# TRAFFIC STATE ANALYSIS OF URBAN AREA FROM WEB-BASED DATA

BY

KIMHEANG LY

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE
(ENGINEERING AND TECHNOLOGY)
SIRINDHORN INTERNATIONAL INSTITUTE OF TECHNOLOGY
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2015

# TRAFFIC STATE ANALYSIS OF URBAN AREA FROM WEB-BASED DATA

BY

KIMHEANG LY

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE
(ENGINEERING AND TECHNOLOGY)
SIRINDHORN INTERNATIONAL INSTITUTE OF TECHNOLOGY
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2015

TRAFFIC STATE ANALYSIS OF URBAN AREA FROM WEB-BASED DATA

A Thesis Presented

By

KIMHEANG LY

Submitted to

Sirindhorn International Institute of Technology

Thammasat University

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE (ENGINEERING AND TECHNOLOGY)

Approved as to style and content by
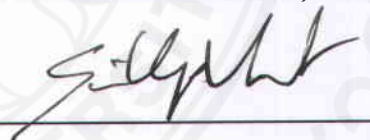
Advisor and Chairperson of Thesis Committee

Assoc. Prof. Mongkut Piantanakulchai, Ph. D.

Committee Member and
Chairperson of Examination Committee

Assoc. Prof. Chawalit Jeenanunta, Ph. D.

Committee Member

Dr. Sornthep Vannarat, Ph. D.

DECEMBER 2015

i

# Abstract

TRAFFIC STATE ANALYSIS OF URBAN AREA FROM WEB-BASED DATA

by

KIMHEANG LY

Bachelor of Architectural Engineering, Institute of Technology of Cambodia, 2013

Traffic state information is known as one of the most important traffic information for the road network. It is a key measure to comprehend traffic situation which can be interpreted into mobility and travel time spending on the road. However, this traffic information is usually the final result of study or the end product information for road user. In this thesis, a number of analyses on traffic state information were carried out. A method to collect the data from web-based mapping service is also proposed. The study covers analysis of temporal traffic state patterns, spatio-temporal traffic state patterns, clustering analysis on structure of data, traffic anomaly detection and last but not least the analysis on traffic state prediction. From these analyses, many interesting results were revealed and discussed. It was concluded that traffic state information from web-based mapping services is very useful for many kind of applications and researches, and the result of study reveal the traffic behavior in the urban area.

**Keywords**: Traffic state pattern, Traffic anomaly detection, Traffic state prediction

# Acknowledgements

First of all, I would like to express my sincere gratitude and appreciation to my advisor Dr. Mongkut Piantanakulchai for the continuous support of my graduate education, for his patience, understanding, motivation, and immense knowledge throughout my study and research. His comments and suggestions helped me drawing the path of my research and writing this thesis. Without his incredible guidance, this research and thesis work would have been a frustrating and overwhelming pursuit.

Besides my advisor, I would like to thank the member of my thesis committee: Dr. Chawalit Jeenanunta, and Dr. Sornthep Vannarat, for their insightful ideas, comments and encouragement, which inspired me to widen my research from various perspectives. In addition, my exceptional thank goes to Dr. Sornthep Vannarat for his initial and brilliant research idea which enlightened my main research topic.

My deep gratitude also extend to ASEAN University Networks/Southeast Asia Engineering Education Development Network (AUN/SEED-Net) program by JICA and SIIT, who provided me an opportunity to continue my graduation study, and financially support me for both academic and living expenses.

I would also like to thank in general to all the people who have helped me with or without my acknowledgment. Special thank goes to my senior and friends in the same research lab as me for their support and friendship all along my research path.

Finally, I would like to thank my family: my parents, my brothers and sister for their unconditional love and support. I would not have been able to complete this thesis without their continuous love and encouragement.
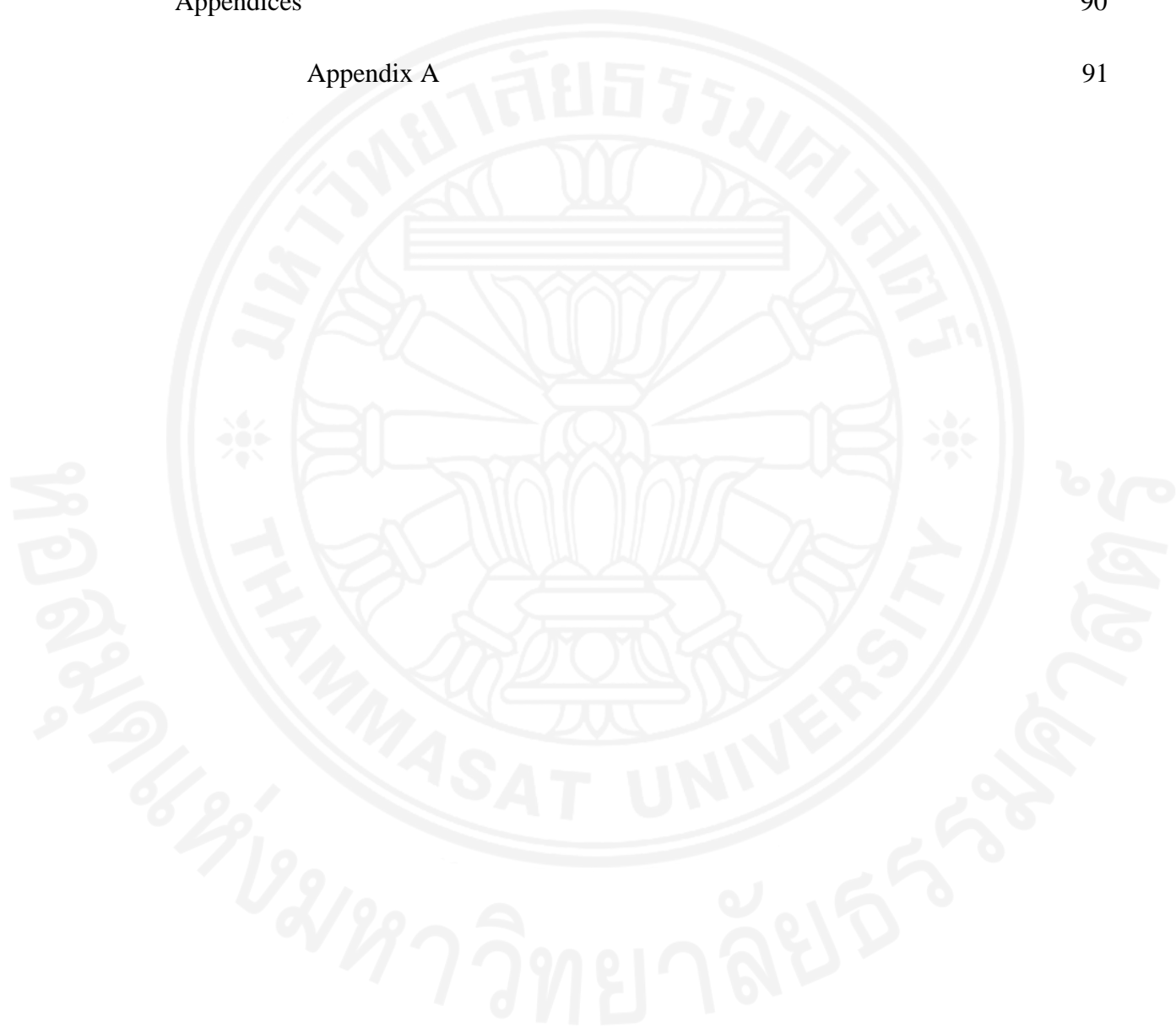
# Table of Contents

# Lists of Figures

# Lists of Tables

# Chapter 1
# Introduction

## 1.1  General

Transportation is one of the major functional systems of modern society in which it provides a service, the movement of goods and people from place to place, that is essential to the functioning of the community as a whole. It cannot be separated from the society because it is an essential feature in the economy and the personal lives of people everywhere, especially in the developed civilization. A highly advanced transportation system can change the way people live by the high levels of personal mobility and also makes possible the abundance and variety of goods.

Transportation system can have a high impact on our society and economic. Most commonly, transportation can be regarded as an economic activity, and major decisions about transportation system are based on economic concerned. Mobility of people and availability of goods and services depend largely on the transportation system. Thus an adequate transportation infrastructure is necessary for a high level of economic activity to exist. Other important roles of transportation system are military and political bases essentially the military movement, and the connection to reach out to people in faraway distance. Despite all of these extensive benefits, public policy still have to place a number of limitations on the transportation system development. The most crucial among them are constraints related to environment such as pollutions, energy consumption and land use.

Road, rail, air and water transport are the major modes of transportation. There are also some particular forms of transport for special purposes such as pipeline transport, mainly for liquid and gas transport, and cable transport such as conveyor belt. These different modes are distinguished in terms of their physical characteristics. Among them, the road transport system is the dominant mode because of its service characteristics such as very high accessibility to almost all potential destinations, direct service with very low door-to-door travel time, moderate speed, and moderate capacity.

1

## 1.2  Background and Motivation

Transportation system exists to meet perceived social and economic needs. It is a major component which gives a lot of advantages and cannot be left out in a developed society. As the social and economic needs change, especially in the urban area, the transportation system also evolves and starts to face many problems and challenges. Similar to the general trend, transportation in urban area is predominantly road transport. Due to high human population density and infrastructure, traffic demand in the urban area is generally high. As the result, road traffic in urban area becomes one of the most challenging issues in the society. Increasing traffic problems in the city such as traffic congestion, safety concern, providing equal access and environmental issues grow into a big obstacle for developing country's economy and dramatically impact life quality of people. These issues create many negative impacts on many aspects such as economic, flexibility and comfort of individual, environment, and society itself as a whole. Alternative approaches to solve and alleviate the problems exist, but in the last recent years, implementation on the advanced traffic management system (ATMS) have been considering as the most preferable approach. Regarded as the traffic management perspective which integrates new technology, this system is set out to improve vehicle traffic flow and safety.

Traffic state information of urban road network, a major contributing factor in advanced traffic management system, is a suitable measure to comprehend traffic situation which can be interpreted into mobility and travel time spending on the road network. With sufficient information, government, road authorities and policy makers can have a better way to understand the traffic in their city such as traffic mobility observation, overall performance of the urban network and investigation of traffic problems; for example, detection of bottleneck in the road network. Moreover, for road users, especially commuters, this information is served as a tool to improve their decision making. By choosing better route and better departure time, both travel time and cost can be reduced. This provides benefit to individual and the economy of the country as a whole. Therefore, the traffic state pattern is becoming a basic and valuable information for understanding and increasing overall performance of traffic network.

2

Last but not least, it has substantial benefit to policy makers, traffic authorities and also to general public.

On the other hand, the opportunity to collect traffic data from web-based mapping services which is considered as a new source of traffic data, is also another motive for this study. The development of new technology in transportation, especially in the field of Intelligent Transportation Systems (ITS), allows us to collect many kind of traffic data and to have many analytical tools. With the advancement of mobile technology, mobility data of road users can be obtained by GPS-enabled devices such as mobile phone and GPS receiver. This possibility allows crowd sourcing technology to collect traffic data which is known as GPS Floating Car Data (FCD). In consequence, many web-based mapping services such as Bing Maps, Google Maps, and OpenStreetMap adopt this technology to give users traffic condition information on the road. Anyone with internet access can get this information from these services. Generally, the study of traffic requires tedious data collection process which includes human resource, time and cost. On the contrary, collecting the traffic condition information from online web-based service is both cost-effective and covers large area of study.

## 1.3 Statement of Problem

As mentioned in the above section, the traffic state pattern especially in urban area is very important in advanced traffic management system (ATMS). So the research problem for this study is finding the traffic state patterns of the urban area with many different considerations such as temporal variation, spatial variation and relationship with land use. Furthermore, a method to collect a good traffic state data from web-based mapping services with large area coverage (whole urban area) is also a topic that should be focused in this study.

Looking as a whole, varieties of problems exist in road network in urban area. There are also many approaches to solve each problem. For instance, the problem of traffic congestion which have long been recognized as a major problem in urban

traffic system. The traditional solution to this problem is to increase the capacity of road (increasing supply) by either expanding the traffic lane or constructing new routes. But due to many constraints such as capital, availability of space, social conflict and potential environmental impacts, this approach needs to have careful considerations before implementation. Furthermore, this method also encourages induced demand. This term refers to the phenomenon when the supply increases, the demand also increases. Therefore, this approach cannot solve congestion problem effectively. Consequently, before implement any countermeasure to the problem, a thorough study and careful considerations must be established because each solution doesn't simply solve the problem but they also affect many more aspects as discussed above. Due to the variability of these factors, knowledge and understanding of traffic behavior and condition of the road network in urban area is a need in order to facilitate not only to policy-makers, planners or engineers in cities, regional entities and on governmental level to set out important decisions, but also to the people in general that use the road network.

In summary, the problem statement of the research could be stated as: "*To investigate the traffic state pattern of urban road network and analyze the traffic state data from web-based mapping services*".

## 1.4 Objective

Regarding as a new opportunity to retrieve traffic condition data from web-based mapping services, a new methodology to analyze and make use of this data is proposed here. The main objective of this research is *"Traffic state analysis from web-based traffic state data"*. The purpose is to study and be able to understand the traffic behavior of the road network in the urban area by proposing a new way to collect traffic state data and method to analyze it. This knowledge would benefit people in general who use urban road network by providing them the information about traffic condition pattern where they commute every day. Furthermore, this information would also contribute to better decision making for road network authority on implementing

countermeasures to road problems. By achieving the main objective, some other sub-objectives can also be obtained such as:

- Develop a method to collect the traffic state data from web-based mapping services.
- Analyze and visualize the traffic state pattern
- Investigate the congestion pattern in the urban area
- Classify the patterns of different day and different location based on the similar characteristic between them.

In addition, this study also focuses on a few other objectives such as finding the correlation of traffic condition between different locations in the urban area and also the development of traffic anomaly detection algorithm as well as the traffic state prediction.

## 1.5  Scope of Study

In order to implement this study, Bangkok, the capital city of Thailand, was chosen as the study area. This research will be focused on vehicular traffic flow on the road network inside Bangkok metropolitan area. This city has two major ring road: inner ring road and outer ring road as shown in Figure 1.1.

- Outer ring road (blue line): Kanchanaphisek Road – Motorway route 9
- Inner Ring Road (red line): Ratchadaphised road (eastern portion), Wong Sawang road (northern portion), Charan Sanit Wong road (western portion).

So the scope of this study is the road network inside the outer ring road perimeter. Additionally, the study of traffic state patterns mentioned in the objective section can only be done in the area based analysis due to the time constraint. The network or corridor based analysis would be a very good extent of this research but it is not included the study.
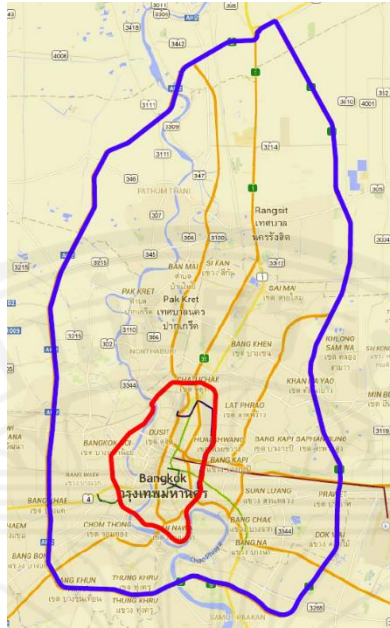
Figure 1.1 Bangkok major ring roads; outer ring road (blue line) and inner ring road (red line)

Image source: Google Maps

## 1.6 Organization of the Report

This thesis is organized into 5 different sections. The first section states the introduction, the statement of problem, the purpose and scope of study. The second section presents a literature review of relevant knowledge, previous research works and developments of the related topic. In the third part, the methodology and procedure of collecting data along with the analyses will be described in detail. The fourth section presents the results of analysis along with the thorough discussion. Finally comes the conclusion which sums up the whole study with recommendation for future study.

# Chapter 2

# Literature Review

## 2.1 Introduction

In this chapter, full theoretical background of the study will be presented. It begins with the introduction to traffic state information. A brief discussion on general concept and influential factors of traffic state is made. Several important classification methods to quantify traffic state are also discussed. After that, available traffic data collection methods are reviewed in order to compare and reflect on the method proposed in this study. Finally, discussion is extended to the theoretical aspect of traffic anomaly detection and traffic state prediction.

## 2.2 Traffic State Information

Traffic flow on the road network is a dynamic phenomenon which keeps changing all the time during the day. Many factors such as change in demand for different time of the day, traffic incident, road work activities, and other special events cause this variation. Traffic state is one of the most useful traffic information that can tell us how the traffic flow on the road is like. It is straightforward and uncomplicated to understand for everyone ranging from specialist in traffic to general road users.

Traffic state measurement of a transportation facility or service is one of the methods to measure the performance or quality of service of that facility. There are many ways to measure and classify this performance and most commonly, the traveler perception is viewed as the main factor of consideration. Many points of view from different agencies such as roadway operators, automobile drivers, pedestrians, bicyclists, decision makers, government and also the community as a whole have different considerations in deciding which measurement to evaluate in order to obtain the traffic state. However, in the study of urban network traffic, only the vehicular

traffic state is focus which is the condition of the traffic on the section of the road network.

According to the Highway Capacity Manual edition 2010 (National Research and Transportation Research, 2010), factors that influence traveler perceived quality of service have been found to include:

- Travel time, speed, and delay;
- Number of stops incurred;
- Travel time reliability;
- Maneuverability (e.g., ease of lane changing, percent time-spent-following other vehicles);
- Comfort (e.g., bicycle and pedestrian interaction with and separation from traffic, transit vehicle crowding, ride comfort);
- Convenience (e.g., directness of route, frequency of transit service);
- Safety (actual or perceived);
- User cost ;
- Availability of facilities and services;
- Facility aesthetics; and
- Information availability (e.g., highway way-finding signage, transit route and schedule information).

As discussed above, the focal point of the study of traffic in urban network is the vehicular movement. Hence the major factors that affect the traffic state mentioned are summarized as travel time, speed of individual vehicle and average speed on the road section, time delay, maneuverability, and comfort aspects of service. Because the traffic condition of the road is a qualitative grading of performance and quality of service, it is generally expressed in different condition such as free flow, moderate flow, heavy flow and congested flow. Similarly, many experimental studies (Kerner (1999), Kim and Keller (2002) & Lee et al. (1998)) have shown distinct dynamic modes of traffic flow:

(i). **Free flow**: cars do not interact much with each other and move with desired speed.

8

(ii). **Synchronized traffic flow**: motorists move with nearly the same speed on the different lanes of the highway.

(iii). **Congested mode**: speed of the traffic is quite low and can fluctuate very much.

(iv). **Jammed mode**: vehicles almost do not move and the flow is very low.

Many concepts have been proposed to classify and interpret these conditions to the quantitative description that give a useful meaning and be informative for the road user and also to the operating agencies. Below sections will describe some of the well-known methods to quantify this traffic condition.

### 2.2.1 Level of Service (LOS)

Level of Service (LOS) is introduced in Highway Capacity Manual (HCM) by Transportation Research Board of The National Academies based in Washington, USA. In the HCM-2010, the traffic state of the road is presented by the concept of the Level of Service (LOS). LOS is a qualitative stratification of a performance measure or measures that represent quality of service. LOS is measured on the alphabetical scale ranging from A to F, in which LOS A represents the best operating conditions and LOS F is the worst from the traveler's perspective as shown in table 2.1.

Table 2.1 Level of Service (LOS) classifications

| Level of Service (LOS) | Traffic Flow | Descriptions |
| --- | --- | --- |
| LOS A | Free flow | Traffic flow at or above the posted speed limit and motorists have complete mobility between lanes and high level of physical and psychological comfort. |
| LOS B | Reasonable flow | Free flow speed is maintained with some restriction on the maneuverability in the traffic stream. The motorists still have a high level of physical and psychological comfort. |
| LOS C | Stable flow | Traffic flow at or near free flow speed. The ability to maneuver through lanes is noticeably restricted and lane changes require more driver awareness. |
| LOS D | Approaching unstable flow | Speed slightly decrease as traffic volume slightly increase. The maneuverability in the traffic stream is much more limited and the driver comfort levels decrease. |
| LOS E | Unstable flow | Traffic operates near or at capacity of the road and the traffic flow becomes irregular. Speed varies drastically and the drivers' levels of comfort drop down to poor. |
| LOS F | Forced or breakdown flow | Traffic flow surpasses the capacity which creates traffic jam frequently. |

The LOS A is generally occurs only late at night in urban areas and frequently in rural areas. While it may be tempting to aim for an LOS A, this is unrealistic especially in the urban areas due to many constraints such as economic

10

factor, land use factor, environmental issues. By knowing which LOS for the road is, one can understand and be informed how the traffic condition on that road is like.

### 2.2.2 Performance Measures

Some studies, i.e. the study of Turner et al. (2004), proved that the traffic condition can be classified by the travel time and it is called "Performance Measures" as shown in Figure 2.1.
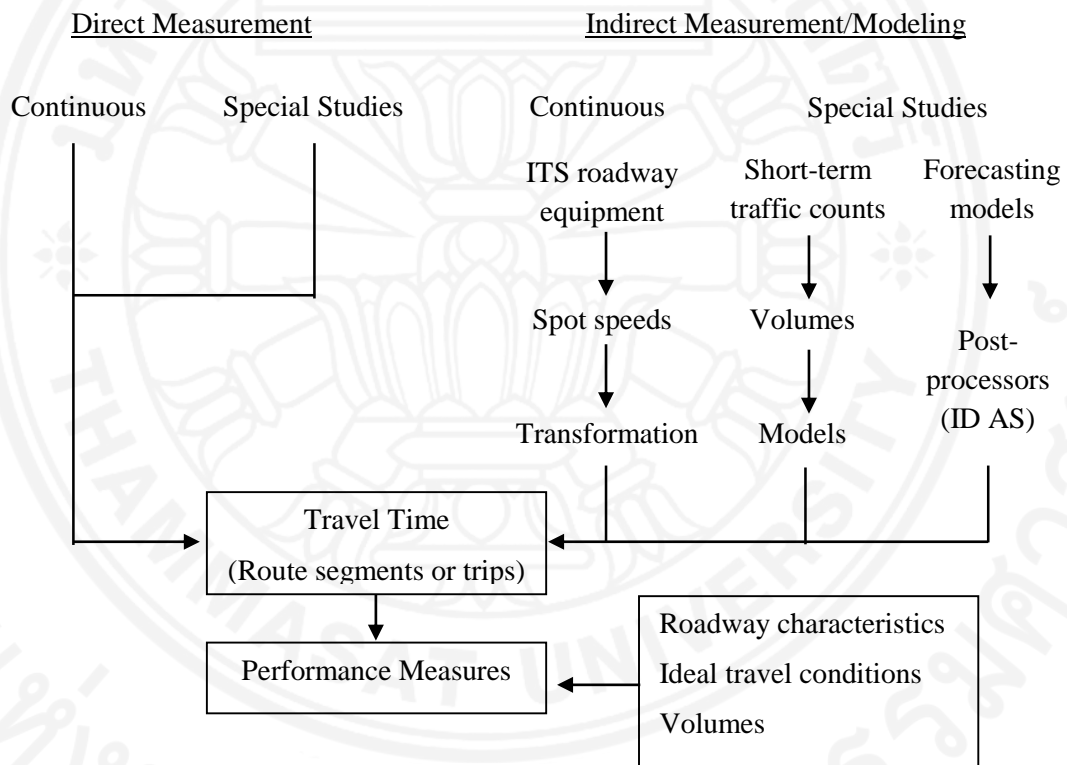


Figure 2.1 Performance Measures as traffic condition classification

Source: Adapted from Turner et al. (2004)

### 2.2.3 Travel Time Index (TTI)

Other research agencies such as Texas A&M Transportation Institute (TTI), the largest transportation agency in the United States, also consider the traffic condition related to travel time. To classify the road condition, they developed an index called "Travel Time Index (TTI)" as in their Urban Mobility Report (Schrank et al., 2012). As for demonstration, Figure 2.2 displays the historical national congestion trend measures in United State with the value of TTI.

| Year | Travel Time Index | Delay per Commuter (hours) | Total Delay (billion hours) | Fuel Wasted (billion gallons) | Total Cost (2011$ billion) |
|------|------|------|------|------|------|
| 1982 | 1.07 | 15.5 | 1.12 | 0.53 | 24.4 |
| 1983 | 1.07 | 17.7 | 1.23 | 0.58 | 26.5 |
| 1984 | 1.08 | 18.8 | 1.34 | 0.65 | 28.9 |
| 1985 | 1.09 | 21.0 | 1.56 | 0.75 | 33.3 |
| 1986 | 1.10 | 23.2 | 1.79 | 0.88 | 37.0 |
| 1987 | 1.11 | 25.4 | 1.99 | 1.00 | 41.2 |
| 1988 | 1.12 | 27.6 | 2.29 | 1.15 | 47.3 |
| 1989 | 1.14 | 29.8 | 2.51 | 1.28 | 52.1 |
| 1990 | 1.14 | 32.0 | 2.66 | 1.36 | 55.2 |
| 1991 | 1.14 | 32.0 | 2.73 | 1.41 | 56.4 |
| 1992 | 1.14 | 32.0 | 2.90 | 1.50 | 60.1 |
| 1993 | 1.15 | 33.1 | 3.06 | 1.57 | 63.1 |
| 1994 | 1.15 | 34.2 | 3.19 | 1.64 | 65.8 |
| 1995 | 1.16 | 35.4 | 3.42 | 1.78 | 71.0 |
| 1996 | 1.17 | 36.5 | 3.64 | 1.90 | 75.9 |
| 1997 | 1.17 | 37.6 | 3.85 | 2.02 | 79.7 |
| 1998 | 1.18 | 37.6 | 4.00 | 2.12 | 81.9 |
| 1999 | 1.19 | 38.7 | 4.30 | 2.28 | 87.9 |
| 2000 | 1.19 | 38.7 | 4.50 | 2.39 | 94.2 |
| 2001 | 1.20 | 39.8 | 4.70 | 2.51 | 98.2 |
| 2002 | 1.21 | 40.9 | 4.97 | 2.67 | 103.7 |
| 2003 | 1.21 | 40.9 | 5.27 | 2.83 | 109.8 |
| 2004 | 1.22 | 43.1 | 5.61 | 3.02 | 119.1 |
| 2005 | 1.23 | 43.1 | 5.91 | 3.17 | 128.5 |
| 2006 | 1.22 | 43.1 | 5.94 | 3.20 | 130.8 |
| 2007 | 1.22 | 42.0 | 5.88 | 3.23 | 131.2 |
| 2008 | 1.18 | 37.6 | 5.23 | 2.76 | 115.3 |
| 2009 | 1.18 | 37.6 | 5.43 | 2.81 | 120.0 |
| 2010 | 1.18 | 37.6 | 5.46 | 2.85 | 120.0 |
| 2011 | 1.18 | 38.0 | 5.52 | 2.88 | 121.2 |

Figure 2.2 Historical national congestion trend measures of United State

Source: 2012 Urban Mobility Report, Texas A&M Transportation Institute (TTI)

Although TTI is a road condition classification index, however it tells only about the overall performance of the road and can act simply as one of the road characteristics because it is the ratio of travel time in the peak period to travel time at free-flow conditions. For this reason, it cannot tell us the traffic condition of road network in real time or of different period of time.

### 2.2.4 Traffic Performance Evaluation Indexes (TPI)

Traffic Performance Evaluation Indexes (TPI) was officially released in the technical standard (Wen et al., 2011) of urban road traffic performance evaluation indexes in Beijing, 2011. The TPI value ranges from 0 to 10, indicating five different congestion levels with the higher value as more serious traffic congestion:

- 0 to 2: smooth traffic
- 2 to 4: basically smooth traffic
- 4 to 6: slight congestion
- 6 to 8: moderate congestion
- 8 to 10: severe congestion

TPI is also used in another research study of Wen et al. (2014). As an example, the Figure 2.3 shows the TPI variation curve of weekdays and weekends of the central area of Beijing city in September 2013.
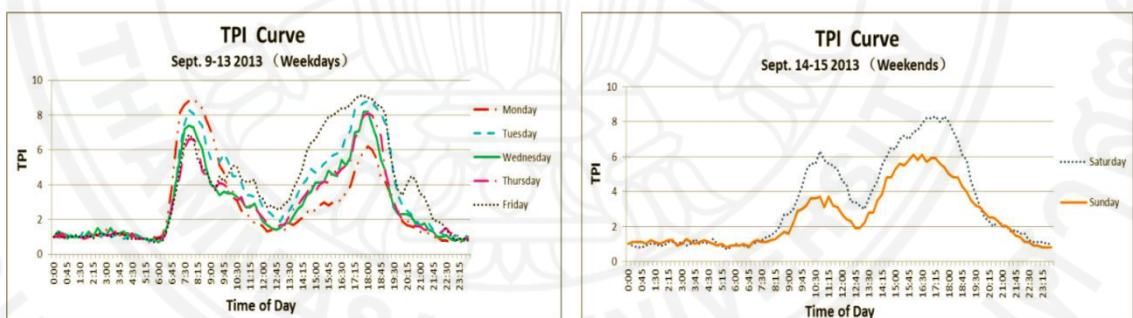


Figure 2.3 TPI variation curve of weekdays and weekends in central area of Beijing city

Source: Wen et al. (2014)

### 2.2.5 Traffic state information by Web Map Service (WMS)

With the rapid development of networking technology and geographic information system (GIS), many private companies have launched their Web Map Service (WMS) which is a protocol for serving geo-referenced map images over the Internet. The two well-known agencies are Google Maps from Google Inc. and Bing Maps from Microsoft Corporation. These services are commonly based on Geographic

Information System (GIS) which give us information on geolocation bases such as terrain situation and road network. From the first establishment of these WMSs, only the basic mapping images were displayed and covered only some part of the world. Presently, much more features are included and the traffic state information is also one among them.

### 2.2.5.1 Google Maps

Google Maps is a desktop and mobile web mapping service application and technology provided by Google Inc., a multinational corporation based in U.S. According to Google Maps official blog, Google Maps has launched a new technology "the traffic layer" as its new feature in 2007. In the beginning, this feature was only available in some major cities in USA. In consequent later years, the company has increased the coverage of the road traffic condition information into more and more countries. This recent feature provides user with the up-to-date traffic condition in real-time by displaying color overlaid on the road network as shown in Figure 2.4 (a). The colors indicate the traffic flow condition of the road section in comparing with its capacity and speed at free flow condition. There are four different colors which represent different traffic conditions as explained in the Table 2.2.

### 2.2.5.2 Bing Maps

Bing Maps is a web mapping service provided by Microsoft Corporation, an American multinational corporation based in Washington. Similar to Google Maps, it offers many features such as street maps (road view, aerial view, 3D maps), driving, walking, and transit direction. In May 2006, Bing Maps released a new update which introduced the real time traffic information on the map. Likewise, this information is represented in the form of color code overlaid on the road network as shown in the Figure 2.4 (b). The interpretation of these colors is the same as those of Google Maps shown in Table 2.2.

14
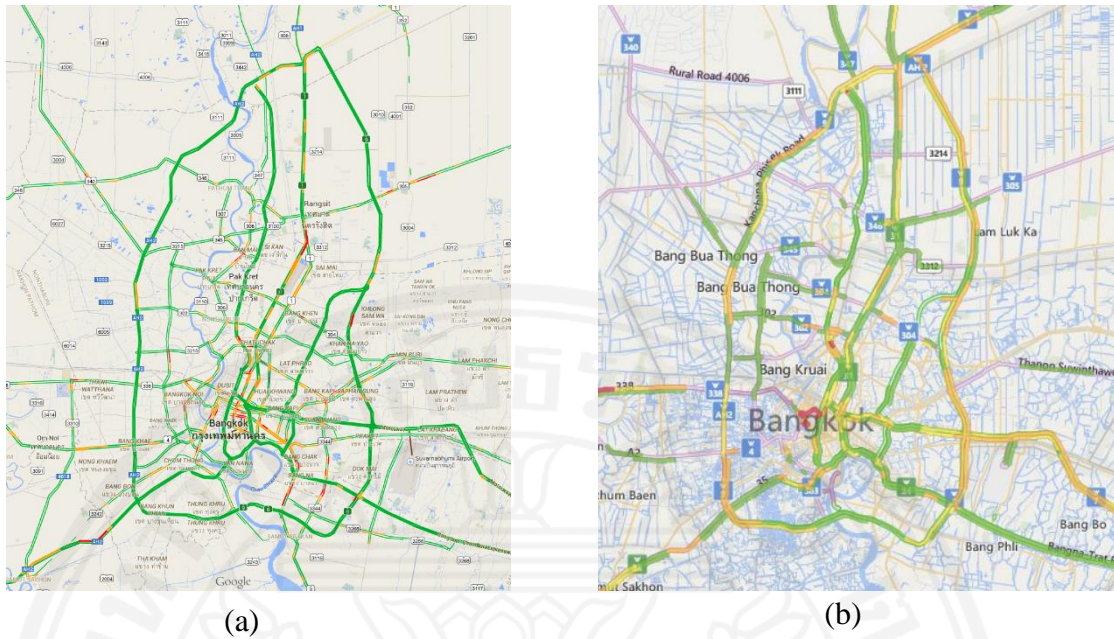
Figure 2.4 Web-based traffic condition by color displays in Bangkok, Thailand.

(a). from Google Maps. (b). from Bing Maps

Source: (a). https://maps.google.com (b). www.bing.com/maps

Table 2.2 Color code of traffic condition classifications

| Color code display | Traffic condition |
|---|---|
| Green | Free flow condition |
| Yellow | Moderate flow condition |
| Light red | Heavy flow condition |
| Dark red | Congested flow condition |
| Grey | No data available |

## 2.3  Traffic Data Collection Methods

The efficiency of traffic studies strongly depends on the availability of traffic data.  Consequently, the data collection method becomes an important aspect for researchers. With the development of Intelligent Transportation Systems (ITS), methods of collecting traffic data have been evolving considerably. According to Leduc

(2008), we can categorize the data collection methods into two groups: conventional data collection method, and Floating Car Data.

The traditional in-situ methods for collecting data such as inductive loops have been employed to get the essential traffic information quite a long time ago. However, these conventional methods cannot reach the expectancy of the study due to their limited coverage and expensive costs of implementation, operation and maintenance. For the last several years, many advanced concepts have been proposed for new methods to collect traffic data in a more efficient manner. As the result, the method based on the vehicle location which is called Floating Car Data (FCD) was discovered. This method is both cost-effective and covers large area of study which can overcome the limitation of the fixed detectors especially for urban road network. The advancement of this technology expands rapidly.

### 2.3.1 Conventional in-situ traffic data collection

Typically, conventional in-situ data collection methods concern about the technologies that collect traffic data by using detectors located along the roadside. These methods can be divided into two categories: the intrusive and non-intrusive methods (Leduc, 2008) .

#### 2.3.1.1 Intrusive methods

The intrusive methods basically refer to the data collection process that employs sensors and data recorders placed in or on the road network and they are generally affect the driver's behavior and traffic stream. These methods have been used in the study of traffic for many years and the most recognized ones are:

- Pneumatic road tubes
- Piezoelectric sensors
- Magnetic loops

### 2.3.1.2 Non-intrusive methods

In contrast to the intrusive methods, the non-intrusive data collection methods cause minimal disruption to normal traffic operations and the driver behavior (Gribbon, 1998). These methods are based on remote observations and the most common among them are:

- Manual counts
- Passive and active infra-red
- Passive magnetic
- Microwave radar
- Ultrasonic and passive acoustic
- Video image detection

### 2.3.2 Floating Car Data (FCD)

Different from the conventional methods, the principle of FCD is to collect real-time traffic data by locating the vehicle via mobile phones or GPS over the entire road network. Each vehicle that carries GPS-enabled device or mobile phone that come with the GPS technology acts as the sensor for road network. Data such as car location, speed and direction of travel are sent anonymously to a central processing center. With sophisticated algorithms and analysis methods in the central processing center, useful traffic information can be generated and extracted, in particular the traffic state information of the road network. Comparing to the traditional methods of collecting traffic data such as fixed detectors (inductive loop detector, video camera and other traffic sensors), the floating car data technique can collect the location, speed information of individual vehicle on the road with more cost effective, larger scale and in real time. Collecting across thousands of devices then combine them together, the actual traffic condition on the road network can be obtained.

Floating Car Data technology is an alternative or rather complement source of high quality data to existing conventional technologies. It is a crucial factor in the development of new Intelligent Transportation Systems (ITS). There are two main types of FCD: GPS-based and cellular-based system.

### 2.3.2.1 FCD based on GPS

In the primary phase of this technology, only a limited number of cars could participate in this system, typically fleet management services (e.g. taxi drivers). Later on, with the reasonable price of GPS-enable device and also the integration of GPS technology in mobile phone, this data collection method becomes more and more popular with larger coverage area. The accuracy of the vehicle location is relatively high as well. Through this method, traffic data obtained from the private vehicles are suitable for highway and road in rural areas. In urban area, signals can be lost due to large buildings, trees, tunnels, or parking garages (Turner et al., 1998).
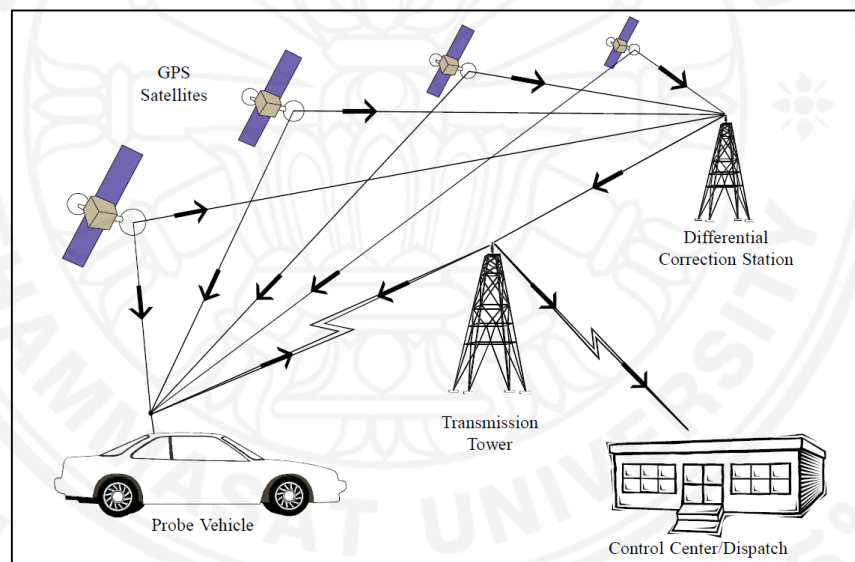


Figure 2.5 FCD based on GPS probe vehicle system

Source: adapted from Perry and Turnbull (1997)

### 2.3.2.2 FCD based on cellular phones

Instead of using the GPS technology, this system collect the location data from cellular service in mobile phone of motorists in the road network. The mobile phone positioning is regularly transmitted to the network usually by means of triangulation as shown in Figure 2.6. No special device, hardware or infrastructure

18

installation is necessary in vehicle and also along the road. Therefore it is more cost effective and since most of the driving vehicles are equipped with at least one mobile phone, this method is great in terms of area coverage. The accuracy of the location information is lower than GPS, but this weakness is compensated by the large number of probe vehicles. This approach is particularly suitable in urban areas because of the closer distance between antennas.
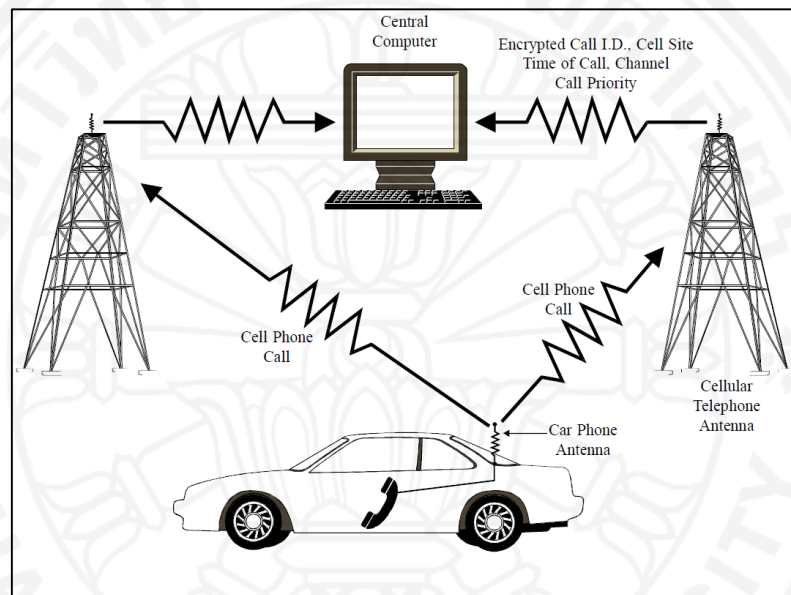


Figure 2.6 FCD based on cellular phones

Source: adapted from Turner et al. (1998)

## 2.4  Traffic Management and Control

In the past few decades, the technology has advanced so rapidly, especially the revolutionary of computation and communication. As the computational cost was decreasing and the existence of microprocessor in variety product ranging from automobile to mobile phone, Intelligent Transportation Systems (ITS) was created based on the advancement in traffic system surveillance and control projects, variable message signing, signal optimization, and simulation. Intelligent Transportation Systems (ITS) are advanced applications which without embodying intelligence as such aim to provide innovative services relating to different modes of transport and traffic

19

management and enable various users to be better informed and make safer, more coordinated, and 'smarter' use of transport networks (EU Directive 2010/40/EU, 2010).

One of the primary subfield within the ITS is the Advanced Traffic Management System (ATMS). This system is intended to integrate technology improving the flow of vehicle traffic and safety. According to the National ITS Architecture, established in 1994 by the United States Department of Transportation, ATMS functional areas are:

- Real-time traffic monitoring
- Dynamic message sign monitoring and control
- Incident monitoring
- Traffic camera monitoring and control
- Active Traffic Management (ATM)
- Chain control
- Ramp meter monitoring and control
- Arterial management
- Traffic signal monitoring and control
- Automated warning systems
- Road Weather Information System (RWIS) monitoring
- Highway advisory radio
- Urban Traffic Management and Control

Most of these functional areas are related directly or indirectly with the traffic state information. Hence, the study associates with the traffic state of the road network would benefit the ATMS. In the following subsections, the review focuses on the traffic state pattern and analysis, traffic anomaly detection and traffic state prediction.

### 2.4.1 Traffic state pattern and analysis

Traffic can have interesting and different patterns at multiple spatial and temporal scales. Traffic state pattern and analysis refer to the investigation of the traffic

state of the road network with different spatial and temporal scales. For different time scales such as daily (e.g., Monday versus Saturday), weekly (e.g., first week versus last week of the month), and monthly (e.g., January versus October), the traffic pattern might behave differently. On the other hand, for different spatial scale, we are interested in the traffic dynamics with various geographical scales and concepts such as at central business district (CBD) area. The traffic state analysis is also the attempt to investigate underlying characteristics of the traffic condition on the road network. These studies are generally performed in the urban road network and there are various approaches to take.

### 2.4.1.1 Graphical method

Graphical method is an approach in scientific or exploratory visualization concept. The most commonly used approach to monitor and recognize the traffic state patterns is the graphical interpretation (Song and Miller, 2012). For traffic analysis, this approach is involved the display of the colors of pixels in the image (map of a location) that represents the macroscopic traffic state of that location. Duan et al. (2009) employed this method to observe and analyze the traffic state of road network by using the pseudo-color map. The authors used the color of pixels on the image to represent the corresponding traffic states condition and the darker the color, the more congested traffic. Finally, congested regions can be defined automatically by reading the image color intensity. Figure 2.7 shows an example of the pseudo-color map that represents the traffic condition for different location.
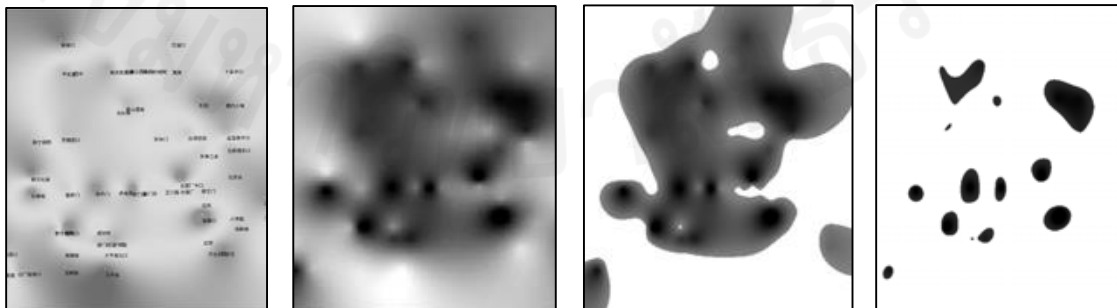


Figure 2.7 Traffic states analysis using pseudo-color map
Source: adapted from Duan et al. (2009)

21

In another study of Song and Miller (2012), the authors used exploratory visualization method to investigate a large traffic flow data set by using space-time plots with the multidimensional data base management technique. The results of study were plotted in visualization graphs that can be seen in Figure 2.8 as an example.
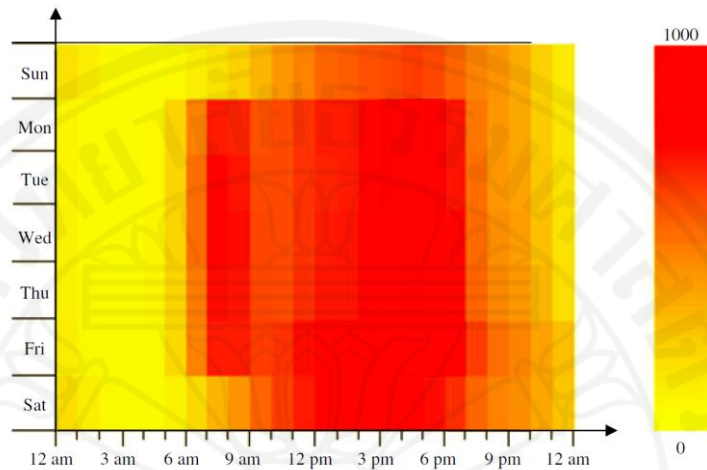


Figure 2.8 Weekly traffic patterns by using exploratory visualization

Source: adapted from Song and Miller (2012)

From their visualization graphs, the authors of the paper revealed the pattern of traffic with different temporal and spatial scaled. Each result comes with reasonable explanation disclosed by authors with backup evident. From this study, we can further confirm the potential of graphical method to reveal patterns about traffic and trends hidden in the database.

**2.4.1.2 Clustering method**

For the former method, it requires human intervention in order to explore and analyze the data. It also faces some challenges such as high data dimensionality and large data volume which create great complexity and visual clutter for the analyst to explore the patterns. On the other hand, another well-known approach in the traffic pattern analysis is the clustering method. This method is an exploratory data mining process and a common method for statistical data analysis. It is an algorithm in which the general task is to identify and group samples of data according to certain similarity requirements (Yildirimoglu and Geroliminis, 2013). Many studies such as the study of

Wen et al. (2014) which focused on the road network inside Beijing city utilized this method. The authors used the floating car data (FCD) from taxi to generate the traffic state of the road network. Then they investigated the traffic state pattern of the city and classify them by different hierarchical clustering methods. As the result, many different categories of traffic condition patterns have been found. Similarly, the study of Payne (1997) also employed the hierarchical clustering to analyze highway traffic flow patterns. From the result of the clustering analysis, the authors revealed distinguished patterns for different days of the week. Lastly, they concluded that the results of traffic patterns can also be used as input for macroscopic traffic models and as the basis for traffic prediction. Another study from Guardiola et al. (2014) used the clustering method along with the functional Principal Component Analysis (PCA) to monitor and recognize the traffic state pattern of the road network.

### 2.4.2 Traffic state estimation

As mentioned above, the traffic state information is very crucial in ATMS. Because the vehicular traffic behavior is highly nonlinear with many complex interactions between vehicles, the problem of having traffic state information is quite a challenge (Helbing, 2001). Moreover, the traffic state is a derived measure from other traffic data, principally speed, volume and occupancy (SVO), so there is a need for method or algorithm in order to define this information. Generally, there are three main key components to estimate the traffic state: a traffic flow model, a state variable representation, and a set of measurement equations (Deng et al., 2013). Many studies have proposed various methods to estimate and predict this information. One of the famous traffic flow models is the Cell Transmission Model (CTM), proposed by Daganzo (1994). This model captures the traffic flow volume between short sections of the road (cells), and keeps updating the flow in traffic stream. This allows us to observe the transitional behavior of traffic and therefore get the traffic state of the road section. Until now, many researchers still putting effort into the study of traffic state estimation models based on CTM such as the study of Sumalee et al. (2011) and Muñoz et al. (2003).

The study of Mihaylova et al. (2007) presented a method to estimate real time traffic state in freeway networks by means of the particle filtering framework. In this study, the authors mentioned the importance of traffic state information and proved the method used in the paper is suitable for it. Another study by Deng et al. (2013) proposed a method to use multiple data sources, including loop detector counts, AVI Bluetooth travel time reading and GPS location samples, to estimate macroscopic traffic states on a homogeneous freeway segment. Antoniou et al. (2013) presented an approach for local traffic state estimation and prediction. This study exploited available traffic information and used data-driven computational approaches. The authors proved the effectiveness of the method to use in the context of mesoscopic traffic simulation models. However, this work can only apply to the freeway without considering whole road network.

### 2.4.3 Traffic anomaly detection

As mentioned in the introduction section, vehicular traffic in urban area is one of the most challenging issues in our modern society. Traffic congestion is a quite common problem that city dwellers as commuters have to deal with every day. According to Akatsuka et al. (2013), there are two different type of congestion: spontaneous and abnormal:

- Spontaneous congestion: recurring traffic congestion, induced by several factors such as road design, urban planning, and high traffic demand. This type of congestion is easier to detect since its time and location can be estimated by statistical methods and by the study of traffic flow (Asakura et al., 2015).
- Abnormal congestion: caused by anomalies in traffic such as traffic incident and special event, which affects the flow of traffic on the road. These anomalies are unpredictable and can cause transport systems to fail.

Therefore, being able to detect anomalies in traffic flow is of a great importance in advance traffic management system (ATMS) and intelligent traffic

24

information. There is a wide body of literature on the traffic anomaly detection. In the early studies of this topic, many algorithms such as standard normal deviation (SND) algorithm by Dudek et al. (1974), California algorithms by Payne (1997), and University of California, Berkeley (UCB) algorithms by Lin and Daganzo (1997) are based on fixed observation system. This means that these algorithms perform on traffic data that is collected from the traditional fixed point observation data. They are discussed briefly as follows. The SND algorithm compares the probability distributions of historical accumulated traffic data with the probability of current traffic data to detect the occurrence of an incident or an anomaly in traffic. For California algorithm, the authors use the space-time variation of current occupancy to compare to a range of value estimated from historical data. In a similar aspect, UCB algorithm also creates a threshold from the past accumulated data in order to compare with the current statistical fluctuations of current time occupancy. The value of current time occupancy that is out of the threshold of the past data signify the occurrence of anomaly.

Later, when the new methods of traffic data collection are available such as floating car data (FCD), there are some more study on this topic in order to apply to the new kind of data. In the study of Thorndike (1953), the authors proposed an algorithm to detect incidents on freeways called mobile sensor and sample-based algorithm (MOSES). It is based on the statistical difference between travel time from two groups of probe vehicle collected before and during an incident. Another study from Zhu et al. (2009) also performed the incident detection by using FCD. This paper introduced the concept of outlier mining into Automatic Incident Detection (AID) in order to detect traffic anomaly not only on freeways but also the urban arterial roads. The authors used the variation of speed in such a way to define the speed differences between neighboring sections and previous time intervals. Finally, using distance-based algorithms, outliers are detected by the distance measures and are considered as traffic anomalies. In a more recent study by Kinoshita et al. (2014), the traffic anomalies are detected by vehicle movements. They developed a method to distinguish the abnormal movements of vehicle from the normal ones and from those occurring during recurring traffic congestion. All of these methods described above have their own advantages and

disadvantages. Choosing one method to implement is rather by the type of data available and also the final application to implement.

### 2.4.4 Traffic state prediction

Similar to the traffic state estimation, the traffic state prediction is referred to the method that can tell and predict the traffic condition on the road network in advance, commonly short distance future by using historical and real-time traffic data. This is also a key factor for the implication of traffic management system. In general, according to Van Lint and Van Hinsbergen (2012), there are two main categories for the traffic prediction method: parametric and nonparametric approaches.

Parametric approaches are methods that use predetermined model structure and parameters based on some theoretical assumptions. Some well-known methods in the literature are kinematic wave model (Newell, 1993), cellular automata model (Nagel and Schreckenberg, 1992), Boris Kerner's traffic model (Kerner, 2004). There are a lot of implementations on these methods and also many modifications as well. The limitation of the methods in this categories is that they are based on ideal assumptions, poor data type support and computational inefficiency.

In contrast to parametric approach, nonparametric approaches are not based on predetermined model structure and parameters. Several modeling approaches have been used by many researcher such as Kalman Filter (Liu et al. (2006); Wang et al. (2006)), support vector machine (SVM) (Wu et al., 2004), neural networks (Vlahogianni et al. (2005); Vlahogianni et al. (2008); Dunne and Ghosh (2011)) and Bayesian prediction (Sun and Sun, 2015). In general, all of the approaches presented above use the conventional traffic data such as traffic flow, speed and density. Apart from the rest, another study from Tostes et al. (2013) investigated on usage of traffic state data from web-based mapping service called Bing Map in order to predict the future traffic state. The authors used graphical interpretation of traffic state data that is collected in form of image data. Then image processing algorithms were developed to capture these traffic state on the road. After that the information was processed by using

logit model to predict the traffic state. Again, choosing a method to implement is mostly done by the analyst decision based on the type of data available and final desired result.

## 2.5 Summary

According to the extensive literature review above, a large number of researches were conducted on the optimization algorithms and strategies for traffic management and control. Most of the studies in this field focuses on the traffic state of the road network. Thus imply that the analysis and observation of the traffic state in the road networks play an important role in efficient dynamic traffic management and control. In the purpose of contribute to the solving process of the urban traffic problem, i.e. traffic congestion, the traffic state analysis that can cover the whole urban area appears to be very crucial. However, there is still a small body of research focused on that problem. Most of the studies on traffic state analysis covers only some corridors or a small area of road network due to some limitations such as:

- Availability of data
- Method of collecting the data
- Budget constraint
- Complicated data analysis process

On the other hand, as presented in the above review, the traffic state data from web-based mapping services becomes a suitable source of data that both cost effective and large area coverage. However, a sophisticate methodology need to be proposed in order to collect and analyze this traffic data.

# Chapter 3

# Methodology

## 3.1 Introduction

In general, the study of traffic behavior is always based on the conventional traffic data such as volume of traffic (e.g. Annual average daily traffic (AADT)), speed of vehicle moving on the road, density, and occupancy ratio of the road, etc. There is a very big body of literature study on various problems and investigate new techniques in the field of vehicular traffic by using these conventional traffic data. Unlike these traffic information, the traffic condition or traffic state of the road is usually the final result of study or the end product information for road user. In contrast to general trend, an attempt to investigate and use traffic state data on the road is proposed here in this study.

In this chapter, detail of the methodology, process of data collection and process of data analysis will be discussed thoroughly. For the purpose of studying the traffic state pattern and analysis, this study intended to use traffic state information from web-based mapping services such as Google Maps, Bing Maps, OpenStreetMap, etc. Because this traffic information is given in the form of image data, there is a need in developing an algorithm to manage this information. Then the process of analysis will be described step by step in the following section.

## 3.2 Online Traffic State Information: new type of traffic data

There are many kind of traffic data that researchers and road authorities collect in order to conduct various analysis as presented in literature review above. Some of them are location data based on GPS, floating car information data, video data from CCTV, and data from traffic sensors such as inductive loop detector, infrared detector, ultrasonic detector and many more, which are being used in wide area of research and applications. However, in the last few years, many web-based mapping

28

services such as Google Maps, Bing Maps, OpenStreetMap, etc., are available for general users to access online. These services provide general mapping application and information related to road traffic. One of the most important traffic information available is the traffic state information. It is shown as color layer superimpose on the road network. There are 4 colors, green, yellow, light red, and dark red, which represent different traffic conditions from free flow, moderate flow, heavy flow and congested flow, respectively. Seen as a good opportunity, online traffic state data appears to be a new source of traffic data that need to be discovered and put into good uses.

This online traffic state data is available through the crowdsourcing technology of Floating Car Data (FCD) particularly from GPS of the vehicle on the road. Online mapping companies collect the location and speed information of individual vehicle on the road via mobile phones with GPS. Then combining across thousands of vehicle moving around in the road network, live traffic conditions can be estimated. These information are continuously collected then display it back as the traffic state information. This technology keeps expanding as the number of cell phones equipped with GPS is rising every day and data plans get less expensive. The crowdsourcing method is an efficient way to harness the location data as the more people participate (more probe vehicles), the better the resulting traffic reports get.

## 3.3 Data Collecting Method and Area of Study

The online traffic state information is in the form of color representative as mentioned in above section. By accessing this information, it displays only color of image on road network in the map without any additional numerical data. So the data we can obtain is image data. The image captured need to be adjusted in an appropriate zoom level in order to see not only the detail of the road conditions on major roads, but also for collectors and arterial roads inside the city. This data is generally obtainable for urban areas. It is also important to know that online traffic state information is only available for several big cities around the world. But the availability of this data keep expanding as there are more and more people participate in this technology by the mean of increasing number of smart phone and vehicle on the street.

For the purpose of this study, Bangkok, capital city of Thailand, and its surrounding area is taken into the area of study. Covering nearly whole urban areas of Bangkok, a study area of 35 km x 70 km containing road network inside the outer ring road as mentioned in the scope of study has been selected as shown in Figure 3.1.
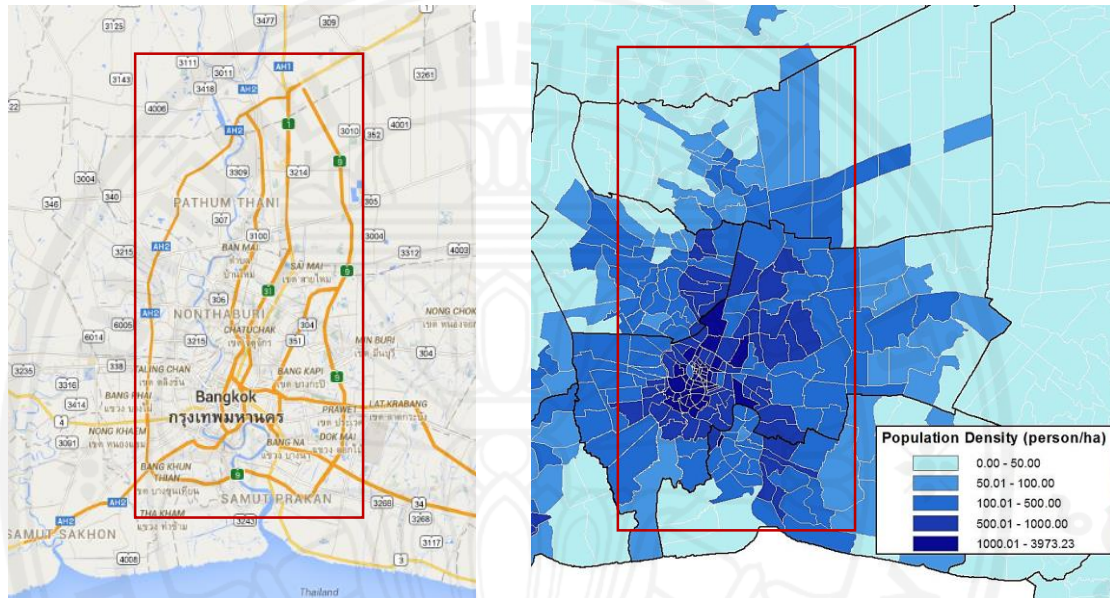


Figure 3.1 (a) Study area (35 km x 70 km), within outer ring road of Bangkok (image source: Google Maps). (b) Population density map of Bangkok, Thailand (adapted from Transport Fundamental Geographic Data Set (FGDS), Ministry of Transport of Thailand, 2007)

The duration for data collection is this study is full one year starting from September $1^{st}$, 2014 to August $31^{st}$, 2015. In our data collection process, there are two main steps needed to be accomplished:

- Develop an algorithm to call for the display of map on the area of study (web crawler);
- A program to capture the image displayed.

### 3.3.1   Data acquisition algorithm (web crawler)

In order to obtain the image data, an algorithm to call for webpage to open the map and enable the traffic layer is needed. Real-time traffic state data is collected

for every 10 minutes. Likewise, a web crawler algorithm is developed to reload itself for every 10-minute time step. The process of this algorithm is shown in the Figure 3.2 and explained as following. First of all, the algorithm requests to a web browser (any web browser according to analyst preference) to access to the website of web-based mapping service. Then it requests the location by the input of coordination. The map display is centered by that coordination and appropriate zoom level is set. After that, in order to get the traffic state information, the traffic layer is turn on. With the purpose of getting the traffic information (color layer), the background (image of the map) is faded off which can facilitate the image analysis process later on. Until this step however, there are still many displayed objects which often obstruct the view of our data, so various features appeared on the map such as labels, icons and toggle switches are turn off from the display. Finally, plain traffic layer is displayed on the screen, ready to be captured and after that close itself. Thereafter, the algorithm reloads itself every 10 minutes and the process is repeated continuously.

This study is intended not only to cover highways and major roads, but also collector roads and local roads inside urban network. Thus an appropriate zoom level (zoom level equals to 16 in our algorithm) is needed in order to display these traffic layer correctly. To cover the whole study area, one screen from the web page is not enough, so multiple screen displays are necessary. With this zoom level, a total of 32 displays is required. In the former method, 8 OS Virtual Machines from Oracle called "VirtualBox" are used, each with 4 displays, resulting 32 displays as needed. In this method, there was a high requirement processing and heavy load on the computer resulting errors and missing data. Later on, this method is replaced by a new method by compiling script in Python to mimic the behavior of the first method. This method require less performance from the computer and produce less error.
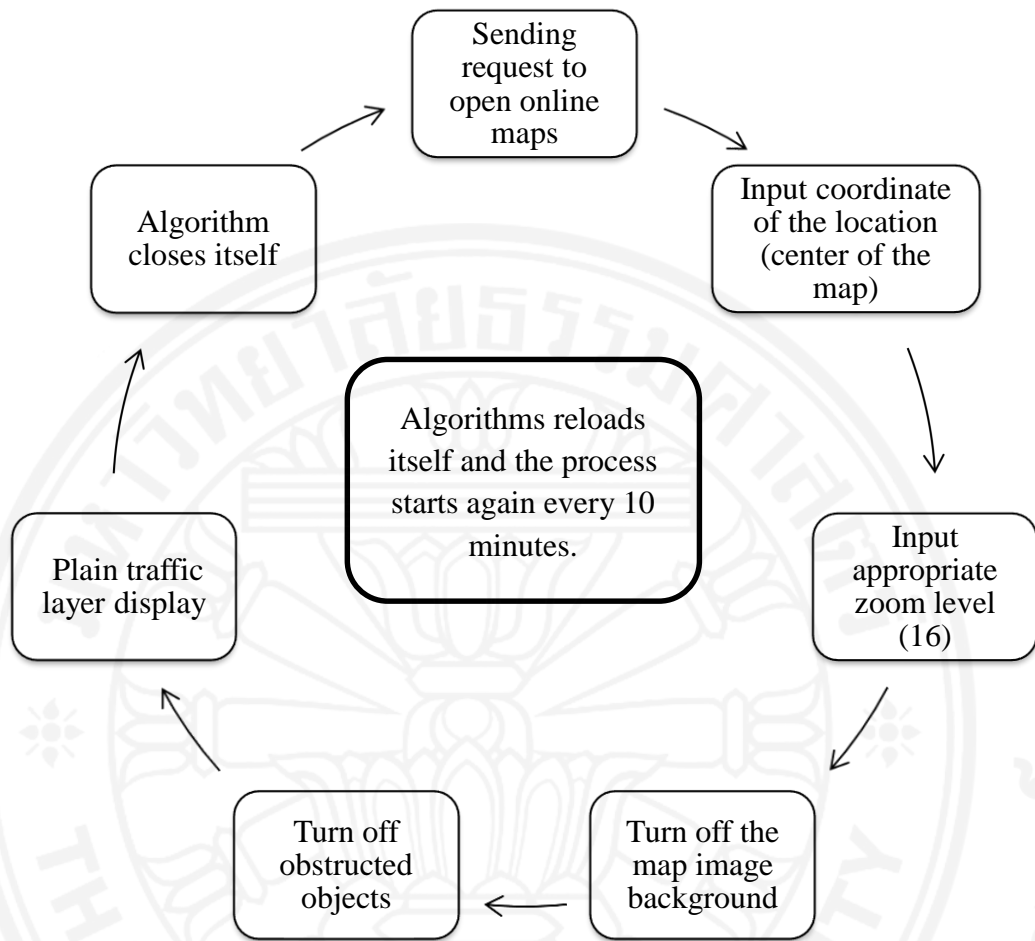
Figure 3.2 Process of web crawler algorithm

### 3.3.2 Image data acquiring method

After having all the displays correctly placed on various determined locations, the image data of each screen is captured and stored in a database. Each of the image has the date and time of capturing as file name. For one screen there are in total 144 images corresponding to 144 10-minute time bin for a single day. An example of the image is shown in Figure 3.3.

32

Figure 3.3 An example of image data captured. (a). traffic state data on one screen display captured on Dec 16, 2014 at 5:00 PM. (b). zoom in image on the Victory Monument roundabout (ellipse in the center).

## 3.4  Image Processing

After getting the image data from all the displays, the traffic state data of the whole urban network can be obtained. In order to capture these information for further analysis, these images need to be processed into quantitative data. The process of transforming these image information into quantitative information is presented in detail in the following sub-sections.

### 3.4.1  Road network masks and virtual sensor placement

With the traffic layer in image data as an example shown in Figure 3.3, all the road network were traced down manually with the help of CAD program (Computer-aided Drawing program). The complete drawing of study area in Bangkok urban network is shown in Figure 3.4. In total, more than 4000 km length of road was traced and created a whole urban network mask.

33

Figure 3.4 Study area road network mask

Different from the study of Tostes et al. (2013), who used an image processing software to extract the road traffic intensity by saving the percentage of different color pixels, this study intended to extract the traffic intensity by developing virtual sensors along the road network. In the topic of traffic sensor placement, typically, higher density of sensors produces higher overall accuracy of the measurement. This means a maximal number of sensors is a good way to achieve the best possible measurements (Danczyk and Liu, 2011). Unfortunately, due to the budget constraints, such work cannot be attained. In the most common case for many metropolitan highway in the United States, traffic sensors (e.g. inductance loop detectors) have been placed uniformly along the road and the distance between two neighboring sensors is about 1/3 to 1/2 miles (0.5 to 0.8 km). Furthermore, according to practical application such as The NaviGAtor, the Intelligent Transportation System of Georgia Department of Transportation, the traffic sensors, in this case Video

Detection System (VDS), are installed approximately every 1/3 mile (about 0.5 km) along the major highways. Another type of traffic sensor, in this system also, Closed-Circuit Television (CCTV) cameras are positioned about every 1 mile (1.6 km) (Ni, 2013).

For this study however, higher density of traffic sensors to detect the traffic state along the road can be achieved by the mean of virtual sensors. Points with the distance of 250 meters apart from each other were set on the road network masks as shown in Figure 3.5. A total number of more than 17,000 points were distributed all over the road network to record the data. Therefore, color of the pixel those points located can be detected and later on converted into traffic condition information.



Figure 3.5 An example of an area (same location to the Figure 3-3) with points act as artificial traffic sensors (blue dots).

### 3.4.2 Quantitative classification of traffic state data

As mentioned above, the artificial sensor can detect only the color of the pixel it locates in, so there is a need for an image processing algorithm to interpret these color into numerical indexes. Table 3.1 show the classification index according to our type of data. This conversion algorithms is shown in Appendix A and it is run in the program MATLAB.

Table 3.1 Traffic state classification indexes

| Traffic State | Color | Index assigned |
|---|---|---|
| Free flow condition | Green | 1 |
| Moderate flow condition | Yellow | 2 |
| Heavy flow condition | Light red | 3 |
| Congested flow condition | Dark red | 4 |
| No data available | Gray | N/A |

The data collection process is an ongoing process for now; images are being captured and the image processing algorithm converts those image data to quantitative data and then store them all in one big database.

## 3.5  Area-Based Traffic State Analysis

The traffic state analysis is the study of the traffic condition that investigate the behavior of traffic with both temporal and spatial variations. It presents the common patterns of the traffic condition and examines how the traffic state changes from one state to another. Because the traffic state data of whole area of the urban network can be collected, there are many approaches to extend the study such as:

- Area based analysis: discretize the whole into smaller grid, then analyze the data based on grid system
- Corridor based analysis: some important corridors are selected for the study
- Network based analysis: study based on the road network

However as mentioned in the scope of study, due to the time constraint, this study is limited to only the area-based analysis. Similar to the study of Liu et al. (2012), the area of this study was discretized into smaller grid system. With the dimension of 35 km by 70 km, a total number of 2450 small cells with the size of 1 km x 1 km were

generated as shown in Figure 3.6. Each small cell contains different number of artificial sensor; some cells has up to 50 points, some contains only a few point, while some others don't cover any point (blank cell). Analysis approaches of this study is based on these small cells. Each cell gets the traffic state value by aggregating the value in the method of arithmetic mean to find the average value of the points it contains.

$$\overline{I}_{ij} = \frac{\sum\limits_{k} I_k}{K} \tag{3.1}$$

Where: $i = 1,\ldots, 35$ number of columns

$j = 1,\ldots, 70$ number of rows

$k = 1,\ldots, K$

$K$: number of detector points inside cell $ij$

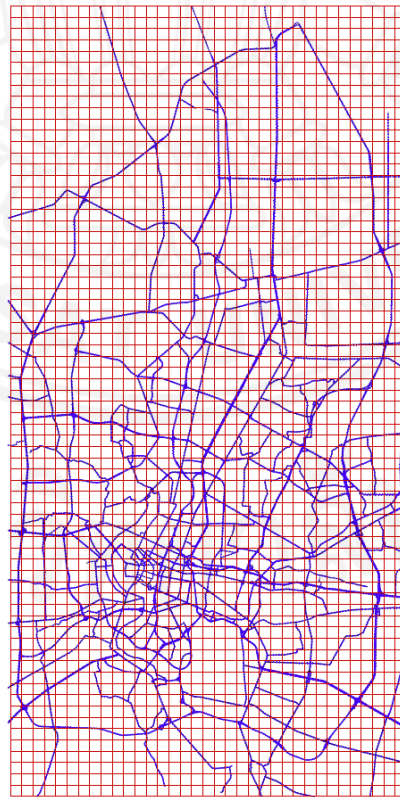$\overline{I}_{ij}$ : average intensity of traffic of cell $ij$



Figure 3.6 Small 1 km x 1 km grid system

### 3.5.1 Temporal pattern of traffic condition

Traffic flow conditions vary widely throughout the day. The study on temporal pattern of the traffic condition give an overview of how traffic evolve. For this purpose, traffic state data is collected for every 10 minutes. The whole area traffic pattern can be obtained by averaging the value of all cells.

$$I' = \frac{\sum_i \sum_j \overline{I}_{ij}}{N} \tag{3.2}$$

Where $N$: number of cells with $\overline{I}_{ij} \neq 0$

$I'$: whole area traffic intensity at one instant

$\overline{I}_{ij}$: average intensity of traffic of cell $ij$

Patterns of different day of the week and different month of the year will be investigated.

### 3.5.2 Spatio-temporal pattern of traffic condition

At the same instance, traffic flow is also not uniform with different location. Thus this section is intended to investigate this behavior. Three different considerations are proposed here:

- Spatio-temporal pattern of the whole urban network
- Spatio-temporal pattern based on the distance from Central Business District of the city

### 3.5.2.1 Spatio-temporal pattern of the whole urban network

This study aims to investigate the concentric behavior of traffic in the urban area. Heat map will be used to represent the traffic intensity over the whole road

38

network. The spatio-temporal pattern of the city will be discussed and presented in form of heat maps. An example of heat map of the city is shown in Figure 3.7.
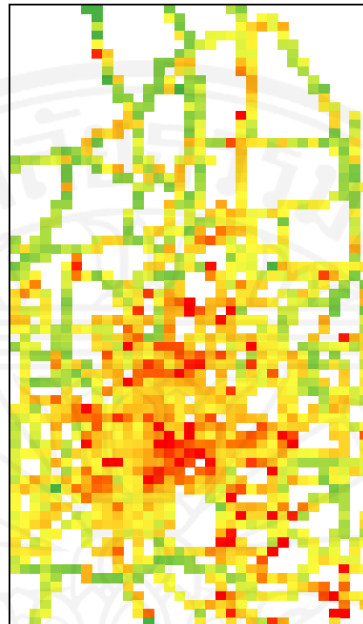


Figure 3.7 Heat map representing the spatial traffic intensity of the urban network

### 3.5.2.2 Spatio-temporal pattern based on the distance from the Central Business District of the city

Many study such as of Maat et al. (2005), Ahas et al. (2010) and Liu et al. (2012) examined and discovered the interdependence between land use and travel behavior. This can be confirmed that land use type is one important factor affecting the traffic state which in result different traffic patterns. This study intended to further investigate this behavior with our data and also apply for our study area (Bangkok).

A Central Business District (CBD) is an important land mark for all urban area. It is generally a concentrate location of retail and office buildings in the center of the city where most social, business and economic activity take place. The CBD usually has an urban density higher than the other area of the city and thus lead to higher density of traffic flow. The purpose of this analysis is to reveal the difference of the traffic intensity pattern by location of the area in terms of the distance from central business

district (CBD) of Bangkok. Silom area, cross mark in Figure 3.8, is considered as the CBD because of its infrastructures such as office building, transport facilities and mass transit; BTS Skytrain and MRT subway which are the two main mass transits in Bangkok. This area can be further confirmed by the study of Tipakornkiat et al. (2012), in which the author mentioned that Silom area is the CBD of Bangkok. Different zone will be divided according to the distance from this location. Then the traffic patterns of these zone will be presented and compared to each other.
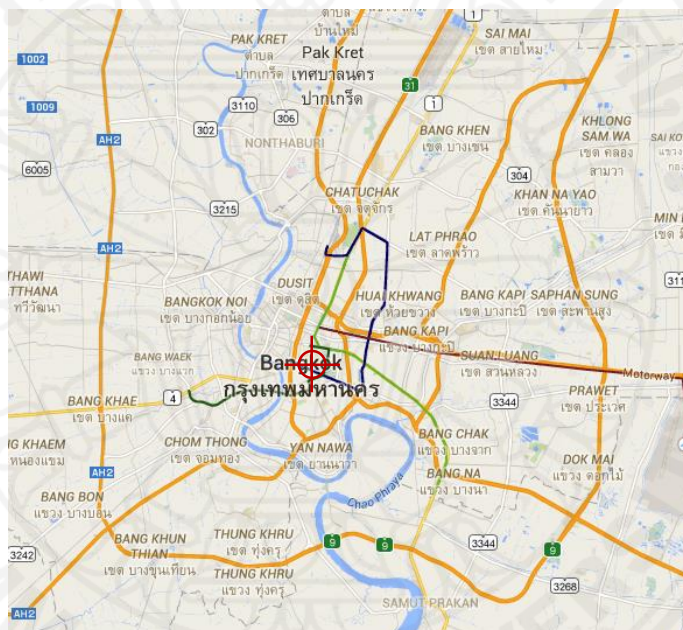


Figure 3.8 Silom area (cross mark), CBD of Bangkok

Image source: Google Maps

## 3.6 Clustering Analysis

In this analysis, we try to investigate the structure of data and obtain traffic state pattern by using clustering algorithms. First of all, time series of traffic state of the whole study area in each day are clustered into different groups by their similarity. The cluster result can be explained by using calendar based visualization technique adapted from Van Wijk and Van Selow (1999). Next, we apply another clustering

algorithm to group patterns based on the area of one square kilometer. In summary, there are two levels of clustering analysis: time-based clustering and area-based clustering. The details of each analysis are presented in the following subsections.

### 3.6.1 Time-based clustering

For time-based clustering, we perform clustering analysis on the data that is sorted into time series data with daily time scale. Because our data has 10 minute-time step, each time series data contains 144 elements where each element is the traffic intensity of the whole study area. The process of analysis is as follows:

- Time series data are organized into daily time scale; in total, there are 365 time series corresponding to 365 days.

- Each time series $Y_i$, i = 1,…, 365 consists of a sequence of pairs ($y_j$, $t_j$), j = 1,…, 144, where $y_j$ is the average traffic intensity at time $t_j$:

$$y_j = \frac{\sum_{n=1}^{N} I_{n,j}}{N}$$

(3.3)

Where $I_n$, j is the traffic intensity of point n at time j, and N is the total number of points excluding the points whose traffic intensity value is zero.

- The whole time series is tailored down to cover only from 7:00 AM until 11:59 PM due to the unavailability of data during the night time. (from 144 to 102 time steps)

- Agglomerative hierarchical clustering (Kaufman and Rousseeuw, 2009) is implemented in our data in order to obtain the clustering tree.

- Since our time series data is uniformly sampled and has the same length, we use Euclidean distance ($d_E$) to calculate the similarity in time between time series data (Warren Liao, 2005). The Euclidean distance between two time series U and V is computed as

41

$$d_E = \sqrt{\sum_{k=1}^{102}(u_k - v_k)^2} \qquad (3.4)$$

Where U = ($u_1$, $u_2$,…, $u_{102}$) and V = ($v_1$, $v_2$,…, $v_{102}$) are two time series with 102 time steps.

- The result is then interpreted by using calendar based visualization (Van Wijk and Van Selow, 1999).

### 3.6.2  Area-based clustering

Different from time-based clustering, we arrange the data into different location based on our grid system introduced before. This analysis is performed after getting results from the time-based clustering. Main steps in area-based analysis are:

- The whole study area is discretized into small grid cells with the size of 1 km x 1 km
- From results of clustering in the previous step, each cluster is combined to get the representative traffic state pattern for each cell.
- The agglomerative hierarchical clustering algorithm is implemented on the time series data of each cell in order to group them according to their similarity.
- Consistently, the similarity between time series is calculated by Euclidean distance as shown in equation (2).
- Incremental number of cluster is generated to identify the change of member of cluster.
- Finally, the clustering result is plotted in color-coded maps to visualize and get insight into the spatial distribution of clusters.

## 3.7 Anomaly Detection

Apart from the analysis of traffic state information to get various pattern, an attempt to cover also the traffic anomaly detection from our data is also proposed. The whole period of data is divided by 2 classifications; day of the week and working day or non-working day as shown in Figure 3.9.
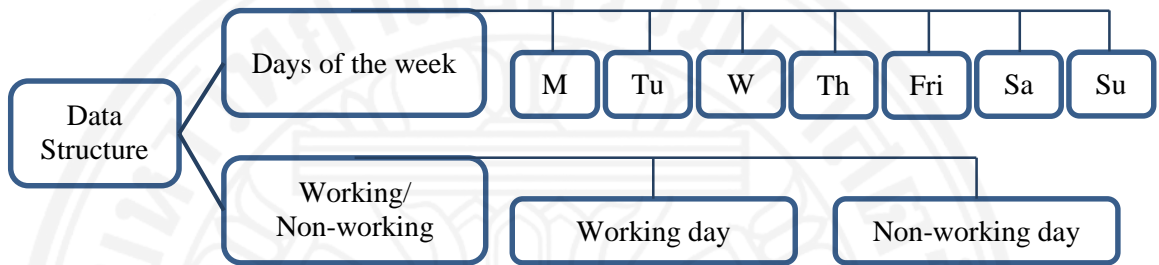


Figure 3.9 Data structure classification

Both time and location of anomalies can be detected by comparing the actual traffic intensity of any specific day that we want to study to the expected traffic intensity of the same classification calculated from historical data. The expected traffic intensity ($I_e$) is obtained from the average of traffic intensity from historical data. Standard deviation is also calculated in order to create thresholds of normal traffic intensity.

$$I_{e,ijdt} = \frac{\sum_n I_{n,ij}}{N} \tag{3.5}$$

$$\sigma_{e,ijdt} = \sqrt{\frac{1}{N-1}\sum_n^N (I_{n,ij} - \bar{I}_{e,ijdt})^2} \tag{3.6}$$

Where $I_{e,ijdt}$: expected traffic intensity of cell Cij

$I_{n,ij}$: intensity of traffic in cell Cij from data in the same class

$N$: total number of day from data in the same class

$\sigma_{e,ijdt}$: standard deviation of the expected traffic intensity

43

The anomaly happens when the difference between actual and expected traffic intensity (D) is higher than sensitivity factor (α) multiply by standard deviation value (σ) or lower than sensitivity factor (α) multiply by minus standard deviation value (-σ). It is noteworthy that in this anomaly detection application, we leave the freedom of threshold of detection (α) to users.

$$D = I_{ijdt} - I_{e,ijdt} \tag{3.7}$$

And the anomaly detected when: $D > \alpha.\sigma$ or $D < - \alpha.\sigma$ (3.8)

## 3.8 Traffic State Prediction

As discussed in literature review section, traffic state prediction is an ongoing trend research at the moment. It provides a lot of benefits to road users, road authority and also policy makers. However, most of the studies focus on using conventional traffic data such as traffic flow, speed, density, etc. to predict the future traffic state. This creates an idea to make use of available traffic state data to predict the future value.

For this study, the Bayesian prediction method is selected to implement on our data. Bayesian prediction is a probability model that based on Bayesian probability theory. It follows Bayes' rule where we want to evaluate the relative truth of a hypothesis H given the data D by the following formula:

$$P(H \mid D) = \frac{P(D \mid H) \cdot P(H)}{P(D)} \tag{3.8}$$

$P(H \mid D)$ is the probability of the hypothesis after consideration of given data and is called posterior probability. The term $P(D \mid H)$ is known as likelihood function which is the probability of the observed data D given the hypothesis H. $P(H)$ is called prior probability. It is the probability of the hypothesis H before the data D is considered and generally taken from historical knowledge. For $P(D)$, it is the normalizing factor that doesn't involve any hypothesis H and usually ignorable in determining the relative probabilities of different hypotheses.

In most of the previous research in the prediction field, they mostly focused on the short-term prediction and keep updating the forecast value when the new data is available. For this study, we would like to predict the traffic state value for the rest of the day given an observed value of any time during that day. This mean that by having an observed value of traffic state combining the historical data, we can make a prediction for each time step during that day. By this definition, equation 3.7 can be written as follows,

$$P(I_t \mid I_{t_0}) = \frac{P(I_{t_0} \mid I_t) \cdot P(I_t)}{P(I_{t_0})} \tag{3.9}$$

Where $I_t$ is the traffic state intensity at time $t$, $t = t_0+1, t_0+2, \ldots, 144$

$I_{t0}$ is the observed traffic state intensity at time $t_0$

In this sense, we want to predict the value of traffic state intensity from an observation with noise variation. Therefore, our prediction model is defined by

$$I_t = I_{t_0} + n \tag{3.10}$$

Where $n$ is additive Gaussian noise

The prediction is the process of making the best guess of value $I_t$ given observed value $I_{t0}$. If the probability distribution of $I_t$ given $I_{t0}$, $P(I_t \mid I_{t0})$, is available, the predicted value is the value that maximizes the distribution.

$$I_t = \arg \max_{I_t} P(I_t \mid I_{t_0}) \tag{3.11}$$

We have,     $$P(I_{t_0} \mid I_t) = P(n + I_t \mid I_t)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(I_{t_0} - I_t)^2}{2\sigma_n^2}}$$

And
$$P(I_t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(I_t - \mu)^2}{2\sigma^2}}$$

Where $\sigma_n^2$ is the variance of the noise

$\sigma^2$ is the variance of predicted value from historical data

$\mu$ is the mean of predicted value from historical data

The posterior from Equation 3.9 can be derived as follows,

$$P(I_t \mid I_{t_0}) \propto P(I_{t_0} \mid I_t) \cdot P(I_t)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(I_{t_0} - I_t)^2}{2\sigma_n^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(I_t - \mu)^2}{2\sigma^2}}$$

$$= \frac{1}{2\pi\sigma_n\sigma} e^{-\frac{(I_{t_0} - I_t)^2}{2\sigma_n^2} - \frac{(I_t - \mu)^2}{2\sigma^2}}$$

$$= \frac{1}{2\pi\sigma_n\sigma} e^{-\frac{1}{2}\left[\frac{(I_{t_0} - I_t)^2}{\sigma_n^2} - \frac{(I_t - \mu)^2}{\sigma^2}\right]}$$

In order to maximize $P(I_t \mid I_{t0})$, we have to determine value of $I_t$ which minimizes the exponent in the brackets. That value $I_t$ is our predicted value at time $t$. With this algorithm, we can predict the traffic intensity for the whole day with only one given observed data. Finally, we get a set of predicted values $[I_1, I_2, ..., I_{144-t0}]$.

Moreover, in case that the data of the next time step is available, we can use it to update our prediction to get a better accurate result. From the previous result, we have a set of predicted values $[I_1, I_2, ..., I_{144-t0}]$. Whenever another observed value at the next time step $I_{t0}'$ is obtained, we can use the previous predicted values as our input in the equation instead of using the historical data alone. Same as previous formula (3.11):

$$I_t ' = \arg \max_{I_t '} P(I_t ' \mid I_{t_0} ')$$

$$P(I_t ' \mid I_{t_0} ') \propto P(I_{t_0} ' \mid I_t ') \cdot P(I_t ')$$

From this process, we will get another set of result $[I_2 ', I_3 ', \ldots, I_{144\text{-}t0} ']$. There is no first time step since we obtain that value from observation for the second time step already. This process can be repeated since the observed values keep coming in.

Lastly, we employ several statistical methods to measure the prediction accuracy. According to Hyndman and Koehler (2006), statistical method to determine the accuracy of prediction can be separated into two different types: scale-dependent and not scale-dependent. The most commonly used methods that are scale-dependent are Mean Square Error (MSE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). For not scale-dependent mode, the most well-known method is Mean Absolute Percentage Error (MAPE). The prediction error (E) is the difference between the predicted values (P) and the observed values (O). And the percentage error (PE) is the proportion of error (E) to the observed value (O).

$$E = P - O \tag{3.12}$$

$$PE = (E / O) \times 100 \tag{3.13}$$

With different formulation, all of them give different meaning to the accuracy and they are listed in the table below:

Table 3.2 Statistical methods to measure accuracy of prediction

| Model | Method | Formula |
|---|---|---|
| Scale-dependent | Mean Square Error (MSE) | $MSE = \sum E^2 / n$ |
| | Mean Absolute Error (MAE) | $MAE = \sum |E| / n$ |
| | Root Mean Square Error (RMSE) | $RMSE = \sqrt{\sum E^2 / n}$ |
| Not scale-dependent | Mean Absolute Percentage Error (MAPE) | $MAPE = \sum |PE| / n$ |

## 3.9  Summary

In conclusion, our methodology can be separated into two main phases. Phase one is the process to collect data which is the traffic state information from web-based mapping services. Phase two concerns about the method to analyze and make use of our data such as the study of traffic state pattern, the usage of unsupervised clustering, traffic anomaly detection and finally the traffic state prediction. We sum up the whole methodology by using a flowing diagraph as our research framework as shown in Figure 3.10.
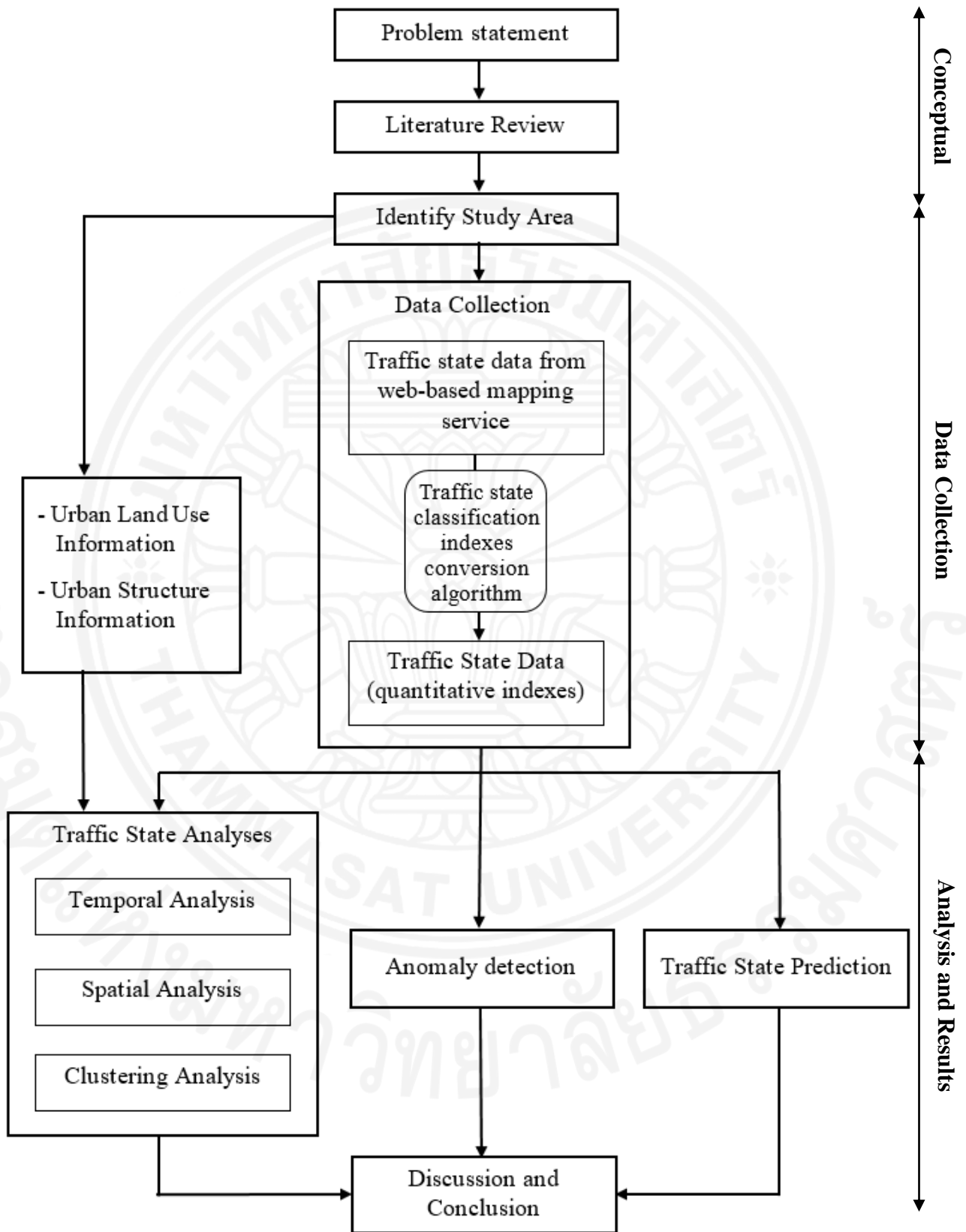
Figure 3.10 Research framework

49

# Chapter 4

# Results and Discussion

## 4.1 Introduction

By following the methodology proposed in this report, the traffic state data from web-based mapping service is being collected every day. For the purpose of this study, a total of one year data, from September 01, 2014 to August 31, 2015, has been collected.

## 4.2 Temporal Traffic State Patterns of Urban Network

To visualize the time variation pattern of the whole area, the traffic intensity of each cell was calculated, and then combined to find an average value which represent the overall intensity of the traffic by equation 4.1:

$$I' = \frac{\sum_{i}^{35}\sum_{j}^{70}\overline{I}_{ij}}{N}$$

(4.1)

Where $N$ the total number of cells with $\overline{I}_{ij} \neq 0$

### 4.2.1 Temporal patterns of urban network for whole duration

The result of traffic state pattern with 10 minutes-time step of the whole area for one year is plotted in Figure 4.1. As can be seen in the graph, there are some blank parts to the curve signified missing data due to technical problem from data collection instruments and internet connection. 3 portions of missing data are: 14th to 17th November, 2014; 21st to 29th April, 2015 and 25th to 28th July, 2015.

Apart from the missing parts, there are some interesting patterns. For the whole duration, we can see the traffic intensity fluctuates up and down around its mean value (1.3 to 1.4). This is in accordance with general concept of traffic which has a

50

dynamic behavior. Additionally, we also notice two strange patterns in which the intensity is lower than usual. In order to simplify and see the trend of the pattern, the data of 10 minute-time step was aggregated to 1 full day and plotted in the same time scale in Figure 4.2. In this graph, the abnormal patterns can be seen clearly. They are during the end of December, 2014 to the early January, 2015 and during mid-April. With further investigation, these two periods of time are related to long holiday in Thailand. The first period (end of the year) is the New Year holiday and the second period is the Songkran festival in Thailand.

In order to investigate more about the daily traffic state pattern, results in a finer scale are needed. Thus in the following sections, we will extract some part of the data to do further analysis. Traffic state patterns of different day will be observed and detail discussion will be made.
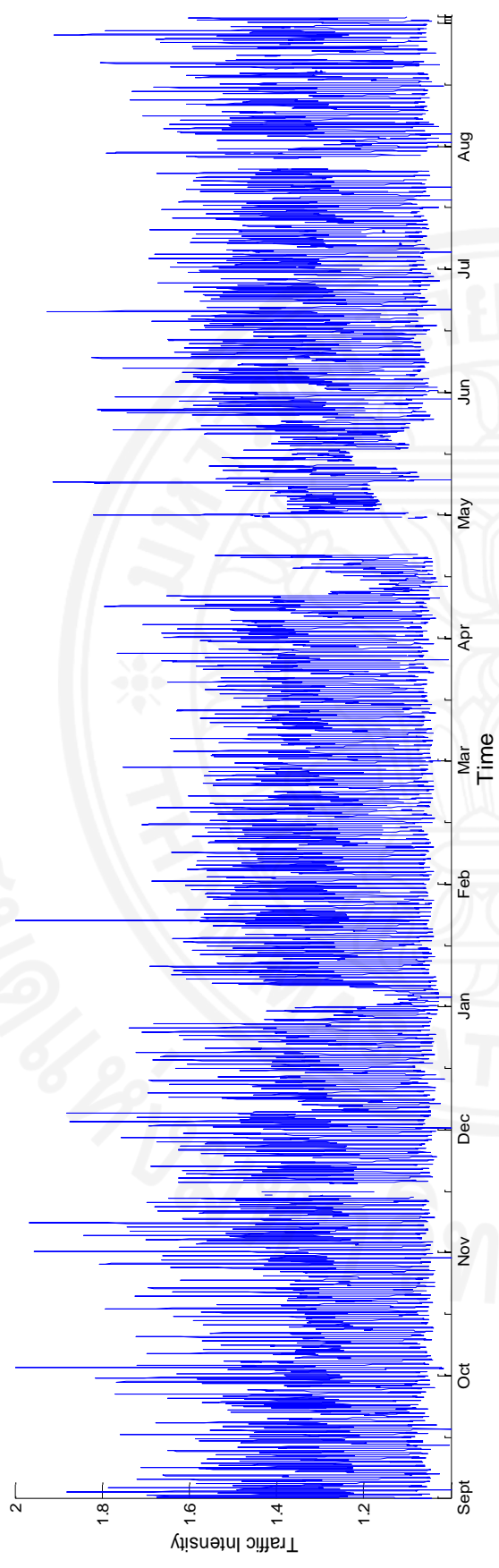
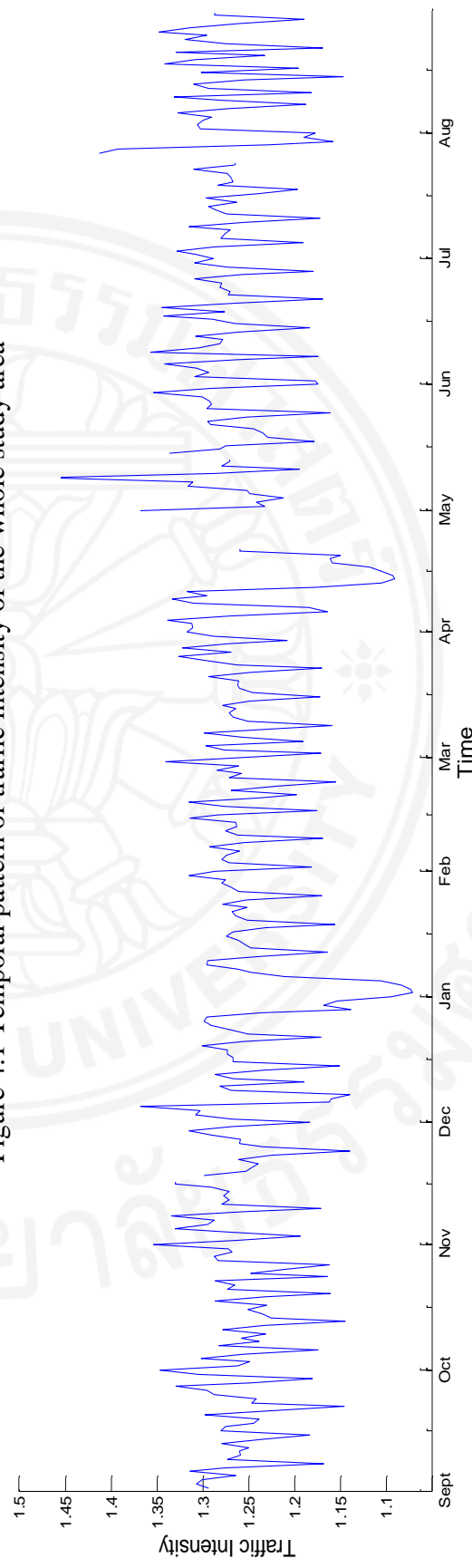Figure 4.1 Temporal pattern of traffic intensity of the whole study area



Figure 4.2 Temporal pattern of traffic intensity of the whole study area with one day aggregation

52

### 4.2.2    Temporal patterns of urban network for one month duration

In this section, we will study on different period of data. September, 2014 is chosen to investigate the normal traffic state pattern because during this month, there is no public holiday. After that, according to previous result, we will look into the period of data that contain the abnormal patterns.

#### 4.2.2.1 Temporal patterns of September

By plotting only the data of one month, the detail daily patterns can be seen in Figure 4.3. A total number of 30 temporal sequences of traffic state in the study area with a one full day cycle can be identified. This can represent that the traffic repeats itself within one full day. In most of the repeating sequences, there are some interesting characteristics (2 peaks in one day), while a few others come out to be different (only one peak in one day). This behavior will be discussed in the next section.
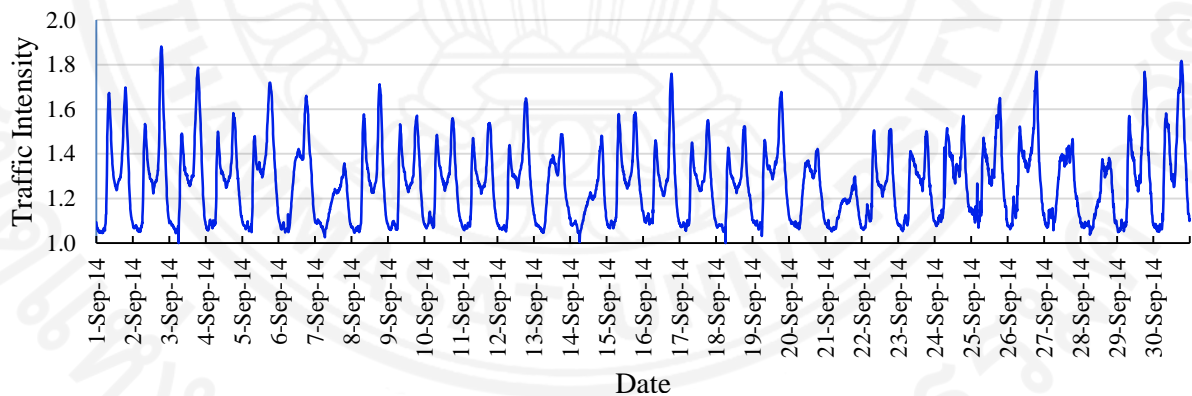


Figure 4.3 Temporal pattern of traffic intensity for September

#### 4.2.2.2 Temporal patterns during New Year holiday

For the New Year holiday, the end of year 2014 and the beginning of year 2015, the official holiday in Thailand is 31st December (New Year 's Eve) and 1st January (New Year day). By further observing the calendar, we found out that it is in

53

the middle of the week (Wednesday and Thursday). Therefore, most people took a long whole week holiday starting from Saturday, 27th December 2014 to Sunday, 4th January 2015. As a result, the traffic intensity of the whole city changed abruptly during that week from the normal week that can be seen in Figure 4.4.
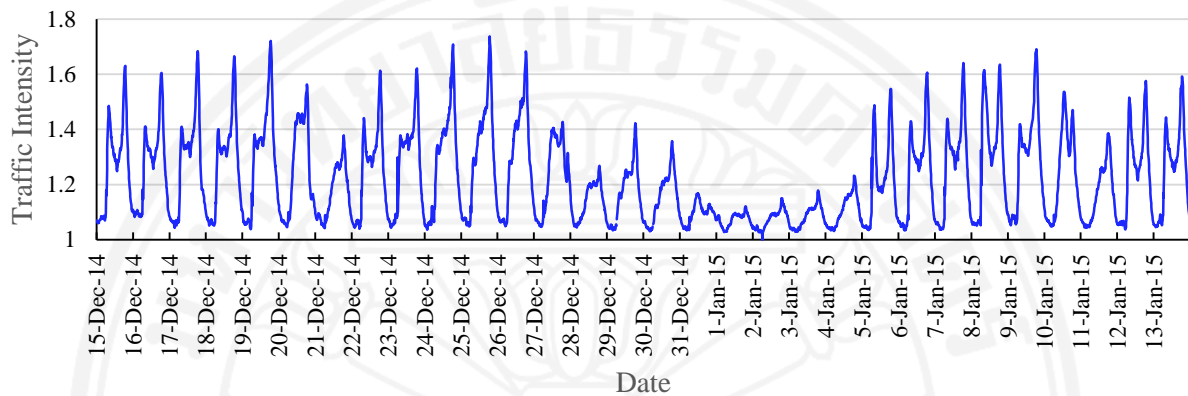


Figure 4.4 Temporal pattern of traffic intensity during New Year holiday

### 4.2.2.3 Temporal patterns during Songkran festival holiday

Songkran is a festival celebrating the traditional New Year in Thailand and some other country such as Cambodia, Myanmar and some part of India. There are some slightly differences of the exact date between different countries, but for Thailand, the official holiday for this festival is on 13th, 14th and 15th April every year.

The traffic state pattern during April, where Songkran festival holiday takes place, is plotted in Figure 4.5. It is also worth to mention again that during this month, we have a portion of data missing starting from 21st to 29th April. Thus the graph shows only 20 days data from 1st to 20th April. Similar to previous result, very low traffic intensity for the whole week can be observed. Starting from 11th and 12th April, which are the weekends from previous week, until 19th April. Although the holiday happened on Monday, Tuesday and Wednesday, but people generally tend to extend their holiday by skipping work for both the following Thursday and Friday. For this festival, most Thai people go back to their hometown which makes the traffic in the city quite low.

54

From the same graph, we also notice a strange pattern on 6$^{th}$ April. It is Monday, but the traffic intensity is quite low comparing to other weekday. By further inspection, we found out that it is Chakri Day which is an official holiday in Thailand. It is the date to commemorate the establishment of the Chakri Dynasty and to recognize the contributions of all the kings in the dynasty. By this reason, this specific Monday has a traffic pattern similar to the one of Sunday.
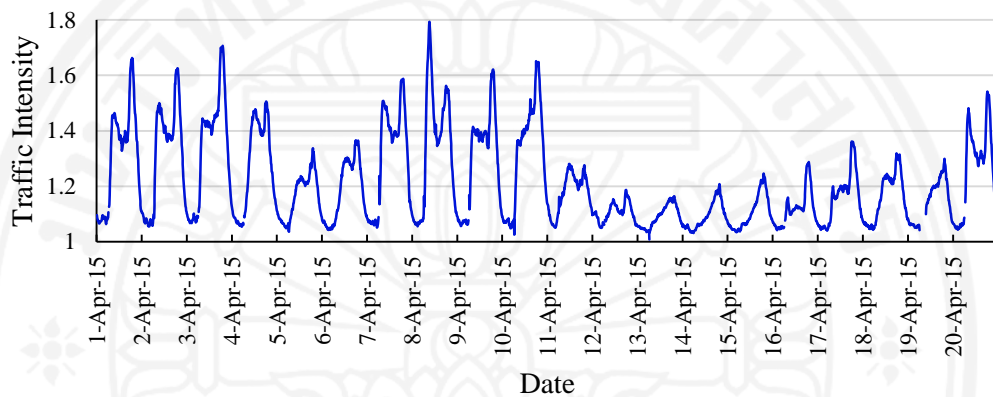


Figure 4.5 Temporal pattern of traffic intensity during Songkran festival

With this observation, we will also investigate different pattern according to different feature of the day by using machine learning in the later section.

## 4.3  Traffic Patterns for Different Days

From the result of temporal traffic state patterns that revealed different patterns for different days, detail investigation is made in this section. For the purpose of this study, we extract one month of data during September 2014 same as previous section. By differentiating between weekday and weekend in this month, a significant difference between traffic patterns can be found. During weekday as shown in Figure 4.6, the temporal sequence shares similar characteristics; two main peak corresponding to the morning peak which starts around 7:30 am and ends around 09:30 am and the evening peak from 5:00 pm to 8:00 pm. This pattern can be associated to the trip purpose of the weekday which is non-recreational trips (working trip, school trip, etc.). On the other hand, the diurnal traffic pattern of the weekend is different from the one

55

of the weekday. Figure 4.7 and 4.8 show the average traffic intensity on Saturday and Sunday along with its variations, respectively. There is only the evening peak which starts around 5:00 pm and ends around 8:00 pm. The morning peak disappears due to the non-working day and people tend to start their trips a bit late (around 10:00 am). However, there is also a distinctive difference between Saturday and Sunday. There is more traffic on Saturday, especially high evening peak. These trips can be associated to the trip with recreational purposes.
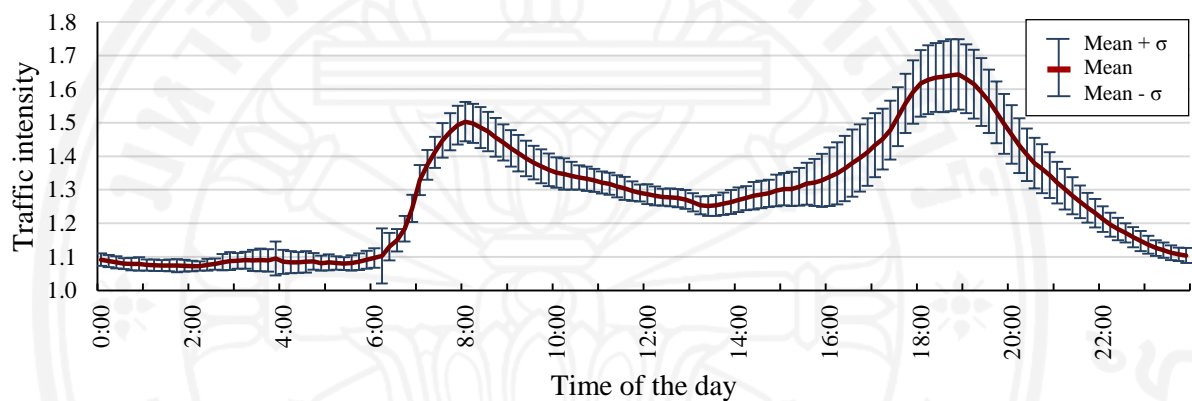


Figure 4.6 Temporal pattern of traffic intensity during weekday
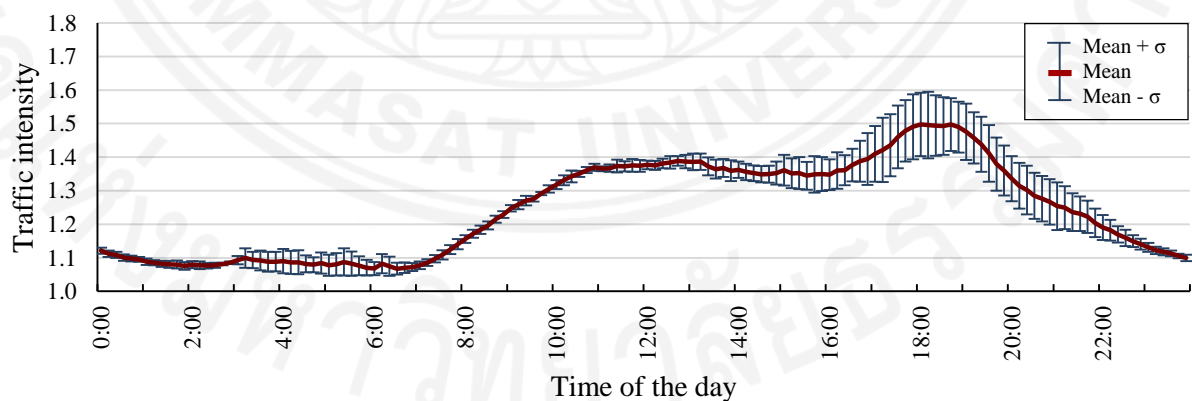


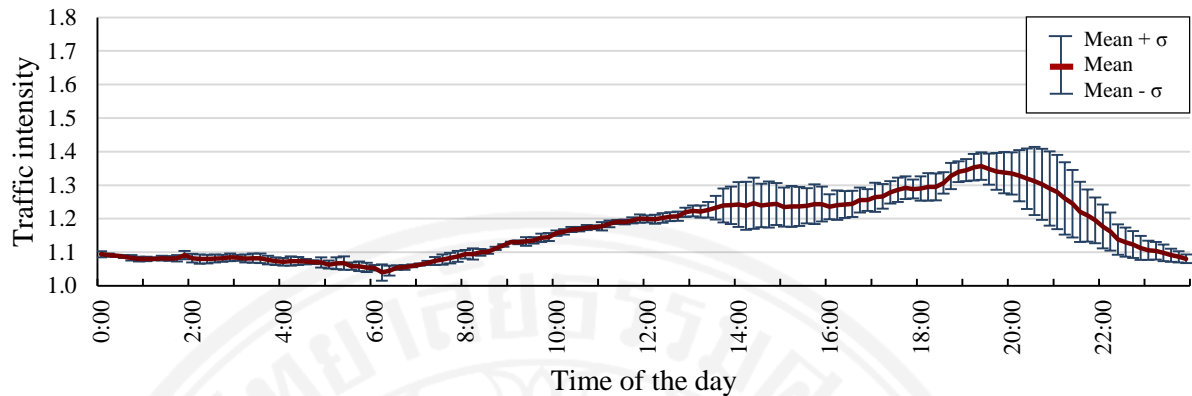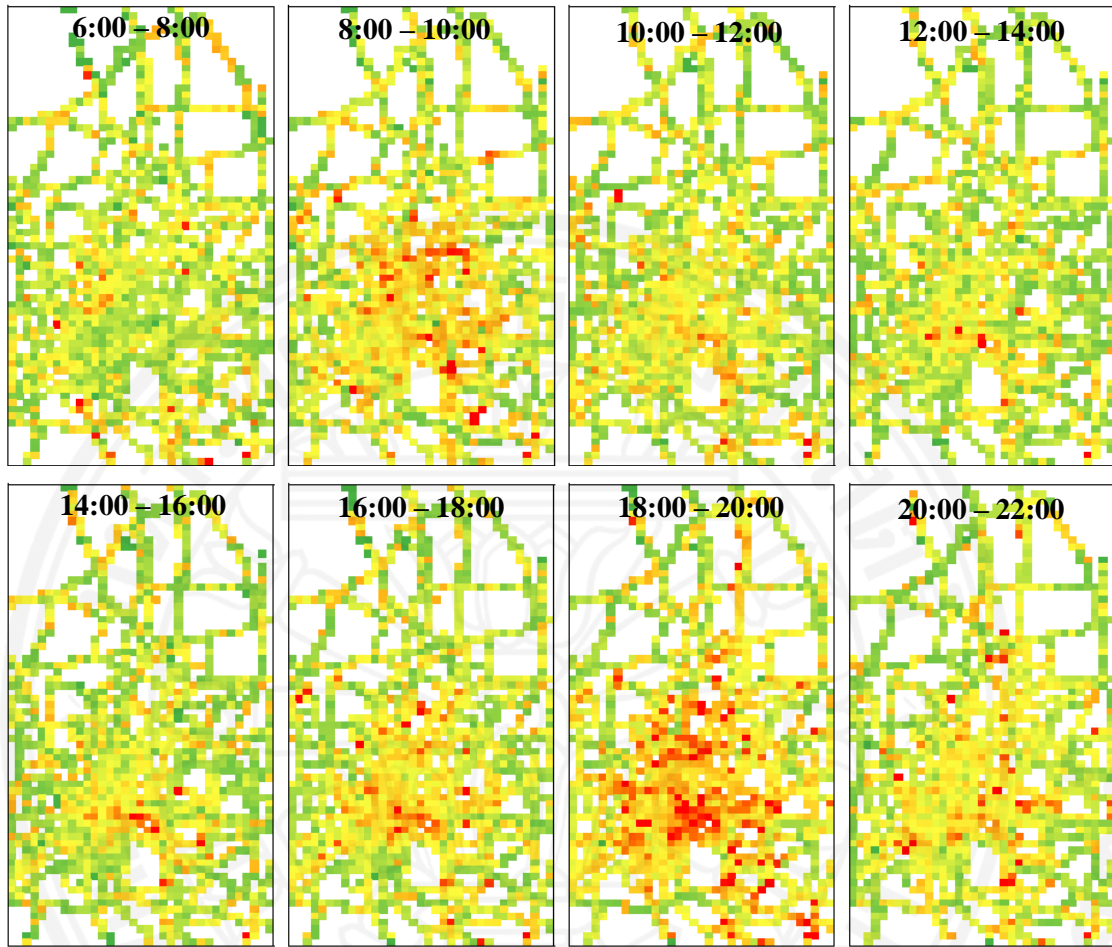Figure 4.7 Temporal pattern of traffic intensity during Saturday

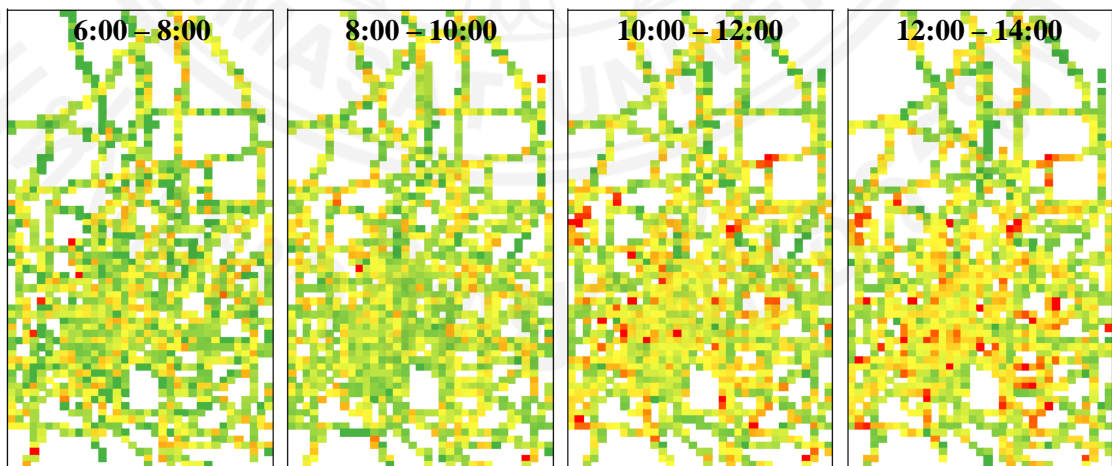Figure 4.8 Temporal pattern of traffic intensity during Sunday

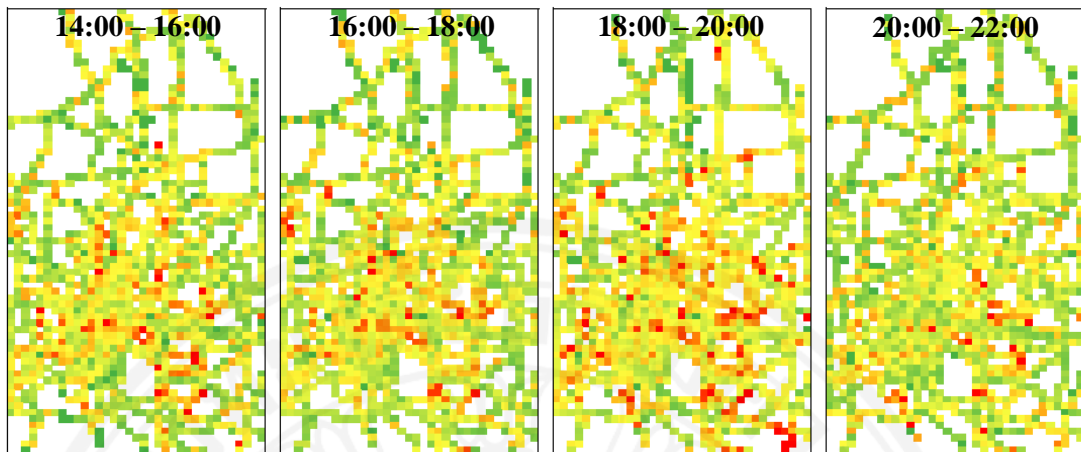## 4.4 Spatio-Temporal Patterns of Traffic Condition

### 4.4.1 Spatio-temporal traffic patterns for different days

With understanding of the difference traffic patterns between different days of the week, we would like to further investigate how the traffic evolve in terms of not only time but also location. As described in methodology section, the whole study area was discretized into 2450 small cells. Then the average traffic intensity was computed for each cell with the different days of the week. For the consistency, the data of one month, September 2014, was use in this analysis. As a result, the heat map of the whole area was plotted. Figure 4.9 (a), (b) and (c) show the spatio-temporal evolution of traffic intensity on weekdays, Saturday and Sunday, respectively. We aggregated the data of 2 hours for each heat map and the intensity of traffic for each cell is represented by the color from green color to red color depicted low to high traffic intensity. In addition to the diurnal temporal pattern of traffic, the spatial patterns are also clearly seen. This spatio-temporal patterns of traffic is consistent with the temporal patterns discussed above.

57

| 6:00 – 8:00 | 8:00 – 10:00 | 10:00 – 12:00 | 12:00 – 14:00 |

| 14:00 – 16:00 | 16:00 – 18:00 | 18:00 – 20:00 | 20:00 – 22:00 |

**(a)**

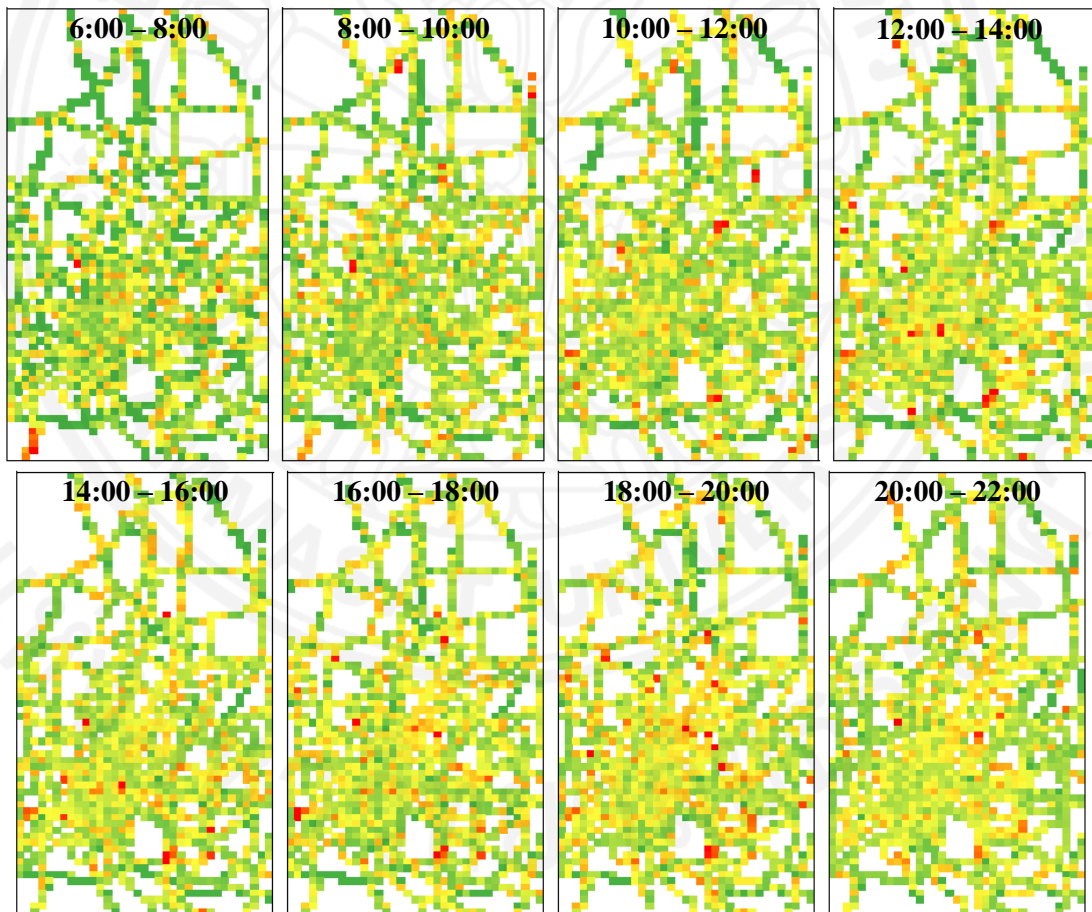| 6:00 – 8:00 | 8:00 – 10:00 | 10:00 – 12:00 | 12:00 – 14:00 |

**(b)**



**(c)**

Figure 4.9 Heat maps represent spatio-temporal patterns of traffic intensity from 6:00 AM to 10:00 PM during (a). Weekdays (b). Saturday (c). Sunday

From the heat maps, discussions for the result can be made as following. For weekday pattern, the result is consistent with the temporal pattern we investigated earlier. There are two high traffic intensities during morning between 8:00 to 10:00 and during evening period between 18:00 to 20:00. Moreover, we can see that the high traffic intensity is generally concentrated in the central area. With urban land use information, the concentrated area is actually the central business district of Bangkok, called Silom area.

Additionally, the result of Saturdays and Sundays reveals the patterns as expected from the temporal patterns. There is no high traffic intensity during the morning period. The overall intensity is also lower comparing to weekday. There is no concentrated behavior in the CBD area as well. The high intensity areas are sparse all over the map. These locations are considered to be related to the recreational activities.

### 4.4.2 Traffic state pattern related with the distance from Central Business District (CBD)

The purpose of this analysis is to reveal the difference of the traffic intensity by the location of area in terms of the distance from the central business district (CBD) of Bangkok. In Figure 4.10, four circles with the same center have been drawn on the map in order to classify the area into four different zones; zone 1, zone 2, zone 3 and zone 4 with the radius of 2 km, 6 km, 10 km and 14 km from the center point, respectively. For each zone, the average intensity of traffic is computed and varied by time of the day as shown in Figure 4.11. It can be seen that the pattern of each zone is different from one another. In the morning time, there is not much variation between each zone, but the difference is clear after the middle of the day (12:00 PM). This can be interpreted as the traffic in the CBD area is the most congested, and this intensity is decreasing as the distance from the center increasing. This traffic behavior is consistent with the study of Liu et al. (2012) which found out the concentric urban structure by investigating the trip pattern.
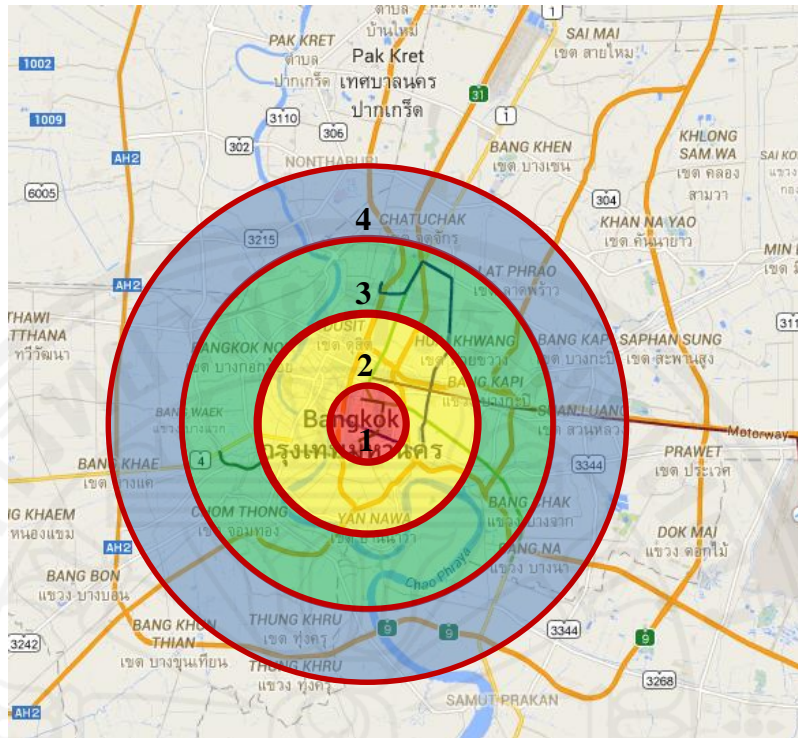
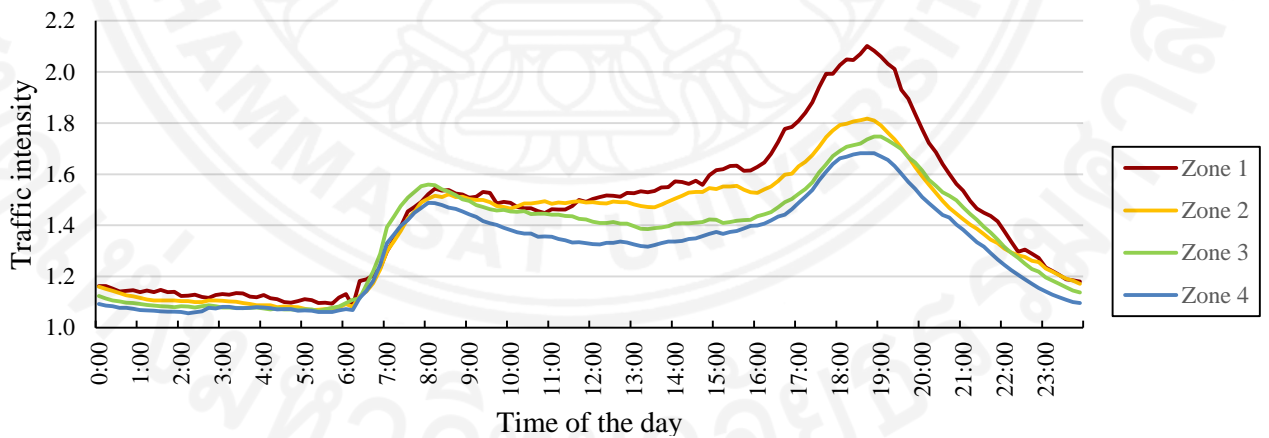Figure 4.10 Four zones defined by four circles with radius of 2 km, 6 km, 10 km and 14 km



Figure 4.11 Temporal pattern of average traffic intensity for each zone

## 4.5 Clustering Analysis

### 4.5.1 Time-based clustering

By using hierarchical clustering, we can create a hierarchy of clusters. The final number of cluster is based on the level of detail and clustering resolution the analyst want to get. Theoretically, the optimal number of cluster is the tradeoff between maximum compression of data and maximum accuracy of each element in the clusters. Therefore, there are some criterions to define this number. For this study, we use elbow method and plot the result in scree plot, firstly introduced by Thorndike (1953).

As mentioned before, our data has some missing parts, so these parts are excluded from our analysis. From the plot result in Figure 4.12, we can see that the optimal number of cluster is 2 clusters. By increasing the number of clusters, it doesn't give much better modeling of the data. Including the missing data, we have in total 3 different clusters. In Figure 4.13, calendar based visualization is used to show the member of each cluster; Cluster A and Cluster B, by different color along with the representative of traffic state pattern for respective cluster.
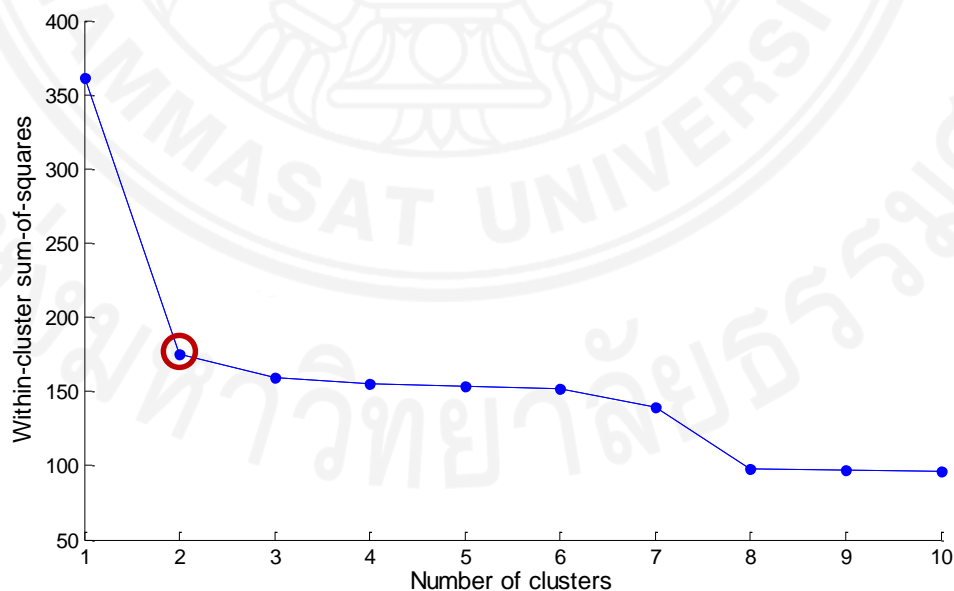


Figure 4.12 Scree plot and elbow point (red cycle) to indicate optimal number
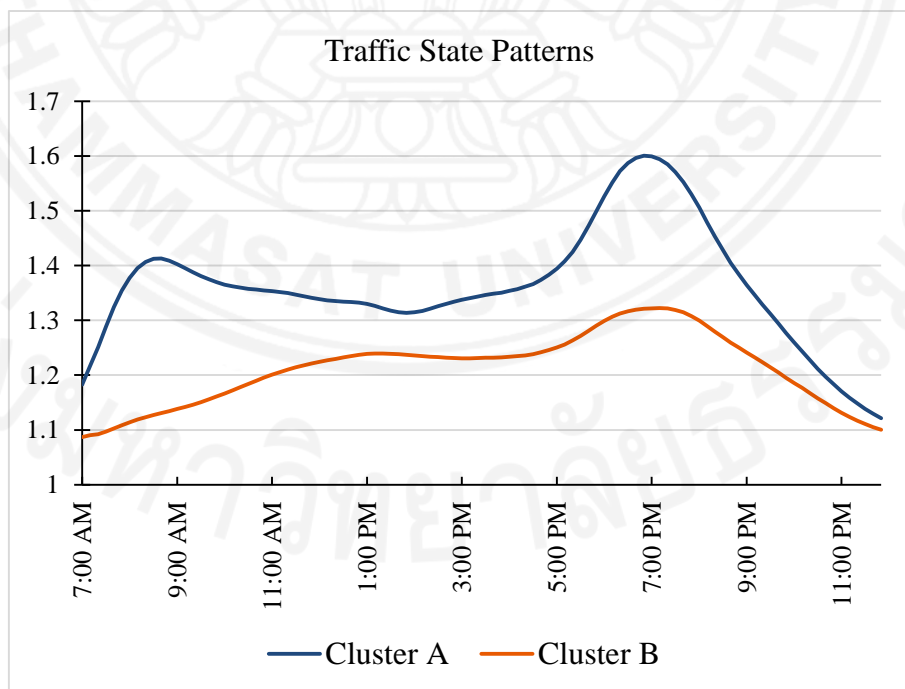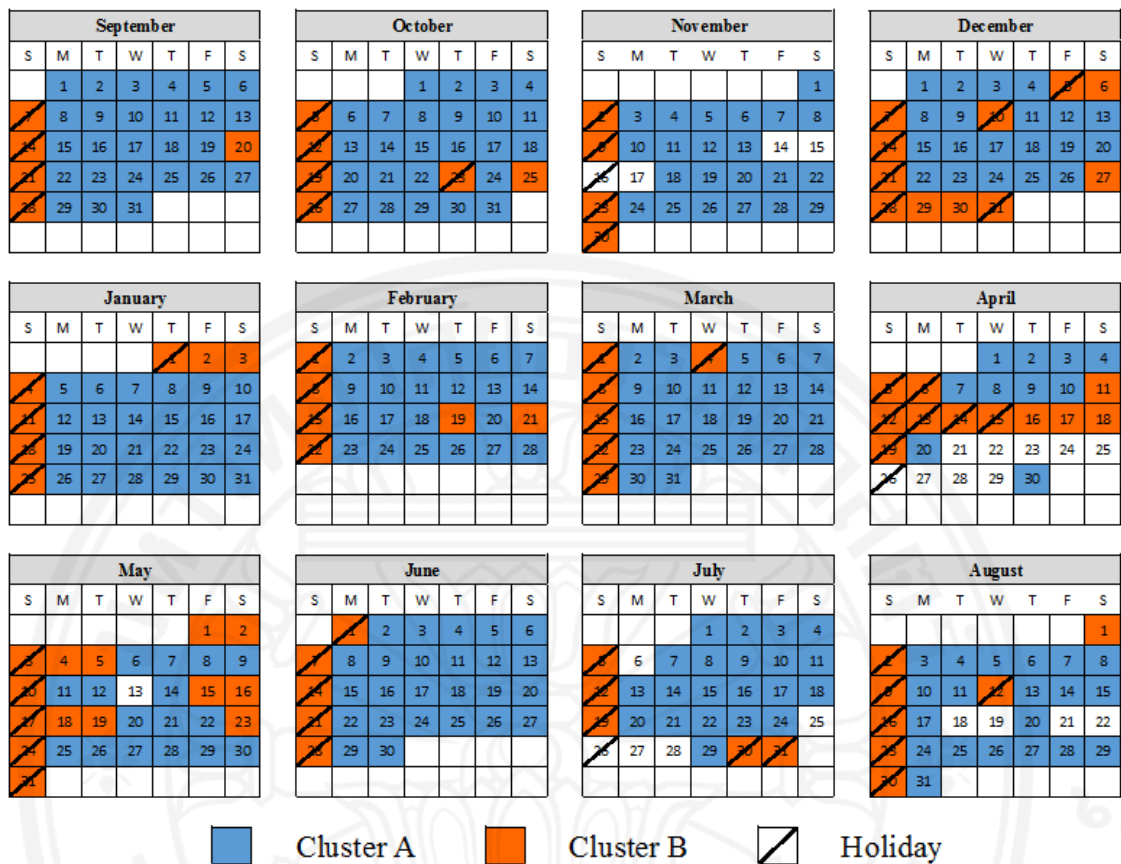of cluster

62

Figure 4.13 Calendar-based visualization and representative traffic state
patterns of each cluster

63

From the figure above, several results can be discussed as follows:

- Cluster A is the cluster that consists of the time series from most of the weekdays (working day). The traffic pattern for this group of data reveals two peaks of traffic intensity; morning peak, from 7:00 AM to 9:00 AM and evening peak, starting around 5:00 PM until 9:00 PM. This behavior can be explained by the traffic activity of the city during working days. Both morning and evening peak correspond to traffic activity for non-recreational purposes such as going to school and working place. However, the evening peak intensity is higher than morning peak due to extra traffic for the recreational purposes.

- Cluster B is the cluster that consists of time series from weekends and also some days from weekdays. Upon further investigation, we found out that those weekdays are actually official holidays in Thailand. For example, during the New Year's week and Songkran's week as we have discussed already in previous section. The traffic state pattern of this cluster is significantly different from cluster A because there is only one peak in the evening. The morning peak disappears due to the non-working day and people tend to start their trips late (around 10:00 AM). The high intensity traffic in the evening can be associated to the trips with recreational purposes.

- The members in Cluster B are generally matching to the holidays, only some exceptions such as extend of the holiday and special events.

- Days which have no highlight color on it are the missing data. There is no information on those days.

### 4.5.2 Area-based clustering

The color-coded maps from different number of cluster are shown along with respective traffic state patterns in Table 2. The data is initially grouped into two different clusters from the result of the city-level clustering namely: cluster A and cluster B.

64

Table 4.1 Area-based clustering for working day cluster (cluster A)

| Cluster A | Traffic State Patterns |
|---|---|
| **2 sub-clusters** |  |
| **3 sub-clusters** |  |
| **4 sub-clusters** |  |
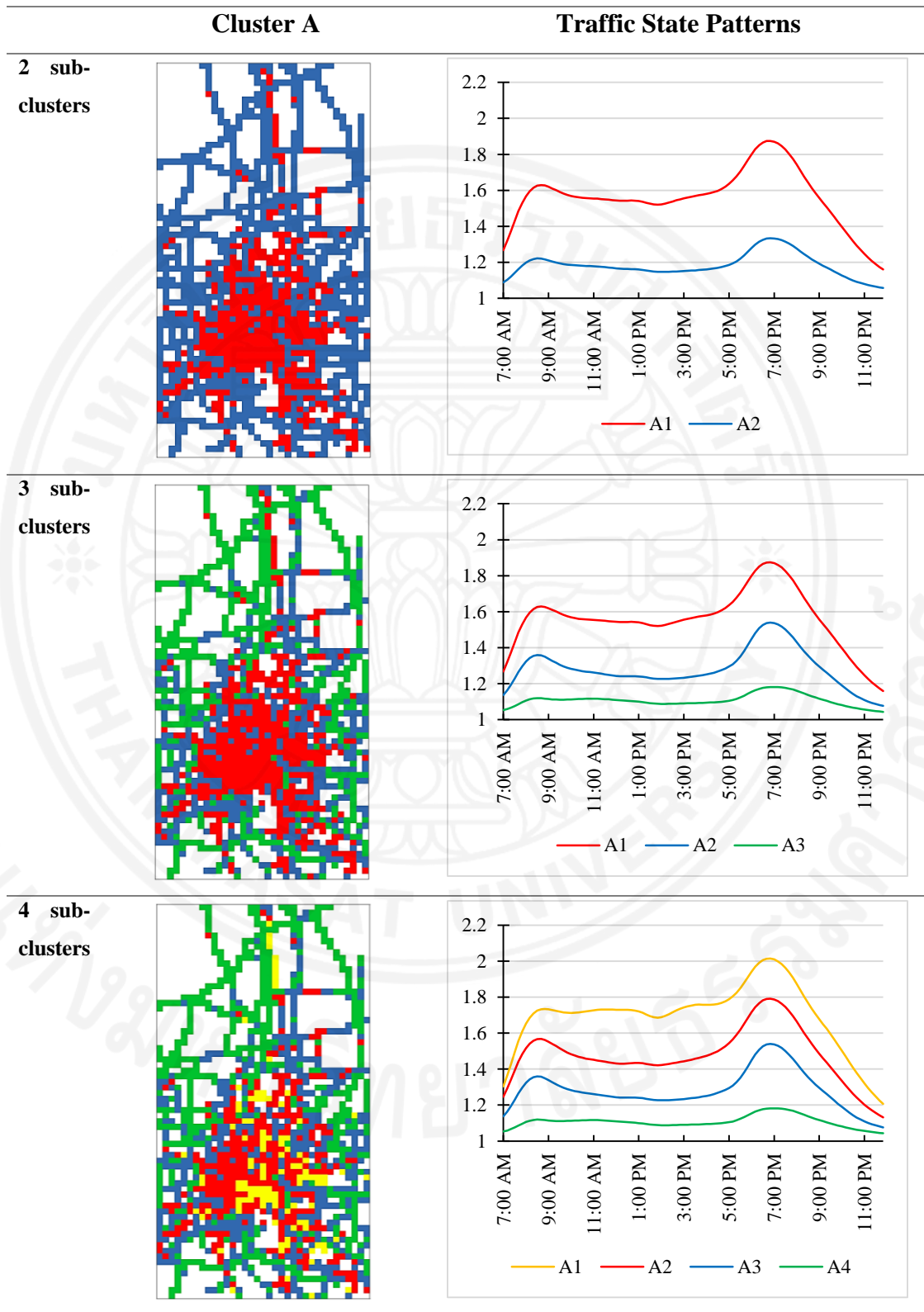
Table 4.2 Area-based clustering for non-working day cluster (cluster B)

| Cluster B | Traffic State Patterns |
|---|---|
| 2 sub-clusters |  |
| 3 sub-clusters |  |
| 4 sub-clusters |  |
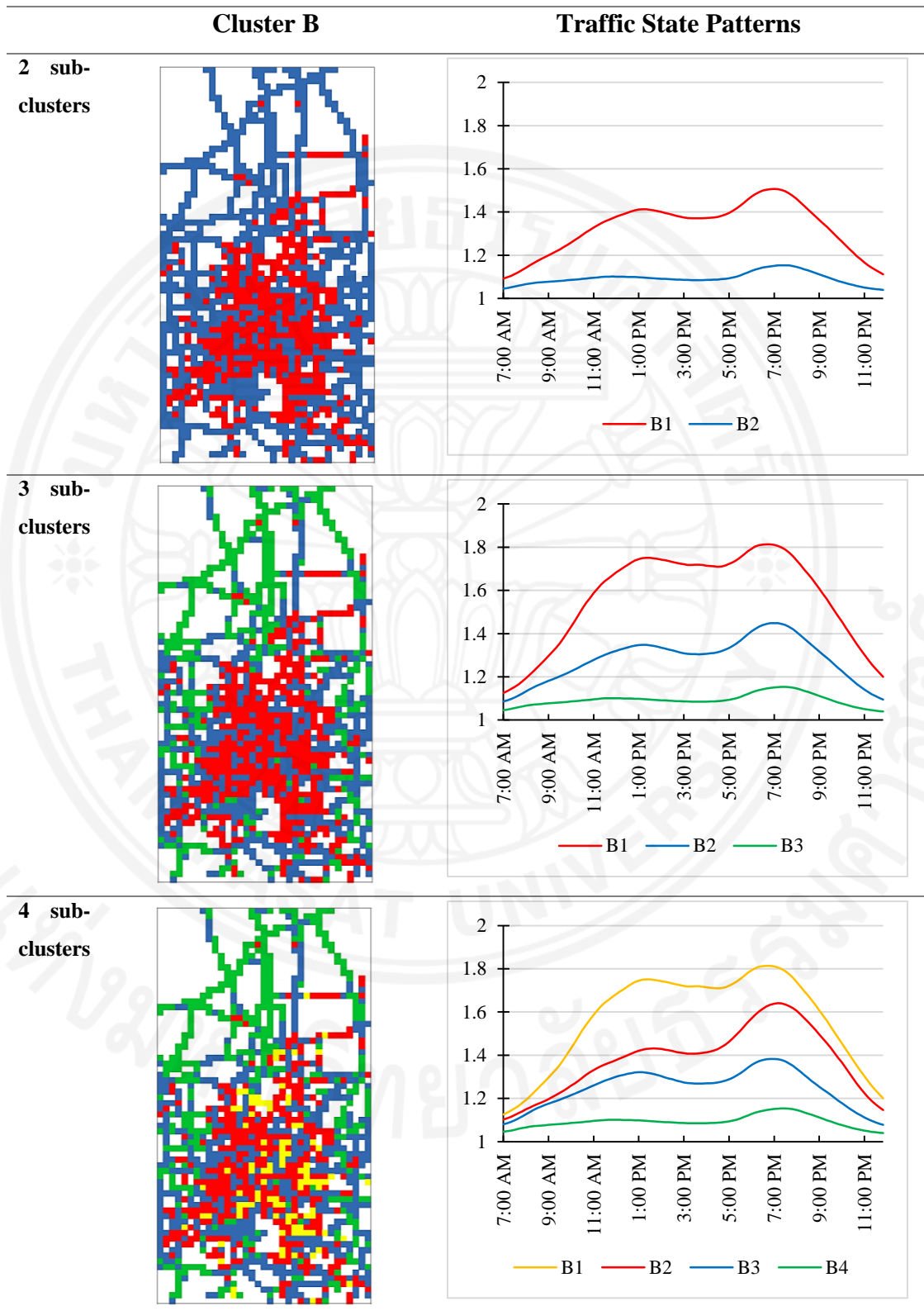
According to above results, various conclusions can be made as follows:

- Further clustering of traffic patterns at city level (both A and B) revealed the concentric behavior of traffic of the city. This is consistent with the finding of Li et al. (2007).
- For cluster A, the distribution of members of clusters (cells) is organized by the distance from the center of the city (Central Business District) due to the pattern of working days. In contrast, the spatial distribution from cluster B is more scattered all over the area due to the non-working day patterns.
- The intensity of traffic is higher when the area are closer to the center.

## 4.6 Traffic Anomaly Detection

The method proposed in the methodology for traffic anomaly detection was implemented on a case study of a random day to test the detection. Monday, 13th October 2014 was chosen. This day can be classified in our data structure as normal working day. Figure 4.14 shows the traffic intensity of that day and expected traffic intensity from historical data for an area (a single cell). The difference between actual and expected traffic intensity along with the standard deviation are presented in Figure 4.15.
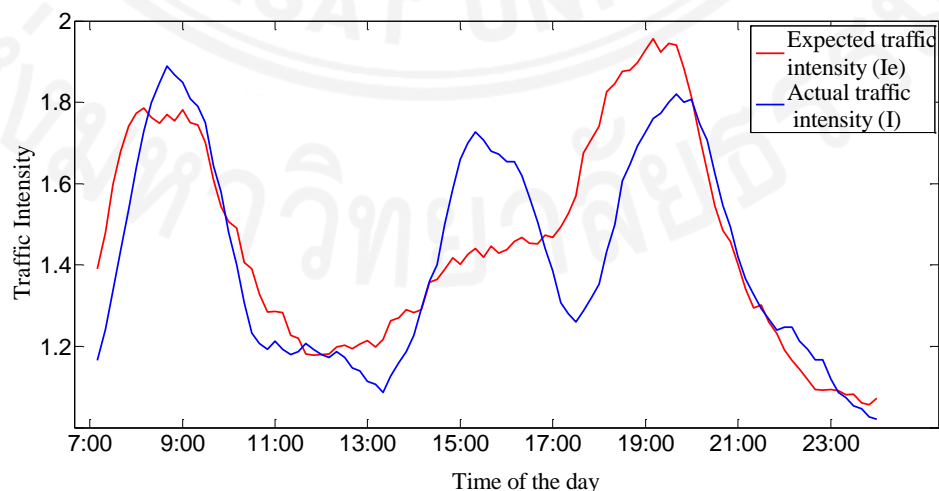


Figure 4.14 Actual and expected traffic intensity on 13th October 2014
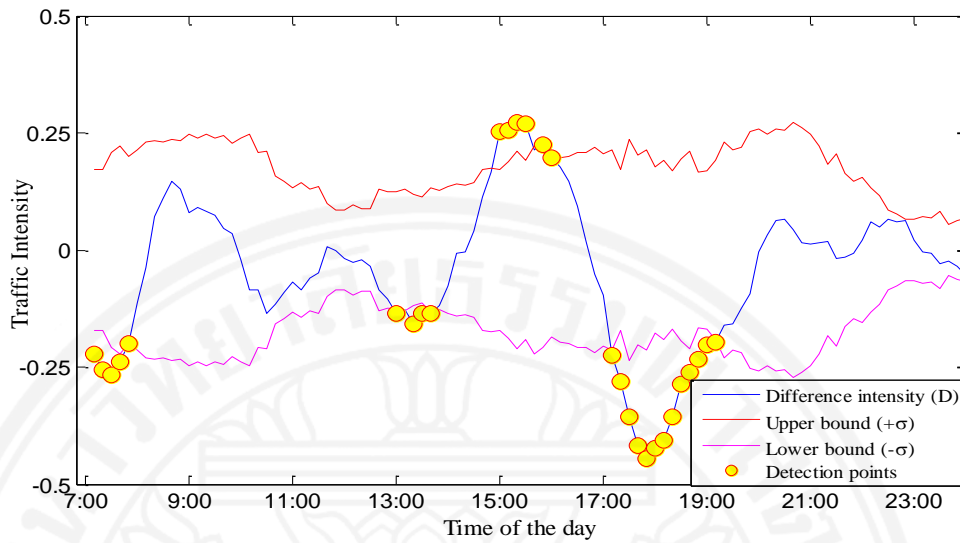
67

Figure 4.15 Difference between actual and expected intensity compare to standard deviation

The yellow dots in Figure 4.15 above represent the traffic anomalies that our algorithm detected. The sensitivity of detection is based on sensitivity factor (α).For the whole study area, the results of spatio-temporal detection are shown in form of map where the red dots representing the anomaly detected in Figure 4.16 and Figure 4.17 with the sensitivity factor of 1 and 2, respectively. The detection maps from the algorithm can detect traffic anomalies for every 10 minutes, but the results show only some periods of time from the whole period.

From the result in Figure 4.16, we can see that the detected areas are excessive with the sensitivity factor α=1. On the contrary, those with α=2 in Figure 4.17 seem to be more acceptable. In brief, the detection sensitivity in this application is based fully on the user to input the factor α.

From this detection method, some useful modifications can be done to increase the confidence in detecting anomaly. One among them is to detect only the anomalies that last equal to or more than 3 time-steps (30 minutes). By this method, we can eliminate the spontaneous traffic breakdown just for a few minutes from our anomaly. Same as the result in Figure 4.17, with sensitivity factor α=2, this method reduces the amount of detected cases. Maps of result are plotted in Figure 4.18.

68
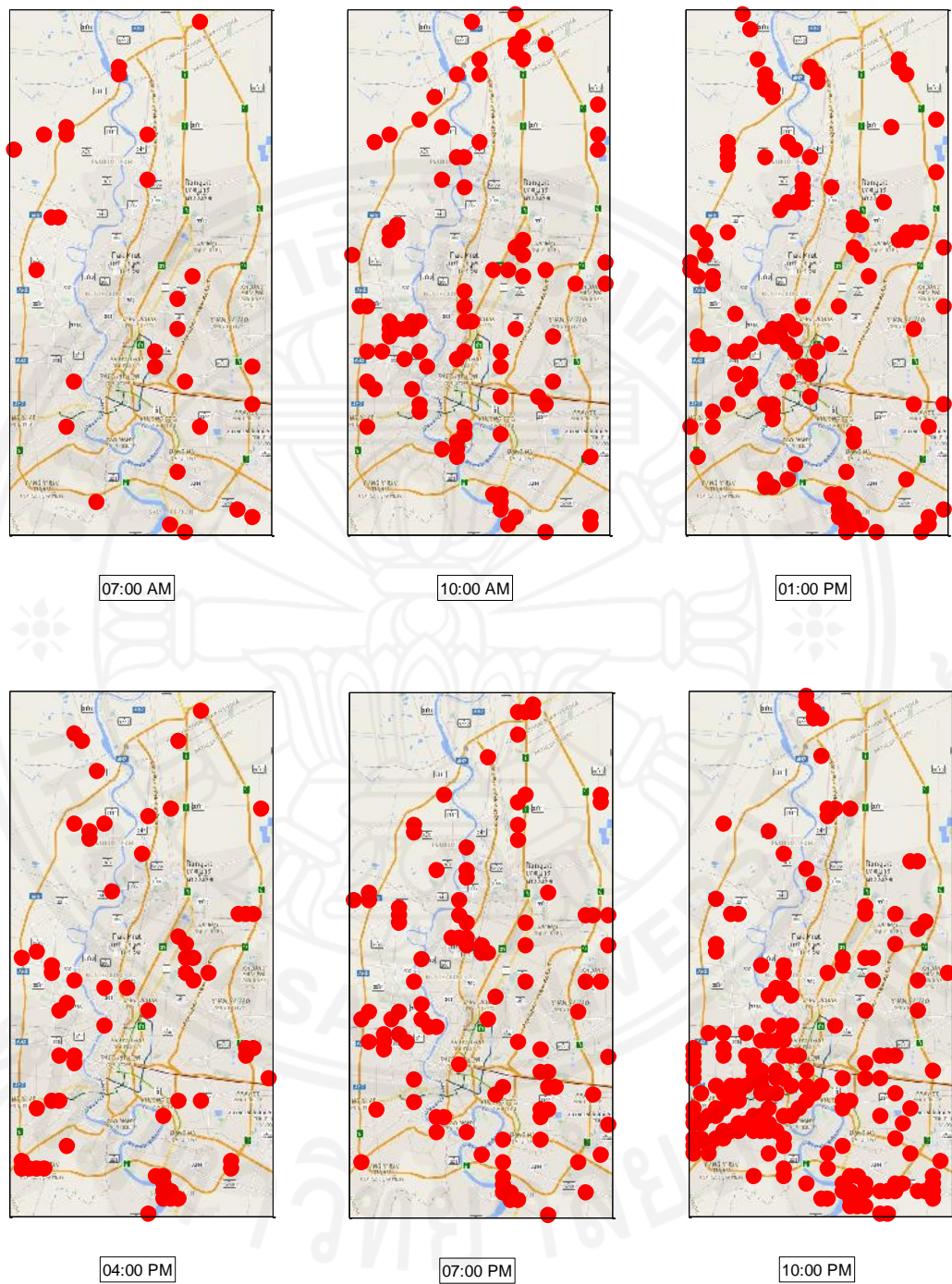
Figure 4.16 Spatio-temporal anomaly detection maps for different time of the day

with sensitivity factor α=1

69
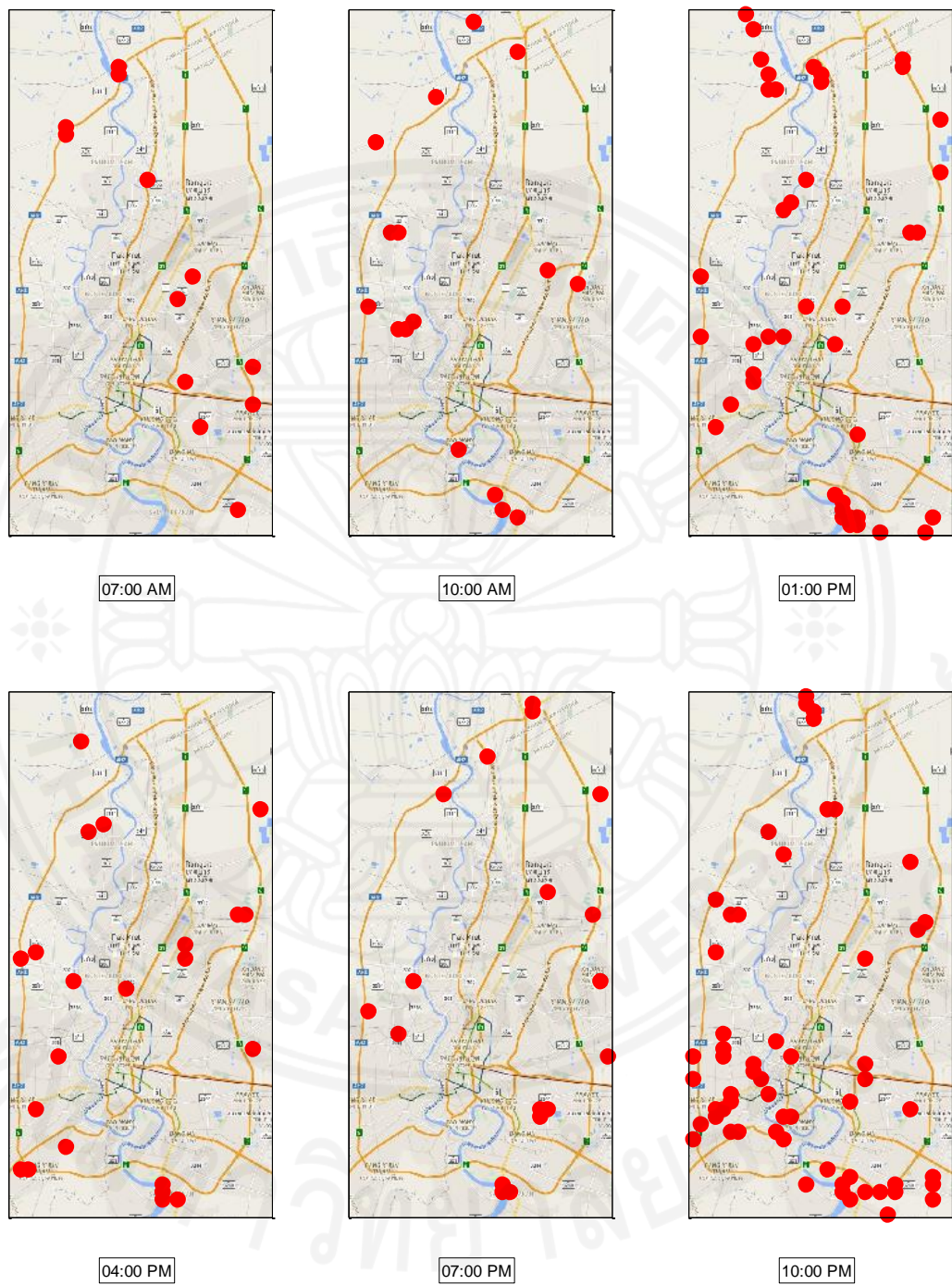
Figure 4.17 Spatio-temporal anomaly detection maps for different time of the day
with sensitivity factor α=2

07:00 AM     10:00 AM     01:00 PM

04:00 PM     07:00 PM     10:00 PM
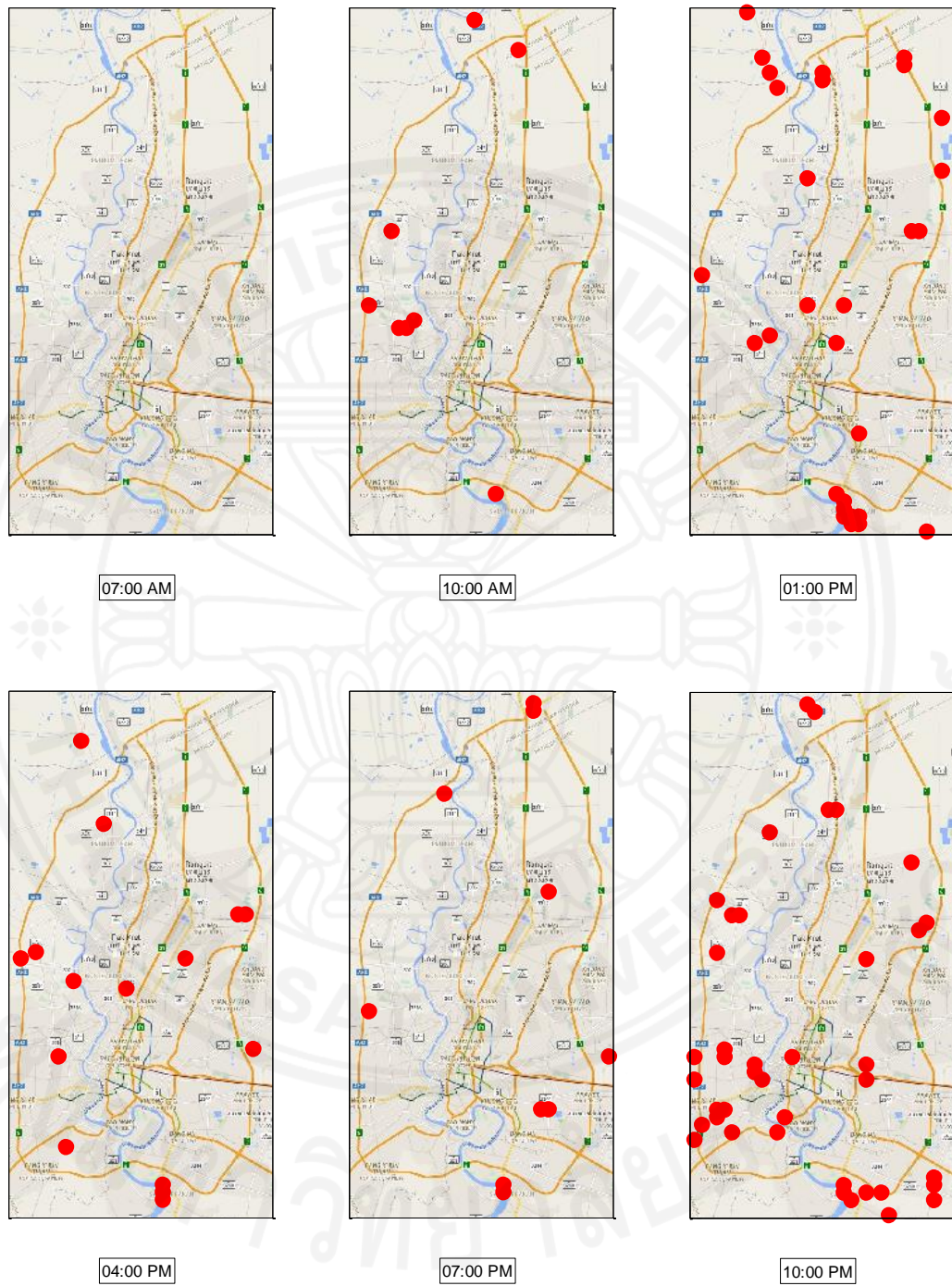
Figure 4.18 Spatio-temporal anomaly detection maps for different time of the day

with sensitivity factor α=2 and 3 time-steps delay (30 minutes)

In order to get insight to the variation of detection affected by the value of α, we generated the number of detections based on the value of α in Table 4.3. Figure 4.19. The detail values are shown in. As mentioned above, only the anomalies that last longer than 30 minutes (3 time-steps) are considered in this case.

Table 4.3 Number of detection based on value of α

| α | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Number of detection | 6537 | 1582 | 359 | 100 | 36 | 20 | 14 | 13 | 12 | 12 |



Figure 4.19 Number of detection based on value of α

In the next subsections, a specific setting, α = 3 and 3 time-steps lag (30 minutes detection) is proposed to extend our analysis to the application perspective.

### 4.6.1 Day with the highest number of traffic anomaly

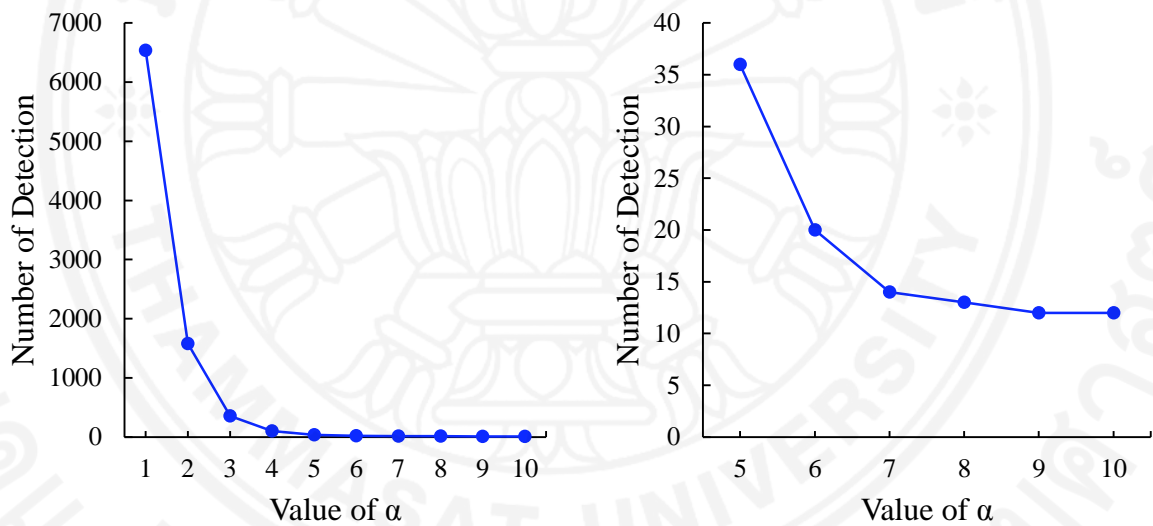With the above setting, our anomaly detection algorithm can perform on the database in order to determine the day that has the highest number of traffic anomaly happened. We can apply it to the whole period of data (1 year) or to some specific period of time, for example during one month. The result of 3 days with the highest amount of anomaly happened is shown in Table 4.4.

Table 4.4 Days with the highest number of traffic anomaly happened

| Date | Number of Anomaly happened |
|------|----------------------------|
| 29th July 2015 | 4967 |
| 28th September 2014 | 4904 |
| 30th September 2014 | 4772 |

We can also select any period of time to apply this method. For instance, we tested the data of June 2015 and tried to find the day that has the highest number of anomaly happened. The result is on 08th June 2015 with the number of detections of 4429 points for the whole day.

### 4.6.2 Area with the number of traffic anomalies

In the same context, we can also detect the area that are most likely prone to traffic anomaly. After analyzed the data for the whole year, the number of traffic anomalies happened in each area (cell) can be generated. The result is shown in the form of color heat map ranging from bright yellow to dark red representing low to high number of detection shown in Figure 4.20. The distribution of number of area associated with the number of traffic anomalies is plotted in the histogram in Figure 4.21.
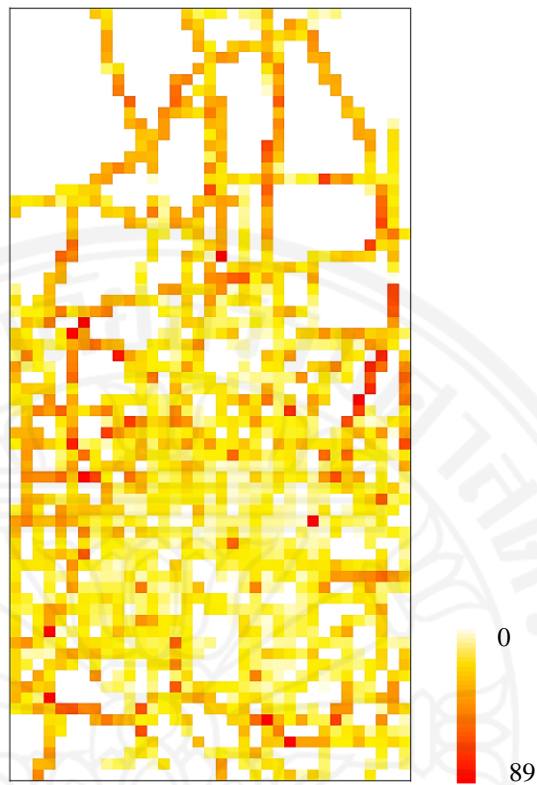
Figure 4.20 Area with the number of traffic anomalies happened
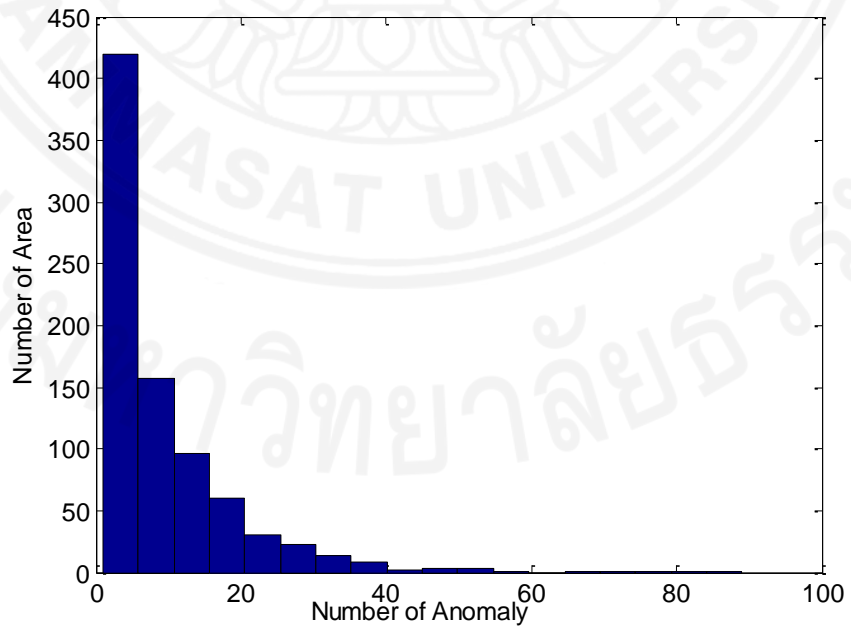


Figure 4.21 Histogram of number of area related to number of traffic anomaly

74

### 4.6.3 Special event assessment: Bike for Mom event

Bike for Mom, a remarkable event to celebrate queen's 83rd birthday in Thailand, was held on 16th August 2015. It is a special, kind of its own event because it is celebrated by bicycle riding from hundreds of thousands cyclists. In Bangkok, a big group of cyclists took the street shown in Figure 4.22 where the motorists are advised to avoid those sections of the road. This event created a high congestion in these areas. Therefore, we would like to see whether the traffic anomaly detection algorithm can detect this specific event.
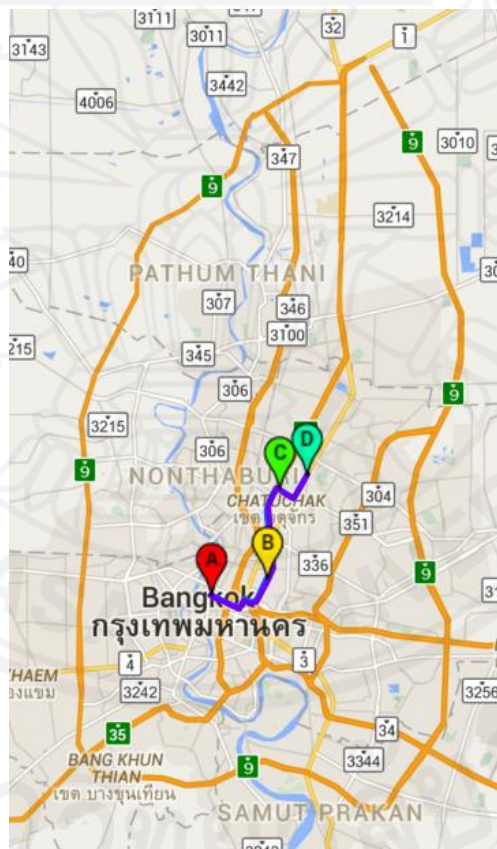


Figure 4.22 Road map of Bike for Mom event in Bangkok

Similar to previous analysis, we set the sensitivity factor, $\alpha = 3$ with 30-minute time lag and let the algorithm run for the full day. From the result of 10-minute detection, the traffic was mostly normal for the full day except there were anomalies

happened starting around 4:00 PM until 9:00 PM. The detected maps during these period are plotted in Figure 4.23.
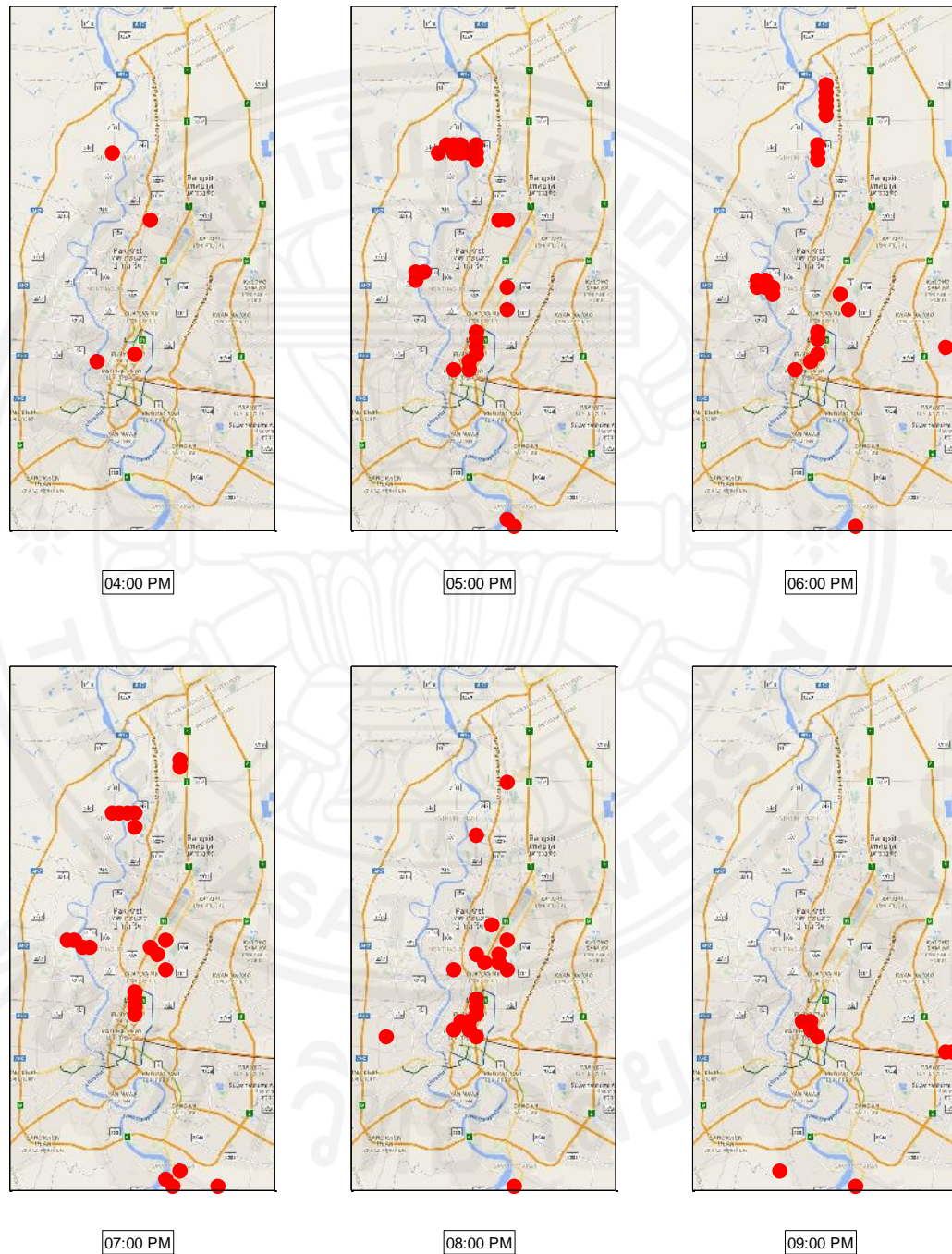


Figure 4.23 Spatio-temporal anomaly detection maps during Bike for Mom event in Bangkok

Both time and area of detection is consistent with the event. The area that anomaly happened is around Victory monument and Don Muang Tollway. The time of detection started around 4 PM and disappeared around 9 PM.

## 4.7 Traffic State Prediction

In order to test our traffic state prediction model, we have to implement it on a case study. From the whole study area, a single cell (1 km x 1 km area) that covers Silom area was chosen as the case study. In addition, we also chose a single day data in order to perform the prediction. Figure 4.24 shows the traffic intensity of Silom area for every Monday of the whole study period. We selected a random day in order to perform our prediction algorithm. This day is Monday, 20th January 2015, and its traffic state pattern is plotted in Figure 4.25. With high level of fluctuation, we also applied the moving average to smooth the data and plotted in the same graph.
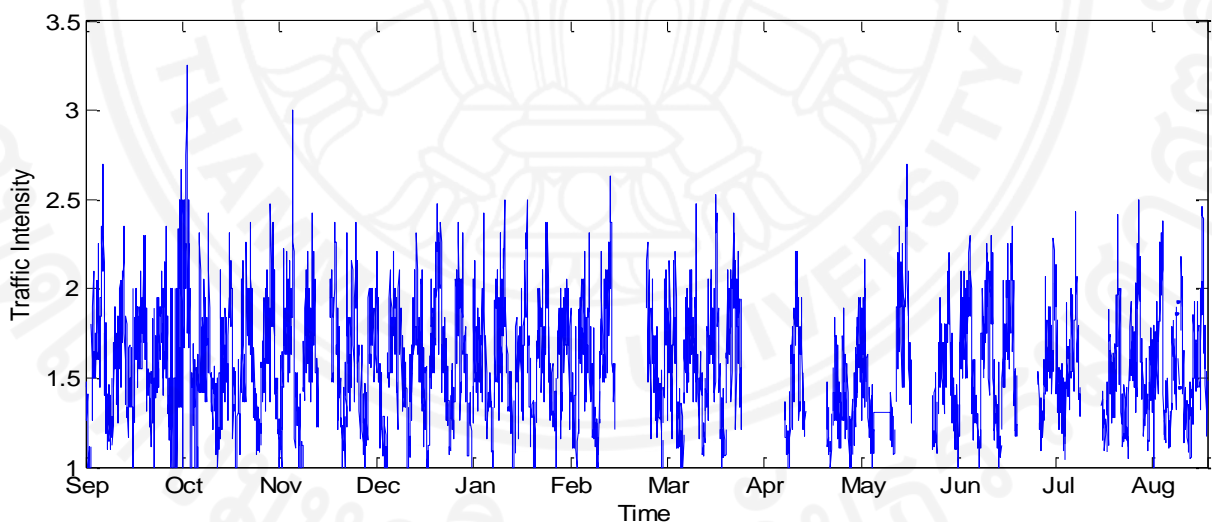


Figure 4.24 Traffic state pattern for Silom area on Monday of the whole study period

We select a traffic intensity value from a random time, for this case study, 07:00 AM, as our observation (given data). Then we used it as the input to our prediction model. The result of prediction is plotted against the actual value along with the average value of the whole year in Figure 4.25.

77
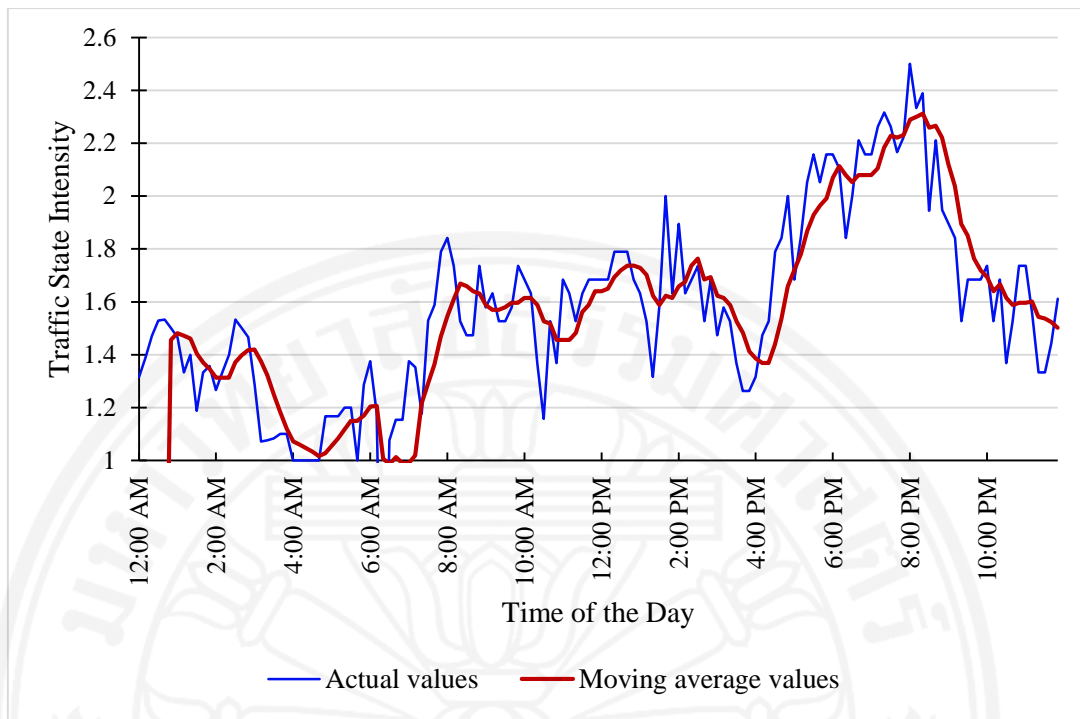
Figure 4.25 Traffic state pattern on Monday, 20th January 2015
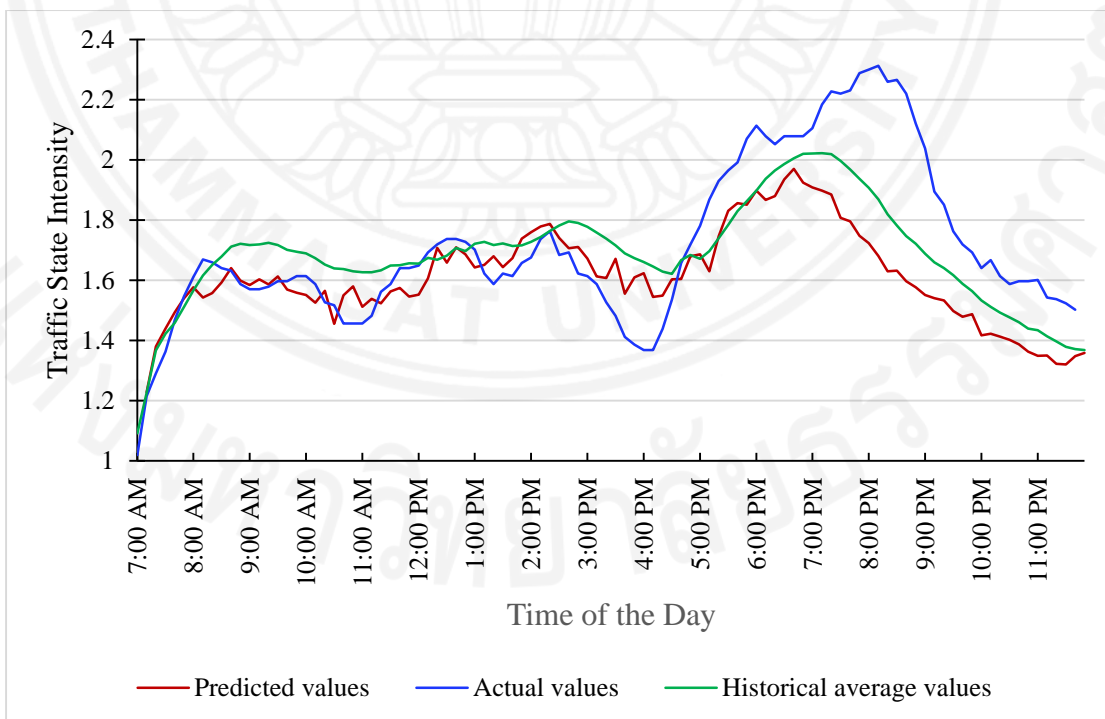


Figure 4.26 Traffic intensity from prediction, actual values and historical average

values

78

First of all, from the graph in Figure 4.25, we can see that our predicted values give a good result in terms of short period prediction. For long term prediction, the predicted values have a high error rate. Therefore, with an observed value, we can make a prediction of traffic state for the rest of the day with a better result than using the average of historical data alone.

In case we have new observed data, we can update our prediction for every time step. For this study, assuming that we obtain the new data for every time steps, we can have a much accurate result comparing to previous result. The result of the 10-minute prediction with update for every time steps is shown in Figure 4.27 and the accuracy measurement in Table 4.5.
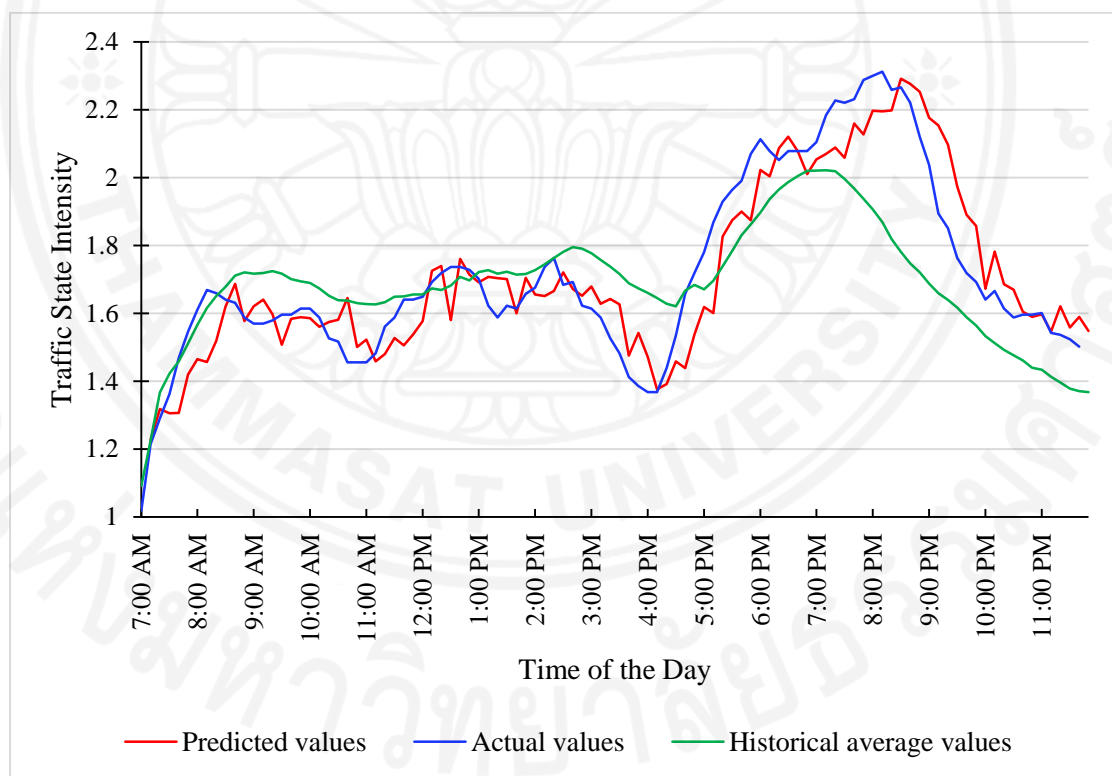


Figure 4.27 10-minutes traffic state prediction update from every time step

79

Table 4.5 Prediction error measurements for update model

| Model | MAE | MSE | RMSE | MAPE |
|---|---|---|---|---|
| Historical Average model | 0.1701 | 0.0477 | 0.2185 | 9.9532 |
| Updated prediction model | **0.0639** | **0.0064** | **0.0798** | **3.8011** |

The result from the measurement error shows that our prediction performs better than using historical average model alone.

Similarly, for every observation, we analyze different distance of prediction: 10 minutes, 20 minutes, 30 minutes, 1 hour, 2 hours, and 6 hours. The result is shown in Figure 4.28, 4.29 and 4.30.
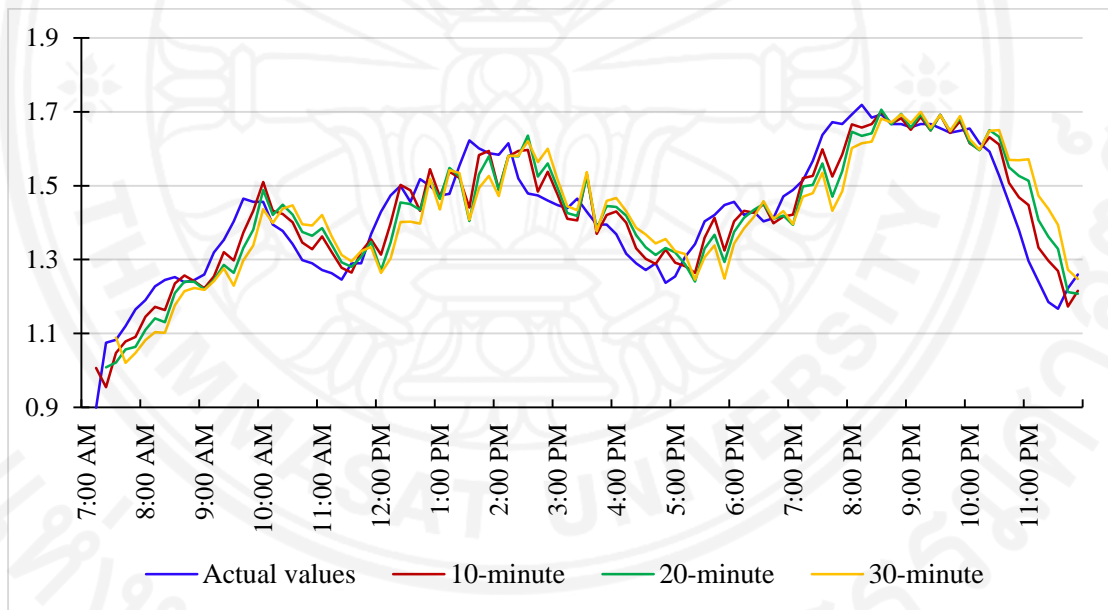


Figure 4.28 Traffic state prediction update from every time step with 10-minute, 20-minute and 30-minute prediction
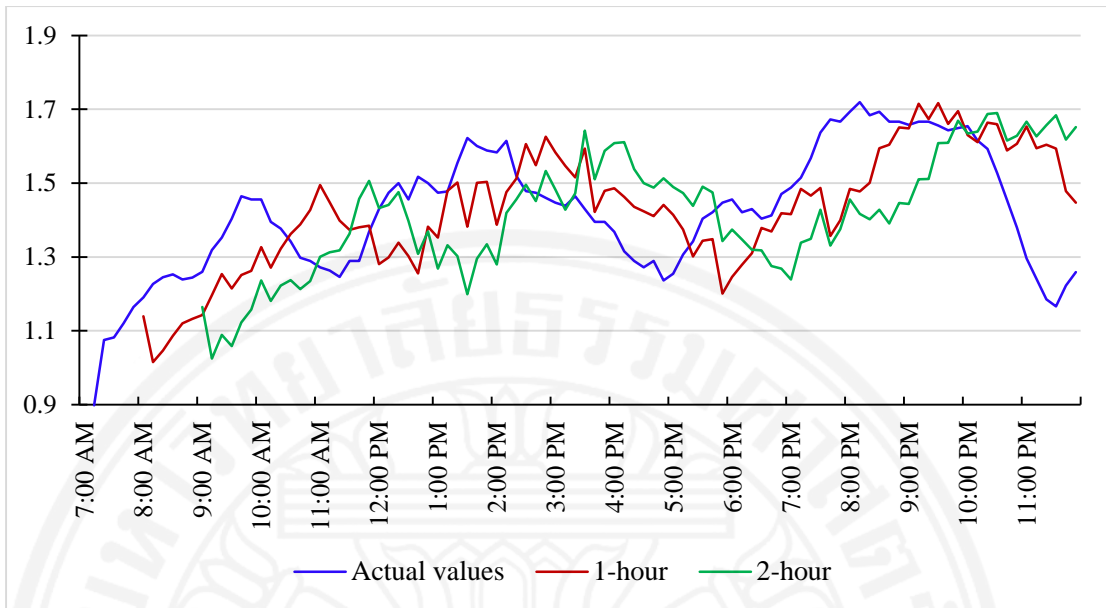
Figure 4.29 Traffic state prediction update from every time step with 1-hour and 2-hour prediction
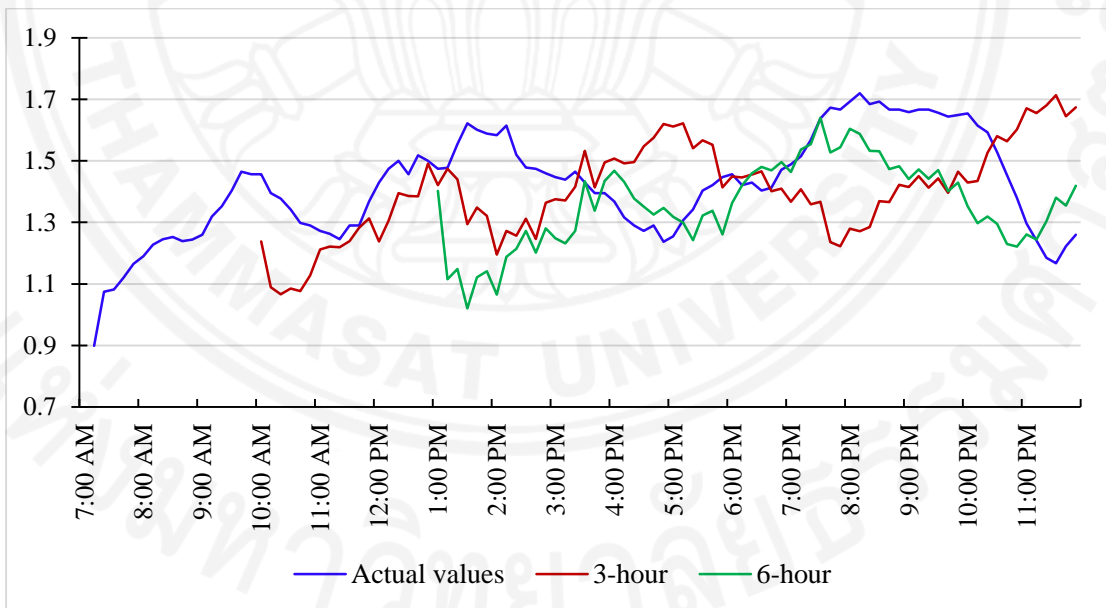


Figure 4.30 Traffic state prediction update from every time step with 3-hour and 6-hour prediction

We perform this prediction with 10 random days and generate the average of measurement errors shown in Table 4.6.

Table 4.6 Prediction error measurements for all prediction model

| Model | MAE | MSE | RMSE | MAPE |
|---|---|---|---|---|
| Historical Average model | 0.1701 | 0.0477 | 0.2185 | 9.9532 |
| **10-minute prediction model** | **0.0718** | **0.0117** | **0.0925** | **4.1297** |
| 20-minute prediction model | 0.0946 | 0.0160 | 0.1140 | 5.4703 |
| 30-minute prediction model | 0.1304 | 0.0340 | 0.1555 | 7.858 |
| 1-hour prediction model | 0.1716 | 0.0517 | 0.2065 | 10.0446 |
| 2-hour prediction model | 0.215 | 0.0794 | 0.2633 | 12.0798 |
| 3-hour prediction model | 0.2445 | 0.1099 | 0.3076 | 13.6405 |
| 6-hour prediction model | 0.2551 | 0.1042 | 0.2997 | 14.9018 |

From the table above, we can clearly see that the prediction model with 10-minute prediction perform the best. All of the error measurements from this model have the lowest values which signify the best predicted values to the actual values. This result also revealed the error is getting bigger when we try to predict farther distance from the observed values.

# Chapter 5

## Conclusions and Recommendations

As thoroughly discussed in the first chapter, traffic state information is a very important traffic information, especially for urban area. It is a key measure to comprehend traffic situation which can be interpreted into mobility and travel time spending on the road network. One of the important factors in the advanced traffic management system, traffic state information has proved itself to be a basic and valuable information for policy makers, traffic authorities and also for general public. Combining with the advantages of collecting the data from online mapping services, many analyses were carried out and revealed several interesting results in this study. Those results are temporal traffic state patterns, spatio-temporal traffic state patterns, structure of data revealed by clustering, traffic anomaly detection implementation and last but not least traffic state prediction. In the following parts, we will outline the contributions of this study along with the recommendations and future work that can be extended from this study.

### 5.1 Contribution of the Thesis

In general, the study of traffic behavior is always based on the volume of traffic or the speed of vehicle moving on the road. The traffic condition or traffic intensity of the road is usually the final result of the study or the end product information for road user. In contrast, for this study, the traffic condition on the road is used as the data input for the study. Using the traffic state data from the web-based mapping services, researchers will not face the budget constraint from collecting data, and it is also a remarkable data source for its large area coverage and accessible in every instance. After the analyses above, we can demonstrate that this traffic information is valuable for both the research and practical purposes.

For the research community, first of all this study proposed a method in collecting traffic state data from new source which is the traffic state data from web-

based mapping service. This new kind of traffic state data would open up many more opportunities to researchers to continue and use it for their own researches in the purpose of primary data input or secondary data input for verification of data. Secondly this research revealed many interesting knowledge and implementations in the field of traffic state analysis and traffic state patterns.

For practitioners such as road authority, policy maker or even the public road user, this study is beneficial in terms of giving information on traffic state patterns of the urban area where the most of the traffic problems taking place. It gives planners and responders a comprehensive look at the state of the urban road network. Furthermore, this type of analysis can be applied in any place in the world where the online traffic state data is available.

Moreover, this research is able to further clarify the high level connection between traffic flow movement and urban structure of the city. The outcome of the study can be served as a part in Advanced Transport Management System (ATMS) which is a major concern in the civilized society. Having enough information would give us more possibilities in resolving the traffic problems, especially the traffic congestion. Many more benefits from the study can be summarized as following:

- Gain awareness of the road network performance
- Enable urban traffic planners and transportation system operators to prepare and use critical information throughout the existing transportation network.
- Provide a timely-fashion information to road network planners, urban planners and citizens about current traffic condition pattern in the urban road network and also to anticipate and find solution to problems.
- Track peak-hour and congestion on the most critical areas in the network as a function of location and time.
- Provide critical information about traffic pattern in the city.

## 5.2 Recommendations and Further Study

Although there are several interesting result from the study, there are still some other parts that we can extend and further improve. First of all, our traffic state data is collected from web-based mapping services for the whole urban area, this can allow us to do many kind of analyses as mentioned before in the scope of study section. In here, we focus only on the area-based analysis (grid cells) due to the time constraints, while there are other methods such as individual point-based, corridor-based and network-based analyses which are a very nice extend of the study.

For the analyses, there are some limitations in the current study and also some recommendations for further research:

- Data collection process is not yet really flawless. There are still some issues mostly due to the internet connection and also defective instruments.

- In case that the land use data is available, detail analyses can be carried out on the relationship of urban land use with the traffic state pattern.

- For the traffic anomaly detection in this study, we leave the sensitivity factor $\alpha$ to the user to choose. However, in case that we have a complete report of anomalies on the road such as traffic accident or special events, model calibration can be performed and the appropriate sensitivity factor can be suggested.

# References

AHAS, R., AASA, A., SILM, S. & TIRU, M. 2010. Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C: Emerging Technologies,* 18**,** 45-54.

AKATSUKA, H., TAKASU, A., AIHARA, K. & ADACHI, J. Highway Incident Detection Based on Probe Car Data.  International Conference on Information Systems 2013, 2013 Lisbon, Portugal. 103-110.

ANTONIOU, C., KOUTSOPOULOS, H. N. & YANNIS, G. 2013. Dynamic data-driven local traffic state estimation and prediction. *Transportation Research Part C: Emerging Technologies,* 34**,** 89-107.

ASAKURA, Y., KUSAKABE, T., LONG, N. X. & USHIKI, T. 2015. Incident Detection Methods Using Probe Vehicles with On-board GPS Equipment. *Transportation Research Procedia,* 6**,** 17-27.

DAGANZO, C. F. 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological,* 28**,** 269-287.

DANCZYK, A. & LIU, H. X. 2011. A mixed-integer linear program for optimizing sensor locations along freeway corridors. *Transportation Research Part B: Methodological,* 45**,** 208-217.

DENG, W., LEI, H. & ZHOU, X. 2013. Traffic state estimation and uncertainty quantification based on heterogeneous data sources: A three detector approach. *Transportation Research Part B: Methodological,* 57**,** 132-157.

DUAN, H., LI, Z., LI, L., ZHANG, Y. & YIN, S. 2009. Network-Wide Traffic State Observation and Analysis Method Using Pseudo-Color Map. *Journal of Transportation Systems Engineering and Information Technology,* 9**,** 46-52.

DUDEK, C. L., MESSER, C. J. & NUCKLES, N. B. 1974. Incident detection on urban freeways. *Transportation Research Record,* 495**,** 12-24.

DUNNE, S. & GHOSH, B. 2011. Regime-based short-term multivariate traffic condition forecasting algorithm. *Journal of Transportation Engineering,* 138**,** 455-466.

GRIBBON, A. P. 1998. Field test of nonintrusive traffic detection technologies. *Mathematical and Computer Modelling,* 27**,** 349-352.

GUARDIOLA, I. G., LEON, T. & MALLOR, F. 2014. A functional approach to monitor and recognize patterns of daily traffic profiles. *Transportation Research Part B: Methodological,* 65**,** 119-136.

HELBING, D. 2001. Traffic and related self-driven many-particle systems. *Reviews of modern physics,* 73**,** 1067.

HYNDMAN, R. J. & KOEHLER, A. B. 2006. Another look at measures of forecast accuracy. *International journal of forecasting,* 22**,** 679-688.

KAUFMAN, L. & ROUSSEEUW, P. J. 2009. *Finding groups in data: an introduction to cluster analysis*, John Wiley & Sons.

KERNER, B. Theory of congested traffic flow: self-organization without bottlenecks. 14th International Symposium on Transportation and Traffic Theory, 1999.

KERNER, B. 2004. The Physics of Traffic—Empirical Freeway Pattern Features, Engineering Applications, and Theory, Understanding Complex Systems. Springer.

KIM, Y. & KELLER, H. 2002. Online traffic flow model applying the dynamic flow-density relation.

KINOSHITA, A., TAKASU, A. & ADACHI, J. Traffic Incident Detection Using Probabilistic Topic Model. EDBT/ICDT Workshops, 2014. Citeseer, 323-330.

LEDUC, G. 2008. Road traffic data: Collection methods and applications. *Working Papers on Energy, Transport and Climate Change,* 1**,** 55.

LEE, H., LEE, H.-W. & KIM, D. 1998. Origin of synchronized traffic flow on highways and its dynamic phase transitions. *Physical Review Letters,* 81**,** 1130.

LI, Z., WU, F. & GAO, X. 2007. Polarization of the global city and sociospatial differentiation in Shanghai. *Scientia Geographica Sinica,* 27**,** 304.

LIN, W.-H. & DAGANZO, C. F. 1997. A simple detection scheme for delay-inducing freeway incidents. *Transportation Research Part A: Policy and Practice,* 31**,** 141-155.

LIU, H., VAN ZUYLEN, H., VAN LINT, H. & SALOMONS, M. 2006. Predicting urban arterial travel time with state-space neural networks and Kalman filters. *Transportation Research Record: Journal of the Transportation Research Board,* 1968**,** 99-108.

LIU, Y., WANG, F., XIAO, Y. & GAO, S. 2012. Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning,* 106**,** 73-87.

MAAT, K., VAN WEE, B. & STEAD, D. 2005. Land use and travel behaviour: expected effects from the perspective of utility theory and activity-based theories. *Environment and Planning B: Planning and Design,* 32**,** 33-46.

MIHAYLOVA, L., BOEL, R. & HEGYI, A. 2007. Freeway traffic estimation within particle filtering framework. *Automatica,* 43**,** 290-300.

MUÑOZ, L., SUN, X., HOROWITZ, R. & ALVAREZ, L. Traffic density estimation with the cell transmission model. American Control Conference, 2003. Proceedings of the 2003, 2003. IEEE, 3750-3755.

NAGEL, K. & SCHRECKENBERG, M. 1992. A cellular automaton model for freeway traffic. *Journal de physique I,* 2**,** 2221-2229.

NATIONAL RESEARCH, C. & TRANSPORTATION RESEARCH, B. 2010. *HCM 2010 : highway capacity manual* [Online]. Washington, D.C.: Transportation Research Board. Available: http://www.hcm2010.org/.

NEWELL, G. F. 1993. A simplified theory of kinematic waves in highway traffic, part I: General theory. *Transportation Research Part B: Methodological,* 27**,** 281-287.

NI, D. 2013. Chapter 1: Traffic Sensing Technologies. *Lecture Notes in Traffic Flow Theory.* Department of Civil and Environmental Engineering, University of Massachusetts Amherst, MA 01003, USA.

PAYNE, H. J. 1997. Development and testing of operational incident detection algorithms: executive summary. 19p.

PERRY, C. & TURNBULL, K. Implementation and Assessment of An Automated Vehicle Location System for Rural Transit Using GPS and UHF Radio Technologies. Merging the Transportation and Communications Revolutions. Abstracts for ITS America Seventh Annual Meeting and Exposition, 1997.

SCHRANK, D., EISELE, B. & LOMAX, T. 2012. TTI's 2012 urban mobility report. *Proceedings of the 2012 annual urban mobility report. Texas A&M Transportation Institute, Texas, USA.*

SONG, Y. & MILLER, H. 2012. Exploring traffic flow databases using space-time plots and data cubes. *Transportation,* 39**,** 215-234.

SUMALEE, A., ZHONG, R. X., PAN, T. L. & SZETO, W. Y. 2011. Stochastic cell transmission model (SCTM): A stochastic dynamic traffic model for traffic state surveillance and assignment. *Transportation Research Part B: Methodological,* 45**,** 507-533.

SUN, J. & SUN, J. 2015. A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data. *Transportation Research Part C: Emerging Technologies,* 54**,** 176-186.

THORNDIKE, R. L. 1953. Who belongs in the family? *Psychometrika,* 18**,** 267-276.

TIPAKORNKIAT, C., LIMANOND, T. & KIM, H. 2012. Determining an influencing area affecting walking speed on footpath: A case study of a footpath in CBD Bangkok, Thailand. *Physica A: Statistical Mechanics and its Applications,* 391**,** 5453-5464.

TOSTES, A. I. J., DE LP DUARTE-FIGUEIREDO, F., ASSUNÇÃO, R., SALLES, J. & LOUREIRO, A. A. From data to knowledge: city-wide traffic flows analysis and prediction using bing maps. Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, 2013. ACM, 12.

TURNER, S., MARGIOTTA, R. & LOMAX, T. 2004. Lessons learned: monitoring highway congestion and reliability using archived traffic detector data.

TURNER, S. M., EISELE, W. L., BENZ, R. J. & HOLDENER, D. J. 1998. Travel time data collection handbook.

VAN LINT, J. & VAN HINSBERGEN, C. 2012. Short-term traffic and travel time prediction models. *Artificial Intelligence Applications to Critical Transportation Issues,* 22.

VAN WIJK, J. J. & VAN SELOW, E. R. Cluster and calendar based visualization of time series data. Information Visualization, 1999. (Info Vis '99) Proceedings. 1999 IEEE Symposium on, 1999 1999. 4-9, 140.

VLAHOGIANNI, E. I., KARLAFTIS, M. G. & GOLIAS, J. C. 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. *Transportation Research Part C: Emerging Technologies,* 13**,** 211-234.

VLAHOGIANNI, E. I., KARLAFTIS, M. G. & GOLIAS, J. C. 2008. Temporal Evolution of Short-Term Urban Traffic Flow: A Nonlinear Dynamics Approach. *Computer‐Aided Civil and Infrastructure Engineering,* 23**,** 536-548.

WANG, Y., PAPAGEORGIOU, M. & MESSMER, A. 2006. A real-time freeway network traffic surveillance tool. *Control Systems Technology, IEEE Transactions on,* 14**,** 18-32.

WARREN LIAO, T. 2005. Clustering of time series data—a survey. *Pattern Recognition,* 38**,** 1857-1874.

WEN, H., SUN, J. & ZHANG, X. 2014. Study on Traffic Congestion Patterns of Large City in China Taking Beijing as an Example. *Procedia - Social and Behavioral Sciences,* 138**,** 482-491.

WU, C.-H., HO, J.-M. & LEE, D.-T. 2004. Travel-time prediction with support vector regression. *Intelligent Transportation Systems, IEEE Transactions on,* 5**,** 276-281.

YILDIRIMOGLU, M. & GEROLIMINIS, N. 2013. Experienced travel time prediction for congested freeways. *Transportation Research Part B: Methodological,* 53**,** 45-63.

ZHU, T., WANG, J. & LV, W. 2009. Outlier mining based Automatic Incident Detection on urban arterial road. *Proceedings of the 6th International Conference on Mobile Technology, Application \&\#38; Systems.* Nice, France: ACM.

**Appendices**

# Appendix A

## Algorithm 1: Color-Indexes conversion algorithm

```matlab
function Result= Code_11(x,y,Folder)

Answer = zeros (length(x) ,8);

Cell_ID_X = ceil((x-141)/433);
Cell_ID_Y = ceil(y/433);
Concat = [];
color = zeros (length(x),1);
Imgs = dir([Folder '/' '*.png']);
NumImgs = size(Imgs,1);

%% Sorting the picture files
Imgsfields = fieldnames(Imgs);
Imgscell = struct2cell(Imgs);
sz = size(Imgscell);

% Convert to a matrix
Imgscell = reshape(Imgscell, sz(1), []);

% Make each field a column
Imgscell = Imgscell';

% Sort by first field "name"
Imgscell = sortrows(Imgscell, 2);

% Put back into original cell array format
Imgscell = reshape(Imgscell', sz);

% Convert to Struct
Imgssorted = cell2struct(Imgscell, Imgsfields, 1);
```

```matlab
%% Read idividual files
tic
for k=1:NumImgs
    Ans =double(imread([Folder '/' Imgssorted(k).name]));
    pixel=impixel(Ans,x,y);

    %write to matrix
    Answer (:,5:7,k) =   pixel;

    %Read color into code
    for i=1:length(pixel);
        if (pixel(i,1)<240) && (pixel(i,2)>170) && (pixel(i,3)<210);
            color(i)=1;
        elseif (pixel(i,1)>240) && (pixel(i,2)>150) && (pixel(i,3)<200);
            color(i)=2;
        elseif (pixel(i,1)>240) && (pixel(i,2)<150) && (pixel(i,3)<200);
            color(i)=3;
        elseif (pixel(i,1)<240) && (pixel(i,2)<170) && (pixel(i,3)<200);
            color(i)=4;
        else color(i)=0;
        end
    end

    %write to matrix
    Answer (:,1,k) = x;
    Answer (:,2,k) = y;
    Answer (:,3,k) = Cell_ID_X;
    Answer (:,4,k) = Cell_ID_Y;
    Answer (:,8,k) = color;
    clear pixel
    clear Ans
    Concat = horzcat(Concat, Answer(:,8,k));
end

Result = horzcat(Answer(:,1:4), Concat);
toc
end
```