facilitates the users to simulate the traffic and extract results in batch mode without the need for GUI. This makes the simulation process much faster and can be used for simulating larger networks.

SUMO has several applications that can be used for different purposes like importing/creating/preparing/visualising/computing network or traffic demand. Some of the applications included in the package are:

1. Importing: netconvert and polyconvert.
2. Visualising: sumo, sumo-gui.
3. Generating/creating: netedit, netgenerate, activitygen, emissionMap.
4. Computing: duarouter, jtrrouter, dfrouter, od2trips, emissionsDrivingCycle.

All these applications can be called and defined by just typing their name in the command line in a simple and easy way. It is also possible to import the data files or network files from other software like VISSIM, MATsim and so on. The software is utilised for different urban traffic scenarios, traffic managements, intermodal or multimodal traffic simulations, vehicle communications, emissions calculations and many more (Krajzewicz, Erdmann, Behrisch, & Bieker-Walz, 2012).

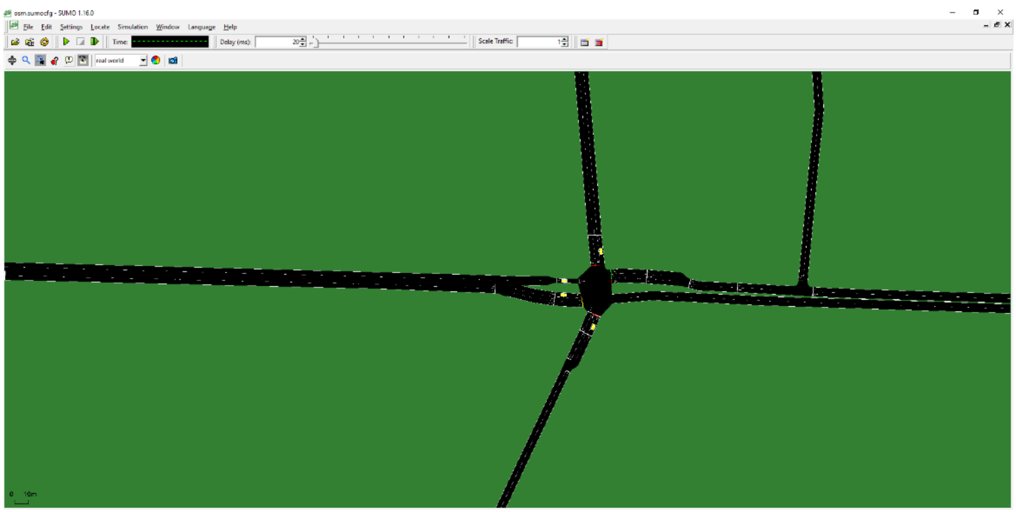

Figure 2.11 SUMO interface

2.3.1 SUMO simulation models

Simulation software is built based on certain models to describe the nature of the traffic stream. In SUMO, the car following and lane changing models are used to replicate the real-world traffic condition. By default, Standard Krauss model is used as the car following model (Krauss, 1998) and LC2013 as the lane changing model (Erdmann, 2015). SUMO also gives the option for sub-lane simulation with SL2015 as default model where the lateral resolution during lane-changing operations can be described with this model (SUMO Proceedings, 2016). SUMO has a vast library of different car following and lane-changing models. The default car following model is implemented with the main objectives of collision avoidance even when vehicles drive faster i.e., faster traffic stream with perfect safety. Thus, deceleration and safe velocity are mainly focused to maintain safety.

2.3.2 Parameters

The models used in SUMO are developed with a set of parameters that can be used to define within a certain range of values to describe the traffic stream exactly. The parameters are defined with default values but also can be adjusted according to the particular traffic behaviour. Some of the parameters of car-following models are minimum gap, acceleration, deceleration, maximum deceleration during an emergency, delay time to start driving after a stop, driver imperfection etc., and for lane changing model: eagerness for lane changing, willingness for cooperative lane changing, eagerness to gain speed and many more. All these parameters can be modified from default values and experimented to calibrate the model. The list of parameters and their values used for simulation in this thesis work are tabulated in Table 5.1.

2.3.3 Inputs

For every traffic simulation, the fundamental data which are extracted from the real world have to be given as input to exactly replicate the real-world traffic conditions. Some of the input data are:

1) Data from the network: number of roads, number of lanes, width of lanes, location of bus stops etc.

2) Traffic control system: signal, pedestrian crossings, traffic signs, priority lanes and so on.

3) Traffic data: demand or traffic flow, vehicle composition, speed distributions etc.

To account for the diversity of vehicles, a number of vehicle types can be defined with various parameters for each vehicle type. Thus, the vehicle composition of heterogeneous /mixed traffic stream can be represented in the simulation. Additionally, each vehicle type/vehicle can be assigned a particular route by defining routes and route distributions.

Thus, public transportation routes within urban cities and their arrival time, waiting times etc., can be also included in the simulation, making it a multimodal simulation.

2.3.4 Simulation

After defining the simulation environment through the above-mentioned models and parameters, the simulation properties must be defined. All the files defining the network, traffic demand, route and other additional files can be called through this process. Additional files include the definition of detectors, variable speed signs, traffic signal programs and bus stops. Simulation duration and time step length are also defined before starting a simulation. The default time step length value is one second and by reducing the time step length value increases the time taken to complete the simulation.

2.3.5 Outputs

A variety of outputs can be extracted from a single simulation. Apart from the basic outputs like vehicle positions, trip information, etc., many other additional outputs can be defined in recent developments. Some of the common outputs are floating car data, emission output, inductive loop detector measurements, edge/lane traffic values, queue output, collision output and so on. SUMO has several real-time applications and one of the recent trending topics is simulation of vehicular communications, which will be handled in this thesis.

2.3.6 Simpla Plugin

Platooning (SUMO: Simpla, 2023) is one of the recent trending research topics which are commonly tested in simulations. SUMO allows formation of platoons with a leader and follower vehicles with the development of a plugin named Simpla. The simulation of platoons can be configured through TraCI (Traffic Control Interface) Python client. Additional vehicle types should be defined for Simpla to represent leader, follower, catch-up and catch-up follower modes. Other than this, a set of parameters can be configured and listed as maximum number of vehicles in platoon, maximum gap between platooned vehicles, maximum platoon headway and catch headway etc. Table 5.1 shows the list of parameters used for platoon formation by using this plugin.

CHAPTER 3

METHODOLOGY

Before starting the process of traffic forecasting, the following questions should be answered to formulate the methodology and describe a suitable problem definition. 1. "Which traffic parameter has to be forecasted? (like speed or flow or density or travel time, etc.,)", 2. "Why it is forecasted?", 3. "How can it be applied for further usage?". This chapter answers all these questions, and the formulated methodology will also be explained step by step.

3.1 CONCEPT

Real-time smart solutions are becoming a vital factor in handling urban traffic related problems. In particular, the real-time and near future traffic predictions gained much considerable attractions for reducing congestion and travel time. For e.g., if there is congestion building up in the current route, the driver should be immediately updated and suggested with alternate routes if there is any. The traffic control system must be equipped not only to provide real-time updates to drivers and networks based on current traffic conditions but also to anticipate near-future traffic situations. This foresight is crucial due to the possibility of traffic conditions changing by the time a driver reaches a specific location. Hence not only the real-time traffic, but also the near future traffic prediction is also important. Traffic predictions have received much attention in the last years due to the widespread, readily available and variety of huge traffic data. Most studies have only focused on traffic predictions on highways or freeways. Urban traffic needs a special approach where the various external dependencies applicable to the locality and overall city/area like day of week, time of day, weather, holidays and so on should be considered in the prediction process.

The traffic flow fluctuation during the weekday and weekend or holiday is different. Hence the influence of such external factors should also be identified by the prediction model. In addition, the traffic flow before/during/after a severe weather like heavy rainfall or snowfall also influences the traffic flow. Since there is a chance of postponing the trip or planning early before the rainfall/snowfall and also choosing other modes/routes of transport. Hence the usual traffic will not be there due to the severe weather condition. Hence for most updated prediction system, the models should be able to capture the effect of weather, for better traffic network updating and handling. Thus, ultimately influencing the real-time or near-term future traffic predictions in the strategies of real-time solutions. This chapter proposes a methodology to predict urban traffic and simulate the predicted traffic. Thus, real-time traffic simulation will be obtained as the final output which can further be used for various applications.

Development of urban traffic forecasting system requires various phases of working.

1. Identifying and defining the traffic forecasting problem.
2. Gathering in-depth knowledge about the real-world traffic data and behaviour.
3. Selection of an appropriate and efficient forecasting model after comparing several models' performances.
4. Final model evaluation and summarizing the scope and limitations if any.
5. Transferability to another location or huge network.

3.2 FRAMEWORK

The common steps involved in prediction processes are listed along with the important points to consider during each step.

Step 1: Problem Definition

- What data should be forecasted?
- Data collection duration
- Interval of data collection and prediction
- Data type

Step 2: Data Collection

- Various available methods
- Merits and demerits of each method
- Suitable method of data collection
- Comparison of data collection techniques (optional)
- Selection of suitable data collection method according to problem definitions
- Extraction of data

Step 3: Preliminary Data Analysis

- Drawing time series plots
- Visual analysis
- Stationarity and seasonality check by visual inspections
- Preliminary understanding about the historical data

Step 4: Model selection

- Number of models for current data type
- Selection of appropriate models

Step 5: Model fitting and forecasting

- Order Estimation
- Training and testing the models
- Forecasting

Step 6: Model Evaluation

- Evaluation of metric/error values calculations
- Comparison and evaluation of models

Thus, the predicted system can be integrated with a simulation-readable file and modify the format of predicted results according to the appropriate applications. The flowchart for the development of traffic flow prediction systems is shown in Figure 3.1 and explained in depth in the following subsections.

3.2.1 Problem definition

Problem definition is the first and foremost step in forecasting problems in which most of the important items have to be defined. The knowledge about purpose of the prediction and final expected outputs are mandatory for better definition of problem. This makes the further process of prediction much faster and easier. This thesis mainly focuses to forecast the traffic flow/volume in a single urban intersection. This forecasting system then can be extended to a bigger network and can be used in driving simulator which requires current traffic conditions in numeric values to replicate exact traffic conditions.

Hence the traffic flow was decided as the main forecasting variable. Apart from traffic flow, it is also possible to predict other parameters like traffic speed, density, travel time and waiting time etc. The next essential thing in problem definition is the forecasting horizon i.e., the duration of predictions up to which the model must estimate the traffic flow. In this thesis, the goal was set to predict the traffic flow for the next 24 hours in order to simulate the traffic patterns for one full day in the future. Prediction frequency is another critical thing, where the frequency of running the prediction model must be decided. This can be decided according to the requirement, scope and final application. For e.g., for real-time intelligent solutions, prediction models are run more frequently to know the current and near future traffic and to implement control strategies immediately. But for driving simulators, the prediction models are called only when there is a requirement of simulation for vehicle dynamics analysis.

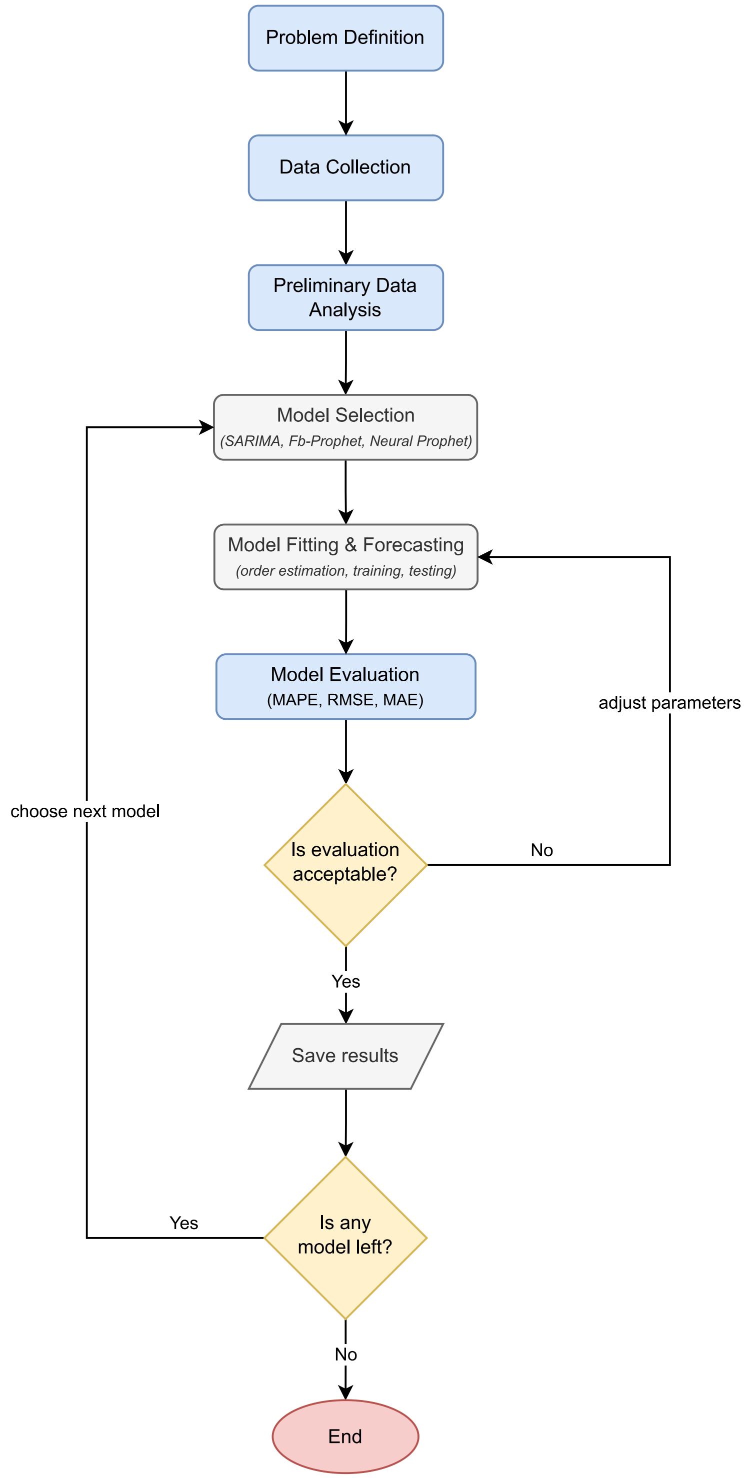

Figure 3.1 Traffic flow prediction methodology

3.2.2 Data collection

Once the traffic prediction problem is defined, the type of data and method of data collection, and the duration of data collection etc., should be decided. The frequency of data collection, i.e., every one minute or 5 minutes, should also be decided. The available data collection techniques should also be analysed, if possible, to acquire more accurate traffic data. Nowadays, traffic data are collected regularly for various purposes. Hence the data collection must have already been done in such cases and only data acquisition must be carried out.

Data collection includes collecting historical traffic information and also other basic knowledge about the local traffic conditions and its influencing factors. It is also very important to collect very recent data so that the prediction models perform better with more accuracy. For our prediction problem, the traffic data was collected along with time series and at regular intervals of every one-minute. The minute wise data was then extracted to hourly traffic data for more convenience and accuracy.

This research work required huge traffic data for prediction of future traffic flows. Such historical data was collected from inductive loop detectors and video detection technique.

3.2.3 Preliminary data analysis

Preliminary analysis is usually carried out with graphs and visual analysis. The trends and variations are noticed, and the stationarity and seasonality of the dataset are decided during preliminary analysis. The influence of weekends and holidays were evidently noticed in the traffic flow time series plots. Preliminary data analysis helps to find the properties of the traffic data and also useful for the selection of both model and its parameters.

3.2.4 Model selection

There are several models developed in the past decades for prediction problems. Depending on the nature of data (times series) and the influencing factors like holidays, the models can be selected and compared for better prediction. The type and size of historical traffic databases influence the selection of prediction methods or model. If there is only limited data available, then certain models can handle such smaller traffic data e.g., SARIMA (Kumar & Vanajakshi, 2015). If there is missing data, then the performance of certain models varies accordingly, nevertheless, certain models like Fb-Prophet and NeuralProphet, can handle missing data effectively. The type of data also plays a vital role in selection of the prediction model. Traffic flow data are numerical information with timestamps, thus making it as time series data and time series forecasting models can handle such databases effectively.

3.2.5 Model fitting and forecasting

The gathered traffic data is divided into training and testing datasets. Typically, 80% of the data is used for training the model, while the remaining 20% is allocated for testing the model's performance. Every model has certain assumptions and a set of parameters which must be estimated by fitting the training dataset into the model. Then after model fitting, the forecasting for the testing period is carried out and the error values are calculated.

3.2.6 Model evaluation

The performance of a prediction model for a particular database depends on:

1. Size of database: the ability of model to handle both huge and limited database
2. Data pre-processing: type of data, data format, data aggregation
3. Seasonality: repetitive pattern in daily/weekly/monthly/yearly/seasonally
4. Versatility: to include the various components like trends, change points, events, holidays
5. Computational cost: the processing time and the hardware requirements
6. Complexity of the model: parameter definitions and order estimation of models which requires expertise in model fitting, more complex models require more computation time

7. External factors: ability to process and include the influence of external factors

The performance of a prediction model can be evaluated with different evaluation metrics. By considering both the actual and predicted values, the error values can be calculated and interpreted to find the accuracy level of the prediction models. Some of the error values used in this thesis are:

- Mean Absolute Percentage Error (MAPE)
- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)

CHAPTER 4

TRAFFIC FLOW PREDICTIONS

Versatility in predictions means that the model should be able to capture the changing patterns during various conditions. The changing pattern in the past is assumed to be followed also in the future. If the changes are highly unpredictable, such volatile changes should also be captured along with information like how volatile it was, when and how the volatile change occurs. This analysis gives a basic understanding of the historical database and gives better predictions. This chapter explains the process of forecasting based on time series models.

4.1 PREDICTION MODELLING FRAMEWORK

The original patterns of the past data should be captured by the prediction models. For e.g., the morning and evening peak hour traffic patterns are repeating every day. But the traffic flow values depend on several factors, one of them is the weekday/weekend influences. The peak/off peak hour traffic pattern must be captured but with weekend traffic flow. Such flexibility in estimations to differentiate the random fluctuations and the original fluctuations should be mainly focused on, especially in traffic predictions. The flowchart depicted in Figure 4.1 illustrates the process of prediction modelling, starting from reading input data and proceeding through the various stages until reaching the forecasting step.

4.2 TRAFFIC DATA COLLECTION

Historical traffic data serves as the backbone of traffic predictions. The accuracy, size, frequency and type of data influences the performance of a prediction model. In this section, the acquirement of data will be clearly explained. Considerable attention was paid when collecting the data for better accuracy. As mentioned in the previous chapter, to determine the best suited traffic data collection method, a quality assessment was carried out to verify the performance of video detection technique and inductive loop detectors.

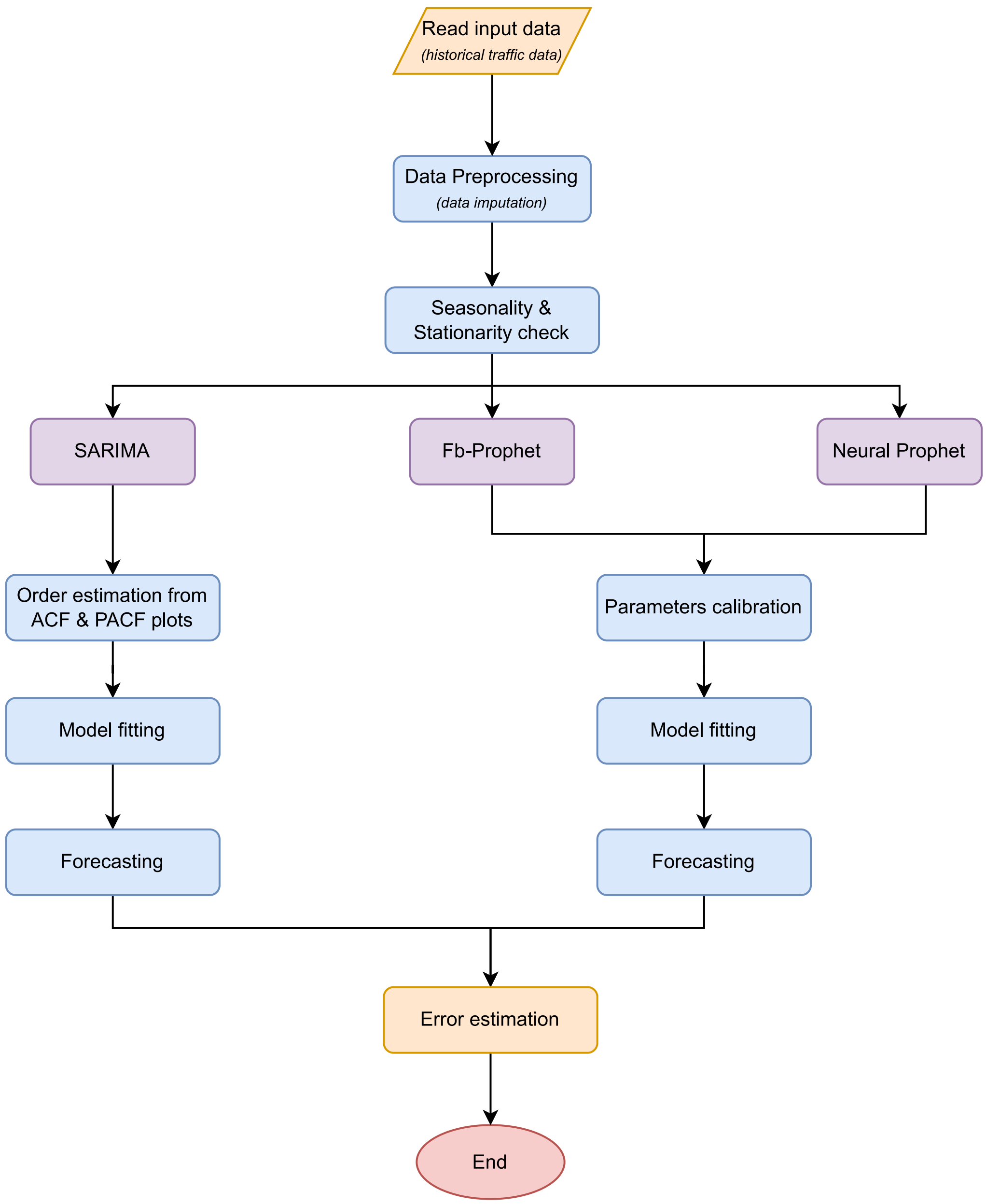

Figure 4.1 Methodology for comparison of traffic flow prediction models

4.2.1 Test bed overview

Duisburg is the 15th largest city in Germany with around half a million inhabitants (Duisburg, 2023). According to the data from the Duisburg office for elections and information logistics as of 31/12/2023, there are 507 thousands of people residing in the city (Duisburg in Zahlen, 2023). The city is renowned in Europe for its significant steel production industries. Duisburg is one of the transport hubs of European significance which can be reached through roads, rails and waterways. Being a centre for international trade and logistics, Duisburg had attained 23rd GDP (Gross Domestics Product) ranking among German cities by economic output (Duisburg Verkehr: Wikipedia, 2023). Having such a noteworthy profile, Duisburg is always dependent on effective mobility options. Not only for the goods, but also the residents have increasing demand for everyday mobility. The traffic demand is increasing day by day, which also influences the economic development and the quality of life of the inhabitants. The city of Duisburg tries to handle this increasing demand with strategies “Duisburg2027” (Duisburg: Verkehrkonzepte, 2023). Hence formulation of real-time simulation of traffic scenarios, helps to make smart decisions and better plannings for the future traffic conditions.

The test bed chosen for data collection of traffic flow is at an urban intersection in Duisburg city, Germany as shown in Figure 4.2. The multimodal intersection has four approaches with a number of lanes for each turning movements and a tramline running through it. It also has a considerable amount of pedestrian and cycle users, where the traffic is controlled by signal lights. During the traffic data collection process, careful consideration was given to collect data for each lane. This approach was adopted to enable the microscopic simulation of real-world traffic scenarios in virtual environments. The traffic signal light data was also obtained from the city administration. The chosen intersection is located very close to the city center and the University of Duisburg-Essen. The test bed is highlighted in Figure 4.3. There is also an Autobahn (in 1.1 km) and a Zoo (in 600 m), where the vehicles enter and exit the city. Thus, there is always a huge idling traffic with higher waiting time. Hence, it was also kept in mind to capture the congestion so that it can be incorporated into the model and then predicted.

Figure 4.2 Images of test field

Source: (Website: Google Maps, 2019)

4.2.2 Field data collection and data analysis

In order to achieve the formulated objectives or research gaps as in Section 1.2, the test bed was selected in such a way to collect traffic flow data effectively. The location of test bed was selected after several inspections by considering various points such as

- Possibilities to capture all complex vehicle movements
- Possibilities to mount cameras for capturing all traffic inflows and turning movements
- Possibilities to collect traffic data continuously for long-term duration

A comparative analysis on the performance of three different data collection methods were conducted and will be explained in this subsection. As explained in the previous chapter, advanced manual counting, inductive loop detectors and videographic detection were considered for analysis.

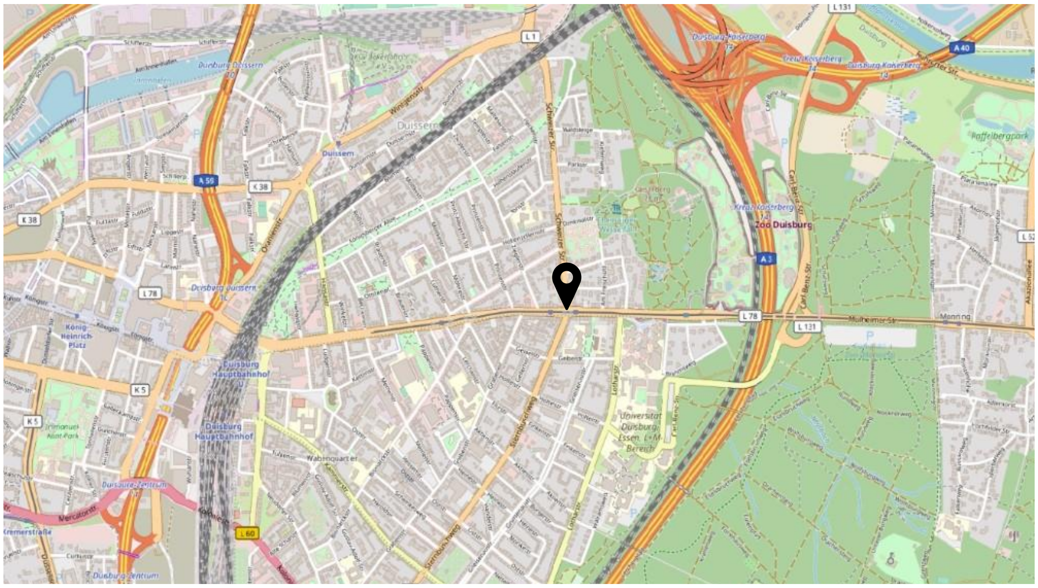

Figure 4.3 Snapshot of OSM map showing the two roads crossing the intersection

Source: (Openstreetmap, 2023)

Figure 4.4 depicts the layout of the intersection with indication of the location of inductive loop detectors. The approaches are numbered for east, south, west, and northbound as 1, 2, 3 and 4 respectively. The comparative analysis was conducted in two phases (short-term and long-term). The first phase of analysis focused on comparing the traffic data from MCME and inductive loops. Three samples of traffic data were collected randomly on 05/03/2019 (two 10-minute samples) and 29/04/2019 (one 15-minute sample). The second phase of analysis was conducted to compare the videographic detection and inductive loop detection techniques. For this phase, one week (02/04/2019 to 08/04/2019) data was collected and the error values were calculated. It took many trials to define and configure the virtual sensors in videographic detection. The distinct virtual sensors were drawn for each lane at an optimum position to reduce error due to shades from nearby lanes and also gates for each virtual sensor were defined to indicate the data collection at a particular direction.

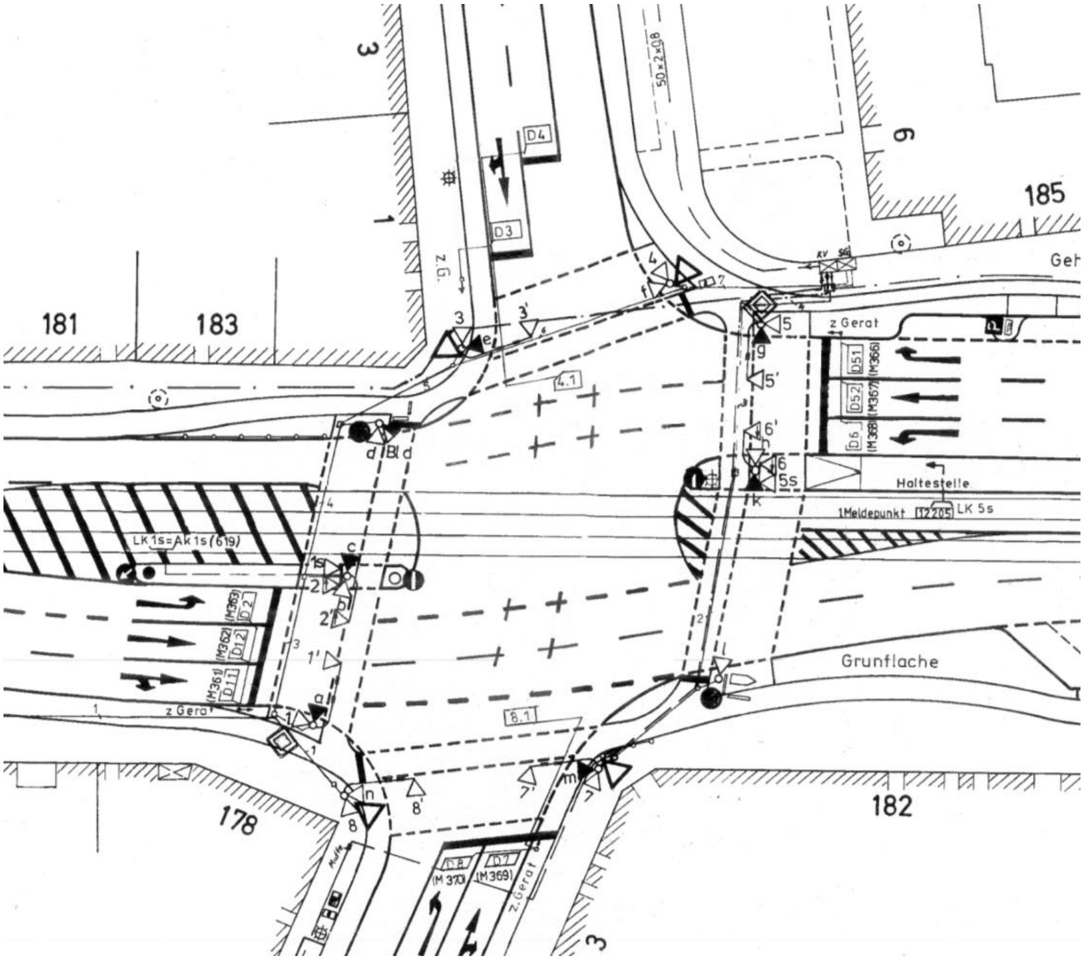

Figure 4.4 Layout of the intersection

Source: (Wirtschaftsbetreibe Duisburg, 2019)

The video recordings were not stored, instead only the collected traffic flow data was stored which in turn supports the data protection policies and also lesser storage space was required. The raw data was then extracted through the network (offline) and simultaneously the data from inductive loop detectors was also extracted.

4.2.3 Quality assessment

Time synchronization between two sensors was done carefully and the absolute percentage error values for every one-hour data was calculated. With the help of MCME, the traffic data was extracted from three videos and considered as the ground data. Then the traffic flow from inductive loop detectors were compared with the MCME values. Table 4.1 indicates that the inductive loops gave accurate traffic flow values except 3 detectors (D5.1, D7, D3). Those detectors with three error values were also with lesser error values (<5%). Therefore, the traffic flow data from ILD being very close to MCME values, they were considered as original traffic flow (ground truth) for comparing the accuracy of data collected from vehicle counters.

Table 4.1 Percentage error values of ILD data

Source: (Pragalathan & Schramm, 2019)

Date	Time(hrs)		Duration	Approach No.	Lane / Detector No.	MCME	Loop data	Percentage Error (%)
	From	To	(minutes)			counts in numbers	counts in numbers	
05.03.2019	15:00	15:10	10	3	right / D5.1	25	24	4.00
				3	straight / D5.2	49	49	0.00
				4	straight/right / D3	56	56	0.00
				4	left / D4	14	14	0.00
05.03.2019	15:13	15:23	10	2	straight/right / D7	60	61	1.67
				2	left / D8	5	5	0.00
				1	straight / D1.2	44	44	0.00
29.04.2019	10:25	10:40	15	3	straight / D5.2	75	75	0.00
				3	left / D6	16	16	0.00
				4	straight/right / D3	74	75	1.35
				4	left / D4	18	18	0.00

The phase one analysis was mainly focused on for short term data collection. Whereas Phase 2, aimed at comparing the data for a longer duration. Data collection with manual counting is very much tedious and needs more manpower. For our case, the traffic data is needed for continuous and long-term data duration. This objective can only be persuaded with video detection or inductive loop detectors. But there is need to find which is more accurate for urban scenarios. Hence, the phase 2 analysis focused on these two collection techniques specifically. Table 4.2, Table 4.3 and Table 4.4 tabulate the error value for every one hour vehicle counts where in the column headings VC stands for Vehicle Counter (video detection) and ILD for Inductive Loop Detectors.

The error values were calculated for everyday in a week and for each lane separately. The time value in each row represents the vehicle count that were collected in the previous one hour. The error values are indicated with green-yellow-red colour scale variations, where green: low error and red: high error. The results indicates that error values vary over period of time in a day. For example, the increased traffic flow during peak hours or during daytime are captured more accurately than the nighttime traffic. The influence of sunlight and the shadows of nearby lane vehicles also affected the accuracy of video detection technique. Another noticeable observation is that the error values are higher during nighttime (after 9:00 pm) with considerable number of vehicles. Thus, the study evidently proved that for long-term traffic data collection for traffic predictions over day and nighttime, inductive loop detectors are more likely suitable and accurate. Thus, the further work in this thesis will use the data from inductive loop detectors for traffic predictions.

Table 4.2 Absolute error values for right turning lane

Source: (Pragalathan & Schramm, 2019)

Time (hrs)	Vehicle counts and Absolute Error values for D5.1:Approach 3 Right Turning																				
	02/04/2019			03/04/2019			04/04/2019			05/04/2019			06/04/2019			07/04/2019			08/04/2019		
	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error
	(counts in numbers)			(counts in numbers)			(counts in numbers)			(counts in numbers)			(counts in numbers)			(counts in numbers)			(counts in numbers)		
1.00	3	3	0	5	2	3	9	4	5	9	6	3	32	17	15	33	15	18	7	4	3
2.00	5	3	2	7	3	4	8	3	5	5	1	4	21	10	11	17	9	8	5	3	2
3.00	3	1	2	0	0	0	3	1	2	5	1	4	2	2	0	12	5	7	0	0	0
4.00	1	1	0	4	3	1	5	3	2	5	3	2	9	5	4	14	9	5	4	3	1
5.00	6	3	3	6	2	4	7	4	3	7	2	5	8	5	3	4	2	2	8	4	4
6.00	36	17	19	37	17	20	40	21	19	36	20	16	12	6	6	6	3	3	36	17	19
7.00	83	49	34	69	38	31	73	41	32	82	52	30	22	12	10	19	10	9	90	54	36
8.00	88	82	6	84	82	2	105	100	5	116	104	12	15	13	2	4	5	1	92	97	5
9.00	123	112	11	98	108	10	96	103	7	89	93	4	25	25	0	19	18	1	88	92	4
10.00	82	80	2	72	67	5	72	75	3	102	110	8	73	72	1	37	37	0	68	106	38
11.00	61	62	1	92	91	1	71	68	3	71	68	3	89	85	4	50	45	5	80	72	8
12.00	76	77	1	80	77	3	72	77	5	86	85	1	90	92	2	90	74	16	61	80	19
13.00	71	73	2	93	93	0	86	83	3	91	92	1	88	85	3	102	101	1	88	82	6
14.00	77	76	1	89	85	4	77	77	0	98	111	13	68	67	1	114	110	4	90	85	5
15.00	113	110	3	92	92	0	100	103	3	181	177	4	91	88	3	100	93	7	132	128	4
16.00	167	181	14	137	139	2	143	155	12	156	157	1	133	131	2	97	108	11	160	151	9
17.00	207	200	7	195	191	4	235	224	11	156	155	1	90	87	3	106	104	2	149	141	8
18.00	155	169	14	148	181	33	172	187	15	133	137	4	90	90	0	89	93	4	169	161	8
19.00	105	113	8	106	130	24	113	119	6	107	106	1	87	92	5	119	108	11	123	119	4
20.00	66	66	0	104	94	10	99	95	4	89	84	5	59	61	2	71	69	2	73	71	2
21.00	74	58	16	74	51	23	70	53	17	77	44	33	47	35	12	56	42	14	62	45	17
22.00	76	37	39	47	23	24	84	41	43	78	41	37	80	38	42	42	19	23	96	42	54
23.00	53	26	27	42	22	20	37	22	15	68	36	32	71	33	38	37	17	20	33	15	18
24.00	11	6	5	16	10	6	22	12	10	60	25	35	38	19	19	11	7	4	18	8	10

Table 4.3 Absolute error values for straight lane

Source: (Pragalathan & Schramm, 2019)

Time (hrs)	Vehicle counts and Absolute Error values for D5.2 :Approach 3 Straight																				
	02/04/2019			03/04/2019			04/04/2019			05/04/2019			06/04/2019			07/04/2019			08/04/2019		
	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error
	(counts in numbers)			(counts in numbers)			(counts in numbers)			(counts in numbers)			(counts in numbers)			(counts in numbers)			(counts in numbers)		
1.00	55	33	22	53	24	29	29	25	4	35	26	9	116	80	36	174	111	63	64	39	25
2.00	33	20	13	25	14	11	21	18	3	30	20	10	76	52	24	121	78	43	27	19	8
3.00	17	12	5	17	13	4	18	14	4	10	8	2	51	36	15	77	46	31	20	13	7
4.00	13	9	4	6	6	0	14	8	6	14	8	6	24	16	8	48	32	16	6	5	1
5.00	31	21	10	19	13	6	27	17	10	18	14	4	11	7	4	35	22	13	30	17	13
6.00	127	75	52	113	77	36	134	78	56	111	63	48	40	23	17	30	19	11	105	62	43
7.00	344	221	123	302	205	97	329	212	117	323	208	115	33	26	7	22	17	5	288	211	77
8.00	455	435	20	434	426	8	450	436	14	430	422	8	93	91	2	29	27	2	441	414	27
9.00	471	453	18	459	455	4	454	449	5	436	413	23	126	121	5	52	48	4	438	427	11
10.00	305	280	25	313	303	10	312	311	1	335	292	43	172	184	12	90	78	12	362	321	41
11.00	306	266	40	278	274	4	262	260	2	257	247	10	222	223	1	174	130	44	316	285	31
12.00	286	246	40	254	246	8	233	226	7	249	251	2	242	236	6	227	176	51	270	239	31
13.00	227	223	4	219	221	2	229	225	4	237	235	2	272	267	5	258	181	77	343	272	71
14.00	229	228	1	251	246	5	262	260	2	272	271	1	248	242	6	254	204	50	269	232	37
15.00	275	270	5	275	272	3	303	298	5	314	309	5	260	255	5	288	241	47	286	283	3
16.00	361	362	1	291	291	0	346	345	1	354	348	6	257	246	11	272	245	27	401	379	22
17.00	406	403	3	369	365	4	378	375	3	363	363	0	231	224	7	327	287	40	542	405	137
18.00	518	404	114	362	360	2	398	387	11	332	330	2	278	258	20	380	280	100	504	367	137
19.00	334	313	21	320	322	2	304	304	0	250	249	1	250	235	15	326	260	66	285	284	1
20.00	249	226	23	203	200	3	226	223	3	218	215	3	204	191	13	197	188	9	213	208	5
21.00	199	151	48	242	161	81	232	178	54	233	151	82	201	140	61	202	157	45	213	157	56
22.00	216	129	87	201	111	90	252	145	107	223	128	95	233	141	92	251	144	107	253	144	109
23.00	194	107	87	174	101	73	169	103	66	190	119	71	230	138	92	134	93	41	173	101	72
24.00	68	47	21	94	63	31	90	63	27	183	109	74	186	107	79	98	64	34	86	57	29

Table 4.4 Absolute error values for left turning lane

Source: (Pragalathan & Schramm, 2019)

Time (hrs)	Vehicle counts and Absolute error vlaues for D8 :Approach 2 Left Turning																				
	02/04/2019			03/04/2019			04/04/2019			05/04/2019			06/04/2019			07/04/2019			08/04/2019		
	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error	VC	ILD	Error
	(counts in numbers)			(counts in numbers)			(counts in numbers)			(counts in numbers)			(counts in numbers)			(counts in numbers)			(counts in numbers)		
1.00	9	4	5	3	5	2	8	6	2	13	7	6	34	17	17	54	29	25	21	10	11
2.00	5	3	2	5	4	1	13	7	6	18	10	8	20	11	9	39	21	18	4	2	2
3.00	1	1	0	3	3	0	4	3	1	3	3	0	14	10	4	19	13	6	4	3	1
4.00	1	0	1	1	1	0	1	0	1	4	0	4	19	9	10	17	11	6	1	1	0
5.00	13	6	7	9	4	5	8	4	4	13	6	7	15	9	6	20	13	7	12	5	7
6.00	49	18	31	41	18	23	45	13	32	34	12	22	25	8	17	12	7	5	39	10	29
7.00	44	24	20	38	14	24	53	20	33	65	39	26	28	9	19	14	8	6	45	19	26
8.00	72	51	21	75	64	11	71	68	3	66	54	12	29	22	7	14	10	4	73	56	17
9.00	84	73	11	77	66	11	76	70	6	103	94	9	37	32	5	24	21	3	79	72	7
10.00	77	68	9	74	63	11	82	65	17	78	72	6	65	57	8	32	27	5	102	67	35
11.00	89	79	10	91	74	17	83	74	9	90	76	14	79	74	5	44	37	7	97	80	17
12.00	93	69	24	91	77	14	107	91	16	78	77	1	81	89	8	55	49	6	96	80	16
13.00	104	87	17	85	70	15	100	91	9	99	87	12	115	111	4	60	51	9	102	90	12
14.00	75	73	2	80	71	9	89	79	10	101	82	19	82	80	2	66	56	10	101	83	18
15.00	102	87	15	85	84	1	86	71	15	111	96	15	90	85	5	56	42	14	81	71	10
16.00	84	72	12	104	89	15	80	75	5	92	73	19	113	112	1	55	62	7	81	85	4
17.00	87	74	13	83	68	15	94	87	7	76	81	5	95	88	7	68	55	13	98	87	11
18.00	89	83	6	95	85	10	76	68	8	98	102	4	69	77	8	59	50	9	87	91	4
19.00	78	68	10	80	73	7	96	94	2	93	88	5	75	69	6	63	59	4	100	80	20
20.00	64	60	4	182	76	106	79	68	11	82	84	2	74	64	10	65	61	4	73	66	7
21.00	63	43	20	156	51	105	80	58	22	87	61	26	75	54	21	68	49	19	79	60	19
22.00	67	42	25	114	35	79	70	37	33	86	48	38	84	48	36	65	36	29	89	43	46
23.00	32	29	3	95	29	66	57	32	25	45	26	19	31	38	7	60	27	33	65	36	29
24.00	32	23	9	29	7	22	38	19	19	55	27	28	0	23	23	23	12	11	36	17	19

4.3 PREDICTION MODELLING

As explained in the previous chapter in Section 2.2.2, the traffic flow will be predicted using two recent models (Fb-Prophet and NeuralProphet) and their performance will be compared with classical model named SARIMA. This section explains the methodology followed to estimate the urban traffic using these models and analyse the pros and cons of each model. Parts of results of this section were already published in the paper (Pragalathan & Schramm, Apr 2023).

4.3.1 Testing and training dataset

The traffic flow data from inductive loop detectors over a period of three months (February 2017 – May 2017) were utilised for training and testing the models. Forecasting models require sequence of data attached with corresponding timestamps as inputs. Excluding speed and other data from the available minute wise database, only the traffic flow along with timestamp including date and time has been extracted. The original data from inductive loop detectors were at one-minute intervals. Additionally, the data was collected lane by lane, which were then pre-processed by summing the values on all three lanes so that the traffic at overall road was found. The minute wise vehicle counts were actually derived for hourly traffic counts and then split into training and testing dataset (16/05/2017).

4.3.2 Model fitting and forecasting

SARIMA Model Identification.

The Box-Jenkins method (Box & Jenkins, 1976) was followed for the model formulation and order estimation of SARIMA model. Stationarity is one of the basic requirements of SARIMA model where the time series variable should have constant statistical properties i.e., mean-variance over the period. This means that the mean-variance of the time series is same over the training period of the model. Figure 4.5 depicts the stationarity of the current dataset considered for training the model. For better performance of the model, the traffic data should be collected and arranged at regular time intervals. In the collected traffic database, there were few missing values. Proper handling of such missed data is crucial for accuracy and reliability in time series forecasting. Missed data can be completed in 2 ways: either by dropping the complete day's data or imputing the missed values. It was found while modelling that the latter method performed better than the former one. Imputing can also be done in many ways by taking a) previous values b) the following values or c) mean values. Based on the dataset knowledge, it has been decided to use the previous week's same-day data. For example, the missing values in a Thursday can be filled by taking corresponding values from the previous Thursday. Hence by doing this, even the missed data during weekends can also be filled meaningfully.

Stationarity test or unit root test

The current traffic flow dataset did not have any increasing or decreasing pattern but rather showed seasonal behaviour over a 24-hour time duration. Simultaneously, the ADF test results showed that the p-values were always lesser than 0.05 i.e., it rejects the null hypothesis. As explained earlier, seasonal differencing must be done to eliminate non-stationarity.

The overall training dataset is depicted in Figure 4.5. Visual inspection of the time series plot confirmed that the available traffic flow database did not have any increasing or decreasing statistical properties i.e., it showed stationarity. If the whole dataset is non-stationary with increasing or decreasing trend over time, then it should be differenced one or more times accordingly to remove the trend. Augmented Dicky Fuller (ADF) test was also done to check stationarity. The verification of p-values from the ADF test results showed that p-values were always lesser than 0.05 and the test statistics were lesser than critical values. Therefore, the traffic flow data considered for modelling rejected the null hypothesis and confirmed stationarity.

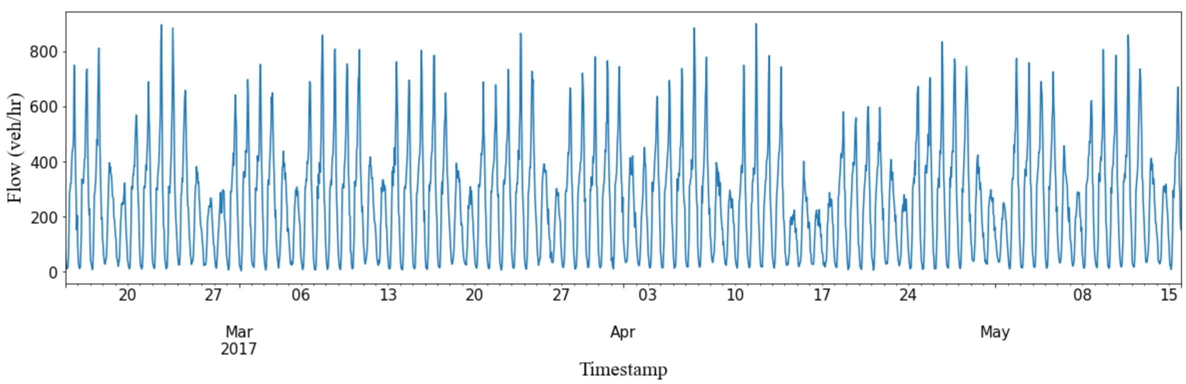

Figure 4.5. Time series plot for overall training period

Source: (Pragalathan & Schramm, Apr 2023)

Repetitive patterns were also observed in the visual inspection of Figure 4.5. This proves the seasonality of the database. The traffic flow values had a pattern that repeats every 24 hours. This property of repeating patterns is termed as seasonality. In this database, a 24-hour season was identified and then incorporated into the model by taking the ‘s’ value as 24 i.e., number of datapoints in a season. In a day with a repeating pattern, there are 24 hours, resulting in 24 data points within a single season..

Moving on to model fitting and forecasting of the SARIMA model, the main objective was to find the values for the parameters (both seasonal and non-seasonal) such as p,d,q,P,D,Q. For this purpose, ACF and PACF plots were used. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots are basically used to do Autocorrelation Analysis in a time series data. The correlation between a time series and a lagged version of itself is called autocorrelation. The ACF plot begins with a lag of zero, which is the time series' correlation with itself and gives it a correlation of one.

Since the non-seasonal differencing was not done due to stationarity, the value for d= 0 was taken. If the dataset has a strong seasonal pattern throughout the time period, then seasonal differencing should be done, and it is better to avoid more than one degree of seasonal differencing. At the same time, the sum of the seasonal and non-seasonal degree of differencing should not exceed 2 (Box & Jenkins, 1976; Rules for ARIMA, 2022). But because of seasonality, first-order seasonal differencing was done, thus the value for D=1 was taken. Figure 4.6 shows the seasonally differenced data plot. For finalising the values for other parameters, ACF and PACF plots should be drawn with this seasonally differenced database.

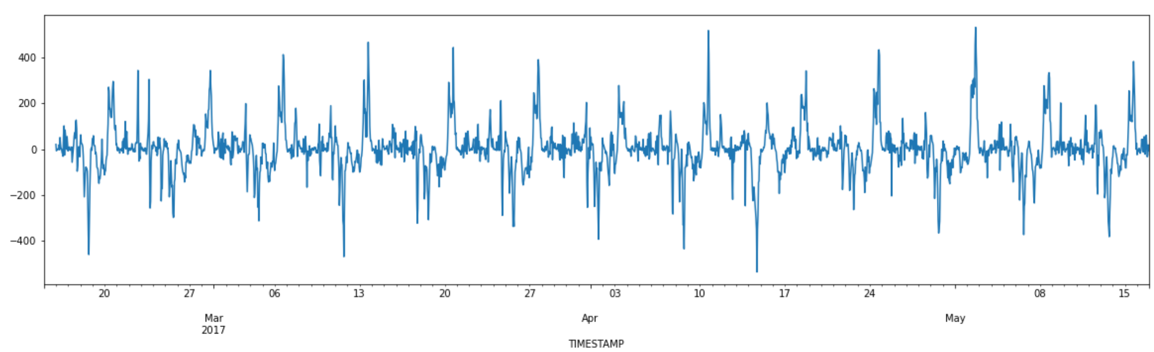

Figure 4.6 Time series plot of seasonally differenced data

The visual inspection of the ACF (Figure 4.7) and PACF (Figure 4.8) plots according to Box-Jenkins procedure (Forecasting homepage: rules for ARIMA, 2022), indicates the following points.

- Gradually decreasing pattern towards zero in ACF
- Sharp cut-off in PACF

If the autocorrelation function (ACF) of the differenced series exhibits a sharp cutoff and/or the lag-1 autocorrelation is negative, then it is the indication of adding a non-seasonal MA term to the model. In our case, the gradually decreasing pattern in ACF indicates that the model has no MA terms. Similarly, if the partial autocorrelation function (PACF) of the differenced series displays a sharp cutoff and/or the lag-1 autocorrelation is positive, then it is the indication of adding one or more AR terms to the model in the non-seasonal part. The lag beyond which the PACF cuts off is taken as the value for order of AR terms. By considering this rule, the order of AR can be found in the PACF plot as p value p = 1 or 2.

Moving on to seasonal part of the model parameters, if the series has a strong and consistent seasonal pattern, then it is necessary to include an order of seasonal differencing in the model. Else, the model would incorrectly assume that the seasonal pattern diminishes over time, leading to lesser accuracy. However, it is important to avoid more than one order of seasonal differencing or more than 2 orders of total differencing (seasonal + nonseasonal) for avoiding complexity.

Seasonal differentiation has already been carried out on this basis. Therefore, the value D=1 is chosen. If the autocorrelation of the appropriately differenced series is positive at lag s, where s is the number of periods in a season, then consider adding an SAR term to the model. If the autocorrelation of the differenced series is negative at lag s, consider adding an SMA term to the model. In our case, significant negative PACF values at corresponding lag 's' = 24 in Figure 4.8 indicated that there is no seasonal-AR (SAR) term, rather a Seasonal-MA terms (SMA) should be included in the model. Therefore, the order of SAR is P=0 and SMA is Q=1. The final representation of the model is given as $(1,0,0)\ (0,1,1)_{24}$. The order estimation of the model is followed by model training and then forecasting. The training dataset was fitted into the model and then the traffic flow values for the next 24 hours were forecasted and tabulated in Table 4.5.

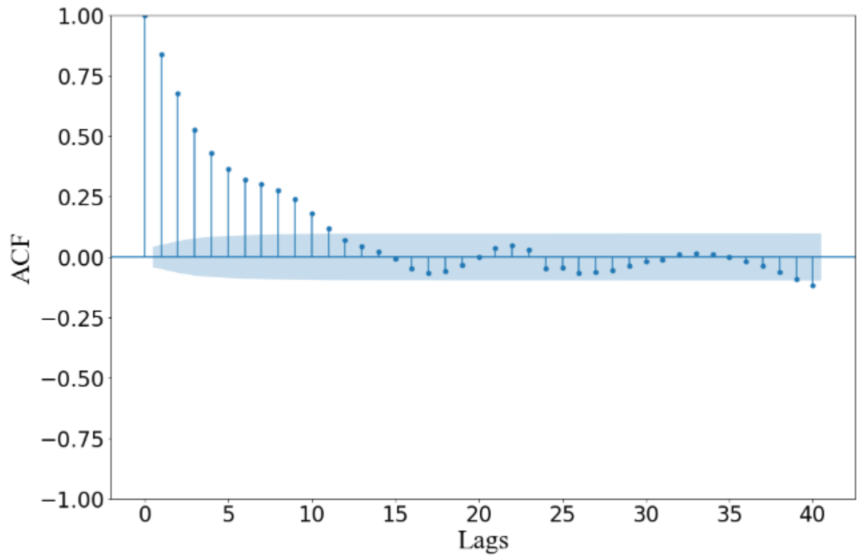

Figure 4.7. Autocorrelation Plot

Source: (Pragalathan & Schramm, Apr 2023)

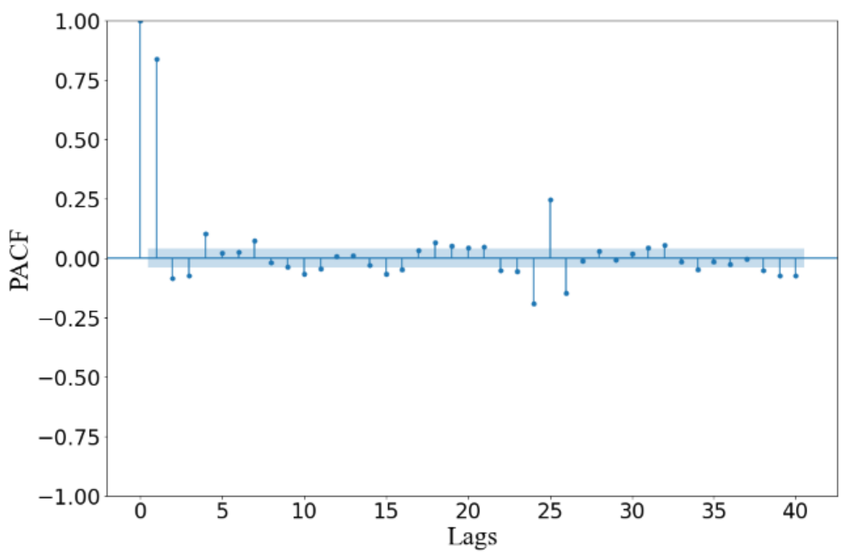

Figure 4.8. Partial Autocorrelation Plot

Source: (Pragalathan & Schramm, Apr 2023)

Fb-Prophet model

As explained in Section 2.2.4, Fb-Prophet model is based on an additive model where the effects of seasonality (yearly/monthly/weekly/daily) and also the effects of holidays and events can be incorporated in the model. This model is very much applicable to time series with strong seasonality. Thus, it is been chosen for traffic flow predictions in this thesis. This model can also handle missed data and outliers in a better way. Additionally, the model is an open-source prediction tool implemented in Python. Hence the available traffic database was used for training Fb-Prophet model. Initially, the installation of Fb-Prophet model was carried out in Anaconda using pip. Thus, all the libraries, functions and plots can be easily called for fitting the database into the model. The current training dataset (Feb 2017- May2017) was given as input for training.

One of the specific requirements of the Fb-Prophet model is that the data given for training should have only two columns named as ds and y. Here ds represents the column with timestamp in datetime datatype and in a specific date format: YYYY-MM-DD HH:MM:SS. Another requirement of this model is that the given data should be of numeric datatype. Thus, after completing the initial data processing, the next step of model fitting should be done. After importing the data in specific format, the complete dataset was split into training and testing dataset (16/05/2017). Thus, if needed certain intuitive parameters for changepoints, seasonality and so on can be defined after visual inspection of the time series plot. Equation 2.3 showed the components of the Fb-Prophet model, based on which the parameters are decided automatically for initial training. Thus, after model initialisation and model fitting, the traffic data was forecasted for future testing period.

According to the accuracy and trend capture, the components can be adjusted, added or neglected. In our case, forecasting the following 24-hour traffic data was not within adequate precision. However, it captured the daily/weekly seasonal patterns very well. Therefore, some of the default parameters were modified and the changes in the forecasting was also verified. Finally, default values of parameters: weekly seasonality = 3 and changepoint prior scale = 0.05 , were modified as weekly seasonality = 6 and changepoint prior scale = 0.001. The strong seasonal pattern indicated the growth parameter as ‘flat’ (Prophet: Additional Topics, 2022). The final forecasting results for the testing period of 24 hours is given in Table 4.5.

NeuralProphet model

Being a successor of Fb-Prophet model, most of the procedures were same for NeuralProphet model for fitting and forecasting the time series data. Similar to previous model, NeuralProphet model was also installed in Anaconda and all the required libraries, functions and plots were imported. The traffic database was imported which was then followed by required initial data processing such as having only two columns, having column heads as ds and y, datatype as datetime and numeric and finally changing the timestamp format. Moving on to model training, the data was given as input for the model to read and capture the trends, seasonality and many other components as mentioned in the Equation 2.4. The extra components like non-linear deep layers, autoregression and covariates can also be defined whenever needed after initial model fitting. Initial fitting was carried out with automatic components definitions with default parameters. Like Fb-Prophet model, the NeuralProphet model also captured the seasonal patterns of the traffic data. But to increase more accuracy, autoregressive component was included in the model with a parameter “n_lags” with values 2*24 (after many trials) i.e., the model took two previous days data for autoregression. Finally, the forecasting for testing period was done and the model was evaluated with MAPE values, and the error values are given in Table 4.5.

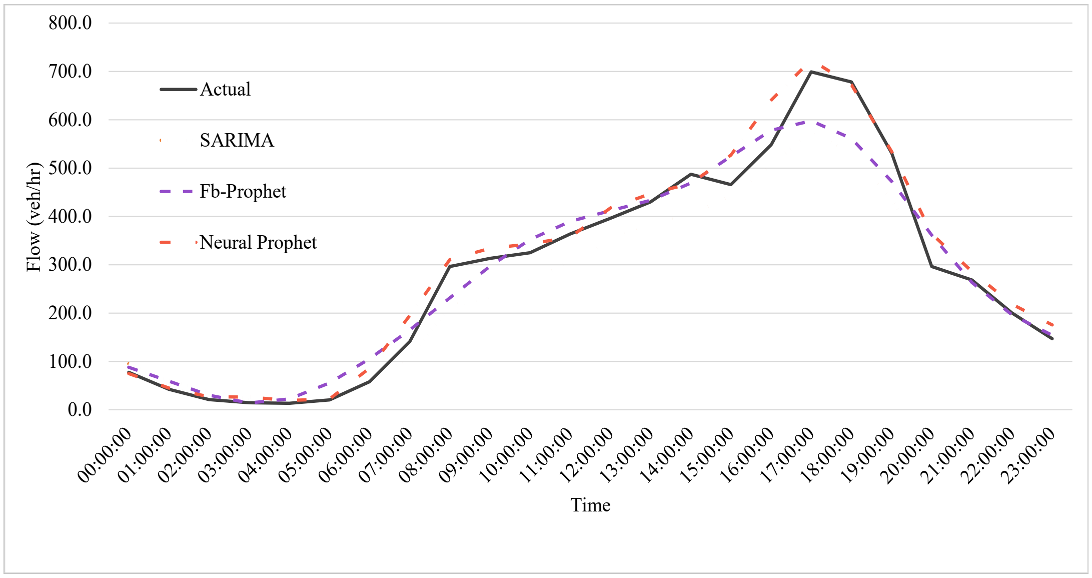

Figure 4.9. Comparison of actual and predicted traffic flow

Source: (Pragalathan & Schramm, Apr 2023)

4.3.3 Model evaluation

Model evaluation was done to compare the performance of all the three models: SARIMA, Fb-Prophet and NeuralProphet Model. The forecasted values for the testing period of 24 hours are given in the in Table 4.5. The accuracy was measured in terms of Mean Absolute Percentage Error (MAPE). To be more comprehensive, other metrics such as MAE and RMSE for model accuracy were also calculated and are given in Table 4.6. The graph shows the comparison plots of actual and predicted traffic flow calculated from all the three models. The plot highlights how the efficiency of the models changes over a day in Figure 4.9. In order to find the computational costs of the models, the running time required for the model to fit, and forecast was also recorded and given in Table 4.7.

Table 4.5 MAPE values for SARIMA, Fb-Prophet and NeuralProphet models

Timestamp (16/05/2017)	Actual (veh/hr)	Predictions (veh/hr)			Absolute Percentage Error (APE) (%)		
		SARIMA	Fb-Prophet	Neural-Prophet	SARIMA	Fb-Prophet	Neural-Prophet
00:00:00	77	95	88	75	23.4	14.3	3.0
01:00:00	42	59	59	45	40.5	40.5	6.0
02:00:00	21	38	30	27	81.0	42.9	28.1
03:00:00	14	23	14	26	64.3	0.0	84.6
04:00:00	13	16	22	19	23.1	69.2	47.3
05:00:00	20	21	55	22	5.0	175.0	9.7
06:00:00	58	56	105	86	3.4	81.0	47.6
07:00:00	141	106	165	195	24.8	17.0	38.1
08:00:00	296	215	231	311	27.4	22.0	4.9
09:00:00	313	232	297	334	25.9	5.1	6.6
10:00:00	325	270	352	343	16.9	8.3	5.4
11:00:00	364	305	390	357	16.2	7.1	1.8
12:00:00	396	352	412	418	11.1	4.0	5.6
13:00:00	430	384	433	447	10.7	0.7	3.9
14:00:00	487	405	469	468	16.8	3.7	3.8
15:00:00	466	442	524	527	5.2	12.4	13.0
16:00:00	548	504	578	640	8.0	5.5	16.9
17:00:00	699	577	598	724	17.5	14.4	3.5
18:00:00	678	531	561	672	21.7	17.3	0.9
19:00:00	531	434	472	534	18.3	11.1	0.6
20:00:00	296	310	361	364	4.7	22.0	22.9
21:00:00	268	252	263	286	6.0	1.9	6.6
22:00:00	200	193	196	218	3.5	2.0	9.1
23:00:00	147	162	154	175	10.2	4.8	19.1
MAPE (%)					20.2	24.3	16.2
MAPE (%) Daytime (06:00 - 22:00)					14.7	9.7	9.0

Table 4.6 Model evaluation metrics

Source: (Pragalathan & Schramm, Apr 2023)

Metrics	SARIMA	Fb-Prophet	NeuralProphet
MAPE (%)	20.2	24.3	**16.2**
MAE (veh/hr)	43	32	**23**
RMSE (veh/hr)	58	44	**32**

The significance of manual order estimation and model definition in SARIMA modelling was confirmed with the results in Table 4.5. The predictions with all the three models are shown in Figure 4.9 and Table 4.5 which confirms that the error values are within the acceptable range. As per the Lewis' scale of interpretation (Kenneth & Ronald, 1982) for evaluating accuracy, the MAPE values in Table 4.5 indicated that both the models (Fb-Prophet and NeuralProphet) gave highly accurate predictions with error value lesser than 10% during daytime and overall 24- hour predictions resulted in reasonable error value (11% - 20%).

The MAE values in Table 4.6 revealed that the NeuralProphet model performed better with minimum MAE (23 veh/hr) which is also lesser than 10% of the average traffic flow value (285 veh/hr). In similar way, RMSE values in Table 4.6 also proved the better performance of NeuralProphet model with minimum error value (32 veh/hr). The range of traffic flow of that particular day is 686 veh/hr which is the difference between minimum (13 veh/hr) and maximum (699 veh/hr). Thus, the normalised RMSE values of Fb-Prophet is 0.06 and for NeuralProphet is 0.04 which again confirmed the models' performances. Because the normalized RMSE value for NeuralProphet is the lowest, it implies that it predicts with better accuracy. Subsequently the MAPE, MAE and RMSE values of the NeuralProphet model are the lowest, the best model in terms of accuracy is the NeuralProphet.

The efficacy of the models were also studied by comparing the computational time for model fitting and forecasting and the results are given in Table 4.7. Although the

NeuralProphet model provided better accuracy, it took more time for computation especially for model fitting. This confirms that for achieving more accurate results, the complexity and time consumption of the model are also increased.

Table 4.7 Computational time for model fitting and forecasting

Source: (Pragalathan & Schramm, Apr 2023)

Models	Running time in seconds
SARIMA	5
Fb-Prophet	**3**
Neural-Prophet	98

4.3.4 Conclusion

The comparative study has emphasized the urban traffic flow predictions with SARIMA, Fb-Prophet and NeuralProphet models. Other than accuracy, the model used for traffic predictions should be simple, fast and automated to make it more ideal for urban traffic scenarios. The effect of peak/off-peak hour traffic should also be captured which is repeatedly noticed in the urban traffic conditions. The prediction model should also be expandable for bigger network and should be faster at the same time. Thus, the results given in the previous subsection proved the efficiency of both Fb-Prophet and NeuralProphet model for forecasting a busy urban traffic. Both the models gave optimum accuracy within acceptable range and additionally they were fast and automated. One can prefer to Fb-Prophet model for quickest predictions and optimum accuracy. On the other side, NeuralProphet model can be preferred for maximum accuracy. Both the models have options to handle the effect of holidays and weather which will be discussed in the following Section 4.4.

4.4 PREDICTION WITH EXOGENOUS VARIABLES

Predictability of data with regularity or certainty is much higher than the data with uncertainties. Estimation of such uncertain predictions also helps to plan and prepare for uncertain scenarios. Urban road traffic is influenced by various spatio-temporal factors and external influences. Predictions with inclusion of these external factors are also used for simulations where the future traffic during those uncertain situations like severe weather conditions can be replicated in virtual scenarios. This section deals with traffic flow predictions with the effect of weather (rainy) and holidays (Pragalathan & Schramm, 2024). Numerous research works has been carried out to find the impact of exogenous factors on traffic flow parameters. A number of external factors influences the traffic flow forecasting. Some of them are effect of working day/holiday, seasons, weather. The accuracy of the database used for predictions also affects the models' performances.

The effect of weather is very much predominant over road traffic which indirectly influences various parameters like speed, flow, level of service, capacity and so on. While implementing various traffic management systems, traffic demand should be considered along with the weather data for more effectiveness (Calvert & Snelder, 2016). A profound analysis was carried to find the influence of factors like rain, snow, temperature and wind over road traffic demand. It was found that the capacity and demand of road traffic was reduced due to rain. Another study (Xu, He, Sha, Zhuang, & Sun, 2013) specifically examined the effect of rainfall over traffic parameters (flow, speed & density) and the evening peak hour road traffic was found more influenced because of rainfall.

The rainfall was also quantified into three different classes (light, moderate and heavy) and based on that a study (Yuan-qing & Jing, 2017) was conducted to find their impact on traffic parameters. The results of the study proved that the maximum flow rate and free flow speed were reduced because of adverse weather (rainfall). It was also found traffic flow had more impact due to rainfall than speed. The maximum flow rate during moderate rainfall (5-10 mm/hr) and heavy rainfall (>10 mm/hr) was found decreased even up to 19% and 33% respectively when comparing to normal weather. The research works carried out by (Calvert & Snelder, 2016; Xu, He, Sha, Zhuang, & Sun, 2013; Akin, Sisiopiku, & Skabardonis, 2011) to obtain the influence of weather on traffic flow revealed the significance of weather while considering traffic.

Similar research work (Akin, Sisiopiku, & Skabardonis, 2011) also analysed the impact of weather on traffic flow of an urban freeway. It was found that the rainfall reduced the capacity by 7-8% and light snow impacted demand leading to a significant reduction in traffic volume. The user's decision for postponing or cancelling a trip, selection of travel mode and route directly or indirectly influences the urban traffic flow which should be included in the prediction model. Even though weather influenced the urban traffic flow significantly, very few research works considered it as an exogenous factor for predictions (Lana, Ser, Velez, & Vlahogianni, 2018). Adverse weather conditions can also shift peak hour if the drivers chose to leave early or later.

Next to weather, calendar information was found very influential over urban traffic flow. Hence considering the information like day of week, holidays and so on during urban road traffic predictions increase the accuracy (Laña, Ser, & Olabarrieta, 2016). Some of the literatures (Liu, Li, Xi, & Tang, 2015; Su, Dong, Jia, Qin, & Tian, 2016) considered exogenous variables for traffic predictions by only considering calendar information and by neglecting the impact of weather. The research work done by (Oh, Kim, & Hong, 2015) tried to include rainfall for traffic prediction modelling. Inclusion of both weather and holidays into prediction model was done by very few works (Zhang, Yao, Du, & Ye, 2021; Al-Selwi, Aziz, Abas, Hamzah, & Mahmud, 2022).

4.4.1 NeuralProphet model with external factors

NeuralProphet model being the successor of Fb-Prophet model, gives more options for traffic flow predictions to include the effect of holidays and events and uncertainties. Hence in this thesis, NeuralProphet model was used to forecast the urban traffic flow with effect of rainy days and holidays. The basic model information and the mathematical equation (Equation 2.4) are already explained in the Section 2.2.5. The observations in the previous section found that NeuralProphet predicts with maximum accuracy. Moreover, NeuralProphet model is also capable of handling holidays and events into the predictions. Knowing the significance of such exogenous variables, it is a necessary to forecast the urban road traffic by using NeuralProphet model.

The research work (Becker, Rust, & Ulbrich, 2022) based on meteorological variables for traffic predictions confirmed the improvements in the accuracy. The mean squared error was reduced by up to 60% in the analysis conducted for different vehicle types, because of the inclusion of weather data like precipitation, temperature, cloud cover, and wind speed data. Acknowledging the importance of traffic predictions with the inclusion of external data, this current thesis primarily focused on evaluating the performance of the NeuralProphet model for urban traffic predictions. The study considered factors such as holidays and precipitation data to enhance the accuracy and robustness of the traffic forecasting model. Therefore, the fluctuations in traffic flow patterns will be analysed considering the influence of holidays and weather conditions, specifically precipitation.

The presence of precise historical traffic data significantly influences traffic predictions. Accurate past traffic information is essential for developing reliable Traffic Flow Prediction Models (TFPMs), leading to better traffic management strategies and enhanced transportation efficiency. The traffic data collected from the test bed over the period of 2017 to 2019, were utilized to train the prediction models for traffic flow forecasting. Traffic flow estimation is also possible with limited historical data collected few days (Pragalathan & Schramm, Apr 2023; ChikkaKrishna, Rachakonda, & Tallam, 2022).

In this thesis, a huge database was employed to incorporate the impact of weather and holidays on traffic predictions. The extensive dataset allowed for a comprehensive analysis of how these external factors influence traffic patterns, enabling a more accurate and detailed evaluation of traffic flow dynamics under varying conditions. Simultaneously the data for local/regional holidays were also acquired (List of Holidays in NRW, 2019) and tabulated in Table 4.8. Climate data was collected by Climate Data Center (CDC) of the Deutscher Wetterdienst (DWD) at 400 climate stations and gave free access to historical meteorological data (Climate data for direct download, 2020).

Table 4.8 List of holidays in North Rhein Westphalia, Germany
Source: (Pragalathan & Schramm, 2024)

Holidays	Dates		
New Year's Day	01/01/2017	01/01/2018	01/01/2019
Good Friday	14/04/2017	30/03/2018	19/04/2019
Easter Monday	17/04/2017	02/04/2018	22/04/2019
May Day	01/05/2017	01/05/2018	01/05/2019
Ascension Day	25/05/2017	10/05/2018	30/05/2019
White Monday	05/06/2017	21/05/2018	10/06/2019
Day of German Unity	03/10/2017	03/10/2018	03/10/2019
Christmas Day	25/12/2017	25/12/2018	25/12/2019
St.Stephan's Day	26/12/2017	26/12/2018	26/12/2019

Several parameters like humidity, temperature, precipitation, wind speed and so on were collected and given at different time intervals (10-minute, hourly, daily, and monthly). In this study, both traffic data and weather data (precipitation) were utilized at hourly intervals for prediction modelling. Specifically, the modelling considered hourly precipitation events with maximum severity (greater than 5mm), and the corresponding dates for these events are listed in Table 4.9. By incorporating both traffic and weather data at a detailed hourly level, the research aimed to develop more accurate and insightful traffic flow prediction models, capturing the impact of weather events on traffic conditions.

The NeuralProphet model excels in capturing the impact of holidays, events, or other exogenous factors when forecasting time series data, even at higher sub-daily frequencies. Its adaptability and flexibility allow for effective integration of external influences that may affect the time series, resulting in more accurate and dependable predictions. By considering exogenous variables such as holidays and events, the

NeuralProphet model produces forecasts that accurately reflect specific patterns and variations in the data, leading to improved traffic flow predictions, particularly when these external factors significantly influence the outcomes.

Table 4.9 List of days with heavy rainfall

Source: (Pragalathan & Schramm, 2024)

List of days with severe precipitation (>5mm/hr)		
22/02/2017 19:00	05/08/2017 08:00	08/05/2019 17:00
16/04/2017 12:00	18/08/2017 20:00	10/06/2019 16:00
12/05/2017 21:00	30/09/2017 02:00	11/07/2019 15:00
15/06/2017 15:00	30/09/2017 11:00	20/07/2019 21:00
15/06/2017 16:00	16/05/2018 15:00	27/07/2019 16:00
28/06/2017 13:00	29/08/2018 22:00	02/08/2019 19:00
12/07/2017 12:00	29/08/2018 23:00	29/08/2019 04:00
14/07/2017 16:00	30/10/2018 03:00	29/09/2019 19:00
20/07/2017 05:00	08/12/2018 21:00	01/10/2019 08:00
23/07/2017 00:00	02/05/2019 16:00	

The components of the NeuralProphet model are auto-regression, trend, seasonality, lagged regression, future regression, and events. As an open-source model, it provides the flexibility to define a list of days to include holidays and events specific to a particular country or region. The model's open-source nature empowers researchers and practitioners to adapt and enhance it to suit the needs of their specific geographical area or application, making it a versatile and valuable tool for traffic prediction tasks. The traffic data obtained from inductive loop detectors were divided into two separate datasets: a training dataset and a testing dataset. The training dataset was utilized for model fitting and calibration, while the testing dataset was used for forecasting and evaluating the model's performance.

The inherent capabilities of the model allowed it to determine suitable parameters on its own, reducing the need for manual fine-tuning and streamlining the modelling process. This automation not only saves time and effort but also enhances the efficiency and accuracy of the traffic flow predictions (NeuralProphet, 2021). Following the calibration process, the model successfully captured the repetitive patterns of everyday traffic flow during morning and evening peak hours, as demonstrated in Figure 4.10 and also the pattern repeated over a week as shown in Figure 4.11. Figure 4.12 shows how the impact of holidays are repetitive with both positive and negative coefficients over every year.

The impacts are significantly noticed by comparing the list of holidays (Table 4.8) and rainfall (Table 4.9) and the plot Figure 4.12 at corresponding dates. Moreover, the plot clearly illustrates the adverse impact of hourly precipitation on traffic flows, validating the findings of previous research works. Thus, the estimation of future 24 hours (01/10/2019) traffic data was carried out by using NeuralProphet model with the incorporation of weather and holiday data. The model also demonstrated its capability to capture the influence of holidays and weather over the previous time span, resulting in increased accuracy of overall predictions.

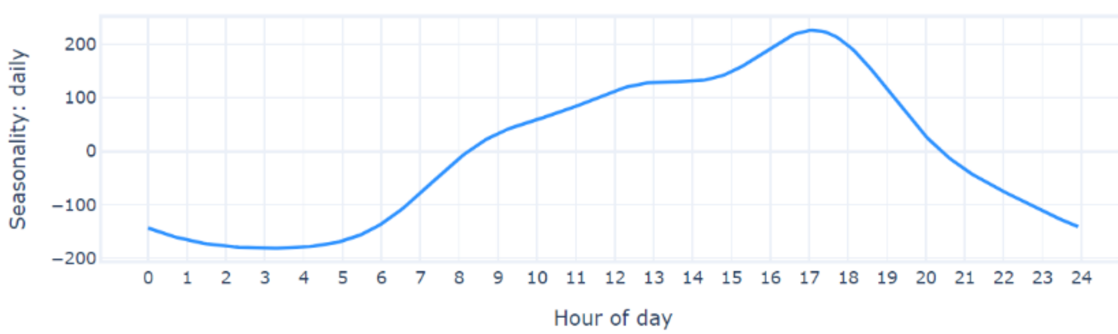

Figure 4.10 Daily seasonality captured by NeuralProphet model

Source: (Pragalathan & Schramm, 2024)

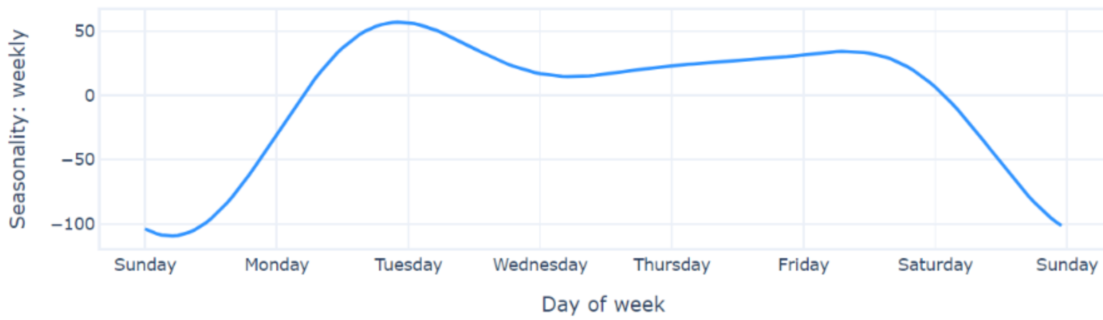

Figure 4.11 Weekly seasonality captured by NeuralProphet model

Source: (Pragalathan & Schramm, 2024)

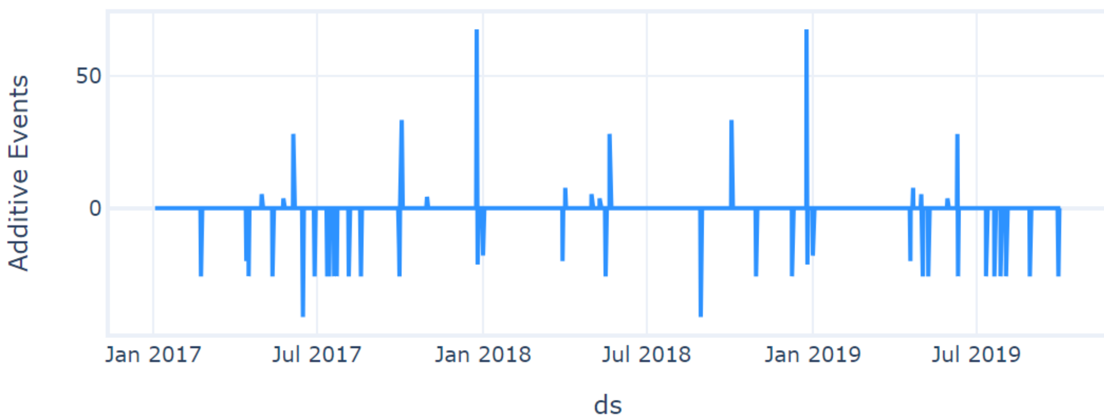

Figure 4.12 Effect of holidays and rainfall captured by NeuralProphet model

Source: (Pragalathan & Schramm, 2024)

4.4.2 Results and discussions

Traffic flow prediction models offer forecasts of traffic volume or flow, indicating the number of vehicles passing through a specific location within a given unit of time, for the subsequent 24 hours. Table 4.10 and Table 4.11 give the error values calculated from the actual traffic and the predicted values without and with external factors (holidays/weather) respectively. The evaluation metrics, including Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), were tabulated to assess the efficiency and accuracy of the model.

Table 4.10 Calculation of error values for predictions without holiday and rainfall data

Timestamp	Actual	Predicted	Absolute Percentage error	Absolute error	Squared errors
01/10/2019 00:00:00	72	52	27.95	20.13	405.05
01/10/2019 01:00:00	45	25	43.68	19.66	386.40
01/10/2019 02:00:00	22	24	9.01	1.98	3.93
01/10/2019 03:00:00	19	23	22.58	4.29	18.41
01/10/2019 04:00:00	22	14	37.56	8.26	68.29
01/10/2019 05:00:00	27	6	78.20	21.11	445.83
01/10/2019 06:00:00	67	55	18.18	12.18	148.29
01/10/2019 07:00:00	134	145	8.30	11.12	123.75
01/10/2019 08:00:00	283	255	10.01	28.32	802.09
01/10/2019 09:00:00	289	293	1.24	3.57	12.75
01/10/2019 10:00:00	328	292	10.87	35.66	1271.36
01/10/2019 11:00:00	315	321	1.76	5.55	30.82
01/10/2019 12:00:00	356	381	7.04	25.06	628.18
01/10/2019 13:00:00	443	428	3.41	15.09	227.80
01/10/2019 14:00:00	418	461	10.36	43.29	1874.17
01/10/2019 15:00:00	527	516	2.02	10.64	113.11
01/10/2019 16:00:00	630	611	3.08	19.37	375.36
01/10/2019 17:00:00	730	700	4.06	29.62	877.28
01/10/2019 18:00:00	728	632	13.22	96.23	9260.41
01/10/2019 19:00:00	470	485	3.26	15.31	234.29
01/10/2019 20:00:00	347	339	2.19	7.60	57.78
01/10/2019 21:00:00	285	261	8.47	24.14	582.85
01/10/2019 22:00:00	209	209	0.22	0.47	0.22
01/10/2019 23:00:00	148	167	12.57	18.60	346.11
Mean or rooted mean*	288.08	278.93	14.13	19.89	27.61*
			MAPE	MAE	RMSE

Table 4.11 Calculation of error values for predictions with holiday and rainfall data

Timestamp		Actual	Predicted	Absolute Percentage error	Absolute error	Squared errors
01/10/2019	00:00:00	72	52	27.81	20.02	400.97
01/10/2019	01:00:00	45	26	42.13	18.96	359.35
01/10/2019	02:00:00	22	25	13.97	3.07	9.44
01/10/2019	03:00:00	19	25	29.52	5.61	31.47
01/10/2019	04:00:00	22	15	30.22	6.65	44.20
01/10/2019	05:00:00	27	8	71.83	19.39	376.14
01/10/2019	06:00:00	67	56	16.80	11.26	126.77
01/10/2019	07:00:00	134	146	8.75	11.72	137.42
01/10/2019	08:00:00	283	255	9.89	27.98	783.02
01/10/2019	09:00:00	289	293	1.29	3.74	13.97
01/10/2019	10:00:00	328	293	10.70	35.10	1231.98
01/10/2019	11:00:00	315	321	1.94	6.10	37.17
01/10/2019	12:00:00	356	382	7.22	25.69	660.06
01/10/2019	13:00:00	443	429	3.18	14.09	198.40
01/10/2019	14:00:00	418	462	10.49	43.83	1920.93
01/10/2019	15:00:00	527	517	1.97	10.36	107.39
01/10/2019	16:00:00	630	611	2.99	18.83	354.39
01/10/2019	17:00:00	730	700	4.04	29.51	870.69
01/10/2019	18:00:00	728	632	13.22	96.26	9266.53
01/10/2019	19:00:00	470	485	3.09	14.54	211.28
01/10/2019	20:00:00	347	339	2.25	7.80	60.78
01/10/2019	21:00:00	285	261	8.38	23.87	569.78
01/10/2019	22:00:00	209	209	0.04	0.08	0.01
01/10/2019	23:00:00	148	167	12.85	19.03	361.96
Mean or rooted mean*		288.08	279.47	13.94	19.73	27.49*
				MAPE	MAE	RMSE

Table 4.12 Evaluation metrics

Source: (Pragalathan & Schramm, 2024)

Metrics	Error Values (%)	Interpretations
MAPE	13.94	10%-20%
MAE	19.73	< 10% of 288 veh/hr(average flow)
RMSE	27.49	Rooted RMSE = 0.04, very much nearby 0

The results and analysis presented in Table 4.11 demonstrate that the model's performance significantly improved with the incorporation of external factors. All the error values were reduced, and the accuracy was increased with incorporation of weather and holidays. Table 4.12 indicates that the predictions are within the acceptable range of accuracy. Thus, the interpretations of the error values confirmed the better performance of NeuralProphet model.

CHAPTER 5

REAL-TIME TRAFFIC SIMULATIONS AND APPLICATIONS

Traffic simulation is the process of replication of real-world traffic conditions in the virtual world based on certain models like car following, lane changing models and so on. Simulation can be run at different levels of details: microscopic, mesoscopic, macroscopic simulations depending on the purposes. This chapter explains the procedure followed for the creation of real-time simulation based on predicted values. A sample study will be explained to analyse the impact of platooned vehicles over the traffic at an intersection using SUMO microscopic simulation software.

5.1 REAL-TIME SIMULATIONS

There are numerous traffic simulation software available, offering a combination of models, parameters, and a wide range of applications. Some of the traffic simulation software are AIMSUN, CORSIM, PARAMICS, SUMO, TRANSIM, VISSIM. Development of urban traffic forecasting system for real-time simulations involves various phases of working.

- Identifying and defining traffic forecasting problem
- Expertise in traffic data and also the real-world traffic behaviour
- Selection of appropriate forecasting model
- Final model evaluation
- Transferability to another location or huge network

A similar work done by (Weber, et al., 2004) provided an advanced traffic information system called OLSIM (OnLine Traffic SIMulation) which was capable of providing the current as well as near future traffic state of North Rhein Westphalia region freeways (Wahle, Chrobok, Pottmeier, & Schreckenberg, 2002). The main purpose of their work was to provide estimates like travel time to the road users about both current and future traffic states. This is known as Advanced Traveller Information System (ATIS), a strategy in ITS. For achieving this in OLSIM, three components were mainly focused.

- Firstly, the traffic data were collected from inductive loop detectors (4000 detectors) and further estimated by forecasting algorithms.
- Secondly, simulation of the traffic state by cellular automata traffic flow model (Nagel & Schreckenberg, 1992) with variation in cell size.
- Finally, the results were provided to the users through online graphical user interface.

In this thesis, the development of traffic flow prediction system is mainly concentrated which can be further given as an input to a traffic simulation. The developed prediction system of this thesis can be used for various applications as mentioned in Figure 5.11, not only for ATIS, but also for numerous other applications. SUMO based simulation is used in this thesis, instead of Cellular Automation (CA) traffic flow model (time and space discrete model) as used in OLSIM. The major advantage is the flexibility of SUMO simulation for using various car following models and variety of vehicle groups in a single simulation. SUMO has a list of car following models which can be defined and implemented according to the requirements. Additionally, the TraCI facility of SUMO provides opportunity to contact a particular vehicle or vehicle type and to manipulate the behaviour even during simulation. This is very much beneficial while using in a dynamic driving simulator to find the vehicle dynamics.

Another main reason for using SUMO is the requirement of the driving simulator (Maas, Benjamin, Martin, & Schramm, 2014) in which this research work can be further applied uses already SUMO based simulation system. Considering these points, the predicted traffic is simulated with SUMO software and the procedure of simulation will be explained in this section.

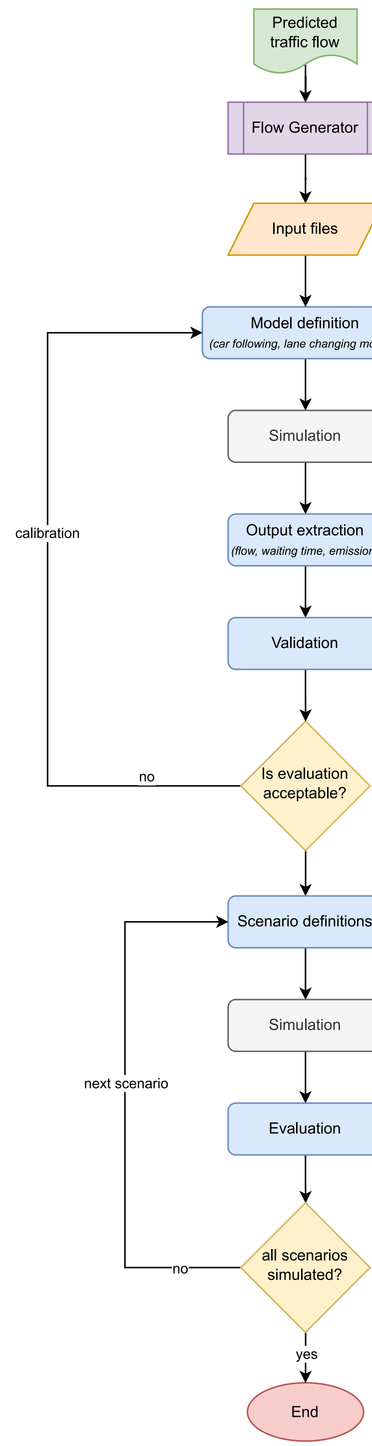

Figure 5.1 Real-time/near future traffic simulation in SUMO

The flow chart in Figure 5.1 explains the process of developing the simulation with SUMO software as per Forschungsgesellschaft fuer Strassen- und Verkehrswesen (FGSV) (Trapp, et al., December 2006). The flowchart can be split into three phases. Firstly, interfacing predicted traffic flow into SUMO. Secondly, microscopic simulation, calibration, and validation of SUMO. Thirdly, the simulation of scenarios with platoon formations.

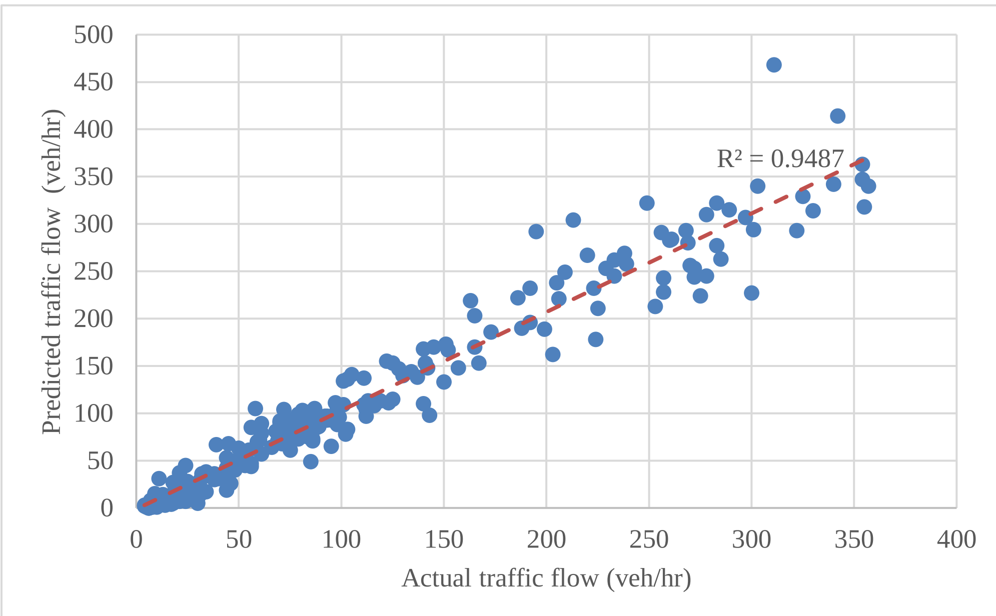

Figure 5.2 Comparison of actual and predicted traffic flow

5.2 FLOW GENERATOR INTERFACING TO SUMO

As discussed in the previous chapter in Section 4.3.4, the observations about the performance of Fb-Prophet and NeuralProphet models for traffic predictions suggested that for faster and multiple traffic predictions for larger network, Fb-Prophet model is much suitable with optimum accuracy, due to the lesser computation time (Pragalathan & Schramm, Apr 2023). Hence for further process of simulations, Fb-Prophet model was used for predictions. Two months of traffic data (Jan 2020 – Feb 2020) were used for model training and the traffic flow on 28/02/2020 was forecasted. To check the model performance, a different traffic data was used when compared to the previous chapter.

The predicted results of Fb-Prophet model is compared with the hourly actual traffic flow in the plot shown in Figure 5.2 with R-squared value of 0.94, where each point represents the corresponding predicted and actual values. The values of all the detectors with actual and predicted traffic flow is given in Appendix A.

The predicted data in CSV format were then given into a flow generator for creating SUMO readable input file in XML format. The file was then given into SUMO for creating the real-world traffic scenario. Figure 5.3 explains the process of interfacing prediction system and simulation system. Appendix B.1 shows the source code of the flow generator tool. And finally, the traffic demand was obtained from the flow generator as a .xml file.

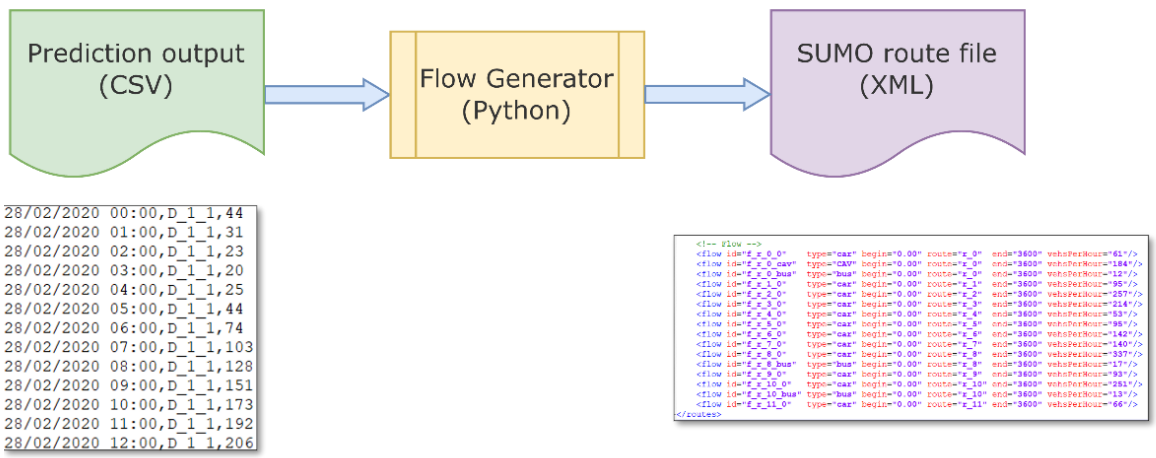

Figure 5.3 Interfacing prediction system and simulation system

Source: (Pragalatahan & Schramm, May 2023)

5.3 SIMULATIONS IN SUMO

5.3.1 Network building

The first step for SUMO simulation is network building to create the junctions and roads of actual traffic network. In SUMO, junctions are called as “nodes” and roads or streets are termed as “edges”. SUMO provides flexibility in network building with various options. One can create the traffic network manually with the tool netedit or import from other sources like Openstreetmap (Openstreetmap, 2023) or software like VISSIM (Verkehrssimulation Software PTV Vissim, 2023). In this thesis, OpenStreetMap networks were used for network building through OsmWebWizard (OSMWebWizard SUMO, 2023). The network data and all other physical data were downloaded by selecting an area in osmWebWizard (OSM Webwizard, 2023). Figure 5.4 shows the osmWebWizard screen shot while network building.

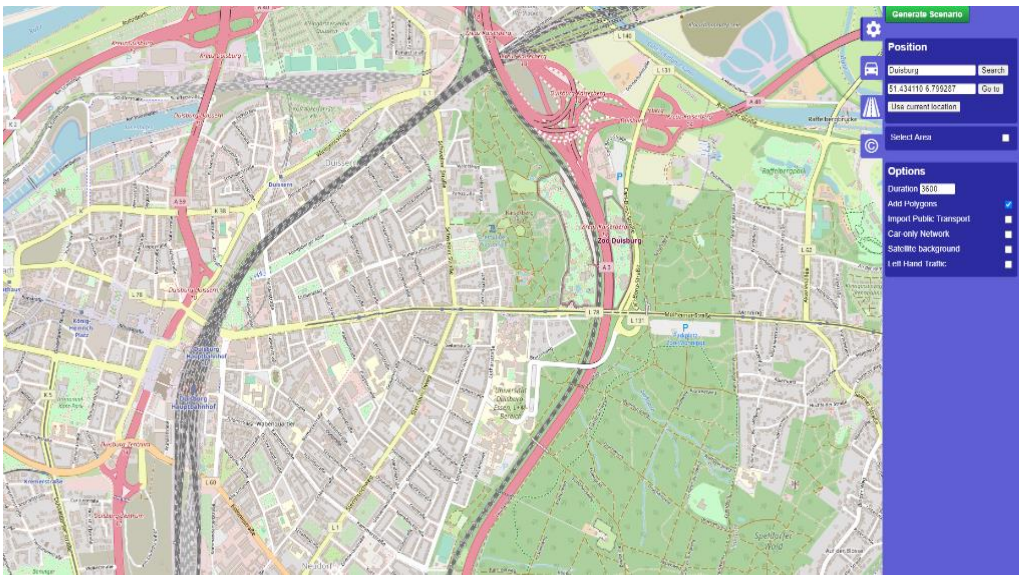

Figure 5.4 OSM web wizard

5.3.2 Definition of route files and traffic demand

The routing files were created based on the real-world traffic to replicate the percentage of traffic going through and taking turning at the intersection. The routes for cars, bicycles, public transportations and pedestrians were verified and then modified if there were any need for corrections. The traffic demand for each routes were taken from the real world or from the predicted results and was used as input file for simulation. The traffic demand file for the SUMO simulation was created from the traffic prediction system's output using the flow generator tool (Appendix B.1). Thus, the network after defining traffic signal program and inductive loop detectors were visible in the SUMO-GUI (Figure 5.5) where the simulation can be run manually.

Figure 5.5 Network created in SUMO GUI

5.3.3 Additional input files

SUMO simulation requires input files which defines the real-world data. All the values and SUMO readable files are given in Appendix B. Two major input files are the network files (.net.xml) and route files with traffic demand (.rou.xml). Apart from these two files, generally SUMO can be provided with additional files to define additional information like the inductive loop detectors, dynamics simulation controls and many other extra inputs to represent the real world. For the current simulation, the inductive loop detectors (.det.xml) were defined in the lanes for validation of the simulation. The Traffic Signal Program was also taken from the field and given in to simulation by defining it in the input file. Figure 5.6 explains a sample signal light program with the green duration and the complete signal cycle. Figure 5.7 shows the defined traffic signal program according to SUMO readable format and included in main network file as a static signal program. SUMO gives a number of options for output extraction at different levels (vehicle-wise, trip-wise, edgewise/lane-wise). For validation of simulation inductive loop detectors were defined and then the traffic flow values were extracted.

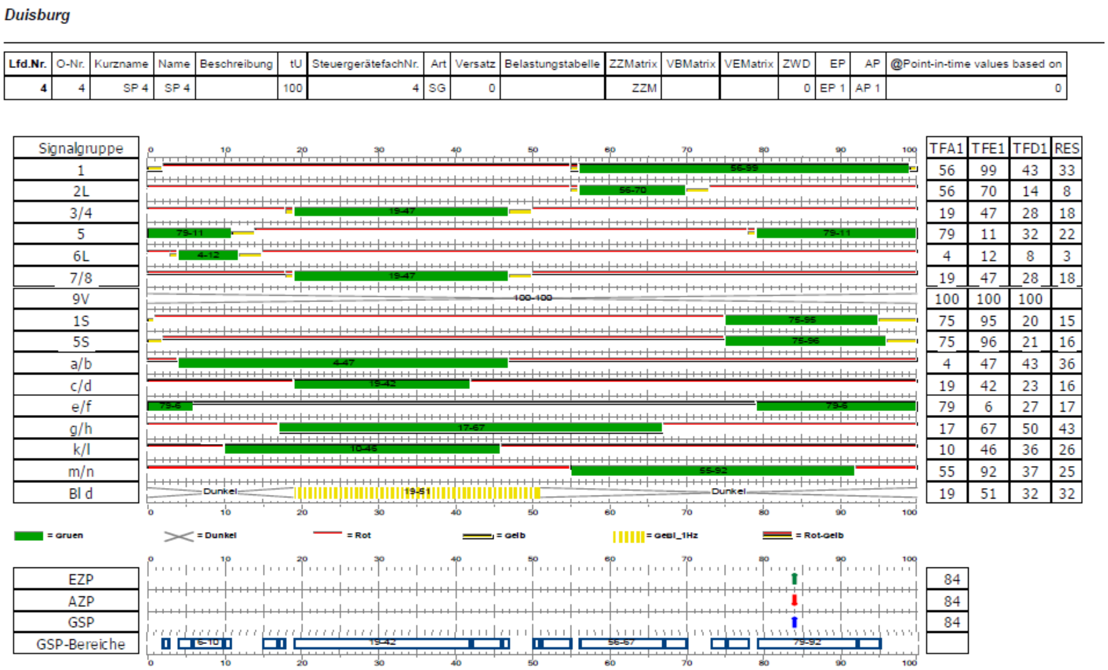

Figure 5.6 Traffic signal program

Source: (Wirtschaftsbetreibe Duisburg, 2019)

```
<tlLogic id="GS_cluster_1621664904_291764491_34093024_35230026" type="static" programID="0" offset="0">
    <param key="cycleTime" value="100"/>
    <phase duration="2" state="Grrryrrr"/>
    <phase duration="1" state="Grrrrrrr"/>
    <phase duration="1" state="Gyrrrrrr"/>
    <phase duration="7" state="GGrrrrrr"/>
    <phase duration="1" state="yGrrrrrr"/>
    <phase duration="2" state="yyrrrrrr"/>
    <phase duration="1" state="ryrrrrrr"/>
    <phase duration="3" state="rrrrrrrr"/>
    <phase duration="1" state="rryyrryy"/>
    <phase duration="28" state="rrGgrrGg"/>
    <phase duration="3" state="rryyrryy"/>
    <phase duration="5" state="rrrrrrrr"/>
    <phase duration="1" state="rrrryyrr"/>
    <phase duration="14" state="rrrrGGrr"/>
    <phase duration="3" state="rrrrGyrr"/>
    <phase duration="5" state="rrrrGrrr"/>
    <phase duration="1" state="rrrrGrrr"/>
    <phase duration="20" state="GrrrGrrr"/>
    <phase duration="1" state="Grrryrrr"/>
</tlLogic>
```

Figure 5.7 Defined traffic signal program in SUMO

5.3.4 Configuration file

The simulation parameters and the input files were then defined in the configuration file (Appendix B.3). For calibration and validation, the actual traffic scenario was simulated, and the simulated traffic flow values were extracted from defined loop detectors during simulation. The total simulation duration and other simulation parameters were also defined, and the simulation was run for 3600 secs.

5.4 SIMULATION OF PLATOON FORMATIONS

In this section, a sample simulation study will be explained for finding the impact of platoons at an urban intersection. Platooning involves a group of vehicles traveling closely together in a coordinated manner, taking advantage of communication between the vehicles and automated driving capabilities. This formation can lead to various benefits, such as reduced aerodynamic drag, improved fuel efficiency, and enhanced traffic flow on highways. Platooning is considered one of the promising applications of connected vehicle technology in improving overall transportation efficiency and safety. It is very much efficient at the highways where there is an uninterrupted traffic flow for a longer distance (Martínez-Díaz, Al-Haddad, Soriguera, & Antoniou, 2021). Such platoon's performance might vary at a busy urban intersection where there a numerous factors affecting platoon formations. The formation of platoons at signalized urban intersections is a highly significant subject of interest in numerous research studies (Hardes & Sommer, 2019).

One significant advantage of connected vehicles is their ability to communicate and exchange information not only with the infrastructure (V2I) but also with the nearby vehicles (V2V). Due to this capability of information exchange, it becomes feasible to form a coordinated driving group of vehicles. By achieving smaller spacing between vehicles and enabling high-speed travel, the formation of such platoons increases road capacity significantly (especially medium lengths) (Martínez-Díaz, Al-Haddad, Soriguera, & Antoniou, 2021). Majority of studies on platoons have primarily focused on highways, aiming to understand their impact on traffic efficiency and environmental considerations. Previous research has indicated that truck platooning can lead to positive effects, including reduced fuel consumption and subsequent lower harmful emissions. There is a necessity to evaluate the impacts in urban areas with complex traffic movements, particularly at signalised intersections (Hardes & Sommer, 2019).

In Section 5.2, initially the process of traffic flow prediction using timeseries prediction models and then the interfacing of the prediction results with the microscopic simulation network was clarified. Then the process of network building and traffic demand definition and output extraction were explained in the Section 5.3. Now with the developed simulations, the impact of platoons at the intersection will be analysed and discussed. The impact of CAV over overall traffic with traditional and smart vehicles were studied (Pragalatahan & Schramm, May 2023) and is explained in this subsection.

5.4.1 Platoon formation

SUMO offers a configurable plugin called Simpla, which enables the spontaneous formation of vehicle platooning. This plugin allows to define a vehicle’s behaviour even when it is inside a platoon. This is possible by defining various vehicle types to denote different platoon vehicle modes like leader, follower and catchup modes. The default setting in SUMO 1.16.0 (Vehicle Type Parameter Defaults, 2023) was employed to set up human driving vehicles. The vehicle length is 5m, the minimal gap between vehicles is 2.5m, and the Krauss model (Krauß, Wagner, & Gawron, 1997), (Car-Following Models, 2023) is used as car following model. Subsequently, the definition of Connected Autonomous Vehicles was done with assumption that the vehicles were fully automated with an automation level of 5 and the parameters were defined according to the list of parameters given in Table 5.1. Moreover, these CAVs were also capable of forming platoons with leader and follower vehicles. For emissions, the default model used in SUMO is HBEFA3/PC_G_EU4 by considering the vehicles as average passenger car based on the Handbook of Emission Factors for Road Transport (HBEFA) (HBEFA, 2023).

Table 5.1 List of parameters with default and altered values for platoon formation

Attribute	Values		
	Default/human driving	Leader	Follower/catchup
Simulation Parameters			
Step Length	0.1	0.5	
Vehicle type parameters			
Length (m)	5.0		
Width (m)	1.8		
Height (m)	1.5		
Acceleration (m/s^2)	2.6		
Deceleration (m/s^2)	4.5		
Emergency deceleration	9.0		
Maximum speed (m/s)	55.55	50	50
Minimum gap (m)	2.5		0.5
Drivers desired minimum time headway-tau (s)	1		0.5
Platoon parameters			
Control rate (s)	1.0		
Verbosity	4		
Maximum number of vehicles in a platoon (Nos.)	10		
Maximum platoon gap (m)	15		
Catchup distribution (m)	50.0	100	
Switch impatience factor	0.1	-1	
Platoon split time (s)	3.0		
Lane changing mode	594	594	514
Speed factor	1.0	1	1.2

5.4.2 Scenarios simulations

The effects of platoon formations at signalised intersections were simulated in SUMO with the penetration of CAVs at different rates (75% HV + 25% CAV, 50% HV + 50% CAV, 25% HV + 75% CAV). As a result, the impacts of mixed traffic with both human-driven vehicles and CAVs were studied, using the human driving scenario as the baseline for comparison. Figure 5.8 shows the screenshot of SUMO simulation with human driving, and platooned vehicle. The vehicles are differentiated with different colours where yellow- leader vehicle, green- follower vehicle , red- catchup vehicle, cyan- human driven vehicles. The study involved analysing the overall traffic efficiency and environmental impacts, with a focus on extracting delay or waiting time (i.e., the time in which the vehicles speed was below or equal to 3.6 kmph.), fuel consumption and emission (i.e., the amount of fuel the vehicle consumed, and the amount of CO and CO_2 emitted by the vehicle during the trip respectively) as metrics for the assessment.

Figure 5.9 illustrates the percentage of improvement in fuel consumption and emissions resulting from the increase in the penetration of CAVs. Similarly, Figure 5.10 shows the improvements in average waiting time. This can be justified because of smoother driving behaviour of CAVs (especially the follower vehicles). This confirms that the increase in penetration of CAVs have positive impact on both overall traffic efficiency and environment.

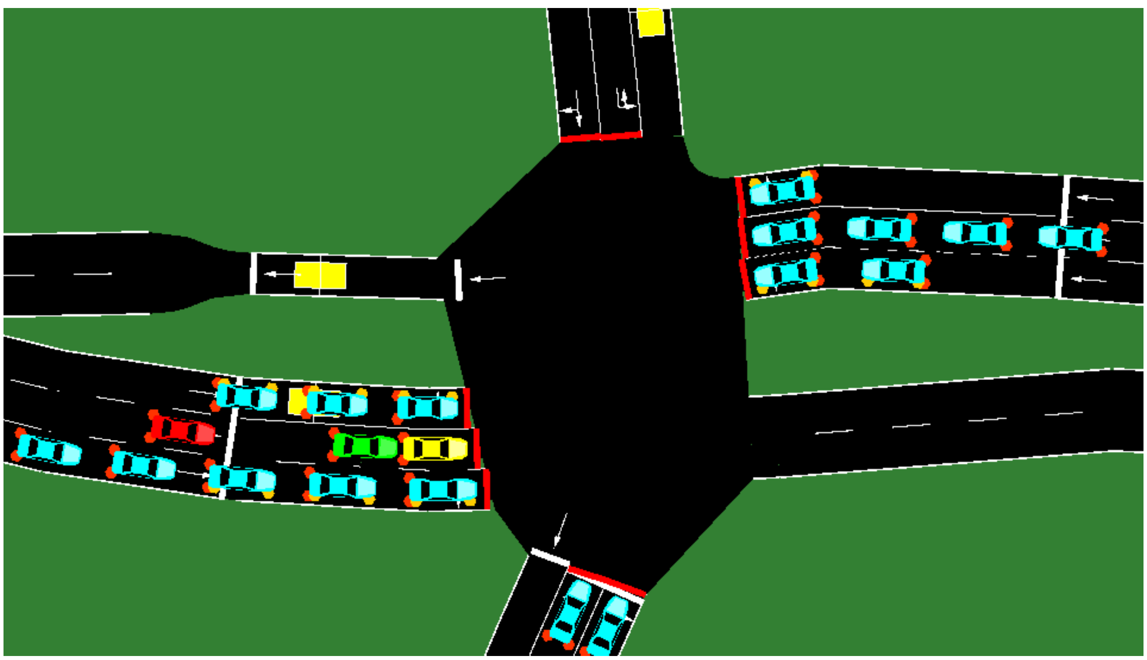

Figure 5.8 SUMO simulation with platoons (Vehicles colour indication – yellow: leader, green: follower, red: catchup, cyan: human driven)

Source: (Pragalatahan & Schramm, May 2023)

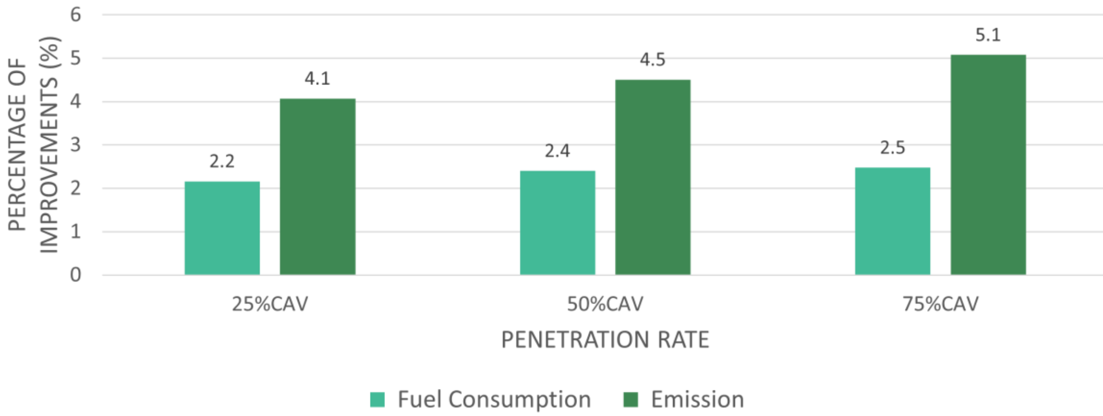

Figure 5.9 Environmental impacts

Source: (Pragalatahan & Schramm, May 2023)

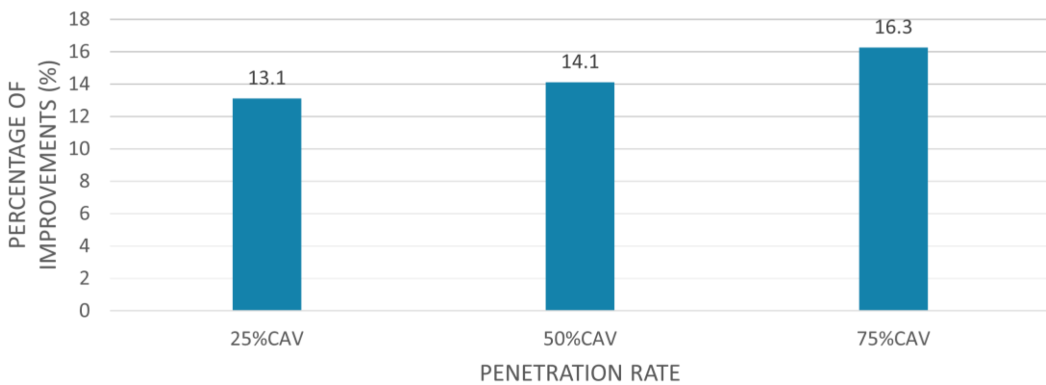

Figure 5.10 Network Impacts: Improvements in traffic delays

Source: (Pragalatahan & Schramm, May 2023)

The results obtained from the simulations provide a clear depiction of how the overall traffic efficiency is influenced by the presence of connected vehicles. These simulations offer valuable insights into the impact of connected vehicle technologies on traffic flow, congestion, fuel consumption, and emissions. Such information is vital for making informed decisions and implementing effective strategies to enhance transportation system's efficiency and sustainability.

5.4.3 Future scope

Faster and more efficient solutions are possible with the recent advancements in traffic data collection methods and communication systems. Future smart vehicles heavily rely on real-time and near-future traffic information to operate efficiently and safely. Traffic Flow Prediction Models (TFPMs) play a crucial role in providing this essential information. The availability of vast databases containing historical traffic data, coupled with advancements in communication technologies, greatly facilitates the possibility of making accurate traffic predictions. These data sources provide valuable insights into traffic patterns, behaviour, and trends, enabling traffic engineers and researchers to develop sophisticated Traffic Flow Prediction Models. Real-time simulation systems offer the advantage of evaluating various scenarios in a cost and time-effective manner. In this study, the impacts of platoon formations at a signalized intersection have been investigated using real-time simulations. GLOSA (Green Light Optimal Speed Advisory) is another application of connected vehicles, where vehicles are equipped to receive signal light timings from the infrastructure.

With this information, vehicles can adjust their speed to optimize their approach and reach the signal during green timing. In the future, the incorporation of GLOSA into simulations and traffic flow predictions, along with weather data (such as rainy or snowy conditions) as an exogenous variable, can be carried out. This integration of real-time simulation can explore the potential benefits and challenges of GLOSA and platooned vehicles in urban scenarios in various weather conditions and assess its impact on traffic efficiency and safety. By considering weather as a variable in the simulations, researchers can gain valuable insights into the performance of connected vehicles and adaptive traffic management strategies under different weather scenarios, ultimately leading to more robust and effective transportation solutions.

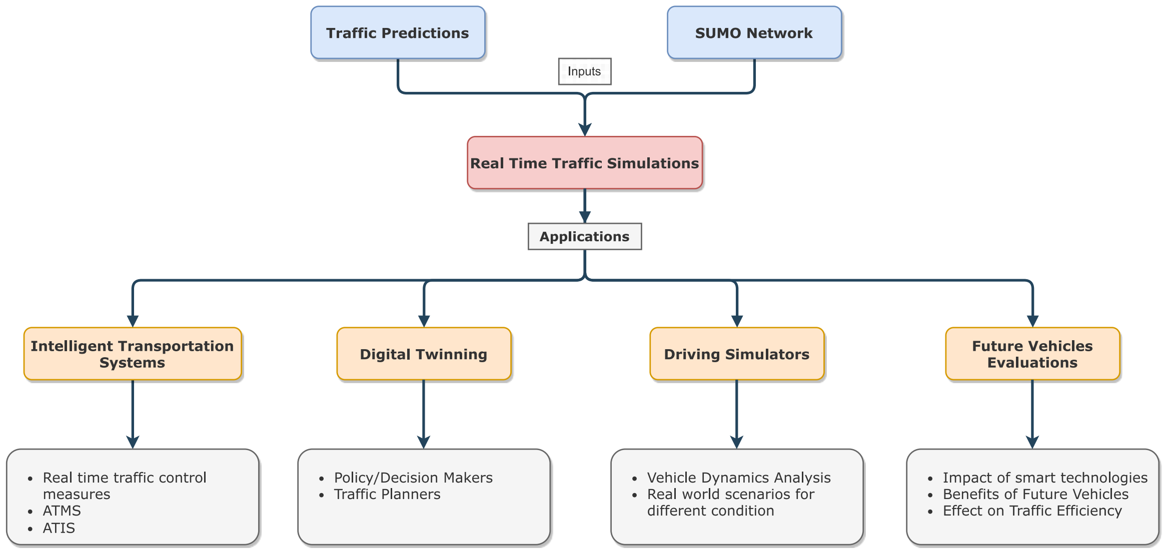

Figure 5.11 Applications of real-time traffic simulations

Source: (Pragalatahan & Schramm, May 2023)

5.5 APPLICATIONS OF REAL-TIME TRAFFIC SIMULATIONS

Thus, the major focus of creating real-time traffic simulation is achieved and further application of such simulations will be discussed in this section. Such traffic flow predictions and simulations can be practically used in various circumstances. Figure 5.11 explains some of the possible applications. In this current study, traffic flow predictions are conducted for the near future or short-term duration. Predicting city traffic conditions encounters several challenges, including dealing with networks with a greater number of shorter links and the high computational cost associated with predicting numerous links (Liang, et al., 2018). To support real-time functionalities, it is essential to continuously acquire dynamically varying traffic information. There are many techniques available for such long-term continuous data acquisition. Some of them are inductive loops, videographic detection and vehicle tracking devices like GPS trackers. The acquired traffic information must be continuously updated for building a Dynamic map. By utilizing simulation, this dynamic map can assist in developing tools for real-time traveller information, local and global traffic control, and other traffic management strategies (Chmiel, et al., 2016).

Real-time traffic simulation finds application in various ITS approaches, providing valuable real-time measures for traffic management and optimization. However, incorporating Connected Vehicles demands a significantly more intricate simulation system capable of predicting various scenarios in detail. The shortcomings of the existing real-time simulation systems are outlined in the work (Sippl, Schwab, Kielar, & Djanatliev, 2018). Their focus has been on simulating urban traffic scenarios, particularly incorporating automated driving vehicles.

Real-time traffic simulation offers numerous benefits, including the ability to analyse traffic efficiency, implement real-time traffic control measures, and evaluate overall traffic conditions, considering the integration of connected autonomous vehicles at various penetration rates and automation levels. All these studies can be conducted in the virtual world through driving simulators and real-time traffic simulations. This virtual environment allows researchers to explore and analyse various scenarios to identify optimum solutions for reducing congestion and minimizing delays in urban traffic. Thanks to the recent advancements in simulation software, it has become feasible to analyse fuel consumption and emissions. Various measures of effectiveness such as average speed, stopping duration, number of stops, queue lengths also can be obtained from simulations.

5.5.1 Intelligent Transportation Systems

Intelligent Transportation Systems (ITS) are strategies to provide smart and cost-effective solutions for traffic related problems. It aims to provide innovative solutions for transportation management and enable safer and smarter usage of traffic networks. In recent decades, ITS plays a major role in traffic engineering and also automobile engineering because of availability of huge data and developments in information-communication technologies. Nowadays, both government and people are looking for smart things, solutions, cities, and everything smarter. ITS become indispensable element in the transport policies to tackle challenges of traffic related problems of 21st century (Federal Minstry of Digital and Transport: ITS, 2023). Most of the nations, especially the

European Commission mainly focus on ITS based innovative projects to tackle emission and congestions (European Commission: Mobility and Transport, 2023). Information and Communication Technologies (ICT) help in such smarter way of living and thinking. It helps to analyse past data and use the knowledge for time and cost-effective solutions with available resources. Intelligent Transportation Systems handle traffic related problems with such smarter strategies. Thus, making use of ITS makes any city “a smart city” with following advantages.

- More communication of traffic data
- Prediction of future traffic
- Smart traffic controls based on the prediction values
- Simulating the urban traffic to digitalize the current and future traffic scenarios

Figure 5.12 shows various strategic research areas which act like the building blocks for developing an innovative ITS system for integrating advanced technologies in the transportation field. Both the vehicles and traffic infrastructure focuses on smart applications to make more efficient transportation system. Traffic flow predictions plays a vital role in most of the ITS strategies for real-time solutions. Especially traffic predictions supports the real-time traffic strategies for Advanced Traffic Management Systems (ATMS) and Advanced Traveller Information systems (ATIS). It updates both the network and the user with real-time traffic and near future traffic, thus congestion can be avoided or reduced with such faster traffic updates.

The rapid urbanization and population expansion in cities have led to higher congestion levels, longer travel times, and increased environmental impacts. The urban mobility is facing a multitude of challenges as a result of population growth and increasing traffic demand. Intelligent Transportation System (ITS) strategies are designed to tackle the challenges posed by urban traffic, including congested road networks and limited road capacity. These challenges are considered as the primary barriers for attaining efficient road traffic streams.

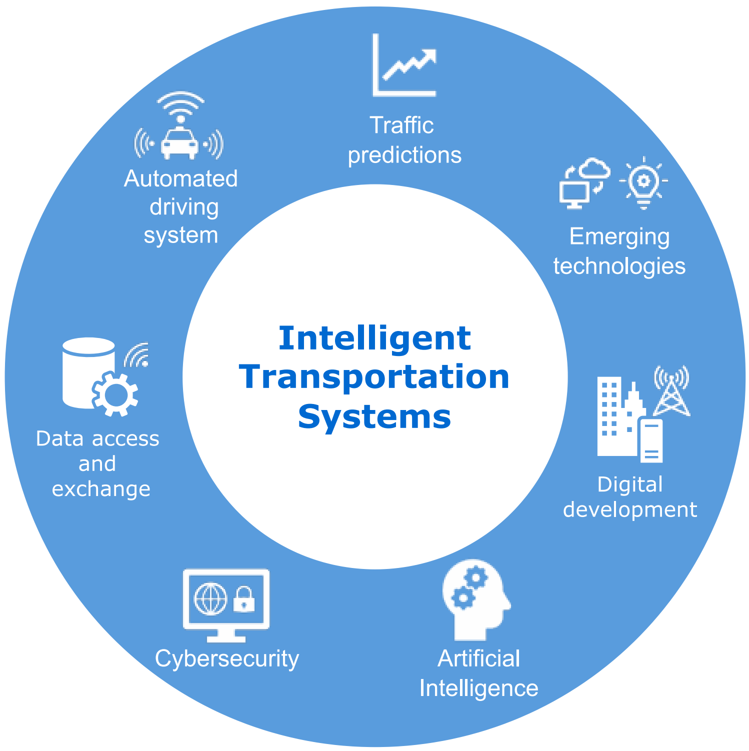

Figure 5.12 Intelligent Transportation Systems

ITS encompasses a diverse range of technologies that can improve the urban traffic efficiency. These technologies serve various phases of ITS such as traffic data collection and storage, traffic predictions and communication, which work together to develop real-time traffic management tools. Based on the purposes, the functionalities of ITS such as route planning, real-time traffic management and the integration of connected vehicles (V2V, V2I) can be decided (Chmiel, et al., 2016).

A similar experiment in the urban traffic environment was conducted by simulating the scenario with the dataset obtained from the project Ko-PER at a signalised intersection in Aschaffenburg, Germany. Subsequently. The simulation was employed to analyse the interactions between pedestrian behaviours and self-driving vehicles (Sippl, Schwab, Kielar, & Djanatliev, 2018). Such analyses can be carried out with the help of real-time traffic simulation systems where both ITS tools and CAVs can be analysed and real-time solutions can be provided. Given the significance of real-time traffic simulations, particularly in high traffic volume areas, this work concentrated on formulating an urban traffic simulation system. The majority of real-time solutions prioritize updating the network and road users with both current and future traffic data. For such applications, real-time or near-future traffic flow predictions play a crucial role as one of the fundamental elements. Furthermore, predictive traffic information proves to be more effective than real-time traffic data, as it can account for the rapidly changing and dynamic nature of traffic flow caused by spatial and temporal influences (Dong, Mahmassani, & Lu, 2006).

The two major real-time strategies in Intelligent Transportation Systems (ITS) are Advanced Traffic Management Systems (ATMS) and Advanced Traveller Information Systems (ATIS). Both can be used with the help of traffic flow prediction models which subsequently enable them to inform both the transport network (Liang & Wakahara, 2013) and the traveller about current and future traffic scenarios. Especially, the optimization of traffic signal lights and variable message signs particularly relies on near-future traffic values more than the current traffic state (Kato, Kim, Suzuki, & Okuma, 2005). Likewise, in the case of Advanced Traveller Information Systems (ATIS), route guidance is provided to travellers to minimize travel delays, considering the near-future traffic conditions (Chen & Underwood, 1991).

5.5.2 Digital twinning

In the upcoming years, advancements in digitalization within transportation systems are anticipated to make significant progress (European Commission: Mobility and Transport, 2023). EU's digital strategy is trying to achieve more digital transformation by 2050 (European Commision: A Europe fit for the digital age, 2023). Europe's "Digital Decade" is the major focus of European Commission to handle data, technology and infrastructure. The cities can be replicated in the virtual scenarios with all the physical details and traffic information that are predicted through models and resulting in the digital twin. Such development of digital twinning systems helps to make real-time decisions and traffic control measures like varying speed values, lane closures. Real-time simulations help to bridge the actual world and virtual scenarios to develop and analyse alternate smart solutions (PWC: Digital Twins, 2023). Digital twinning requires constant updating of traffic data about the real world and near future traffic conditions. Such predictions are obtained through the previously explained prediction models (Section 2.2.2). This has a wide range of application also for the future years to make decisions and allocate budgets accordingly. Thus, traffic planners are most benefited through this digital twinning. One such digital twinning was done for Geneva Motorway in the project Digital Twin Geneva Motorway(DT-GM) with SUMO software. With this project DT-GM, real-time traffic data from traffic sensors were used to simulate the real-world traffic. (Kušić, Schumann, & Ivanjko, 2023).

With the help of SUMO numerous city level simulations were done in several projects. The development of one such simulation scenario for Cologne city in Germany was carried out with SUMO. It was utilised to simulate and collect real-time traffic volume data for individual roads in the city (TAPASCologne, 2023). The German capital city, Berlin was virtually replicated in a whole city level simulation for over 24 hours duration. The simulated scenario was called as Berlin Sumo Traffic (BeST) (Berlin Sumo Traffic Scenario, 2022) and was done for motorised private transport with 2,25 million trips which is one of the largest microscopic traffic simulations. Likewise such simulations were done for cities like Braunschweig/Brunswick (Armellini & Bieker-Walz, 2020), Ingolstadt- InTAS (Lobo, Neumeier, Fernandez, & Facchi, 2020), Stuttgart (Förster, Löhr, Grätz, Petit, & Kargl, 2017) and Hamburg-TAVF (David & Plötz, 2020). Similar bigger network sized simulations were also done in other countries. Some of the cities and their corresponding projects are Bologna- iTETRIS (Bieker, Krajzewicz, Morra, Michelacci, & Cartolano, 2014), Dublin (Gueriau & Dusparic, 2020), Luxembourg-LuST (Codeca, Frank, & Engel, 2015), Monaco-MoST (Codeca & Härri, 2017), Turin-TuST (Rapelli, Casetti, & Gagliardi, 2021), Boston, Lisbon, Los Angeles, Rio de Janero and San Francisco (Ambühl, Menendez, & González, 2022). Similarly, the work from this thesis can be scaled up to create a complete Duisburg SUMO scenario which can be provided to the driving simulator developed at Uni DUE, Mechatronics. This will be explained in detail in the Section 5.5.3.

5.5.3 Driving simulators

The automobile industry is facing more progress with automated applications. Driving simulators are used for several analysis before implementing a new technology in the real world. For example, Figure 5.13 shows the dynamic driving simulator developed at the Chair of Mechatronics, University of Duisburg-Essen, which is a human-centred one which can give tactile feedback to the drivers (Maas, Benjamin, Martin, & Schramm, 2014). This simulator setup is made of driver's cabin and a curved screen for realistic visualisation. The driver's cabin is capable of capturing the movement of the cabin with the help of motion platform. The screen before the driver's cabin is fully covering the driver's vision with angle of 250^{o}.

One of the recent studies with this driving simulator was to obtain data for fuzzy control model for human drivers (Ma, 2020). Then the parameters from the driver model and the vehicle model were used for studying different levels of automation. Some of the other common usages of such driving simulators are testing of Advanced Driver Assistance Systems (ADAS), testing of Automated Vehicles and for various advanced research purposes. Such simulations can be made with real-time traffic by using traffic prediction systems and make the analysis more reliable. For e.g., predicted traffic data exchange between the infrastructure and the vehicle can be analysed and also the driver's reaction to the updated traffic information can be studied with driving simulators. Prediction of traffic during certain weather conditions can be carried out and the influence of the weather over the vehicle dynamics can be evaluated for different precipitations (wet/dry road surface conditions) (Weber T. , 2021).

Figure 5.13 Dynamic driving simulator

Source: (Chair of Mechatronics, University of Duisburg-Essen, 2023)

5.5.4 Future vehicle evaluation

Various research works are being done to evaluate the effect of automated vehicles over traditional human driving vehicles especially in urban in scenarios (Ma, 2020). Smart vehicles also require updated information about the traffic from the nearby vehicle and also from the infrastructure. For this purpose, the traffic predictions can serve the near future traffic information for better routing decisions to avoid travel delays and waiting time. Environmentally friendly traffic is obtained by having lesser emissions and lesser congestion. This is evaluated with the help of simulations to study the fuel consumption and emissions. The deployment of Cooperative-ITS makes the transport sector move ahead with automated solutions with support of wireless technologies. Thus, vehicles can connect to each other and also to infrastructure through wireless communications. In this study, the simulation of platoon formations with different percentages of Connected Automated Vehicles was carried out and the traffic efficiency is analysed.

Indeed, the advancements in automotive and transportation hold enormous potential, but they must be thoroughly tested and evaluated before being deployed for road users' safety and efficiency. In the current transportation era, there is a strong emphasis on integrating emerging technologies to equip advanced Connected Autonomous Vehicles (CAVs). These cutting-edge technologies create safer and more efficient transportation system by allowing these vehicles to communicate between each other (V2V) and the surrounding infrastructure (V2I) (Uhlemann, March 2015). The combination of Intelligent Transportation Systems (ITS) and Connected Autonomous Vehicles (CAVs) resulted in the creation of a Cooperative Intelligent Transportation System (C-ITS). This system has an added advantage to both broadcast and receive real-time traffic information and thereby facilitates real-time traffic controls and management measures (Qu, Li, Wang, & Dixit, 2017).

It is strongly believed that there will be significant increase in connected autonomous vehicles in the coming decades (Schoitsch, 2016) (Shaping Europe's digital future, 2022). As connected autonomous vehicles are expected to increase in the future, it is necessary to equip road networks with the required infrastructure to facilitate efficient traffic management. Urban road networks offer a wide array of scenarios that can be analysed to determine the critical penetration rate and automation level required for connected autonomous vehicles to operate effectively. The impact of connected vehicles on urban traffic scenarios varies depending on their penetration rate. Therefore, studying the traffic flow involving both human-driven vehicles and automated vehicles (at different levels), with or without connectivity(V2V/V2I) is essential.

CHAPTER 6

SUMMARY OF THE STUDY

Huge amount of traffic data has been collected and stored everyday with help of advancement in the technologies. The usage of such data is an essential need of the current digitally transforming world, especially the transportation field. Many real-time and near-future data-based advancements will reach both the physical network and the users. This chapter summarises the different phases of the real-time and near future traffic forecasting and simulation processes. Finally, the future scope of this thesis is also explained.

6.1 CONCLUSIONS

This thesis explained how the state-of-the-art time series models potentially performed for urban traffic flow predictions and efficiently captured the effect of exogenous variables. Thus, the main goal of this thesis as explained in Section 1.2 is achieved to develop a prediction system that can read the raw traffic data and then capture the fluctuations and variations of everyday traffic flow at an urban location. To provide a traffic flow prediction system for urban traffic conditions, the following goals were focused:

- Appropriate and accurate traffic data collection
- Selection of traffic flow prediction model to estimate the real-time and near future traffic
- Generation of SUMO traffic simulation based on the predicted traffic flow values

The summary of the overall observations while addressing each objective and the final conclusions and future scope will be explained in the following subsections.

6.1.1 Quality assessment of traffic data collections

The results shown in the Section 4.2.3, are in line with the previous literature works proving that the performance of video detection technique is based on the camera's mounting (height/location) and the frame of the camera. Apart from this, other external factors like lighting (sunlight), weather (cloudy/rainy/snowy) and shadows of nearby lane traffic also influences the accuracy of video detection technique. The error due to limited lighting at night times can be eradicated by using thermal cameras (Fu, Stipancic, Zangenehpour, Miranda-Moreno, & Saunier, 2017). On the other side, inductive loop detectors gave much accurate results and very much suitable for long-term traffic data collections. For forecasting the urban traffic flow, the already installed inductive loop detectors are very much suitable for data collection for a continuous data exchange.

6.1.2 Traffic flow prediction models

The results given in the Section 4.3.3 indicate that the two recent models (Fb-Prophet and NeuralProphet) are very accurate for urban traffic flow predictions. The two models can capture the trend and seasonal patterns of the urban traffic with influence of weekday/weekend traffic and peak/off-peak hour traffic. The NeuralProphet models are also capable of estimating traffic with effect of weather and holidays. The models gave acceptable traffic data estimation with faster and automated computations. Adding to this, lesser human intervention is needed for model definitions and order estimations. Since the model is open source and available in Python/R , traffic prediction for research purposes can be carried out with more expandability.

Most of the time, the impacts of external dynamic factors were neglected while establishing traffic flow prediction models. Weather has impact on various traffic flow parameters like speed, density and volume and other network operations and traffic managements. The effect of weather especially precipitation (rainfall) on urban traffic flow was incorporated into traffic flow prediction model successfully. This thesis hence also proved the efficiency of NeuralProphet model for prediction of any special events or atypical conditions like severe weather (Section 4.4.1).

6.1.3 Real-time traffic simulations in SUMO

Traffic Simulations play a major role for analysis of alternate smart solutions to reduce congestion and travel time. Instead of random day traffic flow as demand input for simulations, the traffic flow predictions can be used for simulations even with the effect of external factors like weather and holidays. This research work has developed a framework (Section 5.1) that can predict the urban traffic flow and generate a SUMO readable traffic demand file. This input can be given into SUMO networks for real-time simulations. The developed flow generator tool (Section 5.2) helps this process more automated with lesser human intervention. A sample simulation was run to analyse the impact of platooned vehicle in a mixed traffic with both human driving and connected automated vehicles.

6.2 SCIENTIFIC CONTRIBUTION OF THE STUDY

In Germany, most of the cities have inductive loop detectors (ILD) as the basic source of real-time traffic data. The accuracy of urban traffic data collection is doubtful because of various inconsistencies. The reliability of data collection of urban traffic using ILD and videographic method was studied in this thesis. Thus, the findings explained in Section 6.1.1, would contribute to the researchers or the traffic data users to choose the data collection method more appropriately. Secondly, this thesis explored the state-of-the-art time series prediction models for urban traffic simulation applications. Thirdly, the potentiality of the models in capturing exogenous factors was also checked. Thus, the traffic flow prediction systems developed in this thesis was made as multi-efficient, automated and multi-applicable one for real-time traffic simulations.

Various models have been developed since decades to forecast diverse parameters of traffic stream such as flow/volume, speed, congestion, delay and so on. Among others, traffic flow is one of the important parameter which provide information about urban traffic condition. Traffic flow prediction system developed in this thesis can serve as one of the foundations for ITS strategies. Especially in urban areas, the traffic forecasting along with simulations helps the traffic planners, engineers or researchers to provide facilities in a way to reduce delay and congestion. With the help of huge traffic flow data that have been recorded every day, traffic flow in a particular location can be forecasted and simulated and can be applied in various circumstances as listed:

- To update travellers or road users
- For the evaluation of existing traffic control measures
- To provide efficient traffic management strategies
- To analyse the influence of future or smart technologies, etc.

Advanced Traffic Management Systems (ATMS) and Advanced Traveller Information Systems (ATIS) are two noticeable real-time strategies of ITS which gets more benefitted with the help of traffic flow prediction models. After predictions, these strategies subsequently serve to inform both the network and the traveller about current and future traffic scenarios. Hence travel delays and vehicle idling can be reduced, leading to a reduction in waiting time and also fuel consumption. This research work will also help the traffic planners and decision makers to have an idea about the process of forecasting the future scenarios for making cost effective and efficient developments.

6.3 LIMITATIONS

This thesis mainly addressed urban traffic at an intersection. The steps in the process of prediction modelling might be similar for highway traffic and the traffic at a bigger network level. But the model selection and order or parameter estimation has to be finalised according to the corresponding historical traffic data. The model selection has to be carried out between Fb-Prophet or NeuralProphet model depending on both the factors: accuracy and computation time. Computation time also plays a major role in the model selection since the real-time applications require repetition of running the prediction models very frequently. Thus, a pre-modelling process should be required for deciding the model and its parameters before forming a prediction system for highway and a network level traffic.

6.4 FUTURE SCOPE

European commission is targeting Intelligent Transportation Systems for smart solutions and cost-effective decision making (European Commission: Mobility and Transport, 2023; Federal Minstry of Digital and Transport: ITS, 2023). Traffic predictions serves as the foundation for most of the real-time solutions. Thus, some of the applications mentioned in the Section 5.5 would be trending research topics in the coming decades. With support of global modelling, the traffic flow prediction for a bigger network can be done for a bigger network where numerous locations has to be predicted simultaneously. The performance of the model for traffic prediction at uncertain conditions can be also evaluated (NeuralProphet: Uncertainty, 2023). With availability of historical external data, it is possible to incorporate the exogenous variables into the model and predict the atypical or adverse conditions. In future work, the prediction results at different weather conditions can be used for simulation of urban traffic at severe weather conditions and can be also used in dynamic driving simulators for analysis of vehicle dynamics at different weather conditions. Adding to this, this thesis mainly focused on urban intersection. Hence the methodology followed in this work can be extended to bigger network and also to highway traffic.

APPENDIX A

Table A.1 Actual and predicted traffic flow using Fb-Prophet model for all the detectors

Timestamp	Detectors	Actual traffic flow (veh/hr)	Predicted traffic flow (veh/hr)
28/02/2020 00:00	D_1_1	44	19
28/02/2020 01:00	D_1_1	31	14
28/02/2020 02:00	D_1_1	23	11
28/02/2020 03:00	D_1_1	20	15
28/02/2020 04:00	D_1_1	25	11
28/02/2020 05:00	D_1_1	44	53
28/02/2020 06:00	D_1_1	74	69
28/02/2020 07:00	D_1_1	103	136
28/02/2020 08:00	D_1_1	128	147
28/02/2020 09:00	D_1_1	151	173
28/02/2020 10:00	D_1_1	173	186
28/02/2020 11:00	D_1_1	192	196
28/02/2020 12:00	D_1_1	206	221
28/02/2020 13:00	D_1_1	220	267
28/02/2020 14:00	D_1_1	238	269
28/02/2020 15:00	D_1_1	260	283
28/02/2020 16:00	D_1_1	270	256
28/02/2020 17:00	D_1_1	257	228
28/02/2020 18:00	D_1_1	225	211
28/02/2020 19:00	D_1_1	188	190
28/02/2020 20:00	D_1_1	157	148
28/02/2020 21:00	D_1_1	130	140
28/02/2020 22:00	D_1_1	101	134
28/02/2020 23:00	D_1_1	71	68
28/02/2020 00:00	D_1_2	30	5
28/02/2020 01:00	D_1_2	24	7
28/02/2020 02:00	D_1_2	21	7
28/02/2020 03:00	D_1_2	14	5
28/02/2020 04:00	D_1_2	8	6

28/02/2020 05:00	D_1_2	18	15
28/02/2020 06:00	D_1_2	45	40
28/02/2020 07:00	D_1_2	72	104
28/02/2020 08:00	D_1_2	87	105
28/02/2020 09:00	D_1_2	96	95
28/02/2020 10:00	D_1_2	113	113
28/02/2020 11:00	D_1_2	140	168
28/02/2020 12:00	D_1_2	165	203
28/02/2020 13:00	D_1_2	186	222
28/02/2020 14:00	D_1_2	213	304
28/02/2020 15:00	D_1_2	249	322
28/02/2020 16:00	D_1_2	272	244
28/02/2020 17:00	D_1_2	257	243
28/02/2020 18:00	D_1_2	203	162
28/02/2020 19:00	D_1_2	143	98
28/02/2020 20:00	D_1_2	103	83
28/02/2020 21:00	D_1_2	85	49
28/02/2020 22:00	D_1_2	69	71
28/02/2020 23:00	D_1_2	46	26
28/02/2020 00:00	D_2	13	5
28/02/2020 01:00	D_2	10	1
28/02/2020 02:00	D_2	9	2
28/02/2020 03:00	D_2	6	0
28/02/2020 04:00	D_2	4	2
28/02/2020 05:00	D_2	9	7
28/02/2020 06:00	D_2	23	20
28/02/2020 07:00	D_2	39	67
28/02/2020 08:00	D_2	50	52
28/02/2020 09:00	D_2	55	61
28/02/2020 10:00	D_2	61	79
28/02/2020 11:00	D_2	68	81
28/02/2020 12:00	D_2	74	75
28/02/2020 13:00	D_2	79	83
28/02/2020 14:00	D_2	86	98
28/02/2020 15:00	D_2	96	93
28/02/2020 16:00	D_2	102	78
28/02/2020 17:00	D_2	95	93
28/02/2020 18:00	D_2	75	61

28/02/2020 19:00	D_2	53	45
28/02/2020 20:00	D_2	38	30
28/02/2020 21:00	D_2	31	29
28/02/2020 22:00	D_2	25	28
28/02/2020 23:00	D_2	16	11
28/02/2020 00:00	D_3	14	11
28/02/2020 01:00	D_3	14	8
28/02/2020 02:00	D_3	14	3
28/02/2020 03:00	D_3	7	4
28/02/2020 04:00	D_3	10	10
28/02/2020 05:00	D_3	50	63
28/02/2020 06:00	D_3	125	115
28/02/2020 07:00	D_3	195	292
28/02/2020 08:00	D_3	229	253
28/02/2020 09:00	D_3	233	245
28/02/2020 10:00	D_3	239	258
28/02/2020 11:00	D_3	261	284
28/02/2020 12:00	D_3	283	322
28/02/2020 13:00	D_3	285	263
28/02/2020 14:00	D_3	278	310
28/02/2020 15:00	D_3	278	245
28/02/2020 16:00	D_3	283	277
28/02/2020 17:00	D_3	268	293
28/02/2020 18:00	D_3	224	178
28/02/2020 19:00	D_3	165	170
28/02/2020 20:00	D_3	119	113
28/02/2020 21:00	D_3	86	73
28/02/2020 22:00	D_3	56	85
28/02/2020 23:00	D_3	24	45
28/02/2020 00:00	D_4	8	5
28/02/2020 01:00	D_4	6	2
28/02/2020 02:00	D_4	6	0
28/02/2020 03:00	D_4	5	1
28/02/2020 04:00	D_4	9	15
28/02/2020 05:00	D_4	23	29
28/02/2020 06:00	D_4	47	44
28/02/2020 07:00	D_4	71	84
28/02/2020 08:00	D_4	84	91

28/02/2020 09:00	D_4	86	71
28/02/2020 10:00	D_4	85	83
28/02/2020 11:00	D_4	87	84
28/02/2020 12:00	D_4	93	93
28/02/2020 13:00	D_4	97	96
28/02/2020 14:00	D_4	98	99
28/02/2020 15:00	D_4	98	104
28/02/2020 16:00	D_4	98	88
28/02/2020 17:00	D_4	95	65
28/02/2020 18:00	D_4	84	75
28/02/2020 19:00	D_4	66	64
28/02/2020 20:00	D_4	48	40
28/02/2020 21:00	D_4	33	35
28/02/2020 22:00	D_4	21	37
28/02/2020 23:00	D_4	11	31
28/02/2020 00:00	D_5_1	12	8
28/02/2020 01:00	D_5_1	9	3
28/02/2020 02:00	D_5_1	9	2
28/02/2020 03:00	D_5_1	7	3
28/02/2020 04:00	D_5_1	7	2
28/02/2020 05:00	D_5_1	20	27
28/02/2020 06:00	D_5_1	45	38
28/02/2020 07:00	D_5_1	70	90
28/02/2020 08:00	D_5_1	81	103
28/02/2020 09:00	D_5_1	82	83
28/02/2020 10:00	D_5_1	82	76
28/02/2020 11:00	D_5_1	89	86
28/02/2020 12:00	D_5_1	99	96
28/02/2020 13:00	D_5_1	111	137
28/02/2020 14:00	D_5_1	125	153
28/02/2020 15:00	D_5_1	141	153
28/02/2020 16:00	D_5_1	150	133
28/02/2020 17:00	D_5_1	140	110
28/02/2020 18:00	D_5_1	112	97
28/02/2020 19:00	D_5_1	79	73
28/02/2020 20:00	D_5_1	56	44
28/02/2020 21:00	D_5_1	44	39
28/02/2020 22:00	D_5_1	32	36

28/02/2020 23:00	D_5_1	18	27
28/02/2020 00:00	D_5_2	40	31
28/02/2020 01:00	D_5_2	31	22
28/02/2020 02:00	D_5_2	28	14
28/02/2020 03:00	D_5_2	12	6
28/02/2020 04:00	D_5_2	13	14
28/02/2020 05:00	D_5_2	76	77
28/02/2020 06:00	D_5_2	199	189
28/02/2020 07:00	D_5_2	311	468
28/02/2020 08:00	D_5_2	342	414
28/02/2020 09:00	D_5_2	301	294
28/02/2020 10:00	D_5_2	256	291
28/02/2020 11:00	D_5_2	253	213
28/02/2020 12:00	D_5_2	275	224
28/02/2020 13:00	D_5_2	289	315
28/02/2020 14:00	D_5_2	297	307
28/02/2020 15:00	D_5_2	322	293
28/02/2020 16:00	D_5_2	355	318
28/02/2020 17:00	D_5_2	354	347
28/02/2020 18:00	D_5_2	300	227
28/02/2020 19:00	D_5_2	223	232
28/02/2020 20:00	D_5_2	167	153
28/02/2020 21:00	D_5_2	137	138
28/02/2020 22:00	D_5_2	105	141
28/02/2020 23:00	D_5_2	58	105
28/02/2020 00:00	D_6	17	4
28/02/2020 01:00	D_6	11	9
28/02/2020 02:00	D_6	8	1
28/02/2020 03:00	D_6	6	3
28/02/2020 04:00	D_6	7	8
28/02/2020 05:00	D_6	18	15
28/02/2020 06:00	D_6	38	36
28/02/2020 07:00	D_6	61	89
28/02/2020 08:00	D_6	74	86
28/02/2020 09:00	D_6	79	93
28/02/2020 10:00	D_6	83	85
28/02/2020 11:00	D_6	91	95
28/02/2020 12:00	D_6	101	109

28/02/2020 13:00	D_6	111	109
28/02/2020 14:00	D_6	122	155
28/02/2020 15:00	D_6	134	144
28/02/2020 16:00	D_6	145	170
28/02/2020 17:00	D_6	142	148
28/02/2020 18:00	D_6	123	111
28/02/2020 19:00	D_6	97	111
28/02/2020 20:00	D_6	75	93
28/02/2020 21:00	D_6	61	57
28/02/2020 22:00	D_6	49	50
28/02/2020 23:00	D_6	34	38
28/02/2020 00:00	D_7	34	17
28/02/2020 01:00	D_7	26	12
28/02/2020 02:00	D_7	22	8
28/02/2020 03:00	D_7	18	5
28/02/2020 04:00	D_7	26	24
28/02/2020 05:00	D_7	59	70
28/02/2020 06:00	D_7	112	104
28/02/2020 07:00	D_7	163	219
28/02/2020 08:00	D_7	192	232
28/02/2020 09:00	D_7	209	249
28/02/2020 10:00	D_7	233	262
28/02/2020 11:00	D_7	269	280
28/02/2020 12:00	D_7	303	340
28/02/2020 13:00	D_7	325	329
28/02/2020 14:00	D_7	340	342
28/02/2020 15:00	D_7	354	363
28/02/2020 16:00	D_7	357	340
28/02/2020 17:00	D_7	330	314
28/02/2020 18:00	D_7	272	253
28/02/2020 19:00	D_7	205	238
28/02/2020 20:00	D_7	152	167
28/02/2020 21:00	D_7	116	108
28/02/2020 22:00	D_7	85	91
28/02/2020 23:00	D_7	53	54
28/02/2020 00:00	D_8	12	9
28/02/2020 01:00	D_8	7	7
28/02/2020 02:00	D_8	5	3

28/02/2020 03:00	D_8	4	3
28/02/2020 04:00	D_8	5	3
28/02/2020 05:00	D_8	12	11
28/02/2020 06:00	D_8	27	19
28/02/2020 07:00	D_8	45	68
28/02/2020 08:00	D_8	60	72
28/02/2020 09:00	D_8	70	92
28/02/2020 10:00	D_8	79	99
28/02/2020 11:00	D_8	86	103
28/02/2020 12:00	D_8	90	96
28/02/2020 13:00	D_8	89	86
28/02/2020 14:00	D_8	87	85
28/02/2020 15:00	D_8	88	85
28/02/2020 16:00	D_8	92	97
28/02/2020 17:00	D_8	93	97
28/02/2020 18:00	D_8	86	97
28/02/2020 19:00	D_8	72	75
28/02/2020 20:00	D_8	56	47
28/02/2020 21:00	D_8	44	42
28/02/2020 22:00	D_8	34	34
28/02/2020 23:00	D_8	24	25

APPENDIX B

B.1 SUMO flow generator script

```
import os
import sys

# we need to import python modules from the $SUMO_HOME/tools directory
if 'SUMO_HOME' in os.environ:
    tools = os.path.join(os.environ['SUMO_HOME'], 'tools')
    sys.path.append(tools)
else:
    sys.exit("please declare environment variable 'SUMO_HOME'")

from sumolib import checkBinary  # noqa
import traci  # noqa

# function to obtain hourly predictions
def get_hourly_count(detector_id, t):
    with open("FP_hourly_pred_all_detectors.csv", "r") as counts:
        for line in counts:
            date, detector, flow = line.split(",")
            timestamp = date.split(" ")[1]
            if detector_id in detector and timestamp == t:
                return float(flow)

# function to generate SUMO route file
def generate_routefile():
    N = 7200  # number of time steps
    qPKV = 300.00
    print("generate_routefile")
    #
    route_detector_map = {"r_0" : (1.0, "D_1_2"),
                          "r_1" : (1.0, "D_2"),
                          "r_2" : (1.0, "D_1_1"),
                          "r_3" : (0.8, "D_3"),
                          "r_4" : (0.2, "D_3"),
                          "r_5" : (1.0, "D_4"),
                          "r_6" : (1.0, "D_6"),
                          "r_7" : (1.0, "D_5_1"),
                          "r_8" : (1.0, "D_5_2"),
                          "r_9" : (1.0, "D_8"),
                          "r_10": (0.8, "D_7"),
                          "r_11": (0.2, "D_7")}
```

```
with open("lanewise_flow_prediction.rou.xml", "w") as routes:
        print("""<routes>
                 <!-- Routes -->
                 <route id="r_0" edges="209506777#0 231111841 231111849#0
                 32555580#0"/>
                 <route id="r_1" edges="209506777#0 231111841 231111849#0 -
                 238185352 -231111851"/>
                 <route id="r_2" edges="209506777#0 231111841 231111849#0 -
                 238185357 -231111854#1 -231111853 -236596230#7"/>
                 <route id="r_3" edges="231111851 231111852#0 -238185357 -
                 231111854#1 -231111853 -236596230#7"/>
                 <route id="r_4" edges="231111851 231111852#0 62194327
                 238185349 231111838"/>
                 <route id="r_5" edges="231111851 231111852#0 32555580#0"/>
                 <route id="r_6" edges="231111846#1 231111848 231111842
                 231111850#0 -238185357 -231111854#1 -231111853 -
                 236596230#7"/>
                 <route id="r_7" edges="231111846#0 231111846#1 231111848
                 231111842 231111850#0 -238185352"/>
                 <route id="r_8" edges="231111846#0 231111846#1 231111848
                 231111842 231111850#0 62194327 238185349 231111838"/>
                 <route id="r_9" edges="236596230#0 231111853 231111854#0
                 238185357 62194327 238185349 231111838"/>
                 <route id="r_10" edges="236596230#0 231111853 231111854#0
                 238185357 -238185352 -231111851"/>
                 <route id="r_11" edges="236596230#0 231111853 231111854#0
                 238185357 32555580#0"/>
                 <!-- Flow -->""", file=routes)
        i = 0
        time = "17:00"
        for route in route_detector_map:
            factor, det = route_detector_map[route]
            print(route, factor, det)
            print('    <flow id="f_%s_%d" begin="0.00" route="%s" end="3600"
            vehsPerHour="%i"/>'
            % (route, i, route, factor * get_hourly_count(det, time)),
            file=routes)
        i += 1
        print("</routes>", file=routes)

# this is the main entry point of this script
if __name__ == "__main__":
    generate_routefile()
```

B.2 SUMO runner script

```
import os
import sys
import traci
import sumolib
import simpla

if "SUMO_HOME" in os.environ:
    sys.path.append(os.path.join(os.environ['SUMO_HOME'], 'tools'))

binary = 'sumo-gui'
if 'nogui' in sys.argv:
    binary = 'sumo'

traci.start([sumolib.checkBinary(binary),
             '-c', 'osm.sumocfg',
             '--step-length', '0.5',
             '--fcd-output', 'fcd.xml',
             '--fcd-output.max-leader-distance', '100',
             '--tripinfo-output', 'tripinfo.xml',
             '--queue-output', 'queueoutput.xml'])

simpla.load("simpla.cfg.xml")

while traci.simulation.getMinExpectedNumber() > 0:
    traci.simulationStep()

traci.close()
```

B.3 SUMO configuration file

```xml
<?xml version="1.0" encoding="UTF-8"?>

<!-- generated on 2023-04-03 17:47:29 by Eclipse SUMO sumo Version 1.16.0

<route-files value="lanewise_flow_prediction.rou.xml"/>
<route-files value="manual_route.rou.xml"/>

-->

<configuration xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="http://sumo.dlr.de/xsd/sumoConfiguration.xsd">

    <input>
        <net-file value="osm_tls.net.xml"/>
        <route-files value="lanewise_flow_prediction.rou.xml"/>
        <additional-files value="detector.add.xml"/>
    </input>

    <processing>
        <ignore-route-errors value="true"/>
    </processing>

    <routing>
        <device.rerouting.adaptation-steps value="18"/>
        <device.rerouting.adaptation-interval value="10"/>
    </routing>

    <report>
        <verbose value="true"/>
        <duration-log.statistics value="true"/>
        <no-step-log value="true"/>
    </report>

    <gui_only>
        <gui-settings-file value="osm.view.xml"/>
    </gui_only>

</configuration>
```

B.4 SUMO Simpla plugin configuration file

```
<configuration>
    <vTypeMap original="CAV" leader="t_leader" follower="t_follower"
catchup="t_catchup" catchupFollower="t_catchupFollower" />
    <!-- controll all vehicles -->
    <vehicleSelectors value="CAV" />
    <verbosity value="4" ></verbosity>
    <controlRate value="1" />
    <maxPlatoonGap value="15" />
    <catchupDist value="100" />
    <useHeadway value="false" />
    <!-- deactivated -->
    <switchImpatienceFactor value="-1" />
    <platoonSplitTime value="3" ></platoonSplitTime>
    <lcMode original="594" leader="594" follower="514" catchupFollower="514"/>
    <speedFactor original="-1" leader="1" follower="1.2" catchup="1.2"/>
</configuration>
```

B.5 SUMO route file

```
<routes>
    <!-- Routes -->
      <route id="r_0" edges="209506777#0 231111841 231111849#0 32555580#0"/>
      <route id="r_1" edges="209506777#0 231111841 231111849#0 -238185352 -
      231111851"/>
      <route id="r_2" edges="209506777#0 231111841 231111849#0 -238185357 -
      231111854#1 -231111853 -236596230#7"/>
      <route id="r_3" edges="231111851 231111852#0 -238185357 -231111854#1 -
      231111853 -236596230#7"/>
      <route id="r_4" edges="231111851 231111852#0 62194327 238185349
      231111838"/>
      <route id="r_5" edges="231111851 231111852#0 32555580#0"/>
      <route id="r_6" edges="231111846#1 231111848 231111842 231111850#0 -
      238185357 -231111854#1 -231111853 -236596230#7"/>
      <route id="r_7" edges="231111846#0 231111846#1 231111848 231111842
      231111850#0 -238185352"/>
      <route id="r_8" edges="231111846#0 231111846#1 231111848 231111842
      231111850#0 62194327 238185349 231111838"/>
      <route id="r_9" edges="236596230#0 231111853 231111854#0 238185357
      62194327 238185349 231111838"/>
      <route id="r_10" edges="236596230#0 231111853 231111854#0 238185357 -
      238185352 -231111851"/>
      <route id="r_11" edges="236596230#0 231111853 231111854#0 238185357
      32555580#0"/>

    <!-- VTypes -->
    <vType id="car" color="cyan" maxSpeed="50" >
        <param key="has.emissions.device" value="true"/>
    </vType>

    <vType id="bus" color="magenta" maxSpeed="50" vClass="bus"/>
    <vType id="CAV" color="white" maxSpeed="50">
        <param key="has.glosa.device" value="false"/>
        <param key="has.emissions.device" value="true"/>
    </vType>
    <vType id="t_catchup" minGap="0.50" color="red" tau="0.5" maxSpeed="50"/>
     <vType id="t_catchupFollower" minGap="0.50" color="blue" tau="0.5"
     maxSpeed="50"/>
     <vType id="t_follower" minGap="0.50" color="green" tau="0.5"
     maxSpeed="50"/>
    <vType id="t_leader" color="yellow" maxSpeed="50"/>

    <!-- Flow -->
     <flow id="f_r_0_0"    type="car" begin="0.00" route="r_0"  end="3600"
     vehsPerHour="245"/>
     <flow id="f_r_0_bus"  type="bus" begin="0.00" route="r_0"  end="3600"
     vehsPerHour="12"/>
```

```
    <flow id="f_r_1_0"    type="car" begin="0.00" route="r_1"  end="3600"
    vehsPerHour="95"/>
    <flow id="f_r_2_0"    type="car" begin="0.00" route="r_2"  end="3600"
    vehsPerHour="257"/>
    <flow id="f_r_3_0"    type="car" begin="0.00" route="r_3"  end="3600"
    vehsPerHour="54"/>
    <flow id="f_r_3_cav"  type="CAV" begin="0.00" route="r_3"  end="3600"
    vehsPerHour="160"/>
    <flow id="f_r_4_0"    type="car" begin="0.00" route="r_4"  end="3600"
    vehsPerHour="13"/>
    <flow id="f_r_4_cav"  type="CAV" begin="0.00" route="r_4"  end="3600"
    vehsPerHour="40"/>
    <flow id="f_r_5_0"    type="car" begin="0.00" route="r_5"  end="3600"
    vehsPerHour="95"/>
    <flow id="f_r_6_0"    type="car" begin="0.00" route="r_6"  end="3600"
    vehsPerHour="142"/>
    <flow id="f_r_7_0"    type="car" begin="0.00" route="r_7"  end="3600"
    vehsPerHour="140"/>
    <flow id="f_r_8_0"    type="car" begin="0.00" route="r_8"  end="3600"
    vehsPerHour="337"/>
    <flow id="f_r_8_bus"  type="bus" begin="0.00" route="r_8"  end="3600"
    vehsPerHour="17"/>
    <flow id="f_r_9_0"    type="car" begin="0.00" route="r_9"  end="3600"
    vehsPerHour="93"/>
    <flow id="f_r_10_0"   type="car" begin="0.00" route="r_10" end="3600"
    vehsPerHour="251"/>
    <flow id="f_r_10_bus" type="bus" begin="0.00" route="r_10" end="3600"
    vehsPerHour="13"/>
    <flow id="f_r_11_0"   type="car" begin="0.00" route="r_11" end="3600"
    vehsPerHour="66"/>
</routes>
```

REFERENCES

(2023, April). Retrieved from Vehicle Type Parameter Defaults: https://sumo.dlr.de/docs/Vehicle_Type_Parameter_Defaults.html

(2023, April). Retrieved from Car-Following Models: https://sumo.dlr.de/docs/Car-Following-Models.html

(2023). Retrieved from HBEFA: http://www.hbefa.net/e/index.html

Akin, D., Sisiopiku, V. P., & Skabardonis, A. (2011). Impacts of Weather on Traffic Flow Characteristics of Urban Freeways in Istanbul. *Procedia - Social and Behavioral Sciences, 16*, 89-99. doi:10.1016/j.sbspro.2011.04.432

Al-Selwi, H. F., Aziz, A. B., Abas, F. S., Hamzah, N. A., & Mahmud, A. B. (2022). The impact of weather data on traffic flow prediction models. *IAES International Journal of Artificial Intelligence (IJ-AI), 11*(4), 1223-1231. doi:10.11591/ijai.v11.i4.pp1223-1231

Ambühl, L., Menendez, M., & González, M. C. (2022). Traffic Simulations for Boston, Lisbon, Los Angeles, Rio de Janeiro, San Francisco. ETH Zurich. Retrieved from https://www.research-collection.ethz.ch/handle/20.500.11850/584669

Armellini, M. G., & Bieker-Walz, L. (2020). Simulation of a Demand Responsive Transport feeder system: A case study of Brunswick. *SUMO User Conference 2020*, (p. Vol.1). doi:https://doi.org/10.52825/scp.v1i.101

Barros, J., Araujo, M., & Rossetti, R. J. (2015). Short-term real-time traffic prediction methods: A survey. *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)* (pp. 132-139). Budapest, Hungary: IEEE. doi:10.1109/MTITS.2015.7223248

Becker, N., Rust, H. W., & Ulbrich, U. (2022). Modeling hourly weather-related road traffic variations for different vehicle types in Germany. *European Transport Research Review, 14*(1). doi:10.1186/s12544-022-00539-0

Berlin Sumo Traffic Scenario. (2022, Apr). Retrieved from https://github.com/mosaic-addons/best-scenario

Bieker, L., Krajzewicz, D., Morra, A. P., Michelacci, C., & Cartolano, F. (2014). Traffic simulation for all: a real world traffic scenario from the city of Bologna. *SUMO 2014*, (pp. 15-16). Berlin. doi: 10.1007/978-3-319-15024-6_4

BL., S., BM, W., & RK, O. (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, *10*(4), 303-321. Retrieved from https://doi.org/10.1016/S0968-090X(02)00009-8

Box, & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day.

Calvert, S., & Snelder, M. (2016). Influence of Weather on Traffic Flow: an Extensive Stochastic Multi-effect Capacity and Demand Analysis. *European Transport\Trasporti Europei*.

Cetiner, B., Sari, M., & Borat, O. (2010). A Neural Network Based Traffic-Flow Prediction Model. *Mathematical and Computational Applications*, *15*(2), 269-278. Retrieved from https://doi.org/10.3390/mca15020269

Chair of Mechatronics, University of Duisburg-Essen. (2023, September). Retrieved from https://www.uni-due.de/mechatronik/index_en.php

Chen, C., Hu, J., Meng, Q., & Zhang, Y. (2011). Short-time traffic flow prediction with ARIMA-GARCH model. *2011 IEEE Intelligent Vehicles Symposium (IV).* Baden-Baden, Germany.

Chen, K., & Underwood, S. (1991). Research on anticipatory route guidance. *Vehicle Navigation and Information Systems Conference* (pp. 427-439). Troy, MI, USA: IEEE. doi:10.1109/VNIS.1991.205790

Chikkakrishna, N., Hardik, C., Deepika, K., & Sparsha, N. (2019). Short-Term Traffic Prediction Using Sarima and FbPROPHET. *IEEE 16th India Council International Conference (INDICON)*, (pp. 1-4). Rajkot, India. doi:10.1109/INDICON47234.2019.9028937

ChikkaKrishna, N., Rachakonda, P., & Tallam, T. (2022). Short - Term Traffic Prediction Using Fb-PROPHET and Neural-PROPHET. *IEEE Delhi Section Conference (DELCON)*, (pp. 1-4). New Delhi. doi:10.1109/DELCON54057.2022.9753459

Chmiel, W., Dańda, J., Dziech, A., Ernst, S., Kadłuczka, P., Mikrut, Z., . . . Wojnicki, I. (2016). INSIGMA: an intelligent transportation system for urban mobility enhancement. *Multimed Tools Appl, 75*, 10529–10560. doi:10.1007/s11042-016-3367-5

Climate data for direct download. (2020). Retrieved Mar 2023, from https://www.dwd.de/EN/ourservices/cdc/cdc_ueberblick-klimadaten_en.html

Codeca, L., & Härri, J. (2017). Towards Multimodal Mobility Simulation of C-ITS: The Monaco SUMO Traffic Scenario. *VNC 2017, IEEE Vehicular Networking Conference.*

Codeca, L., Frank, R., & Engel, T. (2015). Luxembourg SUMO Traffic (LuST) Scenario: 24 Hours of Mobility for Vehicular Networking Research. *7th IEEE Vehicular Networking Conference (VNC15).*

David, H., & Plötz, P.-A. (2020). Test track for automated and connected driving in Hamburg (TAVF-HH). *Virtual ITS European Congress.*

Deng, S., Jia, S., & Chen, J. (2019). Exploring spatial–temporal relations via deep convolutional neural networks for traffic flow prediction with incomplete data. *Applied Soft Computing, 78*, 712-721. doi:10.1016/j.asoc.2018.09.040

Dong, H., Jia, L., Sun, X., Li, C., & Qin, Y. (2009). Road Traffic Flow Prediction with a Time-Oriented ARIMA Model. *2009 Fifth International Joint Conference on INC, IMS and IDC*, (pp. 1649-1652). Seoul, Korea (South).

Dong, J., Mahmassani, H. S., & Lu, C.-C. (2006). How Reliable is this Route?: Predictive Travel Time and Reliability for Anticipatory Traveler Information Systems. *Transportation Research Record, 1980*(1), 117-125. doi:10.1177/0361198106198000116

Duisburg. (2023, June 19). Retrieved from Wohnen un d Leben: https://duisburg.de/wohnenleben/index.php#:~:text=Duisburg%20ist%20die%20westlichste%20Gro%C3%9Fstadt,Einwohnern%2015%2Dgr%C3%B6%C3%9Fte%20Stadt%20Deutschlands.

Duisburg in Zahlen. (2023). Retrieved June 2023

Duisburg Verkehr: Wikipedia. (2023). Retrieved June 2023, from https://de.wikipedia.org/wiki/Duisburg#Verkehr

Duisburg: Verkehrkonzepte. (2023). Retrieved June 2023, from https://duisburg.de/microsites/pbv/verkehr/verkehrskonzepte.php

Eclipse Homepage: SUMO Simulator. (2023). Retrieved June 2023, from https://eclipse.dev/sumo/

Erdmann, J. (2015). SUMO’s Lane-changing model. *LECTURE NOTES IN CONTROL AND INFORMATION SCIENCES, 2nd SUMO User Conference* (pp. 105-123). Berlin: Springer Verlag. Retrieved from https://elib.dlr.de/102254/

Esugo, M., Lu, Q., & Haas, O. (2022). Short-Term Traffic Flow Forecasting A wide and deep approach with periodic feature selection. *TechRxiv*. doi:https://doi.org/10.36227/techrxiv.19763008.v1

European Commision: A Europe fit for the digital age. (2023). Retrieved July 2023, from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age_en

European Commission: Mobility and Transport. (2023, June 20). Retrieved from Intelligent transport systems: https://transport.ec.europa.eu/transport-themes/intelligent-transport-systems_en

Federal Minstry of Digital and Transport: ITS. (2023). Retrieved July 2023, from https://bmdv.bund.de/SharedDocs/EN/Articles/DG/intelligent-transport-systems.html

Feng, B., Xu, J., Lin, Y., & Li, P. (2020). A Period-Specific Combined Traffic Flow Prediction Based on Travel Speed Clustering. *IEEE Access, 85880-85889*, 8. doi:10.1109/ACCESS.2020.2992657

Förster, D., Löhr, H., Grätz, A., Petit, J., & Kargl, F. (2017). An Evaluation of Pseudonym Changes for Vehicular Networks in Large-scale, Realistic Traffic Scenarios. *IEEE Transactions on Intelligent Transportation Systems*, 3400-3405. doi: 10.1109/TITS.2017.2775519

Fu, T., Stipancic, J., Zangenehpour, S., Miranda-Moreno, L., & Saunier, N. (2017). Automatic Traffic Data Collection under Varying Lighting and Temperature Conditions in Multimodal Environments: Thermal versus Visible Spectrum Video-Based Systems. *Journal of Advanced Transportation*. doi:10.1155/2017/5142732

George, E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day.

Ghosh, B., Basu, B., & O'Mahony, M. (2009). Multivariate Short-Term Traffic Flow Forecasting Using Time-Series Analysis. *IEEE Transactions on Intelligent Transportation Systems, 10*(2), 246-254. doi:10.1109/TITS.2009.2021448

Gueriau, M., & Dusparic, I. (2020). Quantifying the impact of connected and autonomous vehicles on traffic efficiency and safety in mixed traffic. *23rd IEEE International Conference on Intelligent Transportation Systems (ITSC 2020).*

Hardes, T., & Sommer, C. (2019). Dynamic Platoon Formation at Urban Intersections. *2019 IEEE 44th Conference on Local Computer Networks (LCN)* (pp. 101-104). Osnabrueck, Germany: IEEE. doi:10.1109/LCN44214.2019.8990846

HBEFA. (2022). Retrieved April 2023, from http://www.hbefa.net/e/index.html

Hidas, P., & Wagner, P. (2004). Review of Data Collection Methods for Microscopic Traffic Simulation. *Proceedings of World Conference on Transport Research (WCTR).* Istanbul, Turkey.

Hoyer, R., Feldges, M., Kemper, C., Kochs, A., Leonhardt, A., Listl, G., . . . Wallas, D. (April 2012). *Hinweise zur Verkehrprognose in strassenverkehrstechnischen Anwendungen* (Vol. FGSV 382/1 W1). Cologne: Forschungsgesellschaft fuer Strassen- und Verkehrswesen (FGSV).

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting principles and practice 2nd Edition.* Otexts.

Intelligent Transportation Systems: Joint Program Office. (2023, June 29). Retrieved from https://www.its.dot.gov/

ITS JPO Strategic Plan 2020-2025. (2020, May 06). Retrieved from https://www.its.dot.gov/stratplan2020/ITSJPO_StrategicPlan_2020-2025.pdf

Kashyap, A. A., Ravira, S., Devarakonda, A., K, S. R., V, S. K., & Bhat, S. J. (2022). Traffic flow prediction models – A review of deep learning techniques. *Cogent Engineering, 9*(1). doi:10.1080/23311916.2021.2010510

Kato, T., Kim, Y., Suzuki, T., & Okuma, S. (2005). Model Predictive Control of Traffic Flow Based on Hybrid System Modeling. *IEICE Transactions on Fundamentals* *of Electronics Communications and Computer Sciences*, *88-A*(2), 549-560. doi:10.1093/ietfec/E88-A.2.549

Kenneth, D., & Ronald, K. (1982). Advances in business and management forecasting. *Bingley, UK: Emerald Books*.

Khaz'ali, A. R., Emamjomeh, A., & Andayesh, M. (2011). An Accuracy Comparison Between Artificial Neural Network and Some Conventional Empirical Relationships in Estimation of Relative Permeability. *Petroleum Science and Technology*, *29*(15), 1603-1614. Retrieved from https://doi.org/10.1080/10916461003620461

Klein, L., Kelley, M., & Mills, M. (1997). Evaluation of Overhead and In-Ground Vehicle Detector Technologies for Traffic Flow Measurement. *Journal of Testing and Evaluation*, *25*(2), 205-214.

Krajzewicz, D., Erdmann, J., Behrisch, M., & Bieker-Walz, L. (2012). Recent Development and Applications of SUMO - Simulation of Urban MObility. *International Journal On Advances in Systems and Measurements*, *5*(3&4), 128-138.

Krauss, S. (1998, April). *Microscopic Modeling of Traffic Flow: Investigation of Collision Free Vehicle Dynamics.* Retrieved from https://sumo.dlr.de/pdf/KraussDiss.pdf

Krauß, S., Wagner, P., & Gawron, C. (1997). Metastable States in a Microscopic Model of Traffic Flow. *Physical Review*, 5597-5602.

Kumar, S. V., & Vanajakshi, L. (2015). Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *European Transport Research Review* , *7 Article number: 21*.

Kušić, K., Schumann, R., & Ivanjko, E. (2023). A digital twin in transportation: Real-time synergy of traffic data streams and simulation for virtualizing motorway dynamics. *Advanced Engineering Informatics*. doi:10.1016/j.aei.2022.101858

Laña, I., Ser, J. D., & Olabarrieta, I. I. (2016). Understanding daily mobility patterns in urban road networks using traffic flow analytics. *NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium*, (pp. 1157-1162). Istanbul, Turkey,. doi:10.1109/NOMS.2016.7502980

Lana, I., Ser, J. D., Velez, M., & Vlahogianni, E. I. (2018). Road Traffic Forecasting: Recent Advances and New Challenges. *IEEE Intelligent Transportation Systems Magazine (Volume: 10, Issue: 2)*, 93 - 109. doi:10.1109/MITS.2018.2806634

Leduc, G. (2008). *Road Traffic Data: Collection Methods and Applications.* Working papers on Energy, Transport and Climate change, JRC Technical notes: European Commission 1(55).

Li, Q., Tan, H., Wu, Y., Ye, L., & Ding, F. (2020). Traffic Flow Prediction With Missing Data Imputed by Tensor Completion Methods. *IEEE Access, 8*, 63188-63201. doi:10.1109/ACCESS.2020.2984588

Liang, Z., & Wakahara, Y. (2013). City traffic prediction based on real-time traffic information for Intelligent Transport Systems. *2013 13th International Conference on ITS Telecommunications (ITST),* (pp. 378-383). Tampere, Finland. doi:10.1109/ITST.2013.6685576

Liang, Z., Flötteröd, Y.-P., Chen, H., Sohr, A., Bei, X., Bottazzi, M., & Trumpold, J. (2018). Real-Time Microscopic Traffic Simulation and. *International Journal of Electrical and Electronic Engineering & Telecommunications, 1*, 30-37.

List of Holidays in NRW. (2019). Retrieved March 2023, from https://www.officeholidays.com/countries/germany/north-rhine-westphalia/2020

Listl, G., Beck, H.-P., Boelling, F., Custodis, F., Estel, A., Feldges, M., . . . Traeger, R. (Oktober 2019). *Hinweise zur Detektionstechynologien im Strassenverkehr* (Vol. FGSV 312 W1). Cologne: Forschungsgesellschaft fuer Strassen- und Verkehrswesen (FGSV).

Liu, S.-y., Li, D.-w., Xi, Y.-g., & Tang, Q.-f. (2015). A short-term traffic flow forecasting method and its applications. *J. Shanghai Jiaotong Univ. (Sci.), 20*, 156-163. doi:10.1007/s12204-015-1604-0

Lobo, S., Neumeier, S., Fernandez, E., & Facchi, C. (2020). InTAS - The Ingolstadt Traffic Scenario for SUMO. *SUMO User Conference Proceedings, 2020,*. doi:10.52825

Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., . . . Wiessner, E. (2018). Microscopic Traffic Simulation using SUMO. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, (pp. 2575-2582). Maui, HI, USA.

Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. -Y. (2015). Traffic Flow Prediction With Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems, 16*(2), 865-873. doi:10.1109/TITS.2014.2345663

Ma, X. (2020). *Effects of Vehicles with Different Degrees of Automation on Traffic Flow in Urban Areas.* Duisburg: Doctoral Thesis, University of Duisburg-Essen.

Ma, X. (2020). *Effects of Vehicles with Different Degrees of Automation on Traffic Flow in Urban Areas.* Duisburg: Doctoral dissertation: University of Duisburg-Essen.

Maas, N., Benjamin, H., Martin, K., & Schramm, D. (2014). Simulator setup according to use case scenarios-A human-oriented method for virtual development. *2014 IEEE/ASME 10th International Conference on Mechatronic* (pp. 1-6). Senigallia, Italy: IEEE. doi:10.1109/MESA.2014.6935585

Magenta Software Lab: Vehicle Counter. (2019). Retrieved from https://www.magentalab.it/?page_id=189

Martínez-Díaz, M., Al-Haddad, C., Soriguera, F., & Antoniou, C. (2021). Platooning of connected automated vehicles on freeways: a bird's eye view. *Transportation Research Procedia, 58*, 479-486. doi:10.1016/j.trpro.2021.11.064

MCME V2 User Manual, 2012. (n.d.). Retrieved April 2014, from COEUT: http://coeut.iitm.ac.in/MCME%20User%20Manual.pdf

Minge, E., Peterson, S., & Kotzenmacher, J. (2011). Evaluation of Nonintrusive Technologies for Traffic Detection, Phase 3. *Transportation Research Record, 2256*(1), 95-103.

Mousa, S. R., & Bonnette, J. (2016). Impact of Geometric, Traffic, and Environmental Conditions on Video Detection Performance: A Case Study in Baton Rouge Metropolitan Area. *Journal of Traffic and Logistic Engineering, 4*(2), 160-166.

Nagel, K., & Schreckenberg, M. (1992). A Cellular Automaton Model for Freeway Traffic. *Journal de Physique I 2(12)*, 2221-2229.

NeuralProphet. (2021). Retrieved from https://neuralprophet.com/contents.html

NeuralProphet: Uncertainty. (2023). Retrieved July 2023, from https://neuralprophet.com/tutorials/tutorial08.html

Oh, S.-d., Kim, Y.-j., & Hong, J.-s. (2015). Urban Traffic Flow Prediction System Using a Multifactor Pattern Recognition Model. *IEEE Transactions on Intelligent Transportation Systems, 16*(5), 2744-2755. doi:10.1109/TITS.2015.2419614

Openstreetmap. (2023). Retrieved June 06, 2023, from https://www.openstreetmap.org/note/3711603#map=16/51.4342/6.7950&layers=HN

Osipov, V., Nikiforov, V., & Zhukova, N. (2020). Urban traffic flows forecasting by recurrent neural networks with spiral structures of layers. *Neural Computing and*

Applications, *32*, 14885–14897. Retrieved from https://doi.org/10.1007/s00521-020-04843-5

OSM Webwizard. (2023). Retrieved from https://sumo.dlr.de/docs/Tools/Import/OSM.html

OSMWebWizard SUMO. (2023, Januray). Retrieved from https://sumo.dlr.de/docs/Tutorials/OSMWebWizard.html

Ponnaluri, R., & Alluri, P. (2021). *Connected and Automated Vehicles-Developing Policies, Designing Programs, and Deploying Projects: From Policy to Practice.* Elsevier. Retrieved January 2023

Pragalatahan, J., & Schramm, D. (May 2023). Applications of traffic flow predictions and simulations in urban mobility with connected automated vehicles and advanced traffic management systems. *15. Wissenschaftsforum Mobilität "Next Chapter in Mobility".* Duisburg, Germany: Springer.(in press)

Pragalathan, J., & Schramm, D. (2019). Quality Assessment of Long-Term Traffic Data Collection by Video Detection Technique in a Busy Urban Intersection. *Journal of Traffic and Logistics Engineering*, *7*(2), 35-40. doi:10.18178/jtle.7.2.35-40

Pragalathan, J., & Schramm, D. (2024). Urban traffic flow predictions with impacts of weather and holidays. ICITT 2023, Madrid, Spain. *Advances in transdisciplinary engineering* 274-281. IOS Press. doi: 10.3233/ATDE240042

Pragalathan, J., & Schramm, D. (Apr 2023). Comparison of SARIMA, Fb-Prophet and Neural Prophet Models for Traffic Flow Predictions at a Busy Urban Intersection. *Sustainable Infrastructure: Innovations, Opportunities and Challenges-2023 (SIIOC-2023).* Springer. (in press)

Prophet: Additional Topics. (2022). Retrieved November 2022, from https://facebook.github.io/prophet/docs/additional_topics.html

PWC: Digital Twins. (2023). Retrieved July 2023, from https://www.pwc.com/m1/en/publications/documents/how-digital-twins-can-make-smart-cities-better.pdf

Qu, X., Li, X., Wang, M., & Dixit, V. (2017, 11 19). Advances in Modelling Connected and Automated Vehicles. *Journal of Advanced Transportation, 2017*(Article ID 3268371), 0197-6729. doi:10.1155/2017/3268371

Radhakrishnan, S., & Menon, L. (2019, May). *COEUT IITM.* Retrieved from https://coeut.iitm.ac.in/MCME%20User%20Manual.pdf

Rapelli, M., Casetti, C., & Gagliardi, G. (2021). Vehicular Traffic Simulation in the City of Turin from Raw Data. *IEEE Transactions on Mobile Computing.*

Rules for ARIMA. (2022). Retrieved Oct 2022, from https://people.duke.edu/~rnau/arimrule.htm

Schoitsch, E. (2016). Autonomous vehicles and automated driving status, perspectives and societal impact. *IDIMT Proceedings 2016*, (pp. 405-423).

Shaping Europe's digital future. (2022, June). Retrieved August 2023, from https://digital-strategy.ec.europa.eu/en/policies/connected-and-automated-mobility

Sippl, C., Schwab, B., Kielar, P., & Djanatliev, A. (2018). Distributed Real-Time Traffic Simulation for Autonomous Vehicle Testing in Urban Environments. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 2562-2567). Maui, HI, USA: IEEE. doi:10.1109/ITSC.2018.8569544

Skszek, S. L. (2001). *"State-of-the-Art" Report on Non-Traditional Traffic Counting Methods.* Retrieved May 2019, from https://rosap.ntl.bts.gov/view/dot/14764

Smith, B. L., Williams, B. M., & Oswald, R. K. (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies, 10*(4), 303-321.

Somasekhar, G., Shirabadagi, S., & Hegadi, R. S. (2014). High density traffic management using image background subtraction algorithm. *International Journal of Computer Applications*, 10-15.

Song, Z., Guo, Y., Wu, Y., & Ma, J. (2019). Short-term traffic speed prediction under different data collection time intervals using a SARIMA-SDGM hybrid prediction model. *PLoS ONE, 14*(6). Retrieved from https://doi.org/10.1371/journal.pone.0218626

Su, F., Dong, H., Jia, L., Qin, Y., & Tian, Z. (2016). Long-term forecasting oriented to urban expressway traffic situation. *Advances in Mechanical Engineering, 8*(1). doi:10.1177/1687814016628397

SUMO Proceedings. (2016). *2016 SUMO Conference Proceedings.* Berlin. Retrieved from https://elib.dlr.de/106342/1/SUMOconference_proceedings_2016.pdf

SUMO User Documentation. (2023). Retrieved February 2023, from https://sumo.dlr.de/docs/index.html

SUMO: Car-Following Models. (2023). Retrieved April 2023, from https://sumo.dlr.de/docs/Car-Following-Models.html

SUMO: Simpla. (2023). Retrieved April 2023, from https://sumo.dlr.de/docs/Simpla.html

SUMO: Vehicle Type Parameter Defaults. (2023). Retrieved April 2023, from https://sumo.dlr.de/docs/Vehicle_Type_Parameter_Defaults.html

Sun, S., Wu, H., & Xiang, L. (2020). City-Wide Traffic Flow Forecasting Using a Deep Convolutional Neural Network. *Sensors*, *20*(2), 421.

TAPASCologne. (2023, September). Retrieved Jan 2023, from https://sumo.dlr.de/docs/Data/Scenarios/TAPASCologne.html

Taylor, S., & Letham, B. (2017). Forecasting at scale. *PeerJ Preprints, 5*(e3190v2). Retrieved from https://doi.org/10.7287/peerj.preprints.3190v2

Tian, Y., & Pan, L. (2015). Predicting Short-Term Traffic Flow by Long Short-Term Memory Recurrent Neural Network. *International Conference on Smart City/SocialCom/SustainCom(SmartCity)* (pp. 153-158). Chengdu, China: IEEE. doi:10.1109/SmartCity.2015.63

Trapp, R., Arbter, B., Baier, M. M., Baumert, R., Breser, C., Harding, J., . . . Wagner, P. (December 2006). *Hinweise zur mikroskopischen Verkehrsflussisimulation-Grundlagen und Anwendung* (Vol. FGSV 388 W1). Cologne: Forschungsgesellschaft fuer Strassen- und Verkehrswesen (FGSV).

Triebe, O., Hewamalage, H., Pilyugina, P., Laptev, N., Bergmeir, C., & Rajagopal, R. (2021). NeuralProphet: Explainable Forecasting at Scale. *arXiv*. doi:10.48550/ARXIV.2111.15397

Triebe, O., Laptev, N., & Rajagopal, R. (2019). AR-Net: A simple Auto-Regressive Neural Network for time-series. *CoRR, abs/1911.12436*.

Tselentis, D. I., Vlahogianni, E. I., & Karlaftis, M. G. (2015). Improving short-term traffic forecasts: to combine models or not to combine? *IET Intelligent Transport Systems, 9*(2), 193-201.

Uhlemann, E. (March 2015). Introducing Connected Vehicles [Connected Vehicles]. *IEEE Vehicular Technology Magazine, 10*(1), 23-31. doi:10.1109/MVT.2015.2390920

Van Der Voort, M., Dougherty, M., & Watson, S. (1996). Combining kohonen maps with arima time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies, 4*(5), 307-318.

Verkehrssimulation Software PTV Vissim. (2023, January). Retrieved from https://www.ptvgroup.com/de/produkte/ptv-vissim

Vlahogianni, E. I., & Karlaftis, M. G. (2013). Testing and Comparing Neural Network and Statistical Approaches for Predicting Transportation Time Series.

Transportation Research Record: Journal of the Transportation Research Board, 2399(1), 9-22.

Wahle, J., Chrobok, R., Pottmeier, A., & Schreckenberg, M. (2002). A Microscopic Simulator for Freeway Traffic. *Networks and Spatial Economics, 2*, 371-386.

Wang, Y., Li, L., & Xu, X. (2017). A Piecewise Hybrid of ARIMA and SVMs for Short-Term Traffic Flow Prediction. *International Conference on Neural Information Processing*, (pp. 493-502). Guangzhou, China.

Weber, D., Chrobok, R., Hafstein, S., Mazur, F., Pottmeier, A., & Schreckenberg, M. (2004). OLSIM: Inter-urban Traffic Information. *Innovative Internet Community Systems, IICS 2004* (pp. 296-306). Mexico: Springer.

Weber, T. (2021). *On the Potential of a Weather-Related Road Surface Condition Sensor Using an Adaptive Generic Framework in the Context of Future Vehicle Technology*. Duisburg: Doctoral Thesis, University of Duisburg-Essen.

Website: Google Maps. (2019). Retrieved June 2019, from http://www.google.co.in/maps

Williams, B. (2001). Multivariate Vehicular Traffic Flow Prediction: Evaluation of ARIMAX Modeling. *Transportation Research Record 1776*, 194-200.

Williams, B. M., & Hoel, L. A. (2003). Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. *Journal of Transportation Engineering, 129*(6), 664-672. doi:10.1061/(ASCE)0733-947X(2003)129:6(664)

Wirtschaftsbetreibe Duisburg, W. (2019). *Traffic Signal Light Programs.*

Xu, C., Li, Z., & Wang, W. (2016). Short-term traffic flow prediction using a methodology based on autoregressive integrated moving average and genetic programming. *Transport, 31*(3), 343–358.

Xu, F., He, Z., Sha, Z., Zhuang, L., & Sun, W. (2013). Assessing the Impact of Rainfall on Traffic Operation of Urban Road Network. *Procedia - Social and Behavioral Sciences, 96*, 82-89. doi:10.1016/j.sbspro.2013.08.012

Yang, D., Li, S., Peng, Z., Wang, P., Wang, J., & Yang, H. (2019). MF-CNN: Traffic Flow Prediction Using Convolutional Neural Network and Multi-Features Fusion. *IEICE TRANSACTIONS on Information and Systems, E102-D*(8), 1526-1536. doi:10.1587/transinf.2018EDP7330

Yuan-qing, W., & Jing, L. (2017). Study of Rainfall Impacts on Freeway Traffic Flow Characteristics. *Transportation Research Procedia, 25*, 1533-1543. doi:10.1016/j.trpro.2017.05.180

Zhang, G., Avery, R. P., & Wang, Y. (2007). Video-Based Vehicle Detection and Classification System for Real-Time Traffic Data Collection Using Uncalibrated Video Cameras. *Journal of Trabnsportation Research Record, 1993*(1). doi:10.3141/1993-19

Zhang, W., Yao, R., Du, X., & Ye, J. (2021). Hybrid Deep Spatio-Temporal Models for Traffic Flow Prediction on Holidays and Under Adverse Weather. *IEEE Access, 9*, 157165-157181. doi:10.1109/ACCESS.2021.3127584

Zhang, Y. (2020). Short-Term Traffic Flow Prediction Methods: A Survey. *Journal of Physics: Conference Series, 1486*, 052018.

PUBLICATIONS OF THE AUTHOR RELATED TO THE DISSERTATION

2024

Pragalathan, J., & Schramm, D. (2024). Urban traffic flow predictions with impacts of weather and holidays. ICITT 2023, Madrid, Spain. *Advances in transdisciplinary engineering*, 274-281. IOS Press. doi: 10.3233/ATDE240042

2023

Pragalatahan, J., & Schramm, D. (May 2023). Applications of traffic flow predictions and simulations in urban mobility with connected automated vehicles and advanced traffic management systems. *15. Wissenschaftsforum Mobilität "Next Chapter in Mobility"*. Duisburg, Germany: Springer. (In press)

Pragalathan, J., & Schramm, D. (Apr 2023). Comparison of SARIMA, Fb-Prophet and NeuralProphet Models for Traffic Flow Predictions at a Busy Urban Intersection. *Sustainable Infrastructure: Innovations, Opportunities and Challenges-2023 (SIIOC-2023)*. Springer. (In press)

2019

Pragalathan, J., & Schramm, D. (2019). Quality Assessment of Long-Term Traffic Data Collection by Video Detection Technique in a Busy Urban Intersection. *Journal of Traffic and Logistics Engineering*, *7*(2), 35-40. doi:10.18178/jtle.7.2.35-40

DOI: 10.17185/duepublico/82102
URN: urn:nbn:de:hbz:465-20240701-151509-3