NOTTINGHAM TRENT UNIVERSITY

A Pervasive Prediction Model for Vehicular Ad-hoc NETwork (VANET)

Qiming Zhang

School of Science and Technology

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of Doctor of Philosophy

August 2017

Copyright statement

Acknowledgements

Around of four years and following of my research life, I am indeed honored to be able to learning and living in Nottingham Trent University which is an extraordinary experience of my lifetime. In here, sincerely express my gratitude for many people directly or indirectly giving me supervisions, advice, encouragement and funding for my research time, including my supervisors, research group colleagues, friends and my family.

I would like to acknowledge my gratitude to my supervisor Dr. Evtim Peytchev whose gave me an opportunity for this research and was always being constant advice, inspirations, guidelines and encouragement of my research. I cannot reach the ultimate goal without his supports. He always encouraged me to participate the various activities in academic such as conference and workshop, in order to communicate with many more researchers that made me had right and wide research ideas. Here, I again extend my most sincere thanks to him.

Also many thanks go to Dr. Taha Osman, Dr. Joanna Hartley and Dr. XiaoQi Ma who played an important role in my research work for giving assessments and proposals. It has been a pleasure to work with all of the members of Intelligent Simulation, Modeling and Networking Group and thanks to all for their knowledge sharing and discussions.

I would like to thank Dr. Yueyue Li, Dr. Emadeddien Gamati and Dr. Yasser M. Alroqi, they were my research colleagues and have successfully completed their research. They gave a lot of help and advice since the start of my research. They were kind of share their research ideas and very patiently to improve my research ideas. Particularly, I am grateful to Dr. Yueyue Li. Her suggestions and assistance contributed significantly to my research work during the last phase of simulation experiments. She also introduced me to a lot of new knowledge and advanced technical resources when the start of my research, that offer a shortcut to me and saving a lot of time for selection of tools.

During my stay in the university, I have made some great friends, especially, Nnamdi Anyameluhor, Ayman Abufanas and Gregory Albiston who are my research colleagues at current, also Liming Yao, Yingxiang Long and Yuandong Liu who are my best friends in my PhD life and always gave me a lot of encouragement and support. We have always discussed knowledge and shared information for our research. They also gave me many assistances for reading and proofreading my writing and helped me to make great progress in English language. It would not be out of place to thank them for their support, encouragement and friendship during all these years.

Supporting every worker is an understanding family. I am obliged to express gratitude to my parents for their support, understanding and funding. I owe the largest debt of gratitude to them. They made the decision and support me to study in the UK and provided me with great opportunities to enjoy and experience different things of my lifetime. They always stood by me during these many years and gave me motivation, encouragement and love.

Abstract

The growth of city traffic has contributed to severe traffic congestion and traffic accidents in the most of the cities in the world. Since people's travel demand rise at a rate usually greater than the addition of road capacity to lead many other issues, such as environmental problems and the quality of life. Intelligent Transportation System (ITS) is committed to solving the worsening traffic problems. Wide deployment of such ITS can eventually provide more dynamic, real-time and efficient solutions to transportation problems. ITS uses a variety of high technologies, especially electronic information technology and data communications technology to improve road traffic efficiency, road traffic safety and environmental protection. A number of researchers have depended on the wireless mobile communication to improve data collection and utilisation. The data could be used for early warning and forecasting traffic conditions in real-time.

The benefit of wireless mobile communication research, especially Car to Car (C2C) communication is to abandon the expensive wireline-deployed and central processing units. Through the interconnection of many personal mobile devices, a low-cost freely extended, high-performance and parallel system can be formed. Car to Car communication can make possible efficient and reliable data transmission by wireless links in a traffic area. It is based on principles of mobile ad-hoc network (MANET) and applies to the domain of vehicles, being Vehicular ad-hoc network (VANET) which is a key component of ITS. The C2C communication system has become essential for driving safety and comfort and also for improving road condition. Also, the traffic prediction system is also an important part of ITS, traffic condition prediction can be regarded as an extension application of VANET. It provides traffic condition in advance via a variety of prediction models and helps the people make better driving safety, travel decisions and route selections regarding departure or driving time.

The challenge of wireless traffic prediction technology is the uncertainty of traffic and real-time traffic data collection. It is widely known that urban transport system is a participatory, time-varying and complex nonlinear system. This uncertainty comes not only from the natural causes, such as seasonal and weather factors, but also from human factors, such as traffic accidents, emergencies and driver's behaviour. In particular, the

short-term traffic prediction is more affected by random interference factors. Current wireless traffic prediction research is usually based on a combination of wireless technology and traditional prediction model. The predictable traffic conditions include travel speed, travel time, traffic density, traffic accident, congestion level. However, in a large network environment, as the number of nodes increases, the transmission performance degrades and the prediction accuracy decreases because the prediction model does not obtain enough data.

In this thesis, a novel traffic prediction framework (PPM-C2C) is proposed – Pervasive Prediction Model (PPM) based on the C2C communication. The framework utilises ad-hoc data via C2C communications for a short time traffic prediction in each car.

This project builds and investigates the behaviour of a pervasive traffic simulation model in Ad-hoc network, with a particular part of it embedded into each vehicle's equipment. It includes the data collection, aggregation and application aim to be running in all individual cars so that cars have up to date information on the traffic at all times. Moreover, those cars could predict the traffic conditions of a road section in a short time through the proposed prediction framework, especially travel speed prediction. When the car receives the current traffic information about other vehicles, the prediction system will incorporate the information, analyse the data and predict the traffic conditions of this road section for a future time. The design does not depend on any roadside communications infrastructure. It is a simple and flexible car communication and processing technology to collect real-time traffic information. This process will be aided by car to car wireless communication technology available nowadays. To achieve this goal, a mobility model adapted to VANET needs to be generated that a realistic city scenario based on the actual traffic traces is carried out through simulation. Based on this, we investigate the necessary influencing factors for predicted results. The simulation results illustrate that the prediction model can be applied to wireless network environment for a short time prediction, and our results demonstrate the viability and effectiveness of the proposed prediction framework over Car to Car communications. Furthermore, the wireless environment and derived factors can result in decreased application performance.

Contents

List of Figures

List of Tables

Abbreviations

ARIMA-log series	ARIMA model with log-transformed series
AU	Application Units
C2B	Car to Broadband Could
C2C	Car to Car Communications
C2I	Car to Infrastructure Communications
C2X	Car-to-Car/Car-to-Infrastructure Communications
GPS	Global Positioning System
IEEE 802.11	Wireless access standard
ITS	Intelligent Transportation System
MAC	Media Access Control
MANET	Mobile Ad-hoc NETwork
MRErc	MRE remove cases
NS2/NS3	Network Simulator 2/ Network Simulator 3
OBU	On-Board Unit
OSI	Open System Interconnection
Otcl	Object-oriented Tool Command Language
PDR	Packet Delivery Ratio
PPM	Pervasive Prediction Model
PPM-C2C	The prediction framework which PPM based on the C2C communication
PRAWMA-log series	PRAWMA scheme with log-transformed series
QoS	Quality of Service
RSUs	Road-Side Units
TCP	Transmission Control Protocol
TSM	Traffic Speed Message
UDP	User Datagram Protocol
V2B	Vehicle to Broadband Cloud
V2I	Vehicle to Infrastructure
V2V	Vehicle to Vehicle
VANET	Vehicle Ad-hoc NETwork
WAVE	Wireless Access for Vehicle Environments
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WLAN	Wireless Local Area Network
y/5, y/10…..	Means observation data from SUMO by time period

1 Chapter 1

Introduction

This research investigates how to predict the traffic conditions of road sections in Vehicular Ad-hoc Networks (VANET), this includes the design of a traffic prediction framework with a novel prediction model for traffic conditions based on ad-hoc data in each vehicle, especially travel speed along road sections; the design of a traffic message delivery algorithm with real city mobility scenario; investigation and evaluation of influencing factors for road traffic prediction in a wireless environment.

1.1 Introduction to Research Project

The development of the transport industry is an important symbol of national prosperity. The rapid development of the transport industry not only promotes the exchange of goods and people travelling, it also makes greatly shortens the travel time and improves work efficiency. Simultaneously, traffic congestion, road construction and traffic accidents are plaguing the world's cities. To improve the efficiency of road networks, ease traffic pressure and guarantee traffic safety, many countries have been studied urban transport planning and traffic control. The traffic system control mode changed from Isolated Intersection control and arterial wireless control to area-wide control, now intelligent traffic systems (ITS) have been developed. ITS is based on information technology, data communications technology, electronic control technology and computer processing techniques to establish a large-scale, comprehensive, real-time, accurate and efficient traffic management system (Xu and Fu 2009). Until the early 1980s, only about 300 cities in the world have traffic control centres with some sort of ITS implemented. The most representative of these control systems are the urban road traffic control systems TRANSYT (US) and SCOOT (UK) by the Transport Research Laboratory (TRL), and the SCATS system from Australia (Xu and Fu 2009). Those transportation systems can be applied for different purposes, such as road safety, traffic monitors, sharing public information (parking, weather, buses or navigation map) and so on.

In recent years, traffic information prediction has increasingly demand and study of traffic prediction emerge intelligently and combination since the developments of ITS systems. The study includes traffic information collection, predictability analysis, predictive modelling and prediction system design. Meanwhile, traffic prediction can serve intelligent traffic control, traffic safety, traffic guidance, traffic information service and some subsystems of ITS. Therefore, the research on traffic prediction is of significant academic value and practical significance and provides a theoretical base to improve urban traffic management and solve urban traffic issue. (Laisheng, et al. 2009)

1.1.1 Vehicular Ad-hoc NETwork (VANET)

Vehicular Ad-hoc NETwork (VANET) is a relatively new technology in the mobile wireless network as an important role for ITS which is data sharing between vehicles equipped with wireless interfaces (Martinez, et al. 2011). Moving vehicles and transport facilities use wireless communication technology to form a mobile network. Within this network, vehicles will act as wireless nodes or wireless routers. Vehicles within a short or medium distance generate a link to each other to create a large-scale network. When a vehicle is out of signal range, other vehicles still can join in the network continue to link to each other to create a new mobile network.

In recent years, scholars have been extensively studying various transmission problems of VANET. VANET is a special network environment with unusual mobility and special applications, its information transmission method is different from a traditional mobile wireless network, and the design of its transmission control protocol becomes more challenging and originality. In VANET, efficient and reliable transmission control protocol has great significance to acquire the geographic information, channel quality and path state; but at the same time, the transmission control protocol also needs to face great difficulties and challenges of the communication channel narrow, high mobility, high density and so on. (Nzouonta, et al. 2009)

The applications of VANET can be divided into safety applications and user applications. Safety applications can significantly reduce the number of traffic accidents that 60% of accidents can be avoided if the driver gets a warning in 0.5 seconds before the collision (Xu and Fu 2009). Meanwhile, safety applications can also be used for timely to remind drivers that distance is too close to next one. According to the US Department of Transportation in 2003, the amount of collision at crossroads accounted for 45% of all registered vehicles and accounted for 21% of all traffic accidents (Xu and Fu 2009). If safety applications can inform drivers before the crash occurs and they can take measures to prevent the collision, then the accident will be greatly reduced. In addition, safety applications can also help drivers to choose the best route to reach the destination, which contributes to reducing congestion on the road, keeping traffic smooth, thereby increasing the traffic capacity of the road to avoid traffic jams. On the other hand, the user application can provide advertising, entertainment and other information to the passenger during the journey. The people's demand of accessing the internet has increased, thus, there is necessary to provide internet connectivity services for vehicle users and VANET network applications. In VANET, Peer to Peer (P2P) application are also an interesting approach for boredom, the car passengers (non-driver) can share, upload and download music, movies etc., also, can chat with each other and play games on long-distance trips. Some commercial organizations can push commercials to nearby vehicles through VANET, such as the contents of products or services, prices, open hours and location, these businesses could include restaurants, hotels, gas stations, tourist attractions. For example, supermarkets and chain stores can offer to customer orders through the VANET system, or help customers quickly find parking and cheaper fuel. These communication activities of information exchange are achieved by Car-to-Infrastructure (C2I) or Car-to-Car (C2C) (Eichler, Schroth and Eberspächer 2006). The latter overcome some shortcomings of the former, such as infrastructure deployment, connection and cost. It also focuses on message delivery between vehicles without centralised support. So whether from a safety perspective or commercial development, the study of VANET has great significance and value.

1.1.2 Traffic Prediction

Traffic information prediction is an important part of the modern science of traffic and ITS. It is based on history or existing transportation factors and statistics, then uses the intelligent computational method to forecast the transport system state in the target area for the future. Traffic prediction includes traffic volume, travel speed, travel time and traffic event. The traffic prediction method can be mainly categorized into two types: one is based on the dynamic Origination-Destination (OD) that is accordance with the rules of the traffic flow to forecast traffic conditions; another builds on the observational data of specific road section to forecast traffic conditions. The former requires a lot of OD data and road network infrastructure data, and OD data collect by onboard Global Positioning System (GPS) and Geographic Information System (GIS). The dynamic OD method can serve the entire transport network, and also requires a lot of advanced equipment and personnel. The latter serves a particular road section, requires less equipment and does not need people to cooperate. (Zhu, Wang and Xiang 2008)

There is already a variety of technologies used in traffic information collection, such as GPS, loop detector, visual detector or wireless sensor networks (WSN), and traffic prediction is based on the conventional testing equipment. These technologies have their advantages and disadvantages: GPS with wireless transceiver devices can monitor and collect the traffic information, but requires vehicles to install such relates to user privacy information; visual detector can acquire rich contents but it is vulnerability to environmental impacts such as weather or light intensity, and video information processing is difficult with requiring high bandwidth and only using data cable to transfer to the traffic control centre that is not conducive to expansion; loop detector is more traditional and simple equipment with high precision and has been widely used, but roads will be destroyed for laying and maintenance; WSN technology can provide information collection and transmission for ITS through roadside sensors to monitor traffic volume and speed, and the sensors self-organizing Mesh network by multi-hops and aggregated into a gateway node, then gateway node sends the collected data to the control centre by 3G or 4G network, however deploying a larger number of sensors for monitoring requires a lot of cost overheads. Most of the traffic predictions are used for centralised support method from the traffic centre to drivers. Typically traffic information is collected by those technologies and transmitted to the traffic control centre for forecasting and analysis, then the predicted results are dispersed and spread to the vehicles.

1.2 Outline of the Problems – Utilising C2C Communication Facility

The overall scope of the proposed research project provides innovative methods to handle various traffic problems, in line with efforts of scholars of various countries who have been committed to research and resolve traffic problems in the field of ITS. They tried to start from the hardware and software to introduce new concepts and technologies which will improve urban traffic. Wilson in his research “phantom traffic jam”, found that accident may not be caused by traffic jam alone and it may be due to a driver (Wilson 2008). He said that the main reason for traffic congestion is braking, unnecessary lane changing and overtaking other road users. Under normal circumstances, the effect of a personal bad driving or accidents could extend to 50 miles outside of that traffic confusion on the motorway. So if other drivers are prompted by information and prepared in advance, it can reduce congestion in a domino effect. Drivers usually want to know what the route situation they have chosen, or what the situation may be at a later time, so that they can choose an optimal and reasonable route for travel.

Traffic prediction technology is an important part of ITS which includes hardware technologies (data collection and data transmission) and software algorithms (prediction system and prediction model). However, the key to most of the prediction systems success lies in their prediction model. The software will combine the real-time data from roadside sensors, cameras and GPS transponders within the vehicle, and also need to have historical traffic information, road construction and query of weather conditions. The prediction system requires calibration of statistical data of recent a few weeks for prediction model adjusting. The system will transmit the alert information to roadside electronic nameplate and car satellite navigation systems. Even some of prediction system can also give an indication when the traffic of a congested road will return to normal (e.g. IBM traffic prediction system). Thus, each route of transportation can be balanced; the traffic will not become mechanical. Although it can provide a long term traffic prediction, the trouble is when the prediction system plays a role, many drivers are already travelling on the road, or simply been caught in a traffic jam or were on the way to congested roads. Each observation point has facilities running for optimising navigation effect to achieve unified by wire or wireless communication. In fact, the prompting of a simple road congestion information cannot avoid congestion fundamentally, if the prompting is misused it could lead to an increase in congestion levels. For example, there is a buildup of traffic on a highway, many drivers who receive an alert message for this case may move to highway 2. So highway 2 is now congested. Therefore, engineers need to be clever in designing the modelling system to decide if some drivers are sent the congestion information to achieve the best balance traffic flow of two roads. For example, sending information to only 30% to 50% of drivers or providing a real-time prediction of speed and traffic flow to drivers, this will prevent 100% of the drivers turning to highway 2.

The scope of this research will only focus on three parts of traffic condition prediction framework based on the ad-hoc data: data collection, data transmission and data application. We use the prediction framework to demonstrate these three parts. Therefore, their working performance and influence on the predictions are our concern.

A large of amount of existing prediction models and wireless communication algorithms can be directly applied to traffic conditions. We suggest a combination of both to improve traffic conditions and driving safety and estimate traffic conditions more effectively. Therefore, we are interested in "How to share and collect data using ad-hoc network"; "how to achieve traffic prediction by using ad-hoc data easily and efficiently in each vehicle"; finally, "To find out influencing factors for traffic prediction in the ad-hoc network to solve the traffic problems in the future".

In VANET, traffic prediction between vehicles is very challenging because the number of vehicles on the road network can change the topology and network fast. Thus, it requires an appropriate prediction model and algorithm for message delivery. Most of the existing prediction models are based on a large number of historical data and roadside infrastructure. A large number of software and hardware components require prediction models of so complex algorithms that using traditional methods of prediction model building will be error prone and time-consuming. Moreover, the predicted results cannot be passed to the participants promptly. We propose to utilise ad-hoc data via Car to Car (C2C) communication for traffic prediction in a short time. The proposed traffic prediction framework can be understood that the prediction model is an agent and C2C communication is a platform. Therefore, there are the relative advantages of traffic prediction here. Meanwhile, from the point of view of VANET performance, there are introduced the challenges for short time traffic condition prediction of a road section in two aspects that are data collection and prediction modelling based on the C2C communication:

Data collection and Data transmission

As previous introductions, the primary means of data collection is by roadside facility communication with moving cars (C2I) or stationary sensors and then aggregated data to traffic control centre (Leduc 2008, Xu and Fu 2009). Its advantage is high accuracy and capacity of data collection, mature technology and strong capacity of storage, operation and analysis on the server. To be economically feasible, it is an ideal system for data collection and data transmission (C2I) (Li 2013). However, the real-time data will become obsolete after modelling, comparison, analysis and transmission from traffic control centre. This is caused by the mobility of vehicles with uncontrollable, which means it is the driver’s intention rather than the computers that control the route of vehicles. Therefore, the data collection via a wireless network is more efficient than stationary sensors for the real-time traffic problems. To achieve high-level data collection, we also need to data transmission and data processing efficiently with the cooperation by each user.

For data transmission, as we mentioned above, since vehicles movement fast and uncontrollable, the topology of road network changes frequently, causing the temporary interrupt of communications. Researchers believe that vehicular networks have heterogeneity and have developed many distributed routing protocols for VANET. Therefore, how to use the changeable, unstable and heterogeneous vehicular network more effectively is a challenge. However with the development of C2C communication and onboard equipment, the vehicles can be used as an individual to receive, transmit and analyse data to achieve the exchange and utilisation of real-time traffic data. The wireless technology can enable the establishment of direct communication between the vehicles namely wireless ad-hoc network. Also, if each vehicle can embed a traffic prediction system or algorithm to analyse and forecast the traffic condition, these can make exchanging data between vehicles more efficient, real-time, increase coverage area and less expensive than the roadside infrastructure of data communication and collection. Although authors in (Raya, Papadimitratos and Hubaux 2006) stated the potential security threat of vehicular communications, it is able to cope with the threats based on a robust security architecture to achieve efficient and secure data collections and communications. Moreover, C2C communication still spaces to improve, especially shared with the dynamic real-time traffic conditions in the complex and busy scenario.

Prediction modelling

We demonstrate an application of C2C communication through the prediction modelling. Many factors need to be considered when generating prediction model. The role of prediction model must be clear which is the speed prediction, traffic volume prediction or traffic event prediction, it will maximise the relevance of collected data (van Hinsbergen and Sanders 2007, Min and Wynter 2011). Traffic environment changes with the higher speed of mobility over time while the adaptability of the prediction model and sample size need to be taking into account. In the previous discussion, the traffic control centre needs to proofread data in order to adjust the prediction model. If the prediction model is embedded in each vehicle, the prediction model needs to have the ability of self-adjustment and dynamic prediction. Data storage and processing power of onboard equipment are inferior to the server of a traffic control centre. Meanwhile, C2C communication can ensure that data up to date and cover more vehicles. However, the prediction algorithm based on data collection by C2C should not be too complicated. Therefore, the challenges of prediction framework based on the C2C communication are "how to build an adaptable prediction model with a limited number of samples", and "how to reduce the traffic prediction modelling difficulty with a large scenario".

On the whole, C2C communication could transmit data directly between vehicles and cover a large area for collecting the traffic data, while providing the real-time data to the prediction model embedded in each vehicle. Data transmissions may not reach every vehicle due to the high mobility of vehicles and transmission protocols. The data transmission is usually affected by the cost of connections, the speed of links, the size of data, volume of data or coverage range (Li 2013). Usually, the transmission speed is small when vehicles move with a high speed that will result in a small amount of data exchanges and time delay. Those reasons will impact to the predicted results. So far, many related projects and researchers have contributed more to C2I communications and traffic predictions by utilising current infrastructures and innovative algorithms. Also in C2C front, researchers have focused on improving the efficiency of communication to better share information. In fact, the C2C applications have broad prospects for the cooperation of C2C communication and traffic prediction. This can take sufficient advantage of the C2C to help people travel and solve traffic problems. However, it is almost impossible for real world deployment with many vehicles, so that we intend to use simulation to in lieu of real world deployment, thereby investigate the possibility and performance of the traffic prediction framework based on the C2C communication. Therefore, how to establish a large-scale simulation similar to the real world is also a challenge. Moreover, we also investigate several traffic parameters and the related influencing factors to demonstrate the advantages and disadvantages of our proposed solutions.

1.3 Research Aim and Objectives

This project deals with a framework to build traffic condition prediction models based on the C2C communication, which requires the integration of multiple technologies, including modelling and wireless communication. The overall aim of the project is to build a novel traffic prediction framework for the traffic conditions of a road section based on wireless ad-hoc network, with a specific part of prediction model embedded into each vehicle's equipment. When a vehicle equipment receives the traffic conditions around it, the prediction model will be analysing traffic data and forecasting traffic condition for a short time. This process will be aided by Car to Car (C2C) wireless communication method. The predicted results will be primarily coming from the vehicles themselves. Thereby, the related factors and influencing factors for the prediction model can be investigated.

C2C applications are the most appropriate for the local information exchanges and solve problems from traffic congestion, safety warning and so on (Li 2013). This project introduces a prediction method as a C2C application, and according to the information exchange between all vehicles in the road network, the prediction model can use this information to forecast the traffic conditions of road sections, especially average travel speed. So a broadcasting mechanism needs to be used in this project which includes sending, receiving and forwarding messages. The broadcast is an efficient method to deliver messages in a local area between vehicles. One vehicle sends its information to other vehicles with multi-hops technology; meanwhile, it also can receive information from other vehicles with broadcast routing protocols. Some of the routing protocols have been proposed and proved to be suitable in the vehicular ad-hoc network environment for data transmission. However, the reliability of them will be affected by traffic density, vehicle speed, road topology and other reasons. So routing protocol is the key to ensuring effective data transmission between vehicles. But information sharing of C2C communication is still weaker than centralised controls. In this cases, the prediction method is established by the small amount of historical data. On the other hand, the speed of vehicles is changed with time, so that the prediction model is based on the time series and traffic environment is usually a dynamic nonlinear system. All those reasons need to be considered when the pervasive prediction framework is designed.

The research consists of proposing an in-car pervasive traffic prediction model for VANET. A novel traffic prediction framework includes an in-car traffic simulation model based on the ad-hoc data and the Traffic Message Delivery Algorithm (TMDA) based on the C2C communications. To prove that this pervasive prediction framework based on the C2C communication for a future time prediction is practical and workable. This project proposes a Pervasive Prediction Model (PPM) as an example prediction model for the traffic condition prediction and adopts C2C communication to collect the information of vehicles (e.g. time, car id, speed and location). Thus, we name this traffic prediction framework PPM-C2C. Furthermore, influencing factors of prediction model performance and comparing routing protocols of communication performance will be simulated and evaluated with a real city scenario (Nottingham city centre, UK).

As a summary of the discussion above the objectives of this project could be identified: design pervasive traffic simulation model and message delivery in wireless ad-hoc networks for traffic conditions prediction in a short time. The proposed work will be carried out following the steps below:

- Review the existing traffic prediction model and research studies in the area of VANET with C2C communications.
- Generate pervasive traffic simulation model with real city scenario where the mobile nodes have random movement and route behaviour.
- Propose the Pervasive Prediction Model (PPM) as an example prediction

model for traffic conditions prediction in a short time, especially average speed prediction, and compare with existing prediction model.

- Present a design of a prediction framework based on the C2C communication (PPM-C2C) where the pervasive prediction model is included
- Create a simulation environment for testing and evaluating of the proposed framework working performance and influencing factors for predicted results.

1.4 Methodology

The research process in this thesis is designed in five stages:

- Establish traffic prediction framework (PPM-C2C) with a time series prediction model for ad-hoc data
- Compare and test new prediction model (PPM) with existing prediction model, thus, determine that the new model is feasibility.
- Build the traffic model and road network in real city scenarios.
- Import the network topologies and new prediction model into wireless network environment to establish the car to car communication with the routing protocols, also compare the performance of the using route protocols.
- Verify and evaluate the prediction results with influencing factors, and prove workable of the prediction framework in the wireless environment.

In the first stage, a traffic prediction framework with a time series prediction model for the traffic condition prediction is created for ad-hoc data in short-time and the parameters of prediction model will be defined. The new prediction model can use less real-time data car to car communication to forecast trend and value of the traffic conditions. The new prediction model needs ad-hoc data to forecast traffic conditions for a road section in the real city environment. Firstly the model determines and classifies the road section based on the vehicle’s position, then judges the validity of message according to vehicle’s id and message time, finally uses the vehicle’s speed to calculate the traffic conditions for the road section. Thus, the prediction model can be created in the VANET for short-time prediction because wireless transmission technology can guarantee efficient data transmission between each vehicle. For this aim, the research work will implement the prediction model by a real city scenario with C2C communication. Also, some of the factors can affect the performance of prediction model. These factors are expected to negatively affect the prediction model because of the wireless ad-hoc network environment. In this stage (Chapter3), we assume that the wireless environment has the infinite speed for data transmission. In this case, the data is received instantaneously without packet loss and time delay that we name “no wireless environment” for initial designing. Thus we use the observation data from SUMO as ad-hoc data in this stage.

In the second stage, we choose a known time series prediction model to compare with the new prediction model, in order to determine the feasibility of new prediction model. This implementation will use no wireless data and wireless ad-hoc data. The performance of new prediction model is expected to be better than existing prediction model.

In the third stage, to achieve the first and second stage described above, the simulation work will implement a real city scenario inspired by traffic simulation and network simulation in Simulation of Urban Mobility (SUMO) and Network Simulator (NS3) which describe in section 2.6.2 and 2.6.3. SUMO can generate the road topology network for a real city and routes of vehicles by traffic model and initialize individual vehicle from 0 simulation time as a mobility pattern. If we record the output of vehicle and traffic information from SUMO by every second which is called observation data. This output file describes the car id, speed, position and simulation time for the prediction model required. When the prediction model uses observation data that means no factors influencing it from the wireless network environment. No wireless environment can be seen as a complete statistical and record data, to investigate the influence by wireless network environment to the prediction model. However other reasons maybe impact the predicted results, for example, the number of vehicles or the length of the road.

In the fourth stage, after the real city scenario mobility pattern is generated, the implementation of the new prediction model and real city scenario is imported into a wireless network environment by NS3. We assume that each vehicle has the proposed prediction model. Moreover each vehicle sends their information to others with routing protocols for C2C communications. Then the prediction model can start working after vehicles receive data. In view of the characteristic of moving vehicle and mobility pattern, that time of operation of vehicles in the network is limited. Thus, data collection is different for individual vehicle and is might not enough information for prediction model required in some of the vehicles. For this reason, some of the observation static nodes are assumed to collect data from other vehicles, they are same functions (send and receive information) and properties (have car id, speed and position) with other vehicles, the only difference is their speed which is 0. The wireless environment will impact predicted results because of throughput, packet loss or time delay. Those metrics depend on the performance of routing protocol, which is the key to data transmission. We will use existing routing protocol for VANET to achieve data transmission between vehicles.

In the fifth stage, we will verify and evaluate the influencing factors to prediction model, and analyse the advantages and disadvantages of the proposed prediction model in a wireless environment and no wireless environment. Meanwhile, we will prove that the pervasive prediction framework based on the C2C communication for the future time is practical and workable.

1.5 Original Contributions

1.5.1 Design of an In-Car Traffic Simulation Model based on ad-hoc data

Chapter 3 describes a novel traffic simulation model as an agent of the proposed traffic prediction framework for traffic conditions based on ad-hoc data in short time prediction, namely Pervasive Prediction Model (PPM). It consists of components found in a real traffic network, such as vehicle and road section to achieve reliable and efficient prediction by car to car communications. Vehicle's activities include movement, broadcasting, receiving and computing to ensure adequate performance of communications and predictions. The prediction model can accommodate the dynamic and uncontrollable topologies with real-time ad-hoc data. Novelties incorporated in the prediction model include:

Usage of Combination Prediction

It is well-known that combination prediction method can be used for traffic prediction. In traffic prediction, there are few studies that use the in-car prediction model for the traffic condition (changing of trend and value), or many prediction models usually require massive historical data for traffic prediction. However, the PPM is based on less historical data for traffic condition integration and predicts the traffic at a short-term. Furthermore, the vehicle has many effects such as driver behaviour, vehicle, weather, road condition and so on in urban traffic. In this case, each vehicle might affect the traffic conditions of the road section. The PPM as a combination prediction model can provide various possible predicted results that are closer to the traffic conditions.

Usage of Adjusting Range

The part of adjusting range is an empirical and “fuzzy” aspect in the Pervasive Prediction Model (PPM) which dynamically determine the validity of polynomial regression results. Adjusting range can prevent too large or negative predicted results to make big errors while ensuring that the results are in reasonable range. Because the prediction model is designed for the wireless ad-hoc data with small sample sizes, in this case the traffic condition is updated instantly, so that there is ensured that the prediction model uses the latest data. Also according to fluid dynamics, the transportation is a continuous motion. Therefore, traffic condition is taking into account continuum assumption rather than discrete. The traffic condition is treated as a single point, and from one point to another is continuous change. This change needs to be a reasonable interval in a certain period whether it is accelerating or decelerating.

Usage of Ad-hoc Data

In short-term traffic prediction, the prediction framework and model does not need training with large historical data but only needs limited ad-hoc data which will be possible to forecast the traffic conditions for a road section. It aims especially to the performance of C2C communication via wireless technology in city transport, in particular there is without any traditional wired connections or roadside infrastructure. The messages will be created in the real-time and data transmission occurs between vehicles via a routing protocol.

1.5.2 Generation of a Traffic Message Delivery Algorithm with Real City Scenario Mobility Model

A real city scenario will be used to validate and evaluate the performance of traffic prediction framework (PPM-C2C), in order to obtain the ad-hoc data. The TMDA by C2C communication will be a platform of the proposed traffic prediction framework to deliver the messages, and the traffic prediction model achieves the goals of efficient traffic prediction. The evaluations of the proposed prediction framework and data transmission in this project are based on the simulation because actual implementation testing is financially infeasible. So we study on SUMO and NS3 as traffic and network simulators. There is drawn according to the map of Nottingham city centre by SUMO. In the traffic network, the nodes are assumed to vehicles with original dynamic movement. These nodes are distributed in the road network with different travel routes by the mobility model. Mobility pattern consists of vehicle information, road information, route information and time. The mobility pattern needs to import into network simulator (NS3) as a real city wireless ad hoc scenario. The responsibility of each vehicle is sending their information (e.g. position, time, id and speed) and forwarding others information. In this case, an algorithm of message delivery needs to be created to accommodate the wireless communication environment and estimate message behaviour (e.g. forwarding, dropping or using). The message should be delivered in an efficient and reliable approach via a routing protocol. Wireless data transmissions often process data collision and simultaneous or redundant broadcast to degrade communication performance. According to different routing protocols in our scenario, we will verify their effects on the traffic prediction framework. Thus, studies on routing protocols will be focused on delivery time delay and delivery successes. The simulation provides realistic results to analysis and evaluation.

1.5.3 Evaluation of the influencing factors in the Traffic Prediction Framework

In the traffic prediction research, there is no one prediction model and algorithm can be adapted to all conditions. Therefore, studies on the practicability of the PPM-C2C prediction framework in a wireless environment and appropriate applying conditions for this framework are mainly discussed in this research. The predicted results are probably constrained and influenced by many factors. The evaluations of underlying determinants will be listed and considered in no wireless and wireless environment for the proposed prediction framework and model, such as routing protocol, mobility pattern, observation location, objective road, time period and particular hours. Meanwhile the predictable traffic conditions include average travel speed (space mean speed), travel time and density. According to the experimental evaluation, the influencing factors could be adjusted that will be shown in section 5.3.

1.6 Thesis Structure

Chapter 1 introduces the research field and outlines the research problems that are addressed in this project. It also defines aims and objectives of this project, then lists methodologies and all major research contributions in this project.

Chapter 2 provides an overview of the area of the vehicular ad-hoc network (VANET) and its related works which include a comparison of MANET and VANET, VANET architectures, C2C communication and mobility model in VANET. Second, it also provides an overview of standard traffic prediction model, performance evaluation of prediction model and prediction model for VANET. The evaluation of traffic prediction model is assessed by reviewing published works. Additionally, we introduce the application of ad-hoc routing protocols for C2C communication which are proactive routing protocol and reactive routing protocols. There are also current simulation advantages and problems provided including various simulators and methodology explanations. Finally, we discuss the current applications of traffic prediction for VANET.

Chapter 3 presents modelling details of a Pervasive Prediction Model (PPM) based on ad-hoc data for traffic prediction of a road section. The chapter starts with a discussion on the classification of traffic prediction model for the prediction framework, and then describes the principle, component, processing and demonstration of the prediction model, evaluates and verifies the prediction model, finally, compares with existing prediction model in theoretic discussion and experimental discussion.

Chapter 4 presents the technical details of a message delivery algorithm with a particular emphasis on C2C communication for the prediction framework (PPM-C2C). The Chapter discusses on traffic messages delivery in VANET, meanwhile defines category, format, encapsulation, access and delivery scheme of traffic message at the beginning. Then a discussion on the routing protocol in VANET based on the real city scenario mobility pattern via simulation experiments. Their results are presented in the end part of this chapter to evaluate and compare them for the following chapter selections.

Chapter 5 describes the simulation methodologies and shows the results of experiments. This chapter emphasises simulation models of proposed PPM-C2C prediction framework for VANET and the implementations of C2C communication based on studies of specific real city scenario mobility models. There is a comprehensive and in-depth analysis of the various influencing factors in Pervasive Prediction Model (PPM) with wireless and no wireless environments.

Chapter 6 summarises the primary results and concludes this whole project presented in the previous chapters and proposes a baseline for the future works.

2 Chapter 2

Literature Review

2.1 Overview

Our aim is new knowledge in the area of in-car traffic prediction by C2C communications. Our concerns are the workable and performance of in-car prediction framework that includes the data collection, data transmission and data application through C2C communications.

In the Intelligent Transportation System (ITS) operating environment, the traffic analysis and prediction method are different from traditional concepts. "Real-time" and "Dynamic" are the most basic characteristics and the most different key features from the traditional transport analysis methods. In order to make ITS strategy to achieve immediate control and guide traffic, researchers and managers need to obtain the dynamics of traffic changes in the decision-making process and phenomena of interaction behaviour and control strategies. Therefore, a real-time, dynamic, accurate and efficient traffic control system is essential to aim to guide traffic flow and reduce congestion. This project provides a traffic speed prediction model based on the wireless ad-hoc network, thus the main focus on the field of vehicle wireless communication and traffic prediction in the literature review.

Vehicle Ad-hoc NETwork (VANET) has wide applications in ITS. VANET is usually combined with Global Positioning System (GPS) and wireless communication networks, such as Wireless Local Area Network (WLAN). VANET provides a high rate of data access networks in the fast movement of the vehicle, thus provides a possible solution for safe driving, data communications, accounting management, traffic management and vehicle entertainment. With the development of mobile wireless communication technology, derived from a variety of application based on the wireless Ad-hoc network in public transport.

Dynamic Traffic Prediction is a real time dynamic traffic information that it provides traffic predictions and travel guidance (Ben-Akiva, et al. 1998). The aim of Dynamic Traffic Prediction is intended to make the prediction algorithm with a change of traffic condition, thereby continue to adjust and update the prediction model. The data of prediction model demand can be collected quickly through proper message delivery approach, namely Car to Car (C2C) communications (Eichler, Schroth and Eberspächer 2006).

On the basis of studies in VANET filed (Dashtinezhad, et al. 2004, Nzouonta, et al. 2009, Nandan, et al. 2005)and traffic prediction filed (Ben-Akiva, et al. 1998, Sun, et al. 2003, Xie, Wu and Yang 2011), Dynamic Traffic Prediction will be an application based on the VANET architecture, and it is characterised by wireless mobile technologies which are C2C communications.

This chapter presents brief reviews of VANET, standard traffic prediction models and their taxonomy in section 2.2 and 2.3. Then section 2.4 introduces a most important research area concerning new generation prediction models for VANET and related works. Section 2.5 refers the routing protocols ranging from general classifications to the particular environments applied in VANET. Section 2.6 discusses the simulation techniques and simulation tools of traffic simulators and network simulators. Section 2.7 discusses the knowledge of related works for traffic prediction in VANET. The last section 2.8 gives a summary of literature review and raises coming chapter.

2.2 Vehicular Ad-hoc NETwork (VANET)

2.2.1 Overview of MANET and VANET

Mobile Ad-hoc NETwork (MANET) has characteristics different from the traditional network that is mobility, easy arrangement, high fault tolerance, self-organization and multi-hops. MANET originated from military mobile communication or disaster relief communications at the beginning. It is a wireless communication technology which cannot rely on the ground communication station and infrastructure, also can quickly layout in extreme conditions and environments. MANET has most popular applications; for instance, Sensor Network for data collection (Sohrabi, et al. 2000), Peer-to-Peer Network (P2P) for data sharing (Ding and Bhargava 2004) and

Vehicular Ad Hoc Network (VANET). Traditional Sensor Network architecture is directly data transmission between sensor and sink. However, sensor and sink may not be established a connection in certain terrain and environment. So MANET is applied to make sensor motion detection, complete self-organization and network architecture, transmit data (Tilak, Abu-Ghazaleh and Heinzelman 2002). P2P network is a distributed application network architecture that between Client and Server, particularly suitable for MANET's network properties that directly connect to the nodes without the base station, it can easily achieve data sharing, data transmission, distributed computing, even some nodes often join in or leave in P2P network(Schollmeier 2001). Finally, VANET is a special application and typical of MANET (Li and Wang 2007), also the basis for ITS.

The vehicular network can be formed in three ways which are cellular network, road infrastructure or Car to Car communications (Nzouonta-Domgang 2009), namely heterogeneous wireless networks shows in Fig. 2.1.

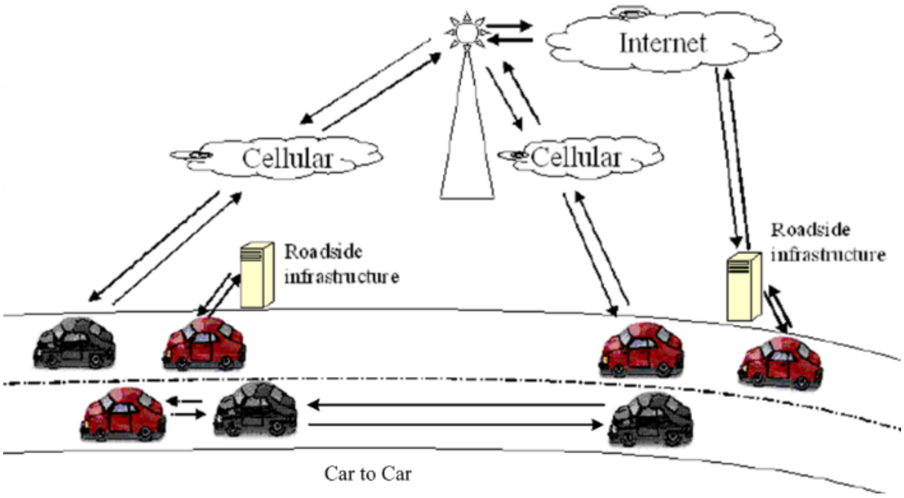


Fig. 2.1 Vehicular heterogeneous wireless networks

VANET's node is a vehicle and uses multi-hops wireless communication only between cars. The established of VANET has been becoming possible by the latest developments in communication technologies such as hardware and software. VANET involves two huge industries which are automotive industry and communication industry, and the relevant principals include government departments such as U.S Department of Transportation (U.S.DOT), Federal Communication Commission (FCC) and European Commission (Qian and Moayeri 2008, Federal Communications

Commission 2002) , car manufacturers such as GM, Ford, BMW, International Organization for Standardization (ISO) such as IEEE, European Telecommunications Standards Institute (ETSI) and Car to Car Communication Consortium (C2C-CC) (Hartenstein and Laberteaux 2008), information technology companies such as Google, HuaWei (Bisio, et al. 2013)and research institutes.

2.2.2 Characteristics of MANET and VANET

MANET is one of the popular application of mobile wireless network in the world. Its characteristics that differentiate it from other traditional wire and wireless networks are:

- **Duality of mobile terminal:** In MANET, the mobile node’s feature can be varied. When a node is used as a host to run the application for the user, as a router for data forwarding.(Laneman 2006)
- **Dynamic Network Topology:** The user terminal is free to move in MANET. So network topology is changed by node’s transmission range, radio transmission power and Geographical.(Wu 2005)
- **Self-Organization:** There is no control centre in MANET, each node has the same status. Each node is responsible for discovering other nodes around it. Any node failure does not affect the entire network, with more robustness and survivability.(Sohrabi, et al. 2000)
- **Multi-hops:** There are no any roadside infrastructures in the network; therefore, the nodes should be located close to communicate with each other neighbour nodes. Certain remote nodes need Multi-hop to surmount the distance and range problems. In this way network coverage can be continued to expand. Meanwhile, the receiver and sender can be used much smaller power than direct connectivity, thus reduce system power consumption. (Chlamtac, Conti and Liu 2003)

VANET has many similarities with the MANET, such as multi-hops transmission, dynamic topology changes, self-organization, etc. But VANET also has distinct features different with MANET (Mohimani, et al. 2009):

- The node distribution is restricted by road topology. (2L/(2+∆r)) data packets can be simultaneously transmitted in L length road, and r is the average distance of one hop.
- Nodes move fast and network topology changes fast and it leads to short link's life.
- Unstable radio channel quality, and by a variety of factors, including roadside buildings, road conditions, vehicle types, and the relative vehicle speed.
- Network load will change by traffic density, and therefore, the node should be able to adapt to this change.
- Nodes in a VANET are vehicles or roadside units so that they are going with plenty of energy supply, good wireless communications, storage and computing.
- Node obtains a wealth of external auxiliary information such as GPS or GIS.
- Node moves regularly so that node's travelling direction, speed and road link can be predicted.

Compared to traditional wireless networks, VANET has special behaviour and characteristics that including advantage and disadvantage aspects (Martinez, et al. 2011):

Table 2.1 Evaluation of VANET compares with traditional wireless networks

Advantage	Unlimited Transmit Power	Vehicle self-generating, adequate energy
	Higher Capability	More device space and computation power
	Predictable Mobility	Vehicle regularly movement in one direction
	Well Equipped Vehicles	Vehicle has been equipped with a variety of sensor and communications device
Disadvantage	Large Scale	May cover the entire road network, the scalability of VANET's protocol is an issue
	Partitioned Network	May become a separate branch network in sparsely populated scenario that not connect to the entire network
	High Mobility	Vehicle's move and route complicated, channel dramatic changes, wireless range and node's density impact of network connection, the behaviour of connection and disconnection are very frequently, rapid changes in the topology

VANET has a high dynamic network structure, the vehicle's security applications also have very strict requirements and rely heavily on Quality of Service (QoS) so that design of VANET's protocol an enormous challenge revolves around the adaptions of network features and demand of applications (Xu, et al. 2013).

2.2.3 VANET Architectures

This section introduces the architecture of VANET. C2C-CC provides a guiding framework for VANET, there are three domains: In-vehicle Domain, Ad Hoc Domain and Infrastructure Domain (Fonseca, et al. 2007, Kosch, et al. 2012), shows in Fig. 2.2 (Car to Car Communication Consortium 2007).

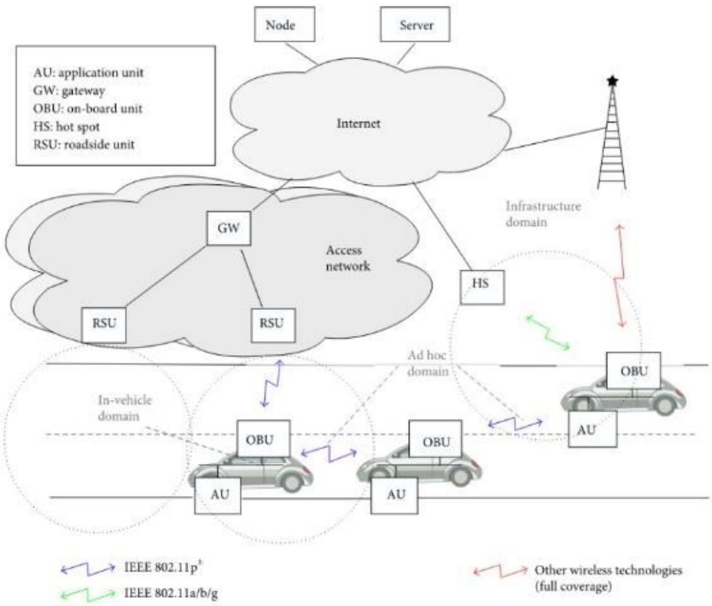


Fig. 2.2 C2C-CC reference architecture

In these three domains, the wireless access scheme may include: Global Navigation Satellite System (GNSS) belongs to Infrastructure Domain, Cellular Networks 2G, 3G, 4G also belongs to Infrastructure Domain, WLAN (IEEE 802.11 a/b/g/n/ac) can be applied in Infrastructure Domain and In-vehicle Domain (Crow, et al. 1997), but IEEE 802.11p uses into Ad hoc Domain (Jiang and Delgrossi 2008) and Bluetooth belongs to In-vehicle Domain. In these cases, the application of VANET might be used not only one communication way. For example, eCall uses the both GPS navigation and 2G network (Grzeszczyk, et al. 2009). When the airbag deployment and other major sensor signal are received by the eCall device, eCall will automatically start communication module, call the emergency number, report vehicle's GPS coordinates, accident time, license plate number and other information, also establish voice communications with the local rescue team at the accident first time.

As shown in Fig. 2.2 and Table 2.2, the In-vehicle domain is composed of an On-Board Unit (OBU) and an Application Units (AU). The OBU is responsible for C2X communications, and also provides communication services to AU and forwards data on behalf of other OBUs in the Ad hoc Domain. The connection between an OBU and an AU usually connects with a wired or wireless in sometimes. The Ad hoc Domain is a network composed of vehicles equipped with OBUs and Road-Side Units (RSUs) that are fixed units along the road. OBUs can be seen as nodes of MANET, and RSUs are stationary or static nodes, so that if no direct connectivity from the source to destination, one OBU forwarded the data to another allowing multi-hop communication. An RUS can be attached to an infrastructure work via the gateway and may allow OBUs to access the infrastructure. (Olariu and Weigle 2009) Infrastructure Domain access consists of RUSs and Hot Spots (HSs), and consequently to be connected to the internet. OBUs can communication capabilities of cellular radio networks (4G, WiMax, GSM, GPRS and UMTS) if without RSUs and HSs. (Moustafa and Zhang 2009, Padmavathi and Maneendhar 2012)

Table 2.2 Three Domains in VANET

In-vehicle Domain	Connect between OBU and AUs. AUs can be a specific device (e.g. cell phone) Or, an integrated module is permanently connected to an OBU. The connection is wired or wireless such as Wi-Fi and Bluetooth.
Ad Hoc Domain	Wireless communication between OBUs and RSUs can be direct communication or multi-hop, usually namely C2C or C2I
Infrastructure Domain	The communication between OBU, RSUs and infrastructure, such as satellite, Hot Spots, 3G and 4G to complete access to the internet. For RSU, the connection can be wired.

VANET can be categorized into four communication types, and it is closely linked VANET key components as the discussion above. The VANET key components and functions of four types are shown in Fig. 2.3: (Faezipour, et al. 2012, Liang, et al. 2014)

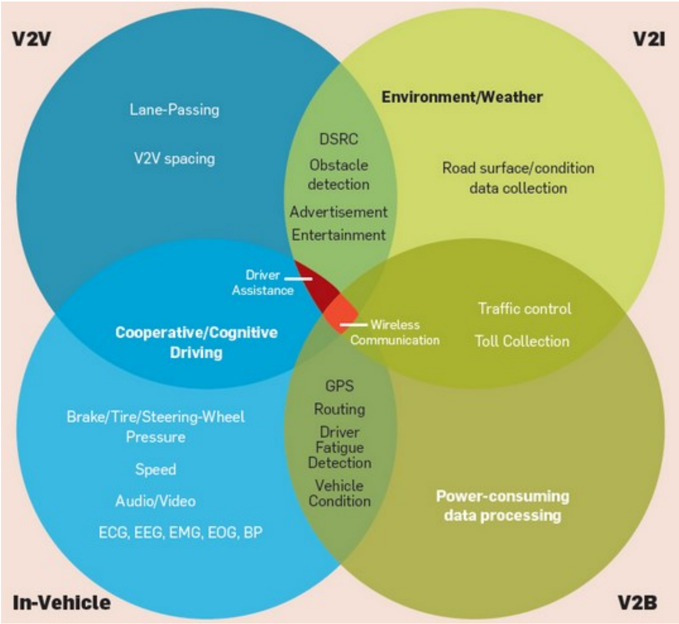


Fig. 2.3 VANET key components and functions

- Vehicle to Vehicle (V2V) Or Car to Car (C2C) communication provides a data assistance exchange platform to broadcast and share information or warning messages for the drivers, refers to the Ad-hoc Domain.
- Vehicle to Infrastructure (V2I) Or Car to Infrastructure (C2I) communication provides the traffic and weather updates information in the real-time to the drivers.
- In-Vehicle communication refers to the In-vehicle domain. It can detect the performance of the vehicle, process the data from share information and provide necessary safety functions for drivers and public. It is more and more essential and important in VANET research.
- Vehicle to Broadband Cloud (V2B) or Car to Broadband Could (C2B) communication is seen as V2I, but the information and data stored in the broadband cloud, rather than road-side infrastructure, via wireless broadband mechanisms such as 3G or 4G from vehicles.

2.2.4 Standards for Wireless Access in VANET

For diverse VANET's applications, each protocol layers vary greatly and tend to consider cross-layer or simplify layer (Shakkottai, Rappaport and Karlsson 2003). The layered architecture of VANET has been divided into seven logical layers, and the Open System Interconnection (OSI) model groups into one of them (Zimmermann 1980). Each layer in one network node communicates with the same layer in another node, and each layer serves the layer above it and applies the service of the layer below it. The presentation layer and session layer are often ignored, and IEEE gives the Wireless Access for Vehicle Environments (WAVE) protocol stack that an authoritative framework agreement, as illustrated in Fig. 2.4 (Uzcategui and Acosta-Marum 2009).

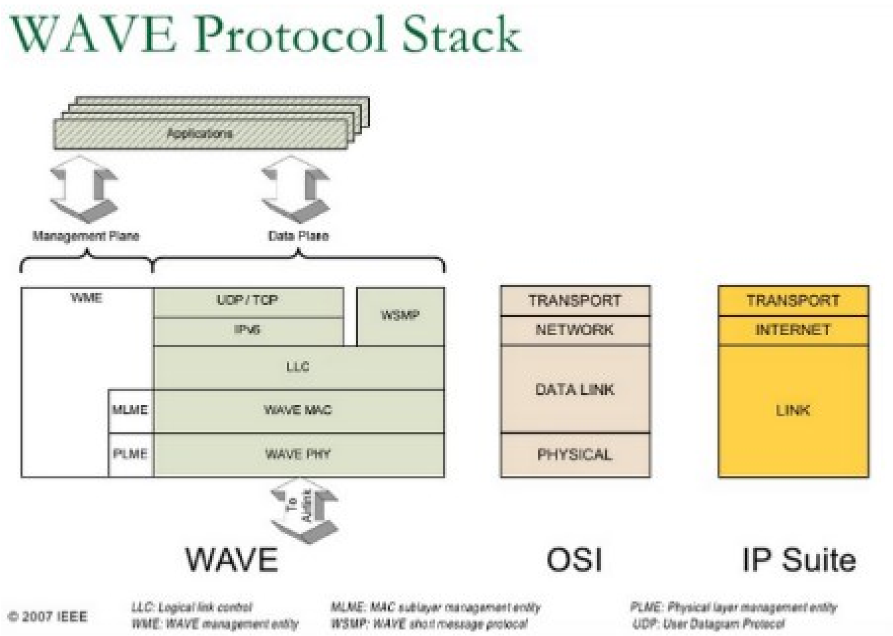


Fig. 2.4 WAVE Protocol Stack

WAVE accommodates two protocols stacks standard internet protocol (IPv6) and WAVE Short Message Protocol (WSMP). The IP messages (IPv6) may be sent only on service channel (SCH) and allows access to the general application and network. The system management frames are sent on the control channel (CCH). CHH reserves for short, high-priority application and system control message. SCH supports application data transmission for the general purposes. WAVE Short Message (WSM) might use in any channel in order to enhance the transmission efficiency in the WAVE. WSMP allows applications to directly control PHY characteristics, such as channel number and transmitter power for data transmission (Uzcategui and Acosta-Marum 2009). WSMP only supports one-hop transmission and does not support multi-hop transmission in the general ad hoc networks. This makes it more like a transport layer protocol rather than routing issues in the network layer.

WSMP data will be transferred from LLC layer to the MAC layer. Next MAC will route the packet to the appropriate buffer, which is referred to the channel number in WSMP header. If the channel number in WSMP header is invalid, then the packet is discarded, and the so-called invalid channel number means that it does not correspond to the CCH number or SCH number. As for IP message, before the IP packet is transmitted, there is a registration to MLME. The content of registration needs to include power level, data rate, SCH number. Then the IP packet is transferred from LLC layer to the MAC layer, and MAC layer will forward the packet to SCH buffer.

In the Fig. 2.4, five of seven OSI layer are Physical, Data Link, Network, Transport and Application (Wu, et al. 2010): the Physical layer and Data Link layer are composed of 802.11p, 1609.4 and 802.2; Network layer and Transport layer have two protocols that namely traditional TCP/IP protocol and 1609.3 protocol designed specifically for In-vehicle safety applications; the Application layer is divided into Safety Applications and Non-Safety Applications as shown in Fig. 2.5, and introduce SAE protocol as message sub-layer in Safety Applications, this protocol stack is for Dedicated Short Range Communications (DSRC) and is specifically designed for automotive (Hartenstein and Laberteaux 2010); finally, IEEE 1609.2 protocol is a cross-layer as the security services.

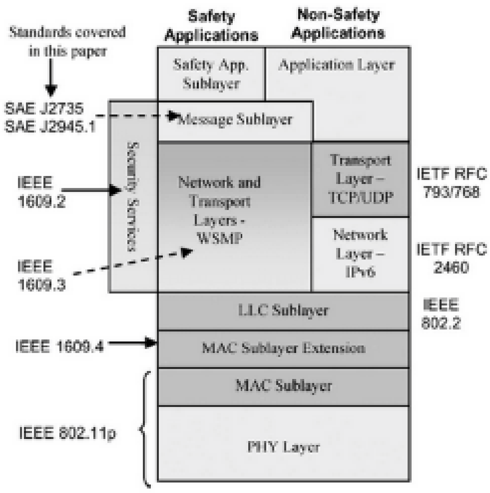


Fig. 2.5 Layered architecture for DSRC-WAVE Protocol Stack

The DSRC is a highly efficient two-way wireless communication technology which can be enabled to identify the high-speed moving target in a specific short to medium range areas (Cheng, et al. 2007), such as C2C or C2I communication for transmitting images, voice and data information in real-time. DSRC widely used in the field of toll collection, access control, information services, vehicle and driver identification and exchange message between network and vehicles. The DSRC supports the message delivery in rapidly changing communication environments where time responses and high rates are required. The American Society for Testing and Material (ASTM) approval the use of DSRC technique based on IEEE 802.11 that assigned 75MHz of the spectrum at the 5.9MHz range in 2003 (Menouar, Filali and Lenardi 2006).

IEEE 802.11 series standard as the well-known standard of WLAN that provides a higher speed of transmissions but shorter distance dependent on different types than the cellular network. IEEE 802.11 standard is achieved to develop PHY and MAC specifications for wireless connections between static, portable and mobile within the local area (Ma and Jia 2005, Li 2013). A comparison of several well-known IEEE 802.11 standard in terms of the frequency band, bandwidth, transmission range, compatibility and modulation as shown in Table 2.3, by referring to (IEEE Computer Society LAN MAN Standards Committee 1997, Jui-Hung, Jyh-Cheng and Chi-Chen 2003, Eichler 2007, Wang, et al. 2008).

Table 2.3 WLAN IEEE 802.11 standard series specifications

	802.11	802.11a	802.11b	802.11g	802.11p
Approval	1997	1999	1999	2003	2010
Frequency(GHz)	2.5	5.15~5.875	2.4~2.5	2.4	5.86~5.925
Bandwidth(Mbps)	1~2	Up to 54	6.5, 11	Up to 54	3, 27
Transmission range (m)	20~100	30~120	35~140	35~140	300~1000
Compatibility	802.11	802.11a	802.11/802.11g	802.11b	802.11a
Modulation	FHSS/DSSS/IR	CCK-OFDM	DSSS/CCK	CCK-OFDM/PBCC	CCK-OFDM
Mac protocol	CSMA	CSMA/CA	CSMA/CA	CSMA/CA	CSMA/CA

IEEE 802.11 physical layer implement by Frequency Hopping Spread Spectrum (FHSS), Direct Sequence Spread Spectrum (DSSS) and Infrared Spectrum (IR) (Crow, et al. 1997). The modulations of Complementary Code Keying (CCK), Orthogonal Frequency Division Multiplexing (OFDM) and Packet Binary Convolution Code (PBCC) apply to the higher data rate and bigger transmission range (Fernandez, et al. 2012).

IEEE 802.11 Media Access Control (MAC) layer “is responsible for wireless media access, network connection, error checking and the confidentiality of data" (Li 2013). Carrier Sense Multi-Access/Collision Avoidance (CSMA/CA) hardware has been used in various environmental conditions in order to obtain good communication quality.

2.2.5 VANET for C2C Communication applications

The automotive applications are driving factor for the development of VANET network, and all applications are in the face of different population, requirements and technologies. Basically, it can be divided into active road safety application, traffic efficiency and management applications, and infotainment applications (Cheng, Shan and Zhuang 2011, Dar, et al. 2010). There into, a strict QoS requirement is appropriate for safety applications, and it needs to be considered in the design for WAVE layer protocols. From the present situation, the safety application will eventually be implemented using WAVE protocol and WAVE Short Message Protocol (WSMP) (Uzcategui and Acosta-Marum 2009), but the application of efficiency and infotainment will be integrated Ad hoc Domain, TCP/IP, and Infrastructure Domain to achieve (Hoebeke, et al. 2004, Baldessari, Festag and Abeillé 2007).

The purpose of safety application is obvious that to avoid accidents and casualties. The message of safety application can be divided periodic message and event-driven message (Gross, Digate and Lee 1994, Hartenstein and Laberteaux 2008). The periodic message is used to detect an unsafe situation to adjust the vehicle's status that in order to avoid a dangerous situation, such as location, velocity and direction. The event-driven message needs to be sent immediately after the dangerous situation to warn and avoid secondary accidents. The message of safety application has to be fast and reliably transferred from the host vehicle to nearby vehicles. Under a very strict QoS requirement, the delay time should below 100ms, and WAVE protocol channel switching to sub-layer strict rules in 50ms (Kim, et al. 2008, Wang 2013).

Non-safety applications (traffic efficiency and management & Infotainment) designed to improve the driver's experience on the road. The transmission capacity of a single node is close to 0 when the number of nodes increases in the network, that's because of the multi-hops network has a capacity limit so that QoS is difficult to ensure the connectivity of the entire network (Li, et al. 2001). The non-safety application has been a hot issue for the industry. For example, the vehicular sensor detects the environmental condition and social activities and report to the appropriate server or vehicle interact with around vehicles to obtain and disseminate messages (Wang 2013). At present, many automotive companies and internet companies have released their operating platform; for instance, Ford and Microsoft together developed a sync system and interoperability through AppLink and cell phone (Cronin 2014). Apple IOS "in car" system (CarPlay) is also using voice and button to control the steering wheel and signed a cooperation agreement with Mercedes and other 12 brands (Cue, et al. June 12, 2013).

While the vehicle application and communication have attracted wide publicity attention to resolve the traffic issues in VANET. VANET consists of a group of the self-organised wireless vehicle without the centralised controls. The vehicle or node exist collaboratively generate and control transmissions in the distribution system with the features of dynamic, flexible, distributed and self-organization. Multi-hop also provides wide road network coverage by message forwarding among independent vehicle mobiles (Ghosekar, Katkar and Ghorpade 2010).

2.2.6 Mobility Model of VANET

Although VANET is a special application of MANET, the topology structures of their difference still exist. The node is a vehicle in VANET topology (Fig. 2.6), it cannot a random and directionless movement so that it has to abide by a scenario and move on the road in a certain direction (Zheng, et al. 2012).

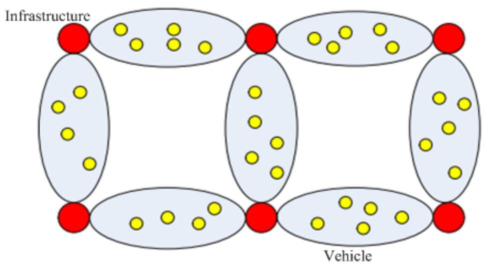


Fig. 2.6 VANET topology structure

In this case, the scenario is crucial which will influence the performance of the VANET, especially the importance of the mobility model is self-evident. In the literature of (Díaz, Mitsche and Pérez-Giménez 2009), they used nodes random move that means random directions in the network to test and analyse the network connectivity of VANET. Also, F.Bai et al, they tested the performance of connectivity and different routing protocols with different mobility models and the mobility models how to influence the network (Bai, Sadagopan and Helmy 2003). An additional remark from M.Fiore and J.Harri provide the topological properties of different mobility model and explaining why model impact the performance of network protocols (Fiore and Härri 2008). An expansion of relates works how to generate a random model in a city scenario for VANET and investigate the node connectivity and communication capability of VANET with this dynamics random model from (Ho, Leung and Polak 2011).

According to (Harri, Filali and Bonnet 2005, Härri, et al. 2006, Karnadi, Mo and Lan 2007, Martinez, et al. 2008, Harri, Filali and Bonnet 2009, Wang 2013), the mobility model of VANET can be divided into four components that address to Table 2.4.

Table 2.4 The Components of Mobility model for VANET

Motion Constraint	Graph	Road map can be randomly generated, based on the real map or user-defined.
	S & D Points	The point of Vehicle’s entering and leave the scene can be randomly generated or based on attraction or repulsion points of interest such as a building or park.
	Intersection	The traffic policy in intersection or crossing, such as stop sign, traffic light even one-way-street
	Collision	Avoid collision between cars, based on the algorithm of mobility model
	Speed	Each lane requires limited speed, especially upper limited
Traffic Generator	Trip	Describe the route of vehicle that from source to destination.
	Path Computation	Gives the full path of the trip from source to destination by the algorithm, should base on the trip length, speed limitation or congestion
	Mobility Patterns	Describe the motion parameters of vehicles, such as speed, vehicle length, distance from the front car, etc.
	Lane Changing	Describe the parameters of lane changing and overtaking, such as a number of the lane, the speed of the front car, etc.
	Intersection Management	Represent the parameter of vehicles under the intersection policy, such as acceleration or deceleration
Time Pattern	Duration	Involve many parameters of time and a different configuration of day or week.
External Influence	Condition	Some effect on vehicle motion such as construction or accidents.

In the literature of (Harri, Filali and Bonnet 2009), they summarised the taxonomy of mobility models available for VANET, and illustrated a framework that proposes a guideline for the generation of realistic vehicular mobility patterns as shown in Fig. 2.7, and the different ways were chosen by the community for the development of vehicular mobility models.

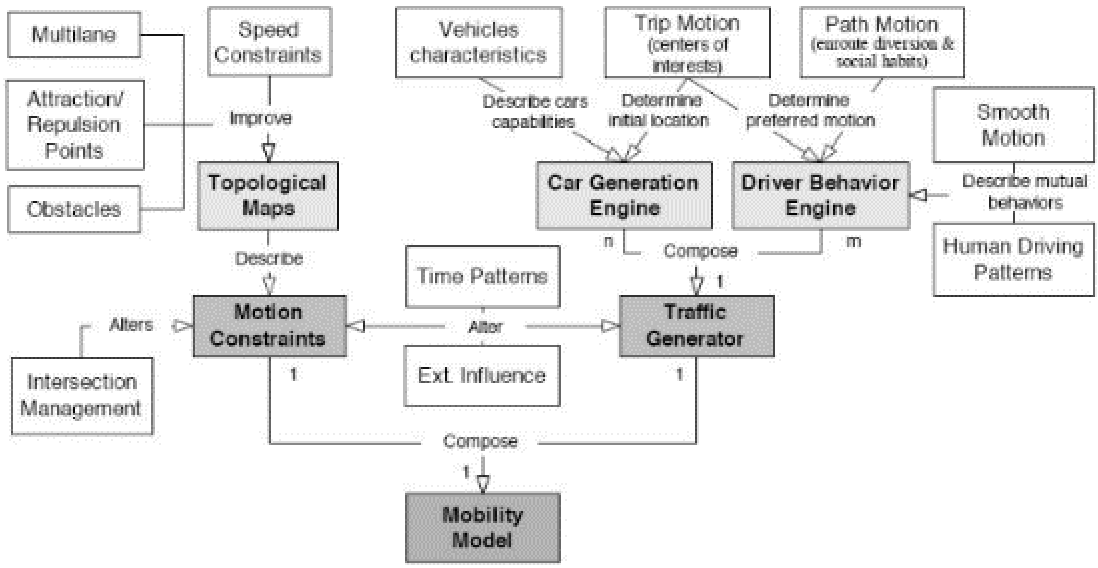


Fig. 2.7 Concept map of realistic vehicular mobility model

The designers should consider to all blocks of as above when they create mobility models. However, this concept map is too complicated rather than actual models that often simplify it for the model generated. For example, the Attraction and Repulsion points are usually absent in Motion Constraints; a simple setting and configuration for Car Generation Engine; the Driver Behaviour only consider the Smooth Acceleration or Deceleration; fix configuration of Time Patterns; the External Influence is often ignored to generating the model, just analyse the scenario.

The mobility patterns, as the most complex part of the vehicular mobility model, can be classified into four categories: Synthetic Models, Survey-Based Models, Trace-Based Models and Traffic Simulator-Based Models (Harri, Filali and Bonnet 2009).

- **Synthetic Models:** It comprises all mathematical models. From the classification of (Fiore 2008), Synthetic Models can be concluded in five classes: stochastic models containing all random motion models, traffic stream models, car following models, queue models and behavioural models, which specific models represent: random waypoint model, highway model, Manhattan model, car following model, intelligent driver model, Krauss model, cellular automata models, etc (Gipps 1981, Bettstetter, Resta and Santi 2003, Maerivoet and De Moor 2005, Kesting, Treiber and Helbing 2010).

- **Survey-Based Models:** The models are based on US Department of Labor Bureau of Labor Statistics time for worker's behaviour and also an important source of macroscopic mobility information. The models reflect the behaviour of real urban traffic changing with the time of work and life. The most typical is UDel Mobility Model. (Kim and Bohacek 2005, Kim, Sridhara and Bohacek 2009)

- **Trace-Based Models:** By recording GPS and including such kind of statistics from transport department as real trace data into mobility model, in order to save tedious computation to obtain an accurate model. However, the recording data is always from special vehicles such as taxi or bus because the models are not the same of private vehicles. Therefore, few number of public statistics data is referred. This method is becoming more popular by using mathematical models to predict the possible trace of mobility patterns, there is much related research such as CRAWDAD, MIT Reality Mining and USC MobiLib (Hsu and Helmy 2005, Scott, et al. 2006, Eagle and Pentland 2006).

- **Traffic Simulator-Based Models:** By real traces and behaviour surveys, some research teams refined the synthetic model to become more realistic traffic simulators. There are well-known microscopic traffic simulators such as SUMO, CORSIM, VISSM or VanetMobiSim (SUMO 2015, CORSIM 2015, VISSM 2015, VanetMobiSim 2015). Although they are able to simulate urban traffic, energy consumption, even pollution or noise level monitoring, however, there is no interface have been developed in each of them, these traffic simulators need to import their mobility patterns with other network simulators, for instance, SUMO + NS3, these two simulators will discuss in section 2.6. (Harri, Filali and Bonnet 2009)

In general, one developed mobility model needs to be validated, namely validation. It means the model needs to compare with the real traces and real topologies. It is also by comparing the motion patterns with other acknowledged simulators if the access to real traces or real topologies is not possible, for example, VISSIM, SUMO, etc. This method is called delegated validation. The validation step usually should be followed before being used to evaluate protocols for vehicular networks. (Tuduce and Gross 2005, Harri, Filali and Bonnet 2009)

The mobility model is necessary and important for VANET research. Whether complex mathematical model or trace-based movement patterns, these approaches should follow and simulate real traffic condition as much as possible. However these approaches are not mutually exclusive, they are interdependent all depend on each other. In order to obtain a high level of precision that researchers also generally use a combination model with different approaches.

2.3 Standard Traffic Prediction Model

2.3.1 Overview of Traffic Prediction

Traffic information prediction is an important part of the modern science of traffic and intelligent transportation systems. It is based on history or existing transportation factors and statistics, then uses the intelligent computational method to forecast the transport system state in the target area in the future. Traffic prediction draws the advanced techniques of prediction and artificial intelligence, especially for the demand of diversity, real-time and accuracy in traffic information. Overall the development of traffic prediction presents intelligent, complex and composite trends. (Smith and Demetsky 1997, van Hinsbergen and Sanders 2007)

The importance of traffic information prediction summarise (Jiang, et al. 2009, Xu and Fu 2009):

(i): Traffic information prediction can grasp the development of the local transport business and carrying trade. This is an essential basis for the development of transport development strategies and policies.

(ii): Its role mainly reflect in the forecast of socioeconomic, travel and other aspects. It is an important part of urban transport plan, that is, the traffic information prediction is the premise of the transportation planning program.

(iii): Traffic information prediction is a technical support for traffic incident detecting, improving the forecasting capability of the traffic incident, reducing the negative impact of the event.

(iv): As an important information processing technology, traffic information prediction can not only constitute independent functional systems, but it is also an important part of many ITS systems. There is a relationship between traffic information prediction and ITS at below:

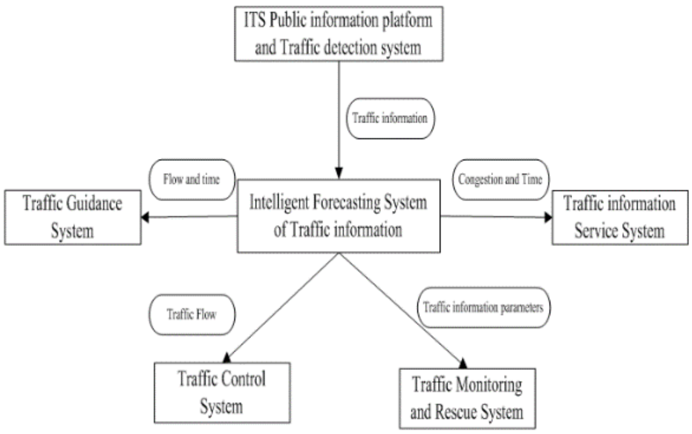


Fig. 2.8 Traffic Prediction in ITS

The object of traffic prediction is traffic information, how to obtain the real and effective traffic information that the primary research in prediction field. The traffic information has differently predictable in a different state of traffic flow or different condition, for example, we can predict the traffic flow at one intersection with different duration in one day according to historical records; or we may predict the time of traffic remission if the traffic occurs congestion or accident on one road. In this case, such reference different prediction methods or models to analyse and evaluate traffic information. (Jessop 1990) There are some related works for traffic collision prediction models evaluation in the literature of (Lovegrove and Sayed 2006), and (Smith and Demetsky 1994) introduce how to create the traffic prediction model and compare three categories of predictive modelling approaches.

2.3.2 Characteristics of Traffic Flow

The cluster of vehicles travelling on the road called traffic flow that represents the running state of the vehicles. (May 1990) The features of traffic flow have variable, wide amplitude and randomness. It can be divided into non-congested traffic flow and congested traffic flow, according to research in the different direction. (Zhu, Wang and Xiang 2008)

For non-congested traffic flow, all vehicles on the road are maintained and sustained stable travelling. Within a certain time interval on a road, the traffic flow is a random variable, it will change as traffic density, and it is nonlinear. Typically present two states which are free and saturation. The state of free means that the drivers have a greater freedom in the choice of lanes and speed, usually occurs at the traffic density of fewer than 12 vehicles/(km*lanes). (Zhu, Wang and Xiang 2008) However in the saturation state, the drivers always follow the vehicle in front, their speed constrained by the vehicle in front and converted lane possible is small than free state. If certain vehicle changes its travelling state, it will spread to the entire traffic flow. The saturation state typically happens at a traffic density between15~40 vehicles/(km*lanes) (Sasaki and Nagatani 2003, Zhu, Wang and Xiang 2008). In the literature of (Ditlevsen 1994), Ove verified the relationship between white-noise and traffic load, paper derived formulas for the mean and intensity of the white-noise traffic load field in terms of traffic parameters and vehicle weight means and variance. From (Zhu, Wang and Xiang 2008), they point out the white-noise has big variance in the short-time free state of traffic flow, otherwise white-noise accounts for a smaller proportion and variance in the saturation state.

The main cause of congestion due to road traffic capacity is less than the traffic demand, the location and time of congestion have regularity, generally occurs in rush-hour traffic and road bottlenecks (Arnott, De Palma and Lindsey 1991). If traffic demand equal or less than traffic capacity, called traffic saturation (Adams 1936). In the opposite, if traffic demand greater than traffic capacity, called traffic congestion or traffic jam. Herewith, there is a gradual transition from non-congestion to traffic congestion, the changing of generation traffic parameters by congestion around is continuous. The traffic parameters (such as traffic flow rate, speed or traffic density) generally exhibit gradually rise curve or gradually down curve. By analysing the trends of these parameters can predict the traffic condition in advance and take measures to avoid the traffic congestion. (Ding, Zhao and Jiao 2002) Another cause of congestion is sporadic incidents, for some reason led to a sharp decline in road capacity such as traffic accident. In this case, the location and time of congestion are uncertain and is difficult to forecast. Therefore, the traffic parameters reflect the discontinuous change and will be a peak or bottom in s short time. The relationships between traffic flow rate, speed and traffic density are shown in Fig. 2.9: (Manual 2000)

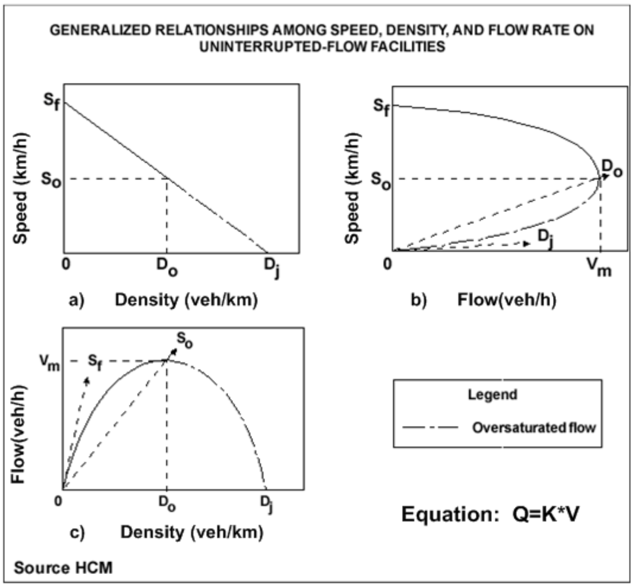


Fig. 2.9 Relationship between Flow rate, Speed and Density

Flow Rate (Q): the equivalent hourly rate at which vehicles pass over a given point during a given time interval of less than one hour. (Okutani and Stephanedes 1984) However, this is different from "traffic volume", which is the number of vehicles observed to pass a point during a specified time interval, such as annual or average daily traffic.(Kalašová and Krchová 2011, Manual 2000)

Speed (V): defined as a rate of motion expressed as distance per unit of time. The curves in Fig. 2.9 utilise "average travel speed", which is the length of the lane divide by average travel time of vehicle passing through it. (Kalašová and Krchová 2011)

Density (K): the number of vehicles occupying a length of a lane during a particular moment. For the curves shown in Fig. 2.9, density is averaged over time, and it is usually expressed as vehicles per kilometer (veh/km). (Kalašová and Krchová 2011)

Flow rate is the product of speed and density from the equation, while the diagrams in Fig. 2.9 show a zero flow rate occurs under two situations when one or both of these terms is zero. One is when density goes to the highest value that means all vehicles cannot move and cannot pass the point on the lane, so speed is zero and flow rate is zero, address to Fig. 2.9 a). Also, the second situation in Fig. 2.9 b), there is no vehicle on the lanes so that the flow rate and density are both zero. Moreover between two extreme points in Fig. 2.9 b) & c), the dynamics of traffic flow appear a maximizing value. As flow increases from zero, density also increases, since more vehicles are on the roadway. When this happens, speed declines because of the each vehicle needs to keep safety distance to avoid a collision and the interaction of vehicles. However, this decline is negligible at low and medium densities and flow rates. As the density further increases in the Fig. 2.9 a) & c), speed decreases significantly just before capacity is achieved, with capacity being defined as the product of density and speed resulting in the maximum flow rate. This situation is shown as optimum speed, so often called critical speed, optimum density(Do) sometimes referred to as critical density, and maximum flow rate is called Vm. In general, this maximum flow rate occurs at a speed between 56 and 80 km/h, furthermore, this situation is also called traffic saturation. (Biham, Middleton and Levine 1992, Manual 2000, Kalašová and Krchová 2011)

While the diagram described an idealized concept and status, in reality the transportation is very complex, the traffic demand is not constant also the traffic capacity. They can change at any moment because of many reasons such as weather, traffic incidents and constructions, even the drivers. Therefore, the actual situation is more ambiguous on a road section in a particular time. In this case, it may be possible to predict the average behaviour and value, but never the precise behaviour or value. (Greenberg 1959, Hall 1996, Manual 2000)

2.3.3 Classification of Traffic Prediction Model

In recent years, dynamic prediction of traffic flow is being more and more attention in the world so that many advanced methods in other scientific fields are applied here to established adaptive prediction model in different locations, road conditions and time periods. The macro-model is mainly dynamic traffic assignment models among them, and micro-model has various measuring model. In the case of the current development of forecast, there are many traffic information prediction models and methods. It includes the Regression Analysis Method, Time Series Prediction, Grey Theory Prediction Model, Markov Prediction Model, Neural Network, Kalman Filter and so on

(Nicholson and Swann 1974, Harvey 1990, Guan-jun 2000, Chen, et al. 2003, Yu, et al. 2003). However, each prediction method contains several prediction models. For example, when prediction model is based primarily on the traditional mathematical and physical methods, it should involve Time Series Prediction Model, Kalman Filter Model, Parameter Regression Model and Exponential Smoothing Model. If the prediction model uses modern science technologies as a research tool, the no-Parameter Regression Model, Wavelet Theory, Multi-fractal Method and Compound Forecasting Model of Neural Network are involved (Riedi, et al. 1999, Peng-Jian and Jin-Sheng 2007). The traffic forecast method reflects diversity and it was classified on the basis of the following six aspects:

Cycle length: The forecast period could be a year, month, day, hour or even minutes for different forecast demand. The corresponding prediction method is divided into long-term, medium and short-term. (Yin, et al. 2002, Papagiannaki, et al. 2003)

Characteristic: It mainly includes the quantitative prediction method and qualitative prediction method. If the characterization of the traffic information is a serial data, the quantitative prediction should be taken. Otherwise, the qualitative prediction method will be used. (Weigend, Huberman and Rumelhart 1990)

State: It is usually divided into static prediction and dynamic prediction. The static prediction does not contain the time changes and predicts the causal relationship of transport phenomena in the same period. Dynamic prediction needs to consider the time changes, and according to the history and current situations of the transport phenomena, when people forecast the traffic trends in the future. (Ben-Akiva, et al. 1998)

Traffic kinetic characteristics: This category includes deterministic prediction, chaotic prediction and stochastic prediction. The urban transport system is composed of people, vehicles, road, weather and other factors. Its essence is a complex non-linear system. In the different traffic conditions, traffic flow has different kinetic behaviour. (Herman, Lam and Prigogine 1971, Disbro and Frame 1989)

Number of prediction models: When people forecast the traffic information, they would be required to choose a single prediction or combination prediction. Single prediction means just use only one model to forecast the traffic information. And the combination prediction can use more than one model to forecast so that it has the complementary advantage features. (Smith and Demetsky 1997)

Technology: It refers to conventional prediction and intelligent prediction. Many methods and models mentioned at above, conventional prediction usually includes Regression analysis Method and Time Series Method of Exponential Smoothing, Moving Average, Seasonal Coefficients and Box-Jenkins. Intelligent prediction comprises Grey Theory, Kalman Filter, Support Vector Machine (SVM) Theory, Chaos Theory, Neural Network and Multi- Agent prediction method. (Disbro and Frame 1989, Williams, Durvasula and Brown 1998, Vapnik and Vapnik 1998, Hong 2011)

The prediction method should be selected according to different demands, objectives, technology and other factors. The prediction methods always conform to the development trend of the ITS, absorb the new advanced technologies to improve themselves and present intelligent, complex and combination of trends. This research is based on the combined model concept to deliver the predicted results.

2.3.4 Traffic Prediction Model Related Work

As early as the 1970s, researcher began to study the dynamic prediction of traffic on the micro-level. In 1974, Nicholson and Swann proposed using the spectral analysis model to predict the traffic flow volumes (Nicholson and Swann 1974). Ahmed and Cook analysed and discussed the freeway traffic time-series data by using Box-Jenkins techniques, but the accuracy of the result is not ideal (Ahmed and Cook 1979). However, in 1980, Nihan and Holmesland reused Box-Jenkins techniques to predict traffic flow for a road section under known conditions of traffic historical data in four years, and they made better results than Ahmed and Cook (Nihan and Holmesland 1980). Thus, it can be seen that this method requires more historical data to generate the model. In 1990, Davis et.al studied with adjustable prediction system to predict the freeway traffic flow volumes, and to estimate whether the traffic congestion (Davis, et al. 1990). At the same year, they also established a high-level adaptive prediction system and applied to real-time prediction and data collection in the urban traffic network. In the next year, they established a nonparametric regression prediction model and think in some cases it is better than the time series model (Davis and Nihan 1991a). Furthermore, other researchers are also trying other prediction methods and models, for example, Iwao Okutani created dynamic prediction of traffic volume by Kalman Filtering Theory

(Okutani and Stephanedes 1984); in 1993, Brain and Michae used and verified that the Backpropagation Neural Network (BP Neural Network) was clearly superior to data-based algorithm and a time-series model in a short-term period (Smith and Demetsky 1994). In the literature of (Van Der Voort, Dougherty and Watson 1996), they used a Kohonen self-organizing map as an initial classifier and associated Auto-Regressive Integrated Moving Average (ARIMA) model and BP Neural Network, so that the time series model such as ARIMA has a wider range of adaptability and portability. Therewith, in the literature of (Clark, Dougherty and Kirby 1993, Kirby, Watson and Dougherty 1997, Chen and Grant-Muller 2001), they used a dynamic time series method and polynomials theory to analyse and describe the traffic flow variation and state for short-term traffic forecasting. Even for the accident prediction models, Mohammed Quddus introduced an integer-valued autoregressive (INAR) Poisson models for the time series analysis of traffic accidents in the UK, the results showed that INAR Poisson models are much better than the ARIMA model for the case of the disaggregated time series traffic accident data where the counts are relatively low (Quddus 2008). Where after his team in the literature of (Wang, Quddus and Ison 2011) used Bayesian spatial model and a mixed logit model as a two-stage model for accident frequency and severity analysis respectively, it used more detailed individual accident level data to predict low-frequency accidents.

One common combinative method used traffic flow- density- speed model which is in the Fig. 2.9 in subsection 2.3.2 and regression method to predict the traffic condition (Davis and Nihan 1991b). The advantages of this method are mature enough in traffic flow research, testing equipment is relatively simple, small quantity and low cost. However, the disadvantages are also obvious, that reflected in poor adaptation and instantaneity. Moreover, the changing of traffic flow is nonlinear in general, purely on the basis of pre-determined regression equation applies only in certain road sections with the specific flow range, and it cannot amend errors. While the results of linear regression prediction have less fluctuation that inconsistent with the actual observations. Therefore, ordinary regression algorithm cannot satisfy with the instantaneity and accuracy of traffic prediction. Hence, many researchers established traffic prediction model by means of time series theory. ARIMA model is simple time series model, its basic idea is (Lee and Fambro 1999):

$$\mathrm{Z(t)} = \beta_0 + \beta_1 Z(t-1) + \beta_2 Z(t-2) + \beta_3 Z(t-3) + \cdots + \beta_n Z(t-n) + \varepsilon \qquad (2.3\text{-}1)$$

In this equation (2.3-1), $Z(t)$ is the predicted result that describes the traffic condition at t time for a road section; $Z(t-1)$, $Z(t-2), \ldots$, $Z(t-n)$ are real data, or observed value; ε is error and β is coefficients that undetermined by creators. This method is based on time series data from historical record to create model, because it can reflect the changes in traffic flow parameters. However, we need to increase the observed data and refine the model in order to improve forecast accuracy. Such as these parametric models use the historical data in search of the mathematical relationship between all factors. (Smith, Williams and Oswald 2002)

Nevertheless, the Non-parametric model is suitable for uncertain and non-linear dynamic traffic and based on the chaos theory with enough historical data also does not require an empirical model. It looks for the similar near points with the current point in the historical data and uses these near points to predict the traffic in the future time. This method considers that the relationships between all factors are implied in the historical data. (Smith, Williams and Oswald 2002) Its essence is to obtain information from the historical data instead of modelling for the data so that there is no smoothing with the data. Therefore, especially when there is a special event, nonparametric model's prediction effects precise than parametric models.

2.3.5 Performance Evaluation of Traffic Prediction Model

Currently, there is a variety of methods and techniques to be applied to the field of short-term traffic prediction. These methods and techniques establish prediction models from a different views, and each model has a certain range of adaptation, application conditions and advantages in precision, real-time and portability. Hence, the evaluation of prediction model needs to be considered many factors. In the short-term traffic prediction, commonly used indicators have an absolute error, relative error, mean squared error (MSE) and correlation (van Lint, Hoogendoorn and van Zuylen 2002, Zhu, Wang and Xiang 2008).

The absolute error is different between the predicted value and observed value. The relative error describes the degree that predicted value deviates from the observed value, and it is better than absolute error when compares the predicted effect of different sequences of observations. (Ishak and Al-Deek 2002) MSE not only reflect the size of the errors but also to describe the degree of error distribution for concentration and discreteness. If MSE gets bigger that means the errors are more discrete, and the effect of prediction are poorly (Dia 2001). The correlation considers the geometry specifications of the predicted curve and observed curve, describes the degree of fitting for the trend with two curves. It refers to uniformity coefficient and the linearity, and they will best approximate to 1. (Zhu, Wang and Xiang 2008)

Different predictions correspond to different predicted results. Such as some prediction models, the mean absolute error is small, but the maximum absolute error is large; even some models have lower mean errors, but the maximum errors are much higher. In the literature of (Zhu, Wang and Xiang 2008), they compared several commonly used prediction model based on the same section of traffic condition, as shown in Table 2.5 and the formula of corresponding evaluation indicators show at Table 2.6.

Table 2.5 Evaluation of Traffic Prediction Model

	MAE1	MAXAE2	MRE3	MAXRE4	MSE	Uniformity Coefficient	Linearity
Moving Average	7.73	18.0	7.95%	31.7%	10.20	0.908	0.952
ARIMA Model	7.01	24.0	7.15%	37.2%	9.3	0.927	0.940
Neural Network	6.55	25.0	6.25%	34.0%	9.47	0.958	0.996
Wavelet Theory	5.19	22	6.0%	34.0%	9.5	0.980	0.976
Fractal Method	7.03	27	6.73%	31.9%	9.14	0.965	0.942

[1]Mean Absolute Error [2]Maximum Absolute Error [3]Mean Relative error [4]Maximum Relative Error

Evaluation of those indicators seemed higher, but they used a large number of the historical data over the same period as observed values, thus predicted the traffic flow with different models. And the model's adaptation conditions are different so that the actual performance will be very different. In the literature of (Zhu, Wang and Xiang 2008), they also compared traffic flow prediction, traffic density prediction and speed prediction, the results showed that the effect of the density was the best, and traffic flow prediction was better than speed prediction. Because for the speed, the changing of traffic density and traffic flow is relatively more stable. Therefore, for the speed prediction, the similarity will be more likely to reflect the difference predicted values and observed values.

Table 2.6 Formula of Traffic Evaluation Indicators

Evaluation Indicator	Formula
Mean Absolute Error	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - x_i\rvert$
Maximum Absolute Error	$\max\{\lvert y_i - x_i\rvert \mid i = 1,2,\dots,n\}$
Mean Relative Error	$\frac{1}{n}\sum_{i=1}^{n}\frac{\lvert y_i - x_i\rvert}{x_i} * 100\%$
Maximum Relative Error	$\max\left\{\frac{\lvert y_i - x_i\rvert}{x_i} * 100\% \mid i = 1,2 \dots, n\right\}$
Mean Squared Error	$\sqrt{\sum_{i=1}^{n}\left(y_i - x_i\right)^2}$
Uniformity Coefficient	$1 - \frac{\sqrt{\sum_{i=1}^{n}(y_i - x_i)^2}}{\sqrt{\sum_{i=1}^{n} y_i^2} + \sqrt{\sum_{i=1}^{n} x_i^2}}$
Linearity	$1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Note: y_i is predicted results, x_i is observed values, n is number of sample.

About similarity comes from the law of cosines, judges the size of an angle to determine the degree of similarity, and the smaller angle is similar. When the angle is close to 0, means similarity close to 1, the predicted value is more similar to the observed value. The similarity formula is (Breese, Heckerman and Kadie 1998):

$$\text{Similarity} = \frac{\sum_{i=1}^{n}(y_i x_i)}{\sqrt{\sum_{i=1}^{n}(y_i)\sum_{i=1}^{n}(x_i)}} \tag{2.3-2}$$

The most common method is cosine similarity as a measure of degree between expression profiles (Tan, Steinbach and Kumar 2006). The cosine similarity is a measure of similarity between two of a vector space that measures the angle between them and focuses on the difference in the direction of two vectors. By the same token, if the angle is small on behalf of more similar of two objectives, conversely, the greater angle represents poor similar of two objectives. In the triangle coefficients, cosine angle is in [-1, 1]. When two vectors have the same direction, the value of cosine similarity is 1; when the angle is 90° of two vectors, the cosine similarity value is 0; when two vectors are in the opposite direction, the cosine similarity is -1. Because in the comparison process, the size of the vector is not considered, only taking into account the direction of the vectors. Therefore, the cosine similarity is typically used in positive space which less than 90° of two vectors, thus the value of cosine similarity should be bounded in [0, 1]. And the formula of cosine similarity represents as:

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{n}(y_i x_i)}{\sqrt{\sum_{i=1}^{n}(y_i)^2 \sum_{i=1}^{n}(x_i)^2}} \tag{2.3-3}$$

Furthermore, the famous statistician Karl Pearson designed the correlation coefficient as a statistical indicator (Pearson 1901), in order to reflect the degree of correlation between closely related variables. Pearson's correlation coefficient is the covariance of two variables divided by the product of their standard deviations and focuses on the linear correlation between two variables. The definition of Pearson's correlation coefficient is (Charles 2015, Microsoft 2015):

$$\text{Pearson's correlation coefficient } r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2.3-4}$$

$$\text{Or } P_{X,Y} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}} \tag{2.3-5}$$

Where is:

r>0, variables have positive correlations.

r<0, variables have negative correlations.

$|r|$=1, variables have complete linear correlations, that is function relationship.

r=0, variables do not have a linear relationship.

0<$|r|$<1, variables has linear relationship. Closer to 1 is best.

It should be noted that the Pearson's correlation coefficient is not the only correlation coefficient, but most common one so that the following explanations of the correlation coefficient or correlation means Pearson's correlation coefficient or Pearson's correlation. The correlation coefficient number can reflect the exact degrees of linear relationship and direction between each variable, and its value is bounded in [-1, 1]. When the linear relationship increases between variables, the correlation coefficient tends to 1 or -1; when one variable increases and another variable also increases, their correlation is positive and correlation coefficient greater than 0; if one variable increases and another variable was reduced, their correlation is negative and the correlation coefficient is less than 0; if the correlation coefficient is 0, it indicates that there is no correlation between them. Therefore, the value of Pearson correlation-based similarity is between -1 and 1. In the literature of (Zou, Tuncali and Silverman 2003), Dr. Zou's team measured the linear and nonlinear relationship between two continuous variables and analysed linear regression between a predictor and an outcome variable via Pearson correlation coefficient, but it was applied to radiologic studies.

By evaluating of similarity is more suitable for values and trends between predicted results and actual values for travel speed prediction, and to verify the speed predicted curve potential fit with the observation speed curve. In this project, the cosine similarity and Pearson correlation-based similarity will be used to evaluating the prediction model.

2.4 Generation Prediction Models for VANET

The literature (Yang, et al. 2010) shows traffic flow is negatively related to average speed. So there are many short-term traffic flow prediction models could also use in prediction of average speed. On the classification of the prediction model that we have mentioned in the previous section. However, the average speed is more useful than traffic flow or traffic density to the drivers. The Moving Average (MA) and Auto-Regressive Integrated Moving Average (ARIMA) are most popular in time series short-term prediction as the mathematical prediction model. They were based on the statistic historical data, and assumed that the future data is a continuation of historical data. In this case, the traffic data collection is necessary for the short-term traffic prediction. The traffic data is aggregated by wired networks, wireless network or monitor sensor network to collect real-time traffic data (Su, Zhang and Yu 2007, Laisheng, et al. 2009). The monitor sensor usually is deployed where are easily congested intersections and interchanges. In general, managers in the traffic control centre use a certain prediction model through real-time data from sensors to rolling predict the traffic. This method as sensor network has been very mature in traffic management (Laisheng, et al. 2009).

Nowadays, wireless device, computing device and navigation system in vehicles are a trend in the ITS system. The short-time prediction for VANET is also a proactive technique in ITS system. It is mainly to share and collect traffic conditions based on the C2C, C2I or C2X communication, and it uses computing equipment in the vehicle to analyse and predict the traffic in the future time. Such as computing equipment certainly embeds in the corresponding prediction models. In this section will introduce the prediction models and algorithms based on the VANET.

2.4.1 Moving Average Model

Moving Average method (MA) is based on the evolvement of historical time series term by term, calculates the average of different subsets of the full historical data in an orderly way to reflect the trend in a term or period. When the historical data changes irregular or undulating, MA could be used to eliminate the influence of these factors. In the concept of MA, each data is seen as equals to others in effect. However, for real-time dynamic prediction, the most recent data contain more information about future events. We need to take into account the importance of each data and give more weight to the most recent data. This is the basic idea of the weighted moving average method, the formula is:

$$M_t = w_1 z_t + w_2 z_{t-1} + \cdots + w_N z_{t-N+1} \ , \qquad t \geq N \tag{2.4-1}$$

$$w_1 + w_2 + \cdots + w_N = 1 \tag{2.4-2}$$

$M_t -$ the weighted moving average at time

$w_\mathrm{i} -$ the weight of z_{t-i+1}

When t forwards by one that means a new data, and we need to add this new data to the model, meanwhile delete the oldest data. Then we can obtain the new average value, namely to achieve the approach of moving forward step by step. In the weighted moving average method, choice of weight is very important and has an empirical. The principle is that the recent data has more weighted, and forward data has small weighted, and specific weighted need a comprehensive analysis of the historical time series (Hunter 1986, Min and Wynter 2011).

2.4.2 ARIMA Model

ARIMA model as a time series prediction model that generated by George E.P. Box and Gwilym M. Jenkinsin in the 1970s, also namely Box-Jenkins method (Anderson 1976, Box and Jenkins 1994). Classical ARIMA (p, d, q) is called Differential Autoregressive Moving Average model where d is differencing operations; p is Autoregressive terms; q is Moving Average terms. In this case, if without the differencing operations that mean d=0, the equation of ARIMA should be followed (Van Der Voort, Dougherty and Watson 1996):

AR(p)	$Y_t = \varepsilon_t + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p}$	(2.4-3)
MA(q)	$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_p \varepsilon_{t-p}$	(2.4-4)
ARIMA(p,d,q)	$Y_t = \varepsilon_t + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_p \varepsilon_{t-p}$	(2.4-5)

Y_t is the prediction of time series Y at t time

$Y_{t-1} \dots Y_{t-p}$ are the previous p values of the time series Y

φ, θ are coefficients; ε_t is white noise

Comparison of three equations at above, when q=0, ARIMA (p,0,0) model is AR(p) model; when p=0, ARIMA(0,0,q) mode is the MA(q) model. In the studies or main cases of prediction, the value of p and q usually take 0, 1 or 2 for the modelling. The ARIMA model applies to short-term traffic prediction if we can defensibly assume that traffic data series can be reflected the change rule of traffic condition (Zhu, Wang and Xiang 2008).

Furthermore, a variation of ARIMA model has been proposed as an extension which is seasonal ARIMA model (Box and Jenkins 1994) that can be for long–term time series. The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model which are ARIMA(p,d,q) × (P,D,Q)S. Let ARIMA(p,d,q) that all of them parameters are non-negative integers. Defined the first-order differential operator ∇ as, then the differential operator $\nabla_{Z_t} = Z_t - Z_{t-1}$, and then the delay operator B is $\nabla = 1 - \mathrm{B}$. ∇^d is differential operator defined in the usual binomial expansion where is $\nabla^d = (1-B)^d$. In the following, let $\{Z_t : t = 0,1,\dots,n\}$ as a non-stationary series, $\{x_t\}$ ARIMA (p, q) ×(P,Q)S sequence. There is a positive integer d, making $x_t = \nabla^d Z_t, t > d$, then has $\Phi(B^S)\varphi(B)(1-B)^d Z_t = \Theta(B^S)\theta(B)d_t$, where $\{d_t : t = 0,1,\dots,n\}$ is the white noise, both non–seasonal and seasonal are polynomials in complex variables and could be written as follow (Van Der Voort, Dougherty and Watson 1996, Zhu, Wang and Xiang 2008, Xu and Fu 2009):

Non-seasonal AR:	$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$	(2.4-6)
Non-seasonal MA:	$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$	(2.4-7)
Seasonal AR:	$\Phi(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \cdots - \Phi_P B^{PS}$	(2.4-8)
Seasonal MA:	$\Theta(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \cdots - \Theta_Q B^{QS}$	(2.4-9)

The theoretical justification for modelling univariate time series of traffic data as ARIMA processes is founded that the future changes of this sequence exist dependency and continuity with forwarding changes. From a practical perspective, seasonal ARIMA models provide linear state transition equations that can be applied recursively to produce single and multiple interval predictions (Williams and Hoel 2003).

Mascha et al. came up with a hybrid method for short-term traffic prediction with and Kohonen maps and ARIMA model, but called KARIMA method (Van Der Voort, Dougherty and Watson 1996). This method applied a Kohonen self-organizing map as an initial classifier and each class has an individually tuned ARIMA model associated with it (Kohonen 1990, Van Der Voort, Dougherty and Watson 1996). They used this method to forecast the traffic flow in a French motorway in an hour and demonstrated the performance of the combined model more superior than pure ARIMA model or BP neural network method. They pointed out that traffic flow data exhibits a very high degree of nonlinear and divide and conquer scheme of the layered models such as Map layer and input layer seems well suited to overcoming this difficulty.

Williams et al. in the literature of (Williams and Hoel 2003) introduced modelling and predicting univariate traffic data as seasonal ARIMA process based on the Wold's decomposition theorem. They verified that a one-week lagged seasonal difference applied to discrete interval traffic data will yield a weekly stationary transformation. They also provided a comparison of the seasonal ARIMA prediction and KARIMA prediction in the same time period, and KARIMA were outperformed by ARIMA. They believed that seasonal and other effect had a substantial impact on the results for their experiments. But they also pointed out that traffic data collected from a peak time in summer holiday rather than normal weekly in their previous statement in 1999 (Williams and Hoel 1999). Therefore, the seasonal ARIMA are not as clearly demonstrated in relation to these atypical data, and not appropriate for general traffic condition prediction. In the literature of (Williams 2001), Williams presented an ARIMAX model to improved prediction performance over univariate prediction model that model assumed constant transfer function parameters. However, the correlation between upstream and downstream observation vary with prevailing traffic conditions, especially traffic speed, so he suggested to extend self-tuning multivariate prediction model for the varying correlations. Dingding et al. (Zhou, Chen and Dong 2013) used a time series forecasting model which is an improved ARMA model, namely ARFIMA (Ravishanker and Ray 2002) in the network traffic prediction. They combined with network traffic of CERNET backbone and the ARFIMA model to compare with ARMA and ARIMA model for China Northeast network centre. The results show that the prediction efficiency and accuracy have increased significantly and not susceptible to sampling. In the literature of (Yu and Zhang 2004), Guoqiang and Chang Shui proposed switching ARIMA model to traffic flow prediction with the real data obtained from Traffic Management Bureau of Beijing. They used a separate ARIMA model to represent different characteristics with of traffic flow. In conclusion, this literature states that the prediction efficiency and precision of ARIMA model have relatively larger research and improvement space.

2.4.3 K-Nearest Neighbour Model

The K-Nearest Neighbour (KNN) model as a machine learning algorithm is a non-parametric method used for classification and regression (Altman 1992). In the KNN classification prediction, the general choice is the majority of voting method, that is, an object is classified by a majority vote of its neighbours, with the object being assigned to the class most common among its K- nearest neighbours. In the KNN regression, the general method uses the average, that is, the regression predicted output is the average of the sample output of its K-nearest neighbours. But the difference is not significant between them in general. The KNN algorithm model consists of three parts: selection of distance metric; selection of K value; selection of classification decision.

A commonly used distance metric for a continuous variable is Euclidean distance, and we will use it in the following chapters. There are also Minkowski distance and Taxicab geometry. The nearest neighbour point is different for the different distance. In the actual implementation, the data might be normalized. The formulas of these three distance:

Euclidean distance: $$dist(X,Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{2.4-10}$$

Minkowski distance: $$dist(X,Y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p} \tag{2.4-11}$$

Taxicab geometry: $$dist(X,Y) = \sum_{i=1}^{n} |x_i - y_i| \tag{2.4-12}$$

Different K values have a significant effect on the KNN model. If the K value is too small, this would be the equivalent of the only sample that is closer to the input object and it will have an effect on the prediction results. The downside is that the prediction results are very sensitive to the sample points of the neighbours, and if the adjacent sample point is just noise points, the prediction results will go wrong. In other words, a decreasing the K value means that the model becomes complicated and overfitting. On the contrary, if the K value is too large, then the prediction results depend on the sample points in the larger neighbour of the input object. The disadvantage is that the prediction results may be disturbed by the distant sample points and the prediction results are inaccurate. In the other words, increasing the K value means that the model becomes simple. Normally, the K value is usually taken with a smaller and odd value, such as 3 or 5.

The classification decision of KNN is also varied, different classification decisions produce different prediction results. The commonly used classification decision is “majority voting”, that is, the classification of the input object is decided in the majority of the classification by the input object of K neighbour sample points. In addition, there are methods such as weighted statistical methods for decisions.

The advantages of KNN are simple and effective; the complexity of training time is lower than support vector machine (SVM); it can be used to classification and regression also for nonlinear classification; there is no assumptions about the data and not sensitive to the abnormal point. The disadvantages of KNN are: there is a large amount of calculation, especially when the number of features is very large; when the sample is not balanced, the prediction accuracy of rare classification is low. Furthermore, the algorithm is more suitable for the classification of the relatively large sample size, and but there is error-prone when using this algorithm for the smaller sample size of the classification.

Some researchers are working on the traffic prediction by KNN algorithm. Zhang et al. (Zhang, et al. 2013) presented a KNN model for the short-term urban expressway traffic flow prediction system which was built in the historical database, the search mechanism and algorithm parameters, and prediction plan. They demonstrated that the feasibility of the average and weighted KNN method were used in short-term traffic flow prediction. In the study of (Yin and Shang 2016), they compared multivariate time series, univariate time series and KNN model in traffic variables prediction. The results showed that multivariate time series prediction was better than univariate time series and KNN model.

2.5 Ad-hoc Network Routing Protocol

The ad-hoc network has highly dynamic topologies with more difficulties and complexities to route messages. In order to ensure timely and accurate transmission of traffic data, so that an efficient routing protocol is a prerequisite for this study. The role of a routing protocol in ad-hoc networks that discoveries and maintains the connectivity between nodes (Karande and Kulkarni 2013). Regarding routing research areas, highly dynamic topologies pose the main challenge. Researchers focused on routing improvement, thereby increasing the network performance (Alotaibi and Mukherjee 2012). The evaluation of routing performance should be following that routing decides which one is the centrally controlled node; routing needs self-healing to re-organise nodes when disconnecting between nodes.

There is a great variety of wireless routing protocols basis of route generation and maintenance that main categories include topology based, geographic position based, geo-casting based, and more details show at Fig. 2.10 (Alotaibi and Mukherjee 2012):

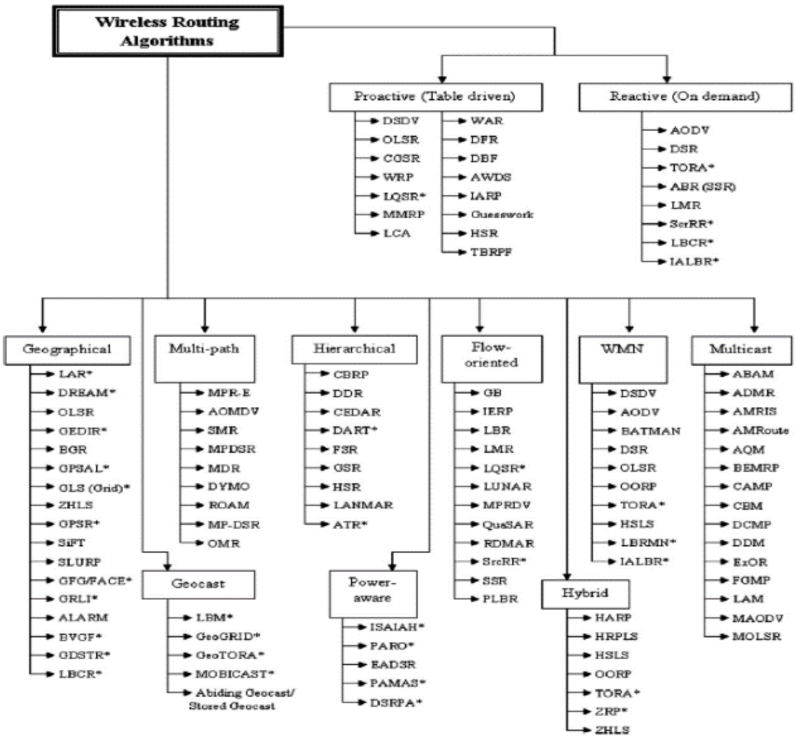


Fig. 2.10 A structure of routing-algorithm categories

According to the structure at above, there are proactive routing protocols and reactive routing protocols for ad hoc networks (Chlamtac, Conti and Liu 2003). The proactive or table-driven routing collects routing information and updates routing table periodically while reactive or on-demand routing builds routes when nodes require, then updates corresponding route entries. MANET and VANET are both no relying on fixed roadside infrastructure for communication and similar features as mentioned in section 2.2.1 and section 2.2.2, thus most ad hoc routing protocols can be accepted for VANET, several existing routing protocols, such as ADOV, DSR, OLSR and DSDV (Ho, Ho and Hua 2008, Karande and Kulkarni 2013). But each protocol has its own advantages and disadvantages for different applications. There is no almighty routing protocol for any environment, all routing protocols need to design be customised or demand.

2.5.1 Proactive Routing Protocols

Proactive routing can build its own routing table based on the information that each node is responsible for a routing table which learns and records by exchanging information among the network's routers, also namely table driven routing. This is achieved through information exchanges among nodes on a regular basis to update the routing table at each node (Alotaibi and Mukherjee 2012). If a sender node consults the routing table and the path information that is immediately updated and used by the node. It is worth noting that although obtaining the path information is fast, updating and maintaining the information table for each node in the network requires high overhead traffic and a significant amount of bandwidth. In addition, the movement has an influence on connection and frequency of even-triggered updates in highly dynamic topologies. The process of maintaining the routes to the reachable nodes is continuous even there is no data transmission on these routes. There are requires additional overhead with high mobility (Royer and Toh 1999).

There are several representatives and well-known proactive routing protocols, such as Destination-Sequenced Distance Vector (DSDV) (Perkins and Bhagwat 1994) and Optimised Link State Routing (OLSR) (Jacquet, et al. 2001b):

DSDV (Destination-Sequenced Distance Vector) is a table driven routing protocol developed by Bellman-Ford algorithm for MANET (Perkins and Bhagwat 1994). Each mobile node needs to maintain a routing table and periodically broadcast its routing information in DSDV. Routing table entry contains the destination node, routing hops, next hop address, destination routing sequence number and time of install. The sequence number is assigned by the destination nodes and uses to determine routing whether the time out and prevent looping. Each node must periodically exchange routing information with a neighbour; furthermore, it also can change the routing table to trigger routing updates. In the case of a new message received, the sequence number must be incremented by the node and update the routing message, but the node cannot change the sequence number of others. The higher sequence number is up-to-date route information, in this case if there are multi-paths to the same destination, the highest sequence number will be used. (Perkins and Bhagwat 1994, Royer and Toh 1999)

OLSR (Optimised Link State Routing) is a table driven and proactive protocol (Jacquet, et al. 2001b), means a routing table protocol table needs to be maintained. Its improvement is that effectively compresses topological information, reducing network overhead when transmission. It also uses Multi-Point Relay (MPR) broadcasting instead of flooding to control topology (Jacquet, et al. 2001a), under the same broadcasting effect to reduce network overhead. So that OLSR can be easier to achieve high operational efficiency. But it uses a part of fix network link to form a routing path and ignores the reliability and busy of nodes. This increases the probability of delay and loss on the busy links. Otherwise, some of the links may be in an idle state as a long-term.

2.5.2 Reactive Routing Protocols

Reactive routing is also called on-demand routing, that different from proactive routing which each node has no pre-establish routing table to be consulted. Reactive routing protocol has low overhead than proactive routing protocols, because the route discovery process is flooding the request messages in the whole network periodically. Due to node's mobility in a wireless network and periodical requests may cause disruption to communication networks. In order to maintain the existing route, the communication between essential nodes is a complex challenge to improve routing performance. This approach usually has transmission delay, because each node has to wait for the discovery process when the node purposes to send a message. The most proverbial reactive routing protocols are Dynamic Source Routing (DSR) (Johnson and Maltz 1996) and Ad Hoc On-Demand Distance Vector (AODV) Routing (Perkins and Royer 1999):

DSR (Dynamic Source Routing) serves as a source routing based on-demand routing protocol that uses source routing algorithm rather than hop routing methods (Johnson and Maltz 1996). It applies the routing cache for storing source routing information, modified to route cache content when a new route is received. There are two processes that route discovery and route maintenance. A source node X sends an RREQ (Route Request) packet by flooding, and the RREQ packet includes source node X, destination address and a unique identifier. While destination node Y receives RREQ packet, in this time RREQ packet should have a record all nodes and node's unique identifier from X to Y, then node Y will send an RREP (Route Reply) packet to node X. This RREP packet should have the route information from X to Y. In addition, other nodes which between X and Y also can optimise or update the routing protocol. Route maintenance is kept to detect any routing operations, and notify error routing message to the source node or sender. The advantage of DSR is using of routing cache to reduce the consuming by route discovery; the node only needs to maintain with its associated route; once route discovery process might generate multiple routes to the destination. However, due to the use of flooding method to request the route, so the adjacent node's RREQ may occur conflict or repeats. Also, expire route in the cache will affect the accuracy of the routing choice.

AODV (Ad Hoc On-Demand Distance Vector Routing) is a reactive routing protocol, improves the algorithm of DSDV and combines with DSR (Perkins and Royer 1999). There are route discovery and route maintenance functions from DSR while adopts multi-hops and the sequence number from DSDV so that AODV reduces overheads based on an on-demand and requires each node to manage a routing table. The AODV includes three types of messages: route request (RREQ), route reply (RREP) and route error (RERR). However, its route discovery is similar to DSR with some differences. The node will stores the route information into its own routing table when it receives an RREQ and discard the repeated RREQ based on the sequence number, and determine any response from other nodes. When RREQ is forwarded, the node will mark its upstream node ID into the routing table to be able to build a reverse route from the destination node to the source node. A new route discovery algorithm will be generated when source node moved. If other nodes moved, its neighbour nodes will find failure link, and send RERR message to upstream node thus spread to the source node, then the source node will renew the route discovery.

2.5.3 Comparison of Proactive routing and Reactive routing

The authors in the literature of (Ade and Tijare 2010) compared the performance of the above four routing protocols in MANET. The performance of AODV is the best consider its feature to maintain communication by periodic exchange messages. The proactive routing protocols DSDV can respond to all transmission request at the required moment, but it still requires much more overhead than AODV and DSR. Its overhead decreases as low node mobility, and increases with the high network payload. For less number of nodes and low mobility environment, the DSDV has superior performance. Additionally, due to DSR uses the caching technology and routing request packet, thereby it greatly reducing the routing overhead, especially at high rates of node mobility DSR saves more than DSDV, AODV has similar performance with DSR at this point. The packet delivery of DSDV is lower than AODV and DSR, this is because DSDV maintains only one route for the destination node if this route fails that will not be able to replace it. Nevertheless, the reactive routing protocols have delay problem, because they need to spend route discovery time coupled with transmission time. Overall, the reactive routing protocols have clear performance advantage to apply for ad hoc network (Li and Peytchev 2010). Table 2.7 shows that a review comparison of the features for these 4 routing protocols:

Table 2.7 Comparison with the protocol property

Property	DSDV	OLSR	DSR	AODV
Distributed	Yes	Yes	Yes	Yes
One-way link support	No	Yes	Yes	No
Multicast	No	Yes	No	Yes
Periodic Broadcast	Yes	Yes	No	Yes
Route Maintained	Route Table	Route Table	Route Cache	Route Table
Reactive	No	No	Yes	Yes
QoS support	No	Yes	No	No

The Fig. 2.11 shows that DSDV protocol broadcasts and updates process, DSR protocol discoveries procedure, AODV protocol sends RREQ and receives RREP, and OLSR protocol uses the MPR (Multipoint Relay) nodes broadcast.

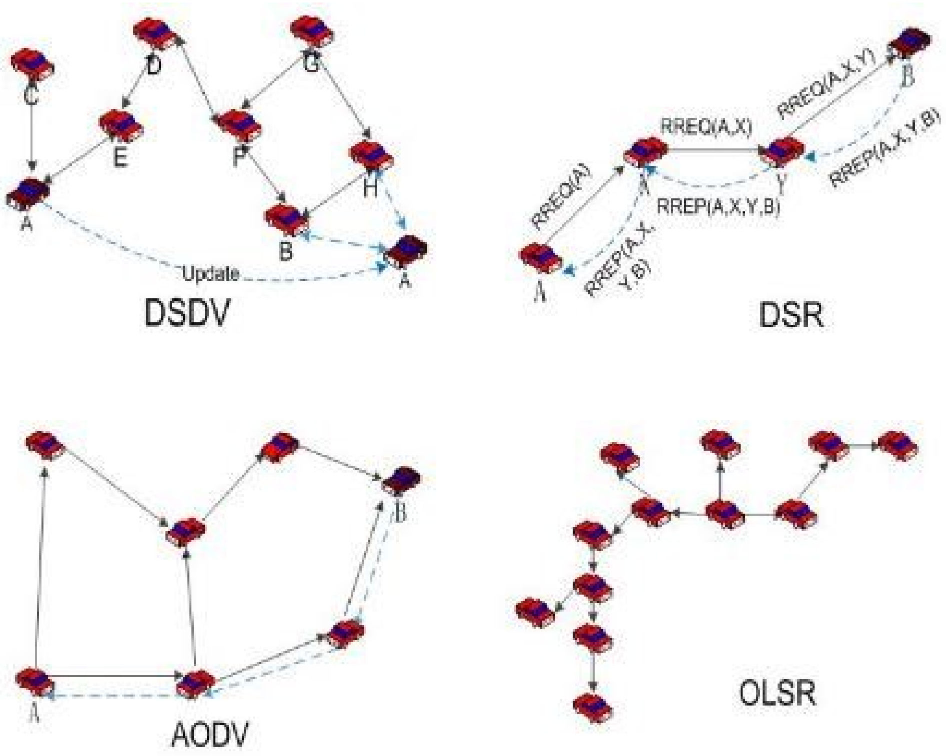


Fig. 2.11 Example of DSDV, DSR, AODV and OLSR

2.6 Traffic Simulation Techniques

Traffic Simulation is the use of computer technology to simulate the changing of the traffic environment (Lieberman and Rathi 1997). The traffic simulation is usually divided into microcosmic traffic simulation and macroscopic traffic simulation, even sub-micron or pure-micro and mesoscope (meso) will be used in some cases (Payne 1979).

Microcosmic traffic simulation has a high description of the traffic element and traffic behaviour details. For instance, micro-simulation describes every single vehicle as the basic unit, when overtaking, car following and changing lanes, the microsimulation can be represented. (Wei, Yang and Cao 2003)

Macroscopic traffic simulation has a lower description of traffic details, compared with micro. Macro-simulation can describe traffic flow, velocity and traffic density, even the relationship between each other. (Payne 1979)

The traffic simulation technology is a very important method to develop, evaluate and optimise ITS project (Pursula 1999). The traffic simulation is based according to the simulation model that describes vehicle behaviour and interaction between vehicles and is mostly as a mathematical model based, such as car-following model and lane-changing model (Gipps 1981, Gipps 1986). Traffic flow simulation technology has many advantages (Pursula 1999, Wei, Yang and Cao 2003):

- Does not need to participate in a real system and without the prohibitive cost.
- Can understand which parameters or variables are important, and how to interact with them.
- Can continue to repeat the random state of traffic flow in some road to traffic conditions.
- Can predict traffic behaviour when the system condition has been changed.
- The traffic demands change by time and space so that the road condition can be forecasted.

There are likewise many advantages of traffic simulation does not mention above, and more details explain in (Pursula 1999). On the other hand, traffic simulation also has its drawbacks and limitations, for instance:

- The simulation model requires a large amount of input data, and some of the data is difficult to obtain in practical problems.
- The simulation model requires validation, calibration and testing.
- Some of the simulation results are limited by the simulation models and simulation software.

2.6.1 Traffic Simulation Model

In traffic simulation, the node is necessary for the transportation network that means vehicles. Moving mode of the vehicle will influence the results of the traffic simulation so that it should be the realistic environment. However, this would involve many factors, such as driving reaction time, driving behaviour and so on (Pursula 1999, Boxill and Yu 2000). Car following model is a model of vehicle moving and is also the theoretical basis of the microscopic traffic simulation modelling. More well-known is GHR model by Gazis, Herman and R.Potts in 1959 (Gazis, Herman and Potts 1959):

$$x''_{n+1}(t+T) = \alpha \frac{x'_{n+1}(t+T)^m}{(x_n(t) - x_{n+1}(t))^l} \times [x'_n(t) - x'_{n+1}(t)] \quad (2.6\text{-}1)$$

$x_n(t)$is the displacement of car n at t time.

$x'_{n+1}(t)$ is the velocity of car n+1 at t time, that is following car.

$x''_{n+1}(t)$ is the acceleration of following car during t time.

T is time length, m and l are constant numbers. m=0~2 and l=1~2.

$\frac{\alpha}{(x_n(t)-x_{n+1}(t))^l}$ is the sensitivity coefficient, and the unit of α is m/s.

Gazis, Herman and Potts proposed that sensitivity coefficient and vehicle's pitch is inversely proportional, also according to the relationship of micro and macro had evolved this car-following model. A feature of this model that behaviour is based on a vehicle driving condition to compute, e.g. the relative position, velocity and acceleration. These are might affect the speed of driving control. Many researchers in the world develop different simulation models and simulators for supporting ITS, more discussion and details are explained in (Boxill and Yu 2000).

2.6.2 Traffic Mobility Simulator- SUMO

Simulation of Urban Mobility (SUMO) is developed by the German Aerospace Centre that is microscopic, multi-modal and continuous road traffic simulation software (Kmjzewicz, et al. 2002). The most important thing which is open source and highly portable, it can be imported a variety of real map database to generate the simulation map. Through a visual graphical interface, the users can observe on the complex transportation network simulation (Behrisch, et al. 2011). Sumo's 'NETCONVERT' can capable many formats and also allows to read many networks from other traffic simulators such as VISSIM or VISUM (Quayle and Urbanik 2008). SUMO also support TIGER network database, Open Street Map or Shapefiles(Hochmair and Zielstra 2011). The '.net.xml' file as a generating file can be imported into the simulator as a simulation scenario (Behrisch, et al. 2011).

For traffic scenarios, 'origin/destination matrices' (O/D matrices) are generally used in SUMO. They describe the movement between traffic assignment zones in vehicle number per time (Behrisch, et al. 2011). There are four applications to deal with the route of the vehicles in road network (Han 2011):

DUROUTER can import or define a route from other simulation packages. And uses the Dijkstra shortest algorithm (Skiena 1990) to calculate the optimal route.

JTRROUTER uses the turn percentages of the intersection to build the static traffic models.

OD2TRIPS can help users to convert the O/D matrices to the trip information.

DFROUTER can compute the route from the given measurement of loop detectors.

The vehicle package should include three information to describe the physical attribute of cars that are a vehicle, vehicle type and vehicle route. The vehicle type and route can be both shared with others.

SUMO utilises the Krauss model as a default model, and this model develops by Stefan Krauss in 1998 (Krauß 1998, Krajzewicz, et al. 2005b). It is a microscopic, space-continuous and time-discrete car-following model. Vehicle behaviour represents the speed of time-discrete. In the mechanism of avoiding a collision, the vehicle's speed is affected by the front of vehicle speed. And in this time, the vehicle speed is called safe velocity, represented by V_{safe}. This is the safe velocity equation (Krajzewicz, et al. 2005b):

$$V_{safe}(t) = V_l(t) + \frac{g(t) - V_l(t)\tau}{\frac{\bar{V}}{b(\bar{V})} + \tau} \tag{2.6-2}$$

$V_l(t)$is speed of the leading vehicle in time t.
$g(t)$is distance from the leading vehicle in time t.
$\bar{V}$is average speed between two vehicles.
b is a deceleration of the vehicle. The unit is m/s^2.
$\tau$is the driver's reaction time (usually default by 1s).

The model should take into account the actual acceleration and road condition. So that the desired or wished maximum speed of normal driving vehicle should be:

$$V_{des}(t) = min\{V_{safe}(t), V(t-1) + a, V_{max}\} \tag{2.6-3}$$

$V_{max}$is the speed limit on the road.
a is the maximum acceleration of the vehicle. The unit is m/s^2.

$V_{des}(t)$will be within the maximum speed and safe velocity.

Finally, the behaviour of the drivers also needs to be considered, because drivers usually cannot reach and maintain the optimal speed in real operation. There are significant coefficients should be used, and the vehicle cannot be backward on travelling, so the minimum speed must be greater than 0. Due to these, the final model equation should be:

$$V(t) = max\{0, rand[V_{des}(t) - \gamma a, V_{des}(t)]\} \quad (2.6\text{-}4)$$

γ is the driver's imperfection in holding the desired speed. (between 0 to 1)

Therefore, the equation (2.6-4) can describe the realistic vehicle speed. The Fig. 2.12 explains how vehicle's movement. The picture intercepted six time periods to show the traffic situation. They are all at a safe distance and safe velocity. And all vehicles stop at an intersection by a lane, because the leader vehicle has limited by traffic lights. The traffic light algorithm or program also is one of the main applications for microscopic traffic flow simulations. The road network can be generated by NETCONVERT or NETGEN tools (SUMO 2015). They are also can generate traffic light program for each intersection. But this program is different with a real traffic light. For actual traffic light program, SUMO accepts the import of external program. And the program can be switched through WAUTs or TraCI (Krajzewicz, et al. 2005a).

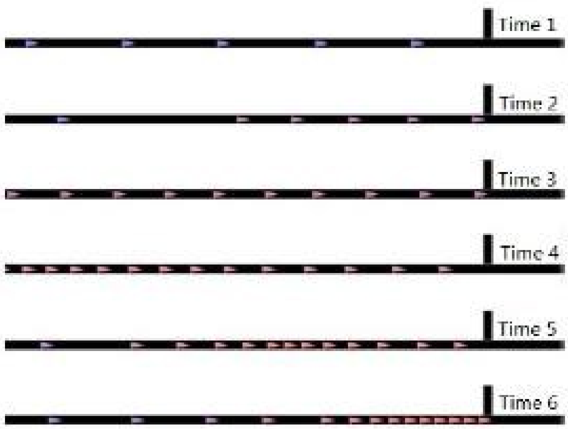


Fig. 2.12 Car's movement in SUMO intercepts 6 time periods

The road network and the vehicle movement are generated by SUMO in this project. The generated network topologies will be imported into the simulator. There are also some attribute parameters of edges and nodes using the simulations of this thesis:

Table 2.8 SUMO's attribute list

Attribute Name	Mandatory	Type	Description
Edge Id	Y	Id(string)	Edge's name
Edge type	N	Type Id	Edge type
Nolanes	N	Int	Number of lanes
Edge Speed	N	Float	The maximum speed of edges.
Priority	N	Int	Priority edge
Length	N	Float	Edge's length (m)
Shape	N	Position list, use x, y. (unit m)	Shape="0,0 0,100" means from point (0,0) to point (0,100). Movement 100m
Spread_type	N	Type Id (right, centre)	How to extend the road lane "centre" means extension in two sides, others mean extent to right
tlLogic Id	Y	Id(string)	Traffic Light Id
tlLogic type	Y	Type Id	Traffic light type(static, actuated, agentbased)
Program ID	Y	Id(string)	tlLogic Id, 'offset' has been retained
Offset	Y	int	Initialization time offset
Phase duration	Y	int	Time
Phase state	Y	Signal status (G, Y, R)	Signal status (Green, Yellow, Red)
Vehicle Id	Y	Id(string)	Vehicle's name
Vehicle type	Y	Type Id	Vehicle type
Route	N	Id	The route of vehicle used
Depart	N	Float	Time of vehicle enters the network
Vehicle Speed	N	Float	Vehicle's max speed

2.6.3 Network Simulator-NS3

Two parts are required when to build VANET simulation system, that is traffic mobility simulation and network simulation (Miao, Luo and Hong 2011). Traffic mobility simulation already introduced in previous subsection and network simulation will be discussed in this subsection.

NS-3 (Network Simulator Version 3) is a discrete-event network simulator, which it has a virtual clock and all simulations are driven by discrete-event (ns-3 2015).

Compared with other network simulators, NS-3 is preferable to others in the completeness, open source, usability and scalability, and is free software. NS-3 has been redesigned on the overall architecture by NS-2 (Fall and Varadhan 2007), abandoned the NS-2 used in Otcl scripts and C++ model structure, and is a completely new software development. In the NS-3 environment, the user can only use C++ and Python to complete the model and topology scene.

NS-3's architecture mainly consists of two parts that Network Components and simulator core (Yu and Zhao 2006), shows in the Fig. 2.13. Scheduler and network simulation support system comprise simulator core. They are the most important kernel in NS-3. Network components are an abstract representation of all part of the network in the simulator.

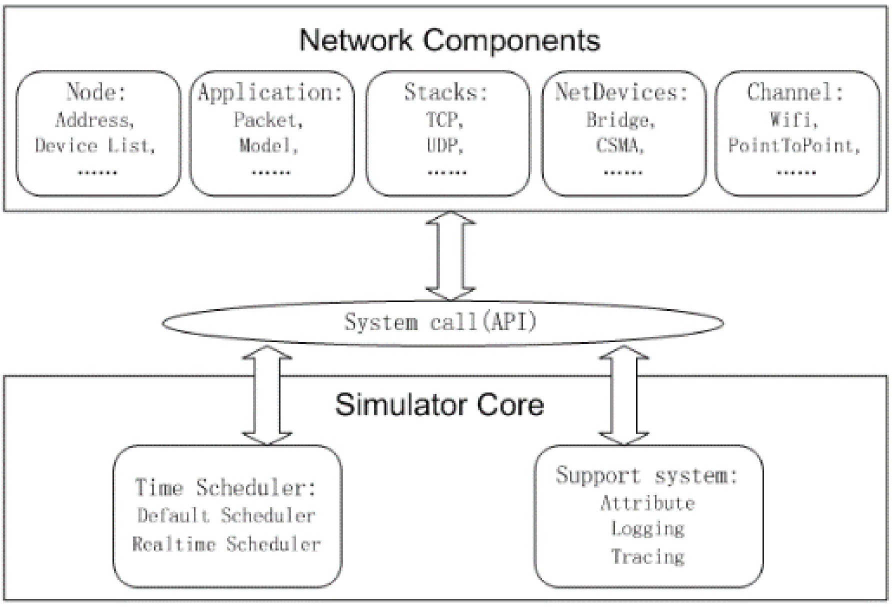


Fig. 2.13 Architecture of NS-3

NS-2 and NS-3 are both discrete-event network simulator (Miao, Luo and Hong 2011). The scheduler is similar with some differences, the real-time scheduler is added in NS-3. Network simulation support system includes Attribute system, Logging system and Tracing system. Attribute system achieves to organise, access and modify the simulation parameters. Access and a set of object attribute values are transferred to the TypeId class, thus, a set of attributes for namespace be formed the base on the string. Logging system is used immediate feedback to user implementation and operation of the system, through set by NS_LOG environment variable, can also be in the function calls. The tracing system is utilised to output the simulation results. Trace source notifies that event occurs, and also can gain access to the data. Trace source must connect with target trace source. Thereby the event and data can be processed. The output format is ASCII and PCAP. (Yu and Zhao 2006)

NS3 widely used in academia with the virtue of its excellent performance and advantages of open source. There is a list of its advantages, and that's why we will use it in this project:

- Simple Configurability: NS3 support C++ language to replace the Otcl in NS2. C++ language is more popularization and understandable. It is responsible for network model component, configurations controlling and parameters scheduling.
- Open source: NS3 can in-depth investigate the structures and behaviour of various model components and protocols. It is also simple to create a new model and modify the existing model. A lot of network models and functions are provided for simulation.
- NetAnim: It is a visual interface of NS3, the user can observe the results with the generated output files. This is a big help for the user to conveniently view the status of the simulations or the movement of nodes.
- Mobility support: NS3 support two methods of mobility pattern generation, one is that creates by C++ in the script with random movement or planning node movement. The other is that supports external trace file to embed into the simulator instead of method one, such as ".tcl" trace file from SUMO.

2.7 Knowledge Gap of Related Work

Our aim is new knowledge in the area of in-car traffic prediction by C2C communications. Our concerns are the workable and performance of in-car prediction framework that includes the data collection, data transmission and data application through C2C communications.

Designing traffic condition estimation for a future time by C2C is one of the open research challenges being investigated in VANET, especially travel speed prediction. The travel speed often impacts by many factors. However, some of the researchers consider that it is significant to detect the density of vehicles. It is widely known that a high-density traffic indicates low travel speed of vehicles, on the other hand, a low-density traffic indicates the high speed of vehicles. But in fact, the traffic condition is affected by multiple factors, and some estimation methods use a large amount of historical data to improve the prediction accuracy and lead to the algorithm is too complex, while ignoring the timeliness of traffic information.

2.7.1 Traffic Speed Estimation-MICE

A traffic speed estimation for VANET is proposed by He et al (He, Cao and Li 2012, He, Cao and Liu 2015) that is a distributed peer-to-peer algorithm to traffic estimation, namely MICE. In MICE, car to car communication is utilised to collect position and speed related data from nearby vehicles. The traffic speed estimation relies on vehicle density model (Underwood) and the traffic information collection uses the algorithm of RTS/CTS forward strategy and Backoff-and-Fork candidate path selection with infrastructure-free so that analyse the traffic information by estimation solution. The data collected is sparse data in the form of floating car data snapshots. Whereas compared to the existing infrastructure-free solutions, this solution can be working for both dense and sparse traffic scenes. Also it does not need training data set and only utilises macroscopic traffic flow model to estimate traffic speed. The solution can outperform some issues in term of network transmission overhead, routing planning effectiveness and traffic overhead through using both testbed based implementation and trace-based simulation, such as it does not have in-network cache that means there is no “warm-up” stage at the beginning.

However, there are also shortcomings with following. The mobility pattern uses the trace-based method in their studies, and the recorded data is always from special vehicles such as taxi or bus. VANET is yet to be deployed widely to have the traces including a large number of vehicles for long-term observation. The model has its limitations for other city scenario or mobility pattern, and the proposed estimation solution is in a specific scenario. There are poor adaptability and extensibility. Their studies use a developed routing based on the geographic routing protocol which is forwarded greedily like GPSR, but this routing is local optimum, not from the overall, limited by node deployment and road topology. Also, GPSR does not be appropriate for city scenarios in VANET (Li and Wang 2007) due to the communications between nodes might influence by buildings or trees to restrict the greedy forwarding. In VANET, topology changes quickly, if each node must maintain a global node information, network scalability will be limited. The frequency of topology change and node increase make up the increased the complexity of routing algorithm, while the threshold distance is not easy to obtain (Karp and Kung 2000). Otherwise, the protocol can be adaptive to different traffic density, but the impact of packet loss or packet delivery ratio is not considered in the model. With a large number of vehicles in the real traffic, this will be a big issue to impact density count when density is high, thus influence to the traffic speed estimation (Huang and Zhang 2013).

2.7.2 Congestion Detection Algorithm

In the study of Milojevic and Rakocevic (Milojevic and Rakocevic 2013), they proposed distributed algorithm to detect and quantify the level of traffic congestion for vehicle speed estimation with infrastructure-free. Their algorithm consists of five procedures which are implemented locally by each vehicle: a) the speed monitoring defines a speed threshold and records the current speed comparing to the speed threshold. b) Congestion detection procedure determines congestion parameter by comparing the size of current speed and speed threshold. c) Localization procedure generates the congestion parameter of the current location into the broadcasting packet. d) The aggregation is defined whether the message is broadcasted. e) The broadcasting includes the message of congestion parameter and road information to sharing with other vehicles. This algorithm defines a congestion parameter and reduces network load comparing to periodic broadcasting; hence it can perform well with different levels of congestion. From the result, that congestion detection can reflect by each vehicle corresponds to actual speed.

However the speed threshold and the congestion parameter should be defined in advance, and each street section has different characteristics of traffic flow. The settings of speed threshold and congestion parameter need to base on the historical data and observations. Also the congestion detection impacts by topology changing, transmission delay and packet loss, the level of traffic congestion did not change significantly when the speed has fluctuations; on the contrary, when the level of traffic congestion increased, the speed response also did not change. The study discussed mainly focus on detecting traffic congestion but very few attempted to estimate the occurrence of traffic congestions or traffic conditions.

2.7.3 Efficient Local Density Estimation

In the research of (Noureddine and Samira 2014), they proposed a segmented-based local density estimation strategy ELDES which improves the redundancy of extended beacons and accuracy of estimation by using linear interpolation. The advantages of this strategy are highly accurate with low overhead for transmission. The strategy defines the identity of each segment on the road. Each vehicle could send an extended beacon that generated by normal beacons or extended beacons. If it did not receive the extended beacons from this road, an estimation value would be generated based on normal beacons. Then the strategy looks a valid data in the nearest vehicle for the corresponding segment, to improve the accuracy of the density. The vehicles only share their information when they travel to the centre of the segment. In this case, the overhead is decreased by avoiding the beacons from the same segment.

This strategy is a very particular scenario, and the density does not fully reflect the vehicle speed with a sparse network, such as a highway. Some cases can happen easily in which the vehicles are concentrated at one end of the road without vehicles in the segment of others; hence the density estimation has an error for this road. Also the communication effect will be very weak in the high-speed and large transmission range. Although a segmented-based is a good opinion in this case, the adaptability and extensibility of this model are poor.

2.7.4 Vehicle-Based Mobile Sensor Network for Traffic Monitoring

A traffic monitoring of vehicle-based mobile sensor network is proposed by Xu et al. (Li, et al. 2009). In this study, the link-based algorithm (LBA) and the vehicle-based algorithm (VBA) has been presented to estimate the real-time mean speed for every road section in Shanghai, which is based on the data from vehicle-based sensor networks. It can provide sparse and incomplete real-time traffic information. The vehicle-based sensor is suited for a long time sampling intervals so that reduce the communication cost and network congestion. The LBA assumed that if a particular link is given, then pairs of sensor data either starting or ending around this link can best reflect the traffic status of this particular link. In contrary the VBA utilises every available data pairs and disseminates them back to all the links travelled to calculate an average travel speed. Thus, a sensor following a vehicle might travel across one or more links which in turn can be associated with one or more roads. The results show that the traffic status can accurately be estimated based on the information from vehicle-based sensor networks.

The advantage of this approach is the accuracy of the measurements and the wide/intensive coverage. Conversely, how to combine sensor networks with traffic control system and deploy sensor at intersections to realize accurate traffic information collection that is important considerations. The sensor network is a special ad-hoc network, in the case of more dense nodes and limited power, how to conserve power to maintaining storage and computing that is also a challenge, for example LBA needs to use and store historical results. Otherwise, the taxis are nodes for data collection and traffic speed estimation in the sensor network in this investigation. The results are susceptible to taxi's distribution and behaviour, such as each road have different densities of taxis or some taxis are within a temporary parking and starting so that these behaviours can affect the accuracy of the data collection, especially link-based algorithm. The results of LBA delivers a less accurate than VBA. Apart from the vehicle aspects, LBA is also affected by road conditions based on the real traffic features, this because of the vehicle across two sensor-deployed intersections during a long sampling interval and there is a delay between two intersections. Also, their algorithm can make more accurate estimation in a congestion than in a travel condition, because most of the vehicles are waiting and stopping so that there are lower speed and a large proportion of the total time of travelling the road. Therefore the accuracy of the results is weak in speed changing. The details of the networking communications are also an important consideration. However, compared with the vehicle-based sensor network for the special vehicle, the ad-hoc network for traffic participants has the advantage of reflecting the complex and diversified traffic environment. For the transmission speed and time delay of the vehicle-based sensor are slower and more time than the ad-hoc network, especially, when the sensor employs GSM to collect the data from round sensor-deployed intersections.

2.7.5 Trajectory-Based data forwarding for VANET

In the study of (Jeong, et al. 2011), Jeong et al. proposed a scheme of trajectory-based data forwarding for sparse traffic network via Vehicle to Infrastructure (V2I). They intended to minimize the packet delay of end-to-end delivery to the Internet access point by one-way delivery. An expected delivery delay (EDD) was presented for their TBD link delay model. Through theoretical analysis and experiments, the carry delay is the primary factor for the packet delay. A privacy-preserving trajectory sharing scheme only provides the expected delay value, to avoid sharing actual trajectory. The solution has lower time delay than existing link delay model for end-to-end delivery.

As the paper mentioned, TBD has less performance when there is a sparse network with low density, because the possibility might increase that each vehicle is not in the communication range, and also the probability is relatively low that the carrying vehicle travels to the destination. When the number of the vehicle is increasing, the performance of TBD is encouraging, that because the possibilities of both might be increased. However, the performance of communication would be decreased to cause packet loss. In general, the travel speed cannot be constant in the real world, and the speed also can be zero when the vehicle is waiting at an intersection or in a congestion queue. If the speed is infinitely close to zero or is zero, the link delay model can be inefficient and invalid. The higher speed leads to less carry delay so that the total delivery delay is short. However, the higher speed also can bring low communication performance. In the real world, the topology is changing very quickly, and the trajectory could be altered at any time, thereby the model needs to adapt to changes in the environment. The model could attempt dynamic transmission rather than dynamic node to static node. Because the deployment of the infrastructure also requires a significant cost.

2.7.6 Summary

To the best of our knowledge, none of the frameworks or approaches in the studies have addressed in-car traffic conditions prediction for the future time based on the C2C communication in the real time. They have not take the traffic prediction from roadside infrastructure into the car. All these approaches suffer from the issues related to continuous data collection and data processing. Also, some of the more complex algorithms in the centre control lead to poor dynamic and timeliness of the computation. Moreover, the adaptability of the model is also a potential problem in the frequent changes of the topology. In response to these questions, we present a novel prediction framework based on C2C communications and verify its feasibility in subsequent chapters.

2.8 Summary

The literature review of this Chapter is regarded as the theoretical basis necessary and the related works of researchers for this research. A review of VANET covering its characteristics, architectures, applications and models, also comparing it with MANET are presented in Section 2.2. The theories and concepts are with the general scope of this project. Section 2.3 and 2.4 presented the prediction models for the traffic, and introduced their classifications based on analysing the features, discuss prediction models for VANET, their algorithms and related works to illustrate the model's advantages and weaknesses. An introduction and comparison of Ad-hoc network routing protocols are in section 2.5 which include proactive routing protocols and reactive routing protocols, covering the routing classifications, features and issues for VANET. Section 2.6 refers to simulation techniques with its advantages and disadvantages meanwhile introduces the well-known simulators that SUMO and NS-3 will be used in this project. Section 2.7 introduces the existing knowledge of traffic prediction for VANET. A summary of the studies and proposes relative deficiency in this area.

3 Chapter 3

Design of a Pervasive Prediction Model (PPM) based on ad-hoc data

3.1 Introduction of Related Prediction Model

Traffic prediction covers a wide range of research topics in academia, such as data collection, data analysis and model generation, with an important application technology of Intelligent Transportation System (ITS). Traffic prediction contributes significantly to the improvements in traffic management, guidance, and control. The results of traffic prediction usually include traffic flow, travel time, travel speed or optimal route to meet the corresponding goals. Currently, in the field of traffic prediction, the leading approaches have a mathematical algorithm, modern science technology or combination prediction method. This chapter will mainly elaborate prediction model proposed by the authors, and compared with the existing classical prediction model.

3.1.1 Mathematical Algorithm

The traffic prediction approach of mathematical algorithm is based primarily on the traditional mathematical and physical methods; there is Time Series Prediction Model, Kalman Filter Model, Parameter Regression Model and Exponential Smoothing Model (Zhu, Wang and Xiang 2008). The most of these mathematical prediction model can be applied not only in the transportation field but can also be used in the field of economic, weather and water conservancies, such as Regression Model and Time Series Prediction. They need to rely on reliable historical data in modelling and usually have less the number of prediction periods. These models have strong theoretical support, meaning clear and easy figures. For example, the regression model as a most common method, it mainly used the relationship between dependent variables (prediction object) and independent variables (influencing factor) to establish the regression model.

Further, ARIMA model and Moving Average Model are Time Series Prediction Model (more details in section 2.4), they also have a strong applicability in short-term traffic prediction. They usually are speculated that the trend of prediction object based on the historical data.

3.1.2 Modern Science Technology

The prediction models are used modern science technologies as a research tool, be namely modern science technology method or intelligent methods, such as the no-Parameter Regression Model, Wavelet Theory, Multi-fractal Method, Grey Theory, Kalman Filter, SVM Theory, Chaos Theory and Multi- Agent prediction method Neural Network.

Some methods can solve the prediction in the case of incomplete data, such as Grey Theory, the essence of Grey Theory is reflected the variation of the prediction object with time by the exponential function. Its core model is GM (1, 1), meanwhile through a multi-stage, higher-order to establish grey dynamic model GM (M, N). (Guan-jun 2000)

Kalman Filter is a filtering method for random signals; it estimates the linear minimum variance of observed data. Kalman Filter is not required to preserve the historical data, after obtaining a new observational data, by the filtering recursion formulas to estimate new value. (Okutani and Stephanedes 1984)

SVM is a statistical theory for the classification and regression problems and is proposed by Vapnik from AT & T Bell Lab (Vapnik and Vapnik 1998). SVM based on Structural Risk Minimization (SRM) principle, and is superior to the traditional Experience Risk Minimization (ERM). Currently, SVM has also been extended to time series analysis.

Chaos theory research uses numerical methods as so far, according to nonlinear time series by chaotic systems extract to predict the future situation of the object. Chaos theory studies the sensitive dependence on the initial value and chaotic systems; topology transmission and Chaos; denseness, randomness and ergodicity of periodic points; positive Lyapunov exponent; fractal dimension and chaotic attractor

(Grassberger and Procaccia 1983, Grassberger and Procaccia 2004).

The number and variety of prediction methods and theories for modern science technologies are far beyond what we cannot cover whole in this thesis, so more on this topic can be found in relevant articles or books if you are interested such as (Xu and Fu 2009).

3.1.3 Combination Prediction Method

Combination Prediction Method means that combines the different prediction models together, utilises information from respective prediction methods to improve the prediction accuracy and increase prediction reliability (Bates and Granger 1969). Combination Prediction Method obtains different information from different sample data, different models and different angles, can more fully grasp the changing characteristics of the system.

By combining different model, the prediction accuracy is usually higher than a single model. This eliminates the process to find and compare the optimum model, but also can remedy defects for single prediction models and reduce the risk. Moreover, combination model has a stronger robustness for changing the structure of the data.

This new prediction theory gradually is developed, since Bates and Granger (1969) were formally proposed combination prediction. Currently, it basically uses linear combination prediction; nonlinear combination prediction is mainly based on Neural Network. However, with the development of ITS system in the world, the using of combination prediction have a broader outlook than a single model to predict the traffic information.

3.2 Pervasive Prediction Model

With the development of traffic condition prediction, the predictable traffic information has become more widespread. More traffic information forecasting can help users and managers to understand the traffic situation, at the same make more rational transportation decisions and leading to establishing efficient and safe travels in urban.

The predictable traffic information typically usually includes traffic volume, travel time, travel speed, traffic density and traffic event. Some of the prediction models can predict different information. For example, Kalman Filter Model can both applied to predict traffic volume and travel time, so that most of the prediction models have interoperability. For the predictable traffic information, the different results provide different reference values. The manager will be more concerned about traffic volume and traffic density, which can originate from the city as a whole viewpoint to consider how to guide the traffic. For the prediction of traffic accidents and events, it also can help managers prepare and take the necessary measures in advance. Whereas travel time and travel speed get more attention for drivers because they would like to learn how to choose the optimal route. This thesis presents the idea of the travel speed prediction, by using car to car communication to collect the real-time traffic data. This chapter will propose a new travel speed prediction model, validate the model and compare the model with existing prediction model. However, how the prediction system is running, transmitting and receiving the messages will be addressed and explained in the next chapter. In particular, comparisons of routing protocols and simulation evaluations in the wireless environments will be presented in Chapter 4 and Chapter 5. Hence, the experimental data and outcomes do not consider communication and network environment in this chapter.

3.2.1 Overview of Pervasive Prediction Model

To provide reliable and efficient predicted results, we propose a novel PPM-C2C prediction framework based on the C2C communication. An innovative pervasive prediction model (PPM) for traffic condition as an agent of the framework is proposed in this chapter. One part of PPM is the data table for traffic data integration and storage, the other part of PPM introduces the Polynomial Regression-Adjusting range and Weight Moving Average as a combination prediction method for the traffic prediction; we name this scheme PRAWMA in the following sections. The PPM allows mobile application platforms or vehicle devices to collect and access data from other participants.

In fact, the PPM is from the traffic data of each node on the road to calculate the traffic condition of this road at the future time, which the part of polynomial regression equations can provide several possible outcomes. Also, the adjusting range part needs to determine the validity of these results because there might be a big difference between some of the results in comparison to real life situation or theoretical assumption. The adjusting range can prevent cases where predicted results are too large or negative while ensuring that the results are in reasonable range. We determine it according to experience and experimental comparison for this case. Weight moving average part can make the data classification because the predicted results are more influential if the data is more recent, at the same time it also provides a trend for prediction. That is, if a vehicle in the road network receives the real-time travel speed information from the cars on a certain road section, then this vehicle utilises PPM to compute the average travel speed of this road section. Also, the PPM is running independently in each vehicle, so that all vehicles only share the travel information and do not share their predicted results.

3.2.2 Theoretical Framework of PPM

This project is dedicated to the development of a pervasive traffic prediction framework for the car to car communication. Each vehicle can independently complete data collection, data processing and computation with the ad-hoc data, while not relying on centre processing and manual analysis. In computer science, we propose a PPM-C2C framework based on real-time C2C communication and design a PPM prediction model as an agent for this framework. Therefore, the framework can introduce the prediction model for different traffic condition in the future.

The proposed PPM is based on the city traffic time-series data in the PPM-C2C framework. However, what is the difference between the time-series regression analysis and the regression analysis? The former refers "time" as the factor that influences the prediction object, that is, the independent variable of the prediction model is "time"; while the latter builds a prediction model based on factors that have a major impact on the prediction object. It is well known that the traffic condition is influenced by many factors such as road width, adhesion coefficient of the vehicle, driver behaviours, driving distance, weather, and so on, which is a complex nonlinear function with many variables. Time-series regression can be used because of the limitation of the analysis method and data acquisition, or the inability to obtain the corresponding data, so as not to meet the applicable criteria of regression prediction. In general, the advantage of regression prediction method is that the modelling is simple and easy to implement. For a large distributed urban traffic simulation, a simple model can ensure the operation speed.

A large number of traffic surveys show that, regardless of the macro level, meso level or micro, the development of traffic phenomena have a certain regularity relative to time (Xu and Fu 2009). When the prediction model object changes with time to show the trend of a curve, the trend prediction model of the prediction object about time can be established. Therefore, the proposed prediction model need to use the historical data for the modelling. However, as the prediction time goes on, the model must be gradually distorted, which leads to a decrease in the accuracy of the prediction. To overcome this shortcoming, we adopt an adaptive prediction method, that is, add a new data then remove the oldest data to ensure that the data size unchanged. Because the historical data comes from the surrounding other vehicles which are transmitted by wireless and stored in the respective vehicles. As long as the vehicle travels in the road network, it can continue to receive new data. Since the new data information is fully utilised, the prediction accuracy can be improved. Otherwise, the regression model is characterised by a simple modelling and a wide range of applications that can be quickly modelled through the ad-hoc data. The collected data of traffic conditions are treated as a dependent variable such as travel speed, travel time, traffic volume or density and the time trend is the independent variable which can be written in time series.

Linear regression and nonlinear regression

Linear regression is a method for modelling the relationship between two or more variables which usually includes one or more independent variable denoted X and one dependent variable denoted Y. The simple linear regression is $y=a*x + e$ with only one independent variable, e is error which follows a normal distribution with zero mean as assumption and the relationship between x, y can be approximated by a straight line. For more than one independent variable, the process is called multiple linear regression.

If the dependent variable is more than one function of the independent variable and the law of regression in the graphics on the various forms of curves, known as nonlinear regression. The basic process method of nonlinear regression is to transform nonlinear regression into linear regression by variable transformation. Normally, a nonlinear expression between the output variable and the input variable has been obtained based on theory or experience, and the coefficient of the expression is determined from observation results according to the principle of least squares, the resulting model is the nonlinear regression model. For example, a nonlinear regression problem is $y = ae^{bx}K,$ with parameter a and b with multiplicative error term K. This can transform into a linear problem by logarithm of both sides, *ln(y) = ln (a) + bx+ ln(K)*.

The fundamental difference between linear regression and nonlinear regression is that what form of model function can be accepted. In particular, linear regression requires linear parameters, while nonlinear regression is not required. When the relationship of variables cannot adequately model using linear parameters, the nonlinear regression should be used instead of linear regression. Since the parameters of a linear regression function must be linear, there is only one basic form of the equation. The parameter is linear when each term is additive and contains only one parameter that is multiplied by the term in the model. Therefore, the basic form is:

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots b_nx_n$$

However, nonlinear equations can take many different forms, in fact, there are infinite possibilities. The specific process of corresponding linearization equation not to reiterate them here.

3.2.3 Principle of PRAWMA Scheme

PRAWMA scheme is one part of PPM and built on or according to Polynomial Regression-Adjusting range and Weighted Moving Average. This method uses historical data of road section by a time series. The historical data is from the communication and transmission between vehicles, and the data is instant speed by every second and will be stored in the data table on their mobile devices through the PRAWMA to estimate the traffic condition. The time series also can be called time period. Therefore, there are three main components in the prediction method:

Polynomial Regression

The most of the traffic problems can be divided into linear and nonlinear as shown in Fig. 3.1. However in general, a simple linear model does not adequately reflect the varied and complex transportation information. The first step is to determine whether there is a linear correlation when analysing the original data. If there is a nonlinear correlation which means the regression function cannot be described by a linear, so that we need to consider to use nonlinear regression function, but we can also build a linear regression model through equivalent transformation. For example, the nonlinear regression model is $c = ae^{bx}\beta$, a and b are parameters, β is random error, then we can use logarithm transformation to change it into $\ln c = \ln a + bx + \ln \beta$, that is a simple linear regression model. Some detailed information about the transformation between linear and nonlinear may refer to (Feigenbaum 1978, Yan 1996).

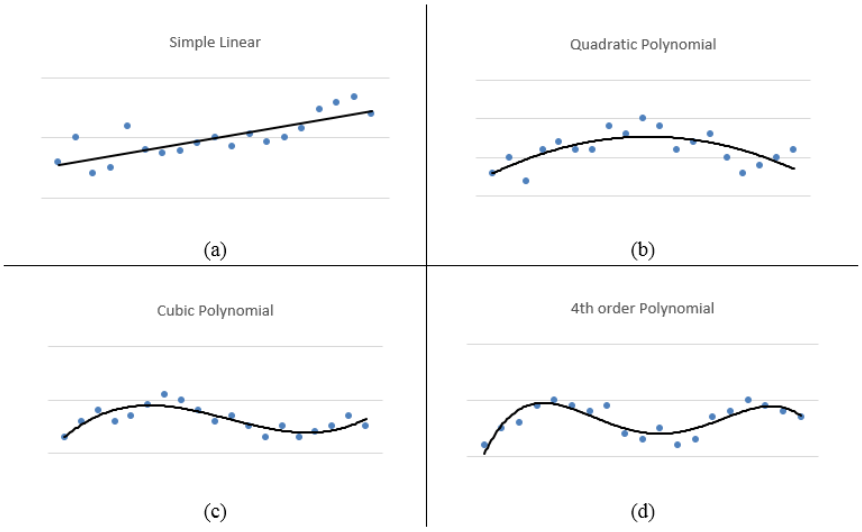


Fig. 3.1 Linear and Polynomial Regression

In the single factor polynomial regression, if the linear relationship cannot be determined via fitting value or scatter plot, meaning that the best-fit line is not a straight line but a curve fitting. Then we can use a polynomial to approximate or fit regression function. In this thesis, the travel speed will be the single factor. If one 'bend' as regression function was observed from scatter plot, so quadratic polynomial will be considered as described in Fig. 3.1(b); if two 'bends' are in there, the cubic polynomial will be considered as shown in Fig. 3.1(c), and so on. The biggest advantages of polynomial regression are met by increasing the high-order term of the actual point to approximate and fit, at the same time the parameters are respected to be linear. If the relationship between y and x is pth degree polynomial and the random error of ε_i ($i=1,2,\ldots,n$) is normally distributed $N(0,\sigma^2)$, so the polynomial model is:

$$y_i = b_0 + b_1 x_i + b_2 x_i^2 + \cdots + b_p x_i^p + \varepsilon_i \qquad i=1,2,\ldots..,n \tag{3.2-1}$$

The parameters of b and ε need to use least squares method (Stigler 1981) to estimate and determine, and then there can be obtained distribution and interval of parameters, various testing statistics and prediction equations. The method of least squares is a mathematical optimization technique, it seeks the best functions of data by minimising the square sum of error (SSE). Least squares method can easily be obtained unknown data, and makes SSE to a minimum between the obtained data and the actual data, also be used for curve fitting. We give a simple example to illustrate how to use the least square to fitting curve. For instance, there are four data samples $(x, y) = (1, 6)$, $(2, 5)$, $(3, 7)$, $(4, 10)$ and the points are indicated in the Fig. 3.2, and we hope to find a line as $y=a + bx$ to match those four points, that find out in some of 'best case' can be broadly consistent with a and b at following an overdetermined system of linear equations (Williams 1990, Krishnavedala 2011):

$$\begin{cases} 6 = a + b \\ 5 = a + 2b \\ 7 = a + 3b \\ 10 = a + 4b \end{cases}$$

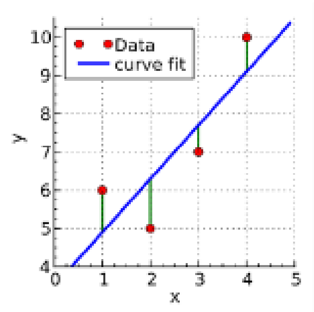


Fig. 3.2 Least squares fitting

The least squares try to make the smallest variance at both sides as much as possible, which is the minimum value of the function:

$$S(a,b) = [6 - (a+b)]^2 + [5 - (a+2b)]^2 + [7 - (a+3b)]^2 + [10 - (a+4b)]^2$$

The smallest variance is the partial derivative of a and b via $S(a,b)$, then make them to zero:

$$\frac{\partial S}{\partial a} = 0 = 8a + 20b - 56$$

$$\frac{\partial S}{\partial b} = 0 = 20a + 60b - 154$$

There are two unknown value equations, so that $a=3.5$ and $b=1.4$, that means the best line is $y=3.5+1.4x$ for this question and shown in the Fig. 3.2. The least squares method can also be applied in higher-order regression and matrix issues with the same principle, more details about it can see in the literature of (Durbin and Watson 1951, Whittle 1963, Marquardt 1963).

The changing tendency of travel speed is usually presented as a waveform with traffic volume, traffic light or traffic event in the urban road section. Therefore, the higher-order polynomial can be qualified for the description of speed dynamic changes. This thesis intends to build six polynomial regression equations as the first part of the prediction model, which means the highest polynomial has five ‘bends’ on the scatter plot, and all polynomial regression equations can provide six kinds of predicted results. However, we need to further determine the reasonableness of these six predicted results through the second part of the prediction model which is adjusting range, and all reasonable results will be averaged. This result is one of the predicted results. The second predicted result comes from Weight Moving Average part.

Furthermore, the number of samples is noteworthy for regression modelling. It is generally known that the quality of the samples decides to the quality of regression results in statistics. The large sample size can better describe the variable, but the calculation might more complicate. If the sample size is small that it does not reflect the objective changes. The sample in this thesis is the average speed by time series. So we can control time period length, and how many orders of average speed can be used in modelling. For example, if we have a travel speed list for 5 minutes by the second. In this case, we can set the time period which is 10 seconds, then we have 30 samples of average travel speed by a time series, and the predicted result is average speed between 5mins to 5mins10s. On the second way, we change the time period to 30 seconds, which means 10 samples will be used to predict the average speed between 5mins to 5mins30s. Or maybe we can continue to refine the time period, and only take the first 10 samples, which means the time series from beginning to 1min40s, to predict the average speed between 1min40s to 1min50s, to keep 10 samples according to a moving method that a new sample replaces the oldest one. The literature of (Nan 2000, WenTong and Wei 2004) mentioned some sample selection methods in the nonlinear regression for different cases, but we plan to use 10 samples in a preliminary experiment according to our particular case and easy calculation. If the samples are changed in every time for calculations, that makes the calculation too complicated. Therefore, a fixed number of samples is more conducive to each calculation. In addition, the samples can update in the real time that because our samples come from C2C communications. In the meantime, we use six polynomial regression equations as much as possible to avoid the risk of the predicted results with the sample size. In this case, despite the 10 samples in preliminary, we will discuss the impact of the time period and sample size to the predicted results in following sections and chapters.

Adjusting Range

Although traditional regression is simple and practical, the issues can occur in the following situations: the data set is small; there is difficulty verifying that the error is normally distributed assumption; the relationship between input and output have obscure; inaccuracy or distortion introduced by linearization. However, the fuzzy regression can solve the above issues. Tanaka et.al (Asai 1982) have put forward the concept and methods of fuzzy linear regression. In their view, the deviation between real value and estimated value is caused by the ambiguity of the model parameters. In this case, the prediction based on historical data is not a precise numerical value, but a fuzzy number with a certain fuzzy amplitude. In layman's terms, the prediction results are only "about how many", this is in line with many practical situations. The traditional regression considers observation error as the deviation between the real data and estimated value. In the fuzzy regression, this error is assumed to be the ambiguity of the system structure itself, and the deviation is regarded as the ambiguity of the system parameters. The basic idea of the Tanaka method, can be referred to as possibilistic regression analysis, which is to determine the fuzziness of the model by minimising the total spread of the fuzzy coefficients. In possibilistic regression based on the symmetrical triangular fuzzy number, only the data points involved in the upper and lower bounds can determine the structure of the model, the rest of the data points do not affect the structure.

The predictions of traffic conditions such as travel speed, travel time or density are usually deterministic using regression models. However, in the measurement of traffic related data, there are inevitably many errors caused by the test personnel or equipment; secondly, the choice of vehicle speed or route selection involve a variety of complex factors which has certain uncertainties and ambiguities. The membership function (MF) is used to fuzzify the parameters in the model and the traffic condition is taken as the interval number that is a range of prediction values for the relevant quantities. It can reduce the dependency of the model on the accuracy of the sample data and avoid the wrong judgment, also is possible to make the prediction result more realistic. In addition, although regression prediction has the characteristics of simple principle and convenient calculation, it is required assumptions that the random error of sample must satisfy the hypothesis of zero mean and normal distribution, otherwise the prediction results may be deviated. The fuzzy regression can meet such requirements.

In this project, we present an approach that it is similar but different from the theory of fuzzy regression and the algorithm of prediction interval in the statistics, namely adjusting range. In regression analysis, the prediction interval is an interval estimate with a certain probability in which future observations will fall. The prediction interval can estimate the interval of one individual value of the dependent variable. Otherwise the prediction interval is different from the confidence interval in statistics, the confidence interval of the probabilistic sample is an interval estimate of a population parameter for the sample. The confidence interval shows the extent to which the real value has a certain probability of falling around the measurement result.

Because our prediction model is based on the wireless ad-hoc network and embedded in each vehicle, in this case, the traffic data is updated instantly, so that there is ensured that the prediction model uses the latest data. Also, according to fluid dynamics, the transportation is a continuous motion. Therefore, the traffic condition especially travel speed is taking into account continuum or regular fluctuation assumption rather than discrete. Take travel speed, which is treated as a single point, and from one point to another is continuous change. This change needs to be a reasonable interval in a certain time period whether the travel speed increases or decreases. In addition, since the model results in different samples per vehicle so that the regression model will certainly be different. On the other hand, the complex statistical results and tests are more difficult to achieve for the model in each vehicle. Therefore, we have to set adjusting range to determine a more reasonable fitting and forecasting results from a couple of polynomial regression.

Based on the above considerations, the adjusting range refers to real-time traffic condition as a centre point, and according to the empiricism formula to calculate the spread of the fuzzy coefficients (feasible range), then the centre point plus and minus the fuzzy coefficients which its left and right spread, this interval is the adjusting range and the predicted results of the polynomial regressions need to be in this range. For example, assume that the real-time latest average speed is 10 mph and the fuzzy coefficient is 4 mph. In that way, the adjusting range should be in [6, 14]. So that we can determine how many predicted results in this range from six outcomes from polynomial regressions. This range will change in real time with the latest data updating. More details about adjusting range empiricism formula will be presented and tested at next section.

Weighted Moving Average

The method of Moving Average (MA) and Weighted Moving Average (WMA) that we have a brief introduction in the section 2.4.1. The WMA is the third part of the prediction model, mainly because it has simple modelling and calculation. It also can quickly analyse the trends in the future which based on a small amount of data, and provides a change trend and one more possible results.

The historical data is in accordance with the time series. However, the influence of result is different for each data. Therefore, the weight parameter is the key to prediction accuracy. The weight refers to certain indicators of the degree of relative importance in the overall evaluation. The greater weight has the higher importance of the indicators and gives higher impact on the overall. Weight needs to meet two conditions: the weight of each indicator lies between 0 and 1; the sum of each indicator must be 1. There are many methods to determine the weight, but we take a subjective evaluation method in this thesis. Because of the traffic data is usually complex and influenced by many factors, the effects of some methods are poorly for the travel speed such as factor analysis and principal component analysis. Hence we evaluate the weight based on our experiences and objective rules, then test the different weights via experiments, finally we need to select the most appropriate value of the weight. Thus, the more recent data should have more weight than prior data, it gives greater impact on the predicted result according to our experiences and objective rules. The WMA model intends to use three weights in this thesis, and the principle is the most recent data has the highest weight. Meanwhile, we determine the weight parameters by experiments as the following section.

Although the weight parameters are based on a large number of experiments, they are fixed. So the weight parameters do not have self-adaptability, they will not change as the data changes. That way, the weight parameters are consistent with this set of experiments, this one road section or even this city, but might not work for another experiment, roads or cities. So if the WMA model is applied in other scenarios, the weight parameters need to verify by experiment at first. This is one weakness for WMA model. With the development of technology and in-depth research, if we can dynamically assign weights so that there will be a great help for model prediction accuracy and model adaptability.

3.2.4 PRAWMA Prediction Scheme

As describe in the last section, the principles of prediction scheme and operation theories provide the basis and conditions for modelling. There are several summaries of the last section when the prediction model is developed and built. Firstly, because the system is based on the Car to Car communication, this ensures that the input data is updated in real-time. Secondly, the combination model with polynomial regression-adjusting range and weight moving average make the predicted results in the confidence interval. Thirdly, the prediction model also has the ability to fast computing, when the new data is constantly refreshed. Finally, in order to reduce the impact of data loss, the prediction model uses the average algorithm.

This section will be presented how to build the PRAWMA prediction scheme and the introduction of all parameters. We assume a set of known data of travel speed for one road section and these data are optimised to give a list of average speed by time for this road section that it will be applied by PRAWMA scheme, and the polynomial regression equations are constructed by the least squares method at below:

$$\left\{\begin{array}{c} Vp(t)_{1r} = c + bt + \varepsilon \\ Vp(t)_{2r} = c + b_1 t + b_2 t^2 + \varepsilon \\ Vp(t)_{3r} = c + b_1 t + b_2 t^2 + b_3 t^3 + \varepsilon \\ Vp(t)_{4r} = c + b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + \varepsilon \\ Vp(t)_{5r} = c + b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5 + \varepsilon \\ Vp(t)_{6r} = c + b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5 + b_6 t^6 + \varepsilon \end{array}\right\} \tag{3.2-2}$$

$Vp(t)_{ir}$ are predicted speed at t with six polynomial regression equations.
t is time period by orders.
c is a constant number of the regression equation, b_i is coefficients.

ε is stochastic error term.

Each polynomial regression equation should have different fitting curves with the data. And each separate curve describes different regression results. There are six polynomial regression equations that mean could be six different results. However, the six different results should be within a reasonable range, such as the results cannot be negative or too large for speed. In this case, the regression results should be within a certain range. Then calculating the average regression value will be within the valid range. A floating adjusting range needs to set up which an empirical formula will be validated in the section 3.2.6, and the predicted results should be within this range that adjusting range is:

$$a = (1 - \mathrm{H}) * (V_{latest} - V_{t-1}) \quad (3.2\text{-}3)$$

a is adjusting range, estimates the upper limit and lower limit of the predicted results.
V_{latest} is speed at latest average speed by second.
V_{t-1} is the average speed at *t-1* time period.
H is the threshold (h-certain). In this thesis, it is 1/2 by empirical testing which is in the section 3.2.6.

The model will exclude results out of range, and calculated the average value of remaining results as polynomial regression predictable consequences. The polynomial regression formula is:

$$Vp(t)_r = \frac{1}{n}\sum_{i=1}^{n} Vp(t)_{ir} (-|a| < Vp(t)_r < |a|) \quad (3.2\text{-}4)$$

$Vp(t)_r$ are predicted speed at t, use by average regression of polynomial equations.
n is the number of possible results after adjusting range screens

The weighted average formula will be found in the next step. The number of weight is setting by experiment. There is a principle which is the weight of latest data is much higher than the previous one, so that the weighted average is:

$$Vp(t)_w = w_1 V_{latest} + w_2 V_{t-1} + w_3 V_{t-2} \quad (3.2\text{-}5)$$

$Vp(t)_w$ is predicted speed at t, use by weight moving average.
V_{t-2} is average speed of penultimate data in the list which is at *t-2* time period.
w_i is weight parameter, and the it will be specifically discussed in the following section by experiment.

The final predicted result is the average value of polynomial regression results and weighted moving average results, so that the predicted result of PRAWMA prediction scheme is:

$$V_p(t) = \frac{Vp(t)_r + Vp(t)_w}{m} \qquad m=1,2 \qquad (3.2\text{-}6)$$

$Vp(t)$ is predicted speed as expect at t time.

m is 1 that means there is no result from polynomial regression part.

m is 2 that means there have both results from two parts.

In the light of the above introduction, the workflow of the prediction model is showed in Fig. 3.3 with five steps. The first step is that six polynomial regression equations should be generated based on the least square method by using Eq.(3.2-1). The second step filtrates possible results of six regression equations according to the adjusting range with Eq.(3.2-3), then the possible results are averaged after screening from six regression equations in step 3 with Eq.(3.2-4) that as one predicted the result for the polynomial regression (PR) part. The step 4 needs to build the weight moving average (WMA) part by using Eq.(3.2-5) as other predicted result. In the last step, two results from PR part and WMA part are averaged as the outcome of PRAWMA prediction scheme.

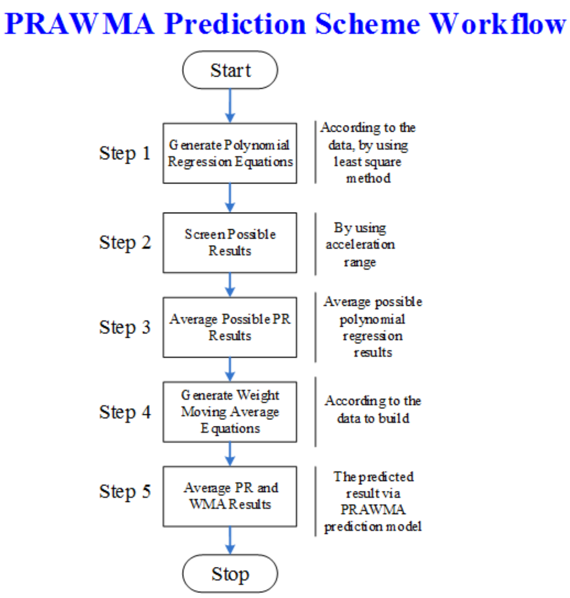


Fig. 3.3 PRAWMA prediction scheme Workflow

However, there is a very special situation that is no possible result after screening in step 2. Then the predicted result is the result from WMA part. Although the probability of this happening is very small and even can be ignored, this will impact the error of one predicted result. Nevertheless, we can increase the number of polynomial equations with offering possible results for screening to solve this problem, if this situation occurs very often in one scenario or experiment.

3.2.5 Model Processing

This section will demonstrate how the PPM prediction model working, and the parameters are proven by experiments. According to the previous section, the workflow of the prediction scheme does not include data compilation, so that the data needs to meet the demands of the PPM prediction model. The PPM requires that the data is arranged in time series, and a list of data should include the time period and travel speed. All data sources are real-time information and based on C2C communications so that the data can be stored in chronological order. Space mean speed will be employed in this project and the all average speed mentioned is space mean speed in this article.

We declare some information to demonstrate how the PPM works. We are setting that each vehicle has a data table to store messages from other vehicles, and the prediction scheme would call the data from Table 3.1, shows at below. This data table is one part of PPM and is only describing a list of information on the Road Section A, and the other road sections would display an identical table. The contents of the table include the name of Road Section, CarID, Time and Speed, and Time/p (time period), speed/ave (average speed) and latest/s (latest average speed by second). The model would compute the average speed in a time period and call the latest received speed from the table. It means that the data table changes in real-time, with data information income. If the table has been given the real information, and it will be like Table 3.2. The data (assumed as ad-hoc data) is recorded from SUMO simulation and picks the Parliament Street from east to west of Nottingham city centre scenario in a period of simulation time. The information of all vehicles driving on this road have been recorded and output in Table 3.2 as a contracted form. Table 3.2 also can expand other parameters, such as using distance and speed to obtain travel time or using car id to get the density of the road. The main purpose of this chapter is proposing, discussing and evaluating the prediction model, and taking travel speed prediction as an example. The most of the simulation parts will be presented in the following chapter.

Table 3.1 Data table

Road Section A					
Car-ID	Time/s	Speed/mph	Time period	Average speed	Latest speed

ID	1	x1	1	v1	x2
ID	2	x2	2	v2	x3
ID	3	x3	3	v3	x4
….	….	….	….	….	….
….	….	….	….	….	….
….	….	….	….	….	….
ID	n-1	x(n-1)	t-1	v(t-1)	x(n)
ID	n	x(n)	t	v(t)	x(n+1)

Table 3.2 Data table from SUMO output

Parliament Street from East to West					
Car-ID	Time/s	Speed/mph	Time/p	Speed/ave	Latest/speed
251	0:06:31	12.91	391-400s	12.55	
49	0:06:31	11.13	401-410s	9.48	…
128	0:06:31	11.90	411-420s	8.30	…
785	0:06:31	12.74	421-430s	11.81	…
251	0:06:32	13.32	431-440s	12.36	…
128	0:06:32	12.44	441-450s	13.21	…
49	0:06:32	11.53	451-460s	12.31	…
251	0:06:33	12.69	461-470s	8.30	…
49	0:06:33	12.29	471-480s	5.75	…
128	0:06:33	12.76	481-490s	7.16	4.52
785	0:06:34	12.61			
251	0:06:34	12.52			
128	0:06:34	12.45			
…	…	…			
305	0:07:40	12.57			
305	0:07:41	12.90			
…	…	…			

Table 3.2 shows the core of parameters in PPM that apply to the PRAWMA prediction scheme, and also affect the prediction results. The predicted results force on the average speed on the road section.

And this example, we are assuming that the data collection is immediate because the outputs come from SUMO simulation for the road section of the Parliament Street. We have not introduced the network communications in this first example, and its purpose is to maintain the integrity and timeliness of the data to make better validation and interpretation for the model. Table 3.2 is a piece of data from SUMO's record, as the explained detail of operation and procedures for the PRAWMA prediction scheme. The left three columns are received messages from other vehicles, there is the output from SUMO's record in this case. And the right three columns will generate the space average speed in the time period of this following Eq.(3.2-7) before running the prediction model (Knoop, Hoogendoorn and van Zuylen 2009):

$$\text{Average speed:} \quad \bar{v} = \frac{1}{\frac{1}{n}\sum_{i=1}\frac{1}{v_i}} \tag{3.2-7}$$

$\bar{v}$ is space mean speed which is thus the harmonic mean of the speeds.
v_i is speed of vehicle in second.
n is the number of vehicles passing the roadway segment.

The right three columns in Table 3.2 are given, and there are arranged in time series with ten samples, hence the PRAWMA prediction scheme can be performed when the parameters are satisfied. The parameters that all demands of the prediction model are retrieved at below:

Table 3.3 Data extract from Table 3.2

Serial No	Time/p	Speed/Ave	Latest speed
1	391-400s	12.55	
2	401-410s	9.48	…
3	411-420s	8.30	…
4	421-430s	11.81	…
5	431-440s	12.36	…
6	441-450s	13.21	…
7	451-460s	12.31	…
8	461-470s	8.30	…
9	471-480s	5.75	…
10	481-490s	7.16	4.52

From above table list, there are the travel speed at 391 seconds to 490 seconds on a road section and divides 10s time period to calculate the average speed, thus, the average speed between 491s -500s will be predicted. According to 10 average speeds to generate optimal curve fitting and polynomial regression group using least square methods:

$$\begin{cases} Vp(t)_{1r} = 11.391 - 0.193t \\ Vp(t)_{2r} = 7.119 + 1.943t - 0.194t^2 \\ Vp(t)_{3r} = -0.039 - 1.614t + 0.4778t^2 - 0.0387t^3 \\ Vp(t)_{4r} = 25.80 - 18.921t + 6.811t^2 - 0.904t^3 + 0.039t^4 \\ Vp(t)_{5r} = 20.242 - 9.744t + 1.993t^2 + 0.173t^3 - 0.0677t^4 + 0.0039t^5 \\ Vp(t)_{6r} = 19.251 - 7.787t + 0.658t^2 + 0.658t^3 - 0.135t^4 + 0.0091t^5 - 0.0002t^6 \end{cases}$$

And the Fig. 3.4 shows the observation data against to six different regression equations, also gives the predicted results at 11^{th} time period for each regression equation.

From the scatter plot, the six fitting curve generate by the least square method and have six different results. As the traditional regression analysis, we need to test the sample and modelling. The observing the following statistical results from Eviews in the Fig. 3.5 supports the following conclusions. In addition, according to the normal probability plots (Q-Q diagram) in the Appendix A. their residuals have already met the normal distribution, because the Q-Q diagram shows that the residuals are distributed on the both sides of the line, except some points seem away from the line in the cubic one. Therefore, data meet to the normality assumption. Besides, as the normal test for small samples is not obvious, so that it cannot be considered in this project.

The t-test is used to test the significance of the regression coefficient. In the case where the assumed coefficient is zero at 5% significance level. The 4^{th} regression is the only one that p-values of all coefficients lower than 0.05, therefore reject the null hypothesis, which means the time has a significant effect on the speed. In the other words, the p-values of other models are higher than 4^{th}. The absolute t-statistic values of 4^{th} regression are bigger than t(df). The t(df) value comes from t-test table, in this case the t(df) is 2.447 at 0.05 significance level with 6 degrees of freedom.

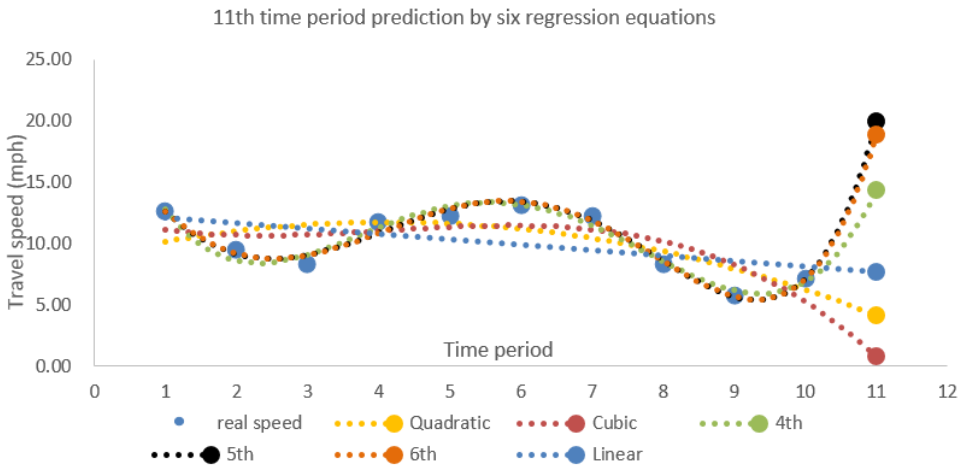


Fig. 3.4 11th time period prediction by six regression equations

The F-test is used to test the significance of the regression model for the prediction. The Prob(F-stat) of 4th, 5th and 6th regression model are obviously less than 0.05, and also the F-statistic value of them are bigger than F critical value (f(4th)22.47> f(0.05,3,6)4.75; f(5th)22.87>f(0.05,4,5)5.19; f(6th)14.35 > f(0.05,5,4)6.26). Meanwhile, the R^2 is over 0.9 in the regression of 4th, 5th and 6th, there is high significance between the time and speed. Therefore, the regression model of 4th, 5th and 6th are significance, and can be a prediction.

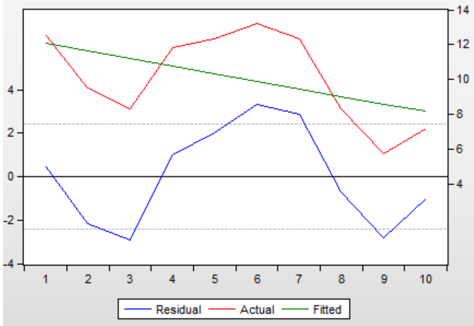
The Durbin-Watson stat value always lies between 0 and 4, and as a rough rule of thumb it is usually between 1.5 and 2.5 what the dependent series does not have autocorrelation (Xu and Fu 2009). Also from the d-w test table, we can estimate that the d-w value of linear regression (0.986) has a strong positively correlated and the d-w value of quadratic and cubic regression (1.524 and 1.694) do not have autocorrelation; the d-w value of 4th, 5th and 6th regression (3.121, 3.488, 3.446) lie between dL and dU at 0.05 significance level.

Also, according to 6 residual graphs, the fitting effects of 4th, 5th and 6th are better than linear, quadratic and cubic. And all residuals of 4th, 5th and 6th are distributed on both sides of 0 and most of them are in the confidence zone, where 4th is the best one. The equation fitting effect of 4th, 5th and 6th are good. Taking into account the statistical results, the model of 5th and 6th may need to be re-fitted, followed by another testing. But our PPM model is based on in-car design, so that it is difficult to delete certain variables to re-fit when the modelling fails. Therefore we recommend removing the failed model and results directly.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.53200	1.659115	7.553426	0.0001
T	-0.438000	0.267391	-1.638054	0.1400

R-squared	0.251162	Mean dependent var	10.12300
Adjusted R-squared	0.157557	S.D. dependent var	2.646079
S.E. of regression	2.428695	Akaike info criterion	4.789442
Sum squared resid	47.18848	Schwarz criterion	4.849959
Log likelihood	-21.94721	Hannan-Quinn criter.	4.723055
F-statistic	2.683219	Durbin-Watson stat	0.986428
Prob(F-statistic)	0.140045		

Linear-stat output

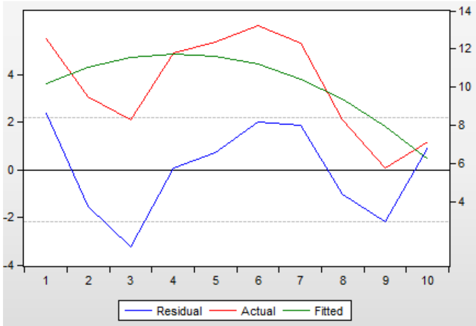


Linear-residual graph

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.988667	2.572671	3.493905	0.0101
T	1.333667	1.074456	1.241249	0.2545
T^2	-0.161061	0.095193	-1.691942	0.1345

R-squared	0.468514	Mean dependent var	10.12300
Adjusted R-squared	0.316661	S.D. dependent var	2.646079
S.E. of regression	2.187363	Akaike info criterion	4.646595
Sum squared resid	33.49189	Schwarz criterion	4.737371
Log likelihood	-20.23298	Hannan-Quinn criter.	4.547015
F-statistic	3.085315	Durbin-Watson stat	1.524167
Prob(F-statistic)	0.109451		

Quadratic-stat output

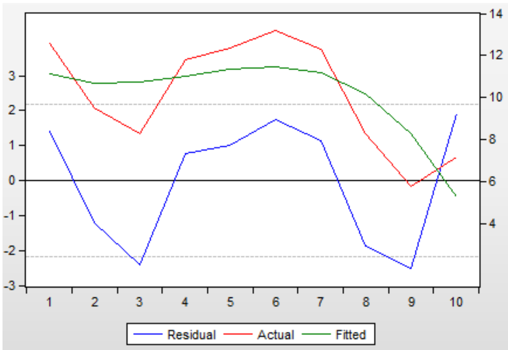


Quadratic- residual graph

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.31967	4.254745	2.895512	0.0275
T	-1.620752	3.189133	-0.508211	0.6294
T^2	0.479516	0.657810	0.728959	0.4935
T^3	-0.038823	0.039446	-0.984209	0.3630

R-squared	0.542393	Mean dependent var	10.12300
Adjusted R-squared	0.313589	S.D. dependent var	2.646079
S.E. of regression	2.192275	Akaike info criterion	4.696931
Sum squared resid	28.83641	Schwarz criterion	4.817965
Log likelihood	-19.48465	Hannan-Quinn criter.	4.564157
F-statistic	2.370559	Durbin-Watson stat	1.693598
Prob(F-statistic)	0.169510		

Cubic-stat output

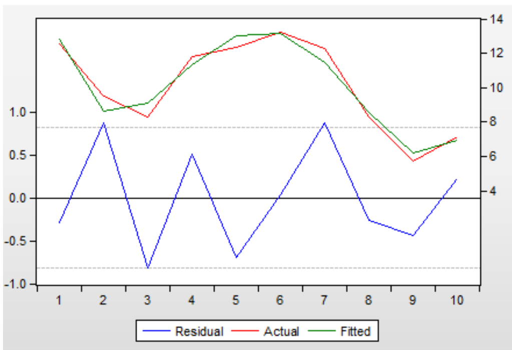


Cubic- residual graph

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	25.82667	2.692519	9.592008	0.0002
T	-18.93742	3.034771	-6.240148	0.0015
T^2	6.815842	1.051061	6.484723	0.0013
T^3	-0.904656	0.140450	-6.441123	0.0013
T^4	0.039356	0.006349	6.198578	0.0016

R-squared	0.947307	Mean dependent var	10.12300
Adjusted R-squared	0.905153	S.D. dependent var	2.646079
S.E. of regression	0.814918	Akaike info criterion	2.735394
Sum squared resid	3.320455	Schwarz criterion	2.886686
Log likelihood	-8.676969	Hannan-Quinn criter.	2.569426
F-statistic	22.47251	Durbin-Watson stat	3.121329
Prob(F-statistic)	0.002147		

4th-stat output

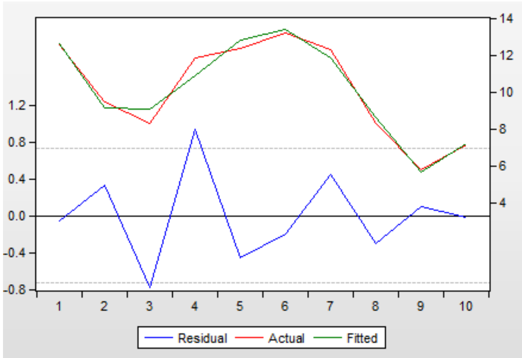


4th- residual graph

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	20.23867	4.446104	4.552000	0.0104
T	-9.725685	6.731216	-1.444863	0.2220
T^2	1.980073	3.367102	0.588064	0.5881
T^3	0.176472	0.733639	0.240543	0.8217
T^4	-0.068105	0.072067	-0.945026	0.3982
T^5	0.003908	0.002612	1.495788	0.2090

R-squared	0.966209	Mean dependent var	10.12300
Adjusted R-squared	0.923969	S.D. dependent var	2.646079
S.E. of regression	0.729622	Akaike info criterion	2.491128
Sum squared resid	2.129390	Schwarz criterion	2.672679
Log likelihood	-6.455639	Hannan-Quinn criter.	2.291967
F-statistic	22.87461	Durbin-Watson stat	3.488438
Prob(F-statistic)	0.004828		

5th-stat output

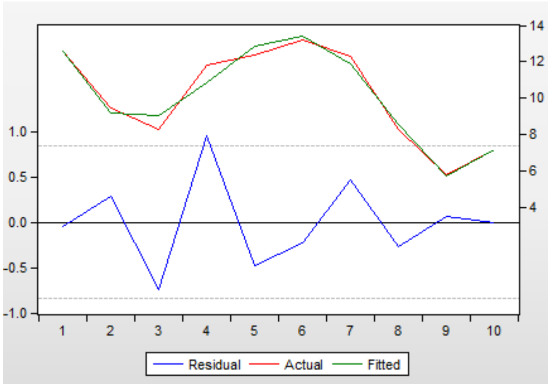


5th- residual graph

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	19.26800	10.67316	1.805275	0.1688
T	-7.810485	20.03582	-0.389826	0.7227
T^2	0.673208	13.18901	0.051043	0.9625
T^3	0.589472	4.072261	0.144753	0.8941
T^4	-0.133934	0.640342	-0.209160	0.8477
T^5	0.009041	0.049604	0.182265	0.8670
T^6	-0.000156	0.001500	-0.103678	0.9240

R-squared	0.966329	Mean dependent var	10.12300
Adjusted R-squared	0.898988	S.D. dependent var	2.646079
S.E. of regression	0.840989	Akaike info criterion	2.687551
Sum squared resid	2.121788	Schwarz criterion	2.899361
Log likelihood	-6.437755	Hannan-Quinn criter.	2.455197
F-statistic	14.34965	Durbin-Watson stat	3.445958
Prob(F-statistic)	0.025952		

6th-stat output



6th- residual graph

Fig. 3.5 Comparison of statistical tests and Residual graph with 10 samples for six estimated equations

According to the statistical tests, we learn that the regression model of 4th, 5th and 6th might be used for the prediction. But only the 4th regression model that the p-value is lower than 0.05. Therefore, from a statistical point of view, we can use the 4th regression model to predict. The predicted results of the six regression model are, 7.71, 4.17, 0.85, 14.35, 19.92 and 18.93, but only 4th regression model pass the significant test and the predicted result is accepted that is 14.35. In the regression predictions, the predicted results can only be made after a significant test for the established model. In this example, the model of 5th and 6th regression could one by one try to remove some variables such as t, t^2 or t^3, and meanwhile re-test the significance of the new model. Repeated and cumbersome tests usually require the professional software and comparisons to complete. However, in our dynamic prediction system and based on the ad-hoc data that is difficult to achieve. Therefore, if the ad-hoc data set is small and regression analysis is difficult, the concept of fuzzy regression is used. The fuzzy coefficient in fuzzy regression is replaced to adjusting range in this project.

The adjusting range can judge and filtrate the possible results through our model, the verification of it will be shown in the section 3.2.6. When t=11, there are 6 results have been given and the adjusting range is [3.91, 5.14] via Eq.(3.2-3) to drop the results out of the adjusting range and get the average number by Eq.(3.2-4), has:

$$Vp(t)_r = 14.35 \quad (Vp(t)_{ir} \in [9.92, 15.43])$$

There are only 1 possible values in the adjusting range which is 4^{th} order polynomial regression equation, and the result is the same with a statistical test. When the weights are set as 0.7, 0.2 and 0.1, because the more recent data have more influence on the prediction result, based on Eq.(3.2-5) so that:

$$Vp(t)_w = 0.7 * 12.68 + 0.2 * 7.16 + 0.1 * 5.75 = 10.88$$

Finally two results have been combined to compute the prediction result as the average speed at t=11 time period which is 391s to 490s on this road section, following Eq.(3.2-6).

$$V_p(t) = \frac{14.35+10.88}{2} = 12.62$$

Therefore, the prediction of average speed is 12.62 mph between 491s-500s, and the result compares with the actual simulation average speed by SUMO's recorded which is 10.92 mph, and it is also in the adjusting range.

The Fig. 3.6 shows an experiment in 20 minutes and the time period is 10s, y/10 means time period is 10s, and P with a means predicted results with adjusting range which is the PRAWMA scheme.

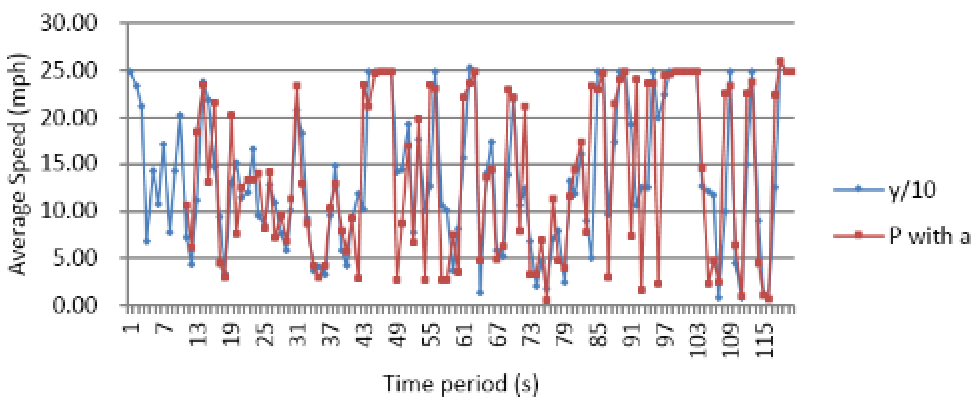


Fig. 3.6 Actual average speed Vs Predicted average speed

Then we will use the correlation coefficient number to compare that how are the fitting of prediction result and actual recording data. The correlation coefficient number can reflect the exact degrees of relationship and direction between each variable. It is calculated by the product moment, based on each variable and deviation of their respective average. Through the deviation multiplied to reflect the degrees of correlation between two variables, that is whether there is a linear relationship between variables thereby estimating the similarity based on the correlation. Also, the similarity of two curves is evaluated by the angle of cosine. Therefore, testing of correlation and cosine similarity are between two curves from Fig. 3.6 that should be following:

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}} = 0.786$$

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{n}(y_i x_i)}{\sqrt{\sum_{i=1}^{n}(y_i)^2\sum_{i=1}^{n}(x_i)^2}} = 0.942$$

The cosine similarity gives a higher evaluation of the predicted results and the similarity based on correlation still show a high degree of similarity although its result is lower than expected. From the scatter plot, the predicted results have a difference with the observation data at the time period of 49^{th} to 54^{th} and 91^{st} to 96^{th}. The errors of predicted results are quite obvious when the average speed changes and fluctuates in continuous. That similarity between predicted results and observation data is a high degree of values, directions and trends.

Specifically, the latest/s should be the latest average speed at 400^{th} second in this case, but only one speed is recorded which is 9.65 mph for the latest average speed at 400^{th} second. However, the latest average speed by the second is not the latest instant speed by received. Because we discuss the average speed of one road section, but the information will be differences on one road section at one time, when we collect the information from each vehicle. For example, vehicle A has 5.5 mph at 100^{th} second on the road nearby the intersection, and vehicle B has 22.6 mph also at 100^{th} second but in the middle of the road, they are on the same road section. In this case, the prediction model does not know which one should be the latest speed, and only can according to the default order, or the simulation time needs to be accurate to the decimal places. Then the predicted results might give large errors with the value but the predicted trend still can be expected. Therefore, we take the average speed at latest second to reduce errors.

On the other side, if there are no vehicles on the road section, so no record of information can be used to predict. In this case, the prediction model will generate the default data to fill the vacancy which is the maximum speed of the road section. We set the maximum speed of all city roads is 30 mph (about 48 km/h or 13m/s) in the simulations of this thesis so that the default speed is 30 mph to fill the vacancy.

This section introduces data type and data management, details process of the PRAWMA prediction scheme and evaluates the predicted result by similarity. The prediction model is proved to be feasible through one example for 20 minutes running. The following section will explain and present the parameters in the model, such as the weight parameters.

3.2.6 Verification of Adjusting Range and Model's Weight

As described and demonstrated in the last section, we only emphasised the principles of the adjusting range and weight parameters but we have not explained the reasons for choosing the formula of adjusting range and the weight parameters by discussing the corresponded experimental results. In this section, we will verify the adjusting range and weight parameters via a set of experiments.

Adjusting Range

The adjusting range is an important part of the PRAWMA scheme as described in the previous sections. It estimates a fuzzy range of the predicted results to determine the results from polynomial regressions. In this thesis, we believe that traffic condition will fluctuate in tendency over time, so that the latest data best reflects the current traffic conditions. Hence, we test by a set of assumptions of adjusting range based on the latest data to arrive at the better one. According to fuzzy regression, the threshold (h-certain) is usually 0.5 that needs to between 0 and 1, so that in our testing case, the threshold sets by 1/4, 1/3, 1/2, 2/3 and 3/4, the adjusting range formula sees also equation (3.2-3). We run 500 nodes in 1 hour with SUMO and divide into four different time periods with 5 seconds, 10 seconds, 20 seconds and 30 seconds. The adjusting range sets to 5 groups with different thresholds at below table, and the error in range and the similarity by correlation-based via the prediction model are verified. The error in range means whether the predicted results are in the range.

Table 3.4 Comparison of Adjusting range with the different thresholds

ACCERLATION RANGE	TEST 1(5S) Error in Range	TEST 1(5S) Similarity	TEST 2(10S) Error in Range	TEST 2(10S) Similarity	TEST 3(20S) Error in Range	TEST 3(20S) Similarity	TEST 4(30S) Error in Range	TEST 4(30S) Similarity
(1-3/4)*$(\boldsymbol{V_{last}} - \boldsymbol{V_{t-1}})$	36.43%	0.915	43.30%	0.863	56.14%	0.750	52.25%	0.733
(1-2/3)*$(\boldsymbol{V_{last}} - \boldsymbol{V_{t-1}})$	33.76%	0.915	40.17%	0.862	53.22%	0.752	50.45%	0.733
(1-1/2)*$(\boldsymbol{V_{last}} - \boldsymbol{V_{t-1}})$	27.99%	0.916	34.76%	0.864	44.95%	0.757	45.95%	0.741
(1-1/3)*$(\boldsymbol{V_{last}} - \boldsymbol{V_{t-1}})$	23.35%	0.916	32.48%	0.861	40.94%	0.756	42.34%	0.741
(1-1/4)*$(\boldsymbol{V_{last}} - \boldsymbol{V_{t-1}})$	20.68%	0.911	29.34%	0.863	39.18%	0.748	40.54%	0.731

The selection of adjusting range has two needs which simultaneously satisfy: the adjusting range is as small as possible; and the accuracy is as high as possible. As the table shown, when the H value is going to small, the error in the range is going to better, this is because the upper and lower bounds of adjusting have broadened and the model will judge whether the polynomial results is in this range. If the adjusting range is too big that means more inconformity results of the statistical analysis will be selected. The aim of adjusting range to replace the complex statistical analysis as reasonably as possible and reduce as far as possible to the inconformity results of the statistical analysis. The similarities of them only have trifling change. Therefore, when H is 1/2 or 2/3, the results satisfy our two needs. Also, considering with the literature (Xie, Jiang and Wei 2014), the H value chooses 1/2 to be universal so that is employed in our case. More details of Adjusting range for the parameter settings are presented in Appendix A.

Model's Weight

The weight is an important parameter in the weight moving average model as described in the section 2.4.1. It can reflect the degree of importance of a certain element in the overall evaluation. The elements that influence the outcome seriously take a significant weight. Hence, the assignment of weights directly affects the effectiveness of the results. In this thesis, the more recent elements have a greater influence on the future, since the speed of one second before decides the speed of next second. Therefore, the most recent speed has the highest weight, and the earlier speed gives poor influence to predicted results.

The WMA model uses three weight parameters. The first element is the speed at latest data by received, usually it is instant speed and has the highest weight. The second element is the average speed of the last time period, usually it is at *t-1* time period and has the middle weight. The third element is the average speed of penultimate time period which is at *t-2* time period and it has the lowest weight. The sum of the weight should 1 by these three elements. We assume a few groups of the feasible weights at first, and then compare the means relative errors (Table 2.6) between predicted results and observation data by the prediction model in the different time periods, in order to determine the optimal weight setting for this project. We run 500 nodes in 1 hour with SUMO, record SUMO's outputs as a sample list for the travel speed, and randomly selected samples in 20 minutes from this list to divide into four different time periods with 5 seconds, 10 seconds, 20 seconds and 30 seconds. The weights are set to 6 groups at below table, and verify their means relative errors and similarity by correlation-based via the prediction model as shown at following charts of Fig. 3.8 and Fig. 3.9 according to the table.

Table 3.5 Comparison of Weights in the different time periods

	TEST 1(5S)		**TEST 2(10S)**		**TEST 3(20S)**		**TEST 4(30S)**	
WEIGHTS	MRE	Similarity	MRE	Similarity	MRE	Similarity	MRE	Similarity
0.9/0.09/0.01	28.17%	0.922	34.31%	0.865	47.94%	0.755	29.90%	0.743
0.8/0.15/0.05	28.72%	0.921	33.87%	0.866	46.08%	0.756	28.50%	0.747
0.7/0.2/0.1	29.69%	0.917	33.66%	0.864	44.21%	0.753	27.20%	0.747
0.6/0.3/0.1	30.05%	0.912	33.30%	0.858	42.39%	0.747	26.05%	0.742
0.5/0.4/0.1	30.48%	0.905	33.07%	0.850	40.62%	0.737	25.12%	0.732
0.4/0.35/0.25	33.50%	0.891	33.63%	0.835	38.66%	0.712	24.07%	0.708

MRE can test the value accuracy between predicted results and observation data. But it will be influenced by special cases thus leading to decrease in average index. For example, in test 2 which is 20 seconds time period with the weight of 0.9/0.09/0.01, the MRE is 47.91% and is quite high with an error. But there is one special case at peak point as the Fig. 3.7 shown which MRE is 349.84%, 270.35% and 200.01%, these errors are senseless and under this situation, these errors also influence the average error for the testing. If this special case is removed so that the MRE should be 24.48%. In order to verify the significance of weight and the feasibility to a prediction by applying the PPM prediction model, so that we do not remove the special cases and retain all the errors in these weights testing in this subsection. But we will use MRE remove cases

(MRErc) for testing the effects of the prediction model in the following sections and chapters.

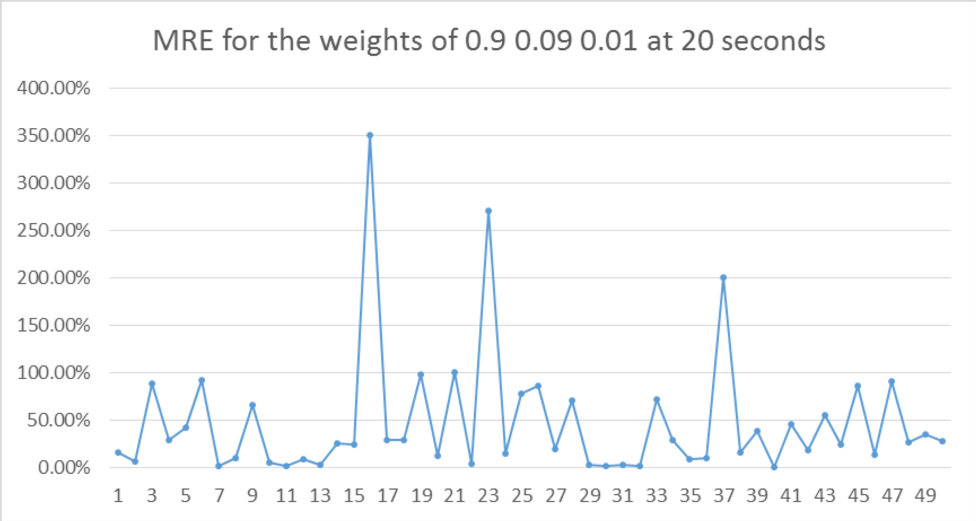


Fig. 3.7 MRE for the weights of 0.9/0.09/0.01 at 20s time period

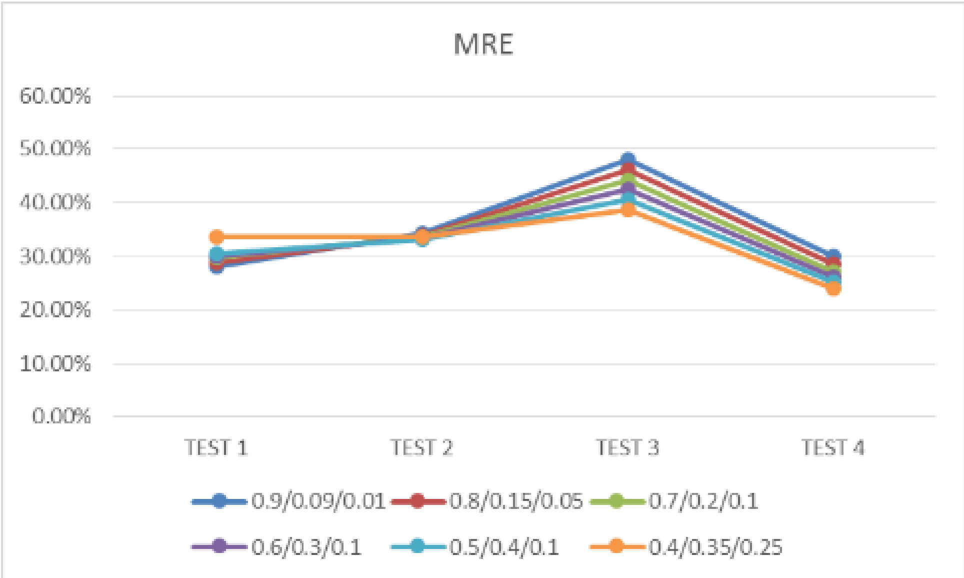


Fig. 3.8 Comparison of the weights in MRE

From the Fig. 3.8, the weights of 0.9/0.09/0.01 have the lower mean relative errors with the better predicted results than other settings in test 1 and test 2 as shorter time periods. But their MREs rise to high when the time period increases. This means that the setting of 0.9/0.09/0.01 is ideally suited to predict in short time period. The MRE does not have an obvious difference in test 1 and test 2, but the MRE of all settings show fluctuation in test 3 and test 4 that there is a significant difference. However, reduction in the first weight is more suitable to the longer time period. In addition, there is small MRE in test 4, because the number is only 30 in 20 minutes.

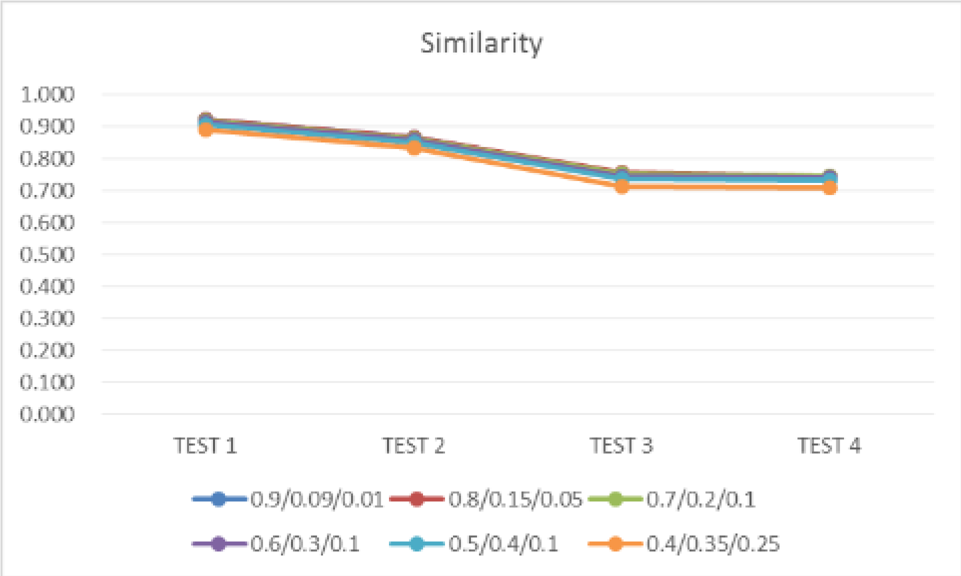


Fig. 3.9 Comparison of the weights in Similarity by correlation-based

As can be seen from the Fig. 3.9, the similarity of all settings show a trend of general fall with the increase of time period. But there is little difference between each other at same time period. We are convinced that the similarity does not vary with settings of weight but only changes with different time period.

From a calculation point to sum up two figures, the weight of 0.9/0.09/0.1 considerably depends on the V_{latest} and best suit for under 10 seconds time period; the weight of 0.8/0.15/0.05/ and 0.7/0.2/0.1 can be seem in a same group to depend on the V_{last} but not stronger than first one, and their changes are relatively stable; when the weights are set 0.6/0.3/0.1 or 0.5/0.4/0.1, the average speed of the latest time period would receive more consideration than the first three settings, and these two settings better suit 20 seconds and 30 seconds time period by determining MRE; the setting of 0.4/0.35/0.25 seems to share the weights more evenly but it is the least effective for the prediction in general. Therefore, general considerate the testing results, we intend to adopt relatively stable weight setting that is 0.7/0.2/0.1 in the following sections. More details of MRE testing figures for the weight settings are presented in Appendix A.

3.3 Performance Evaluation of PRAWMA Scheme

We conducted simulation using SUMO to present the process and weight parameters of PRAWMA scheme in the PPM at the previous sections. We will describe the simulation of SUMO. This chapter focuses on the design and generation of the prediction model for ad-hoc data. We assume that the wireless environment has infinite speed for data transmission. In this case the data is received instantaneously without packet loss and time delay that we name it "no wireless environment". Thus, we use observation data from SUMO as ad-hoc data in this chapter.

3.3.1 Mobility Pattern Generation

Before we tested and verified the performance evaluation the model, the simulated experiment should be prepared. Because of statistical, technical and financial reasons, site investigation and experiment in the real scene might give information shortage or difficulty of measuring, so that most experiments need to be simulated scenarios. The more realistic simulation scenario can give more true and objective effect for the model. Therefore, in order to validate the prediction model in the practical application, the experiment simulates the real city scene which is Nottingham (UK) city centre and the Parliament Street from east to west direction is chosen for objective road section, that it is red line road on the city map in Fig. 3.10.

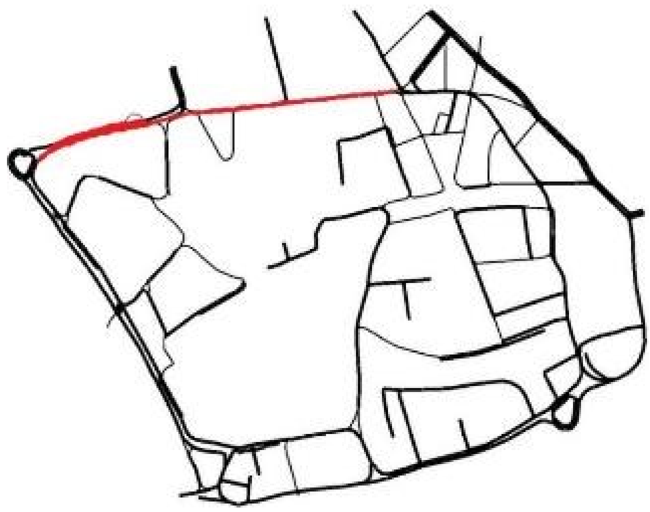


Fig. 3.10 The map of Parliament Street in Nottingham city centre

All experiments will use this scene as a base map in this thesis, and the prediction of the target road section also will be carried out around the Parliament Street from east to west. All vehicles will follow the road to move and comply with the traffic lights at each intersection. The routes of vehicles are generated by randomly. All these of processes depend on SUMO, and this section introduces the generation of mobility pattern, the simulation of scenario and the speed data collection at the request of prediction model by SUMO. The vehicle's route topology is completely random generation by SUMO. The simulation parameters for the scenario are presented in Table 3.6.

Table 3.6 Simulation Parameters with mobility pattern

Parameter	**Value**
Simulation area	1200m x 1000m
Number of vehicles	500
Vehicle's speed	0 - 30 mph (0- 48 km/h)
Mobility Model	SUMO defined
Time Period	5, 10, 20, 30 seconds
Simulation Time	1 hour (3600 seconds)
Network Environments	None
Setting of Weights	0.7/0.2/0.1

Then we elaborate the simulation process by thumbnail codes. First, we should find the Nottingham city map by net.xml from OpenStreetMap (OSM). And clear all mark and colours on the map, because some unnecessary parameters can affect the experimental running speed such as corner shops, street signs and background decoration, so that make the map file resources to minimized. The "netconvert" function can generate the output file map which is nottinghamcenter.net.xml. The road network topology generates by this command at below:

```
netconvert–osm-files nottinghamcenter.osm.xml –o nottinghamcenter.net.xml
```

The file of randomTrips.py in sumo/tools/trip is used to generate the scenario, the .net.xml is an input. The output file should be named as *.trips.xml which is a random trip file. The tool of duarouter can give the route file and determines the movement and route of vehicles with the input file of.net.xml and .trips.xml. The command is:

```
randomTrips.py –n nottinghamcenter.net.xml
duarouter –n nottinghamcenter.net.xml –t nottinghamcenter.trips.xml –o nottinghamcenter.route.xml
```

The main contexts of .net.xml and .trips.xml file show at below, and relative attributes and parameters described in section 2.6.2 Table 2.8.

```
-- nottinghamcenter.net.xml
<edgefunction=”internal” id=”:1061186719_0”><laneid=”:1061186719_0_0” shape=”82.28,660.36
80.28,658.57 80.20,656.04 80.20,653.34 78.44,651.03” length=”10.83” speed=”27.78” index=”0”/>
</edge>
…………
<edgeid=”-102189762” shape=”471.79,108.75 483.83,110.50 494.04,112.92”
type=”highway.unclassified” priority=”5” to=”1179380041” from=”1179380081”><laneid=”-
102189762_0” shape=”472.69,107.22 484.07,108.87 494.42,111.32” length=”22.14” speed=”22.22”
index=”0”/>
</edge>
…………
```

```
--nottinghamcenter.trips.xml
………..
<trip to="152183982" from="-120036351" depart="0.00" id="0"/>
<trip to="-16470506#0" from="46639857" depart="1.00" id="1"/>
<trip to="-4365253#1" from="46639847" depart="2.00" id="2"/>
<trip to="-116231273#9" from="14793202#1" depart="3.00" id="3"/>
<trip to="116231273#14" from="-114898980#1" depart="4.00" id="4"/>
………..
```

```
-- nottinghamcenter.route.xml
<vehicledepart="1.00" id="t1"><routeedges="4565919#2 5205709#0 43080011 43080010#0
43080010#1 43080013#0 43080013#1 36976280#0 36976280#1 36976280#2 34338973#1
31525828#0 31525828#1 -31525828#1"/>
</vehicle>
```

The main contexts include CarId, Time, Speed and Position that we need to record them as ad-hoc data by assumed in this chapter and apply them to the PPM prediction model. The function is an abstract edge where is in <edge function=”internal”……>. It describes how to use the edge and is defined by:

- Normal: this edge is a common part of the road network, such as streets or highways.
- Connector: this edge is a micro-connector and is not part of the road network in the real world.
- Internal: this edge is a part of intersections, it is different with “normal” edges in the simulation.

Now the mobility file can generate by .net.xml and route.xml with the following steps. This process is based on the 'trip' which is SUMO's tool. And there is one car per second is injected in the network.

```
/home/sky/Desktop/sumo/tools/trip/randomTrips.py -n nottinghamcenter.net.xml -r
nottinghamcenter.route.xml -b 0 -e 3600
sumo -n nottinghamcenter.net.xml   -r nottinghamcenter.route.xml --netstate-dump netstate.xml
```

We run two files in sumo's directory, where –b is defined the time at begin and –e is defined the time at the end by second and a netstate.xml file has been generated. Then use the traceExporter to create three .tcl files with this command:

```
$java -jar traceExporter.jar ns2 -n nottinghamcenter.net.xml -t netstate.xml -a activity.tcl -m
mobility.tcl -c config.tcl -p 1 -b 0 -e 3600
```

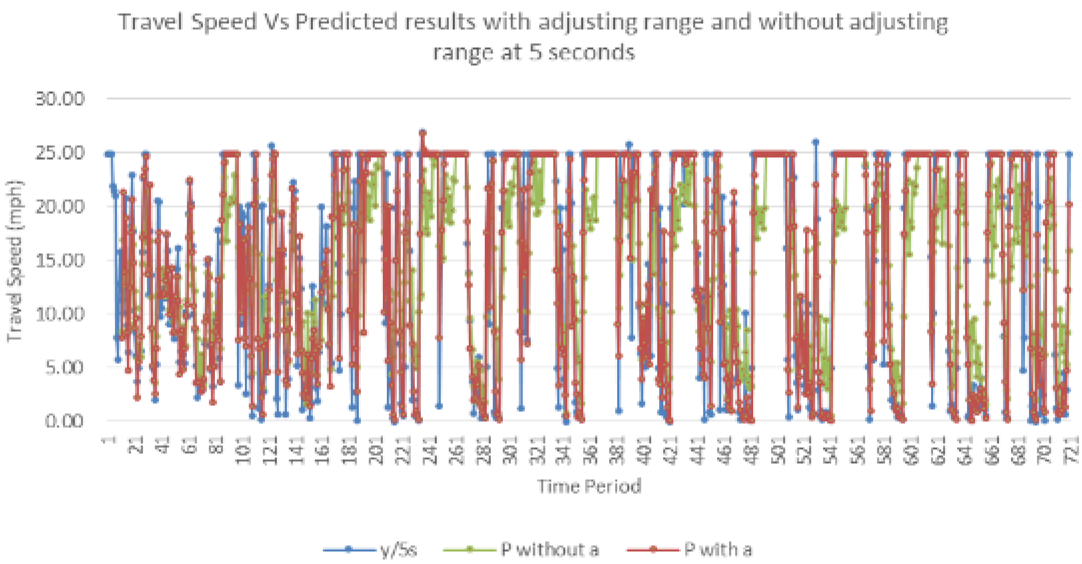
-p is the penetration rate and should be [0, 1]. If it is 0 that no nodes are selected for print out, and if it is 1 that all nodes are selected. Three .tcl files have been created. The Config.tcl holds information about simulation settings like extending, start/stop time and number vehicles in output. The Activity.tcl will be written, when vehicles will be activated and deactivated in the simulation. The Mobility.tcl describes vehicle's movement and their corresponding position, the information for our prediction model required in Table 3.2 in the section 3.2.5 comes from this file. And we need to change the extension of Mobility.tcl to mobility.ns_movments, this will be suitable with NS3 as a mobility trace file for our next step at next chapter. Furthermore, the command for visualization in SUMO which is:

```
$sumo-gui -n nottinghamcenter.net.xml -r nottinghamcenter.route.xml
```

3.3.2 The characteristics of PRAWMA Scheme

PRAWMA prediction scheme divided into three parts which are polynomial regression-adjusting range and weight moving average, the most significant being adjusting range. Because the part of adjusting range enables the model that owns the capabilities of self-judgement and screening during the process of operation, and it is an innovation of the prediction model as increase correct rate of prediction results. In this section, we demonstrate the importance of the adjusting range on behalf of prediction results accuracy by comparing the predicted results with adjusting range and without adjusting range. We record in one-hour continuous running output by second from SUMO with 500 nodes in the scenario of Nottingham centre as described at previous section, and set time periods to 5s, 10s, 20s and 30s, and set weights to 0.7/0.2/0.1 on account of relatively stable for different time periods. The performance of predicted results used correlation Similarity and Cosine Similarity for trend prediction and MRE and MRE remove cases (MRErc) for value prediction which is compared to simulation observation data from SUMO. The Fig. 3.11, Fig. 3.12, Fig. 3.13 and Fig. 3.14 show the predicted values of the prediction model with the adjusting range or not against to the observation data which record by SUMO's output for different time periods.

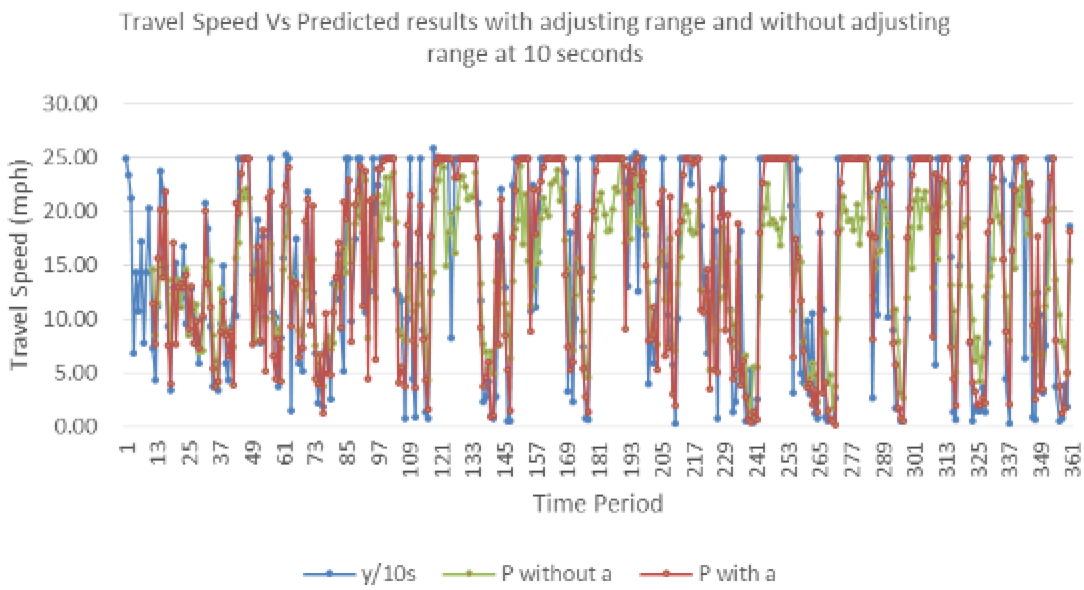
In the Fig. 3.11, the blue curve (y/5) means the observation data by 5 seconds time period; the green curve (P without a) means the predicted results that the prediction model does not include the adjusting range part; the red curve (P with a) means that the predicted results come from the PRAWMA scheme which is included the adjusting range part.



	Correlation Similarity	*Cosine Similarity*	*MRE*	*MRErc*
P without a	0.863	0.960	327.35%	22.81%
P with a	0.945	0.984	66.87%	15.83%

Fig. 3.11 Observation data Vs Predicted results with adjusting range and without adjusting range at 5 seconds

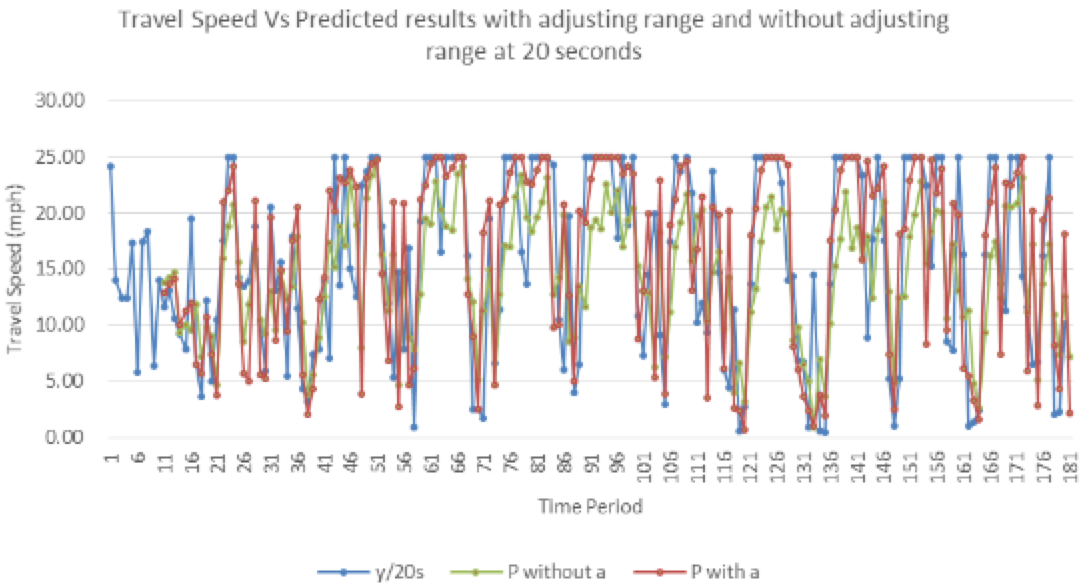
The results show that all performance evaluations of "P with a" are better than "P without a". As the graph shown, the red curve is more fitting to the observation data (blue) than green curve. This green curve will be an obvious fluctuation when the speed changes at the peak point with increasing or decreasing.



	Correlation Similarity	*Cosine Similarity*	*MRE*	*MRErc*
P without a	0.804	0.951	135.87%	25.98%
P with a	0.857	0.964	60.73%	22.72%

Fig. 3.12 Observation data Vs Predicted results with adjusting range and without adjusting range at 10 seconds

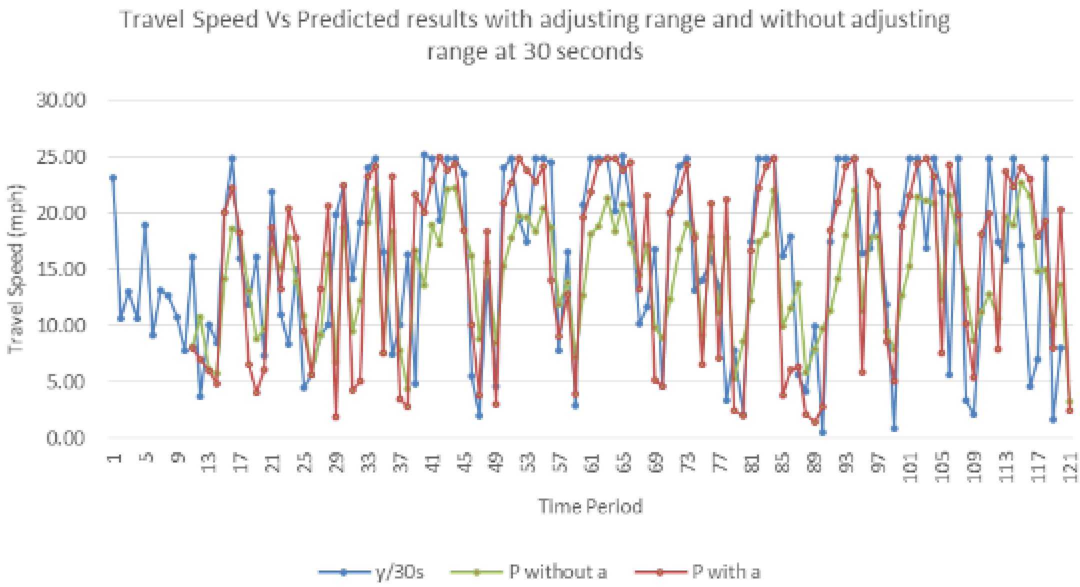
The situations of Fig. 3.12 is similar to the Fig. 3.11, that state the performance evaluation of “P with a” is superior to “P without a”. And because the time period is increased and simulation time does not change so that there are significantly fewer points in the graph, thus we can more clearly observe the fluctuating positions and the discrepancy of green curve nearby the peak. This discrepancy is caused by the error of predicted results.



	Correlation Similarity	*Cosine Similarity*	*MRE*	*MRErc*
P without a	0.654	0.935	78.39%	31.50%
P with a	0.694	0.936	61.57%	28.06%

Fig. 3.13 Observation data Vs Predicted results with adjusting range and without adjusting range at 20 seconds

With the time period increases, the correlation similarity of both results start falling, and “P with a” is still better than “P without a” in the Fig. 3.13. But cosine similarity of “P without a” is higher than “P with a”, and they still maintained over 90%. This shows that two curves have significant similarity at angles and trends. The error with fluctuating positions still around at peak for the green one. However the red curve begins to appear the big errors with some values, such as the time period of 42nd, 48th, 113th and 154th in the Fig. 3.13.



	Correlation Similarity	*Cosine Similarity*	*MRE*	*MRErc*
P without a	0.550	0.927	85.89%	29.40%
P with a	0.610	0.926	63.46%	30.16%

Fig. 3.14 Observation data Vs Predicted results with adjusting range and without adjusting range at 30 seconds

When a further increase in the time period to 30 seconds in the Fig. 3.14, the correlation similarity and cosine similarity of two prediction models continue to decline as same as the Fig. 3.13. The predicted results of green curve appear to more conservative and average than the red one. Because the adjusting range casts off the unreasonable results of the polynomial regression part. The graph and MRE value of "P with a" show that there are many irregular results at the bottom, it means that the predicted values are generally lower than the observation data. Three possible reasons can account for this happen, one is that the inappropriate weights might need to change, or the adjusting range is aggressive when the average speed fluctuates, or only 110 predicted results are less than previous three.

These experiments run in the same environments in this section to introduce the adjusting range to greatly improve the predicted accuracy in general. As expected, the correlation similarity in "P with a" is higher than "P without a" both comparing with observation data, especially the MRE in the first two figures. However, the correlation similarity has decreased an average of 5%-10% with the time period increasing. The cosine similarity has maintained a high level with a small difference of two models ("with a" and "without a"), that shows the advantage in the trend prediction for the model. In next section, we will use the PRAWMA prediction scheme to compare with ARIMA model, and give their performance evaluation based on the experimental outputs.

3.4 Comparison with the existing prediction model

The PRAWMA prediction scheme can be regarded as a similar model to ARIMA model (in section 2.4.2) in the mathematics. We intend to use ARIMA model to compare our PRAWMA scheme in the PPM in this section. Therefore, this section will present and compare their performances in theories and simulation experiments in the PPM for the prediction of average travel speed.

3.4.1 Theoretic Discussion

The ARIMA model is a generalization of ARMA model in time series analysis. It contains an explicit statistical model for handling the irregular portion of time series and allows the irregular portion might autocorrelation, also do not directly consider the changing of other relevant random variables.

The ARIMA model assumes that the time series is stationary. The time series Y(t) is taken from a random process when the features of this random process do not change with time, that the process is stationary; if the random features of this random process change with time, claimed that the process is non-stationary. Through mean, variance and covariance are correlated with time, thereby determining whether the random process is a stationary stochastic process. A stationary time series refers that mean, variance and covariance of the time series have been obtained looking forward to the same as the samples available to the future. And the fitting curve obtained via the time series of samples can still along the existing states inertia continue over a period of the next. Therefore, if there is a non-stationary time series, the first step need to difference for the time series until getting a stationary time series, so that the ARIMA(p, d, q) model will be used, where d is the order of difference (section 2.4.2). The basic prediction program of ARIMA model has six steps:

- According to the scatter plot of time series, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) to identify stability of the time series.

- The non-stationary time series conducts smoothing process until the ACF and PACF are non-significant not to zero.
- The corresponding time series model is established based on the identified features of the time series. After smoothing process, if the PACF is truncation (the parameter is or close to zero) and the ACF is trailing (the parameter is not zero) so that the AR model is established; if the PACF of stationary time series is trailing and the ACF is truncation, the MA model can be concluded; if the PACF and ACF are both trailing, then the time series should be used ARMA model.
- The parameters are estimated and verified whether statistical significance.
- The hypothesis of testing and diagnosis whether the residuals is white noise.
- Using the tested model to analyse and predict.

The contrast of PRAWMA scheme and ARIMA model is simplified at the initial stage. The PRAWMA scheme does not need identification of PACF and ACF, and smoothing process before modelling. Because the polynomial regression and adjusting range will not make much deviation for the predictions regardless of the stationary of the time series. Since the model is based on the wireless network communication between the moving vehicles to predict the average travel speed of one road section, the samples are instantaneous and can be kept continuously updated. In this case, the polynomial regression and adjusting range can dynamically determine the predicted values in each prediction period.

In PRAWMA scheme, the weights are added into the moving average part to ensure that important information is considered to modelling while the predicted results are produced when the polynomial regression part fails in prediction. Moreover, the number of samples of PRAWMA scheme requires being less than ARIMA model. ARIMA model can apply to many scientific fields via years of research, however, the PRAWMA scheme is only designed for the average travel speed and not apply to other fields yet. So the ARIMA model needs to depend on the predicted objective to select the number of samples, but the PRAWMA scheme needs at least 10 samples by our design. We summarise and compare some features between ARIMA model and PRAWMA at Table 3.7.

Table 3.7 Comparison of ARIMA model and PRAWMA scheme

	ARIMA model	PRAWMA scheme
Time series	Yes	Yes
Identification of PACF and ACF	Yes	No
Smoothing process	Yes	No
Screening mechanism	No	Yes
Composition	AutoRegressive + Moving Average	Polynomial Regression + Adjusting range + Weight Moving Average
Prediction periods	One or multi-periods	One period
Number of samples	30 samples	At least 10 samples
Evaluation	Comparison with observation data	Comparison with observation data

3.4.2 Experimental Discussion

ARIMA model is a very widely used prediction model in the field of researches. We tend to use it to compare with PRAWMA scheme in the PPM via experiments in this section. The experimental environments will be same as the section 3.3.1 and 3.3.2, the simulation parameters are shown in Table 3.6. We still record in one-hour continuous running output by second from SUMO with 1000 random movement nodes in the scenario of Nottingham centre as described at previous section, and set time periods to 5s, 10s, 20s and 30s, and set weights to 0.7/0.2/0.1. The number of vehicles is increased and other parameters are same as Table 3.6 in this section. In order to eliminate the non-stationarity, the original data is usually taken as a natural logarithm or differencing. Therefore, we apply a natural logarithm with original data and compare the predicted results with ARIMA, ARIMA model with log-transformed series (ARIMA-log series), PRAWMA and PRAWMA scheme with log-transformed series (PRAWMA-log series) in this section.

Table 3.8 Output ARIMA

Included observations: 10

Model	LogL	AIC*	BIC	HQ
(4,0)	-28.709623	3.470962	3.769682	3.529275
(1,0)	-32.555636	3.555564	3.704923	3.584720
(4,4)	-25.679411	3.567941	4.065807	3.665130
(0,2)	-31.723823	3.572382	3.771529	3.611258
(1,1)	-31.780070	3.578007	3.777153	3.616882
(0,1)	-32.965542	3.596554	3.745914	3.625711

Table 3.8a

Table 3.9 Output with ARIMA-log series

Included observations: 10

Model	LogL	AIC*	BIC	HQ
(0,1)	-7.270175	1.027018	1.176377	1.056174
(1,0)	-7.665166	1.066517	1.215876	1.095673
(1,1)	-7.051093	1.105109	1.304256	1.143985
(0,2)	-7.071420	1.107142	1.306288	1.146017
(4,0)	-5.402782	1.140278	1.438998	1.198591
(2,0)	-7.577631	1.157763	1.356910	1.196639

Table 3.9a

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	20.59565	7.958627	2.587839	0.0608
AR(1)	1.057666	0.487944	2.167596	0.0961
AR(2)	-0.584859	0.719193	-0.813215	0.4617
AR(3)	0.897657	0.703614	1.275781	0.2711
AR(4)	-0.823653	0.548688	-1.501132	0.2077
SIGMASQ	10.40030	14.87910	0.698987	0.5231

R-squared	0.882242	Mean dependent var	18.55400
Adjusted R-squared	0.735043	S.D. dependent var	9.906163
S.E. of regression	5.099092	Akaike info criterion	6.941925
Sum squared resid	104.0030	Schwarz criterion	7.123476
Log likelihood	-28.70962	Hannan-Quinn criter.	6.742763
F-statistic	5.993565	Durbin-Watson stat	1.994889
Prob(F-statistic)	0.053684		

Inverted AR Roots	.88-.40i	.88+.40i	-.35+.87i	-.35-.87i

Table 3.8b

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.721333	0.370013	7.354705	0.0002
MA(1)	0.999989	46904.64	2.13E-05	1.0000
SIGMASQ	0.197183	676.4447	0.000291	0.9998

R-squared	0.551244	Mean dependent var	2.740000
Adjusted R-squared	0.423028	S.D. dependent var	0.698729
S.E. of regression	0.530745	Akaike info criterion	2.054035
Sum squared resid	1.971834	Schwarz criterion	2.144811
Log likelihood	-7.270175	Hannan-Quinn criter.	1.954455
F-statistic	4.299337	Durbin-Watson stat	1.296652
Prob(F-statistic)	0.060539		

Inverted MA Roots	-1.00

Table 3.9b

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	19.05122	7.122334	2.674856	0.0318
AR(1)	0.781820	0.284988	2.743341	0.0288
SIGMASQ	35.82945	17.03577	2.103190	0.0735

R-squared	0.594317	Mean dependent var	18.55400
Adjusted R-squared	0.478408	S.D. dependent var	9.906163
S.E. of regression	7.154365	Akaike info criterion	7.111127
Sum squared resid	358.2945	Schwarz criterion	7.201903
Log likelihood	-32.55564	Hannan-Quinn criter.	7.011547
F-statistic	5.127428	Durbin-Watson stat	1.653377
Prob(F-statistic)	0.042526		

Inverted AR Roots	.78

Table 3.8c

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.762824	0.552942	4.996592	0.0016
AR(1)	0.629468	0.268840	2.341421	0.0517
SIGMASQ	0.257870	0.171440	1.504142	0.1763

R-squared	0.413132	Mean dependent var	2.740000
Adjusted R-squared	0.245455	S.D. dependent var	0.698729
S.E. of regression	0.606948	Akaike info criterion	2.133033
Sum squared resid	2.578699	Schwarz criterion	2.223809
Log likelihood	-7.665166	Hannan-Quinn criter.	2.033453
F-statistic	2.463860	Durbin-Watson stat	1.669047
Prob(F-statistic)	0.154843		

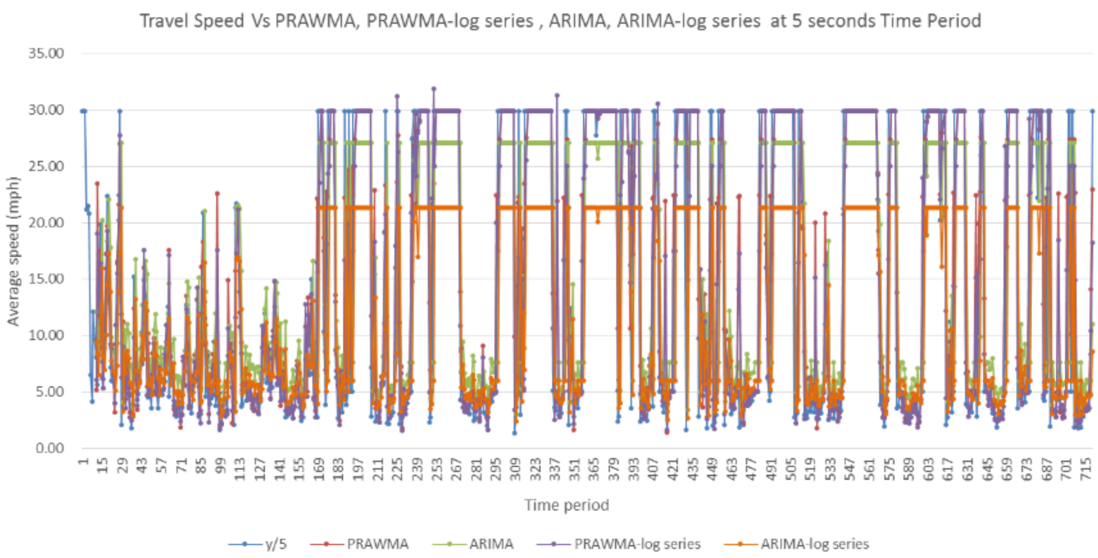
Inverted AR Roots	.63

Table 3.9c

We include an example of how to determine the orders of ARIMA and build the ARIMA model, and we will not explore and discuss it in the future sections. The Table 3.8 and Table 3.9 gives the outputs of ARIMA and ARIMA-log series by the same observation average speed which is 1^{st} time period to 10^{th} time period. Because our data sample size is small, it is not easy to carry out the test of stationarity, where nature logarithm or differencing step on the observation data can be applied to eliminate the non-stationarity. Therefore, we compare directly two results from ARIMA and ARIMA-log series.

From Table 3.8a and Table 3.9a, we try ARIMA (4,0,0) and ARIMA-log series (0,0,1) with smallest AIC value by Eviews. But the outputs of these two models show at Table 3.8b and Table 3.9b. But the coefficients between them are insignificance, which means that the models are not suitable for fitting the ARIMA (4,0,0) and ARIMA-log series (0,0,1). Thus, we consider the ARIMA(1,0,0) and ARIMA-log series (1,0,0) with second smallest of AIC value, where the results show at Table 3.8c and Table 3.9c. The significance level of both are less or close to 5%, and the absolute t-statistic value of ARIMA(1,0,0) is bigger than t(df), and the t-value of ARIMA-log series (1,0,0) close to the t(df). The t(df) value comes from t-test table, in this case the t(df) is 2.306 at 0.05 significance level with 8 degrees of freedom. Therefore, we can reject the null hypothesis, which means there is both fitting AR(1) model for the series. The D-W stat is over 1.5 and tends to 2, which means there is no obvious autocorrelation for the residual. By the consideration of various test results and simplicity of the concept, the model of ARIMA(1,0,0) is better than ARIMA-log series (1,0,0) for 1^{st} to 10^{th} time period. For more information on ARIMA modelling and analysis, please refer to other related literature, we obtain the results directly through the software, such as Eviews.

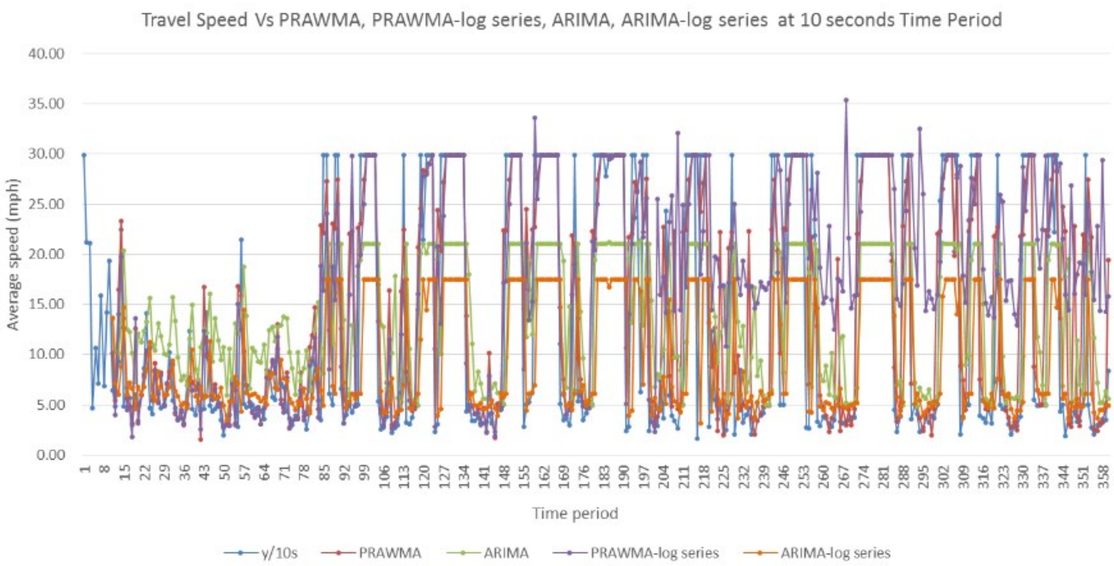
In the Fig. 3.15 with 5 seconds time period, the PRAWMA fitting curve reveal better correlation similarity and MRE than the results in the Fig. 3.11 with 500 nodes. However, the MRErc decreases about 3%.



	Correlation Similarity	*Cosine Similarity*	*MRE*	*MREerc*
PRAWMA	0.918	0.962	54.60%	17.42%
ARIMA	0.759	0.908	90.38%	32.69%
PRAWMA-log series	0.913	0.961	42.88%	17.14%
ARIMA-log series	0.759	0.908	70.52%	33.03%

Fig. 3.15 Travel speed of observation vs 4 predicted results at 5s Time period

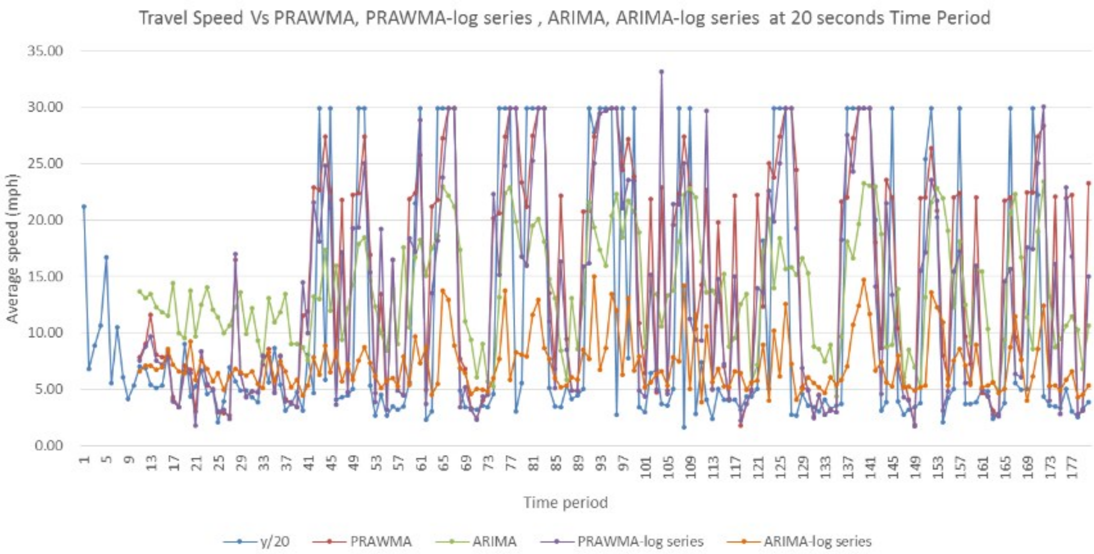
The model of ARIMA (1, 0, 0) and ARIMA-log series (1, 0, 0) have been fit and the predicted equation for each 10 time periods which are: $Y_t = \varepsilon \ + \varphi_1 Y_{t-1}$. As the Fig. 3.15 shown, the predicted results of ARIMA model lags behind the PRAWMA scheme in all aspects of performance evaluations. While the continuous peak position means that there are no vehicles passing on our objective road section during those time periods. So the prediction model uses the default maximum speed (30mph). But the predicted value of ARIMA model significantly lower than expected, although it gives good results for the overall trend. The results of PRAWMA and PRAWMA-log series have a better effect than ARIMA and ARIMA-log series in general.



	Correlation Similarity	*Cosine Similarity*	*MRE*	*MRErc*
PRAWMA	0.828	0.924	84.91%	22.34%
ARIMA	0.604	0.838	134.64%	38.48%
PRAWMA-log series	0.739	0.853	165.85%	23.79%
ARIMA-log series	0.622	0.847	75.35%	40.80%

Fig. 3.16 Travel speed of observation vs 4 predicted results at 10s Time period

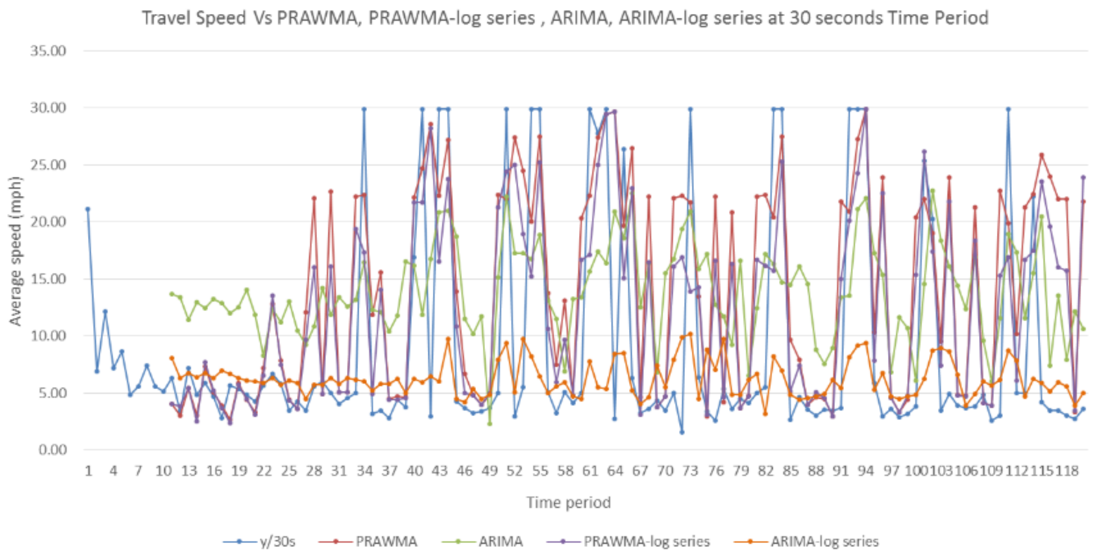
The models of ARIMA (1, 0, 0) and ARIMA-log series (1, 0, 0) are built in the Fig. 3.16 with only AR model and the predicted equations fitted are: $Y_t = \varepsilon \ \ + \varphi_1 Y_{t-1}$. The situations are similar to the Fig. 3.15 that predicted results of ARIMA model are significantly below the results of PRAWMA scheme and still smooth curve at peak. The predicted results of ARIMA model are somewhere approximately between 5 mph and 21 mph, the results of ARIMA-log series are approximately between 5mph and 17mph. The predictions of low speed have not achieved the results we anticipated, such as 23th to 40st time period and 56th to 78th time period. The results of both models present decrease with the time period increasing. The results of PRAWMA is better than PRAWMA-log series with four evaluation indicators, and their predicted effect are still better than ARIMA and ARIMA-log series. The results of PRAWMA scheme do not have that much different than Fig. 3.12 with 500 nodes in four evaluation indicators.



	Correlation Similarity	*Cosine Similarity*	*MRE*	*MRErc*
PRAWMA	0.642	0.837	113.52%	25.96%
ARIMA	0.542	0.789	167.51%	42.67%
PRAWMA-log series	0.626	0.830	120.70%	26.20%
ARIMA-log series	0.462	0.769	70.27%	49.23%

Fig. 3.17 Travel speed of observation vs 4 predicted results at 20s Time period

With increasing of the length of the time period to 20 seconds, the results of default maximum speed decrease at the peak in the Fig. 3.17. The ARIMA (1, 0, 0) and ARIMA-log series (1, 0, 0) have been built as same as previous cases. The ARIMA and ARIMA-log series in sharp contrasts to these indicators of the PRAWMA and PRAWMA-log series, where the results of the first two model are still lower than PRAWMA and PRAWMA-log series. The length of time period increases, and the total simulation time does not change that makes easier to observe the graph. The predicted results of ARIMA and ARIMA-log series present fluctuation when the observation data is smoothing at the peak which is maximum speed, such as 76th to 84th time period, 91st to 99th time period and 136th to 141st. Otherwise, the predictions of low speed have not achieved the results we anticipated with ARIMA model, such as 13th to 39th time period. But the results of PRAWMA and PRAWMA-log series give a better performance during these time periods. Despite all this, the results of PRAWMA scheme have some unexpected results with big errors at the time period of 47th, 59th and 103th and so on in the Fig. 3.17.



	Correlation Similarity	*Cosine Similarity*	*MRE*	*MRErc*
PRAWMA	0.595	0.795	123.99%	26.09%
ARIMA	0.493	0.753	190.74%	47.15%
PRAWMA-log series	0.538	0.784	138.39%	29.08%
ARIMA-log series	0.455	0.739	58.12%	45.79%

Fig. 3.18 Travel speed of observation vs 4 predicted results at 30s Time period

The models of ARIMA (1, 0, 0) and ARIMA-log series (1, 0, 0) are provided in the Fig. 3.18. The results of PRAWMA and PRAWMA-log series are still better than ARIMA and ARIMA-log series, expect the MRErc with a big difference. As the figure shown, the results of ARIMA model with log-transformed series is better than ARIMA in the lower speed with the observation data relatively smooth changing. After that, the results of ARIMA remain fluctuations along with the changing of observation data but the errors of ARIMA and ARIMA-log series are still greater than PRAWMA and PRAWMA-log series in overall. Moreover, the results of PRAWMA and PRAWMA-log series are consistent with the observation data when the observation data trend to the peak or valley in the Fig. 3.18, for example when the time period is 38th to 53th and 81nd to 93th.

This section discussed and compared the difference between ARIMA, ARIMA-log series, PRAWMA and PRAWMA-log series via a series of experiments. The figures and performance evaluations showed that the predicted results of PRAWMA and PRAWMA-log series are more efficiency than ARIMA and ARIMA-log series. However, the performance evaluations of both models tended to fall in response to increasing the length of time period. Based on our small samples modelling for ARIMA, the AR gives 1 and MA give 0 in all order estimations with the smallest AIC (Akaike information criterion) by statistical software, therefore the results of ARIMA and ARIMA-log series show a low level of smooth fluctuations to observation data. The other hands, the results of PRAWMA and PRAWMA-log series are not much difference between them, but they can be adapted to the dynamic prediction for small samples. We apply natural logarithm to eliminate the non-stationarity, but the non-stationarity of the small samples is difficult to determine in general, at the same time, the conversion of logarithm with original data cause susceptible to results distortion. In the case of a small gap between PRAWMA and PRAWMA-log series, we do not intend to logarithm with the original data of PPM in the subsequent example and experiments.

3.5 Summary

This chapter we introduced the concepts and classification of the related prediction model in section 3.1. In section 3.2, we came up with one novel PPM prediction model for traffic condition, especially average travel speed and a scheme as one part of PPM which call Polynomial Regression Adjusting Weight Moving Average (PRAWMA) scheme. Meanwhile, we also explained the modelling principle, model's algorithm, processes of the model and parameter verification. We initially introduced the experimental process how to generate the mobility pattern and emphasised the characteristics of the model for the adjusting range part in section 3.3. In section 3.4, we compared the existing model by ARIMA to our PRAWMA scheme in theories and experiments. The results show that our PRAWMA and PRAWMA-log series in PPM are better than ARIMA and ARIMA-log series for average travel speed prediction in the short time period.

Introducing the adjusting range improves the prediction accuracy of the PPM. However, this approach in this chapter is only based on all known data cases, and assuming all data received immediately, it means all prediction experiments were running without any wireless network. However, in the real case, all data received from others that should spend time. How this influence input data and the prediction results will be in the experiment for the following chapter. In next chapter, we will introduce the PPM-C2C framework in the mobile wireless network, and compare with different routing protocols to test the performance of the network.

4 Chapter 4

Designing a Message Delivery Algorithm in VANET for a Real City Scenario Mobility Model

4.1 Introduction

Based on the descriptions and results presented in Chapter 3, we propose that the necessary and required data for the PPM should be applied to a wireless network via C2C communication to deliver messages. The C2C communication allows non-infrastructure to be deployed in modern wireless mobile networks and can face and cope with comparatively efficient tasks and data disseminations. It challenges other wireless mobile networks through the features of the dynamic and flexible organization. In section 2.2, we introduced some reviews for message delivery with C2C communication in VANET, such as the literature (Ghosekar, Katkar and Ghorpade 2010, Wang 2013). And some solutions are designed by a broadcasting mechanism with MANET, such as (Williams and Camp 2002, Tseng, et al. 2002). The broadcasting mechanism is a proper technology fit for C2C communication that all nodes in the network through one-hop and multi-hops to send and forward the messages. The messages might exchange from one node to another node using one-time delivery or multi-times to reach the destination without any controls in the middle.

In this chapter, a message delivery algorithm with C2C direct communication is introduced. It offers real-time vehicle information for all vehicles based on sharing the data of the individual vehicle. It aims to collect the data by using the data derived from the application to support the prediction framework for traffic condition prediction. Many automobile manufacturers have started to equip their cars with Wi-Fi technologies, for example, Mercedes, BMW and Ford. Ultimately, the behaviour of a pervasive traffic simulation and the traffic condition prediction model in the ad-hoc network will be built and investigated. The prediction model assumes that Wi-Fi is embedded in each vehicle. When the vehicle receives the traffic conditions from others, the algorithm will analyse the data and forecast the traffic condition of objective road section using the prediction model for the future time.

This chapter is organised as: the approaches of dissemination in VANET will be introduced in section 4.1, and how is the multi-hop transmission by C2C communications in section 4.2. In section 4.3, we will define the message type, format and content in this thesis, also design a message delivery scheme for data transmission based on the C2C communication to transfer our data packet which includes the PPM-C2C framework. Meanwhile, we will compare four existing routing protocols to choose two of them to apply in the following experiments in section 4.4.

4.2 Traffic Message Delivery in VANET

The traffic information system is controlled by the urban traffic manager to provide the traffic conditions to drivers or vehicular applications at present. The traffic information system also needs time to collect and centralise the condition, in this case the information is easily delayed. While there is a lack of information coverage due to cost consumptions and limited network resources, such as the sensor network, communication infrastructure or electronic board need to build. (Wischoff, et al. 2003) In contrast, C2C communication technology has relatively small implementation cost, short delays, self-organizations and reused capacities.

The vehicular communication can provide the traffic safety and traffic information to improve the traffic conditions. The traffic safety application encompasses such as emergency event alarming and avoidance of collision (Cheng, Shan and Zhuang 2011). The traffic information application is more widely used in urban transportation, for example, traffic volume collection, accident assistance and sharing of parking information, weather information and jam information in the real-time (Wischoff, et al. 2003, Nzouonta, et al. 2009). However, the different traffic information applications have different demands (e.g. delay, frequency, coverage and receiving rate) for information dissemination with different dissemination approaches. Our target needs to study the influence of these different dissemination approaches and how to affect the prediction model and the predicted results.

4.2.1 Dissemination Approach

The vehicles can provide information about the status of themselves via sensors or communication device, for example the speed, position, acceleration. The information represents the status of certain locations at a certain moment. This thesis provides the traffic condition prediction of the road sections for the drivers. Because they would be interested in how fast they can drive or what’s the situation of their destination. There are four communication types in VANET as described in the Fig. 2.3 and two main types of communication can be applied for traffic information dissemination:

- **Road-Side Unit**: the vehicles need to send their information to the RSU at first, then a central control system collects, manages and forwards the information to other users through a cellular system (Olariu and Weigle 2009). The advantage of this approach is a powerful process capacity with far communication distance. The cellular system needs base station or access point to achieve communications. However, even if the establishment of more base stations, the congestion issues still exist due to the limited frequency resources. In addition, the deployment of a large number of base stations will lead to huge cost while the utilisation of base station is very low.
- **Ad-hoc network**: vehicular ad-hoc network (VANET) is a dedicated network for the traffic based on the multi-hops broadcasting. Its advantages are a good performance on throughput, latency, saver power and lower cost; rather, the disadvantage is not suitable for the wide area network (Nzouonta-Domgang 2009). This thesis will use this approach to deliver messages which is an enhanced version of the wireless technology suitable for VANET.

The routine information is periodically broadcasted through the VANET, such as speed, time and position. Most vehicles send those messages to share the current traffic conditions. Many types of research are based on the information dissemination for vehicle, such as the more common communication technologies for VANET as shown in (Hartenstein and Laberteaux 2008); and in (Wu, et al. 2004), they proposed an algorithm to forward with the information dissemination of vehicle mobility, and it is based on trajectory and geography.

4.2.2 Multi-hop Transmission

Ad-hoc domain supports multi-hop transmissions for more applications in VANET. However, there are some technical difficulties in multi-hops, such as reliability, connectivity, delay, the range of transmission and data security. VANET is facing issues of scalability with a larger number of nodes; connectivity with network disconnected; high mobility that nodes move fast and topologies change fast. Moreover, the C2C communication is not dependent on any fixed infrastructure and management centre. It groups by mobile hosts or nodes, through the cooperation between nodes and self-organization of nodes to connect and transfer data message. The following figure shows how is the ad-hoc network working and how the multi-hops work, finally to cover all nodes in the network step by step. Four sub-figures describe that over time the movement of cars, the change in the number of cars and increasing range of communication.

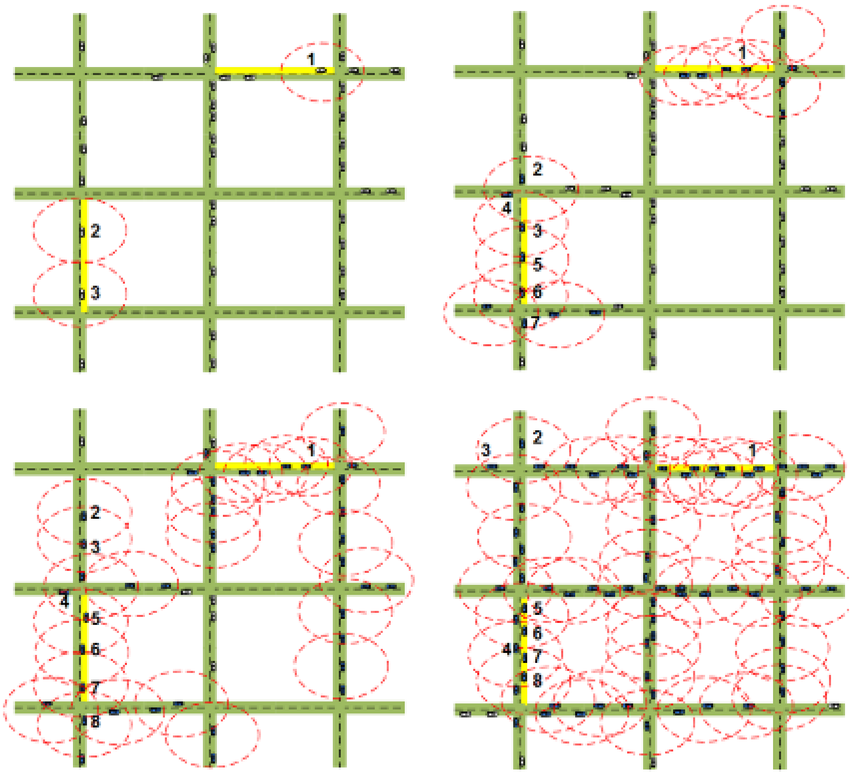


Fig.4.1 Demonstration Steps of C2C Communication

Car to Car communication is derived from the standard IEEE 802.11. Once two or more vehicles are in wireless communication range, they connect automatically and build an ad-hoc network through the routing protocols. Every vehicle can be a sender, receiver and router, also allows delivery messages over multi-hop to further vehicles. To ensure timely and accurate transmission of traffic data, the role of a routing protocol in ad-hoc networks that discoveries and maintains the connectivity between vehicles. With AODV, for example, one vehicle will stores the route information into its routing table when it receives an RREQ and discard the repeated RREQ based on the sequence number, and determine any response from other vehicles. When RREQ is forwarded, the vehicle will mark its upstream vehicle ID into the routing table to be able to build a reverse route from the destination node to the source vehicle. A new route discovery algorithm will be generated when source node moved. If other vehicles move, neighbour vehicles will discover a failed link and send RERR message to upstream node thus spread to the source node; then the source node will renew the route discovery. In the Fig.4.1, the vehicles on the marked road sections (two yellow roads are called horizontal road and vertical road) send messages, including their information such as speed, position, or situation of the vehicles in this demonstration.

In the upper left figure, there are many vehicles running around the horizontal road. Vehicle 1 sends a fault message as a warning message to the around vehicles within the network coverage range (red) through multi-hop transmissions. Vehicle 2 and vehicle 3 travel on the vertical road and try to connect to each other, but they are out of the radio communication range, and the connection attempt failed at the first step.

In the upper right figure, the vehicles around of vehicle 1 start to spread the warning message to other vehicles in their communication range. Moreover, vehicle 2 and vehicle 3 can build a connection, thus send their messages to each other. At the same time, vehicle 3, 5, 6 also build a connection with each other and send their messages to the neighbour vehicles such as vehicle 7 via the ad-hoc network. These messages include time, direction, position and speed as a shared message when they travel on the marked vertical road.

In the lower left figure, over time, the communication coverage gradually spread to the distant locations for the warning message from vehicle 1, and the vehicle 2-8 also form a wireless ad-hoc network to send their shared message to the vehicles on the other neighbour road sections.

In the lower right figure, a large-scale wireless network is formed. All vehicles might receive the warning message from vehicle 1 or the shared messages from the vehicles on the marked vertical road in the communication coverage on this map. When a certain vehicle leaves the communication range, other vehicles will re-establish the wireless network and connect to each other.

If a vehicle wants to pass those road sections, the driver will know the situation and take action before getting there. The information of the marked road section is sent to the neighbour vehicles by unicast and multicast protocols. For example, the vehicles can periodically broadcast "hello" message to maintain the information of their neighbour vehicles, and send the information where the road they travel on. When other vehicles receive the information and discovery neighbour vehicles, the information will be forwarded to their neighbour vehicles. (Williams and Camp 2002)

4.3 Traffic Message

4.3.1 Message Type

The types of message depend on the communication purposes, meanwhile, it decides sizes of the data. There are five classifications for communication purposes: public safety applications; traffic management applications; traffic coordination and assistance applications; traveller information support; comfort applications (Sichitiu and Kihl 2008). In order to simplify the classifications of the message, Y.Y. Li advised using two types of messages which are warning-based (WB) message and shared-based (SB) message (Li 2013).

The features of WB messages is a faster generation, smaller overhead, more brief and accurate than SB messages (Li 2013). The WB message mainly includes traffic event and accident for public traffic safety, coordination and assistance. However, SB message focuses on the traffic information support, control and management, even comfort and entertainment. It is based on the information sharing to satisfy user demands. The literature (GAMATI 2013) provided something very specific for road conditions in details. And those road conditions include message context that is all about time; vehicle context that is all information about a vehicle; information context that is about traveler information update; communication context that is about transmission environment; road context that describes a number of cars or road lanes; the routing context that includes a number of hops and delay. All those road condition identifications are used to multiple cars that need to exchange message, or are used to a single car that process by individual car.

In this thesis, we must clear transmission purposes at first, then select the type of message and determine the message contents. Our purpose is how to predict the average speed of road section so that the type of message should be based on the information sharing.

4.3.2 Message Format

In the WAVE standard (IEEE Vehicular Technology Society, 2010), two type of message formats were defined in vehicular communication environments. One is the "hello" message by periodically exchanging messages and keeping track of neighbour links. The node determines the connectivity of the link by the "hello" messages. As a certain time, the node will check whether it receives a message, if it has not, the node will broadcast a "hello'' message. Therefore, each node will maintain these links via "hello" message be called WAVE Service Advertisement (WSA). Another one is WAVE short message (WSM). WSM specifies the channel numbers, data rate, transmit power for message dissemination. The minimum overhead of one WSM frame header is only 8 octets and the maximum size is 1400 octets. In general, WSM is sent to the CHH or SCH after 802.11p data frame is encapsulated. The frame format of WSM is:

Table 4.1 WSM format

WSM Header					
1	4	Variable	1	2	Variable
Version	PSID	Extension Fields	WSM Wave Element ID	Length	WSM Data (payload)

It contains the Version of protocol; the Provider Service Identifier (PSID) identifies the desired service and its length can be changed; the Extension Fields are optional part, a 1609.3 packet can have 0-n extension fields, however there currently are three extension fields namely of channel number, data rate and transmit power; WSM WAVE Element ID indicates the extension fields and the type of WSM data field contained; Length indicates the length of WSMA data field; WSM data is payload and default 128.

4.3.3 Message Encapsulation

When vehicles want to send a message, they have to encapsulate the packet at first, then broadcast to the neighbour and network. According to the frame format of WSM in Table 4.1, we set the Traffic Speed Message (TSM) is:

Table 4.2 Traffic Speed Message format frame

TSM Header							
1	1	1	1	1	1	1	1
Sender_id	Sender_MAC			Sender_position			
Sender_time	Sender_speed		Sender_direction		Receiver_id	Forwarder_id	Type

Sender_id/MAC: originator's id and MAC address, each vehicle has only one and not same to others.
Sender_time: timestamp when originator broadcast.
Sender_position: originator's position at sender-time.
Sender_direction: originator's direction at sender_time.
Sender_speed: originator's speed at sender_time.
Forwarder_id: current sender's id, it is a relay node.
Receiver_id: destination node's id.
Type: message type is sharing the messages.

The frame above is the broadcasting packet, when the packet is encapsulated, it will be sent out, and our prediction model will use some of those contents about sender's information.

4.3.4 Message Access

We need a container to store the messages when each vehicle received. Here is an assumption that each vehicle has a powerful container which can store enough information for the prediction model. And container has many tables for different road sections, such a table like Table 3.1 and Table 3.2 in the section 3.2.5.

The name or header of these tables is the street name with the direction which depends on the position parameters. So that each street name might have two tables for recording the messages if the street is two-ways street according to GPS device. Otherwise, one-way street only has one table. Our mobility pattern can simulate one-way or two-way situations because it is based on the OpenStreetMap (OSM).

New message will be recorded into the right table by positions and directions.

And the data is ordered by time series which is Sender_time in the previous section. Thus, the contents of the table include:

Position_direction: is table header in the container and decide the data into where. Time: is the Sender_time. Car_id: represents the originator's id, is an identification of each vehicle and not same to others. Speed: is the speed of Car_id at Sender_time.

4.3.5 Message Delivery Scheme in the PPM-C2C framework

Another most important thing is that the information table can help the vehicle to make the choice for forwarding the message or dropping the message. Because the prediction model is not interested in overdue data. In order to avoid data redundancy, save storage space and share the information, the old message needs to dropping and the new message needs to forward. There is the message delivery scheme for sender and receiver:

Table 4.3 Message Delivery Scheme

Sender	Receiver

Loop{
1. Sender encapsulate packet as TSM with speed generated in second
2. Sender broadcasts packet by IEEE 802.11p
}
	Loop{
	1. Add message to table via Sender_position and Sender_direction.
	2. Index Sender_id and Sender_Time from the table
	3. If (new message, valid) then
	4. Record in table
	5. Forward the message
	6. Else if (exist message, valid) then
	7. Forward the message
	8. Else(overdue message, invalid)
	9. Drop the message
	}

Each vehicle can be a sender or a receiver, one vehicle needs to broadcast its speed by each second when it is a sender; meanwhile, it also needs to collect and forward the message when it is a receiver. Each vehicle has a powerful container and many tables for different road sections. The prediction model also needs these tables to generate and process. Therefore, message contents should be identified at first when vehicle receive the messages. Thereby this scheme could process the data with the message details.

Message type can be divided into a valid message and invalid message. And the valid message also includes a new message and existing message. New Msg means the Car_id and time are not in the data table for one road section when a message incomes. Otherwise, the message is Existing Msg. They are both the valid message and will be forwarded and updated into the table. When an incoming message is checked that the time is too staleness with the current time, this message will be called Overdue Msg or Invalid message, then it will be dropped. This time gap depends on the period time setting on prediction model, it needs to meet the demands of the prediction model. For example, if the prediction model needs 10 minutes historical data with 1 minutes time period, the table will keep enough data for the prediction model, and drop the data over 10 minutes. The Fig.4.2 shows message contents and message behaviour as a diagram:

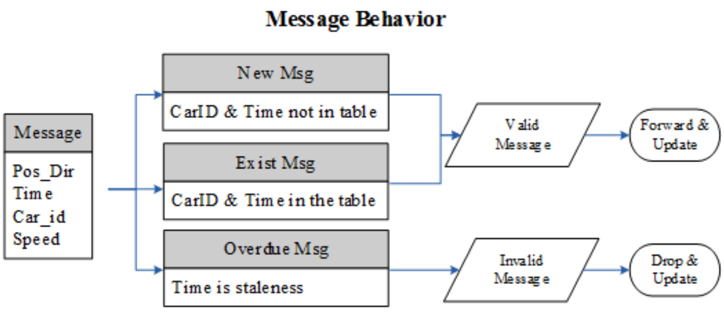


Fig.4.2 Message Contents & Behaviour

Each vehicle is an individual unit in the whole network, and its roles include generated data, making prediction, processing and broadcasting messages. The prediction and communication should be achieved through each individual car. The Fig.4.3 shows the proposed PPM-C2C framework and message behaviour in each vehicle as a constant task.

PPM-C2C framework includes an in-car pervasive prediction model based on the ad-hoc data (PPM) and traffic message delivery algorithm based on the C2C communication, where a PRAWMA prediction scheme is one part of PPM. The role of PPM is data application that includes data sorting with the data table and data computing with the PRAWMA prediction scheme. The data table can set the data sorting plan such as time period and valid of data, or increase the data category such as travel time or traffic density. The PRAWMA prediction scheme can be replaced by the existing prediction model such as ARIMA or KNN. The framework also allows that can use the different routing protocols for C2C communication such as ADOV and

OLSR.

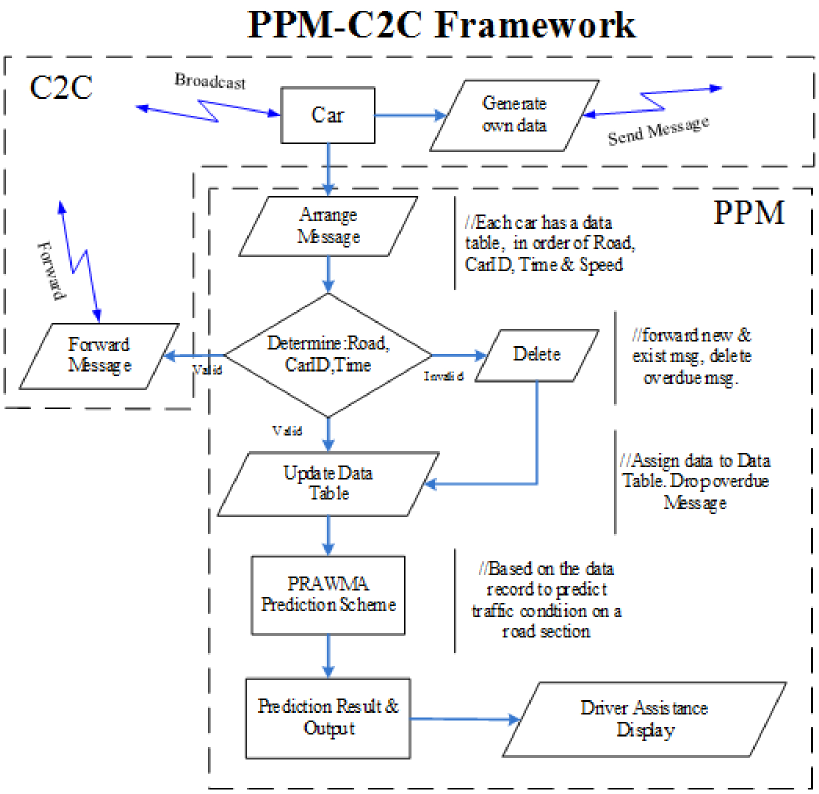


Fig.4.3 PPM-C2C Framework in each Car

As described above, each vehicle can be a sender or a receiver, the message of other vehicle is input; its own data is output message which is line 1 and line 2 in the Sender scheme in Table 4.3. At above Fig.4.3 and the message delivery scheme (Table 4.3), once the vehicle receives the message, the system will add and arrange the message to the data table which is in line 1 and line 2 in the Receiver scheme. The system needs to determine the input data at next step. This function is described in the Fig.4.2 for the message behaviour. The message is divided into three types which are new message, existing message and overdue message. The new message will be forwarded and added into the data table as a valid message which is in line 3-5 in the Receiver scheme; the existing message will be only forwarded to other vehicles which are in line 6-7; the overdue message will be dropped according to the Time and CarID comparison which is in line 8-9. Once the data table update is complete, the prediction model will be generated based on the data table to estimate the average speed of a road section.

4.4 Routing Protocol in VANET

Routing is looking for the path of data transfer from the source node to the destination node. Since the particularity of ad-hoc network, its routing protocol design is very different with the traditional fixed network (section 2.2.2). The reliable routing mechanism of VANET is mostly based on the MANET routing mechanism, researchers have proposed many VANET routing solutions, include location-based, probability-based, broadcasting-based (Nzouonta-Domgang 2009, Li and Peytchev 2010, Li 2013, GAMATI 2013) and so on. For example, a classic location-based routing protocol as GPSR (greedy perimeter stateless routing) (Karp and Kung 2000) algorithm uses geographic information to implement a routing algorithm, which uses a greedy algorithm to establish the route. When node S needs to forward data packets to node D, it selects the nearest node from node D as the next hop data packet in their neighbourhoods at first, and then transmits the data to it. This process is repeated until the packet reaches the destination node D or the best host. However GPSR does not be appropriate for city scenarios in VANET (Li and Wang 2007) due to the communications between nodes might influence by buildings or trees to restrict the greedy forwarding. Also, the mobility may induce packet forwarding to the wrong way leading higher delays. It is often a local optimum and not of the overall consideration while the threshold distance is not easy to obtain (Xu and Xia 2013).

Therefore, the existing routing mechanism is still facing many challenges. The contrast between the routing mechanisms of MANET and VANET needs to face challenges in the following areas: 1. VANET has more dynamic movement and the network consists of vehicles, the topology changes at any time due to the rapid movement of vehicles; 2. VANET has more opportunity for communication, and the connectivity is relatively good between vehicle-intensive areas that easy to implement reliable delivery of messages between vehicles likely point-to-point connectivity. However, the vehicle may not exist point-to-point connectivity in scarce regions. Therefore VANET is the presence of opportunistic routing connectivity; 3. The effective communication time of VANET is short and frequency; 4. The delivery of security message should have backwards transmissibility in VANET and vehicles involved in accidents need to inform all neighbours of it through the wireless broadcasting. Therefore, there is no universal routing to adapt the complex changing of the transportation system. Researchers can take advantage of VANET's feature and their demands to set up the routing protocols, such as vehicle mobility, connectivity, fixed infrastructure, location and dynamic opportunistic.

In this subsection, we will discuss a serval existing routing protocols (AODV, DSDV, OLSR, DSR) in VANET, and apply different routing protocols to transfer our messages, then investigate using different routing protocols how to influence the data collection and our prediction model.

4.4.1 Topology Based Routing Protocol

In the literature of (Li and Wang 2007, Xu and Xia 2013), they are divided into five kinds of routing protocols in VANET according to different characteristics and demands, such as ad-hoc routing, position-based, cluster-based, broadcast and geo-cast routing; or communication-based, movement prediction-based, probabilistic routing, fixed infrastructure-based and location-based. We classify the routes into two categories in this thesis: proactive routing, reactive routing.

Proactive routing

In unicast mode only has one sender and one receiver, or one source node and one destination node. Each node periodically probes the network topology, maintains the real-time routing table based on the current connectivity nodes to obtain the corresponding routing data when sending a message. The classic routing protocols have OLSR and DSDV.

This type of protocols needs to continuously try to update the routing table and adapted to low mobility environments. The proactive routing protocols need a large sum of data for routing maintenance, especially in a high dynamic movement networks. In this case, it consumes a lot of bandwidths and increases overhead when nodes change at any time. Therefore, in urban traffic communication environments, the issues of scalability, connectivity and high mobility restrict the proactive routing developments (Xu and Xia 2013).

Reactive routing

Each node establishes a temporary routing before sends data, thus, the overhead of maintaining routing table is eliminated. This type of protocols depends on flooding with query packet to find the path to the destination nodes. The classic routing protocols have DSR and AODV.

A node receives a packet and continues to broadcast to the neighbourhoods until the destination node received the message. In every node along the path, its address is added to the list of relay nodes carried in the packet. Obviously, this protocol is easy to implement and is a good choice when a smaller number of nodes, traffic notification messages. But for the unicast message, there is only one destination node, the flooding is no longer applicable. Especially, when a number of node broadcast message at the same time, resulting in frequent channel competitions and data collisions which are called broadcast storms (Li and Wang 2007, Xu and Xia 2013).

For the unicast message, the message can be divided into two types of the control packet and data packet to improve the efficiency of flooding. RREQ, RERR and RREP belong to the control packet. The source node broadcasts RREQ packet to all nodes, then receiving nodes detect the destination address in this RREQ packet. If they are not the destination node, then choose the path to send RREQ packet. After the route is determined, the packet can be passed along with this path, if there is an error occurs in the routing establishment so that they will send RERR packet.

In the multi-hop communication, the communication performance decreases rapidly with data and nodes increase. Flooding usually produces large amounts of duplicate packets or broadcast storms, but its performance is considerable when network topology dynamically changes slowly and traffic density is not high (Xu and Xia 2013).

4.4.2 Performance Analysis of Routing Protocols

Combined with the previous analysis (section 2.5 and section 4.4.1), we will test the performance of routing protocols in our Nottingham city centre scenario in this subsection comparison with the routing protocols of DSDV, DSR, AODV and OLSR.

The current of NS3 also supports these protocols with VANET by utilising the WiFi model with IEEE802.11p extensions. The parameters of the simulation are:

Table 4.4 Parameters of experiment

Parameter	Value
Simulator	ns-3.23
Simulation area	1200m x 1000m
Number of vehicles	50
Sink of nodes	5-25
Vehicle's speed	0 - 25 mph (0- 40 km/h)
Channel	Wireless
Network Interface	WirelessPhy
Mac Type	802.11p
WiFi-rate	2 Mb/s
Protocol	AODV, DSR, DSDV, OLSR
Nodes Distribution	SUMO-Mobility
Movement Pattern	Random
Communication type	C2C
Transmit power	7.5dBm
Packet size (bytes)	64,128,256,512,1024
Simulation time (second)	100

In the first experiment, we test the throughputs and number of packet received of the communication protocols which are ADOV, DSR, DSDV and OLSR. The mobility pattern that we use the same one as section 3.3.1. Throughput is that the ratio of the total data transmitted successfully from source node to destination node, and its units are bytes or bit/s. It is an important indicator for the testing of network performance, high throughput represents the efficient network and its formula in ns3 is (Ali and Ali 2009):

$$throughput = \frac{\text{number of packet send}*\text{Packet size}*8}{\text{total simluation time}} \qquad (4.4\text{-}1)$$

The total simulation time is 100 seconds and 50 nodes random walks in Nottingham map road network. In Fig.4.4a and Fig.4.4b, the packet size changes in 64, 128, 256, 512 and 1024 Bytes with 50 nodes and 25 sink nodes; in Fig.4.4c and Fig.4.4d, the number of sink nodes changes in 5, 10, 15, 20 and 25 with 50 nodes and 64 bytes packet; in Fig.4.4e and Fig.4.4f, the number of nodes changes in 10, 20, 30, 40 and 50 with 5, 15, 25, 25, 25 sink nodes and 64 bytes to observe the throughput and packet received changes, as shown at below:

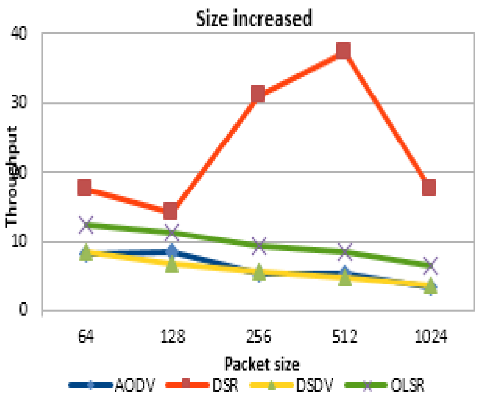


Fig.4.4a

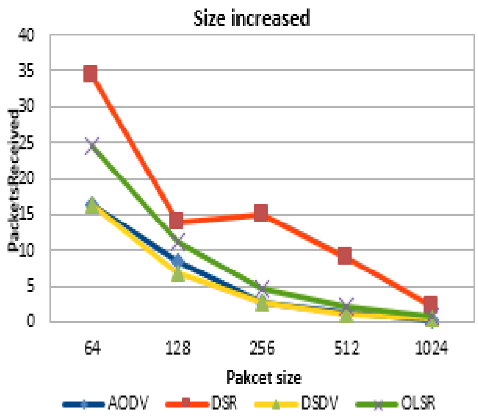


Fig.4.4b

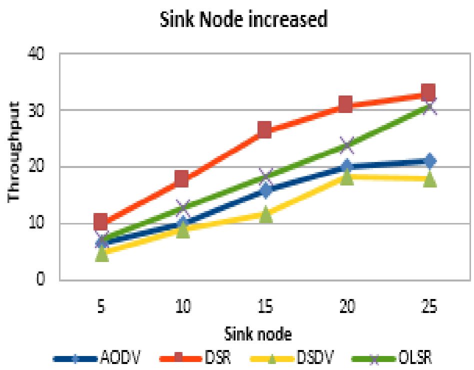


Fig.4.4c

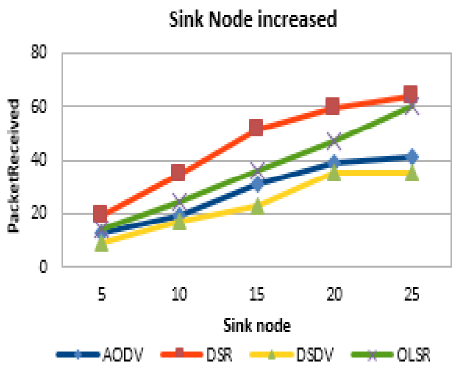


Fig.4.4d

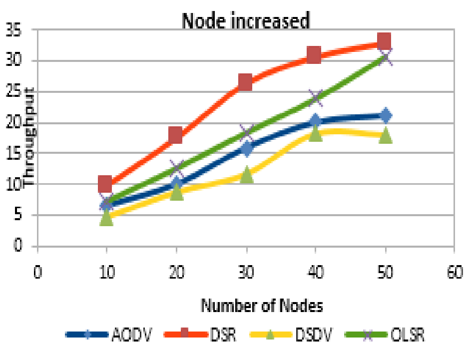


Fig.4.4e

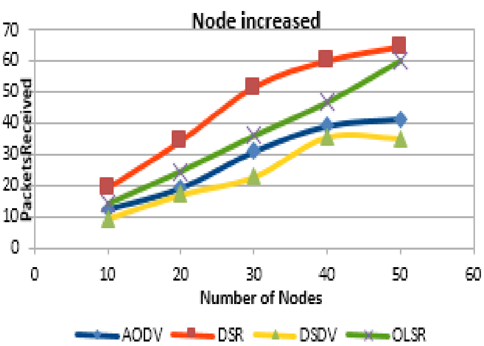


Fig.4.4f

Fig.4.4 Check the throughput and PacketReceived with 4 protocols

Comparison of 4 kinds of routing protocols, DSR has high performance, in this case, the results of throughput and number of packet received are smooth but also have peaks. Because DSR is a source routing based on-demand routing protocol, it needs routing lookups before sending a packet, but if it cannot find a possible routing, the packet will be stored in the cache queue and waiting for send. When a route is discovered, all packets will send out or forward, this the reason why DSR can get very high values in a moment. The flow rate of other three protocols is relatively stable with little fluctuation. In a paper by (Mohapatra and Kanungo 2012), their results in the packet received and throughput is similar to our results. However, they set 10 m/s as max speed and pause time that means the vehicles stop at an intersection or jam in our case and the max speed is similar to our setting (30mph =48 km/h), they also changed the geographic network area but this would not change in our case.

The performance of DSDV is poor than others because there are different with DSR when it does not discover a route. The packet is not stored in the cache or saved but discarded. Moreover, DSDV is a proactive routing protocol and needs to maintain a routing table. When network topology is changed, some routes are lost. However, it still uses lost routes to send packet so that the data cannot arrive at the destination node. Also, the literature of (Ade and Tijare 2010, Jing and Le 2011, Mohapatra and Kanungo 2012) also mentioned that the performance of DSDV is limited in the environment of relatively high speed and large topology due to its characteristics.

AODV and DSR are reaction routing protocols, however, AODV is proposed by DSDV and combined on demand driven of DSR. It differs from the DSR that its header does not carry the routing information, the relay node is based on its routing table by multi-hops. Because in AODV, each node implicitly stores the routing request and routing reply into its routing table, while DSR saves complete routing information in the packet. From our results, the performance of DSR is better than AODV, but AODV is more stable than DSR when the environment is changed. In low speed which usually is 5 m/s, the small size of nodes and big communication cover area, DSR has the best performance than other three. However, when nodes move fast as middle speed around 10 m/s, DSR has poorer performance than AODV. This is proved in the literature of (Ade and Tijare 2010, Jing and Le 2011, LI and TAO). DSR is relying on its cache to find routing information, and updating routing information is slow when the number of nodes increases. By comparison, AODV has less affected when nodes increase because it will frequently update the routing information.

The performance of AODV and OLSR is relatively stable. Moreover, the throughput of OLSR is greater than AODV in each figure on account of OLSR having a lot of “hello” messages in our relatively large topology so that the overhead might be bigger than AODV. Moreover, because of this, OLSR frequently sends a message to neighbourhoods, causing the neighbour nodes can receive many packets in a short time. Meanwhile, in (Mohapatra and Kanungo 2012), they suggested regarding OLSR as a better solution for high mobility condition if the packet delivery ratio and throughput are the prime criteria. The packet delivery ratio (PDR) is that the number of packets received divided by the number of packets transmitted. In our case, we should give more consideration for a packet received. This relates to the quality of data for the prediction model. In the literature of (Haerri, Filali and Bonnet 2006), they tested AODV and OLSR against node density, the results showed that OLSR had smaller routing overhead, and its packet received was better than AODV when some nodes were changed in low density. However, they also stressed that AODV could outperform OLSR when a threshold is reached as the density and number of nodes increased for packet delivery ratio. In the saturated networks, AODV is advantageous.

DSDV periodically broadcasts messages thus takes up bandwidth. DSR excessive dependence on the routing cache and just chooses the shortest path from the cache rather than the latest one when face on the multiple routes, which is likely to pick a failed route. So we suggest AODV which is more stable in our large number of nodes scenario though DSR has better performance with fewer nodes in this experiments. In the following section, we decide to use one proactive routing protocol which is OLSR and one reactive routing protocols which are AODV to run our experiment, thereby possibly impacting the performance profiles of the prediction model. Otherwise, the communication is high performance when sink node increased and packet size decreased. It means our packet size must be controlled as small as possible, and sink node is more the better, but they both have to set in reasonable and realistic situations.

4.4.3 Packet Delivery Ratio and Delay

We discussed the performance of four different routing protocols with our experiments and combined with the experimental results of others, thus chose two protocols as the object of our project which is AODV and OLSR. In this subsection, we will discuss and analyse them in packet delivery ratio and time delay with our Nottingham city centre scenario.

Packet Delivery Ratio (PDR) is the ratio of a total number of packets transmitted successfully from source nodes to destination nodes. The higher PDR gives a better performance of the routing protocols, and the equation of PDR in ns3 is:

$$Packet\ Delivery\ Ratio = \frac{\text{total number of packet recevied}}{\text{total number of packet sent}} *100\% \qquad (4.4\text{-}2)$$

Delay is the time taken when a packet transmits from source node to the destination node that also includes the route discovery wait time, and lower delay means the network or routing protocol has good performance, but when we discuss about the delay, the mean delay is usually used for the entire network, the equation of delay in ns3 is:

$$delay = \text{packet recived time} - \text{packet sent time} \qquad (4.4\text{-}3)$$

The environments of an experiment in this subsection which is the same as in Table 4.4. But the total number of nodes change from 20 to 100, and sink node is always lower than total nodes by 5.

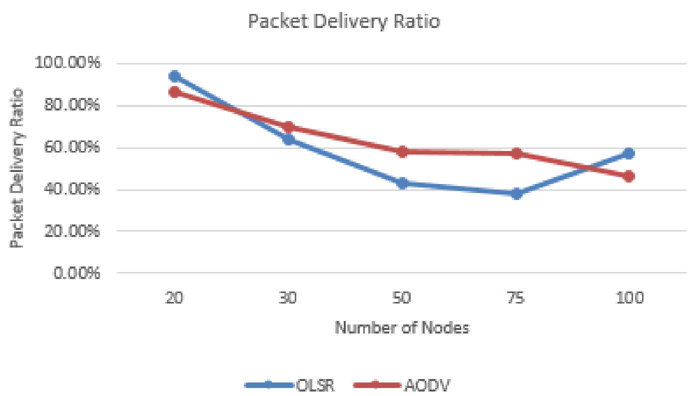


Fig.4.5 PDR of OLSR and AODV with nodes increase

In the Fig.4.5, the PDR of OLSR is a little higher than AODV when the total nodes are 20 and 100. The PDR of OLSR drops faster than AODV when nodes increase from 20 to 50. When nodes increase to 75, the PDR of both protocols not change too much and seems to be stable. However the PDR of AODV is dropped when nodes increase to 100, rather PDR of OLSR gets higher. Perhaps that’s because, the relay nodes become stable to help OLSR improving the packet received when vehicles are increasing and aggregating in intersections.

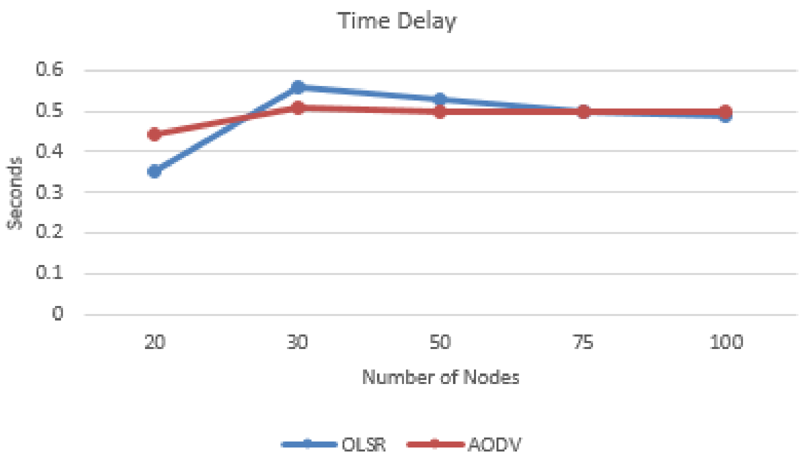


Fig.4.6 Time delay of OLSR and AODV with nodes increase

In the Fig.4.6, we also test the time delay between OLSR and AODV. In general, the average delay is around 0.5 seconds. When nodes increase to 30, the time delay of OLSR rises faster than AODV. Conversely, the time delay of AODV seems to be more stable and the variation is small. AODV as a reactive routing protocol, when RREQ finds a relay node with an active route, then the packet will be sent immediately thus the time delay can be lowered.

4.5 Summary

The Ad-hoc network is flexible, reliable, fast and smooth communication, is not affected by the wired network. Moreover, C2C communication is also not affected by the fixed infrastructure. The dynamic node can be an arbitrary distribution in the network and interconnect by wireless between each other. It is an important field of intelligent transportation applications. The reliable routing protocols achieve the data transmission and higher network performance.

In this chapter, we discussed the approaches of traffic message delivery in VANET, determined that ad-hoc network via multi-hop applied into our project. We also defined the message type, format and content in this thesis which the packet content includes time, vehicle id, speed and positions, then designed a message delivery scheme for data transmission based on the C2C communication. The PPM-C2C framework and its workflow presented in the section 4.3.5. Meanwhile, we compared four routing performance with throughput, PDR and time delay and will apply AODV and OLSR into the following experiments. PDR means how many packets the nodes can receive, thereby how many significant data the prediction model can use. Time delay indicates the lower delay between nodes can give better efficiency to the prediction model thereby reduce errors of predicted results. In the next chapter, we plan to import the mobility pattern to network environments with in-depth research and experimentation. We assume that all vehicles have our prediction model and a wireless network to send the information to each other, then the prediction model will use the information that they have received.

5 Chapter 5

Simulation and Evaluation

5.1 Introduction

VANET has great potential and increasing attention from academia and industry, become a hot research topic. However, because of its characteristics of self-organizing and moving fast, following the rules of the road and a frequently changing network topology, means testing real outdoor conditions is difficult to achieve in studies. Therefore, simulation becomes one of the most important methods of analysis and verification in the current stage of VANET study. From an economic perspective, simulation can reduce the costs of installation, maintenance and replacement for the testing evaluation. The most important thing is that with simulations different parameters and environments can be set. Thus extensive tests can be performed for various possibilities of innovations so that the innovations can be evaluated and verified.

Simulation of VANET includes vehicle simulation modelling and wireless network simulation, where the authenticity of the vehicle mobility model is a key factor to evaluate the reliability of simulation results for VANET. So there are usually two components of separate simulators. The mobility of vehicles is generated by mobility simulators such as SUMO which outputs a trace file for vehicular movements, and also can support real map scenarios and built virtual map scenarios. These generated mobility files are usually based on some mobility models to control speed and avoid collisions, such as Krauss model. The mobility patterns can then be embedded into a network simulator as a trace file, such as NS3. In the network simulator, all vehicular movements follow the mobility trace file to achieve the communication between each vehicle in mobility trace file.

This research involves some data transmission based on real city traffic scenarios which are Nottingham city centre. The simulations are set in different environments and parameters for mobility patterns and wireless network configurations. The investigation around the following aspects: establishing the PPM-C2C framework which includes a pervasive traffic simulation model based on C2C communication; implementing PRAWMA prediction scheme for PPM, and evaluating the predicted results by different environmental influences.

We intend to verify the PPM prediction model with simulation data. In fact, the real-world traffic data is difficult to collect, especially because the PPM is based on the communication environment and all the vehicles can be participants. In reality, the type of vehicle and driving behaviour is different, and not all vehicles are deployed with a communication device to record and send a message to each other in the real-time, for instance, taxi or bus. For this reason, the collection of real-world traffic data may be easily missed or inaccurate. On the other hand, the price of deploying communications equipment is also high, and the spending time of data collected is also considered. Even if the real-world traffic data can be collected, a sample analysis of the data would be another problem. Even if the accuracy of these data is not questioned, these are only a sample, and the sample only reflects a probability or condition. Through simulation, we can directly get the required data and samples, while meeting the different circumstances of the traffic conditions. However, the simulation must be based on reality as much as possible, such as speed limits, road layout, driving directions, routing and rear-end collisions avoidance. Therefore, the simulation data outputs have practical significance, and determine whether the results of prediction model are significant.

A specific description proves that the data and the results of the simulation are in line with the actual traffic situation. A speed limit of 30 mph or 48km/h usually applies in built-up areas in the UK (www.gov.uk/speed-limits). However, in the crowded cities, the vehicle often fail to reach this absolute maximum speed, while the 85th percentile speed is a natural safe speed limit. It indicates that 85% of the vehicle’s speed has this speed or below this speed, only 15% of the vehicle’s speed is above this value in the all observations. This speed is commonly used in traffic management as the maximum speed limit for urban road sections. Therefore in our traffic simulation with SUMO, the vehicle is usually unable to reach the maximum speed of 30mph, and the maximum value is usually around 25mph, and a small number of vehicles will be more than 25mph, which is in line with the actual situation. However, our prediction results will give 30mph as a limit value to describe the space mean speed of the road section, this means that there is no vehicle on the road and it is a free path. Although the main research revolves around predicted travel speed, we will also discuss the practicability and performance of the PPM-C2C framework for the prediction of travel time and traffic density in the experiments.

This chapter is organised as follows: Section 5.2 describes as chosen simulation methodology including traffic mobility models of Nottingham city centre and wireless network models. The simulation assumptions, parameter configurations and influencing factors are also addressed in this section. Section 5.3 provides experimental investigation and evaluation of the performance of PPM-C2C prediction framework and analysis of each influencing factor. Section 5.4 summarises the conclusions about the performance of PPM-C2C prediction framework in different environments.

5.2 Simulation Methodology

5.2.1 Overall Architecture of Simulation

To verify the feasibility of the PPM-C2C prediction framework and the performance of the PPM prediction model, this study has been divided into three important parts: simulate the traffic condition of Nottingham city centre; apply C2C communication in the wireless ad-hoc network; use the data packet to predict the traffic condition. In traffic simulation part, SUMO traffic simulator is used to construct the Nottingham city map and vehicles with the real road network. Then the mobility data which is the topology of generated by SUMO import into NS-3.23. Wireless communication between vehicles is simulated using NS3 and established routing protocols. The PPM is then used to enable cars receiving messages from other cars to arrange and process the messages by area location and time series in order to calculate the traffic condition of a road section by PPM prediction model. The simulation architecture of this study is shown at below:

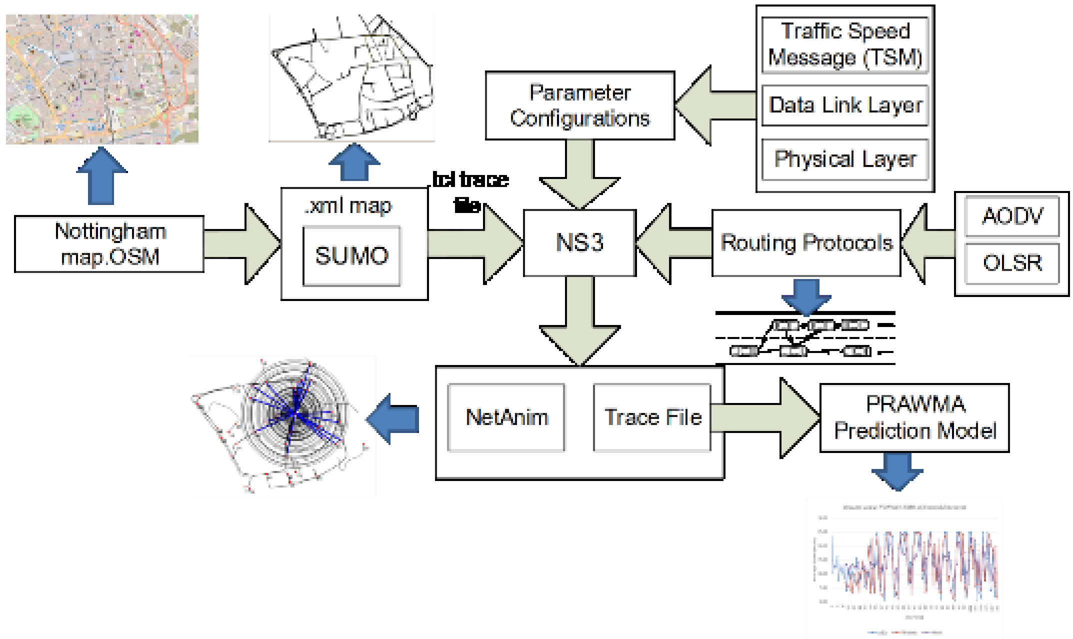


Fig.5.1 Architecture of simulation

The Fig.5.1 shows overall architecture of simulations. OpenStreetMap provides a realistic map to construct a scenario by SUMO, and the realistic mobility patterns of scenario are exported also by SUMO as shown in the section 3.3.1. The mobility pattern is in “.tcl” script and the “duarouter” function can generate different movement routes of mobility pattern with one map scenario that the “.tcl” mobility pattern can import to the NS3. All network settings configure in NS3, which include data packet and WAVE layers. Meanwhile, the routing protocols are integrated into a model in NS3 to achieve the message transmissions based on the mobility pattern from SUMO. NS3 provides a visual interface to show the node’s movements and communications with NetAnim. The outputs as a trace file that include all information for the demands of the prediction model, also contain appropriate information for simulation environments (e.g. time delay, IP/MAC address, PDR). When the information meets the demands of the prediction model, the predicted results can be computed. For the reasons of statistical analysis, much information and graph generations, our equations of PRAWMA prediction scheme (section 3.2.4) are based on functions from the Microsoft Excel 2013 and computed by the steps of the PPM prediction model (section 3.2.5).

5.2.2 Simulation Assumption

Some of the conditions and situations we cannot simulate the actual results, although we use the way of simulation to validate our theories. For example, not all of cars are equipped with GPS, we cannot obtain all data from all cars. In this case, we need some reasonable assumptions to achieve utilisation of C2C communication:

1. No roadside communications infrastructures.
2. All the vehicles are mobile and have their own identification equipped with a GPS device for providing their positions and velocity in real-time, also are able to determine random movement, distribution and density on the road.
3. The packet has complete contents, such as described in the section 4.3.3 and 4.3.4.
4. Each vehicle has the prediction framework embedded, that is enough powerful system for storing data, processing data and computing predicted results.
5. The wireless network allows the time delay, out of delivery range and packet loss.
6. The vehicles can communicate with others using C2C communications standard which is IEEE 802.11p (the current standard for VANET).

In this research, we assume that the simulation data from SUMO is the real time data for the prediction framework. The SUMO provides the Car id, position direction and speed for each node, and also includes the coordinates for each road section. Thus the ground truth data for traffic conditions such as travel speed, travel time and density can be obtained from above data. Then the framework estimates traffic conditions through the simulation data, so that the predicted results can be testified and compared with the ground truth data. In the best case, if the predicted results are in agreement with the changes of traffic condition, we consider that the prediction model can be used for the current traffic conditions. In the worst case, the predicted result is 100% wrong; the prediction model will be defeated. This leads to a decrease in overall prediction accuracy. Moreover, the reasons for the wrong prediction will be discussed in the experiment sections. At the same time, we will focus on the performance of the prediction framework in the C2C communication environment and the impact of the C2C communication environment on the prediction accuracy.

5.2.3 Mobility Pattern

As described in the previous subsection, the first step of simulation is realistic map converted and mobility pattern generation. And we already presented how to generate the mobility pattern in details in the section 3.3.1. The Fig.5.2 shows that a comparison of scenario and mobility about authenticity:

The Fig.5.2a and Fig.5.2b are the real Nottingham city centre with satellite map and road map plan. Fig.5.2b is converted to a simple scenario map in Fig.5.2c by SUMO. The mobility of vehicles generated by SUMO based on the Fig.5.2c. The vehicle mobility traces are displayed in the visual interface NetAnim of NS3 in Fig.5.2d with 200 vehicles.

Fig.5.2a Satellite Map

Fig.5.2b Road Map

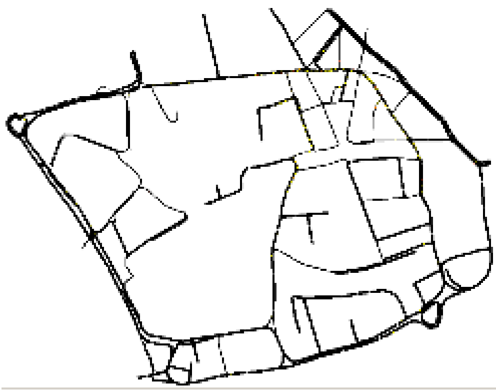


Fig.5.2c SUMO

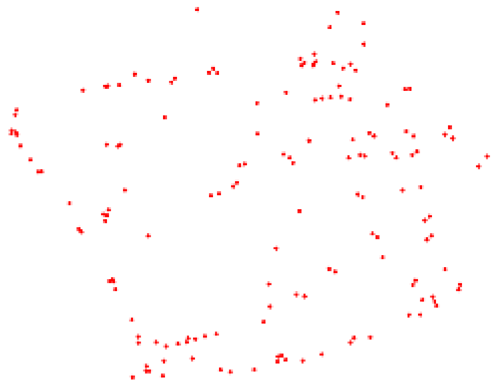


Fig.5.2d NetAnim

Fig.5.2 The comparison of scenario and mobility map

The mobility pattern should contain acceleration, deceleration and route of vehicles, the position of the vehicle movement, in and out of the road network, also street information. All vehicles in our scenario are generated by random routes. The maximum speed of all vehicles is 30mph (48km/h or 13m/s) with the speed limit of city road. Some vehicles could converge in one intersection to cause congestion, while some of the vehicles normal drive on other roads. But they are same effect and part of the network, they need to send information for sharing that sending own information by TSM in every second, receive and forward TSM of other vehicles via one-hop and multi-hop. So the communication between the vehicles is controlled by the NS3.

5.2.4 Scenario Pattern

All vehicles will enter the scene at the different time and leave the scene when they complete their journey with different life times. If we only observe one vehicle, the data might not enough for the prediction model or no enough predicted results for analysis. In this case, we propose to set static observation nodes (SON) in order to collect data and ensure continuity of data. The static observation nodes are same attributes with the mobile node, also can send, forward and receive TSM. Meanwhile, the static observation nodes can also be provided in the different location to predict the average speed of objective road section, as shown at below:

Fig.5.3 Scenario of simulation

The Fig.5.3 shows the scenario of simulation that one TSM packet generated by vehicle V1 is sent from objective road section to the SON.1 via a relay node by two hops at a certain time point. The road section at above with red marked is the objective road section that we plan to predict its traffic condition such as average travel speed. Each vehicle broadcast one TSM packet per second when it drives on this road section. There are six vehicles on this road section which mean 6 TSM packets for describing the speed of this road section are broadcasted in this second so that the static observation node can receive 6 TSM packets for the red road section for this second in theory. There is packet loss and time delay which is usual during data transmissions and those have to be considered into the prediction process. Also, the SONs will be at the edge of the map, away from the main road, and do not affect the normal driving of other vehicles.

5.2.5 Parameter Configuration

There were some experiments with the parameter configurations in the section 3.3.1 and 4.4.2. But they were set by separate for mobility pattern and routing protocol analysis. We will give all parameters and variable values for this thesis in Table 5.1. There includes simulators, mobility details (e.g. number of nodes, speed, movement type and distribution), communication configurations (e.g. channel, layer setting, communication type and packet information) and also weight setting of the prediction model.

Table 5.1 Simulation Parameter

Parameter	**Value**
Simulation area	1200m x 1000m
Simulator	SUMO, ns-3.23
Number of vehicles	500, 1000, 2000, 2800, 6000
Vehicle's speed	0 - 30 mph (0- 48 km/h)
Mobility Model	SUMO defined
Nodes Distribution	SUMO-Mobility with random
Movement Pattern	Random
Simulation Time	3600, 7200, 21600 (Seconds)
Sink of nodes	Less 5 than total
Channel	Wireless
Network Interface	WirelessPhy
Propagation	TwoRayGround
Mac Type	802.11p
WiFi-rate	2 Mb/s

Protocol	AODV, OLSR
Communication type	C2C
Transmit power	7.5dBm
Packet size (bytes)	128
Packet content	TSM
Setting of Weights	0.7/0.2/0.1

5.2.6 Evaluation of Influencing factor

In this thesis, the simulations and evaluation mainly focus on the effect of different time periods, different mobility patterns, different routing protocols, different observation locations, different peak time, different traffic conditions to the PPM with PRAWMA prediction scheme and predicted results. There is also a comparison between PRAWMA scheme, ARIMA model and KNN model in the PPM-C2C framework.

Different routing protocols are another factor that might affect the predicted results. For the C2C communication, the routing protocol ensures timely and accurate transmission of the data. Likewise, for the prediction, a stable communication is an indispensable part of the whole simulation. It can ensure effective delivery and share of information. Therefore, the different routing protocols can give a different impact on network performance, such as time delay and packet loss. We select two routing protocols to perform our experiments, which is AODV as a reactive routing protocol and OLSR as a proactive routing protocol. In the wireless communication, there is bound to be a time delay and packet loss. Moreover, it is a good test for the prediction model due to the incompleteness of data. Thus, an efficient routing protocol is also an influencing factor for the predicted results.

Different mobility patterns give the different road traffic conditions with a different number of nodes and simulation times. When they are increased, the scenario can represent different traffic condition; it looks like it varies in the daytime, night time or peak time which is always changing in density from sparse to dense then sparse of vehicle density. The number of nodes and map size need to be considered when defining the network as sparse or dense. In our case, some nodes are defined from 500 to 5000, this seems busy traffic conditions. However, all 500 nodes do not appear at the same time, some of the nodes go into the network then go out the network when they finish their routes and trips. So, 500 nodes might be a dense network with one hour simulation time. However, the level of sparsity also depends on the simulation time. If the simulation time is 15 minutes with 500 nodes in a mobility pattern, this represents a sparse network. Therefore, we need to consider both of some nodes and simulation time for different mobility patterns. To put it simply, we need to investigate the impact of different vehicle densities to our prediction model.

Different observation locations that set by our observations. Our purpose is that all vehicles in the network can predict the travel speed of a road section according to the information shared. Moreover the route, movement and location of vehicles are random. In this case, we need to know that the predicted results varied significantly in different locations. Because of communication range and multi-hop communication, the vehicles nearby the objective road section might provide more accurately predicted results than a vehicle from far away or the edge of the network in theory. However, we still need to verify it through the experiments.

Different objective road sections determine the different road conditions. As described in section 4.3.4, each vehicle has a powerful container to store the information of various road sections for the prediction model. Moreover, each street might have one or tables for recording the messages. This has resulted in different information for each road sections. On the other hand, each road has its characters such as length of it or its traffic capacity. In this case, we need to investigate if the PPM prediction model varied significantly in different predicted objectives.

Different time periods determine the different length of the time period to the predicted results. It is a parameter for the prediction model. It needs to set down before the prediction model works. It decides how to arrange data for the request of PPM prediction model. The most important that it decided the period of predicted results. We already tested and analysed the predicted results from 5s to 30s time periods in the section 3.3.2 and 3.4.2. We plan to increase the time period as directly influencing factor for the prediction model in the following experiments.

Different peak time determine the peaks and off peaks traffic to predict the traffic conditions. We intend to investigate the workable peak time for PPM and whether the peak and off-peak of the traffic will affect the predicted results of PPM. The temporal distributions and the special routes for the objective road section will be employed which are other than dense and sparse road networks.

Different traffic conditions determine that test the prediction of travel time and traffic density rather than travel speed. The PPM can be evaluated whether the findings apply to other traffic conditions with C2C communications.

Different prediction models are employed into the wireless environment to evaluate that the proposed prediction model with the PPM-C2C framework can improve the prediction efficiency and accuracy. It also proves the effectiveness of the PPM-C2C prediction framework, especially when there are new prediction algorithms in the future that can improve the performance of the framework.

5.3 Evaluation of Experiment

As introduced in the last section, we describe the workflow of simulations, simulation assumptions and preconditions, mobility pattern generation, simulation parameters and influencing factors of analysis. As we are investigating the influencing factors for the prediction model. We will analyse the influence of available parameters and conditions in the following sections.

5.3.1 Analysis of Routing protocol

The C2C communication is based on the multi-hop transmission between vehicles without any infrastructures and centralised controls. In this condition, each vehicle needs to find a suitable transmission path to effective sending and receiving data. Thus, the routing protocol is the key factor for the quality of data. According to previous section 4.4 with the comparisons of routing protocol, we refer the data transmission to AODV and OLSR with a larger number of nodes. We will take three different mobility patterns with two routing protocols which are 500 nodes, 1000 nodes and 2800 nodes in one hour simulation time. In theory, a dense network can give more information than a sparse network, meanwhile keeps many nodes of connections for covering a wider area and whole network. In this case, a dense network can produce more information so that the prediction model uses. Indeed, when the dense network is formed and produce redundant transmissions and simultaneous sending, there is easy to accompany broadcast storms for communications. Since PPM-C2C prediction framework focuses on the traffic condition especially travel speed, so even if there are redundant data generation, it will not affect the predicted results after taking the average. But simultaneous sending can cause network congestion resulting in data loss, this will bring effect on predicted results that cannot be ignored. However, this actually needs experiments to verify the influencing factor of routes to the prediction model.

In this subsection, the experiments are used 500 nodes, 1000 nodes and 2800 nodes in one hour simulation time with 5s, 10s, 20, 30s time period in AODV, OLSR environments. The observation node uses SON.1 in the Fig.5.3 to collect the data. We also use the predicted results of observation data to compare the predicted results of AODV and OLSR, that observation data is assumed without wireless and the methods were shown in the section 3.2.5. In this condition, it can verify the impact of network environments with two different routing protocols to the predicted results. The correlation-based similarity (as shown in section 2.3.5) is used to evaluate the predicted results from PPM. Otherwise, the statistics only focus on the objective road section which is the Parliament Street from east to west in Nottingham city centre as shown in section 5.2.4, such as the packet delivery ratio (PDR) count the packets to SON.1 in regard to objective road section and it is not the information of the entire network.

Table 5.2 PDR and Delay of OLSR and AODV with nodes increasing

PDR

	500	1000	2800
OLSR	72.35%	63.45%	40.36%
AODV	69.89%	68.45%	48.49%

Delay(s)

	500	1000	2800
OLSR	0.35	0.71	0.99
AODV	0.44	0.68	0.89

As above table shown, the PDR of OLSR goes down sharply from 72% to 40% when nodes are increased. This means that SON.1 can receive 72% of the packets from objective road section in a relatively sparse network. But when nodes increase to 2800 as a relatively higher density network, SON.1 only can receive 40% packets and lost 60% packets from the objective road. But we reflect on PDR of AODV, it is not a significant change from 500 nodes to 1000 nodes in the network. Likewise, it significant falls off to 49% and get nearly a half of packet loss from the objective road when nodes increase to 2800. The SON.1 received more packets when AODV is used at 1000 nodes and 2800 nodes.

About time delay, two routing protocols both give results lower than 1 seconds and it is acceptable for the prediction model in spite of delay increasing the number of nodes. The reason of relative time delay that is probably happening is because the scenario size leads to only one-hop or two hops for the communications. So we round the time delay to nearest integral second for the request of prediction model by the integral second. For example, SON.1 receives a packet with 0.35 seconds delay at simulation 380th seconds, the prediction model will regard it as 381th simulation time; if the delay is 1.39 seconds then the prediction model will regard it as 382th simulation time.

The Fig.5.4 shows the green line which the predicted results without wireless are higher than the results with a wireless network in overall. Therefore, we can realize that delay and packet loss make a lot of influence for the predicted results. According to Fig.5.4a, in 5 seconds time period, when the nodes increase, the green line is still maintained at a high level over 90% similarity. And the similarities of AODV and OLSR (blue and red) are almost the same trend. With the PDR of both decrease, the accuracy of predicted results slightly decrease. While nodes are 2800, the PDR of OLSR is 40.36% and PDR of AODV is 49.72%, and similarity of AODV is 0.699 which is higher than OLSR of 0.679 similarities with 5 seconds time period. This basically illustrates that the accuracy of predicted results is affected by PDR. The situations of Fig.5.4b and Fig.5.4c look similar to Fig.5.4a, the similarity of the green line is still better, and keep growth when nodes are increased. When nodes are 500, the similarity of AODV is lower than OLSR due to lower PDR. However with the PDR changing, the similarity AODV is higher than OLSR as higher PDR. According to Fig.5.4d, because the time period increases to lead lower accuracy of predicted results for three curves. Meanwhile, the lower PDR of the dense network (2800 nodes) also makes the accuracy of predicted results at a low level.

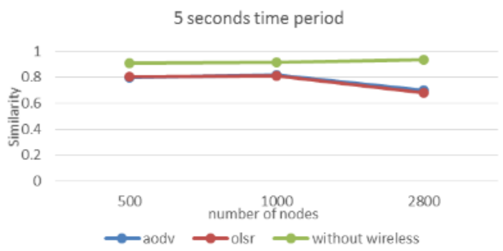


Fig.5.4a

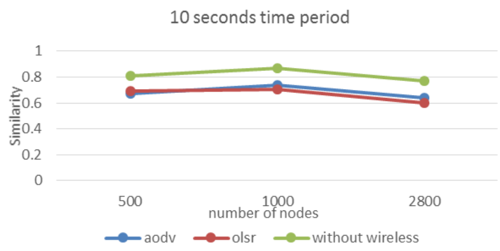


Fig.5.4b

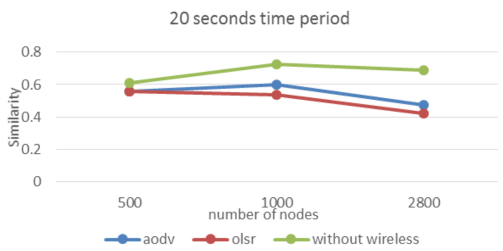


Fig.5.4c

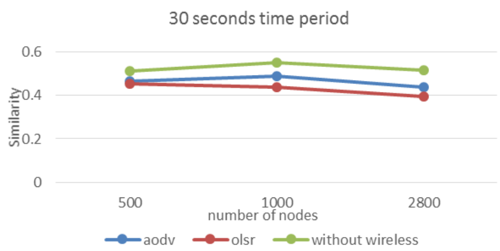


Fig.5.4d

Fig.5.4 The similarity effect of nodes increasing with different routing for PPM predicted results

For similarity results of PPM, the results of AODV and OLSR are not much difference. However, there is a larger difference between the results of wireless environments and no wireless environments which are about 5%-10% for average similarity. Also, there are three density network levels which are a spare network with 500 nodes, medium network with 1000 nodes and dense network with 2800 nodes. In general, the performance of similarity for medium density network is better and higher than the spare and dense network for three curves. This is because the generated data from vehicles might not be enough for the request of the prediction model in the spare network, and if there is not enough data that will affect the accuracy of predicted results. However, in the dense network, the opposite happened: enough vehicles can give enough data to predict, but the packet loss is increased at high density, and hence triggers a situation where there is not enough data to reach the observation node. From this experiment, the medium network present the best predicted results with wireless network environments. However, the influence of different routing protocol between AODV and OLSR is not obvious to the predicted results, the performance of AODV seems a little better than OLSR in about 5%.

The similarity represents the trend accuracy of predicted results, and MRE represents the value accuracy of predicted results. As the same conditions, the Fig.5.5

shows the MRE remove case effect of nodes increasing with different routing protocols and observation data as no wireless environments. In the general, the results of no wireless environment are little better than wireless environments with about 2%-5% in average. In theory, no wireless environments mean that assume the observation node can receive all information of objective road section. In this case, the information should be adequate for the prediction model, and the accuracy of no wireless environments should be higher than wireless environments. This is due to the large cardinal number, and a smaller error can be produced after average them so that the probability of error appears small in no wireless environments. Meanwhile, the cardinal number is less in the wireless network with a relatively small number of useful data, there is not fundamentally different from no wireless environment. Therefore, in general, the wireless environments does not affect the accuracy of predicted result in a value.

Fig.5.5a

Fig.5.5b

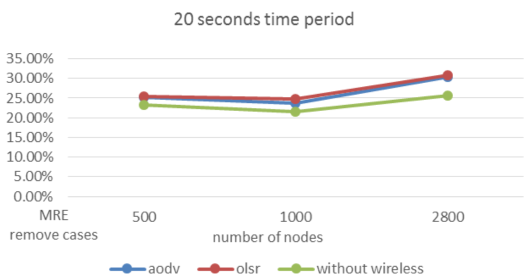


Fig.5.5c

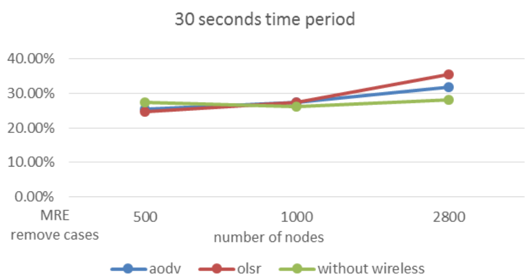


Fig.5.5d

Fig.5.5 The MRErc effect of nodes increasing with different routing for PPM predicted results

The Fig.5.5 also presents that the percentages of MRErc rise when nodes increase in each sub-figure. The best performance of MRErc is in the sparse network with 500 nodes which are about 15% to 20%. When a number of nodes are low, there are no much nodes travel on the objective road, additionally, due to packet loss, the observation node does not receive sufficient reference data to predict. Moreover, PPM deems default max speed (30mph) to replace that if there is no data at this second. In the sparse network, the road section is in the idle state at most of the simulation time, so that the predicted result represents default max speed. In this case, the accuracy of predicted results is high level at the sparse network. On the other hand, we accurate to two decimal places for the average speed, this also can increase the errors rather than integer number. When the network is changed from sparse density to dense density, the road section is not in an idle state anymore. The observation nodes start to receive more information for predicting, and the number default max speed is reduced, then the errors become large. Despite all this, the errors are stable changing when nodes increase from 500 to 1000. At medium density network, the errors are 17%-27% in each sub-figure, higher 5% or less than the sparse network. However, the errors rapid rise when nodes change from medium network to dense network. So the PPM seems better suited to medium network in the wireless environments with 20s and 30s time period (Fig.5.5c and Fig.5.5d).

This subsection tested the performance of PPM in wireless network environments. The use of ad-hoc data with wireless network environments had an impact on the trends of predicted results at 0.05-0.1 in average similarity and didn’t much impact on the accuracy of predicted values. However, AODV is a little better than OLSR in the medium network and dense network. Thus, we will use AODV as wireless environments at following experiments. The best performance of the scenario which is in the medium density network and the errors of predicted results increase when the time period increase. The results of each predicting in this subsection are shown in Appendix B.

5.3.2 Analysis of Mobility Pattern

From the conclusion of the previous subsection, the medium network is the best scenario for the PPM in the wireless environment. However, we have only used three mobility patterns for the wireless environment testing with different routing protocols. In this subsection we purpose to change the mobility patterns, that we will generate the different mobility patterns from SUMO with all nodes random movement in different routes, then each mobility pattern will be imported into NS3 in wireless environments. Eight mobility patterns will be used with four mobility patterns having 1000 nodes in

1 hour with medium density network, and another four mobility patterns are 2800 nodes in 1hour with the dense network. We intend to investigate whether the different mobility patterns will affect the predicted results and whether the PPM-C2C framework is authentically more suitable for medium density network than the dense network.

In this subsection, the experiments have 1000 nodes and 2800 nodes in one hour simulation time for four times per experiment and the time period is still 5s, 10s, 20s and 30s. The observation node uses SON.1 in the Fig.5.3 to collect the data for each mobility pattern. We still use the predicted results of record data to compare the predicted results of AODV, and analysis of the effects of different mobility patterns will be divided to no wireless environment at the first part and wireless environment based on AODV routing protocol at the second part. In the Fig.5.6 the performance of similarity and MRErc for the PPM predicted results with eight different mobility patterns.

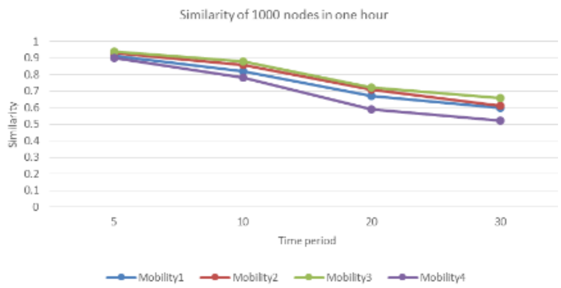


Fig.5.6a

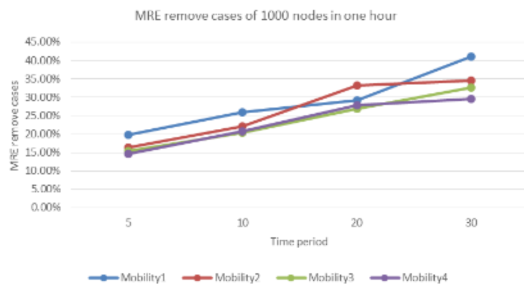


Fig.5.6b

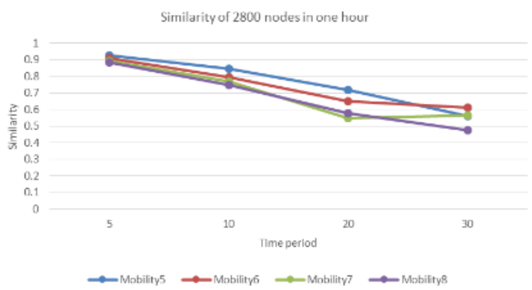


Fig.5.6c

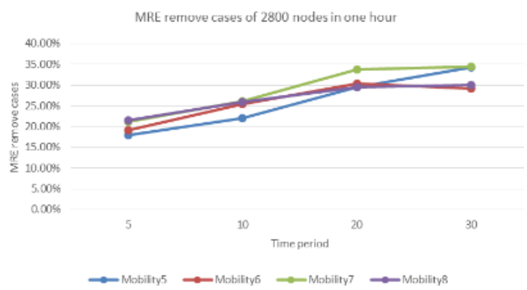


Fig.5.6d

Fig.5.6 The effects of different mobility patterns with no wireless environment for PPM predicted results

According to Fig.5.6a and Fig.5.6b, there is a comparison of four mobility patterns with 1000 nodes (medium density network). In general, the similarities of four mobility patterns present the same downward trends with time period increasing. Moreover, MRErc also presents the same upward trends with time period increasing. These results are consistent with all previous results, where predicted results are influenced by time period. In the Fig.5.6a, the probable interval of similarity is 0.90 to 0.93 in a 5s time period which is over 0.9. The similarity interval of 10s time period is 0.79 to 0.88 which can be seen 0.8 to 0.9; the similarity interval of 20s time period is 0.60 to 0.72 which is about 0.7, and the similarity interval of 30s time period is 0.52 to 0.66 which is lower than 0.7. However, the difference between them is about 0.05 with the same period. Comparatively, mobility 4 is lower than the average similarity of four mobility as 0.70, and the highest similarity is mobility 3 which is 0.80 in the average. As Fig.5.6b shows, time can still be divided into different intervals of time periods for MRErc. When the time period is 5s, the MRE is at 20%, when the time period is 10s, the MRErc is increased to between 20% and 25%. When the time period is 20s, the MRErc is about 30%; and finally, when the time period is 30s, the MRErc is increased to about 30% to 40%. The best two of MRErc are mobility 3 and mobility 4, mobility 1 and mobility 2 are relatively weaker. However, the gaps between them are also not more than 5% in average and in each time period. There is a special case for mobility 4, it has a little weak similarity for the trend of average speed prediction, but has higher level results of MRErc. That is because the mobility 4 produced smaller TSM packets than other three mobility patterns, it only has 3285 TSMs for the objective road section. However, mobility 1 has 4874 TSMs, mobility 2 has 4659 TSMs, and mobility 3 has 6000 TSMs. Although, they have the same number of nodes, there are many vehicles passing by objective road section in mobility 3. Thus, we can easily attain that mobility 3 relatively better results with more reference data. Also, mobility 4 has little TSM packets which lead to its relatively weak similarity. Meanwhile, as we explained at previous subsection, the PPM prediction model deems default max speed (30mph) to filling that if there is no data at this second. So, if there is no enough reference data from the objective road, also may give a satisfactory result with MRErc. In other words, if the objective road remains a longer idle state, the MRErc may be more satisfactory. Overall, the predicted results are not affected by the different mobility patterns with 1000 nodes at no wireless environment.

According to Fig.5.6c and Fig.5.6d, there is a comparison of four mobility patterns with 2800 nodes (dense network). They present the same trend with Fig.5.6a and Fig.5.6b in general, when the time period is increased, the similarity decrease and the

MRErc increase. Moreover, the intervals of similarity and MRErc in Fig.5.6c and Fig.5.6d are also similar to Fig.5.6a and Fig.5.6b. However, the similarities of the dense network are lower 0.04 than the medium density network's average, and the MRErc of the dense network is higher by 2% than medium density network. In similarity, the best performance is mobility 5 which is 0.76 in average, and poor performance of them is mobility 8 which is 0.67 in average. The four mobility patterns are similar to each other which is about 26% errors in average. The same situation with medium density network, the reason for this it still may be some reference data. From mobility 5–8, they also produced the different number of reference data for the objective road section; mobility 5 has 14577 TSMs; mobility 6 has 13768 TSMs; mobility 7 has 12021 TSMs; mobility 8 has 8119 TSMs. This result is obvious as mobility 5 has most of the reference data, and the reference data of mobility 8 is too small that less than nearly one half of mobility 5.

Table 5.3 Comparison of Mobility 3 and Mobility 5

	Similarity					MRErc				
Time Period	5	10	20	30	Average	5	10	20	30	Average
Mobility3	0.9395	0.882	0.7218	0.6574	0.8002	17.50%	22.13%	26.98%	32.60%	24.80%
Mobility5	0.9281	0.8476	0.7205	0.5613	0.7644	17.90%	22.04%	29.61%	34.28%	25.96%

Comparison of mobility 3 (medium network) and mobility 5 (dense network) is described in Table 5.3, they have the best performance for each case. The MRErc results of mobility 3 are close to mobility 5 in general. However, mobility 3 has a full advantage in the results of similarity. From the number of reference data, mobility 3 has 6000 TSMs, and mobility 5 has 14577 TSMs that is much more than mobility 3. Therefore, the performance of similarity can become higher level with more reference data while the MRE remove case might be rise. In general, medium network and the dense network can both give relatively higher level results of similarity and MRErc in no wireless environment when the number of reference data is increased.

The second part of this subsection that we will test the performance of PPM with eight mobility patterns in AODV. Table 5.4 gives the PDR, average time delay, the number of total TSM produced for the objective road section in Fig.5.3 and number of reference data as observation node received.

Table 5.4 PDR and Delay of Different Mobility Pattern with AODV

Nodes	1000 nodes				2800 nodes			
Scenario	Mobility1	Mobility2	Mobility3	Mobility4	Mobility5	Mobility6	Mobility7	Mobility8
PDR	61.88%	67.27%	68.45%	69.38%	48.49%	46.96%	46.77%	49.55%
Delay(s)	0.78	0.73	0.68	0.69	0.89	0.9	0.93	0.87
Total TSM	4874	4659	6000	3285	14577	13768	12021	8119
Rev. TSM	3016	3134	4107	2279	7069	6466	5622	4023

*PDR is the total ratios of SON.1 received packets from objective road section.

*Delay is the average time delay from the nodes travelling on the objective road section to SON.1.

*Total TSM is how many TSM packet from objective road section. (Or reference data in no wireless environment)

*Rev. TSM is how many TSM packet SON.1 received from objective road section. (Or reference data in wireless environment)

The PDR will be reduced due to a number of node increasing and time delay also be increased. In this case, the PDR of 1000 nodes and 2800 nodes have a difference about 10% to 15%. And the average time delay is about 1 second with one-hop or two hops to the observation node. The results of PPM with eight mobility patterns is presented at Fig.5.7 in a wireless environment.

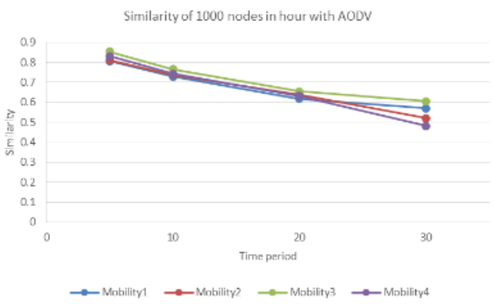


Fig.5.7a

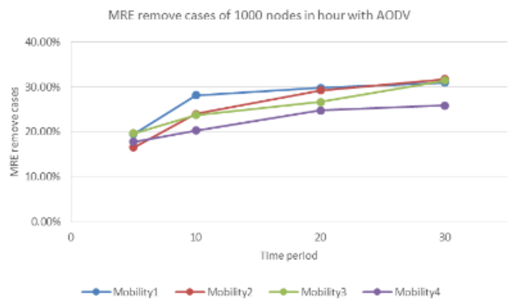


Fig.5.7b

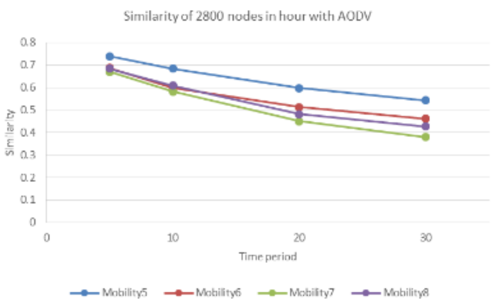


Fig.5.7c

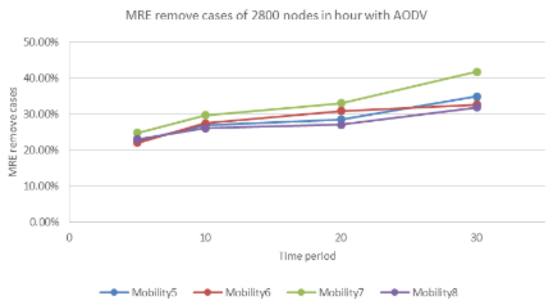


Fig.5.7d

Fig.5.7 The effects of different mobility patterns with AODV for PPM predicted results

According to Fig.5.7, in general, the trends of four sub-figures are consistent with our argument in the section 5.3.1 that the similarity of wireless environment is lower about 5% -10% than similarity of no wireless environment in each time period; and the MRE remove case is not significant change between no wireless and wireless environment.

When mobility patterns have 1000 nodes, the Fig.5.7a shows that the mobility 3 is the relatively best similarity with 0.72 in average, but the difference of other three are not too much which are 0.67, 0.68 and 0.67 in average. It is significant packet loss with about 30% to 40% that produces an impact of the predicted results. However, when we reflect on Fig.5.7b, the MRE remove case of mobility 3 is increased and higher 3% error than mobility 4 on average. In mobility 4, the SON.1 receive less information than no wireless environment, so that the default max speed might accounts for much proportion when the road is free. According to Fig.5.7c, the gaps of similarity between mobility 6, 7 and 8 are not obvious in the dense network with the wireless environment, and the trends of their similarity are decreased with time period increasing. The similarity of mobility 5 is 0.64 average, there is higher than mobility 6 with 0.56, mobility 7 with 0.52 and mobility 8 with 0.55. Also, the similarity of 2800 nodes is also lower 0.12 than 1000 nodes in total average. At the same time, the total average of MRErc (Fig.5.7d) is also higher 4% than 1000 nodes. Especially, the MRErc of mobility 8 continues at a low level with less information.

The wireless network environment leads to an overall decline in predicted results for each mobility pattern. In the same density network, the predicted results of similarity drop about 5% to 10% in each time period with the wireless environment; rather, the predicted results of MRErc are not a significant difference. In the different density network, the predicted results of similarity with 1000 nodes are higher 0.12 than 2800 nodes in total average with wireless environments and is not a significant difference with no wireless environment. This is because the dense network leads to a lot of packet loss. If the mobility pattern likes to mobility 4 and mobility 8, that they produce less information for the objective road to lead relatively big errors. This does not mean that they also produce poor predicted results for other road sections. Mobility 5 and 8 are relatively satisfied mobility pattern for this road section (the Parliament Street from east to west in Nottingham city centre). Mobility pattern provides vehicle distribution and random route of vehicle movement in the road network. Therefore, the predicted results of similarity are influenced by network environment, selection of objective road section and time periods, and less affected by mobility patterns. The full results in this subsection are shown in Appendix C.

5.3.3 Analysis of Observation Location

As the previous two subsections, we tested the performance of PPM in a wireless network environment with different density of networks and different mobility patterns. But the observation node is fixed at the edge of the map which is SON.1 (as shown in Fig.5.3). In this subsection, we will use different observation nodes to collect the data and investigate the influence of different observation locations in the wireless environment to the PPM.

In view of the conclusion of the previous subsection that mobility 3 and mobility 5 are relatively satisfied mobility pattern. Thus, we purpose to use them in this subsection for testing the effect of observation node to the PPM. Each location receives different information, resulting in different data for the prediction model using. We choose three observation nodes (as shown in Fig.5.3) as different locations and still aim at the Parliament Street from east to west in Nottingham city centre. SON.1 is at the left bottom which is far from the objective road section; SON.2 is nearby the objective road section; SON.3 is at the edge of the map which is the upstream position of the objective road section. They are located in different areas so that the received data varies slightly. Each observation node will be applied into mobility 3 and mobility 5 for running. The results of simulations for the static observation nodes are in Table 5.5.

Table 5.5 Performance of static observation nodes in wireless environments

	Mobility 3			Mobility 5		
SON	SON.1	SON.2	SON.3	SON.1	SON.2	SON.3
PDR	68.45%	72.73%	60.93%	48.49%	57.61%	44.30%
Delay	0.68	0.62	0.79	0.89	0.81	0.93
Total TSM	6000	6000	6000	14577	14577	14577
Rev. TSM	4107	4364	3656	7069	8398	6457

The PDR and time delay of SON.2 are better than SON.1 and SON.3 in the both mobility patterns due to its location close to the objective road section. The most of the vehicles travelling on the objective road section can one-hop transfer the TSM packet to SON.2, this greatly reduces the delay while increasing the success rate of data transmission. SON.1 is located at downstream of the objective road, and the data transmission needs one-hop or two hops to get destination. In this case, a time delay and packet loss will be produced to lead volatility of predicted results. SON.3 is located at upstream of objective road section and produces most time delay and packet loss. The vehicles travelling on the objective road need to transmit TSM packet from an opposite direction with movement to SON.3, the communications are usually based on the two-hops or three hops to increase time delay and packet loss. Also, the scenario map (Fig.5.3) addresses, there is a road dense regions between the objective road and SON.3, and a large number of vehicles can concentrate in this area within a short time. This also leads to increase a lot of time delay and packet loss. From the results of Table 5.5, the SON.2 receives more TSM packets than other two observation nodes in each mobility pattern, so that in theory and the conclusions from previous subsections, the predicted results of similarities for SON.2 also should be better than SON.1 and SON.3; the results of MRErc should not much significant difference between them.

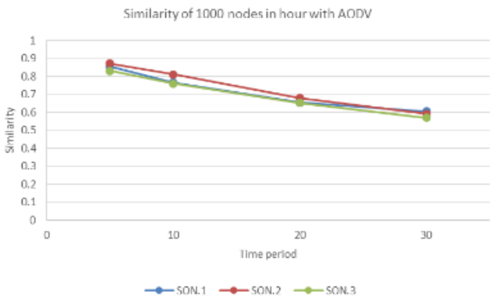


Fig.5.8a

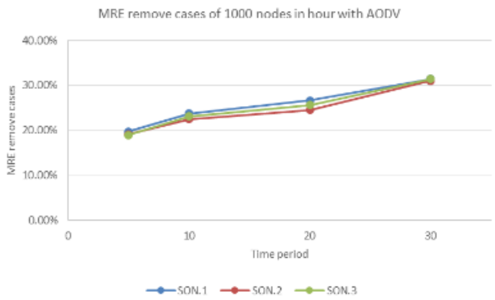


Fig.5.8b

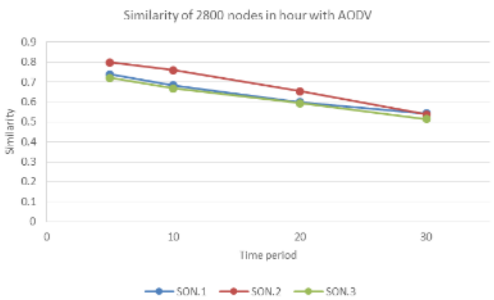


Fig.5.8c

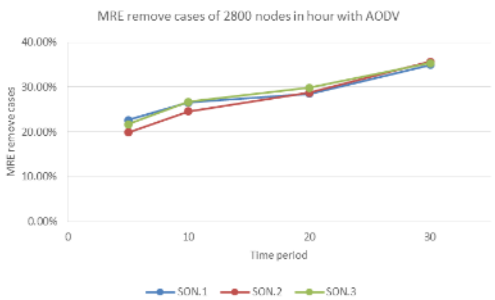


Fig.5.8d

Fig.5.8 The effects of different observation nodes with AODV for PPM predicted results

According to Fig.5.8a, the results of SON.2 are higher 0.04 than SON.3 and SON.1 at 5s and 10s time period. Meanwhile, SON.1 and SON.2 are both better than SON.3. However, the difference is not obvious in medium density network in general. Because the number of received data as the prediction model using of SON.1 is only less 257 than SON.2 and SON.3 are less 708 than SON.2. So there is a difference with the similarity of each observation node, but it is very little. As expected in the Fig.5.8b, the results of MRErc should not the much significant difference between each observation node. The MRErc of SON.2 is greater than SON.1 and SON.3, but the average values of them are very close about 25%.

As the Fig.5.8c shown, the similarity of SON.2 is significantly higher than SON.1 and SON.3 at 5s, 10s and 20s time period, which the difference of average similarity is 0.06 between SON.2 and other two. The reason is that SON.2 receives much more information than SON.1 (1329) and SON.3 (1941). SON.1 and SON.3 are basically same downward trend, except SON.3 appeared abnormal at 20s time period. However, Fig.5.8d shows more closely resembled trend of MRErc from 19% to 35%. Obviously, the MRErc of SON.2 is much better than other two at 5s and 10s time period. However there is no difference for the average value accuracy of predicted result about 25% errors by different observation nodes in the dense network. But because of the number of receive data, the similarity renders the differences in the dense network.

In this subsection, we test the performance of PPM at different observation nodes with medium density network and dense network. In medium density network, the predicted results are little better when the observation node is close to the objective road section, but the difference between each observation node is not much. In a dense network, the predicted results of the observation node nearby the objective road that perform significantly better than the distant observation nodes. The MRE remove case does not have a visible difference in both density network. In general, the different observations have a greater impact on dense networks. The results of each predicting in this subsection are shown in Appendix D.

5.3.4 Analysis of Objective Road Section

In the previous experiments, we only focused on one road section that is the Parliament Street from east to west in Nottingham city centre. This street is the main road of the city and it is significant to study prediction of average travel speed. The length of this street is longer than other roads in Nottingham city centre with 0.4 miles (about 650 meters) with three traffic lights. So there might be affected by each block for average travel speed of the whole street. In this case, we purpose to choose another street as the objective road section. Collin Street west to east (blue) is a busy street at the south of city centre with 0.1 miles length (about 160 meters) as shown in Fig.5.9, it has heavy traffic and small length.

The simulation of this subsection will use 1000 nodes and 2800 nodes in one hour to compare the predicted results of two different objective road sections. Moreover, the predicted results will include no wireless environment and wireless environment with AODV by using SON.2. Because SON.2 is located at the centre of scenario, it can receive more data in the wireless environment. Thus, we can investigate the impact of different objective road sections and road length to predicted results.

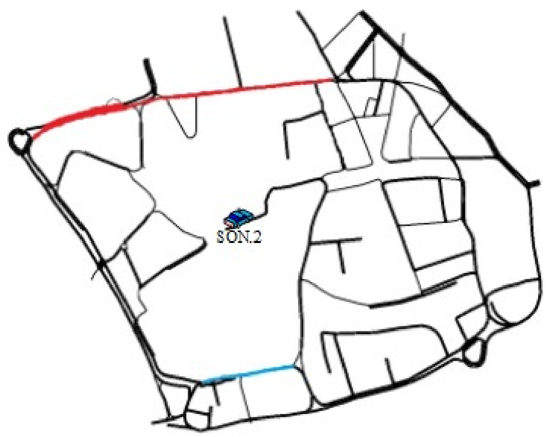


Fig.5.9 Scenario of two streets

The SON.2 is located at the middle of Collin Street (blue) and Parliament Street (red), and receives TSM almost one-hop or two hops. Table 5.6 shows the performance results (PDR, delay, total TSM and received TSM) of wireless environment for Collin Street and Parliament Street in 1000 nodes (medium density network) and 2800 nodes (dense network). The statistical is only limited the two objective road sections, does not mean that the whole network performance.

Table 5.6 Results of wireless environment for Collin Street and Parliament Street

	1000 nodes		2800 nodes	
Street	Collin	Parliament	Collin	Parliament
PDR	69.36%	72.13%	59.45%	61.91%
Delay	0.67	0.62	0.8	0.81
Total TSM	3948	4080	11306	12021
Rev. TSM	2716	2943	6626	7442

According to Table 5.6, the PDR is decreased about 10% with nodes increasing, and PDR of Collin Street is 3% below PDR of Parliament Street in each scenario. The time delays of both streets are under 1 second in average. The total TSM of both streets are almost same in each scenario that means they have very close traffic volume in an hour, but the number of TSM received is affected by the PDR. The difference of reference data is 227 TSMs in 1000 nodes and is 816 TSMs in 2800 nodes, the number of reference data from Parliament Street is greater than Collin Street. In general, the conditions of two streets are still similar to the total TSM and received TSM (reference data) in each scenario. The predicted results are based reference data via PPM for different objective road sections as shown at below:

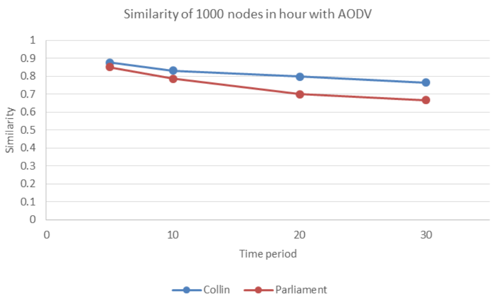


Fig.5.10a

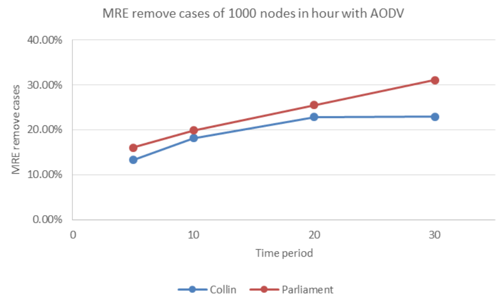


Fig.5.10b

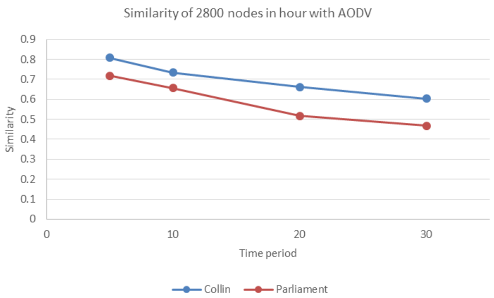


Fig.5.10c

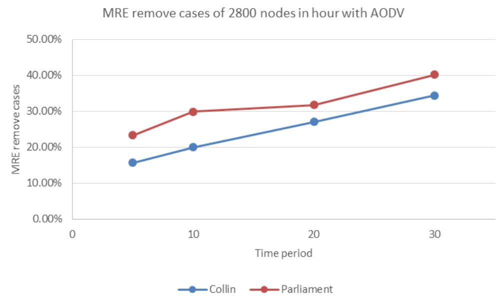


Fig.5.10d

Fig.5.10 The effects of different objective road with AODV at SON.1 for PPM predicted results

According to Fig.5.10a, in 1000 nodes scenario, the similarity of Collin Street is higher than Parliament Street in each time period setting, especially when the time period is 20s and 30s, the difference of similarity is 0.1 between them. As Fig.5.10b shown, the MRErc of Collin Street has remained at a low level and high accuracy with a small value which is 19% in average, and is better than Parliament Street as 5% in average. Overall, when the time period is increased, the similarities of both streets still slow down significantly and the MRErc of both streets rise slowly up.

When nodes are increased to 2800, as the Fig.5.10c and Fig.5.10d shown, the overall trend in line with the previous experimental results. The similarity of 2800 nodes is declined markedly than the similarity of 1000 nodes with 0.1 in general. According to Fig.5.10c, the similarities of Collin Street are apparently higher than Parliament Street with 0.09 at 5s time period, 0.07 at 10s time period, 0.15 at 20s time period and 0.14 at 30s time period. The Fig.5.10d shows that there is a significant difference between two curves where the MRErc of Collin Street is only 24.30% in average and Parliament Street has 31.31% in average.

The Fig.5.11 presents the results of similarity and MRErc for Collin Street and Parliament Street by PPM without routing protocol. This means that the total TSM is reference data for the PPM. We plan to investigate whether there is predicted a difference of two streets in wireless environment or not as all TSM packets obtained.

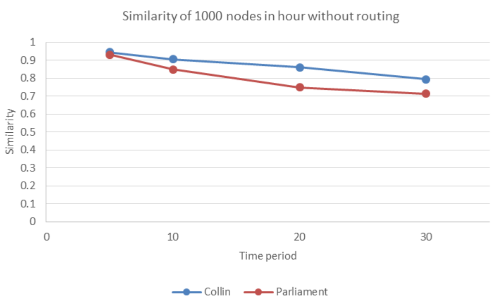


Fig.5.11a

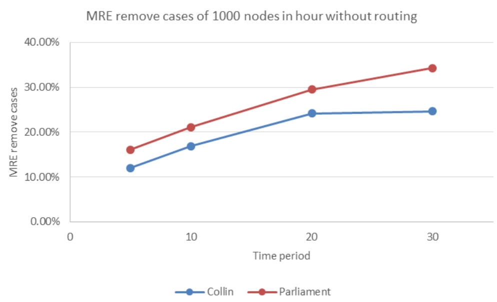


Fig.5.11b

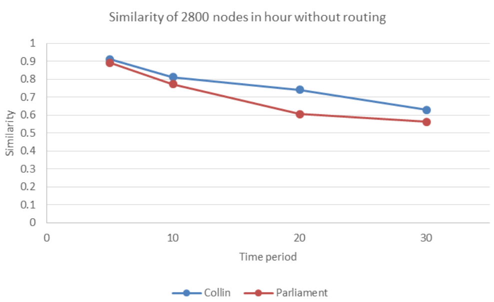


Fig.5.11c

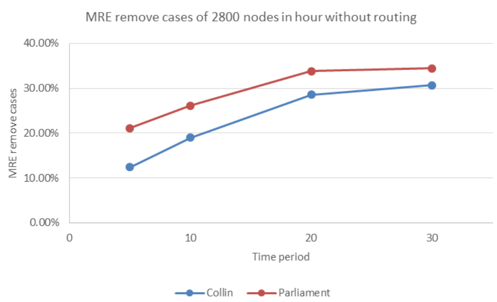


Fig.5.11d

Fig.5.11 The effects of different objective road without routing for PPM predicted results

From Fig.5.11a, the similarities of Collin Street remain at a very high level which is 0.95 at 5s time period, 0.90 at 10s time period, and 0.86 at 20s time period, then drops to 0.79 at 30s time period. The average of similarity is 0.87 for Collin Street. In contrast with that of the similarities of Collin Street and Parliament Street, the results of

Parliament Street are severely lagging behind in each time period, especially at 20s and 30s time period. Fig.5.11b also shows an obvious difference between two curves. The MRErc of Collin Street increases from 12.03% to 24.60% with time period changing and its average is 19.42% which it is at a low level and high accuracy of predicted results. However the MRErc of Parliament Street rapid rises from 16.12% to 34.29% with 25.26% in average, and it significantly underperformed Collin Street in 1000 nodes.

The effectiveness of all results in the dense network is still lower than medium density network in each sub-figures in Fig.5.11. According to the Fig.5.11c, there are the same results with 1000 nodes situation. The similarities of Parliament Street lag behind Collin Street as 0.08 in average in the dense network, the effectiveness to 10s, 20s and 30s time period is particularly outstanding with the difference of 0.04, 0.14 and 0.06. From Fig.5.11d, the difference of MRErc between two streets is 8% in average, and results of Parliament Street is far behind that of Collin Street.

In this subsection, we test the performance of PPM with PRAWMA prediction scheme at different objective road sections with medium density network and dense network in a wireless environment to verify the influence of predicted result by different length of objective road sections. In the same conditions, two streets have almost equivalent reference data for predicting. This means almost same of traffic volume in the both streets. The results of Collin Street exhibit better prediction accuracy than results of Parliament Street in all aspects by the network, a number of nodes and time period changing. As our expected, the Parliament Street is longer than Collin Street, and it has three blocks. Any one of blocks can impact the average speed of whole streets. If vehicle A is decelerating at the start of the road, at the same time another vehicle B is accelerating at the end of the road. The behaviour of vehicle A needs time to transfer to vehicle B due to the length of the road. However, Collin Street is a short street with heavy traffic, and there is only one traffic light. The transmission time of behaviour is greatly reduced between vehicles. And the behaviour of vehicles can be easy to keep consistent with travelling on Collin Street, and the behaviour has more continuity of vehicle movement, thus the prediction accuracy is higher than Parliament Street. For PPM, the predicted results are more accurate when the length of objective road section is short. The results of each predicting in this subsection are shown in Appendix E.

5.3.5 Analysis of Time period

In the previous experiments, we only set the time periods as 5s, 10s, 20s, and 30s and the performance of predicted results fell down with time periods increasing. Moreover, we concluded that each time period setting had an interval of similarity for the predicted results (as discussed in section 5.3.2), and the similarity of predicted results decreases and MRErc are increased when time period increases. So we intend to refine and increase the time period in order to verify the influence of longer time periods to PPM. Meanwhile, the user would like to know what future is happening for a longer time period. For this reason, we need to increase the simulation time and number of nodes to meet the predicted request of the longer time period.

In this subsection, the simulation is used 2800 nodes in 2 hours and 6000 nodes in 6 hours, two objective road sections which are Collin Street and Parliament Street (as shown at Fig.5.9). Due to a large topology using in the experiment, it makes very poor performance of wireless environment, so that we will not use wireless environment in this experiment. The predicted results are only based on the record speed information from SUMO for testing the effects of the longer time period. The time period will be set as 5s, 10s, 20s, 30s, 60s, 90,s, 120s and 180s.

Table 5.7 Total data produced from SUMO

Scenario	2800 nodes in 2 hours		6000 nodes in 6 hours	
Street	Collin	Parliament	Collin	Parliament
Total Data	11680	16740	23825	52193

As Table 5.7 shown, the Parliament Street has more traffic pressure than Collin Street in both scenarios. Those results are from same mobility pattern by SUMO recorded for each scenario. The Fig.5.12 shows the PPM predicted results of similarity and MRErc for Collin and Parliament Street in 2 hours (2800 nodes) and 6 hours (6000 nodes) with time period increasing from 5s to 180s.

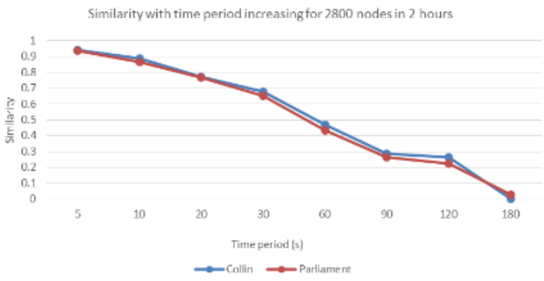


Fig.5.12a

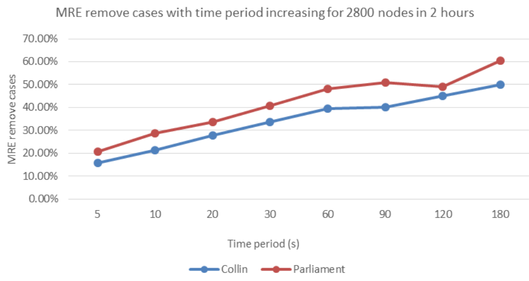


Fig.5.12b

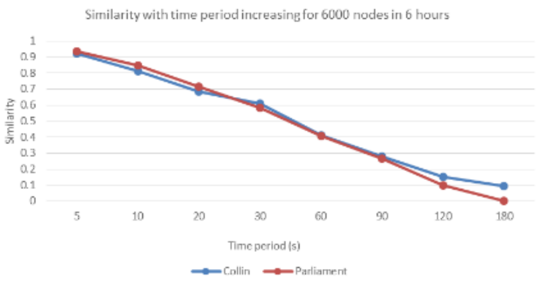


Fig.5.12c

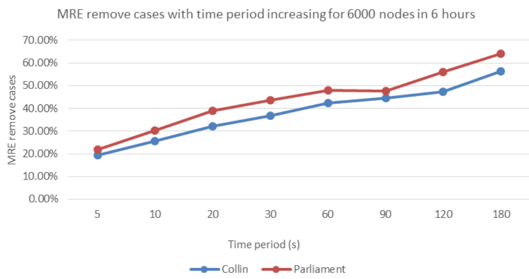


Fig.5.12d

Fig.5.12 The effects of time period increasing without routing for PPM predicted results testing in Collin and Parliament Street

In general, the trend of graphs in line with our conclusions at previous sections that similarity of predicted results decreases and MRErc is increased when time period increases in both scenarios. According to Fig.5.12a, the similarity trend and value of Collin Street is similar to Parliament Street. The similarity can still be maintained at a high level for both streets at 5s time period to 30s time period. But the similarities declines rapidly after 60s time period that are only 0.46 for Collin Street and 0.43 for Parliament Street until the similarities get zero at 180s time period for both streets. The same problem is also reflected in the Fig.5.12b, the errors of predicted results keep rising with time period increasing for both streets. At 60s time period, the MRErc is 39.66% for Collin Street and 48.14% for Parliament Street until the errors get over half at 180s time period, although the results of Collin Street retain a dominant position in Parliament Street. Therefore, the PPM predicted results are not ideal when the time period is 60s or longer time. But it will be presented a high performance of the predicted results within 60s time period.

The Fig.5.12c and Fig.5.12d present the same situation as the scenario of 2800 nodes. A large amount of data does not improve the performance of predicted results for longer time period. The similarities of both streets drop to about 0.41 when the time period is 60s. The MRE remove case of both streets also give the high errors after 60s time period. Moreover, the contrast of the similarities for both streets is not much difference while Parliament Street has a much more information than Collin Street. However, the MRErc of Collin Street is still better than Parliament Street in overall.

From this experiment, we can determine that the PPM predicted results change with the time period. The shorter time period can produce more accuracy of similarity for the average travel speed prediction that it is within 60s. And the accuracy of similarity decreases markedly with the increase of time period when the time period is over 60s. The MRErc continue to rise when the time period is increased. Furthermore, the PPM predicted results are not obviously influenced by topology size (simulation time and nodes increasing). Parliament Street is a long street, the reference data is more much for the predicted model when the traffic flow is small, in this case the similarity changes fast of travel speed and it is poor than Collin Street. However the difference of similarity between them that is not much obviously when Parliament Street has more data than Collin Street. More data mean that Parliament Street has more traffic pressure and vehicles than Collin Street. So the travel speed does not change fast, and it will become gently if there is a congestion on the street. Therefore, the similarity of both streets is very close in a large size of network topology. The results of each predicting in this subsection are shown in Appendix F.

5.3.6 Analysis of Peaks and Off Peaks

We discussed the different mobility patterns from SUMO with all nodes random movement in different routes in the section 5.3.2. In this subsection we intend to investigate whether the peak and off-peak of the traffic will affect the predicted results of PPM. Therefore, we employ temporal distributions and the special routes for the objective road section which are other than dense and sparse road networks. For example, we consciously increase the traffic volumes for the Collin Street as the traffic peak hours in SUMO; instead, the traffic volumes of the road section will be decreased for the off-peak hours.

In this subsection, one experiment is employed 2800 nodes during one hour with a dense road network, but 1000 nodes among them pass the Collin Street with the special routes and the rest of 1800 nodes will be still random movement in the network, hence there are over 1000 nodes that will pass the Collin Street. The second experiment is still employed 2800 nodes with random movement in the network, but we delete some routes containing the Collin Street in SUMO. Therefore, some nodes will not appear on the network for providing off-peak hours at Collin Street. The simulation runs one hour with 5s, 10s, 20s and 30s of time periods. The observation node uses SON. 2 to collect the data. We still discuss the predicted results with AODV and without wireless. The purpose is to test the performance of PPM with PRAWMA scheme in peak and off-peak hours.

Table 5.8 Results of Peak and Off-Peak by AODV for Collin Street

Street	Collin 2800 nodes	
Level	Peak	Off-Peak
PDR	71.16%	79.28%
Delay	0.72s	0.52s
Total TSM	16707	3982
Rev. TSM	11889	3157

According to NS-3 outputs in Table 5.8, the PDR of off-peak is higher than peak. The reason is that we delete some routes through Collin Street to achieve the off-peak hour. In fact, the nodes are less than 2800 in the off-peak scenario. The PDR of the peak is also better than dense network as previous examples. This is because we control the probability for emitting one vehicle per 2 seconds, and also set 1000 vehicles gradually moving into the scene, driving through the Collin Street and leaving the scene. This way does not affect the network performance as the random route. The random route might generate a large number of vehicles in a certain period due to network congestion. On the other hand, this way also ensures that data is generated almost every second. Therefore controlling of the vehicle distributions can enhance communication performance to obtain more effective data. Based on the above-obtained data, the results of PPM are in the following figure:

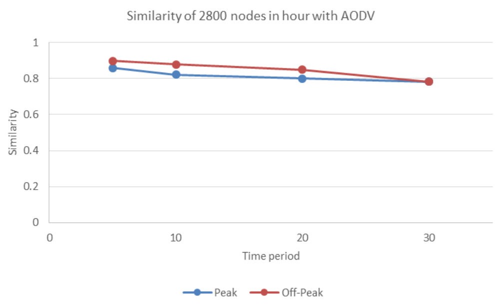


Fig.5.13a

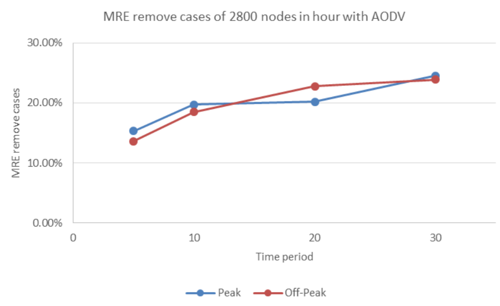


Fig.5.13b

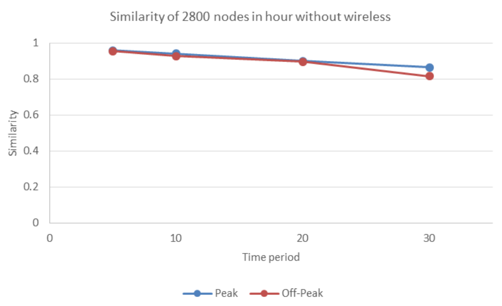


Fig.5.13c

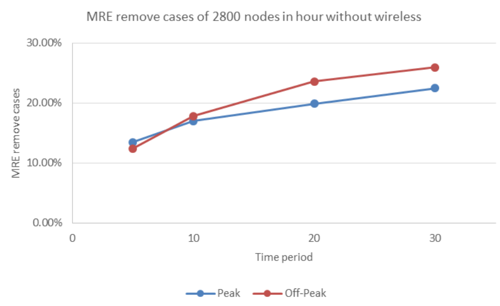


Fig.5.13d

Fig.5.13 The effects of Peak and Off-Peak for PPM predicted results testing with AODV and without wireless at Collin Street

In general, the similarity and MRErc of both are satisfactory results. From Fig.5.13a, the similarity trend and value of off-peak are slightly better than peak. The reason that we discussed in previous sections, a few traffic volumes means that may not have a vehicle on the road at most of the time, so the results are maximum speed by default. The similarities of both conditions are over 0.8 in the average, and the MRErcs are less than 20% error.

The Fig.5.13c and Fig.5.13d show the predicted results without wireless. Without data missing, the predicted results can be maintained at a relatively high level which is 0.9 with similarity and 20% with an error in average. The network environment has a large impact on the peak, and has little effect on the off-peak. It illustrates the outcome of an experiment, like 30s time period at below:

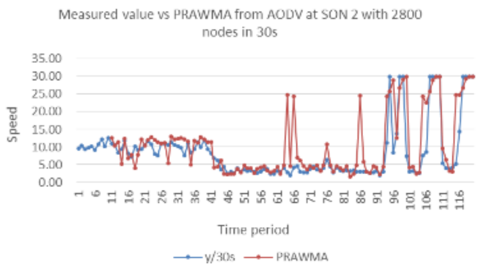


Fig.5.14a

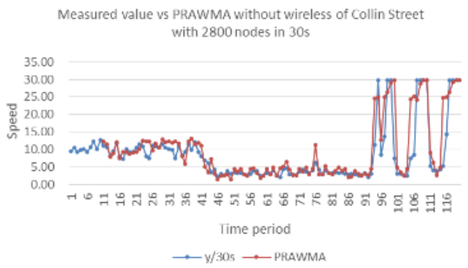


Fig.5.14b

Fig.5.14 Observation speed Vs Predicted speed with AODV and without wireless at 30s time period for the peak hour

The above figure shows space mean speed per 30 seconds at Collin Street during 1 hour simulation time. The traffic is normal to travel in the first 40th time period which is about the first 20 minutes; then the traffic begins to increase which slows the speed down under 5mph from 20 minutes to 45 minutes which is peak time (41st to 90th) ; in the last 15 minutes, the number of vehicles drop and the speed improves. The obvious difference between two figures is at 14th 18th 28th 35th 64th 66th and 86th time period. We can conclude that there is data missing due to communication reasons at these time period, especially the latest speed is absent. Therefore, the predicted results will be better, in the case of data can be continuously obtained especially the latest speed can be obtained.

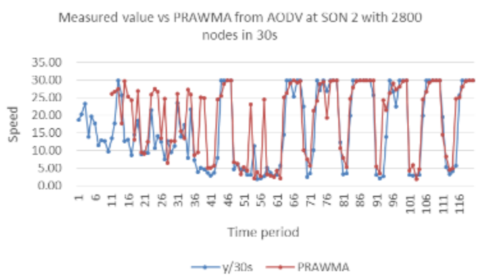


Fig.5.15a

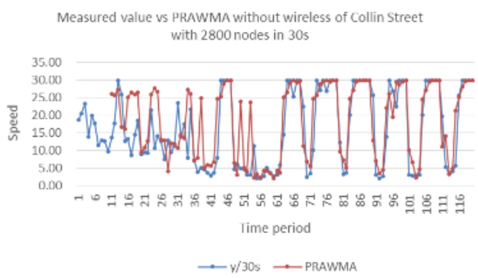


Fig.5.15b

Fig.5.15 Observation speed Vs Predicted speed with AODV and without wireless at 30s time period for the off-peak hour

The above figure presents an off-peak hour traffic condition by average speed. The traffic is still normal state in the first 20 minutes; then a traffic jam happens between 23 minutes to 30 minutes; after that, the traffic condition goes to an off-peak. From both figures, there is a value error for the prediction at normal traffic condition, the observation speed and predicted results all have obvious fluctuations. The model can describe the trend of speed change, but the value estimation is biased. There are two errors that are similar but different in nature, such as the time period of 53rd and 57th in the Fig.5.15a, and the 50th and 53rd in the Fig.5.15b. As the 53rd time period does not obtain the latest data, so there is a problem in both; at 57th time period, the latest data is not available in the communication environment and replace with default speed (30mph); also in communication environment, the delayed data replaces the default speed (if the latest data is missing) with the latest data, so that the error is in at 50th in Fig.5.15b but not in Fig.5.15a as a coincidence. The purpose of this experiments is mainly to verify the operation of the PPM in the off-peak time. Obviously, the predicted results show two curves are fitting very well after 61st time period.

In this subsection, we test the performance of PPM with PRAWMA prediction scheme at peak time and off-peak time with modification routes in the wireless environment. The predicted results of both are greater than normal traffic state. The PPM may have optimum performance, when the data can be obtained continuously, especially the accurate latest data. In the communication environment, if the data can guarantee the amount and continuity, then even if some of data loss will not affect the prediction at the peak time. This problem can improve the routing algorithm or latest data selection. But the model is more suitable for off-peak time, the difference between wireless and no wireless is smaller. This is because less traffic leads to less impact on each other, easy to the free-road situation at the off-peak time. The vehicle speed can reach the approximate maximum speed, even if there is data loss, but the loss data is similar to the default speed. Therefore, the errors will be small with off-peak time prediction. The results of each predicting in this subsection are shown in Appendix G.

5.3.7 Analysis of Travel Time Prediction

We analysed space mean speed of road section based on the PPM-C2C framework with the wireless environment in the previous sections. In this subsection we intend to investigate the effectiveness of PPM-C2C for the travel time as another traffic condition. The most of methods for travel time prediction apply sensor data for collection. In this project, the proposed model is embedded in each vehicle and based on the ad-hoc data, so that this method could be computationally efficient and applicable to the widely available urban road. The current travel time is defined by (Zhang and Rice 2003, Wu, Ho and Lee 2004):

$$T(t,\Delta) = \sum_{i=0}^{L-1} \frac{x_{i+1} - x_i}{v(x_i, t-\Delta)}$$

Where $T(t,\Delta)$ is the current travel time, Δ is the data delay, L is the number of section, $x_{i+1} - x_i$ is the link length, $v(x_i, t-\Delta)$ is the speed at the start of the highway section. $T(t,\Delta)$ is based on the available data that are latest to t temporally.

We run two experiments in this subsection, that 1000 and 2800 nodes will be employed in the network during one hour simulation time with 5s, 10s, 20s and 30s time periods. The objective road section is Parliament street, and observation node uses SON. 2 to collect the data. The purpose is to test the performance of PPM with PRAWMA scheme for travel time prediction.

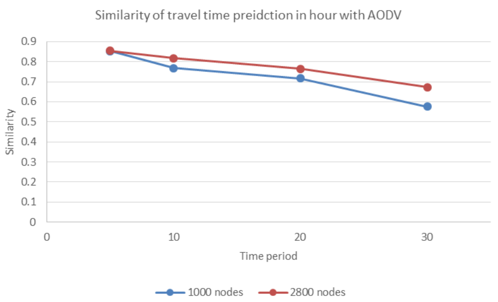


Fig.5.16a

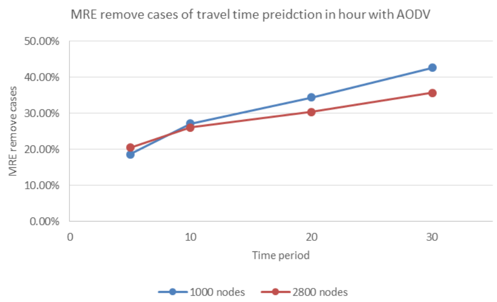


Fig.5.16b

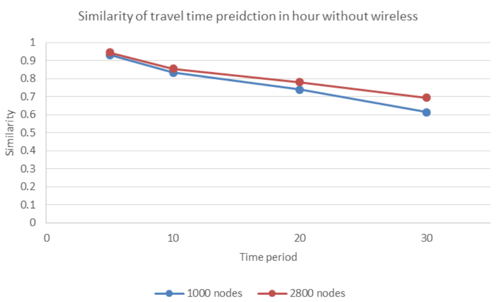


Fig.5.16c

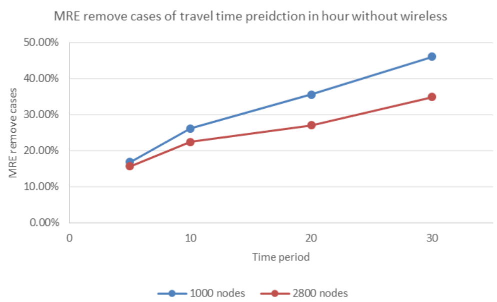


Fig.5.16d

Fig.5.16 The effects of travel time prediction testing with 1000 nodes and 2800 nodes in AODV and without wireless environment at Parliament Street

From the results at above figures, the effectiveness of travel time prediction is strong with 2800 nodes in the both environments. The PDR is 73.03% with 1000 nodes and 58.52% with 2800 nodes. The communication leads to a reduction in the effects of predicted results within a reasonable range which is about 5% with similarity. Based on the small sample and wireless communication, the prediction model decreases significantly with the increase of the time period, which is in line with our previous prediction of the speed. In the case of 10s time period with 2800 nodes:

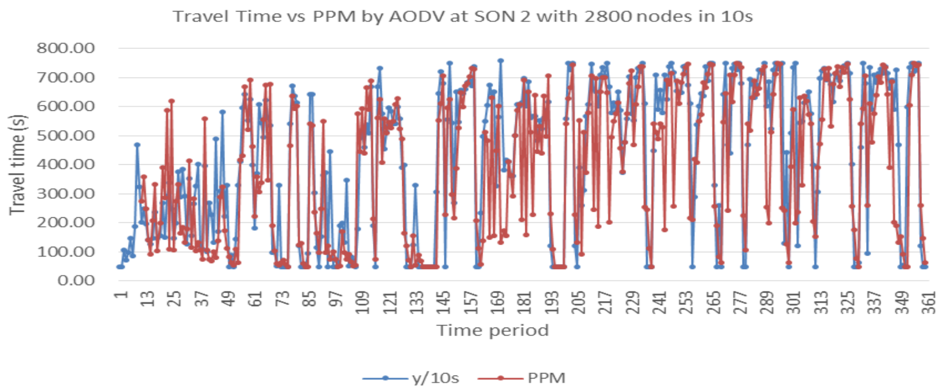


Fig.5.17 Travel Time Vs PPM predicted results with AODV at 10s time period

As above figure shown, the travel time is between 50 seconds and 750 seconds, which means the travel time required to pass the road section at the highest speed or the lowest speed. If the data is missing, the model will be applied the maximum speed, but in fact this may be due to communication problems caused by data loss or free road. Nevertheless, it is difficult to reach and always maintain the maximum speed on the urban road in the real world or in simulation experiments. So the value of lowest travel time (50 seconds) is only prompted that there is no vehicle or data loss. When the speed of a vehicle is under 5 mph, the model can estimate that if this speed is maintained it takes more time to pass through the section. Therefore, through the proposed model to predict the travel time, the instant speed has the greatest impact on the results. The influencing factors of travel time prediction can be considered as equivalent to influencing factors of travel speed prediction. The results of each predicting in this subsection are shown in Appendix H.

5.3.8 Analysis of Traffic Density Prediction

In this subsection we intend to investigate the effectiveness of PPM-C2C framework for the prediction of traffic density. Normally, higher density means longer travel time and lower speed. In addition, the change of traffic density is more random and is also affected by the traffic signal at short-time. Based on the proposed model required, the density of a road section should be collected by a time series data list in each vehicle. The most of the methods for density prediction apply the statistical approach based on short-time observation or through other parameters to estimate density levels such as speed or traffic volume. According to the wireless communication, knowing the route and destination of each vehicle is not required. Any individual in the network only needs to count a number of vehicles at a certain position at a certain time based on the car to car communications. Obviously, the communication could lead to data loss and time delay, which caused insufficient data errors or prediction failure. In this case, we test the capability of the proposed model to make a useful dynamic prediction of the density based on the small sample of ad-hoc data for a short-time prediction. We will not discuss the critical density which is maximum density achievable under free flow, and jam density which is the maximum density achieved under congestion of the road section here. We only focus on the estimation and calculation of density at the future time according to the short-time collection of ad-hoc data. The traffic density (k) is defined as the number of vehicle per unit length of the roadway (in units of vehicles per kilometre) and is calculated by:

$$k = \frac{N}{L}$$

Where N is number of vehicles occupying a road section of length L.

We use the same experiments in the section 5.3.6, that peak and off-peak scenarios will be employed in the network during one hour simulation time with 5s, 10s, 20s and 30s time periods. The objective road section is Collin Street and observation node uses SON. 2 to collect the data. The count of traffic density is the same number of TSM in Table 5.8. The purpose is to test the performance of PPM with PRAWMA scheme for traffic density prediction. The results of PPM is in the following figure:

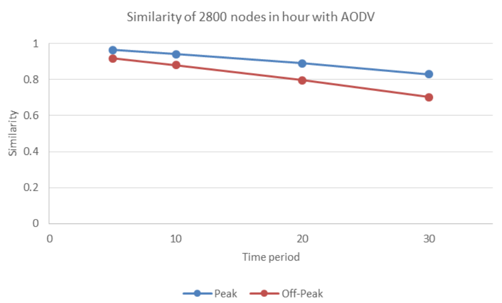


Fig.5.18a

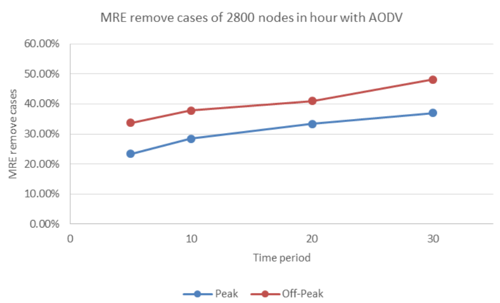


Fig.5.18b

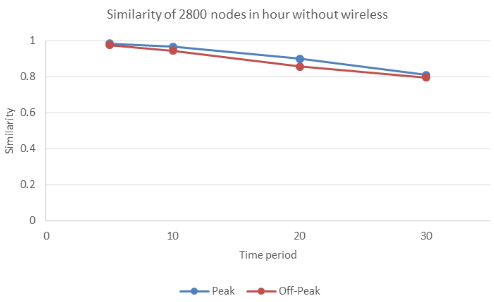


Fig.5.18c

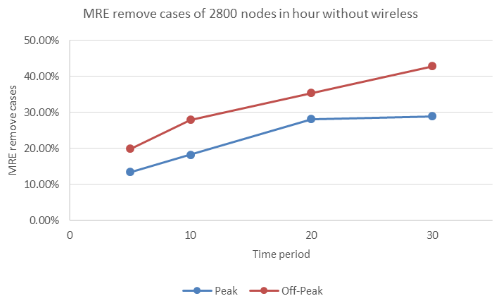


Fig.5.18d

Fig.5.18 The effects of traffic density prediction testing at peak and off-peak hours with AODV and without wireless environment at Collin Street

In general of above figure, the similarity and MRErc of both are satisfactory results, especially the similarities of both conditions that there are over 0.7 with a relatively high level. In the details, the similarity trend and value of peak is better than off-peak in each time period. The MRErcs of the peak is also better than off-peak in average. The network environment has a large impact with MRErcs on the peak hours and off-peak hours. It illustrates the outcome of an experiment, like 30s time period at below:

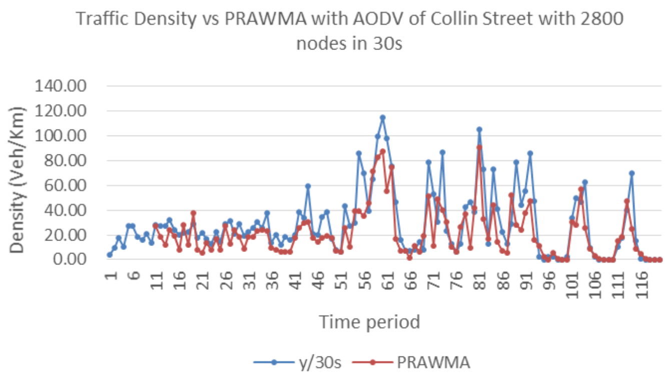


Fig.5.19 Traffic Density Vs Predicted Traffic Density by PPM using AODV during the peak hours at 30s time period

As above figure shown, the value of predicted results is lower than current traffic density in general. This because the wireless communication makes packet loss and forms a count error with density. Accordingly, there is a big difference with MRErcs between AODV and without wireless environment. Because the higher density means lower speed in normally. Three conditions can be distinguished from above figure where the traffic is normal travelling in the first 40^{th} time period with medium density which is about the first 20 minutes; then the traffic begins to increase during the peak time (45st to 95th) with high volatility of density; in the last 15 minutes, the number of vehicles drop with lower density. According to Fig.5.14 in section 5.3.6, by contrast, there is in high density with low speed, thus the higher density can infer lower speed by PPM. As the same time, the change of traffic density is more random and is also affected by the traffic signal at short-time. Therefore, using the latest data has positive consequence of density prediction. The results of each predicting in this subsection are shown in Appendix I.

5.3.9 Analysis of PPM, ARIMA and KNN

In this project, the proposed scheme PRAWMA is one part of PPM model which can be embedded in each vehicle based on the PPM-C2C framework, so that this scheme may be replaced by other prediction models. In this subsection, we intend to investigate the performance of PRAWMA, ARIMA and KNN based on the PPM-C2C framework for the prediction of traffic conditions. The introductions of ARIMA and KNN discuss in section 2.4.2 and section 2.4.3.

We run two experiments in this subsection, that 1000 and 2800 nodes will be employed in the network during one hour simulation time with 5s, 10s, 20s and 30s time periods. The objective road section is Parliament Street and observation node uses SON. 2 to collect the data. The purpose is to test the performance of PRAWMA, ARIMA and KNN based on the PPM-C2C framework for travel speed prediction. We will use Euclidean distance as the distance metric for KNN model, and K-value is 3.

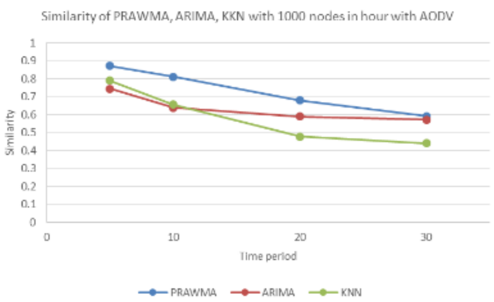


Fig.5.20a

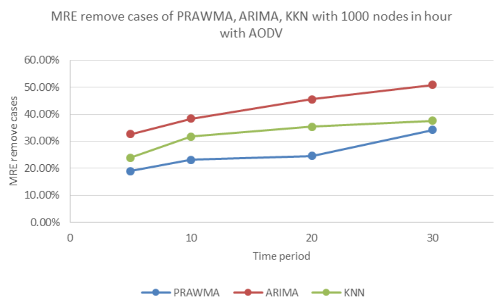


Fig.5.20b

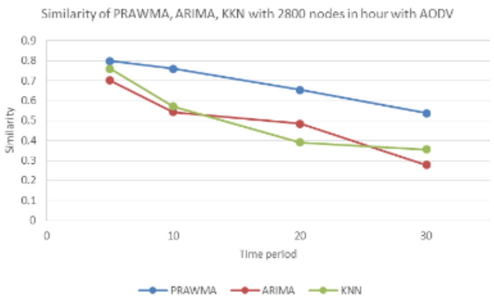


Fig.5.20c

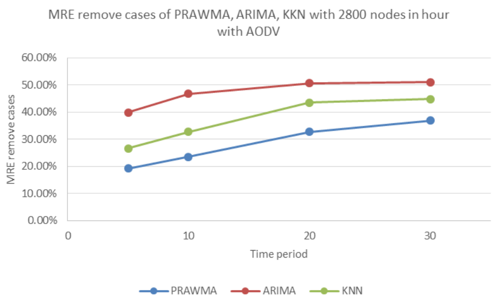


Fig.5.20d

Fig.5.20 Travel Speed is predicted by PRAWMA, ARIMA and KNN based on the PPM-C2C using AODV with 1000 and 2800 nodes

As above figure shown, the results PPM with PRAWMA scheme based on the C2C environment are much better than ARIMA and KNN based on the C2C for different conditions. According to Fig.5.20a, in 1000 nodes scenario, the similarity of PRAWMA is higher than ARIMA and KNN with 0.1 in each time period setting. The MREm of PRAWMA is higher than ARIMA with 15% and KNN with about 5%-10% for each time period in Fig.5.20b. Overall, when the time period is increased, the similarities of ARIMA and KNN slow down significantly and the MREm of both rise slowly up.

When nodes are increased to 2800, as the Fig.5.20c and Fig.5.20d shown, the overall trend in line with the previous experimental results. The similarity of 2800 nodes is declined markedly than the similarity of 1000 nodes with 0.05 to 0.1 in general. According to Fig.5.20c, the similarities of PRAWMA are apparently higher than ARIMA and KNN with the difference of 0.1 and 0.04 at 5s time period, 0.21 and 0.18 at 10s time period, 0.17 and 0.26 at 20s time period, 0.26 and 0.17 at 30s time period. The similarities of ARIMA and KNN fall down significantly when the time period is increased. The Fig.5.20d shows that there is a significant difference between three curves for each time period where the MREm of PRAWMA is only 28.12% in average, ARIMA has 47.12% in average and KNN is 36.97% in average.

Across these experiments, we determine that the PPM with PRAWMA scheme is more suitable than ARIMA and KNN for traffic prediction based on the C2C framework. From the predicted results, we observe that the accuracy of prediction has a positive impact on the performance. The PPM with PRAWMA has 10% improvement for each condition, even if the prediction accuracy decreases as time period increased. The advantage of PPM is that the computation is fast and sensitive to the latest data. Except for wireless communication reasons that we discussed earlier sections, the deficiencies of ARIMA and KNN also affect the predicted results, such as the input data size, the stationarity of data and selection of K-value or distance. The biggest problem is that they usually need to manually use PC software for calculating and testing in each step. For the in-car prediction, adaptabilities of ARIMA and KNN is poor. The results of each predicting in this subsection are shown in Appendix J.

5.4 Summary

This Chapter introduced the experimental process for this research, and provided experimental proofs of PPM working in C2C communication environments with ad-hoc data for a short time traffic conditions, then discussed the influencing factors of routing protocols, mobility patterns, observation locations, objective road sections and time periods to the PPM-C2C framework.

The performance of predicted results used similarity for trend prediction and MRErc for value prediction which is compared to simulation observation data from SUMO. On the other hand, the wireless environment was performed AODV and OLSR protocols for data transmission between vehicles. We also proved that PPM is superior to the existing classic prediction model in C2C environment.

The PPM-C2C framework is designed for traffic prediction in each car, especially travel speed prediction. Through the previous experiments show that the in-car pervasive prediction model based on the C2C communication is feasible for short-term predictions. In general, the prediction accuracy is reduced by 10% in the wireless environment. It also decreases with time period, where take an hour as an example, the similarity is over 0.9, and MRErc is under 20% when time period is 5s; the similarity is between 0.8-0.9 and MRErc is between 20%-30% when time period is 10s; the similarity is between 0.6-0.8 and MRErc is between 30%-40% when time period is 20s; the similarity is under 0.65 and MRErc is over 40% when time period is 30s; the prediction is basically ineffective when the time period is greater than 30 seconds. The results suggest that the prediction of travel time and traffic density is better than travel speed prediction with performance probably higher than 5%-10%. At the same time, the PPM performance is 10% higher than the traditional models based on the PPM-C2C framework.

5.4.1 Similarity of Predicted Results

The similarity is mainly used to evaluate the speed trend changing of predicted results. According to the analysis of section 5.3, there were include five factors might affect the predicted results. The results of Section 5.3.1 showed that the use of ad-hoc data decreasing 0.05-0.1 in average similarities in a wireless environment, on the other hands AODV worked better than OLSR in the medium network and dense network. So the wireless network environment could lead to a decline in predicted results in trend prediction. In section 5.3.2, the predicted results of similarity with medium density network were higher than the dense network in total average with wireless environments but it was not a significant difference with no wireless environment. The results of section 5.3.3, the predicted results of similarities were not much difference between each observation node in medium density network with wireless environments. But there was a difference in the dense network when the observation node was nearby the objective road that performs significantly better than the distant observation node of the objective road. So in the dense network, different observation nodes could affect the predicted results. In section 5.3.4, the results of the shorter objective road showed better prediction accuracy than results of the longer objective road in all aspects by networks, number of nodes and time periods changing. In section 5.3.5, the predicted results of similarities had significant decreasing with time period was increased in each experiment. In section 5.3.6, the predicted results of peak time and off-peak were higher than normal and random traffic distributions. The PPM-C2C framework has the optimum performance of prediction, when the data can be obtained continuously, especially the accurate latest data. In section 5.3.7, through the PPM-C2C framework to predict the travel time, the instant speed has the greatest impact on the results. The influencing factors of travel time prediction can be considered as equivalent to influencing factors of travel speed prediction. In section 5.3.8, the predicted results can be maintained at a relatively high level with traffic density prediction at peak and off-peak hours. In section 5.3.9, the PPM with PRAWMA has a big improvement for the C2C framework in each condition that compares with ARIMA and KNN model in the C2C environment.

Therefore, the PPM-C2C framework can be used for prediction of traffic conditions at a short time such as travel speed, travel time and traffic density.

Meanwhile the similarities of PPM predicted results are influenced by wireless network environment, selection of objective road section and time period, less affected by different mobility patterns, density network and routing protocols.

5.4.2 MRErc of Predicted Results

In respect of MRErc, there was not much impact on the accuracy of predicted values in a wireless environment, not a significant difference with different mobility patterns and also didn't have a visible difference for observation location in different density networks. However, MRErc is influenced by the length of objective road section and time period. The predicted results of MRErc are more accurate when the length of objective road section is short and they will rise when the time period is increased. And also if the data can be obtained continuously, especially the accurate latest data, the predicted results will be more accurate with the longer time period. Finally, the PPM with PRAWMA compares to ARIMA and KNN that it has 10% improvement. The prediction accuracy maintains below 50% for a short-term prediction in general.

5.5 Discussion of Findings

5.5.1 The Finding of the Research

In the current implementation of this work, we proposed a novel traffic prediction framework and designed a pervasive prediction model to traffic conditions estimation based on the Car to Car communication. Each vehicle can collect related traffic data such as position and velocity from other vehicles. All collected data is temporarily stored in the corresponding table with road name by time order, and each table will be momentarily updated as new data become available. The proposed prediction model uses small samples of those data to estimate traffic conditions, especially space mean speed of a road section. In the absence of communication base station on the road, the results obtained using a small sample of ad-hoc data show the feasibility of our approach and proposed prediction model. This approach also can introduce other time series prediction model, at the same time the PPM-C2C framework can be used to address other traffic conditions as prediction parameters according to the demand. If other time series prediction models are used based on this approach, we recommend increasing the number of the samples with their characteristics. The PPM-C2C also can support different routing protocols in VANET to wireless communications. To improve the performance of the PPM with PRAWMA scheme, we suggest a more efficient routing algorithm (Li and Peytchev 2010, Li 2013) to collect the data for it. Therefore, this framework (PPM-C2C) and the proposed model (PPM) have advantages such as simple structure and modelling, good applicability and flexibility, nice commonality, strong expansibility.

Compared to traditional prediction method and current prediction model (Laisheng, et al. 2009, Xie, A and Wei 2013, Zhang, et al. 2013, Xie, Jiang and Wei 2014, Xiang, Xu and He May, 2016), this work presents and provides an approach based on small samples and dynamic predictions in each individual vehicle to avoid the historical data training and statistical testing. In addition to data collection, that literature was based on the roadside sensor collection for short time observations to generate the prediction model. On the other hands, some of the studies (Kim, Sridhara and Bohacek 2009, Alotaibi and Mukherjee 2012) used the mesh network to collect data. For the sources of data, an improved of wireless ad-hoc data collection focus on diversity with the data type, greater flexibility and faster network establishment. Wireless mesh network (WMN) has access nodes and client nodes, but ad-hoc network provides for equality of all nodes with the pervasive algorithm. The ad-hoc network can adapt to the rapid movement of nodes and topology changes to complete the communication between nodes, rather than the user access with lower or static movement. In particular, ad-hoc networks are cheap to deploy and more extensive and flexible, the real-time data by the ad-hoc network is more conducive to traffic conditions prediction and collection in real time. However, the weakness of ad-hoc network is data loss compared against sensor collection. Also, in our experiments, vehicles can communicate with each other, but ad-hoc networks have not been widely used in the real world situations. Thus, we need to consider some vehicles are not in the ad-hoc network, and also prove that incomplete ad-hoc data can be described the traffic conditions of a whole road section in the future studies.

A pervasive prediction model based on the car to car communication that possible would help to collect data and estimate traffic condition in each vehicle. The data collection and computation is not conflicting, the traffic conditions of estimation results will change with the latest data acquisition. The data aggregation function can be found and stored for specific road name and direction with data tables in each vehicle; the model can be generated by this data table for a road section. Because the vehicle cannot always travel in the road network, it would be better to collect the data of each vehicle and estimate for itself, rather than calculate the traffic condition while transmitting. In fact, data aggregation is worth discussing, the literature (He and Zhang 2017) considers that if a data aggregation can be used for the specific data type, the total amount of data to be transmitted can be significantly reduced. We notice from the experiments that the proposed model is over-dependent on the latest data. Therefore, the data aggregation can proceed with a short time period before transmission whenever possible improve the efficiency of the network; meanwhile, the latest data send on demand to improve the performance of the model (AlOrabi, et al. 2016).

Besides the wireless environment that discussed and tested in section 5.3, the predicted results are also affected by some other issues. In the study of (Xie, Jiang and Wei 2014), they established the prediction model by horizontal radius and the longitudinal grade of the road with the observed vehicular speeds on the highway, and also pointed out that the geometric characteristics of the road and the driver's behaviour will have an impact on the traffic condition prediction, such as vertical curve of the road and the visibility of driver. On the other hand, in the real world situations, the road network usually include several special nodes, which will not collaborate with others, or even attempt to break the data accuracy. Such as buses, taxis or trucks will stop anywhere by request, and they still send the data to others, in this case, their data would be different with a normal driver on the same road. We assume that there are no special nodes such as that vehicle that participate the data transmission. However, different weights in the model should be determined to make it practical. Therefore, we emphasise particularly on researching the parameters of traffic conditions for our proposed prediction method and model, rather than the relationships between geometric characteristics of the road and the driver's behaviour with traffic conditions.

For the application problems of the PPM-C2C framework, since each vehicle is involved in transmission and calculation, the potential issues are the computation speed of the device. In fact, the onboard computer with high CPU, big hard drive and integrated graphics deliver a calculation and function almost as powerful as a normal computer. Also, if considering the cost of onboard device, the drivers can also utilise applications in the smartphone, even use mobile cloud partitioning to speed up the computation (Dinh, et al. 2013, Yang, et al. 2013).

5.5.2 The Potential Impact of the Research

The finding of this research, as mentioned previously, the traffic prediction based on the C2C communication becomes a promising prospect since it allows efficient transmissions, economical deployment and large capacity operation via units of vehicles. However, the application of C2C communication is not yet practical to current technology. The aim of this research is to deal with the real-time traffic information based on the C2C communication. Therefore, it could offer a potential way to employ the application of C2C communication in the cars from theory to practice. On the other hand, we also believe that this research has a potential impact on academia, technology, economy and society.

In term of academic impact, the PPM-C2C framework can support different prediction models and different routing protocols in VANET. Future potential research directions can be targeted to develop the model to improve prediction accuracy and extend predicted time period, or develop efficient routing protocol to improve the prediction accuracy. For example, in order to prediction accuracy the weights can be set for different vehicle types, or the model can be added to upstream and downstream road parameters; to maximise data collection, vehicles such as bus and taxi that are in the road network for a long time can be sink nodes and control repeat forwarding, thereby increasing transmission efficiency. Moreover, it is also attractive to use real data to investigate predicted results via simulations. With the increase of deployment rate in the real world, the prediction is made possible using real data via C2C communication.

In term of technical impact, the PPM-C2C framework can implement the C2C communication using any wireless networking technology, such as Wi-Fi, ZigBee, Cellular or LTE. The latest technology for VANET is Visible Light Communication (VLC). The prediction model can be embedded into a variety of mobile wireless communication devices as an expansion function. On the other hand, the autonomous car technology can also introduce the PPM-C2C framework and to help the cars reference to road conditions and to take action in advance with sensing its environment. In particular, when an emergency event occurs ahead, the autonomous system can reduce the speed in advance to avoid a collision.

In term of economic impact, expensive roadside communications facilities can be replaced by low-cost onboard communications equipment. In this way, the costs of urban development and maintenance can be greatly reduced. Personally, the driver can choose the route based on the predicted results of the road sections to avoid congestion, while maintaining a reasonable speed during driving, this will save time and money.

In term of social impact, the traffic prediction is mainly to solve the problem of traffic safety and congestion, and guide people to travel and route selection. Traffic prediction based on C2C communication can provide road conditions efficiently, while significantly improving traffic safety and traffic capacity, also reducing fuel consumption and gas emissions.

6 Chapter 6

Conclusions and Future Work

Overall, this research aimed to propose and evaluate a novel traffic prediction framework with a pervasive prediction model based on the C2C communication in real city scenarios, such as the characteristics of VANET application and external factors that can influence the performance of prediction model. The pervasive prediction model was tested by applying simulation wireless techniques. The simulation and results showed that the prediction model could provide short time prediction for traffic condition and support ad-hoc data from C2C communications with real city scenarios. The research aim and objectives as listed in section 1.3 were investigated, and the viability of the objectives of designing real city scenario, designing a new prediction model and the performance and influencing factors of prediction model was proved via theoretical analysis and experimental evaluation.

6.1 Motivation

The short time traffic prediction is a significant technology of traffic controlling, vehicle guidance and traffic safety in ITS. Previous prediction models for traffic had the disadvantage of lengthy operations with a large amount of historical data and low precision. The data collections are mostly roadside facilities and labour. To improve the timeliness of data and avoid complex operations, traffic prediction is in urgent need of modern wireless communication technology to obtain real-time traffic information. C2C communication has broader prospect and application in technology, efficiency, and economy. It uses VANET technologies to collaborate between vehicles in generating a dynamic, self-organised, flexible, and no roadside unit network, to ensure the efficiency and reliability of ad-hoc data transmission, while data collection, data processing, and prediction calculation performed in the cars. Consequently, the establishment of a traffic prediction model based on the ad-hoc data is proposed and empirically evaluated through using a pervasive prediction model incorporating an instance of Nottingham city centre scenario. Using a real scenario as the tools that reflect reality is the best method for evaluating applications in VANET.

This research proposes a novel traffic prediction framework (PPM-C2C) that is a Pervasive Prediction Model based on the C2C communication. The framework utilises ad-hoc data via C2C communications for short time traffic prediction in each car. It is used to support in C2C applications with less ad-hoc data as historical data to avoid large amounts of computation. Due to the characteristics of moving vehicles, PPM-C2C adopts dynamic identification reasonable range to eliminate invalid results, update the latest data and keep possible results for prediction.

6.2 Review of Contributions

A novel traffic prediction framework (PPM-C2C) is proposed in this project, which is Pervasive Prediction Model (PPM) based on the C2C communication. The framework utilises ad-hoc data via C2C communications for a short time traffic prediction in each car. The contributions are listed in section 1.5 and a review of findings of this thesis are summarised at below.

6.2.1 Design of an In-Car traffic simulation model based on ad-hoc data

Traffic prediction was identified as an important issue, and it needs to consider many influencing factors of each participant (as described in section 2.3.1 and 2.3.2). Most of the traffic prediction models are established for various demands, on the other hand, they need a large amount of historical data and complex modelling, as well as the lack of flexibility which are drawbacks of these models. The PPM prediction model with PRAWMA prediction scheme is a combination prediction model and is designed for traffic condition prediction in short time based on the ad-hoc data in the car; it avoids a significant amount of traffic data and complex modelling by allowing the mechanism of screening multi-results as name the adjusting range implies. The Adjusting range borrows from the fuzzy regression concept and obtains results by empiricism and experiment. It can dynamically determine the validity of polynomial regression results and prevents too large or negative predicted results to make significant errors while ensuring that the results are in reasonable range. The adjusting range cannot guarantee 100% accuracy, but introducing it greatly improves the prediction accuracy as shown in section 3.3.2.

PPM is designed for the wireless ad-hoc data, in this case the traffic data is updated instantly so that there it is ensured that the prediction scheme uses the latest data. In particular, the scheme has predicted results from polynomial regression and weight moving average; thus the predicted results can be determined whether the screening polynomial regression results are successful or not. The historical traffic conditions and current traffic conditions are considered by using the fixed weights to analyse the short time travel speed prediction.

The PRAWMA prediction scheme does not need training with large historical data but only needs a small sample size of the ad-hoc data in real time which will be possible to forecast the traffic conditions for a road section. In the case of each vehicle, as traffic participants have the model when one vehicle receives some data for a road section from other vehicles, the model can start working and forecast the traffic conditions for this road section. During this process, each vehicle is required to send its information, forwarding other vehicle’s information, receiving the information and computing the predicted results. The model is especially aimed at C2C communication via wireless technology in city transport, in particular without any traditional wired connections or roadside infrastructure that data transmission and collection occur between vehicles via an ad-hoc routing protocol.

Additionally, with the same traffic condition and some historical data, the performance of PRAWMA prediction scheme is significantly better than ARIMA as a traditional time series prediction model in all evaluative aspects. The discussions in theory and experiment are shown in section 3.4 for this comparison.

6.2.2 Generation of a Traffic Message Delivery Algorithm with Real City Scenario Mobility Model

The performance of the proposed PPM-C2C framework is validated using a real city scenario. The TMDA by C2C communication will be a platform of the proposed traffic prediction framework to deliver the messages, and the traffic prediction model achieves the goals of efficient traffic prediction. Thus we consider TMDA based on the real city scenario mobility model. We obtain the layout of the Nottingham city centre map from OpenStreetMap and traffic simulations using the well-known SUMO traffic simulator. In the traffic network, the nodes are assumed to be vehicles with original dynamic movement. These nodes are distributed in the road network with different travel routes from the mobility model. Mobility pattern consists of vehicle information, road information, route information and time. The mobility pattern includes sparse, medium and dense traffic in 1200m x 1000m sized area of Nottingham city centre.

The viability of C2C is investigated in above real city traffic scenarios by NS3 simulation technology so that the message is successfully generated and delivered during the 5-layers network model (section 2.2.4). NS3 can provide a trace file for each layer. At the application layer, there is a message format designed for traffic information sharing which is Traffic Speed Message (TSM) (in section 4.3.3). TSM focus on sharing speed information for the prediction model. Meanwhile, an algorithm of message delivery is created to accommodate the wireless communication environment and estimate message behaviour (e.g. forwarding, dropping or using).

Afterwards, the mobility pattern is imported into NS3 to achieve C2C communication. The responsibility of each vehicle is sending their information which is encapsulated in TSM (e.g. position, time, id and speed), forwarding and receiving others information. TSM is delivered in an efficient and reliable approach via routing protocol (AODV and OLSR). Those TSMs are stored in a table in each vehicle and ordered by road sections (in section 3.2.5 and section 4.3.4). Once the vehicle receives TSM data for the prediction model required, the PRAWMA prediction scheme can start to forecast the traffic conditions of the road section. The PPM-C2C framework and its workflow presented in the section 4.3.5.

6.2.3 Evaluation of the influencing factors in the traffic prediction framework

In the research of traffic prediction, there is no single prediction framework and algorithm that can be adapted to all conditions. Therefore, studies on appropriate applying conditions for PPM with PRAWMA prediction scheme in the PPM-C2C framework is mainly discussed in this thesis. The predicted results are probably constrained and influenced by many factors. The evaluations of underlying determinants will be listed and considered in no wireless and wireless environment for the proposed prediction model, such as routing protocol, mobility pattern, observation location, objective road, time period, peak time and other traffic conditions. According to the experimental evaluation, the predicted results via PPM of similarities in the wireless network environment are influenced by the selection of objective road section and time period, less affected by different mobility patterns, density network and routing protocols; The predicted results of MREr c are more accurate when the length of objective road section is short, and they will rise when the time period is increased. Also, traffic density prediction and traffic time prediction are also feasible using by the PPM. With the same traffic condition, the performance of PRAWMA prediction scheme is significantly better than ARIMA and KNN prediction model in all evaluative aspects based on the PPM-C2C framework.

6.3 Conclusion

The application for C2C communication has a promising future in the real world. Data collection, data transmission, and data application are enabling technologies for the current stage. The popularity of these technologies will help to improve traffic conditions, traffic safety, and guided travel. It features low development and operation cost, easy expansion, flexible and fast networking.

For traffic prediction in each vehicle based C2C communication, we have studied in two steps to building the prediction framework in this project. First, we generate pervasive traffic simulation model with real city scenario where the mobile nodes have random movement and route behaviour. In addition to that, a prediction model was proposed based on the data of the mobility pattern for traffic conditions prediction in a short time which the performance of it can be verified with the full data. We have focused on the travel speed prediction of the road section where space mean speed was used.

Secondly, for data collection, data transmission and data application in each vehicle based C2C communication, we have created a wireless simulation environment with the real city mobility scenario to study and compare four routing protocols and chose two of them. We focus on using real-time traffic data to estimate the traffic condition in the future time. Hence, we defined related traffic parameters and designed a data table to store and process the traffic data. As the same time, we designed a forwarding and deletion mechanism to optimise processing ability of the obtain messages. We have examined the possibility of the proposed prediction model in the ad-hoc network for traffic condition prediction. We also pay special attention to deal with the problems introduced by the highly dynamic topology and wireless communication impact, which is crucial to the feasibility of the prediction model. Hence, we have explored several factors that might affect the performance of prediction. Also, we also have conducted the extensive evaluation with traffic density prediction and traffic time prediction. The evaluations have confirmed that the proposed prediction framework and prediction model are feasible for a short time based on the C2C communications.

In summary, this project has advanced the traffic prediction framework and method of VANET applications completely based on C2C communications regarding data collection, data transmission, and data application. It has also demonstrated the influencing factors and drawbacks of the prediction framework and prediction model for the future improvements.

6.4 Future Work

The research work conducted in this thesis indicates that PPM-C2C framework can forecast short-time traffic conditions of a road section with ad-hoc data in a real city scenarios. Although the potential difference between real world and simulation environment needs further investigation, the PPM-C2C can be seen as a promising method for VANET applications. Applying it has produced interesting results which were reported in this thesis. The advantages of experiments are that it can ensure same configurations and conditions, while we can also change the simulation environment, such as the number of vehicles and the communication between vehicles. On the other hand, the simulation is a viable method economically. As wireless mobile technology seems to be consistently adopted by automobile manufacturers and corporations of electronic communications, the traffic prediction model assumes that vehicles will be equipped with in-car devices in the future.

As the results have shown in the thesis, there is some room for improvement to the accuracy of predicted results for longer periods. This means that optimization can consider the road section condition of upstream and downstream, and need to merge other predicted techniques such as Neural Network to improve the predicted accuracy, this is an important area for future research.

In the real world, there are many types of vehicles, such as cars, buses, taxis, trucks and so on. Each unit has its characteristics for movement. Thus, the maximum travel speed of each unit is different. Moreover, each driver has different driving habits and driving experience. For example, when a bus stops at a station and other vehicles are normal driving and sending messages, but the messages for the bus is 0, which may affect the predicted results. Therefore, investigating such as these abnormal data could lead to increased accuracy of predicted results in the real world.

The PPM with PRAWMA prediction scheme uses fixed weight instead of self-adaptive weight. This can increase the computational efficiency, the self-adaptive weighted can be refined classification according to the characteristics of travel speed and vehicle but it could be lead to complex computing and modelling. We have already discussed that weights have a critical influence on the evaluation of predicted results in section 3.2.6. The investigation of weight problem can provide some solutions that increase the accuracy of predicted results. Particularly, the weight of speed is decided by the influence of the different types of vehicles as mentioned previously. For example, based on the current knowledge, the highest weight is given to the recent data, in this case the recent data also can be refined by bus, car, taxi or truck. Cars could be the highest weight than other units because of they usually normal driving and a large number of it. Buses, taxis and trucks might have less weight due to their driving behaviour. However, specific weight assignment still needs to be determined by the investigation in future.

The PPM-C2C framework is designed for the ad-hoc data in the wireless environment. The performance of wireless environment is also important for the accuracy of predicted results as discussed in section 5.3.1 which provides C2C communication for traffic condition prediction. One kind of superior performance routing algorithms can ensure data transmission in timely and efficient. Also, as discussed in the experimental section, the prediction model is too dependent on the latest data. An ad-hoc routing protocol is established for the prediction model that is necessary and potential directions for future research.

In the city traffic scenarios, many factors impact predicted results and message deliveries, as discussed in section 5.3 and section 5.4. The city scenarios are usually congested and travel speed changes fast compared with highways. The effect of stable speed-based is better than might changing anytime and anywhere for the predicted results. Hence, future work could investigate a highway travel speed prediction to estimate in front of the situation in advance for the drivers; this is exciting potential use for traffic safety and an interesting case for future research efforts.

Security issues related to routing and sharing information have not been addressed in this thesis. In fact, some of the drivers do not want their information to be shared. This thesis assumed that all vehicles in the road network could cooperate with others to deliver messages and never hide. In the real world application, the drivers can register a unique virtual account for Car_id to avoid disclosure of real information or even turn off the prediction capabilities and information sharing through the switch of the in-car device if they do not like it. Nevertheless, potential security issues will remain open to further investigations.

In the simulation technology, the simulators SUMO and NS3 are employed in this thesis, and they are compatible in the experiments. However, there are still some problems with lack of the simulators. For example, the vehicle is moving fast in SUMO interface rather than the NS3 interface. On the other hands, SUMO is not perfectly compatible with right-hand drive in the current version of this thesis, but the developers are grappling with this problem for the future versions. With the development of simulation technology and software, the simulation of traffic and traffic communication has broad prospects and practical significance.

Reference

Adams, W.F., 1936. ROAD TRAFFIC CONSIDERED AS A RANDOM SERIES.(INCLUDES PLATES). *Journal of the ICE,* 4 (1), 121-130.

Ade, S., and Tijare, P., 2010. Performance comparison of AODV, DSDV, OLSR and DSR routing protocols in mobile ad hoc networks. *International Journal of Information Technology and Knowledge Management,* 2 (2), 545-548.

Ahmed, M.S., and Cook, A.R., 1979. *Analysis of freeway traffic time-series data by using Box-Jenkins techniques.*

Ali, S., and Ali, A., 2009. Performance analysis of AODV, DSR and OLSR in MANET. *In: Proceedings on Seventh International Conference on Wireless Systems,* pp. 34.

AlOrabi, W.A., Rahman, S.A., El Barachi, M. and Mourad, A., 2016. Towards on Demand Road Condition Monitoring Using Mobile Phone Sensing as a Service. *Procedia Computer Science,* 83, 345-352.

Alotaibi, E., and Mukherjee, B., 2012. A survey on routing algorithms for wireless Ad-Hoc and mesh networks. *Computer Networks,* 56 (2), 940-965.

Altman, N.S., 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician,* 46 (3), 175-185.

Anderson, O.D., 1976. *Time series analysis and forecasting: the Box-Jenkins approach.* Butterworths London.

Arnott, R., De Palma, A. and Lindsey, R., 1991. Does providing information to drivers reduce traffic congestion? *Transportation Research Part A: General,* 25 (5), 309-318.

Asai, H.T.U., 1982. Linear regression analysis with fuzzy model. *IEEE Trans.Systems Man Cybern,* 12, 903-907.

Bai, F., Sadagopan, N. and Helmy, A., 2003. The IMPORTANT framework for analyzing the Impact of Mobility on Performance Of RouTing protocols for Adhoc NeTworks. *Ad Hoc Networks,* 1 (4), 383-403.

Baldessari, R., Festag, A. and Abeillé, J., 2007. Nemo meets vanet: a deployability analysis of network mobility in vehicular communication. *In: Telecommunications, 2007. ITST'07. 7th International Conference on ITS,* IEEE, pp. 1-6.

Bates, J.M., and Granger, C.W., 1969. The combination of forecasts. *Or,* , 451-468.

Behrisch, M., Bieker, L., Erdmann, J. and Krajzewicz, D., 2011. Sumo-simulation of urban mobility-an overview. *In: SIMUL 2011, The Third International Conference on Advances in System Simulation,* pp. 55-60.

Ben-Akiva, M., Bierlaire, M., Koutsopoulos, H. and Mishalani, R., 1998. DynaMIT: a simulation-based system for traffic prediction. *In: DACCORS short term forecasting workshop, The Netherlands,* Citeseer, .

Bettstetter, C., Resta, G. and Santi, P., 2003. The node distribution of the random waypoint mobility model for wireless ad hoc networks. *Mobile Computing, IEEE Transactions On,* 2 (3), 257-269.

Biham, O., Middleton, A.A. and Levine, D., 1992. Self-organization and a dynamical transition in traffic-flow models. *Physical Review A,* 46 (10), R6124.

Bisio, I., Delucchi, S., Lavagetto, F., Marchese, M., Li, L., Jafarkhani, H., Jafar, S.A., Zhang, L., Zhao, G. and Wu, G., 2013. IEEE GLOBECOM 2012 EXPLORES THE "THE MAGIC OF GLOBAL CONNECTIVITY" AT THE DISNEYLAND HOTEL IN ANAHEIM, CALIFORNIA. *IEEE Communications Magazine,* , 17.

Box, J., and Jenkins, G.M., 1994. Reinsel. Time Series Analysis, Forecasting and Control.

Boxill, S.A., and Yu, L., 2000. *An Evaluation of Traffic Simulation Models for Supporting ITS.* Houston.

Breese, J.S., Heckerman, D. and Kadie, C., 1998. Empirical analysis of predictive algorithms for collaborative filtering. *In: Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence,* Morgan Kaufmann Publishers Inc., pp. 43-52.

Car to Car Communication Consortium, 2007. Car 2 car communication consortium manifesto. *Available: Www.Carto-Car.Org,* .

Charles, Z., 2015. ***Basic Concepts of Correlation*** [online]. Real Statistics Using Excel. Available at: http://www.real-statistics.com/correlation/basic-concepts-correlation/ [Accessed 09/2015 2015].

Chen, H., and Grant-Muller, S., 2001. Use of sequential learning for short-term traffic flow forecasting. *Transportation Research Part C: Emerging Technologies,* 9 (5), 319-336.

Chen, S., Qu, G., Wang, X. and Zhang, H., 2003. Traffic flow forecasting based on grey neural network model. *In: Machine Learning and Cybernetics, 2003 International Conference on,* IEEE, pp. 1275-1278.

Cheng, H.T., Shan, H. and Zhuang, W., 2011. Infotainment and road safety service support in vehicular networking: From a communication perspective. *Mechanical Systems and Signal Processing,* 25 (6), 2020-2038.

Cheng, L., Henty, B.E., Stancil, D.D., Bai, F. and Mudalige, P., 2007. Mobile vehicle-to-vehicle narrow-band channel measurement and characterization of the 5.9 GHz dedicated short range communication (DSRC) frequency band. *Selected Areas in Communications, IEEE Journal On,* 25 (8), 1501-1516.

Chlamtac, I., Conti, M. and Liu, J.J., 2003. Mobile ad hoc networking: imperatives and challenges. *Ad Hoc Networks,* 1 (1), 13-64.

Clark, S.D., Dougherty, M.S. and Kirby, H.R., 1993. The use of neural networks and time series models for short term traffic forecasting: a comparative study. *In: TRANSPORTATION PLANNING METHODS. PROCEEDINGS OF SEMINAR D HELD AT THE PTRC EUROPEAN TRANSPORT, HIGHWAYS AND PLANNING 21ST SUMMER ANNUAL MEETING (SEPTEMBER 13-17, 1993), UMIST. VOLUME P363,* .

CORSIM, 2015. *CORSIM: Microscopic Traffic Simulation Model* [online]. University of Florida. Available at: http://www-mctrans.ce.ufl.edu/featured/TSIS/Version5/corsim.htm [Accessed MAY/05 2015].

Cronin, M.J., 2014. Ford Finds Its Connection. *In:* Ford Finds Its Connection. *Top Down Innovation.* Springer, 2014, pp. 13-24.

Crow, B.P., Widjaja, I., Kim, J.G. and Sakai, P.T., 1997. IEEE 802.11 wireless local area networks. *Communications Magazine, IEEE,* 35 (9), 116-126.

Cue, E., Cook, T., Federighi, C. and Ive, J., June 12, 2013. WWDC 2013 Keynote. *In: WWDC 2013 Keynote, Moscone West, San Francisco, California, June 10, 2013.* Apple Inc., .

Dar, K., Bakhouya, M., Gaber, J., Wack, M. and Lorenz, P., 2010. Wireless communication technologies for ITS applications [Topics in Automotive Networking]. *Communications Magazine, IEEE,* 48 (5), 156-162.

Dashtinezhad, S., Nadeem, T., Dorohonceanu, B., Borcea, C., Kang, P. and Iftode, L., 2004. TrafficView: a driver assistant device for traffic monitoring based on car-to-car communication. *In: Vehicular Technology Conference, 2004. VTC 2004-Spring. 2004 IEEE 59th,* IEEE, pp. 2946-2950.

Davis, G.A., and Nihan, N.L., 1991a. Nonparametric regression and short-term freeway traffic forecasting. *Journal of Transportation Engineering,* 117 (2), 178-188.

Davis, G.A., and Nihan, N.L., 1991b. Nonparametric regression and short-term freeway traffic forecasting. *Journal of Transportation Engineering,* 117 (2), 178-188.

Davis, G.A., Nihan, N.L., Hamed, M.M. and Jacobson, L.N., 1990. Adaptive forecasting of freeway traffic congestion. *Transportation Research Record,* (1287).

Dia, H., 2001. An object-oriented neural network approach to short-term traffic forecasting. *European Journal of Operational Research,* 131 (2), 253-261.

Díaz, J., Mitsche, D. and Pérez-Giménez, X., 2009. Large connectivity for dynamic random geometric graphs. *Mobile Computing, IEEE Transactions On,* 8 (6), 821-835.

Ding, A., Zhao, X. and Jiao, L., 2002. Traffic flow time series prediction based on statistics learning theory. *In: Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on,* IEEE, pp. 727-730.

Ding, G., and Bhargava, B., 2004. Peer-to-peer file-sharing over mobile ad hoc networks. *In: Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on,* IEEE, pp. 104-108.

Dinh, H.T., Lee, C., Niyato, D. and Wang, P., 2013. A survey of mobile cloud computing: architecture, applications, and approaches. *Wireless Communications and Mobile Computing,* 13 (18), 1587-1611.

Disbro, J.E., and Frame, M., 1989. *Traffic flow theory and chaotic behavior.*

Ditlevsen, O., 1994. Traffic loads on large bridges modeled as white-noise fields. *Journal of Engineering Mechanics,* 120 (4), 681-694.

Durbin, J., and Watson, G.S., 1951. Testing for serial correlation in least squares regression. II. *Biometrika,* , 159-177.

Eagle, N., and Pentland, A., 2006. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing,* 10 (4), 255-268.

Eichler, S., 2007. Performance evaluation of the IEEE 802.11 p WAVE communication standard. *In: Vehicular Technology Conference, 2007. VTC-2007 Fall. 2007 IEEE 66th,* IEEE, pp. 2199-2203.

Eichler, S., Schroth, C. and Eberspächer, J., 2006. Car-to-car communication. *In: Proceedings of the VDE-Kongress-Innovations for Europe,* Citeseer, pp. 35-40.

Faezipour, M., Nourani, M., Saeed, A. and Addepalli, S., 2012. Progress and challenges in intelligent vehicle area networks. *Communications of the ACM,* 55 (2), 90-100.

Fall, K., and Varadhan, K., 2007. The network simulator–NS-2. *URL: Http://Www.Isi.Edu/Nsnam/Ns,* .

Federal Communications Commission, 2002. In the Matter of Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems. *First Report and Order in ET Docket 98-153,* .

Feigenbaum, M.J., 1978. Quantitative universality for a class of nonlinear transformations. *Journal of Statistical Physics,* 19 (1), 25-52.

Fernandez, J.A., Borries, K., Cheng, L., Kumar, B.V., Stancil, D.D. and Bai, F., 2012. Performance of the 802.11 p physical layer in vehicle-to-vehicle environments. *Vehicular Technology, IEEE Transactions On,* 61 (1), 3-14.

Fiore, M., 2008. Vehicular mobility and network simulation. *Handbook on Vehicular Networks.Abingdon: Taylor and Francis,* .

Fiore, M., and Härri, J., 2008. The networking shape of vehicular mobility. *In: Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing,* ACM, pp. 261-272.

Fonseca, E., Festag, A., Baldessari, R. and Aguiar, R.L., 2007. Support of anonymity in vanets-putting pseudonymity into practice. *In: Wireless Communications and Networking Conference, 2007. WCNC 2007. IEEE,* IEEE, pp. 3400-3405.

GAMATI, E., 2013. *Information Collection Algorithms for mobile ad-hoc networks (Application domain: Urban Traffic Wireless Mobile Ad-Hoc Networks (MANETs)).* NOTTINGHAM TRENT UNIVERSITY.

Gazis, D.C., Herman, R. and Potts, R.B., 1959. Car-following theory of steady-state traffic flow. *Operations Research,* 7 (4), 499-505.

Ghosekar, P., Katkar, G. and Ghorpade, P., 2010. Mobile ad hoc networking: imperatives and challenges. *IJCA Special Issue on "Mobile Ad-Hoc Networks" MANETs, .*

Gipps, P.G., 1986. A model for the structure of lane-changing decisions. *Transportation Research Part B: Methodological,* 20 (5), 403-414.

Gipps, P.G., 1981. A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological,* 15 (2), 105-111.

Grassberger, P., and Procaccia, I., 2004. Measuring the strangeness of strange attractors. *In:* Measuring the strangeness of strange attractors. *The Theory of Chaotic Attractors.* Springer, 2004, pp. 170-189.

Grassberger, P., and Procaccia, I., 1983. Characterization of strange attractors. *Physical Review Letters,* 50 (5), 346.

Greenberg, H., 1959. An analysis of traffic flow. *Operations Research,* 7 (1), 79-85.

Gross, K.C., Digate, C.J. and Lee, E.H., 1994. *Event-Driven Rule-Based Messaging System, .*

Grzeszczyk, R., Merkisz, J., Bogus, P. and Kaminski, T., 2009. Methods and Procedures for Testing the eCall In-Vehicle Unit for the Purpose of Its Performance Assessment and Certification. *In: 21st International Technical Conference on the Enhanced Safety of Vehicles ESV, .*

Guan-jun, T., 2000. The Structure Method and Application of Background Value in Grey System GM (1, 1) Model (Ⅰ)[J]. *Systems Engineering-Theory & Practice,* 4, 018.

Haerri, J., Filali, F. and Bonnet, C., 2006. Performance comparison of AODV and OLSR in VANETs urban environments under realistic mobility patterns. *In: Proceedings of the 5th IFIP mediterranean ad-hoc networking workshop,* pp. 14-17.

Hall, F.L., 1996. Traffic stream characteristics. *Traffic Flow Theory.US Federal Highway Administration, .*

Han, G., 2011. ***Urban Traffic Flow Simulation SUMO-based Platform***. V1.3 ed. Guizhou University, China: School of Computer ScienceNLPLAB.

Harri, J., Filali, F. and Bonnet, C., 2009. Mobility models for vehicular ad hoc networks: a survey and taxonomy. *Communications Surveys & Tutorials, IEEE,* 11 (4), 19-41.

Harri, J., Filali, F. and Bonnet, C., 2005. A framework for mobility models generation and its application to inter-vehicular networks. *In: Wireless Networks, Communications and Mobile Computing, 2005 International Conference on,* IEEE, pp. 42-47.

Härri, J., Filali, F., Bonnet, C. and Fiore, M., 2006. VanetMobiSim: generating realistic mobility patterns for VANETs. *In: Proceedings of the 3rd international workshop on Vehicular ad hoc networks,* ACM, pp. 96-97.

Hartenstein, H., and Laberteaux, K., 2010. *VANET: vehicular applications and inter-networking technologies.* Wiley Online Library.

Hartenstein, H., and Laberteaux, K.P., 2008. A tutorial survey on vehicular ad hoc networks. *Communications Magazine, IEEE,* 46 (6), 164-171.

Harvey, A.C., 1990. *Forecasting, structural time series models and the Kalman filter.* Cambridge university press.

He, Z., Cao, B. and Liu, Y., 2015. Accurate real-time traffic speed estimation using infrastructure-free vehicular networks. *International Journal of Distributed Sensor Networks,* .

He, Z., Cao, J. and Li, T., 2012. Mice: A real-time traffic estimation based vehicular path planning solution using vanets. *In: Connected Vehicles and Expo (ICCVE), 2012 International Conference on,* IEEE, pp. 172-178.

He, Z., and Zhang, D., 2017. Cost-efficient traffic-aware data collection protocol in VANET. *Ad Hoc Networks,* 55, 28-39.

Herman, R., Lam, T.N. and Prigogine, I., 1971. *Kinetic theory of vehicular traffic: comparison with data.* Research Laboratories, General Motors Corporation.

Ho, I.W., Leung, K.K. and Polak, J.W., 2011. Stochastic model and connectivity dynamics for VANETs in signalized road systems. *IEEE/ACM Transactions on Networking (TON),* 19 (1), 195-208.

Ho, Y.H., Ho, A.H. and Hua, K.A., 2008. Routing protocols for inter-vehicular networks: A comparative study in high-mobility and large obstacles environments. *Computer Communications,* 31 (12), 2767-2780.

Hochmair, H.H., and Zielstra, D., 2011. An Introduction to Freely Available Street Network Data.

Hoebeke, J., Moerman, I., Dhoedt, B. and Demeester, P., 2004. An overview of mobile ad hoc networks: Applications and challenges. *Journal-Communications Network,* 3 (3), 60-66.

Hong, W., 2011. Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm. *Neurocomputing,* 74 (12), 2096-2107.

Hsu, W., and Helmy, A., 2005. MobiLib USC WLAN trace data set. *Download from Http://Nile.Cise.Ufl.Edu/MobiLib/USC_trace,* .

Huang, H., and Zhang, S., 2013. A routing algorithm based on dynamic forecast of vehicle speed and position in vanet. *International Journal of Distributed Sensor Networks,* .

Hunter, J.S., 1986. The exponentially weighted moving average. *J.Quality Technol.,* 18 (4), 203-210.

IEEE Computer Society LAN MAN Standards Committee, 1997. Wireless LAN medium access control (MAC) and physical layer (PHY) specifications.

IEEE Vehicular Technology Society, 2010. *IEEE Standard for Wireless Access in Vehicular Environments (WAVE)-Networking Services.* 1609.3-2010 ed. IEEE Std.

Ishak, S., and Al-Deek, H., 2002. Performance evaluation of short-term time-series traffic prediction model. *Journal of Transportation Engineering,* 128 (6), 490-498.

Jacquet, P., Laouiti, A., Minet, P. and Viennot, L., 2001a. Performance analysis of OLSR multipoint relay flooding in two ad hoc wireless network models.

Jacquet, P., Muhlethaler, P., Clausen, T., Laouiti, A., Qayyum, A. and Viennot, L., 2001b. Optimised link state routing protocol for ad hoc networks. *In: Multi Topic Conference, 2001. IEEE INMIC 2001. Technology for the 21st Century. Proceedings. IEEE International,* IEEE, pp. 62-68.

Jeong, J., Guo, S., Gu, Y., He, T. and Du, D.H., 2011. Trajectory-based data forwarding for light-traffic vehicular Ad Hoc networks. *IEEE Transactions on Parallel and Distributed Systems,* 22 (5), 743-757.

Jessop, A., 1990. *DECISION AND FORECASTING MODELS. WITH TRANSPORT APPLICATIONS.*

Jiang, D., and Delgrossi, L., 2008. IEEE 802.11 p: Towards an international standard for wireless access in vehicular environments. *In: Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE,* IEEE, pp. 2036-2040.

Jiang, M., WU, C., Zhang, M. and HU, D., 2009. Research on the comparison of time series models for network traffic prediction. *Acta Electronic Sinica,* 37 (11), 2353-2358.

Jing, W., and Le, H., 2011. The Research of Ad Hoc Routing Protocols and NS2 Simulation (chinese: Ad Hoc 网络路由协议分析与 NS2 仿真模拟, 荆文礼;乐红兵;). *China SciencePaper Online (中国科技论文在线),* 4, 02-14.

Johnson, D.B., and Maltz, D.A., 1996. Dynamic source routing in ad hoc wireless networks. *In:* Dynamic source routing in ad hoc wireless networks. *Mobile computing.* Springer, 1996, pp. 153-181.

Jui-Hung, Y., Jyh-Cheng, C. and Chi-Chen, L., 2003. WLAN Standards: in particular, the IEEE 802.11 family. *Potentials, IEEE,* 22 (4), 16-22.

Kalašová, A., and Krchová, Z., 2011. Effect of aggressive driving on formation of congestion. *In:* Effect of aggressive driving on formation of congestion. *Modern Transport Telematics.* Springer, 2011, pp. 218-225.

Karande, M.N.D., and Kulkarni, M.K.K., 2013. Efficient routing protocols for vehicular adhoc network. *In: International Journal of Engineering Research and Technology,* ESRSA Publications, .

Karnadi, F.K., Mo, Z.H. and Lan, K., 2007. Rapid generation of realistic mobility models for VANET. *In: Wireless Communications and Networking Conference, 2007. WCNC 2007. IEEE,* IEEE, pp. 2506-2511.

Karp, B., and Kung, H., 2000. GPSR: Greedy perimeter stateless routing for wireless networks. *In: Proceedings of the 6th annual international conference on Mobile computing and networking,* ACM, pp. 243-254.

Kesting, A., Treiber, M. and Helbing, D., 2010. Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity. *Philosophical Transactions.Series A, Mathematical, Physical, and Engineering Sciences,* 368 (1928), 4585-4605.

Kim, H., Lee, D.H., Lee, J.M., Lee, K., Lyu, W. and Choi, S., 2008. The QoE evaluation method through the QoS-QoE correlation model. *In: Networked Computing and Advanced Information Management, 2008. NCM'08. Fourth International Conference on,* IEEE, pp. 719-725.

Kim, J., and Bohacek, S., 2005. A survey-based mobility model of people for simulation of urban mesh networks. *Proc.MeshNets,* .

Kim, J., Sridhara, V. and Bohacek, S., 2009. Realistic mobility simulation of urban mesh networks. *Ad Hoc Networks,* 7 (2), 411-430.

Kirby, H.R., Watson, S.M. and Dougherty, M.S., 1997. Should we use neural networks or statistical models for short-term motorway traffic forecasting? *International Journal of Forecasting,* 13 (1), 43-50.

Kmjzewicz, D., Hertkorn, G., Wagner, P. and Rossel, C., 2002. Simulation of Urban Mobility (SUMO): An Opensource Traffic Simulation. *In: Proc. of the 4th Middle East Symposium on Simulation and Modeling. Sharjah, United Arab Emirates:[sn],* .

Knoop, V., Hoogendoorn, S.P. and van Zuylen, H., 2009. Empirical differences between time mean speed and space mean speed. *In:* Empirical differences between time mean speed and space mean speed. *Traffic and Granular Flow'07.* Springer, 2009, pp. 351-356.

Kohonen, T., 1990. The self-organizing map. *Proceedings of the IEEE,* 78 (9), 1464-1480.

Kosch, T., Schroth, C., Strassberger, M. and Bechler, M., 2012. *Automotive Inter-networking.* John Wiley & Sons.

Krajzewicz, D., Brockfeld, E., Mikat, J., Ringel, J., Rössel, C., Tuchscheerer, W., Wagner, P. and Wösler, R., 2005a. Simulation of modern traffic lights control systems using the open source traffic simulation SUMO. *In: Proceedings of the 3rd industrial simulation conference,* .

Krajzewicz, D., Hartinger, M., Hertkorn, G., Mieth, P., Rössel, C., Zimmer, J. and Wagner, P., 2005b. Using the Road Traffic Simulation "SUMO" for educational Purposes. *In:* Using the Road Traffic Simulation "SUMO" for educational Purposes. *Traffic and Granular Flow'03.* Springer, 2005b, pp. 217-222.

Krauß, S., 1998. *Microscopic Modeling of Traffic Flow: Investigation of Collision Free Vehicle Dynamics,* .

Krishnavedala, 2011. *Illustration of least squares fitting.* wikipedia.

Laisheng, X., Xiaohong, P., Zhengxia, W., Bing, X. and Pengzhi, H., 2009. Research on traffic monitoring network and its traffic flow forecast and congestion control model based on wireless sensor networks. *In: Measuring Technology and Mechatronics Automation, 2009. ICMTMA'09. International Conference on,* IEEE, pp. 142-147.

Laneman, J.N., 2006. Cooperative communications in mobile ad hoc networks. *IEEE Signal Processing Magazine,* 1053 (5888/06).

Leduc, G., 2008. Road traffic data: Collection methods and applications. *Working Papers on Energy, Transport and Climate Change,* 1, 55.

Lee, S., and Fambro, D.B., 1999. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. *Transportation Research Record: Journal of the Transportation Research Board,* 1678 (1), 179-188.

Li, F., and Wang, Y., 2007. Routing in vehicular ad hoc networks: A survey. *Vehicular Technology Magazine, IEEE,* 2 (2), 12-22.

Li, J., Blake, C., De Couto, D.S., Lee, H.I. and Morris, R., 2001. Capacity of ad hoc wireless networks. *In: Proceedings of the 7th annual international conference on Mobile computing and networking,* ACM, pp. 61-69.

Li, X., Shu, W., Li, M., Huang, H., Luo, P. and Wu, M., 2009. Performance evaluation of vehicle-based mobile sensor networks for traffic monitoring. *IEEE Transactions on Vehicular Technology,* 58 (4), 1647-1653.

Li, Y., 2013. Dynamic wireless mobile framework for distributed collaborative real-time information generation and control systems. *Dynamic Wireless Mobile Framework for Distributed Collaborative Real-Time Information Generation and Control Systems,* .

Li, Y., and Peytchev, E., 2010. Novel ad-hoc wireless mobile communication network routing model for location based sensor networks, in proceedings of International Symposium on LBS & TeleCartography.

LI, Z.TAO, R., *The research and optimization of DSR route protocol based on NS2 (chinese: 基于NS2 下的DSR 路由协议的研究与优化)*
.

Liang, W., Li, Z., Zhang, H., Wang, S. and Bie, R., 2014. Vehicular Ad Hoc Networks: Architectures, Research Issues, Methodologies, Challenges, and Trends. *International Journal of Distributed Sensor Networks,* .

Lieberman, E., and Rathi, A.K., 1997. Traffic simulation. *Traffic Flow Theory,* .

Lovegrove, G.R., and Sayed, T., 2006. Macro-level collision prediction models for evaluating neighbourhood traffic safety. *Canadian Journal of Civil Engineering,* 33 (5), 609-621.

Ma, L., and Jia, D., 2005. The competition and cooperation of WiMAX, WLAN and 3G.

Maerivoet, S., and De Moor, B., 2005. Cellular automata models of road traffic. *Physics Reports,* 419 (1), 1-64.

Manual, H.C., 2000. Transportation research board. *National Research Council, Washington, DC,* 113.

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics,* 11 (2), 431-441.

Martinez, F.J., Cano, J., Calafate, C.T. and Manzoni, P., 2008. Citymob: a mobility model pattern generator for VANETs. *In: Communications Workshops, 2008. ICC Workshops' 08. IEEE International Conference on,* IEEE, pp. 370-374.

Martinez, F.J., Toh, C.K., Cano, J., Calafate, C.T. and Manzoni, P., 2011. A survey and comparative study of simulators for vehicular ad hoc networks (VANETs). *Wireless Communications and Mobile Computing,* 11 (7), 813-828.

May, A.D., 1990. *Traffic flow fundamentals.*

Menouar, H., Filali, F. and Lenardi, M., 2006. A survey and qualitative analysis of MAC protocols for vehicular ad hoc networks. *Wireless Communications, IEEE,* 13 (5), 30-35.

Miao, X., Luo, Z. and Hong, L., 2011. Routing Protocol Simulation Research Based on SUMO. *Computer Engineering,* 1, 038.

Microsoft, 2015. ***CORREL function*** [online]. Microsoft. Available at: https://support.office.com/en-us/article/CORREL-function-995dcef7-0c0a-4bed-a3fb-239d7b68ca92 [Accessed 09/2015 2015].

Milojevic, M., and Rakocevic, V., 2013. Short Paper: Distributed vehicular traffic congestion detection algorithm for urban environments. *In: Vehicular Networking Conference (VNC), 2013 IEEE,* IEEE, pp. 182-185.

Min, W., and Wynter, L., 2011. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies,* 19 (4), 606-616.

Mohapatra, S., and Kanungo, P., 2012. Performance analysis of AODV, DSR, OLSR and DSDV routing protocols using NS2 Simulator. *Procedia Engineering,* 30, 69-76.

Mohimani, G.H., Ashtiani, F., Javanmard, A. and Hamdi, M., 2009. Mobility modeling, spatial traffic distribution, and probability of connectivity for sparse and dense vehicular ad hoc networks. *Vehicular Technology, IEEE Transactions On,* 58 (4), 1998-2007.

Moustafa, H., and Zhang, Y., 2009. *Vehicular networks: techniques, standards, and applications.* Auerbach publications.

Nan, H., 2000. *Statistic Analysis Tutorial of SPSS for Windows. [SPSS for Windows 统计分析教程] (in Chinese).* Beijing, China: Publishing House of Electronics Industry.

Nandan, A., Das, S., Pau, G., Gerla, M. and Sanadidi, M., 2005. Co-operative downloading in vehicular ad-hoc wireless networks. *In: Wireless On-demand Network Systems and Services, 2005. WONS 2005. Second Annual Conference on,* IEEE, pp. 32-41.

Nicholson, H., and Swann, C., 1974. The prediction of traffic flow volumes based on spectral analysis. *Transportation Research,* 8 (6), 533-538.

Nihan, N.L., and Holmesland, K.O., 1980. Use of the Box and Jenkins time series technique in traffic forecasting. *Transportation,* 9 (2), 125-143.

Noureddine, H., and Samira, M., 2014. Efficient Local Density Estimation Strategy for VANETs. *arXiv Preprint arXiv:1402.4508,* .

ns-3, 2015. *Network Simulator 3* [online]. . Available at: http://www.nsnam.org/ [Accessed May/27 2015].

Nzouonta, J., Rajgure, N., Wang, G. and Borcea, C., 2009. VANET routing on city roads using real-time vehicular traffic information. *Vehicular Technology, IEEE Transactions On,* 58 (7), 3609-3626.

Nzouonta-Domgang, J., 2009. *Road-based routing in vehicular ad hoc networks.* New Jersey Institute of Technology.

Okutani, I., and Stephanedes, Y.J., 1984. Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research Part B: Methodological,* 18 (1), 1-11.

Olariu, S., and Weigle, M.C., 2009. *Vehicular networks: from theory to practice.* Crc Press.

Padmavathi, K., and Maneendhar, R., 2012. A Surveying on Road Safety Using Vehicular Communication Networks. *Journal of Computer Applications, EICA,* .

Papagiannaki, K., Taft, N., Zhang, Z. and Diot, C., 2003. Long-term forecasting of internet backbone traffic: Observations and initial models. *In: INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies,* IEEE, pp. 1178-1188.

Payne, H.J., 1979. FREFLO: A macroscopic simulation model of freeway traffic. *Transportation Research Record,* (722).

Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science,* 2 (11), 559-572.

Peng-Jian, S., and Jin-Sheng, S., 2007. Multi-fractal analysis of highway traffic data. *Chinese Physics,* 16 (2), 365.

Perkins, C.E., and Bhagwat, P., 1994. Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers. *In: ACM SIGCOMM Computer Communication Review,* ACM, pp. 234-244.

Perkins, C.E., and Royer, E.M., 1999. Ad-hoc on-demand distance vector routing. *In: Mobile Computing Systems and Applications, 1999. Proceedings. WMCSA'99. Second IEEE Workshop on,* IEEE, pp. 90-100.

Pursula, M., 1999. Simulation of traffic systems-an overview. *Journal of Geographic Information and Decision Analysis,* 3 (1), 1-8.

Qian, Y., and Moayeri, N., 2008. Design of secure and application-oriented VANETs. *In: Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE,* IEEE, pp. 2794-2799.

Quayle, S., and Urbanik, I., 2008. Integrated Corridor Management: Simulation Model Conversion Between VISUM and VISSIM. *In: Transportation Research Board 87th Annual Meeting,* .

Quddus, M.A., 2008. Time series count data models: an empirical application to traffic accidents. *Accident Analysis & Prevention,* 40 (5), 1732-1741.

Ravishanker, N., and Ray, B.K., 2002. Bayesian prediction for vector ARFIMA processes. *International Journal of Forecasting,* 18 (2), 207-214.

Raya, M., Papadimitratos, P. and Hubaux, J., 2006. Securing vehicular communications. *IEEE Wireless Communications Magazine, Special Issue on Inter-Vehicular Communications,* 13 (LCA-ARTICLE-2006-015), 8-15.

Riedi, R.H., Crouse, M.S., Ribeiro, V.J. and Baraniuk, R.G., 1999. A multifractal wavelet model with application to network traffic. *Information Theory, IEEE Transactions On,* 45 (3), 992-1018.

Royer, E.M., and Toh, C., 1999. A review of current routing protocols for ad hoc mobile wireless networks. *Personal Communications, IEEE,* 6 (2), 46-55.

Sasaki, M., and Nagatani, T., 2003. Transition and saturation of traffic flow controlled by traffic lights. *Physica A: Statistical Mechanics and its Applications,* 325 (3), 531-546.

Schollmeier, R., 2001. A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. *In: Peer-to-Peer Computing, 2001. Proceedings. First International Conference on,* pp. 101-102.

Scott, J., Gass, R., Crowcroft, J., Hui, P., Diot, C. and Chaintreau, A., 2006. CRAWDAD Data Set cambridge/haggle (v. 2006-01-31). *CRAWDAD Wireless Network Data Archive,* .

Shakkottai, S., Rappaport, T.S. and Karlsson, P.C., 2003. Cross-layer design for wireless networks. *Communications Magazine, IEEE,* 41 (10), 74-80.

Sichitiu, M.L., and Kihl, M., 2008. Inter-vehicle communication systems: a survey. *Communications Surveys & Tutorials, IEEE,* 10 (2), 88-105.

Skiena, S., 1990. Dijkstra's Algorithm. *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica, Reading, MA: Addison-Wesley,* , 225-227.

Smith, B.L., and Demetsky, M.J., 1997. Traffic flow forecasting: comparison of modeling approaches. *Journal of Transportation Engineering,* 123 (4), 261-266.

Smith, B.L., and Demetsky, M.J., 1994. Short-term traffic flow prediction: neural network approach. *Transportation Research Record,* (1453).

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

Sohrabi, K., Gao, J., Ailawadhi, V. and Pottie, G.J., 2000. Protocols for self-organization of a wireless sensor network. *IEEE Personal Communications,* 7 (5), 16-27.

Stigler, S.M., 1981. Gauss and the invention of least squares. *The Annals of Statistics,* , 465-474.

Su, H., Zhang, L. and Yu, S., 2007. Short-term traffic flow prediction based on incremental support vector regression. *In: Natural Computation, 2007. ICNC 2007. Third International Conference on,* IEEE, pp. 640-645.

SUMO, 2015. *SUMO - Simulation of Urban Mobility* [online]. DLR - Institute of Transportation Systems. Available at: http://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/ [Accessed MAY/05 2015].

Sun, H., Liu, H.X., Xiao, H., He, R.R. and Ran, B., 2003. Short term traffic forecasting using the local linear regression model. *In: 82nd Annual Meeting of the Transportation Research Board, Washington, DC,* .

Tan, P., Steinbach, M. and Kumar, V., 2006. khdaw. com.

Tilak, S., Abu-Ghazaleh, N.B. and Heinzelman, W., 2002. A taxonomy of wireless micro-sensor network models. *ACM SIGMOBILE Mobile Computing and Communications Review,* 6 (2), 28-36.

Tseng, Y., Ni, S., Chen, Y. and Sheu, J., 2002. The broadcast storm problem in a mobile ad hoc network. *Wireless Networks,* 8 (2-3), 153-167.

Tuduce, C., and Gross, T., 2005. A mobility model based on wlan traces and its validation. *In: INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE,* IEEE, pp. 664-674.

Uzcategui, R., and Acosta-Marum, G., 2009. WAVE: a tutorial. *Communications Magazine, IEEE,* 47 (5), 126-133.

Van Der Voort, M., Dougherty, M. and 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.

van Hinsbergen, J., and Sanders, F., 2007. Short Term Traffic Prediction Models.

van Lint, J.W., Hoogendoorn, S. and van Zuylen, H.J., 2002. Freeway travel time prediction with state-space neural networks: modeling state-space dynamics with recurrent neural networks. *Transportation Research Record: Journal of the Transportation Research Board,* 1811 (1), 30-39.

VanetMobiSim, 2015. *VanetMobiSim Project Home Page* [online]. VanetMobiSim. Available at: http://vanet.eurecom.fr/ [Accessed MAY/05 2015].

Vapnik, V.N., and Vapnik, V., 1998. *Statistical learning theory.* Wiley New York.

VISSM, 2015. *PTV simulation - VISSIM* [online]. PTV GROUP. Available at: http://vision-traffic.ptvgroup.com/en-uk/home/ [Accessed MAY/05 2015].

Wang, C., Quddus, M.A. and Ison, S.G., 2011. Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. *Accident Analysis & Prevention,* 43 (6), 1979-1990.

Wang, T., 2013. Vehicular Ad Hoc Networks - A review of vehicle communications. *In: Peking University, 2013/Sep/10.*

Wang, Y., Ahmed, A., Krishnamachari, B. and Psounis, K., 2008. IEEE 802.11 p performance evaluation and protocol enhancement. *In: Vehicular Electronics and Safety, 2008. ICVES 2008. IEEE International Conference on,* IEEE, pp. 317-322.

Wei, M., Yang, F. and Cao, Z., 2003. A Review of Development and Study on the Traffic Simulation [J]. *Acta Simulata Systematica Sinica,* 8, 035.

Weigend, A.S., Huberman, B.A. and Rumelhart, D.E., 1990. Predicting the future: A connectionist approach. *International Journal of Neural Systems,* 1 (03), 193-209.

WenTong, Z., and Wei, D., 2004. *Learning from the textbook: SPSS Statistical Analysis Advanced Course. [SPSS 统计分析高级教程] (in Chinese).* Beijing, China: Higher Education Press.

Whittle, P., 1963. Prediction and regulation by linear least-square methods.

Williams, B.M., 2001. Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling. *Transportation Research Record: Journal of the Transportation Research Board,* 1776 (1), 194-200.

Williams, B.M., Durvasula, P.K. and Brown, D.E., 1998. Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models. *Transportation Research Record: Journal of the Transportation Research Board,* 1644 (1), 132-141.

Williams, B.M., and 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.

Williams, B.M., and Hoel, L.A., 1999. *Modeling and Forecasting Vehicular Traffic Flow as a Seasonal Stochastic Time Series Process,* .

Williams, B., and Camp, T., 2002. Comparison of broadcasting techniques for mobile ad hoc networks. *In: Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing,* ACM, pp. 194-205.

Williams, G., 1990. Overdetermined systems of linear equations. *American Mathematical Monthly,* , 511-513.

Wilson, R.E., 2008. Mechanisms for spatio-temporal pattern formation in highway traffic models. *Philosophical Transactions.Series A, Mathematical, Physical, and Engineering Sciences,* 366 (1872), 2017-2032.

Wischoff, L., Ebner, A., Rohling, H., Lott, M. and Halfmann, R., 2003. SOTIS-a self-organizing traffic information system. *In: Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual,* IEEE, pp. 2442-2446.

Wu, C., Ho, J. and Lee, D., 2004. Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems,* 5 (4), 276-281.

Wu, H., Fujimoto, R., Guensler, R. and Hunter, M., 2004. MDDV: a mobility-centric data dissemination algorithm for vehicular networks. *In: Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks,* ACM, pp. 47-56.

Wu, J., 2005. *Handbook on theoretical and algorithmic aspects of sensor, ad hoc wireless, and peer-to-peer networks.* CRC Press.

Wu, M., Lu, T., Ling, F., Sun, J. and Du, H., 2010. Research on the architecture of Internet of things. *In: Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on,* IEEE, pp. V5-484-V5-487.

Xiang, H., Xu, T. and He, S., May, 2016. Log-ARIMA-Based Road Operating Speed Prediction Model (in Chinese: 基于 Log-ARIMA 的道路运行车速预测模型). *Journal of Chongqing Normal University (Natural Science),* 33 (3), 121-125.

Xie, J., Wu, W. and Yang, X., 2011. A PDL model used for short-term traffic flow forecasting. *Journal of Tongji University.Natural Science,* 39 (9), 1297-1302.

Xie, S., A, B. and Wei, L., 2013. Comparative analysis of prediction model for vehicle operating speed (in Chinese: 公路运行车速预测模型对比分析). *Journal of Chang'an University (Natural Science Edition) (in Chinese: 长安大学学报 (自然科学版)),* 5, 016.

Xie, S., Jiang, X. and Wei, L., 2014. Prediction of Driving Speed Interval of Passenger Car Based on Fuzzzy Linear Regression (In Chinese: 基于模糊线性回归的小轿车运行车速区间预测). *Journal of Highway and Transportation Research and Development (in Chinese: 公路交通科技),* 31 (8), 126-130.

Xu, L., and Fu, H., 2009. *Intelligent Prediction Theory and Methods of Traffic Information.* China, Beijing: Science Press.

Xu, S., Guo, P., Xu, B. and Zhou, H., 2013. QoS evaluation of VANET routing protocols. *Journal of Networks,* 8 (1), 132-139.

Xu, H., and Xia, C., 2013. VANETs routing reviews (Chiness:VANETs 路由综述). *Application Research of Computers,* 30 (1).

Yan, C., 1996. A simple transformation for nonlinear waves. *Physics Letters A,* 224 (1), 77-84.

Yang, J., Chou, L., Li, Y., Lin, Y., Huang, S., Tseng, G., Wang, T. and Lu, S., 2010. Prediction of short-term average vehicular velocity considering weather factors in urban VANET environments. *In: Machine Learning and Cybernetics (ICMLC), 2010 International Conference on,* IEEE, pp. 3039-3043.

Yang, L., Cao, J., Yuan, Y., Li, T., Han, A. and Chan, A., 2013. A framework for partitioning and execution of data stream applications in mobile cloud computing. *ACM SIGMETRICS Performance Evaluation Review,* 40 (4), 23-32.

Yin, H., Wong, S., Xu, J. and Wong, C., 2002. Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research Part C: Emerging Technologies,* 10 (2), 85-98.

Yin, Y., and Shang, P., 2016. Forecasting traffic time series with multivariate predicting method. *Applied Mathematics and Computation,* 291, 266-278.

Yu, G., Hu, J., Zhang, C., Zhuang, L. and Song, J., 2003. Short-term traffic flow forecasting based on markov chain model. *In: Intelligent Vehicles Symposium, 2003. Proceedings. IEEE,* IEEE, pp. 208-212.

Yu, G., and Zhang, C., 2004. Switching ARIMA model based forecasting for traffic flow. *In: Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on,* IEEE, pp. ii-429-32 vol. 2.

Yu, X.Zhao, W., 2006. *Network Simulation System base on NS-3.* Zhejiang University, China.

Zhang, L., Liu, Q., Yang, W., Wei, N. and Dong, D., 2013. An improved k-nearest neighbor model for short-term traffic flow prediction. *Procedia-Social and Behavioral Sciences,* 96, 653-662.

Zhang, X., and Rice, J.A., 2003. Short-term travel time prediction. *Transportation Research Part C: Emerging Technologies,* 11 (3), 187-210.

Zheng, Y., Zhang, L., Xie, H., Tan, G. and Wang, H., 2012. Topology structure in vehicular ad-hoc network based-on urban scenarios. *Advances in Information Sciences and Service Sciences,* 4 (6), 137-144.

Zhou, D., Chen, S. and Dong, S., 2013. Network traffic prediction based on ARFIMA model. *arXiv Preprint arXiv:1302.6324,* .

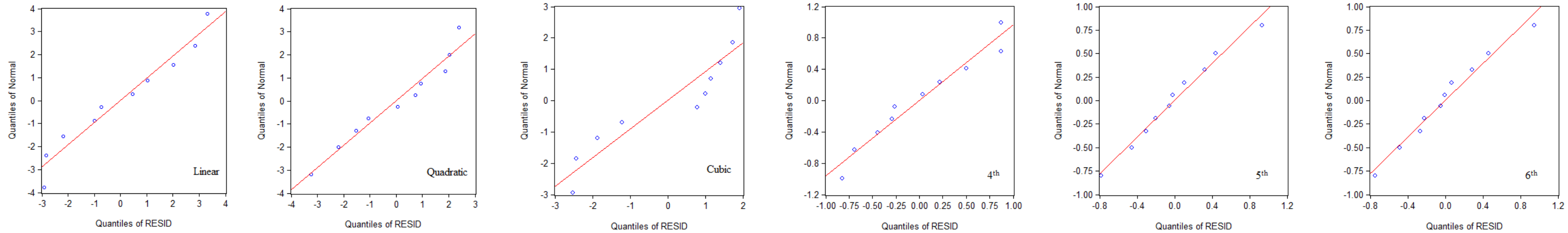
Zhu, S., Wang, H. and Xiang, H., 2008. *Traffic Flow parameter and traffic incident dynamic forecast method.* 1st ed. China, Nanjing: Southeast University Press.

Zimmermann, H., 1980. OSI reference model--The ISO model of architecture for open systems interconnection. *Communications, IEEE Transactions On,* 28 (4), 425-432.

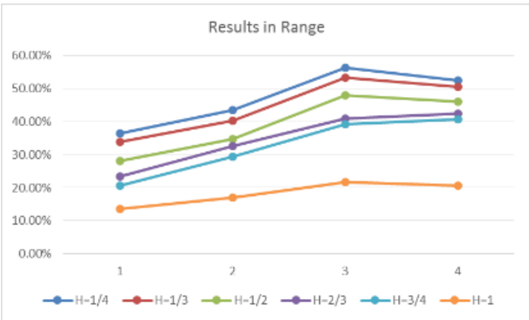
Zou, K.H., Tuncali, K. and Silverman, S.G., 2003. Correlation and simple linear regression 1. *Radiology,* 227 (3), 617-628.

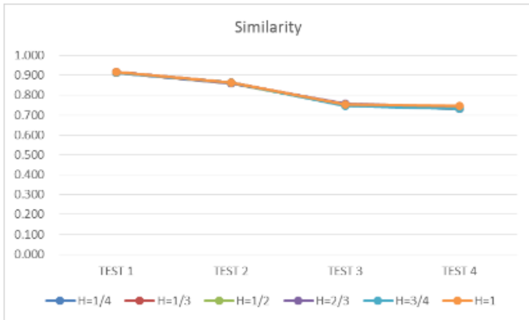
Appendix

Appendix A: The normal probability plot (Q-Q diagram) in section 3.2.5

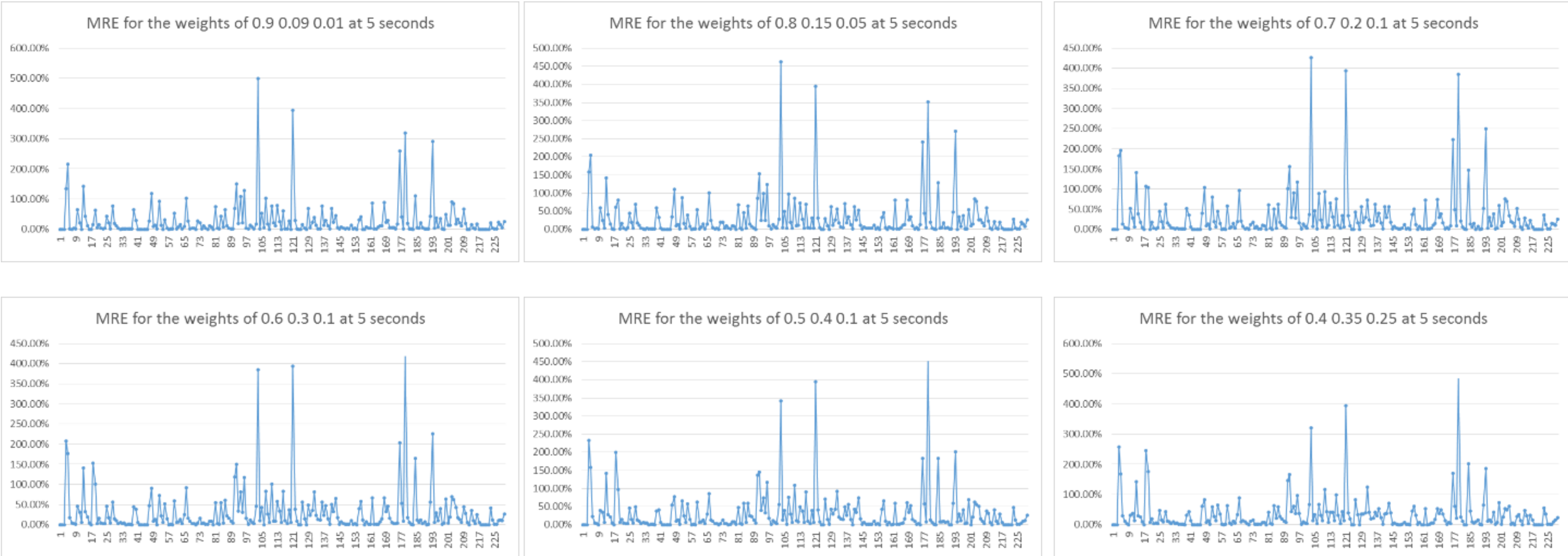


The parameters test for adjusting range (Table 3.4)

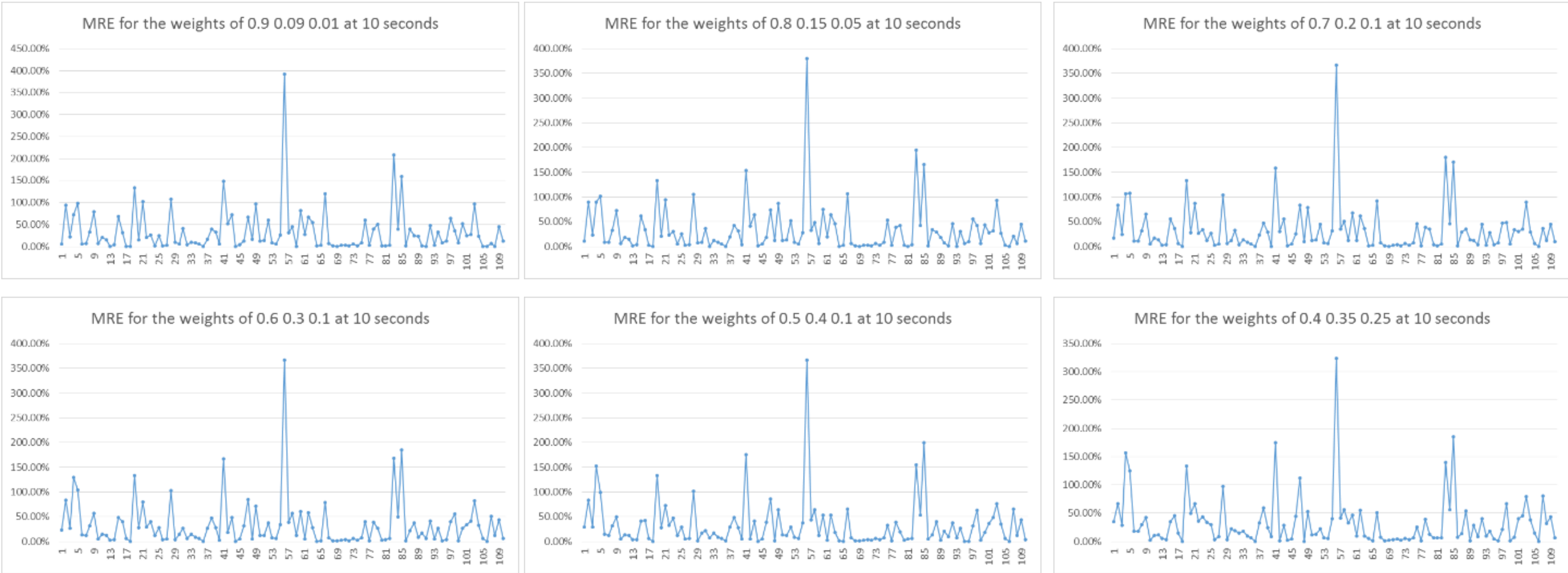




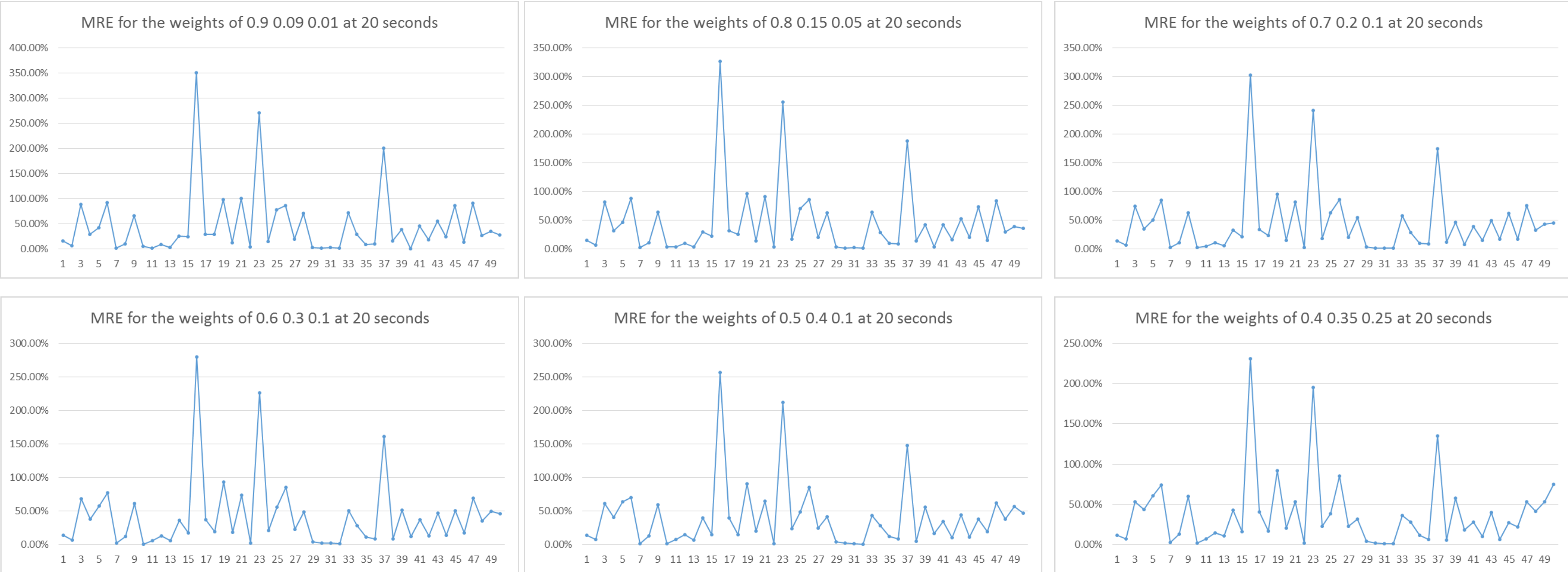
MRE testing figures for the weight settings (Fig. 3.7 and Fig. 3.8)



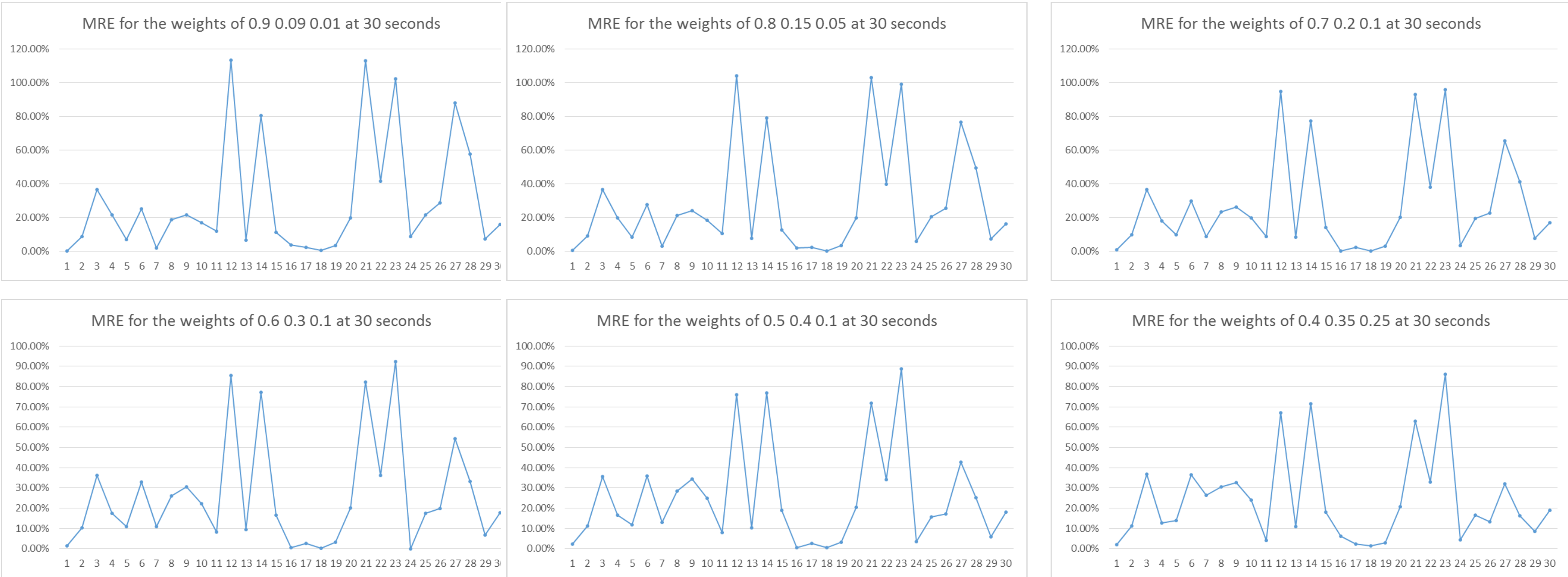
MRE for the weight settings at **5 seconds** time period



MRE for the weight settings at **10 seconds** time period

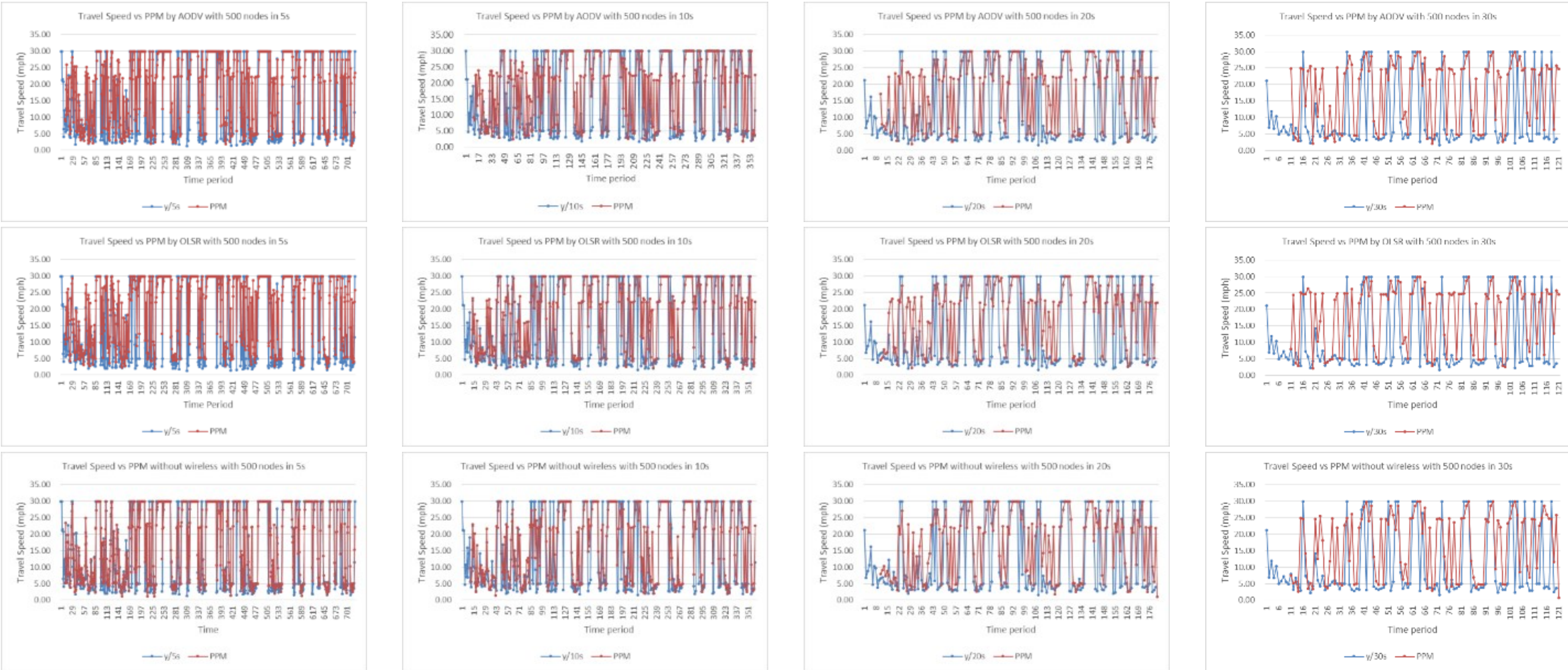


MRE for the weight settings at **20 seconds** time period

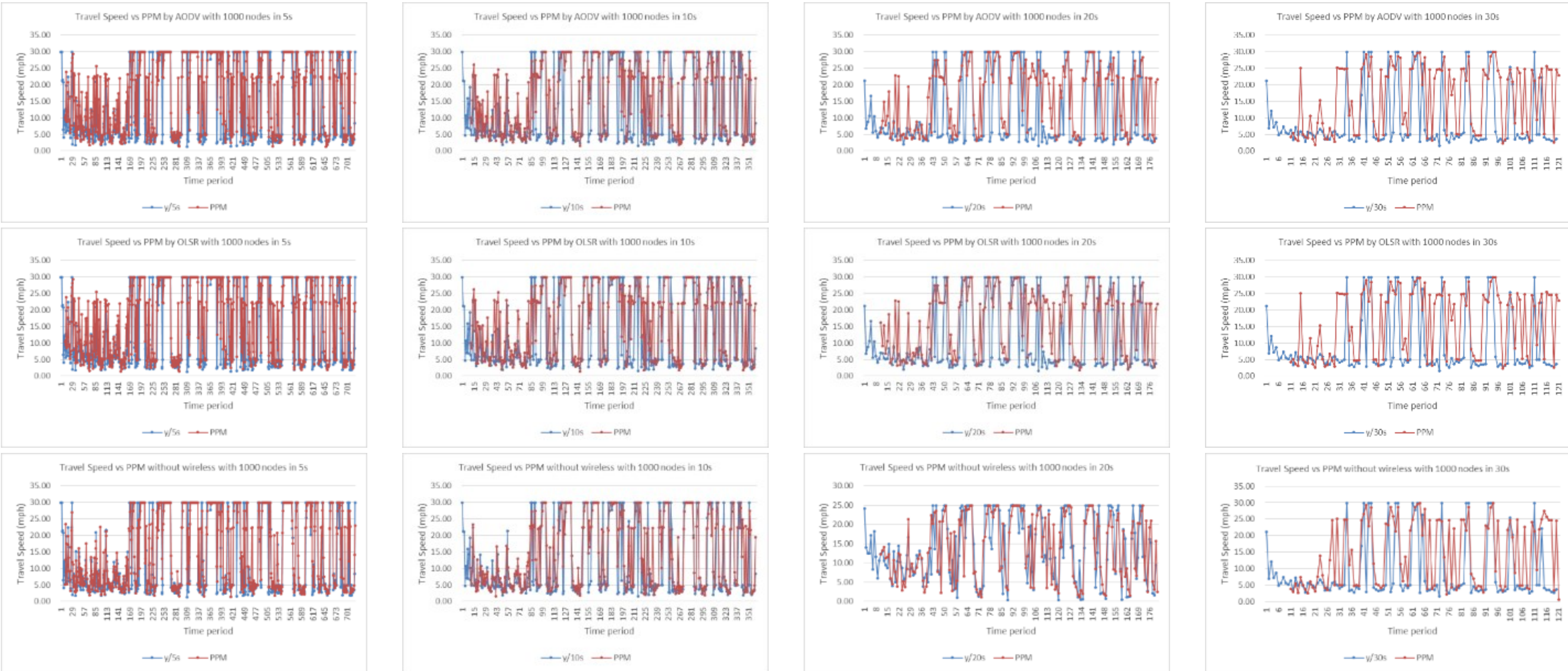


MRE for the weight settings at **30 seconds** time period

Appendix B: The results of analysis of routing protocol in section 5.3.1 (Fig.5.4 and Fig.5.5)



The results of PPM with PRAWMA prediction scheme of AODV, OLSR and without wireless environment compare with observation data in **500 nodes** during 1 hour at 5s, 10s, 20s and 30s time period



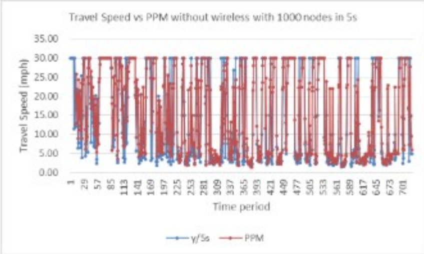
The results of PPM with PRAWMA prediction scheme of AODV, OLSR and without wireless environment compare with observation data in **1000 nodes** during 1 hour at 5s, 10s, 20s and 30s time period

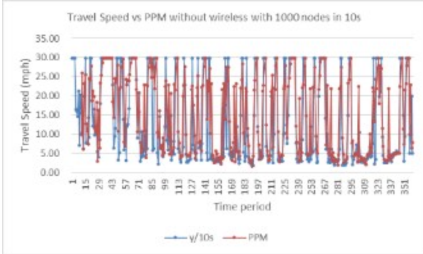
The results of PPM with PRAWMA prediction scheme of AODV, OLSR and without wireless environment compare with observation data in **2800 nodes** during 1 hour at 5s, 10s, 20s and 30s time period

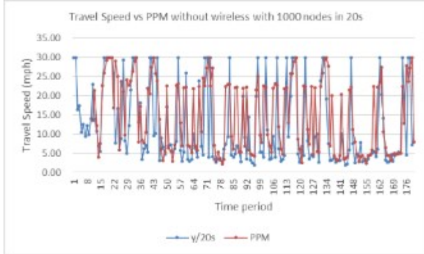
Appendix C: The results of analysis of mobility pattern in section 5.3.2 (Fig.5.6 and Fig.5.7)

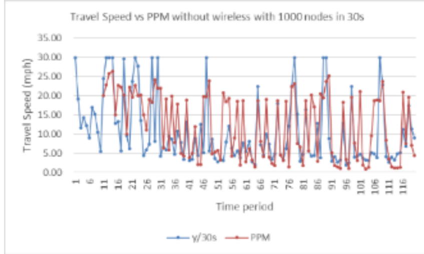
The PPM predicted results without wireless environment of 8 mobility patterns for Parliament street

Mobility 1

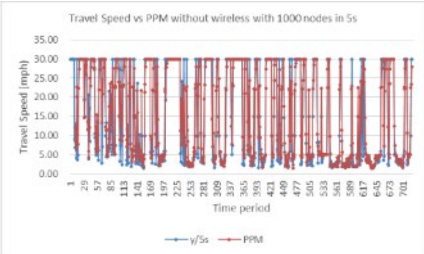


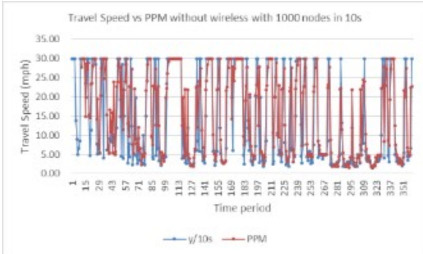


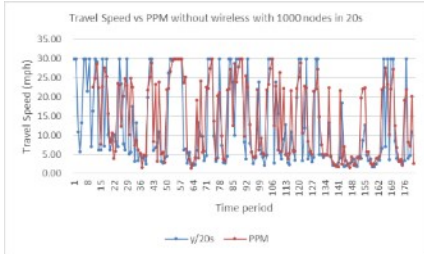


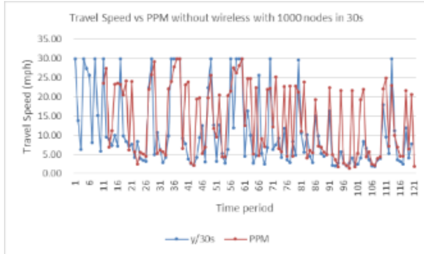


Mobility 2

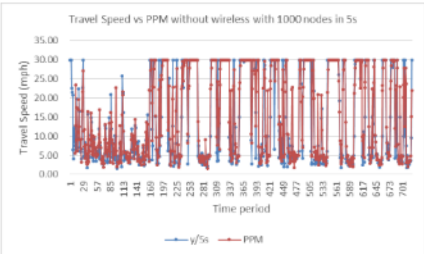


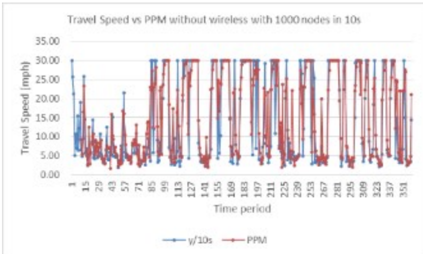


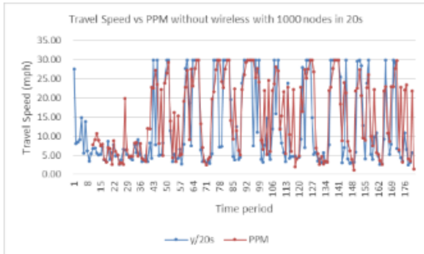


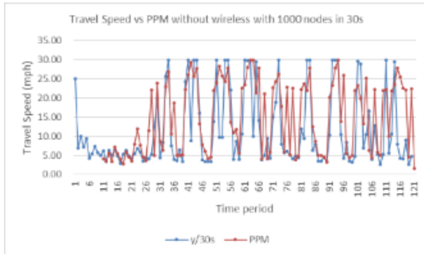


Mobility 3

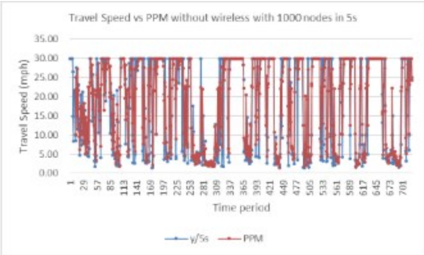


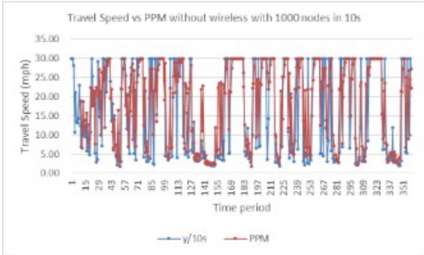


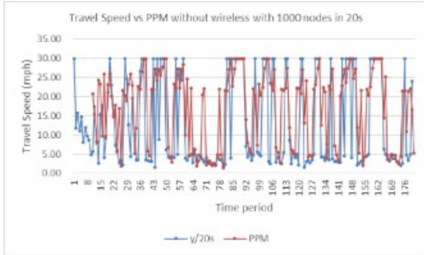


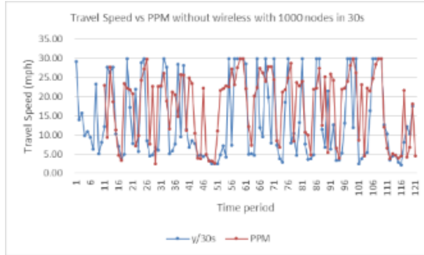


Mobility 4

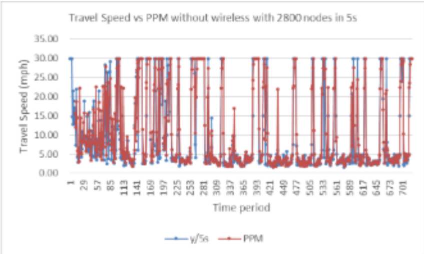


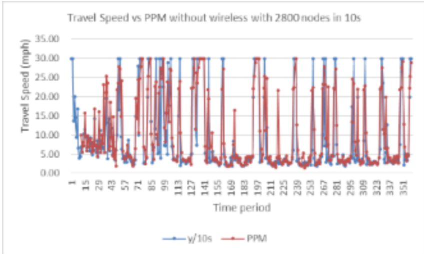


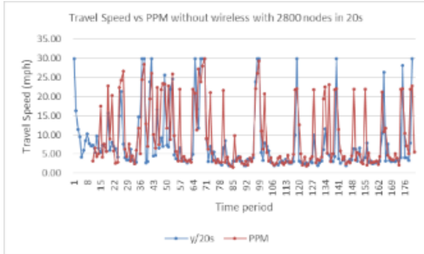


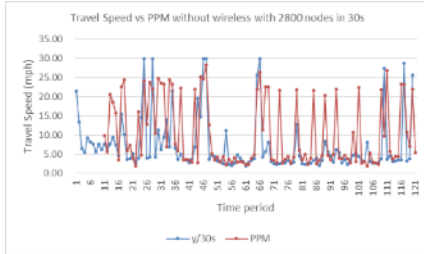


Mobility 5

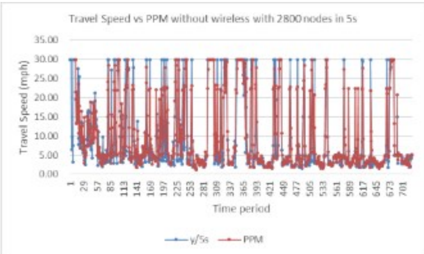


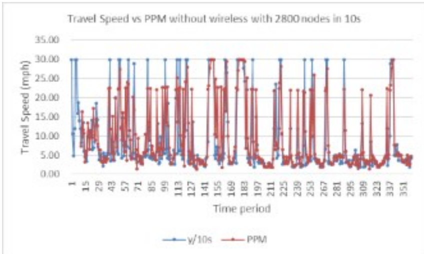


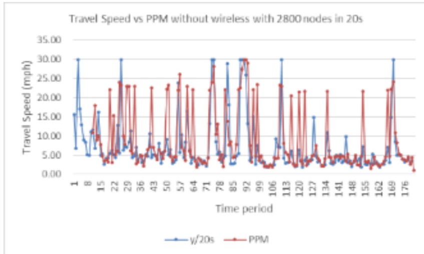


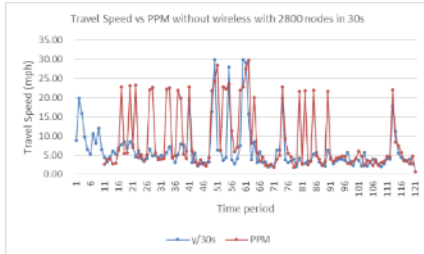


Mobility 6

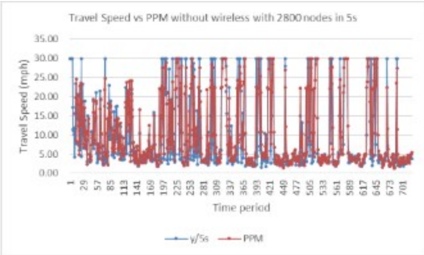


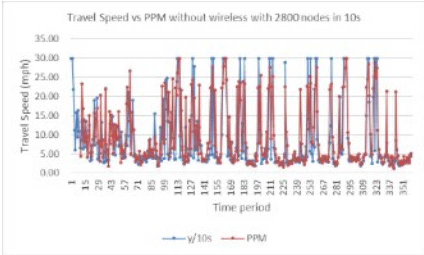


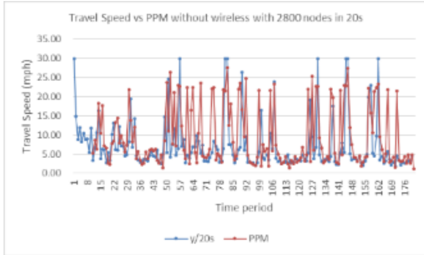


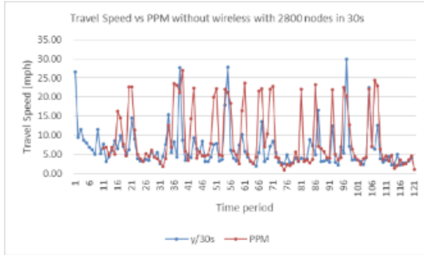


Mobility 7

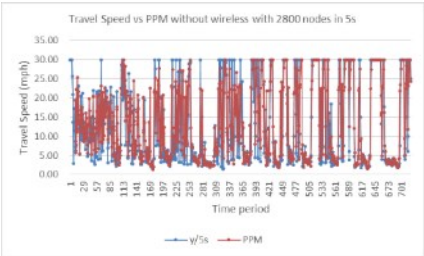


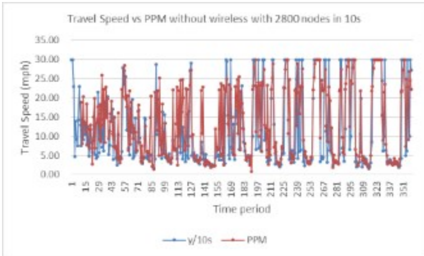


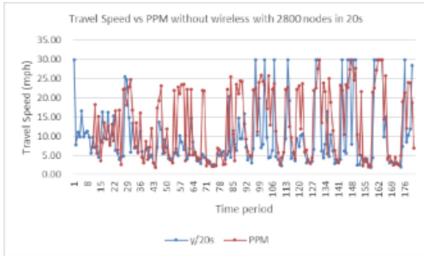


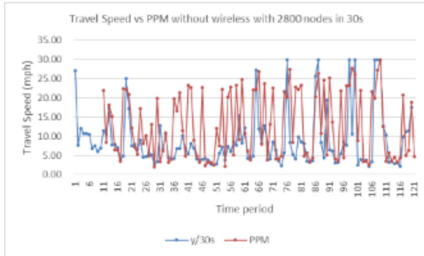


Mobility 8





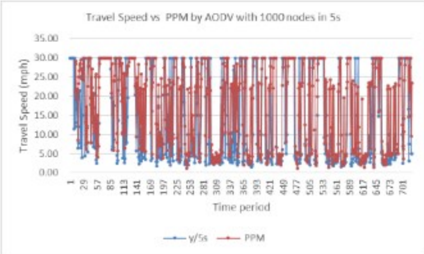


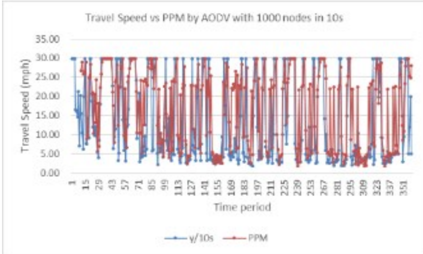


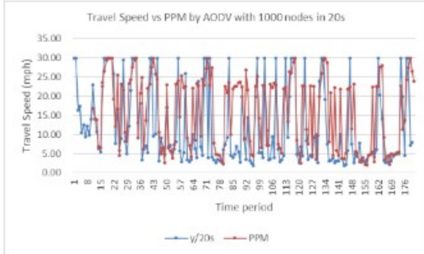
The results of PPM with PRAWMA prediction scheme **without wireless environment** compare with observation data in **1000 nodes (mobility 1-4) and 2800 nodes (mobility 5-8)** during 1 hour at 5s, 10s, 20s and 30s time period

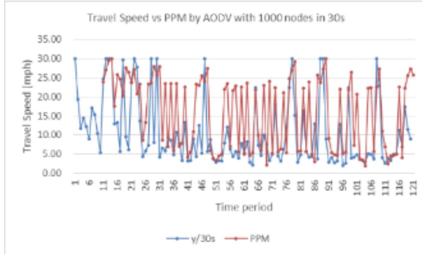
The PPM predicted results of AODV of 8 mobility patterns

Mobility 1

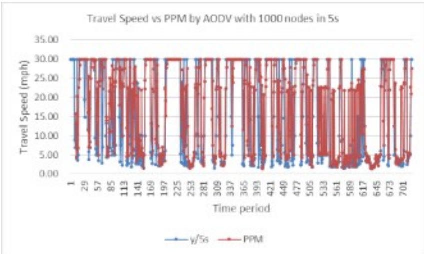


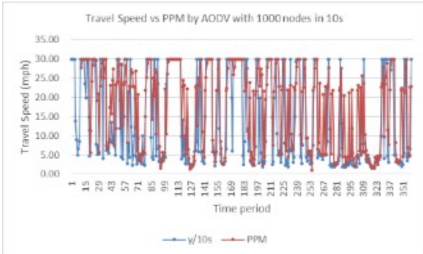


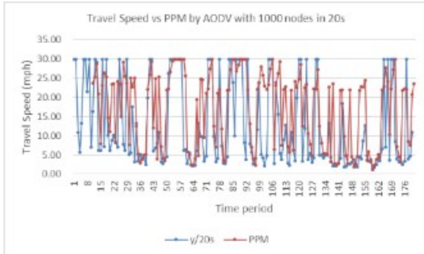


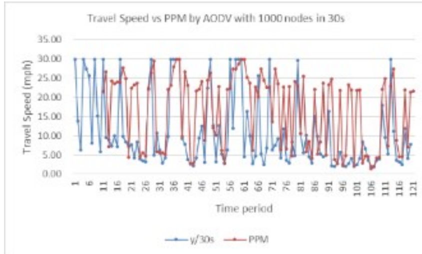


Mobility 2

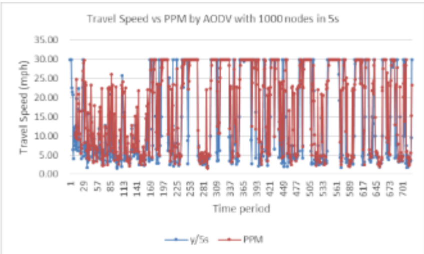


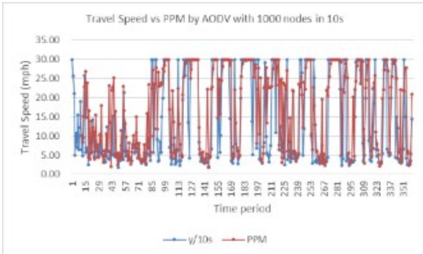


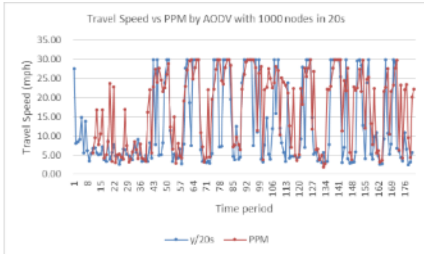


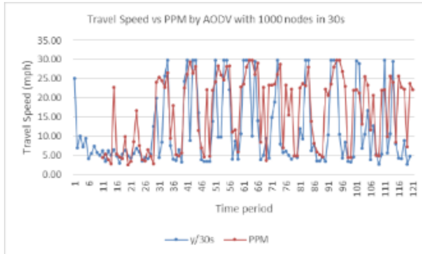


Mobility 3

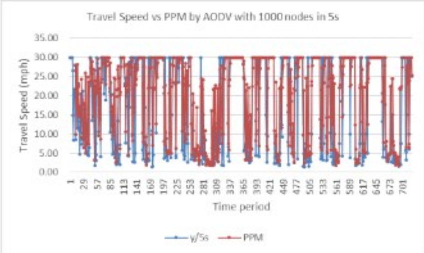


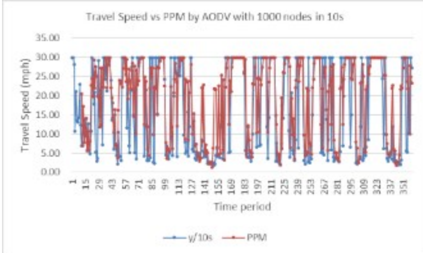


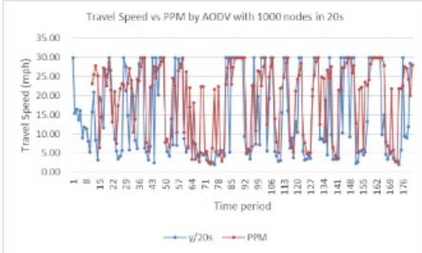


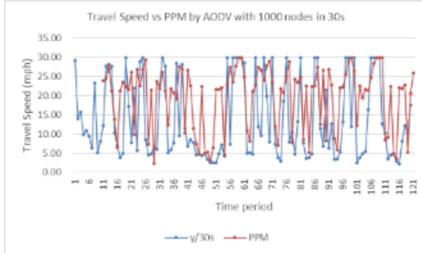


Mobility 4

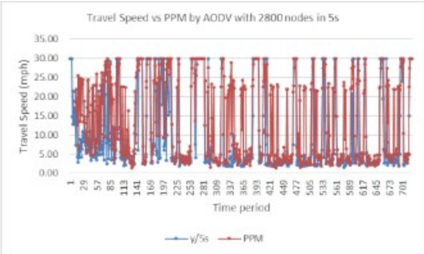


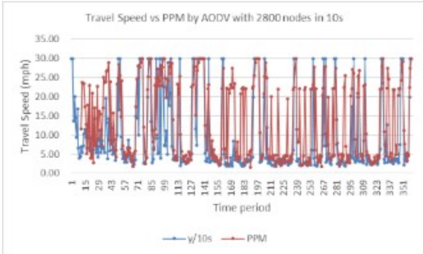


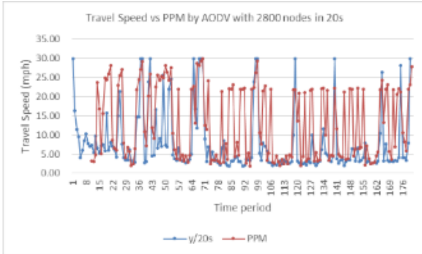


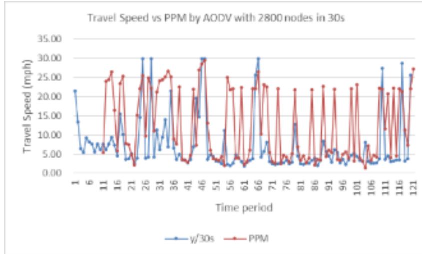


Mobility 5

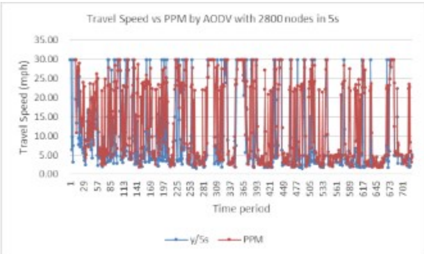


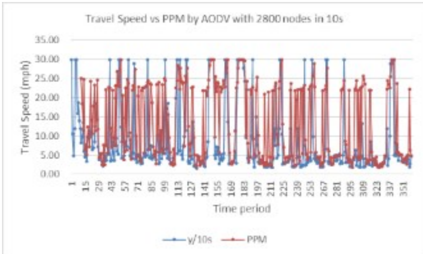


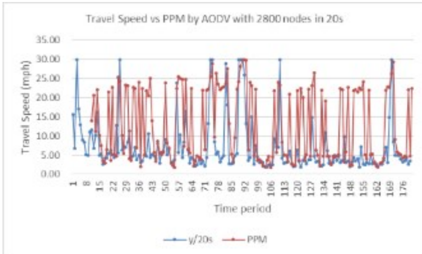


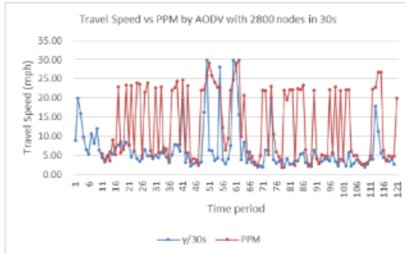


Mobility 6

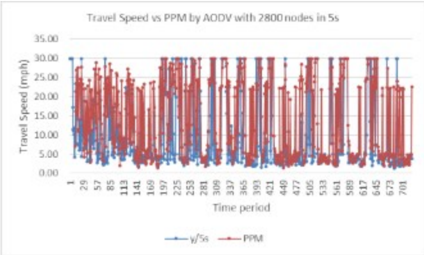


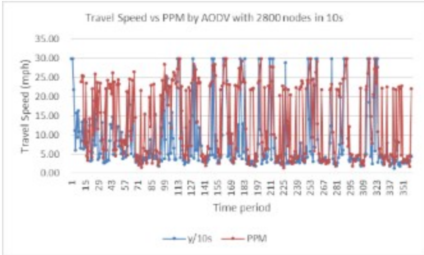


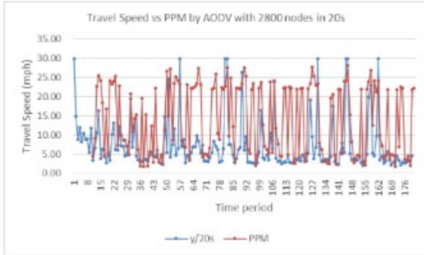


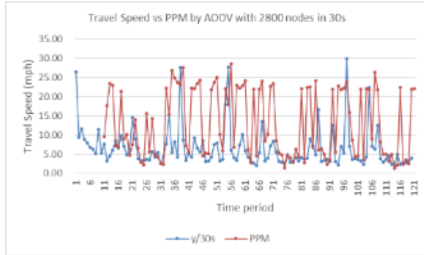


Mobility 7

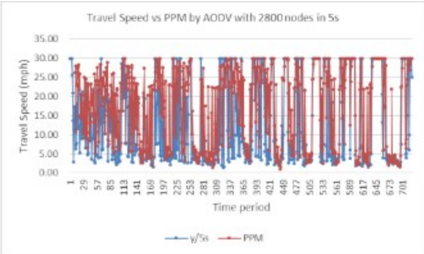


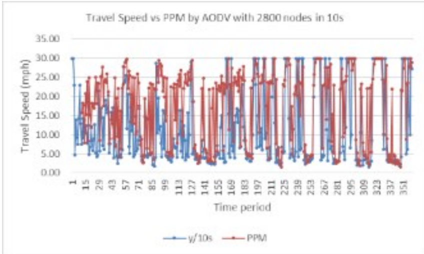


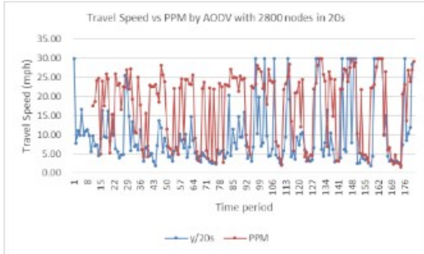


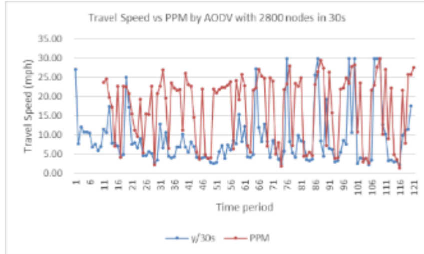


Mobility 8







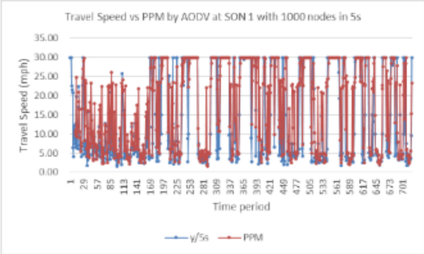


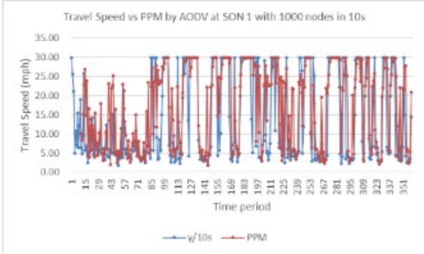
The results of PPM with PRAWMA prediction scheme of **AODV** compare with observation data in **1000 nodes (mobility 1-4) and 2800 nodes (mobility 5-8)** during 1 hour at 5s, 10s, 20s and 30s time period

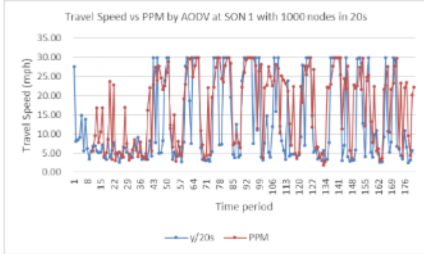
Appendix D: The results of analysis of observation location in section 5.3.3 (Fig.5.8)

The PPM predicted results of AODV of 3 observation nodes in 1000 nodes for Parliament street

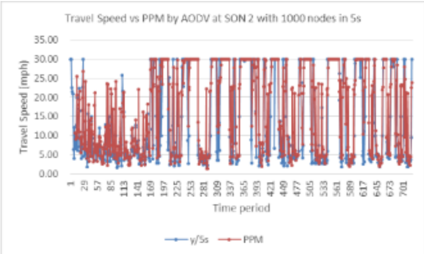
SON 1

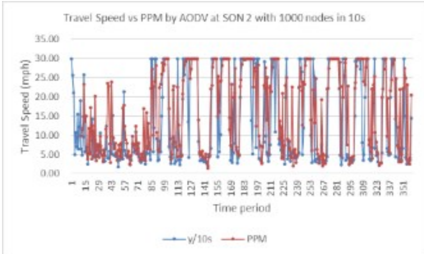


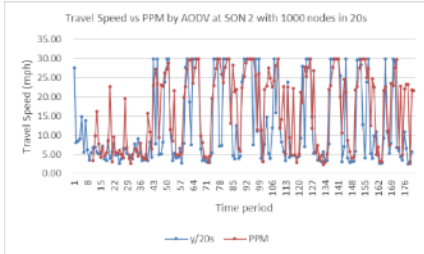


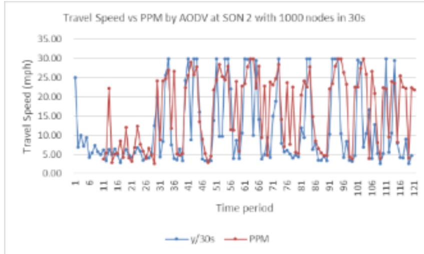


SON 2

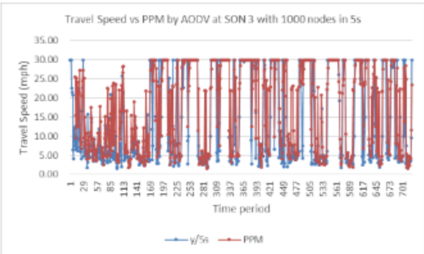


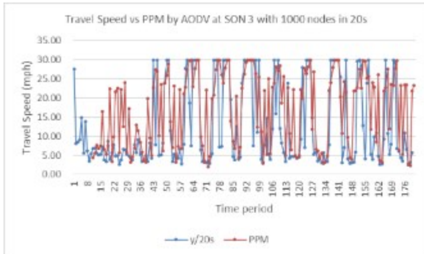


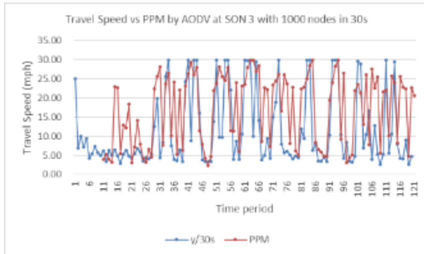




SON 3

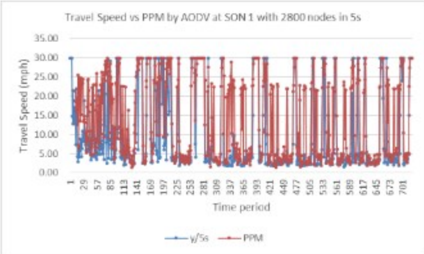


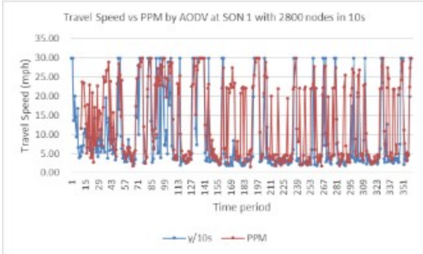


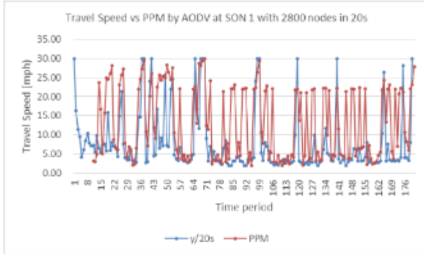


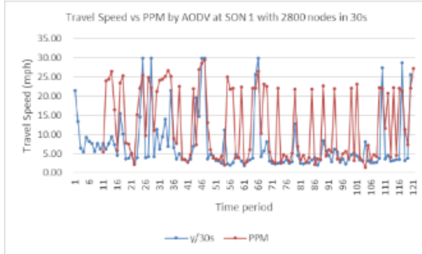
The PPM predicted results of AODV of 3 observation nodes in 2800 nodes

SON 1

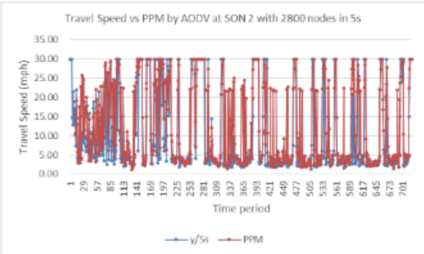


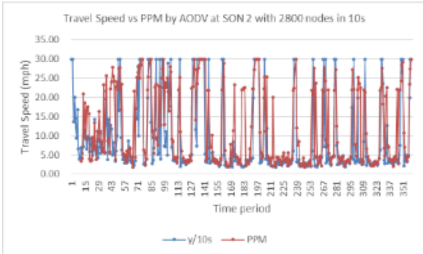


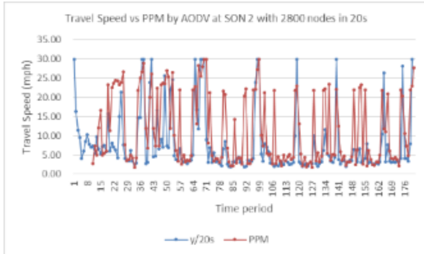


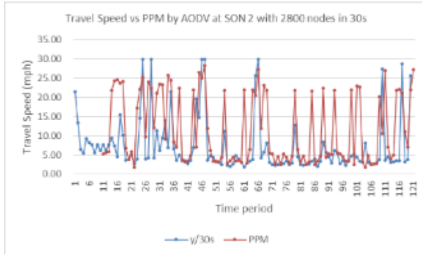


SON 2

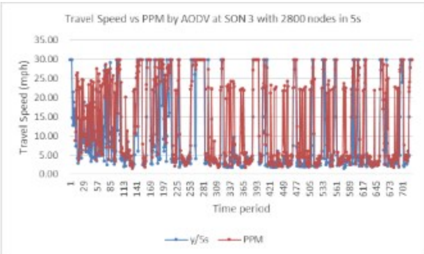


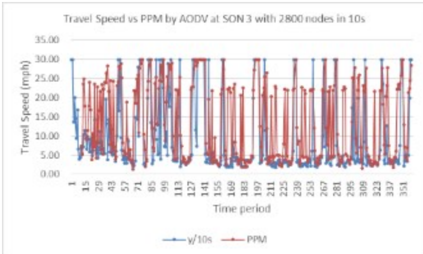


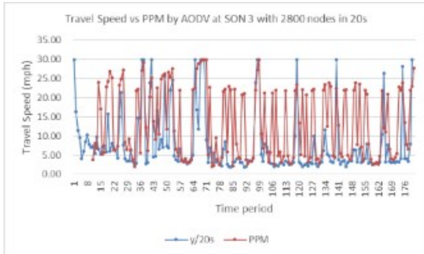


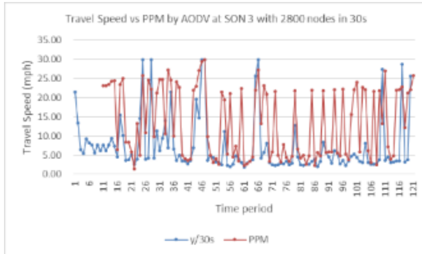


SON 3







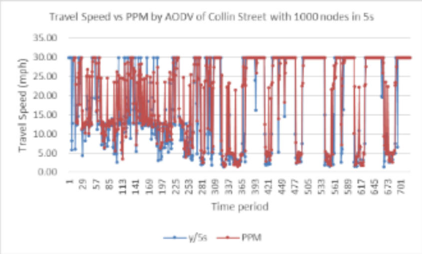


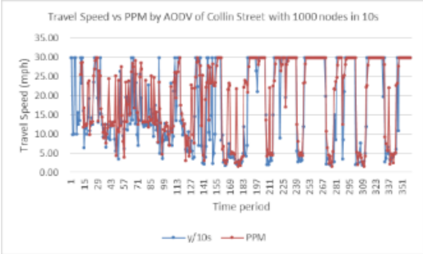
The results of PPM with PRAWMA prediction scheme of **AODV** compare with observation data in **1000 nodes (mobility 3)** and **2800 nodes (mobility 5)** during 1 hour at 5s, 10s, 20s and 30s time period

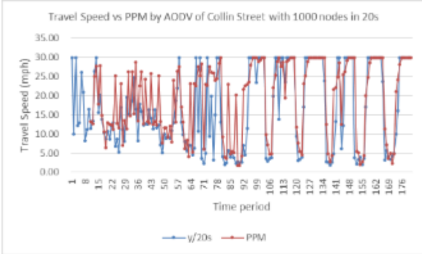
Appendix E: The results of analysis of objective road section in section 5.3.4 (Fig.5.10 and Fig.5.11)

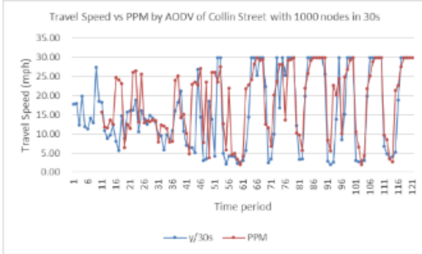
The PPM predicted results of AODV of two streets with 1000 nodes at SON 1

Collin Street

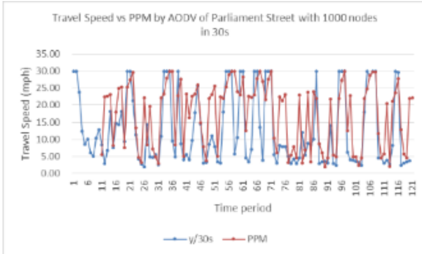








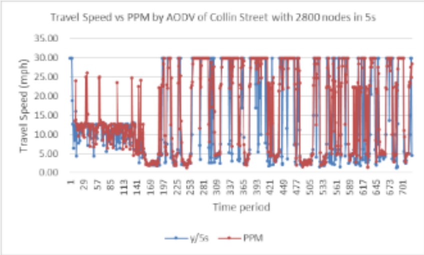
Parliament Street

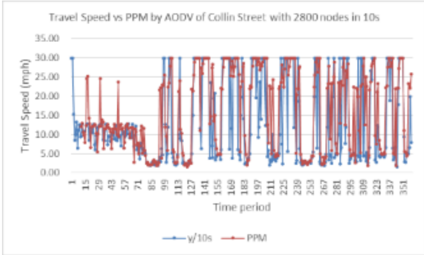


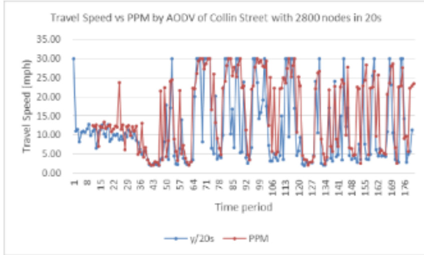
The results of PPM with PRAWMA prediction scheme of **AODV** compare with observation data with **1000 nodes** for Collin Street and Parliament Street during 1 hour at 5s, 10s, 20s and 30s time period

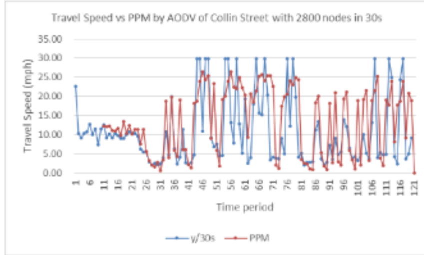
The PPM predicted results of AODV of two streets with 2800 nodes at SON 1

Collin Street

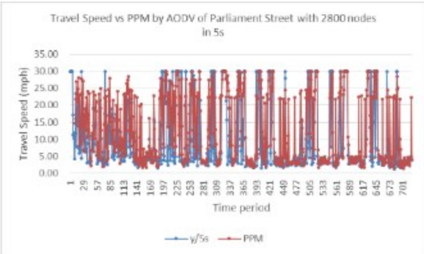


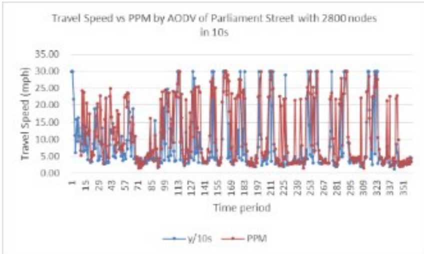


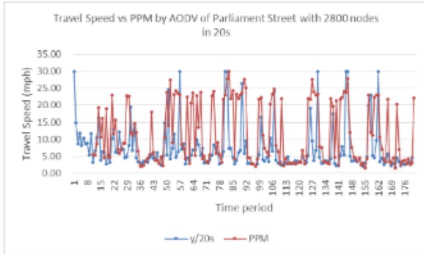


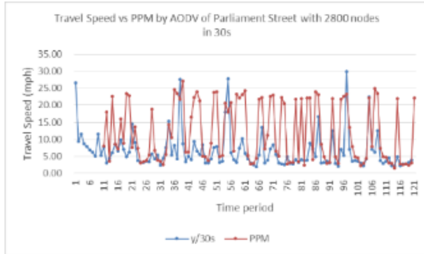


Parliament Street





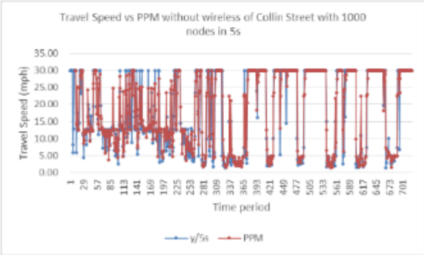


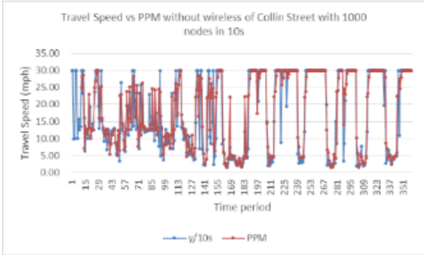


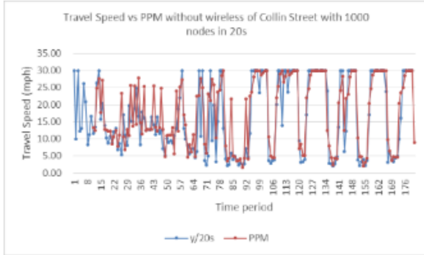
The results of PPM with PRAWMA prediction scheme of **AODV** compare with observation data with **2800 nodes** for Collin Street and Parliament Street during 1 hour at 5s, 10s, 20s and 30s time period

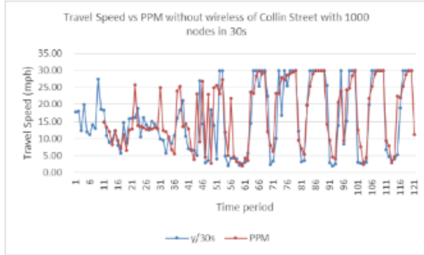
The PPM predicted results without wireless environment of two streets with 1000 nodes

Collin Street

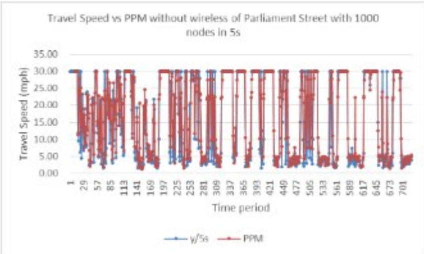


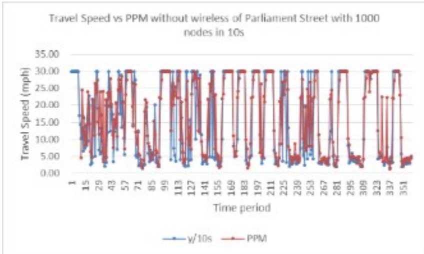


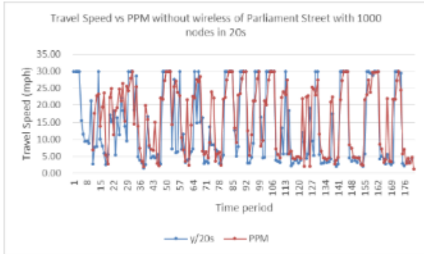


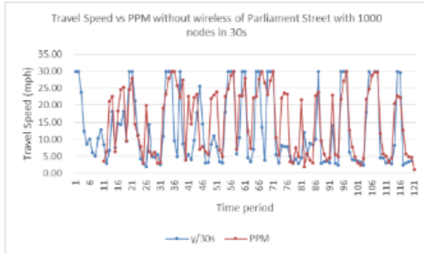


Parliament Street





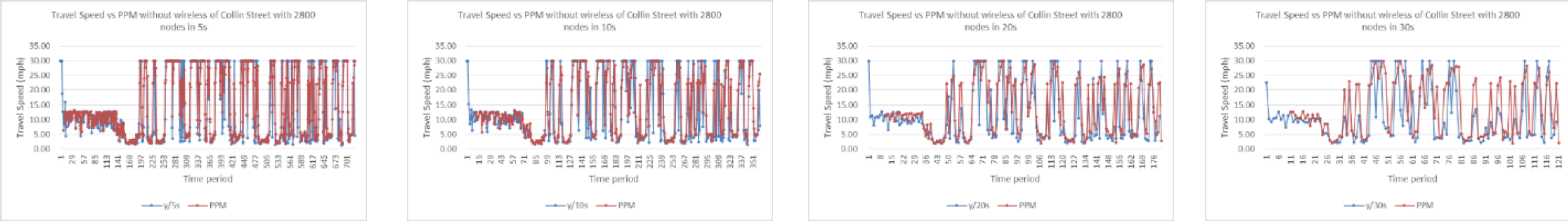




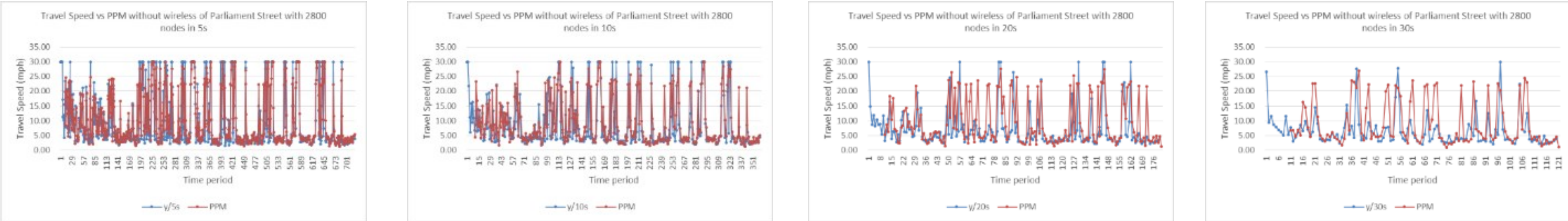
The results of PPM with PRAWMA prediction scheme **without wireless environment** compare with observation data with **1000 nodes** for Collin Street and Parliament Street during 1 hour at 5s, 10s, 20s and 30s time period

The PPM predicted results without wireless environment of two streets with 2800 nodes

Collin Street



Parliament Street

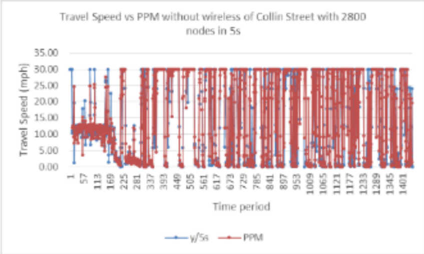


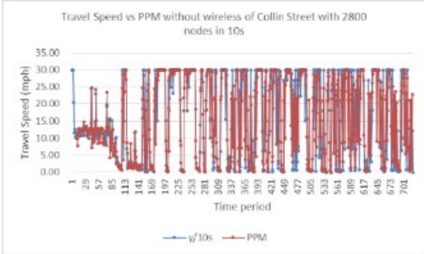
The results of PPM with PRAWMA prediction scheme **without wireless environment** compare with observation data with **2800 nodes** for Collin Street and Parliament Street during 1 hour at 5s, 10s, 20s and 30s time period

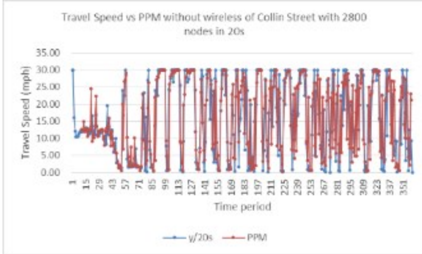
Appendix F: The results of analysis of time period in section 5.3.5 (Fig.5.12)

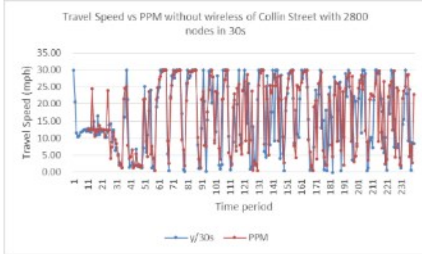
The PPM predicted results without wireless environment of two streets with 2800 nodes during 2 hours

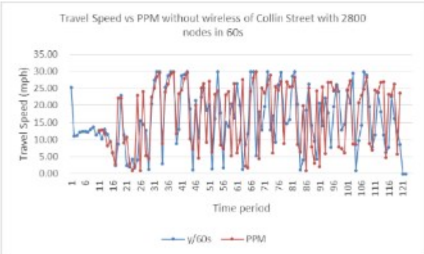
Collin Street

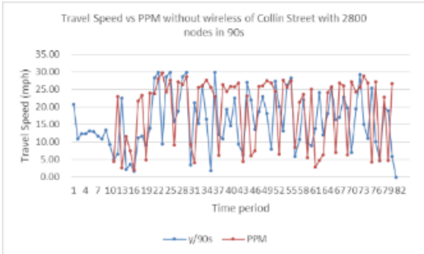


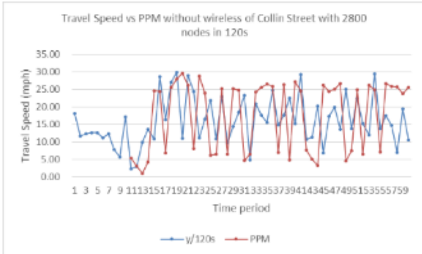


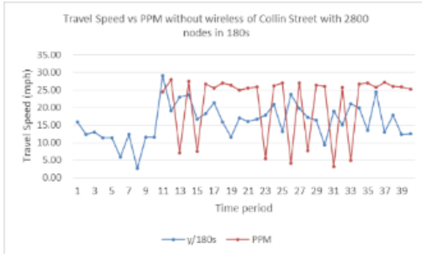




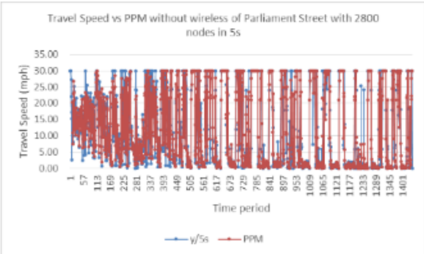


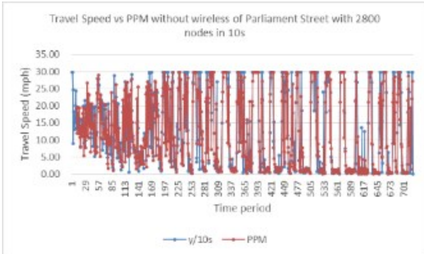


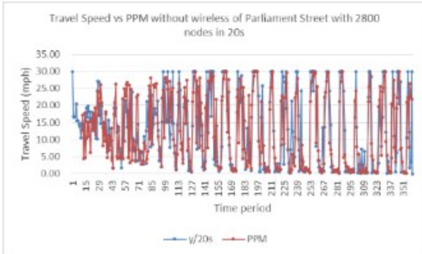


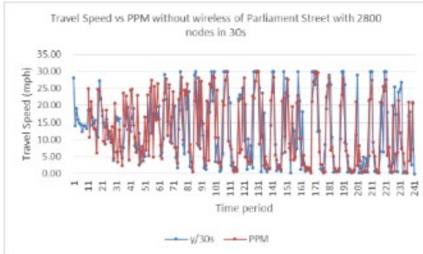


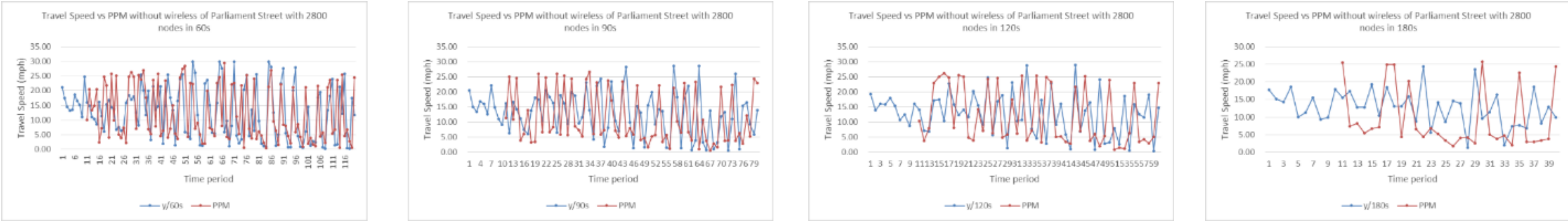
Parliament Street









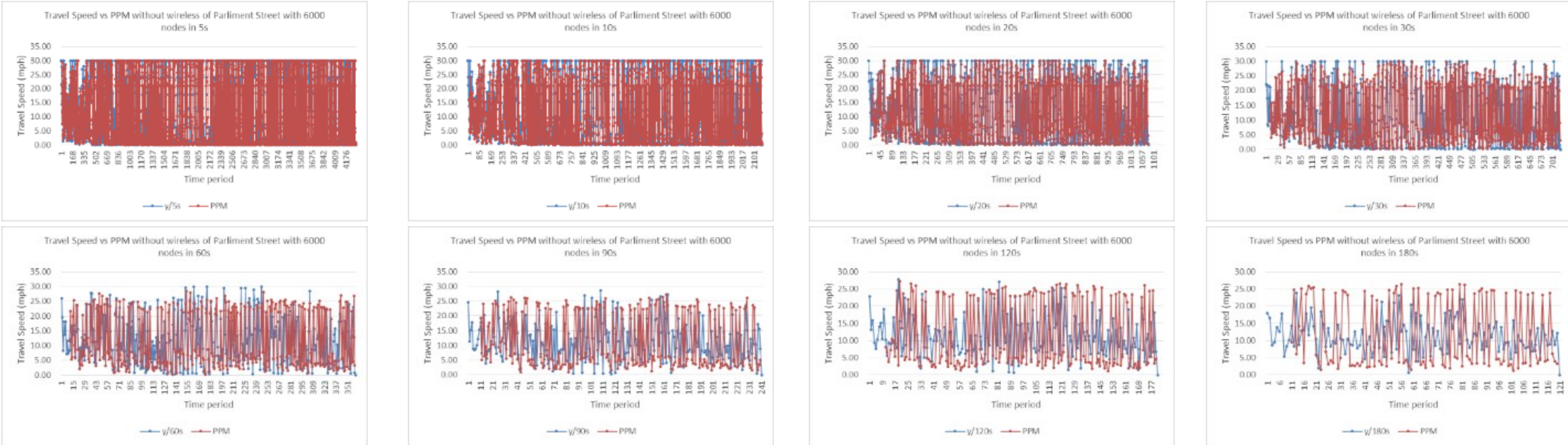


The results of PPM with PRAWMA prediction scheme without wireless environment compare with observation data with **2800 nodes** for Collin Street and Parliament Street during **2 hours** at 5s, 10s, 20s, 30s, 60s, 90s, 120s, and 180s time period

The PPM predicted results without wireless environment of two streets with 6000 nodes during 6 hours

Collin Street

Parliament Street

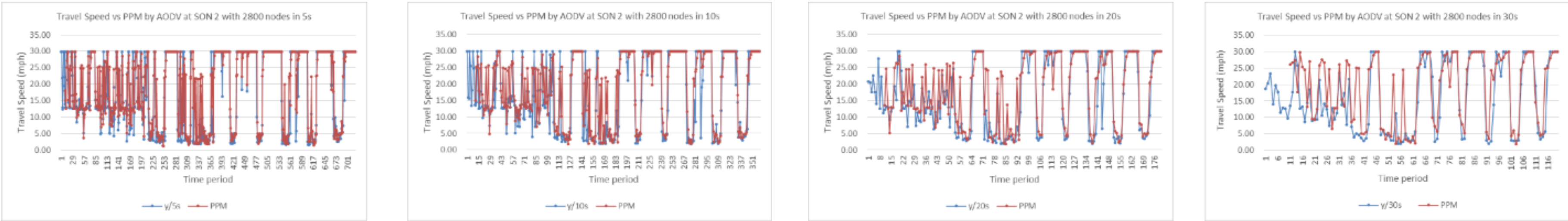


The results of PPM with PRAWMA prediction scheme without wireless environment compare with observation data with **6000 nodes** for Collin Street and Parliament Street during **6 hours** at 5s, 10s, 20s, 30s, 60s, 90s, 120s, and 180s time period

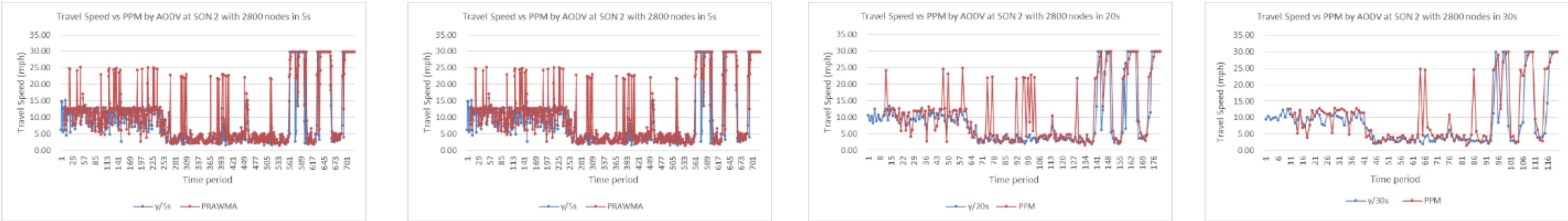
Appendix G: The results of analysis of peaks and off peaks in section 5.3.6 (Fig.5.13)

The PPM predicted results of Peak and Off-Peak by AODV for Collin Street with 2800 nodes at SON.2

Off-peak



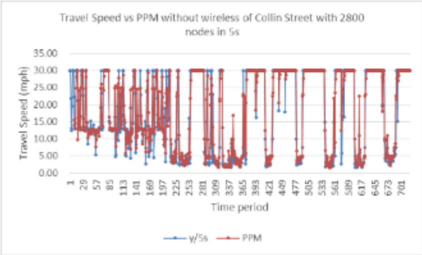
Peaks

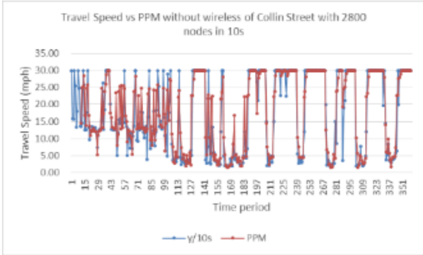


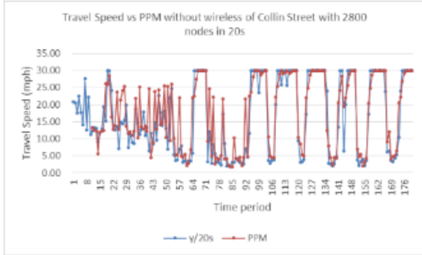
The results of PPM with PRAWMA prediction scheme of **AODV** compare with observation data with **2800 nodes** for Collin Street during 1 hour at 5s, 10s, 20s and 30s time period

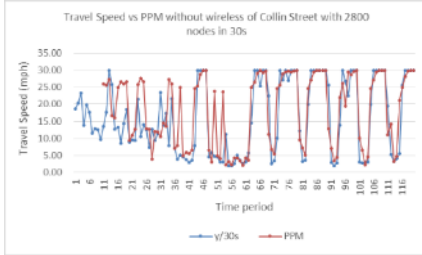
The PPM predicted results of Peak and Off-Peak without wireless for Collin Street with 2800 nodes at SON.2

Off-peak

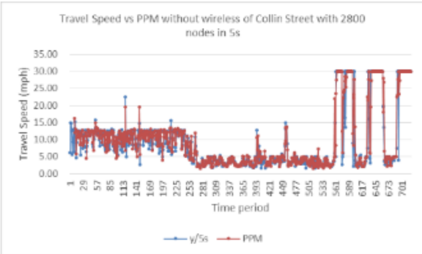


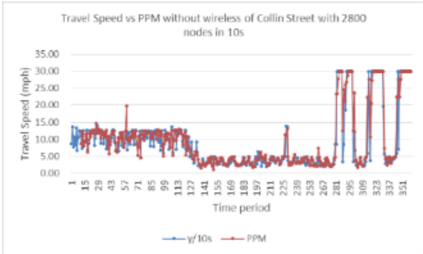


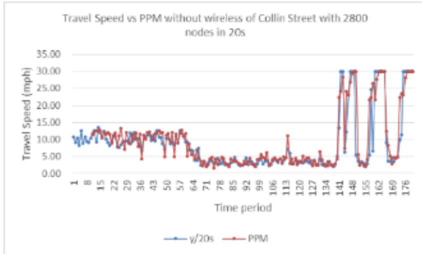


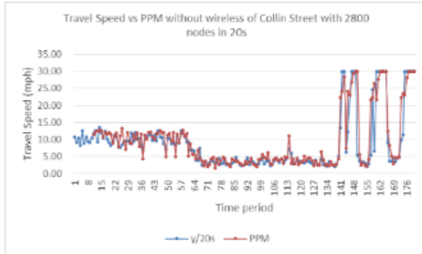


Peaks







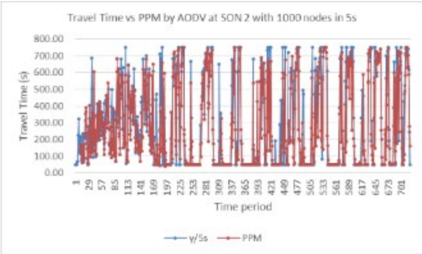


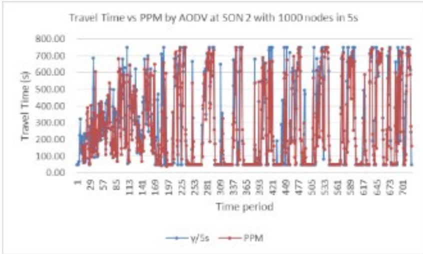
The results of PPM with PRAWMA prediction scheme **without wireless** compare with observation data with **2800 nodes** for Collin Street during 1 hour at 5s, 10s, 20s and 30s time period

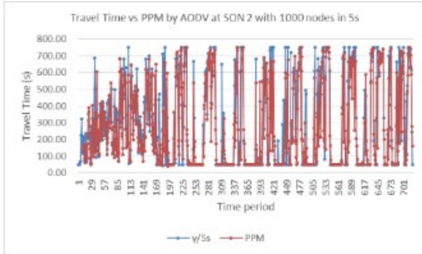
Appendix H: The results of analysis of Travel Time in section 5.3.7 (Fig.5.16)

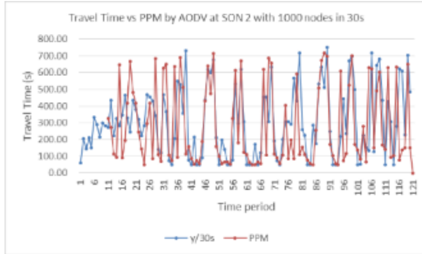
The PPM predicted results of Travel Time for Parliament street with 1000 nodes at SON.2

AODV

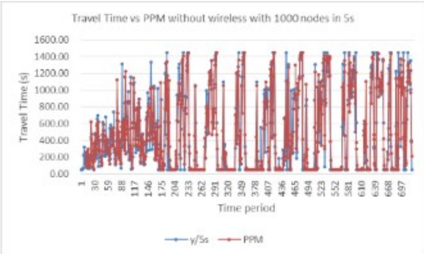


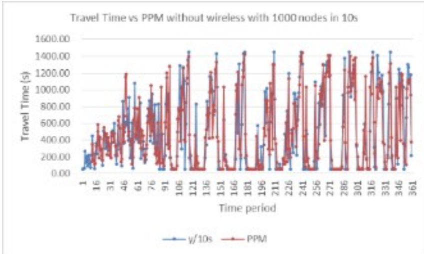


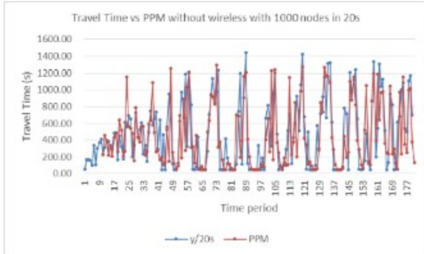


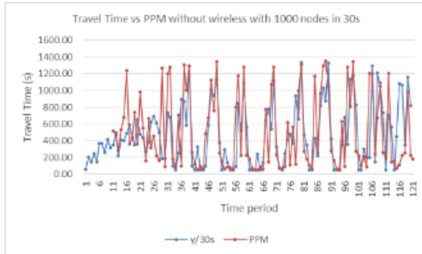


Without Wireless





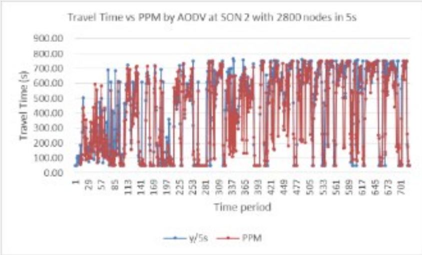


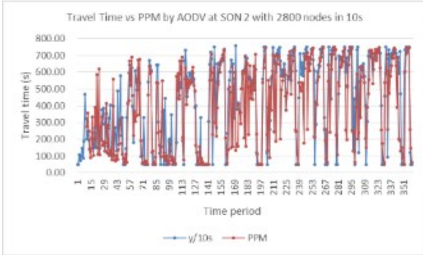


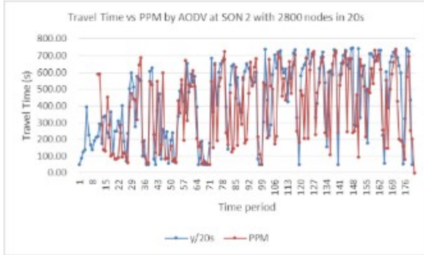
The results of PPM with PRAWMA prediction scheme of **Travel Time** with **1000 nodes** for Parliament Street during 1 hour at 5s, 10s, 20s and 30s time period

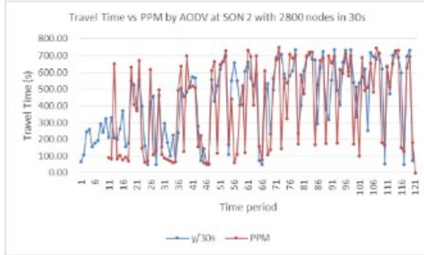
The PPM predicted results of Travel Time for Parliament street with 2800 nodes at SON.2

AODV

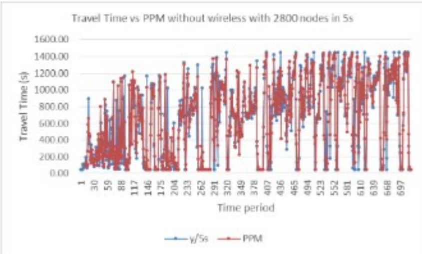


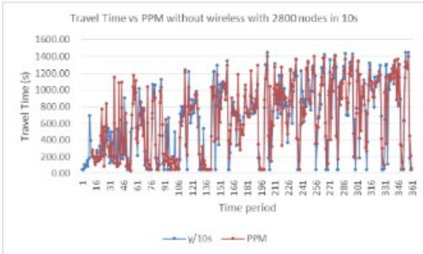


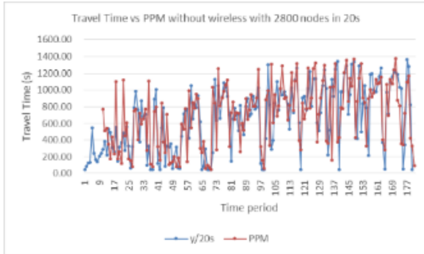


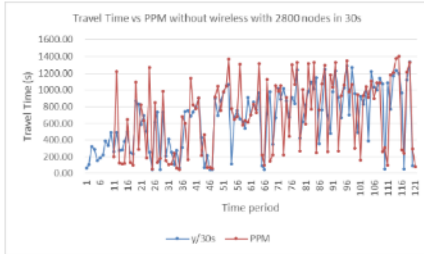


Without Wireless









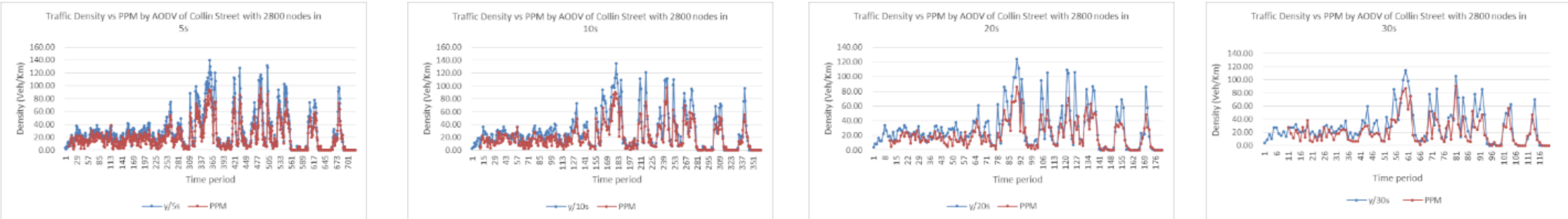
The results of PPM with PRAWMA prediction scheme of **Travel Time** with **2800 nodes** for Parliament Street during 1 hour at 5s, 10s, 20s and 30s time period

Appendix I: The results of analysis of traffic density prediction in section 5.3.8 (Fig.5.18)

The PPM predicted results of Travel Time by AODV for Collin Street with 2800 nodes at SON.2

Off-peak

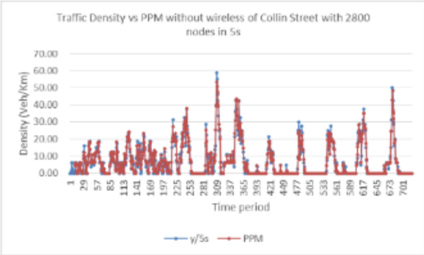
Peaks

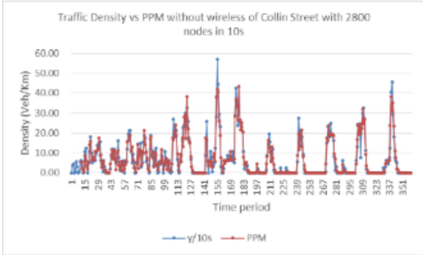


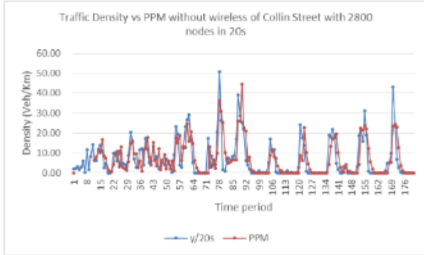
The results of PPM with PRAWMA prediction scheme of **AODV** compare with travel time with **2800 nodes** for Collin Street during 1 hour at 5s, 10s, 20s and 30s time period

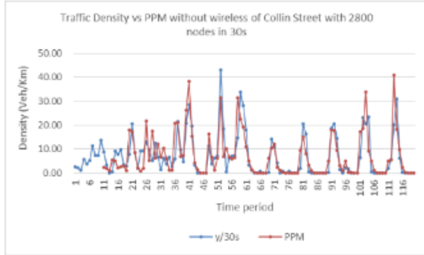
The PPM predicted results of Travel Time without wireless for Collin Street with 2800 nodes at SON.2

Off-peak

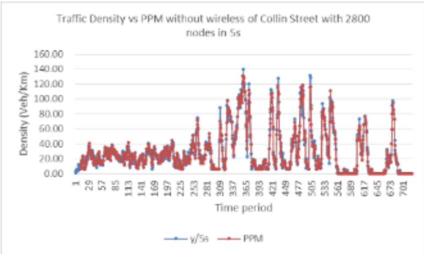


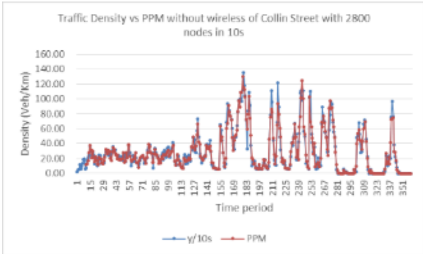


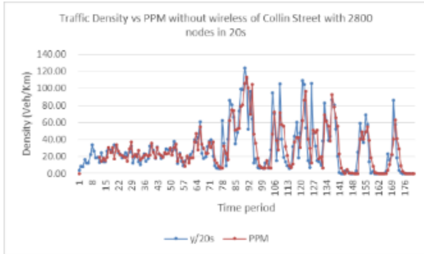


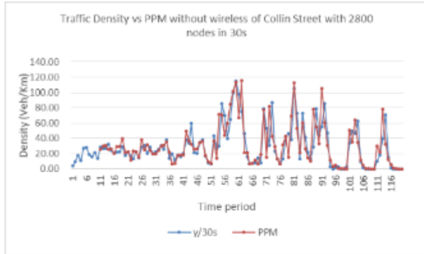


Peaks







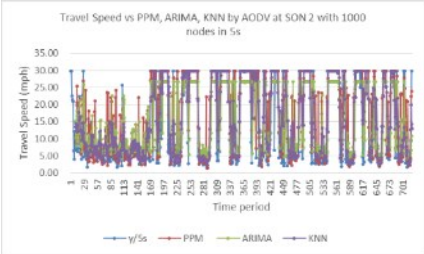


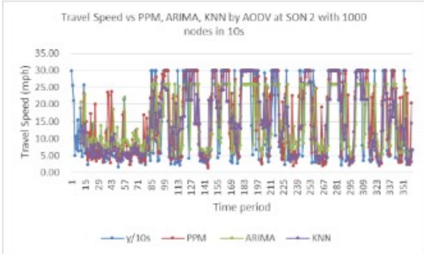
The results of PPM with PRAWMA prediction scheme **without wireless** compare with travel time with **2800 nodes** for Collin Street during 1 hour at 5s, 10s, 20s and 30s time period

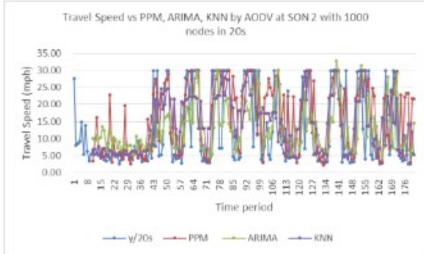
Appendix J: The results of analysis of traffic density prediction in section 5.3.9 (Fig.5.20)

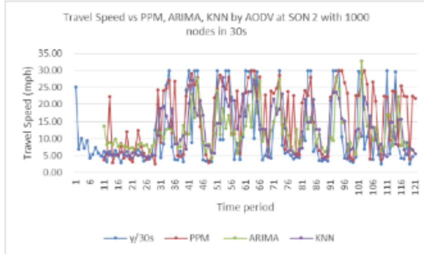
Travel Speed is predicted by PRAWMA, ARIMA and KNN based on the PPM-C2C using AODV for Parliament Street with 1000 and 2800 nodes at SON.2

1000 nodes

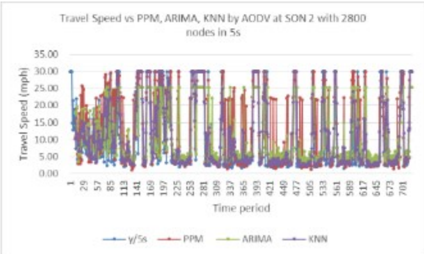


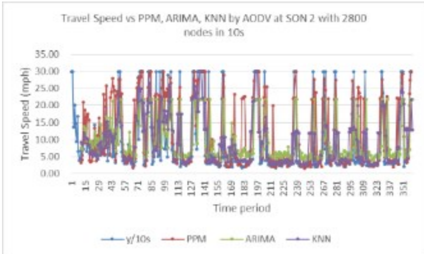


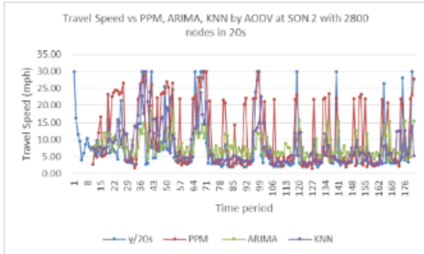


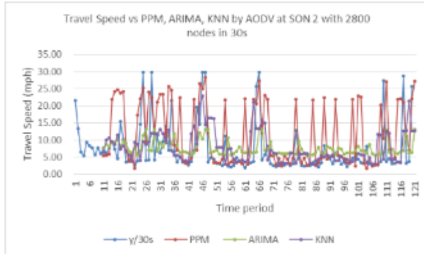


2800 nodes









The results of **PPM** with **PRAWMA** prediction scheme, **ARIMA** and **KNN** using AODV compare with observation data for Parliament Street during 1 hour at 5s, 10s, 20s and 30s time period