Title
Modeling Vehicular Traffic Shock Wave with Machine Learning Approaches

Author
Kim, Jihyoung

Publication Date
2016

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Modeling Vehicular Traffic Shock Wave with Machine Learning Approaches

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Computer Science

by

Jihyoung Kim

2016

Abstract of the Dissertation

Modeling Vehicular Traffic Shock Wave with Machine Learning Approaches

by

Jihyoung Kim

Doctor of Philosophy in Computer Science

University of California, Los Angeles, 2016

Professor Mario Gerla, Chair

The current trend of transforming old cities into smart cities has revealed many issues of the modern cities. One of the issues is the prevailing traffic jam on highways of the modern cities. The vehicular shock wave has been a problem on highways since it is one of the main causes of the traffic jam. The combination of heavy traffic and small traffic perturbations or unexpected driver actions are the main causes of shock waves. In order to alleviate road traffic caused by shock waves, it is crucial to have a system that predicts shock waves and informs them to the drivers. In this dissertation, we analyzed 6 months of freeway traffic data of Los Angeles, CA, provided by CalTrans PeMS (Performance Measurement System) and obtained the vehicular shock wave propagation speed of each freeway. Based on this information, we propose a machine learning approaches to predict shock waves. We utilize Hidden Markov Model (HMM) to predict if the shock wave will occur and propagate based on neighboring lanes' traffic information. Addtionally, HMM is used to estimate the probability of lane change from one lane to other lanes based on the occupancy of a lane. Baum-Welch algorithm is used to predict the parameters (occupancy and state).

We also utilized Deep Learning (DL) in order to predict the shock wave occurrence and propagation. We compared Stacked AutoEncoder (SAE), Deep Belief Networks (DBN), and HMM for the accuracy of the prediction of shock wave occurrences and propagation. These approaches have been tested on the same PeMS data sets and achieved good accuracy.

In the future, our models will be used to include modern collision prevention techniques (e.g., anti-shock wave strategies) to test their efficacy and help to reduce the number of potential accidents and save energy in the process. Also, our models can be used to improve traffic simulators to provide driving patterns that are close to real human's.

The dissertation of Jihyoung Kim is approved.

Jack W. Carlyle

Carlo Zaniolo

Gregory J. Pottie

Mario Gerla, Committee Chair

University of California, Los Angeles

2016

To my father, mother, and Prof. Gerla

the journey has been long...

Thank you for all your support.

Table of Contents

1 Introduction **1**

1.1 Traffic Models 2

1.2 Overview 2

1.3 Contributions 3

1.4 Dissertation Outline 4

2 Vehicular Shock Wave (Vehicular Traffic Shock) **5**

2.1 Vehicular Shock Wave Formation 5

2.1.1 Single-Lane Scenario 6

2.1.2 Multi-Lane Scenario 7

2.2 Traffic Models 8

2.2.1 Microscopic Models 9

2.2.2 Macroscopic Models 11

3 Exploratory Data Analysis **13**

3.1 CalTrans PeMS Data Set 13

3.1.1 Structure of the data sets 13

3.1.2 Station Status 17

3.2 Vehicular Shock Wave Propagation Speed 19

3.2.1 Cross-Validation 30

3.3 Selecting the Right Feature to Infer Shock Wave Occurrence 31

3.4 Summary 34

4 Hidden Markov Model **35**

4.1 Shock Wave Inference 35

4.2 Lane Change Prediction 37

4.2.1 The Model 37

4.2.2 Baum-Welch Algorithm 39

4.3 Prediction Results 41

4.3.1 Shock Wave Prediction Accuracy 41

4.3.2 Lane Switch Prediction 43

4.4 Summary 44

5 Deep Learning **45**

5.1 Encoders 46

5.1.1 Autoencoder 46

5.1.2 Stacked Autoencoder 47

5.2 Boltzmann Machine 48

5.2.1 Restricted Boltzmann Machine 48

5.2.2 Deep Belief Network 50

5.3 Inference Results Comparison 51

5.3.1 Shock Wave Prediction 52

5.4 Summary 54

6 Conclusion **55**

References **57**

List of Figures

2.1 Shock wave forming on a single-lane road 6

2.2 Shock wave forming on a multi-lane road 7

2.3 Velocity Density Chart 12

2.4 Flow Density 12

3.1 The locations of functioning stations on I-405 South 19

3.2 The locations of functioning stations on I-405 North 20

3.3 The locations of functioning stations on I-5 South 20

3.4 The locations of functioning stations on I-5 North 21

3.5 The locations of functioning stations on I-105 East 21

3.6 The locations of functioning stations on I-105 West 22

3.7 The locations of functioning stations on I-10 East 22

3.8 The locations of functioning stations on I-10 West 23

3.9 The locations of functioning stations on I-110 South 23

3.10 The locations of functioning stations on I-110 North 24

3.11 The locations of functioning stations on I-710 South 24

3.12 The locations of functioning stations on I-710 North 25

3.13 The locations of functioning stations on I-605 South 25

3.14 The locations of functioning stations on I-605 North 26

3.15 The locations of functioning stations in LA city 26

3.16 Vehicular shock wave propagation speed on left lane of each station of I-405 South (Jan/4/2016 2:00-8:00 pm) 27

3.17 Vehicular shock wave propagation speed on left lane of each station of I-405 South (Jan/4/2016 5:00-10:00 am) 28

3.18 Vehicular shock wave propagation speed on left lane of each station of I-5 South (Jan/4/2016 5:00-10:00 am) 29

3.19 k-fold cross-validation 30

3.20 Speed vs Occupancy plot and Speed vs Flow plot 32

3.21 Inferring speed using Polynomial Regression using occupancy of the left lane of Station 718296 (I-405) from 1:00-8:00 pm on weekdays . . 32

3.22 Predicted speed of the left lane of Station 718296 (I-405) from 2:00-8:00 pm on weekdays 33

3.23 Vehicular shock wave propagation speed on left lane of each station of I-405 South (Jan/4/2016 2:00-8:00 am) from Occupancy Data 33

4.1 Hidden Markov Model represented as a Dynamic Bayesian Network (One station) 36

4.2 Hidden Markov Model represented as a Dynamic Bayesian Network (Multi stations) 37

5.1 Stacked Autoencoder with a binary Logistic Regression Predictor . . 47

5.2 Restricted Boltzmann Machine 49

5.3 Deep Belief Network composed of Restricted Boltzmann Machine . . 51

List of Tables

3.1 PeMS Station Raw Data Set Features 14

3.2 PeMS Station 5-min Data Set Features 16

3.3 Total number of stations and total number functioning stations on I-405 17

3.4 Total number of stations and total number functioning stations on I-5 18

3.5 Total number of stations and total number functioning stations on I-105 18

3.6 Total number of stations and total number functioning stations on I-10 18

3.7 Total number of stations and total number functioning stations on I-110 18

3.8 Total number of stations and total number functioning stations on I-710 18

3.9 Total number of stations and total number functioning stations on I-605 19

4.1 Confusion Matrix of Lane 1 of I-405 42

4.2 Confusion Matrix of Lane 2 of I-405 42

4.3 Confusion Matrix of Lane 3 of I-405 42

4.4 Confusion Matrix of Lane 4 of I-405 43

4.5 Final Transition Matrix of Lane 1 of I-405 43

4.6 Final Transition Matrix of Lane 2 of I-405 44

4.7 Final Transition Matrix of Lane 3 of I-405 44

5.1 Trainign SAE Algorithm 48

5.2 Trainign RBM Algorithm 50

5.3 Trainign Deep Belief Network 51

5.4 Confusion Matrix of Lane 1 (SAE) 52

5.5 Confusion Matrix of Lane 2 (SAE) 52

5.6 Confusion Matrix of Lane 3 (SAE) 52
5.7 Confusion Matrix of Lane 4 (SAE) 52
5.8 Confusion Matrix of Lane 1 (DBN) 53
5.9 Confusion Matrix of Lane 2 (DBN) 53
5.10 Confusion Matrix of Lane 3 (DBN) 53
5.11 Confusion Matrix of Lane 4 (DBN) 53
5.12 Prediction Accuracy Comparison between SAE, DBN, and HMM . . 53

Vita

2009 Master of Science, University of California, Los Angeles

Publications

Uichin Lee, Jihyoung Kim, Eunhee Yi, Juyup Sung, Mario Gerla, Analyzing Crowd Workers in Mobile Pay-for-Answer Q&A, *in the ACM SIGCHI Conference on Human Factors in Computing Systems*, April 2013.

Sungwon Yang, Jihyoung Kim, Mario Gerla, Clinical Quality Guaranteed Physiological Data Compression in Mobile Health Monitoring, *The Second ACM MobiHoc Workshop on Pervasive Wireless Healthcare In Conjunction With MobiHoc 2012 Conference (ACM MobileHealth 2012)*, June 2012.

Seongwon Han, Sungwon Yang, Jihyoung Kim, Mario Gerla, Eye Guardian: A Framework of Eye Tracking and Blink Detection for Mobile Device Users, *The Thirteenth Workshop on Mobile Computing Systems and Applications (ACM HotMobile 2012)*, February 2012.

Jihyoung Kim, Sungwon Yang, Mario Gerla, StrokeTrack: Wireless Inertial Motion Tracking of Human Arms for Stroke Telerehabilitation, *mHealthSys: First International Workshop on Mobile Systems, Applications, and Services for Healthcare held at ACM SenSys 2011*, November 2011.

Jung Soo Lim, Jihyoung Kim, Jonathan Friedman, Uichin Lee, Luiz Vieira, Diego Rosso, Mario Gerla, Mani B Srivastava, SewerSnort: A Drifting Sensor for Wastewater Collection System Gas Monitoring, *Ad Hoc Networks Journal 2010*.

Jihyoung Kim, Jung Soo Lim, Jonathan Friedman, Uichin Lee, Luiz Vieira, Diego Rosso, Mario Gerla, Mani B Srivastava, SewerSnort: A Drifting Sensor for In-situ Sewer Gas Monitoring, *Sixth Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON 2009)*, June 2009.

CHAPTER 1

Introduction

Recently, transforming metropolises into smart cities has been a global trend. As a part of the smart city project, efficient traffic management also has been emphasized to mitigate heavy traffic issues during rush hours [BMS12]. The traffic demand on major highways, one of the major causes of heavy traffic jam on metropolis areas, has been increasing for the past several decades. The metropolises have resolved the aforementioned issue either by constructing new highways or extending the existing ones [FFG14]. However, due to several reasons, extending existing highways or constructing new highways is no longer a possible option [Bri06, FFG14]. There are studies that state the capacity of highways is not fully exploited [FFG14]. Forster et al. claim by fully exploiting the capacity of highways, the current high traffic demand on highways could be alleviated [FFG14]. The recent advent of autonomous vehicles that equip several sensors and recognize the surrounding environment has antedated the necessity of smart traffic management in smart cities. In order to provide smart traffic information to each vehicle on the road, each vehicle is required to have vehicular network enabled [BMS12]. In this paper, we will focus on providing smart traffic information to vehicles in order to mitigate heavy traffic issue using publicly open data.

1.1 Traffic Models

Traffic models are classified into two big categories, macroscopic traffic models and microscopic traffic models. Microscopic models focus on interactions between individual cars while macroscopic models focus on general overview of traffic status of certain road segment. Forster et al. combined two different traffic models, microscopic traffic model and macroscopic traffic model, to simulate their proposed CADAS [FFG14]. The assumption of their system is that it is possible to obtain both microscopic traffic data and macroscopic data. However, in reality it is very hard to obtain data that contains microscopic traffic information due to privacy issues. Only several companies such as cell phone carriers, Google maps including Waze, Apple maps, Bing maps, and so forth are collecting microscopic traffic information. The aforementioned companies are not disclosing their microscopic traffic data. On the contrary, in the United States, there are many open data sets on macroscopic traffic information such as CalTrans (California Department of Transportation) PeMS (Performance Measurement System) data sets [Tra15]. The challenging aspect of using PeMS data set is the traffic status is not monitored every second. The data set with the finest granularity is the one collected every 30 seconds, i.e., aggregated traffic information. Since the traffic information is an aggregated version, we have to find a way to infer the some information that are not available to the finest granularity data set, e.g., speed (velocity).

1.2 Overview

The goal of this dissertation is to infer aggregated driving patterns as much as possible from the macroscopic traffic information given real world traffic data. There have been many research papers on short term traffic prediction using CalTrans PeMS data sets [Tra15]. However, to the best of our knowledge, this paper is the first pa-

per that applies machine learning algorithms to predict vehicular shock wave and the probability of lane change from one lane to another with real world big traffic data. Especially, we compared state-of-the-art machine learning algorithm such as Deep Learning in order to see the accuracy of our models. We believe our model will contribute to projects that plan on providing real world driving patterns.

1.3 Contributions

The contributions of this dissertation is three-fold.

- We analyzed the PeMS data that was collected for 6 months from January 1, 2016 to June 30, 2016. We removed the stations (loop detectors) that do not collect traffic status in order to obtain accurate data. Through this process we could gain insights of the traffic information collected by stations of each freeway.

- We described what feature of the macroscopic model has the most correlation between shock wave occurrence and propagation, and build an Hidden Markov Model (HMM) that predicts the occurrences of shock wave. We also utilized HMM in order to predict the probability of lane change the lane based on the occupancy. Baum-Welch algorithm was used in order to predict the probability of a car switching lanes of a single lane.

- We utilized Deep Learning (DL) algorithms to predict the shock wave occurrences. We specifially selected Stacked Autoencoder (SAE) and Deep Belief Network among DL algorithms. We compared the accuracy of shock wave prediction from HMM, SAE, and Deep Belief Network.

1.4 Dissertation Outline

The rest of the dissertation is organized as follows. We discuss the formation of vehicular shock wave in both single-lane scenario and multi-lane scenario in Chapter 2. In Chapter 3, we analyze PeMS data sets and obtain average vehicular shock wave propagation speed. Based on this information, we infer the average speed collected by each station every 30 seconds. The 30 second data set does not have aggregated speed information. In Chapter 4, we introduce Hidden Markov Model and Baum-Welch algorithm in order to predict the probability of lane switch of a car for the next 30 second time period. Based on the prediction obtained in Chapter 4, we labelled which lane the majority of cars will stay in the next 30 seconds in Chapter 5, we utilized Stacked Autoencoder and Deep Belief Network to see how accurately Deep Learning algorithms predict the shock wave occurrences. Finally, in Chapter 6, we summarize the results obtained from each model and conclude the dissertation.

CHAPTER 2

Vehicular Shock Wave (Vehicular Traffic Shock)

In this chapter, a brief introduction to vehicular shock wave formation and scenarios of vehicular shock wave occurrence will be provided. Additionally, traffic models that are used for inferring vehicular shock wave (traffic shock) will be introduced. Specifically, two big categories of traffic models, microscopic traffic models and macroscopic traffic models, will be introduced.

2.1 Vehicular Shock Wave Formation

In this section, the mechanism of the vehicular shock wave will be introduced. By definition, vehicular shock waves (vehicular traffic shocks) are traveling disturbances in vehicles distribution on highways. The main cause of the vehicular shock wave is due to the high traffic demand and the unexpected driver actions [FFG14]. Due to the characteristics of the shock wave, the denser traffic on highways, the more aggravating shock wave is formed upward in the traffic stream. That is, a minor change in the driver's behavior will cause a shock wave upward in the traffic stream. Additionally, physical perturbation such as ramps, road works, and small amount of increase and decrease in traffic of a lane cause drivers to change their driving patterns. Moreover, driving behavior such as delayed reaction to traffic changes is prone to cause traffic disruption that leads to traffic congestion [FFG14, Wir78, BSS98].

Consider two scenarios that demonstrate the formation of the shock wave. In

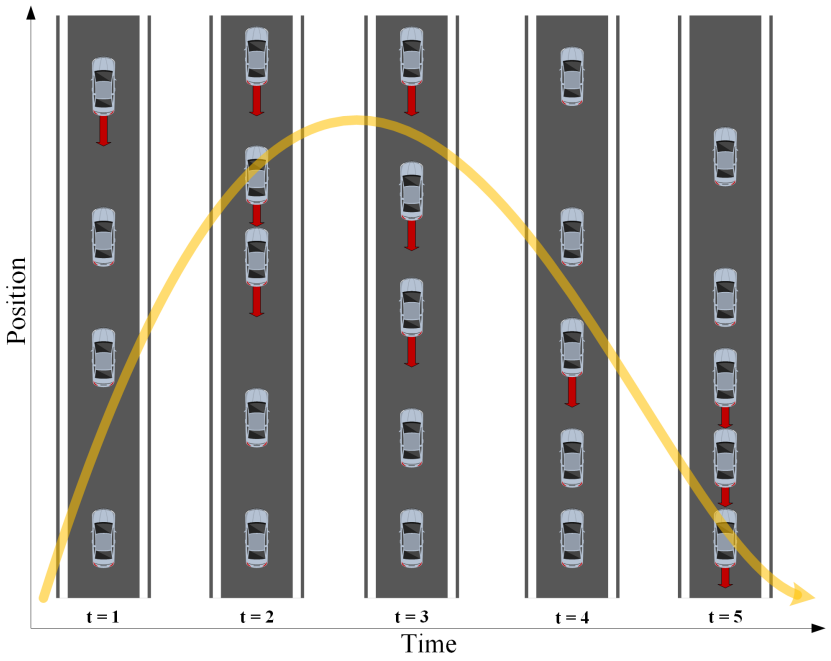

Figure 2.1: Shock wave forming on a single-lane road

particular, a single-lane road scenario and a multi-lane (two-lane) scenario will be covered.

2.1.1 Single-Lane Scenario

In Figure 2.1, the formation of a shock wave on a single-lane road is presented. Each timestamp has the current traffic status a single lane road. The x-axis indicates timestamps and the y-axis indicates the position in the road. At $t = 1$, the leading vehicle is slowing down for unknown reason causing the queueing upstream that affects the trailing vehicles, and the red arrow indicates the influence on the following vehicles. At $t = 2$, the first trailing vehicle has to slow down in order to avoid collision. At the following timestamps, the other trailing vehicles have to slow down in order to avoid accident. The whole phenomenon from $t = 1$ through $t = 5$ forms a vehicular shock wave or traffic shock due to the continuous slow down of all vehicles following

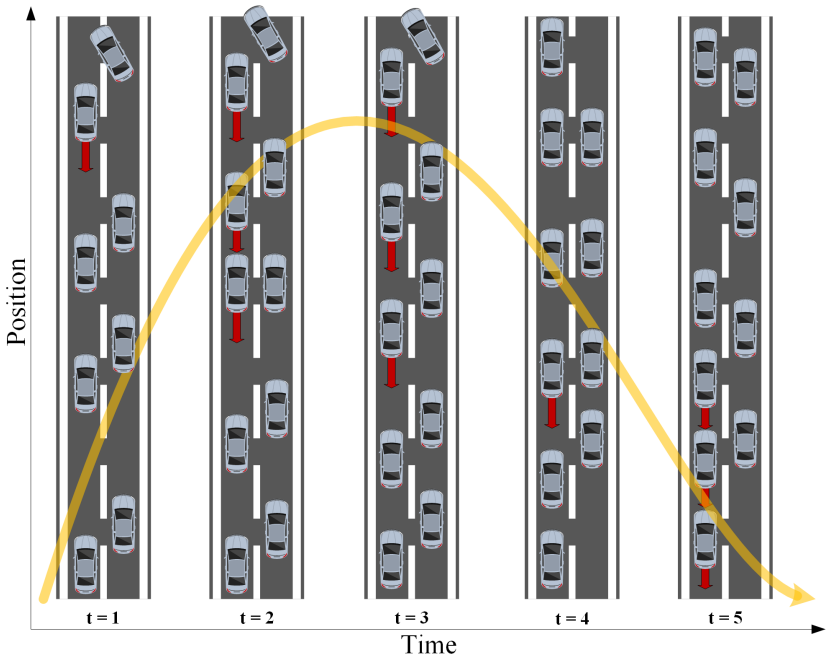

Figure 2.2: Shock wave forming on a multi-lane road

the leading vehicle. The velocity of a following vehicle cannot be greater than that of the car right ahead. Otherwise, collision will occur. Moreover, according to the literature, the aforementioned shock waves occur without a reason and cause traffic jam, and it is referred as *phantom jams* [FFG14, HPM90, Wil08]. Another case that causes traffic jam by the shock wave is the so called "moving bottleneck". This is occurred by the decreased velocity of the leading vehicle and all the trailing vehicles have to adapt to the new velocity, e.g., a fully loaded truck is the leading vehicle and faster vehicles have to slow down on a single-lane road.

2.1.2 Multi-Lane Scenario

Figure 2.2 demonstrates the formation of a shock wave on a two-lane road. Although only a two-lane scenario is presented in this section, the main causes of the formation of a shock wave on a road that has more than two lanes are identical.

Therefore, we can consider the same phenomenon will occur on a multi-lane road. Each timestamp has the current traffic status a single lane road. On multi-lane road, one of the reasons of the formation of the shock wave is the same situation as the single-lane scenario. Hence, the single-lane scenario, i.e., shock wave occurred by the slow down of the leading vehicle, will not be covered in this chapter.

The x-axis indicates timestamps and the y-axis indicates the position in the road. At $t = 1$, the leading vehicle on the right lane is trying to switch lane to the left. The leading vehicle on the left lane has to slow down, otherwise, a collision will occur. At $t = 2$, the first trailing vehicle on the left lane has to slow down in order to avoid collision. At the following timestamps, the other trailing vehicles on the left lane have to slow down in order to avoid accident. The whole phenomenon occurred on left lane of the two-lane road from $t = 1$ through $t = 5$ forms a vehicular shock wave due to the continuous slow down of all vehicles following the leading vehicle.

2.2 Traffic Models

In order to infer vehicular shock wave, it is crucial to have well-established baseline models. These baseline models are built based on empirical data and have proved to operate well in real life [FFG14]. These baseline models can be categorized in two categories, i.e., microscopic models and macroscopic models. Microscopic models describe the reactions of an individual vehicle with respect to its neighboor vehicle while macroscopic models focus on general overview of traffic status of certain road segment [FFG14]. Krauss car-following model will be introduced as an example of the microscopic models. Lighthill-Whitham-Richards (LWR) model will also be covered as an example of the macroscopic models.

2.2.1 Microscopic Models

Krauss car-following model is a time-discrete and space-continuous model [FFG14]. Due to its nature, the states of the model are evaluated for distinct timestamps while the space is not divided into distinct cells [KWG97, Kra98, FFG14]. The model is strongly related to *Cellular Automata* while it uses more realistic acceleration and deceleration values [NS92, FFG14]. Unlike *Cellular Automata*, Krauss car-following model supports spatial continuity, i.e., instead of a vehicle being positioned in a certain discrete cell, the vehicle can be positioned in any position of the continuous road segment. Additionally, a parameter for the length of the vehicle is provided in order to make the model reflect the real world traffic status.

The are three rules in Krauss car-following model, and described as below [FFG14]:

a) Velocity Update: A new velocity is determined by selecting the minimum value among maximum permitted velocity, and it is referred as desired velocity. Formally, this is written as follows:

$$v_i(t+1)^{(i)} = min[v_max, v_i(t) + a_i(v_i(t))\Delta t, v_{i,safe}(t)] \tag{2.1}$$

where i, v_{max}, $v_i(t)$, a_i, Δt, $v_{i,safe}(t)$ denote vehicle i, permitted maximum velocity, velocity of vehicle i at time t, acceleration value of vehicle i, time difference, and safety velocity at time t, respectively. The safety velocity is calculated by the following equation:

$$v_{i,safe}(t) = v_{i-1}(t) + b_i(v_i(t))\frac{g_i(t) - v_{i-1}(t)}{v_i(t) + b_i(v_i(t))} \tag{2.2}$$

where $g_i(t) = x_{i-1}(t) - x_i(t) - 1$, b_i, $v_{i-1}(t)$ denote the distance between vehicle i and $i-1$ while $x_{i-1}(t)$ is position of the preceding vehicle at time t, deceleration value, and the velocity of the preceding vehicle at time t, respectively.

b) Randomization: In order to mimic realistic human's driving behavior, randomization is crucial. The distance between vehicles cannot be identical, e.g., when a car slows down or speeds up the distance between the leading car and trailing car

changes. Randomization provides more realistic description of real traffic on highways by using probabilistic factor, and this is formulated as follows:

$$v_i(t+1)^{(2)} = \begin{cases} v_i(t+1)^{(1)} - b\Delta t, & \text{with } P_d(v(t)) \\ v_i(t+1)^{(1)}, & \text{else} \end{cases} \tag{2.3}$$

where

$$P_d(v(t)) := \begin{cases} p_s, & \text{if } v(t) = 0 \\ p_m, & \text{else} \end{cases} \tag{2.4}$$

with p_m the probability of dallying for moving vehicles and p_s the one for stopped vehicles, respectively. Note that $p_m << p_s$. In Equation2.3 describes vehicle i has to slow down in order to avoid collision by a deceleration value b for a timestamp with a velocity-dependent probability $P_d(v(t))$.

Finally, we have to ensure the updated maximum velocity is not negative. Formally,

$$v_i(t+1) = max\Big[0, v_i(t+1)^{(2)}\Big] \tag{2.5}$$

Therefore, the velocity of vehicle i is not only calculated in deterministic fashion but also in probabilistic fashion using the dally factor.

c) Position Update: After updating velocity of vehicle i, it is important to update the position of vehicle i. The new position vehicle i is computed by adding the old position to the multiplication of the constant time difference and the new velocity. This is calculated by the following formula:

$$x_i(t+1) = x_i(t) + v_i(t+1)\Delta t \tag{2.6}$$

The aforementioned three rules are applied for each timestamp.

2.2.2 Macroscopic Models

Macroscopic models were originated from fluid dynamics under the assumption that traffic streams as a whole are comparable to fluid streams. Since the assumption of the macroscopic models is to view the traffic streams as a whole, they give us the general overview of the traffic status. Lighthill-Whitham-Richards (LWR) model is an example of a Macroscopic Traffic Model. It was originally proposed by Lighthill and Whitham [LW55], and separately by Richards [Ric]. The model is based on continuity equation as follows:

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho V)}{\partial x} = 0 \tag{2.7}$$

where t denotes time, ρ denotes vehicular density, x denotes position, and V denotes flow velocity. The evolution of density over time is given by the spatial evolution of traffic flow. This only applies to homogeneous road segments where inputs and outputs are only available at the borders of the observed segment [FFG14]. A static relationship between the traffic flow $Q(x,t)$ and vehicular density $\rho(x,t)$ is stated by the model as follows:

$$Q(x,t) = \hat{Q}(\rho(x,t)) \tag{2.8}$$

By substituting the flow velocity of Equation 2.7 with the traffic flow of Equation 2.8, we obtain the model of equation of the LWR model, given as below:

$$\frac{\partial \rho}{\partial t} + \frac{d\hat{Q}(\rho)}{d\rho}\frac{\partial \rho}{\partial x} = 0 \tag{2.9}$$

Equation 2.9 is the so called *Transport Equation*, and if we replace the wave function $\rho(x,t)$ with $\rho_0(x - \tilde{c}t, t)$ in Equation 2.9, the equation transforms as follows:

$$v_t + C \times v_x = 0 \tag{2.10}$$

By Equation 2.10, the propagation speed of the shock wave is given as a result of the LWR model [RKP10, FFG14]. The propagation speed of the shock wave is given as

below:

$$\tilde{c} = -\frac{d\hat{Q}(\rho)}{d\rho} \tag{2.11}$$

For more information regarding the LWR model, we direct the readers to the following papers [LW55, Ric].

Finally, Figure 2.3 and 2.4 demonstrate the typical triangular properties of the simplified LWR model [FFG14].

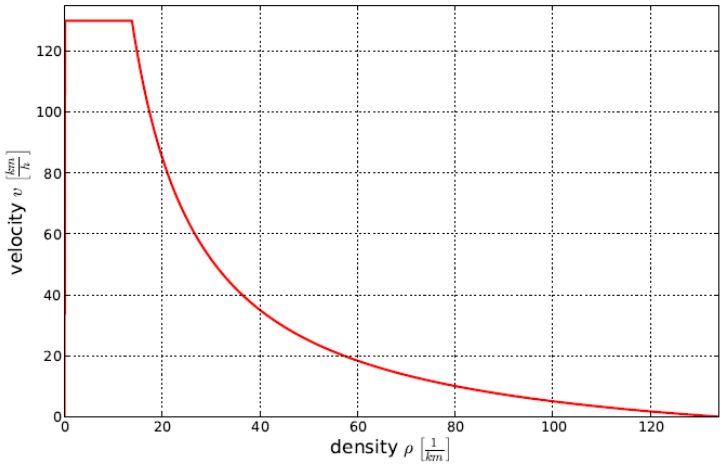

Figure 2.3: Velocity Density Chart

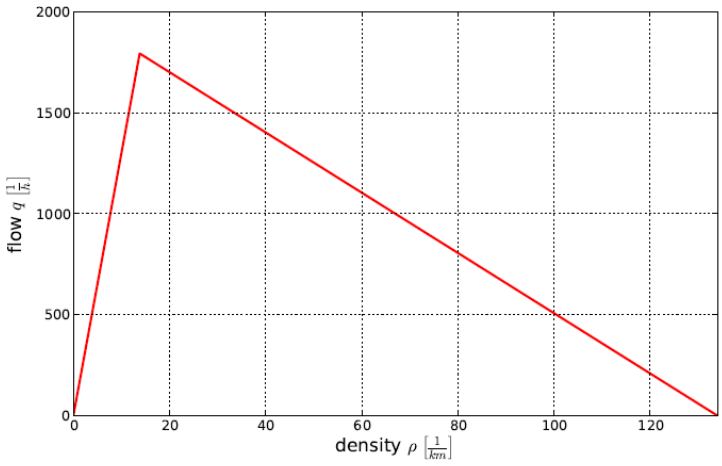

Figure 2.4: Flow Density

CHAPTER 3

Exploratory Data Analysis

3.1 CalTrans PeMS Data Set

3.1.1 Structure of the data sets

In this chapter, the structures of the California Department of Transportation (CalTrans) Performance Measurement System (PeMS) data sets [Tra15] will be introduced. The ideal data set to track drivers' driving patterns and the influence of drivers on shock wave would be the one tracks every driver at a certain location every second. Although there are some data sets that track all drivers who pass a certain location every second, the volume of the data set is very small, i.e., dataset only tracks traffic a single day or several days. On the contrary, the CalTrans PeMS data sets have road traffic tracking granularity of 30 seconds, 5 minutes, an hour, and a day. However, the data sets have been collected for more than a decade [Tra15].

As mentioned in the above paragraph, in order to track shock waves occurred on freeways, it is essential to have the traffic tracking granularity of the data set low. Therefore, the station raw data sets (See Table 3.1 for the structure of the data set) that have road traffic monitored at each station every 30 seconds are the best ones we can use for shock wave (traffic shock). However, the station raw data sets do not have the average speed of the vehicles that pass each station (loop detector) every 30 seconds. The station 5-min data sets (See Table 3.2 for the structure of the data set) have the average speed for each lane.

Attributes	Description
Timestamp	Sample period as reported by the field element as MM/DD/YYYY HH24:MI:SS. This indicates the beginning of the reporting period.
Station ID	Unique station identifier. Use this value to cross-reference with Metadata files.
Freeway	Freeway on which the station is located.
Lane "N" Flow	Number of vehicles that passed over the detector during the sample period (identified by the timestamp)
Lane "N" Occupancy	Measured occupancy for lane N expressed in percentage. Empty if the detector does not report occupancy. N ranges from 1 to the number of lanes at the location. Lanes are numbered from inside (closest to the median) to outside. Units are expressed in percentage.

Table 3.1: PeMS Station Raw Data Set Features

Attributes	Description
Timestamp	Sample period as reported by the field element as MM/DD/YYYY HH24:MI:SS. This indicates the beginning of the reporting period.
Station ID	Unique station identifier. Use this value to cross-reference with Metadata files.
Freeway	Freeway on which the station is located.
Direction of Travel	Direction of travel measured by the station.
Station Type	Type of station. OR = On ramp, FR = Off ramp, ML = Main line, HV = HOV lane, FF = Freeway-to-freeway connector.
Samples	Total number of samples recorded at this station and timestamp. This is the summation of all Lane "N" Samples.
Observed	Percentage of individual lane points at this location that were observed. An individual lane point is a single set of measurements (e.g., flow and occupancy) for a single lane at a single time.
Total Flow	Sum of raw flows over all lanes during the 5-minute time period. Note that this basic 5-minute rollup normalizes flow by the number of good samples received from the controller.
Average Occupancy	Average occupancy across all lanes over the 5-minute period expressed in percentage. Computed as the average of the occupancies of all individual lanes during the 5-minute period.
Average Speed	Flow-weighted average speed over the 5-minute period across all lanes. The flow-weighted average speed is the lane-by-lane average of speeds weighted by their flow.

Lane "N" Samples	Number of good samples received for lane N. N ranges from 1 to the number of lanes at the location. Lanes are numbered from inside (closest to the median) to outside. Units are mph.
Lane "N" Flow	Total flow for lane N over the 5-minute period normalized by the number of good samples. Lanes are numbered from inside (closest to the median) to outside. Units are vehicles/5-min.
Lane "N" Average Occupancy	Average occupancy for lane N over the 5-minute period expressed in percentage. N ranges from 1 to the number of lanes at the location. Lanes are numbered from inside (closest to the median) to outside.
Lane "N" Speed	Flow-weighted average of lane N speeds, which are obtained from raw data by g-factor analysis. N ranges from 1 to the number of lanes. Lanes are numbered from inside (closest to the median) to outside. Units are mph.
Lane "N" Observed	1 indicates observed data, 0 indicates missing data which may be imputed (or "filled in". Lanes are numbered from inside (closest to the median) to outside.

Table 3.2: PeMS Station 5-min Data Set Features

3.1.2 Station Status

The PeMS data sets that are used in this dissertation are the 30-sec data sets and 5-min datasets that were collected from January 1, 2016 to June 30, 2016. Specifically, District 7 (Los Angeles Area) traffic information was used for the data analysis and shock wave inference. Although in the data sets there were many stations (loop detectors) listed, many of the stations were not actually collecting traffic status of the assigned freeway segments. In order to infer vehicular shock wave propagation speed, accurate traffic information of each freeway in District 7 is crucial. Hence, all the stations that do not collect more than a single lane traffic information were excluded. In District 7, all the freeways have more than a single lane, and the stations that do not collect traffic status of more than one lane are considered malfunctioning.

In this section, the total number of stations of each freeway in both directions, i.e., south and north, or west and east, and the total number of functioning stations are listed as follows. Additionally, all functioning stations of each freeway are displayed on a map in order to visualize the distance between each station.

a) I-405 South & North

Directions	total # of stations	total # of functioning stations
South	255	74
North	227	67

Table 3.3: Total number of stations and total number functioning stations on I-405

b) I-5 South & North

c) I-105 East & West

d) I-10 East & West

e) I-110 South & North

Directions	total # of stations	total # of functioning stations
South	146	60
North	109	45

Table 3.4: Total number of stations and total number functioning stations on I-5

Directions	total # of stations	total # of functioning stations
East	82	18
West	90	28

Table 3.5: Total number of stations and total number functioning stations on I-105

Directions	total # of stations	total # of functioning stations
East	269	85
West	286	97

Table 3.6: Total number of stations and total number functioning stations on I-10

Directions	total # of stations	total # of functioning stations
South	83	31
North	83	32

Table 3.7: Total number of stations and total number functioning stations on I-110

f) I-710 South & North

Directions	total # of stations	total # of functioning stations
South	49	11
North	51	16

Table 3.8: Total number of stations and total number functioning stations on I-710

g) I-605 South & North

Directions	total # of stations	total # of functioning stations
South	126	38
North	122	38

Table 3.9: Total number of stations and total number functioning stations on I-605

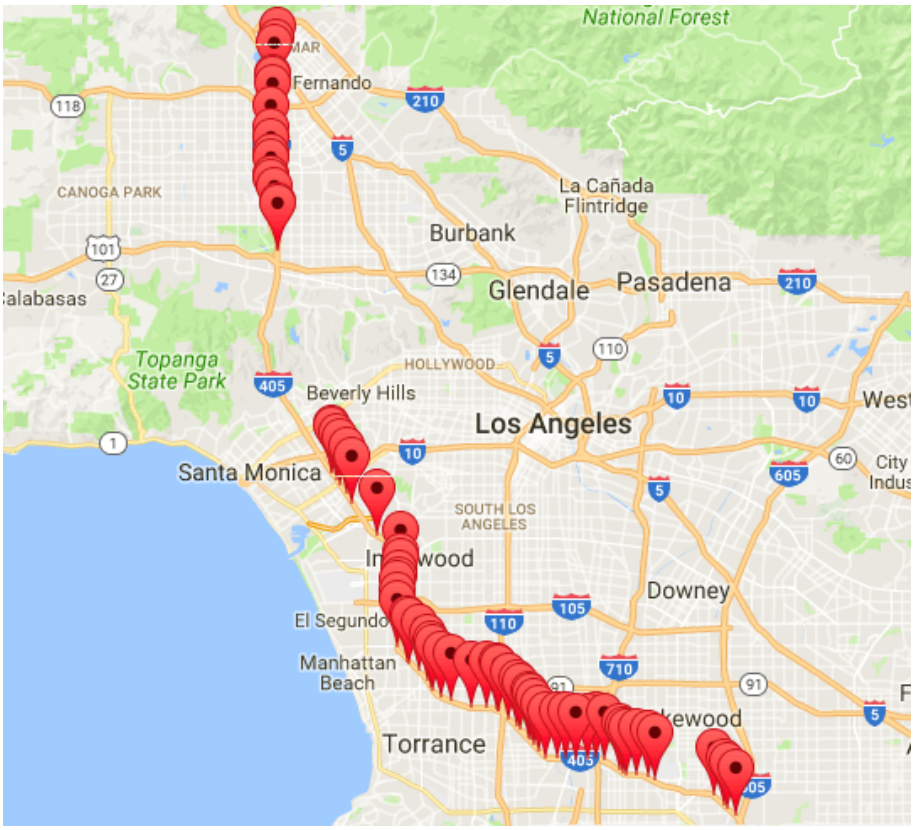

Figure 3.1: The locations of functioning stations on I-405 South

3.2 Vehicular Shock Wave Propagation Speed

In order to see if there are vehicular shock waves propagating on freeways, we plotted a heatmap of each freeway. In Figure 3.16, we can see that there is shock

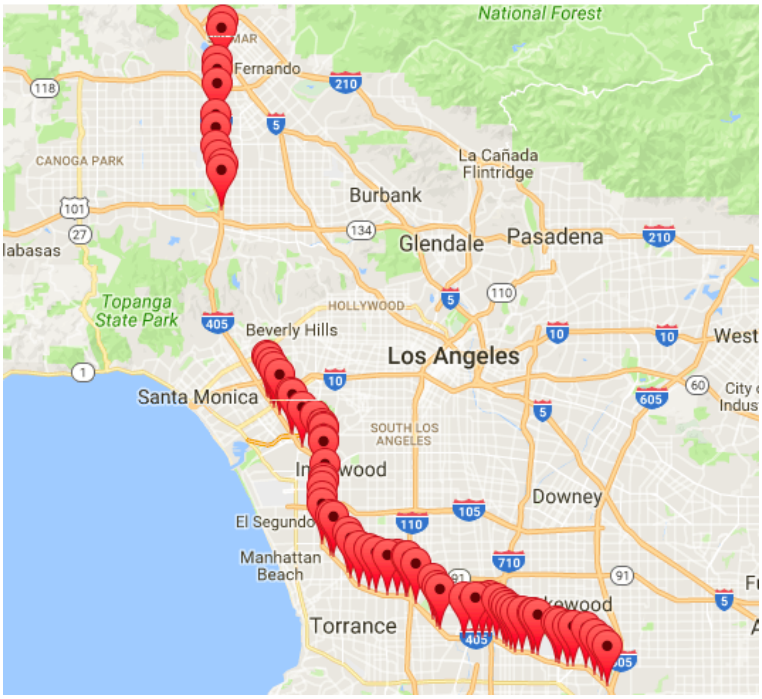

Figure 3.2: The locations of functioning stations on I-405 North

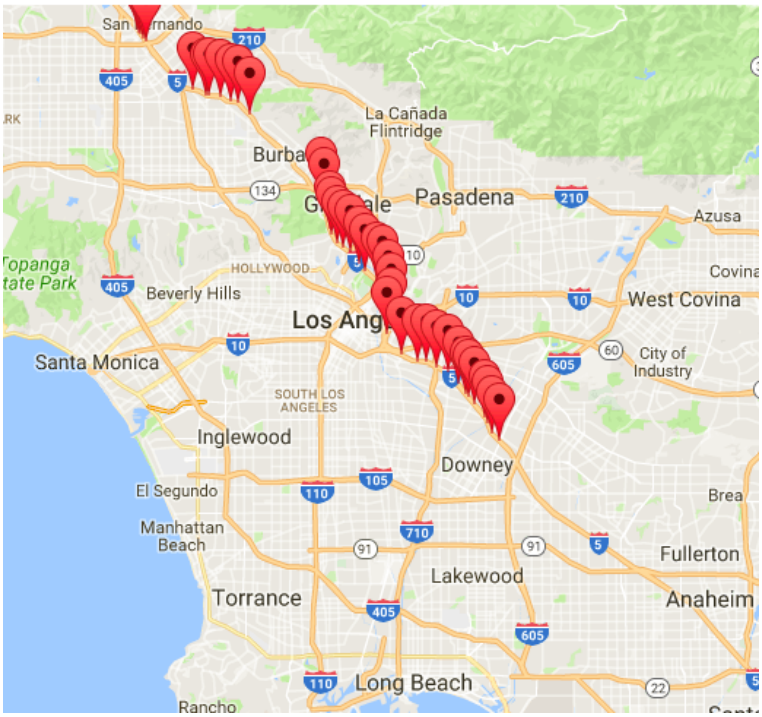

Figure 3.3: The locations of functioning stations on I-5 South

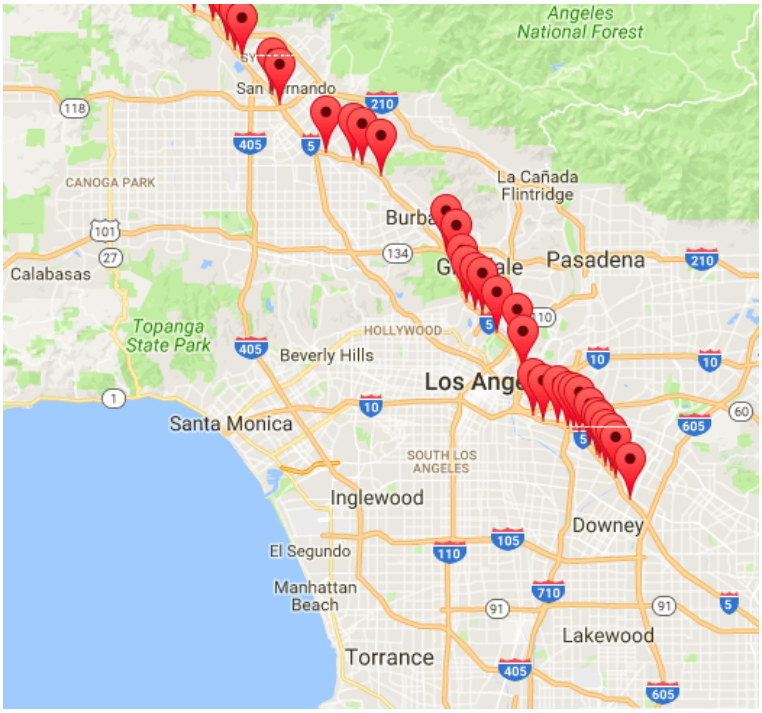

Figure 3.4: The locations of functioning stations on I-5 North

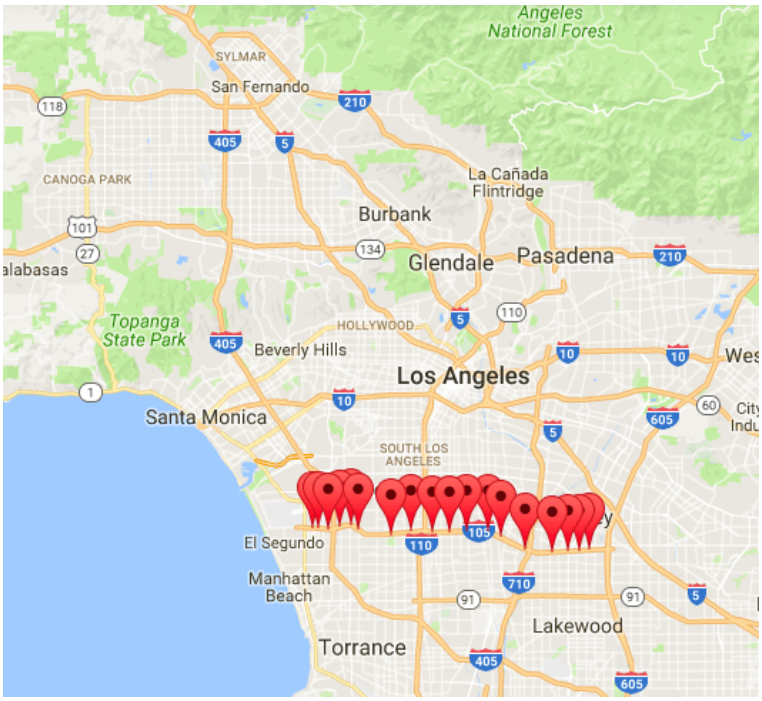

Figure 3.5: The locations of functioning stations on I-105 East

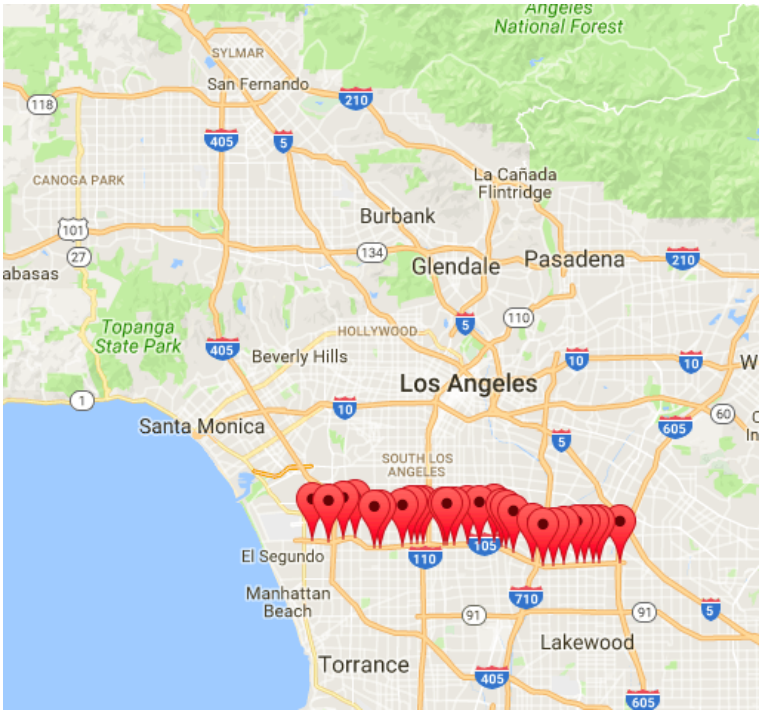

Figure 3.6: The locations of functioning stations on I-105 West

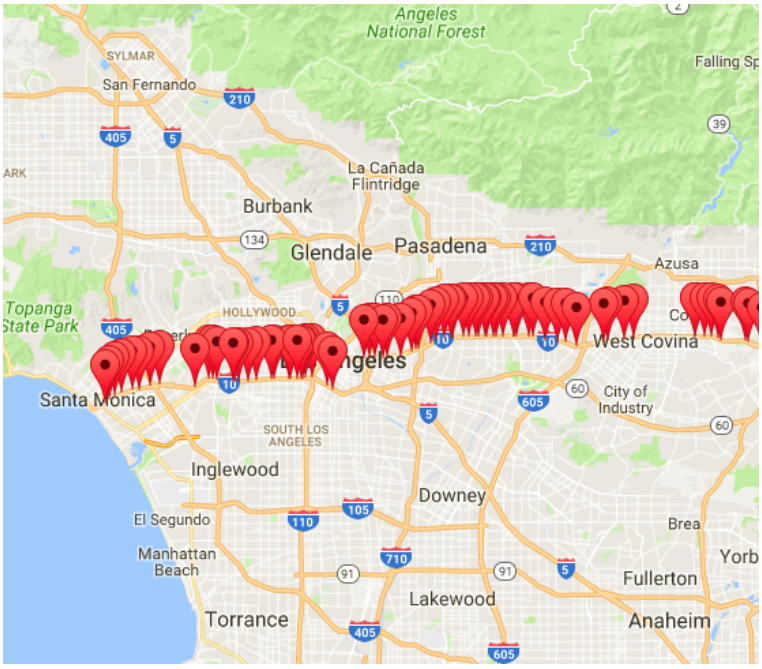

Figure 3.7: The locations of functioning stations on I-10 East

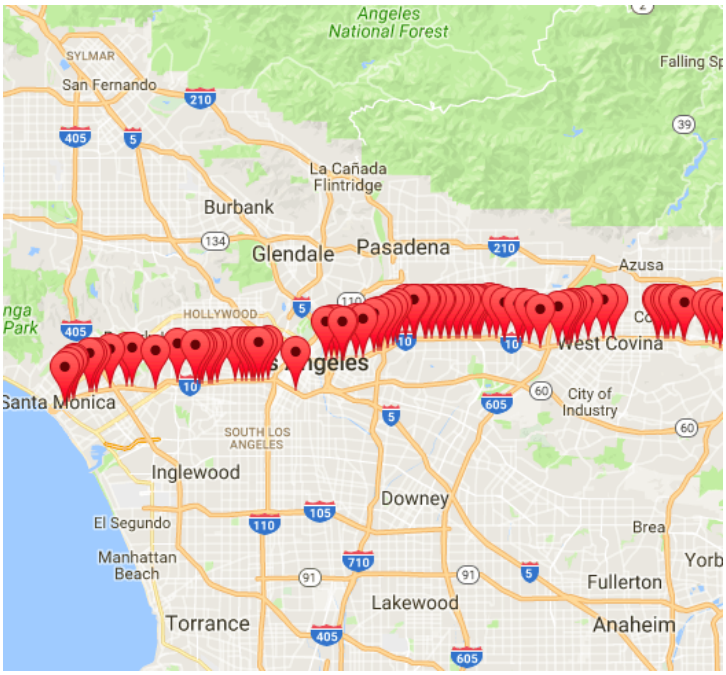

Figure 3.8: The locations of functioning stations on I-10 West

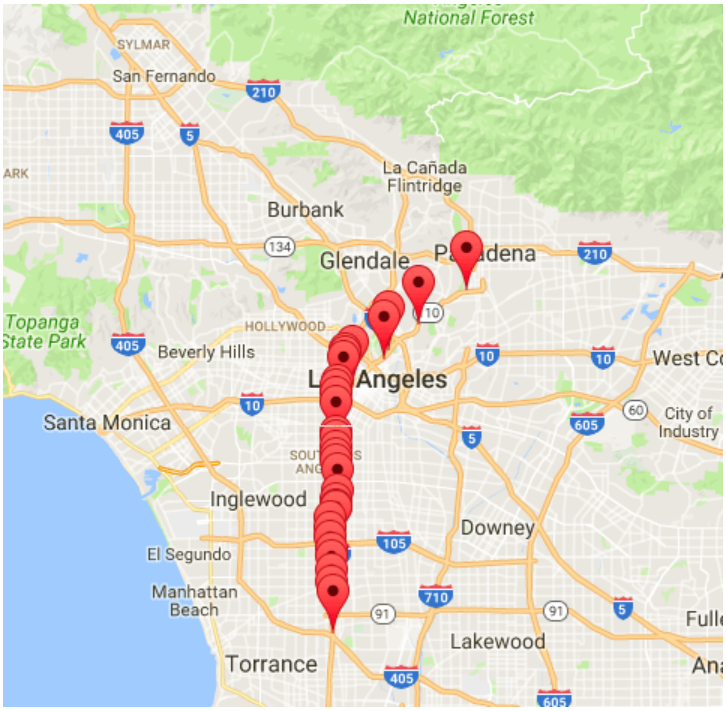

Figure 3.9: The locations of functioning stations on I-110 South

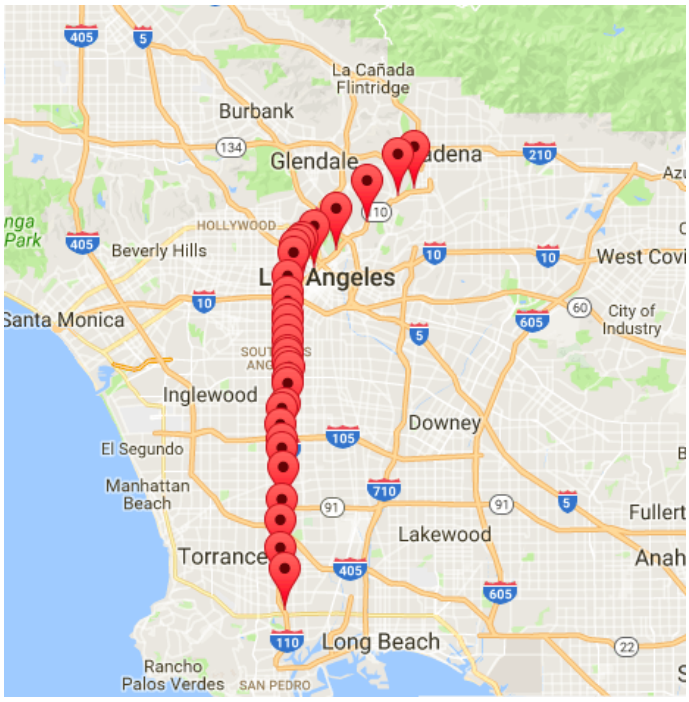

Figure 3.10: The locations of functioning stations on I-110 North

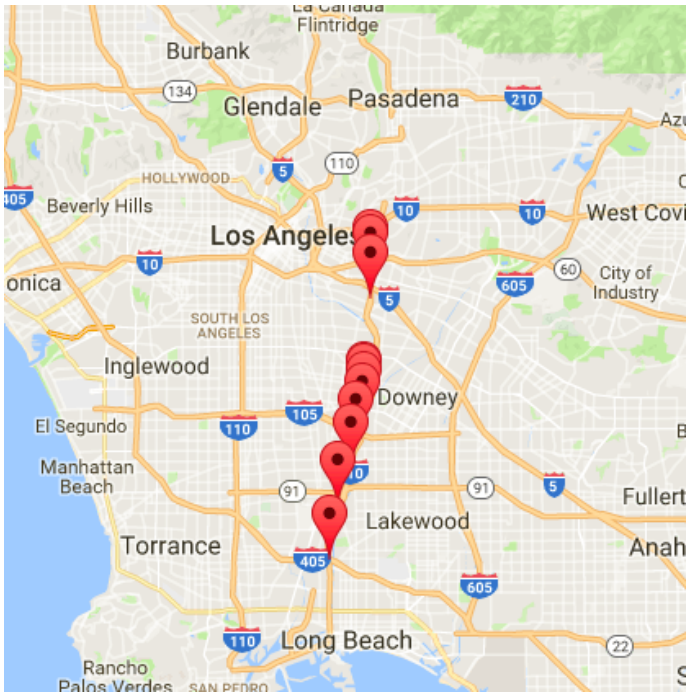

Figure 3.11: The locations of functioning stations on I-710 South

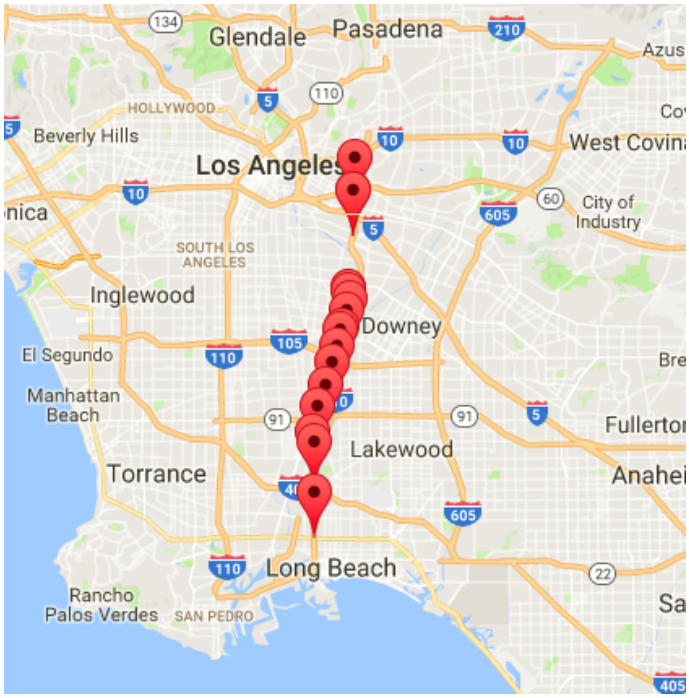

Figure 3.12: The locations of functioning stations on I-710 North

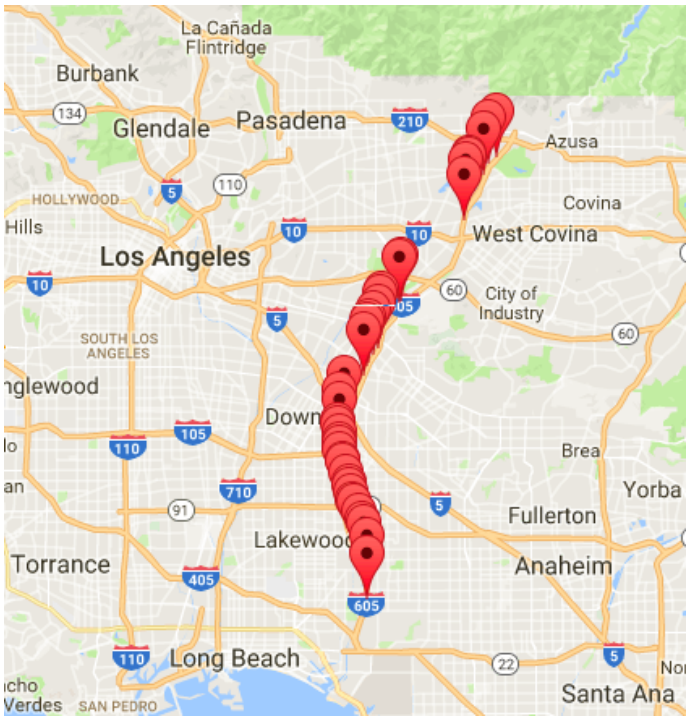

Figure 3.13: The locations of functioning stations on I-605 South

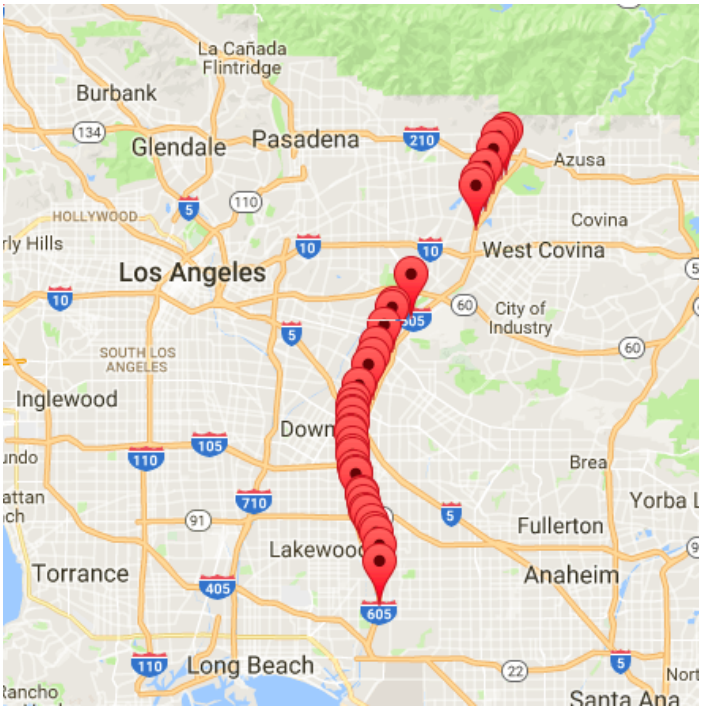

Figure 3.14: The locations of functioning stations on I-605 North

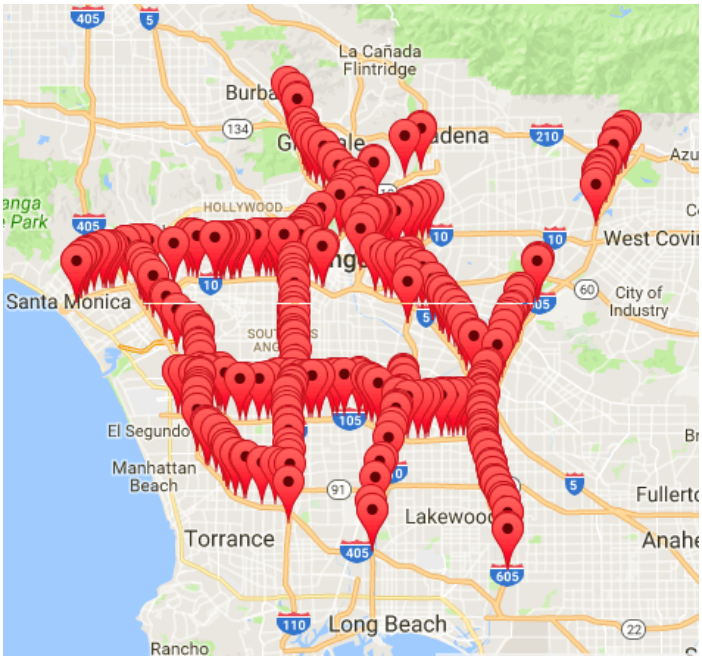

Figure 3.15: The locations of functioning stations in LA city

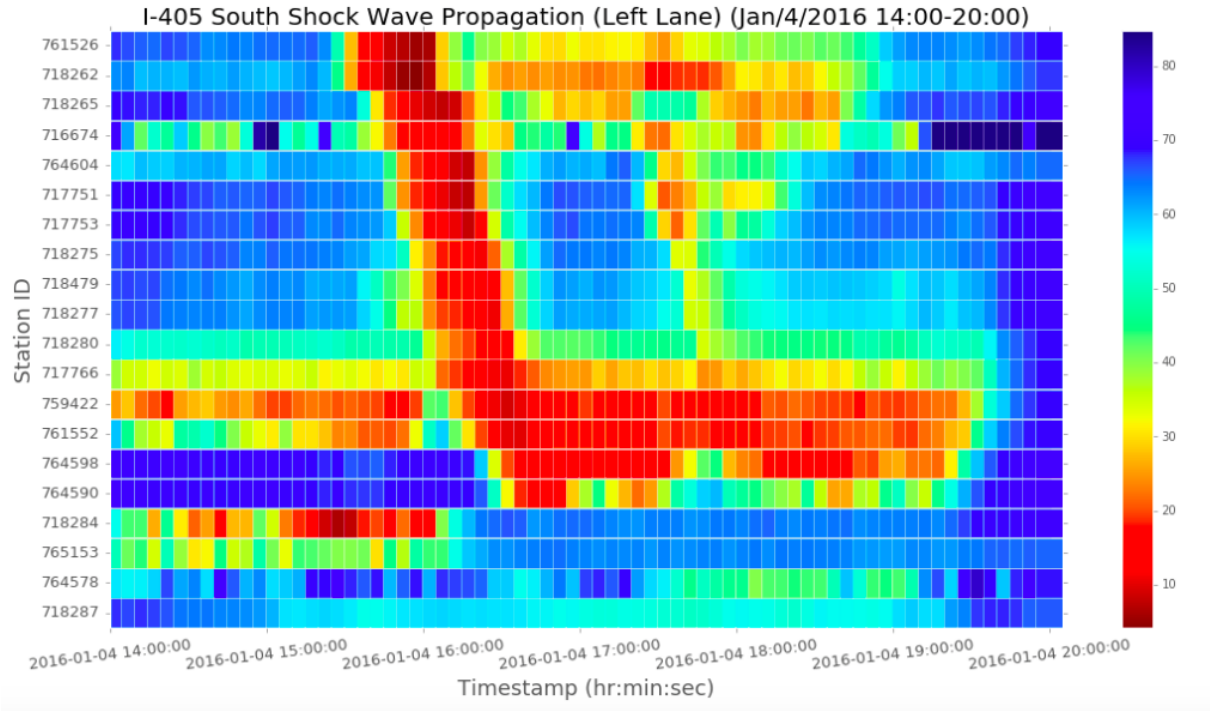

Figure 3.16: Vehicular shock wave propagation speed on left lane of each station of I-405 South (Jan/4/2016 2:00-8:00 pm)

wave propagating from upstream to downstream on the left lane of freeway I-405 South during afternoon rush hour (2:00-8:00 pm). The x-axis is the timestamp and each tick is 5-min, y-axis is the list of active station IDs where the higher y value is a station located at a lower latitude, i.e., station ID with higher y-axis value is more towards south than the one with lower y-axis value. The colorbar on the right side of each graph indicates the average speed of vehicles passing each station in *mi/h*. We can see that at at station a shock wave formed and started to propagate. There are some regions of the freeway that has constant bottleneck, i.e., there are shock waves with very low value of slope propagating very slowly in those regions of the freeway. In Figure 3.16, there are constant bottlenecks from station 717766 to station 759422 from 2:00 pm to 8:00 pm, and we could not find meaningful insights of vehicular wave propagation speed in those road segments.

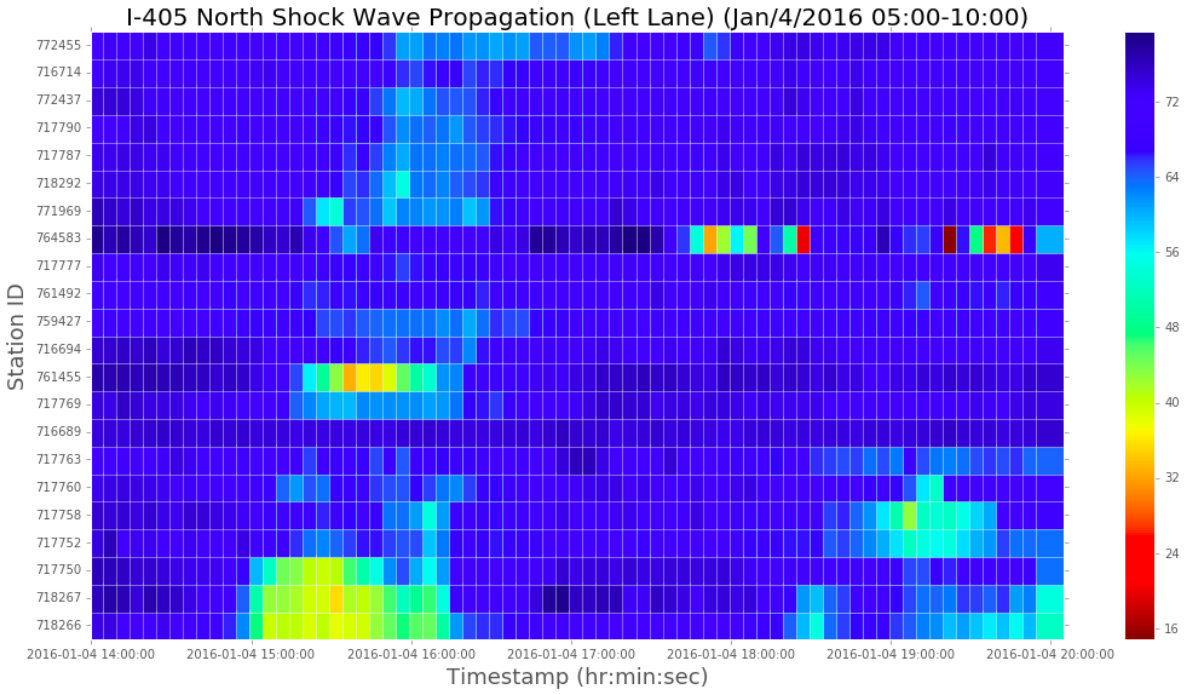

Figure 3.17: Vehicular shock wave propagation speed on left lane of each station of I-405 South (Jan/4/2016 5:00-10:00 am)

Unlike afternoon rush hours, we couldn't find shock wave occurring during morning rush hours (5:00-10:00 am) on any of the lanes of I-405 South (Please refer to Figure 3.17).

After analyzing six months of data of District 7 PeMS datasets, we discovered that freeway such as I-5 has constant traffic bottlenecks during the day at more than half of all stations (See Figure 3.18). Hence, we removed freeways that have the same constant bottlenecks or no bottleneck from data analysis on shock wave propagation. Some research papers state that the shockwave propagates around -9.32 *mi/h* (-15 *km/h*) [RKP10, FFG14] on average on U.S. highways. We verified if the aforementioned value is applicable in District 7 freeways, since the shock wave propagation speed is crucial for inferring the occurrences of the shock waves.

We utilize multiple linear regression on the heatmap (See Figure 3.17) to obtain

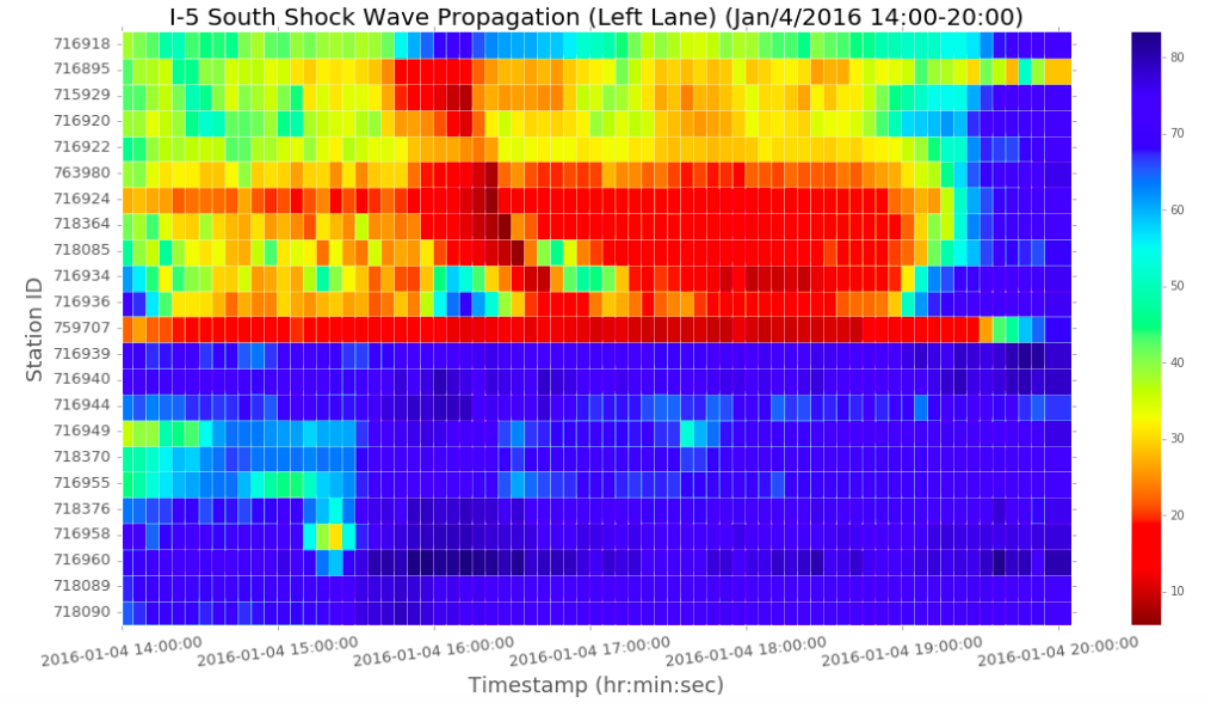

Figure 3.18: Vehicular shock wave propagation speed on left lane of each station of I-5 South (Jan/4/2016 5:00-10:00 am)

the slope (gradient in Section 2.2.2).

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij} \tag{3.1}$$

where y_i is the location of the ith station, x_{ij} is the timestamp, i ranges from 1 to the number of stations, and j ranges from starting time of the shock wave till the end of the shock wave. Although using R^2 as an evaluation metric in linear regression should be avoided, if the variables are timeseries [LJ99], we utilized it with cross-validation in order to obtain the slope. Details on cross-validation are introduced in Section 3.2.1. The average slope (shock wave propagation speed) of freeways from our analysis is -8.51 mi/h (-13.7 km/h) which is considerably close to the previous studies [RKP10, FFG14].

3.2.1 Cross-Validation

Cross-validation is a popular technique that is used for validating the results of a statistical analysis to see how it will generalize to an independent data set [Gei93, Koh95, DK82]. In a prediction problem, the statistical model is given a known dataset for training, and an unkown dataset for testing. In cross-validation, testing the model is performed during the training phase. The purpose of the cross-validation is to prevent overfitting. In this section and Section 3.3, we utilize k-fold cross-validation.

In k-fold cross-validation, the sample data set is divided into k chunks. During the first iteration the first chunk becomes the test data set and the other chunks are the training data set. In the second iteration, the second chunk is the test data set and the other chunks are the training data set. This process repeats until the k-th iteration. In Figure 3.19, test data set is colored as orange and the rest as white. Additionally, k is set to 5 in Figure 3.19, however, it is very common to set $k = 10$.

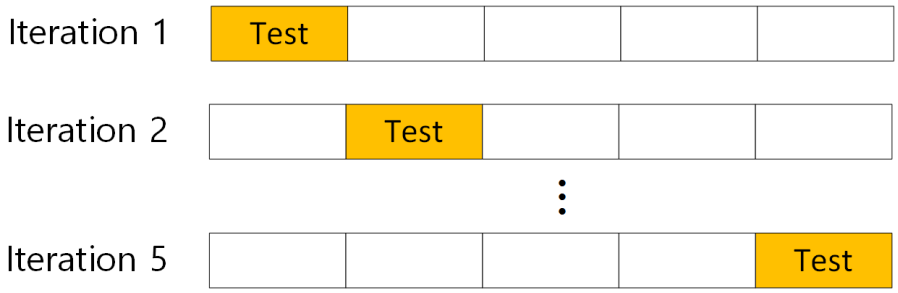

Figure 3.19: k-fold cross-validation

During each iteration, evaluation metric such as $RMSE, MSE, ME$, and R^2 can be utilized in Linear Regression model.

3.3 Selecting the Right Feature to Infer Shock Wave Occurrence

As mentioned in Section 3.1.1, the finer the granularity the better to see the influence of the shock wave on traffic status. However, the since the 30-second data set only has occupancy and flow value collected, it is hard to check whether shock wave occurred or not using those two predictors. This is because the heatmap (See Figure 3.16) uses speed to check the propagation of the shock wave.

We initially thought inferring speed from either flow or occupancy would address the issue of not having speed in the finer granularity data set. Since there was a polynomial characteristic in the correlation between speed and occupancy plot (See Figure 3.20), we decided to use Polynomial Regression to infer speed from occupancy (See Figure 3.21). We used cross-validation and regularization in order to avoid overfitting. However, when we actually predicted the speed from occupancy, the results didn't look very reliable. It had many spikes which implies the accuracy of the prediction is not very good (See Figure 3.22).

Therefore, instead of inferring speed from occupancy, we investigated if the occupancy itself could be used to infer shock wave occurrence. We again plotted a heatmap of I-405 South using occupancy instead of speed. In Figure 3.23, we could see the heatmap pattern is exactly the same as the heatmap generated from speed data (See Figure 3.16). Hence, we can conclude under certain threshold of occupancy, shock wave occurs and propagates if the occupancy of a lane at the previous timestamp is high. We will expand what we have discovered in Chapter 4.

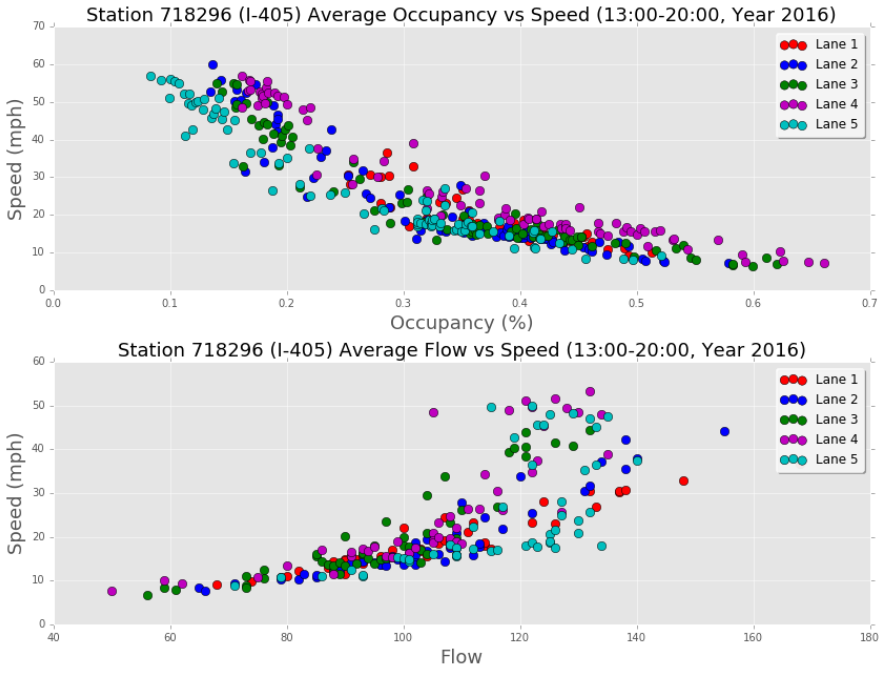

Figure 3.20: Speed vs Occupancy plot and Speed vs Flow plot

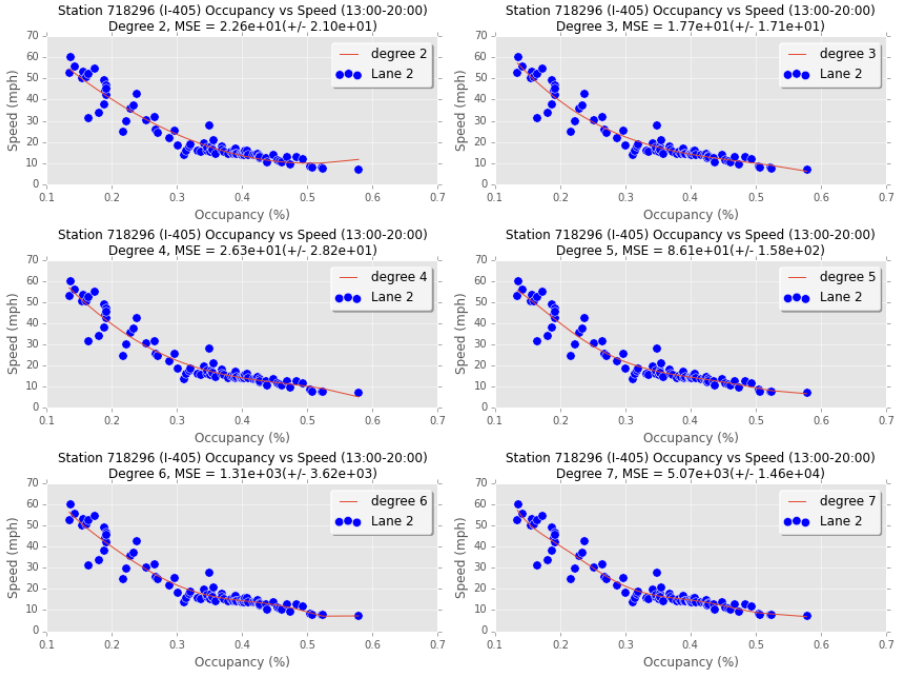

Figure 3.21: Inferring speed using Polynomial Regression using occupancy of the left lane of Station 718296 (I-405) from 1:00-8:00 pm on weekdays

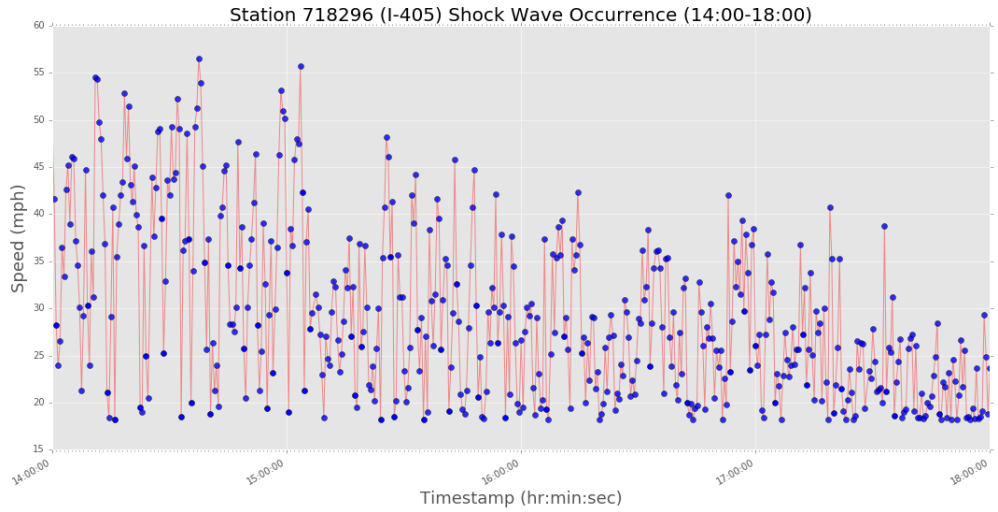

Figure 3.22: Predicted speed of the left lane of Station 718296 (I-405) from 2:00-8:00 pm on weekdays

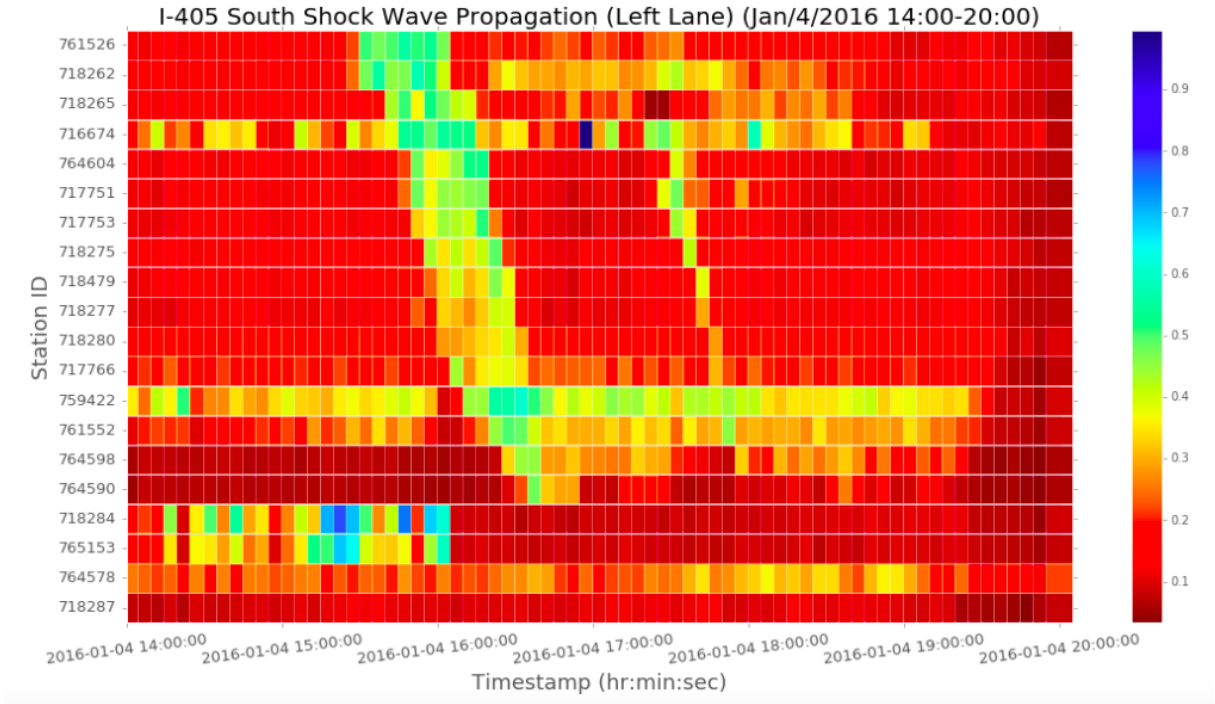

Figure 3.23: Vehicular shock wave propagation speed on left lane of each station of I-405 South (Jan/4/2016 2:00-8:00 am) from Occupancy Data

3.4 Summary

In this chapter, we analyzed the PeMS data set and plotted the occurrences of the vehicular shock wave on a heatmap. In order to calculate the shock wave propagation speed from the heatmap, linear regression and cross validation were adopted. We also attempted to infer speed from occupancy since the 30-second data set only has occupancy and flow. However, we realized the prediction was inaccurate and discovered that occupancy can be used for shock wave prediction since the heatmap from speed data and the heatmap from occupancy data had the same distribution.

CHAPTER 4

Hidden Markov Model

In this chapter, we discuss how to infer the occurrence of a shock wave of a lane on a freeway. Additionally, inferring the probability of lane changes of vehicles from one lane to another based on the macroscopic traffic information is introduced. There are a few studies that utilize Dynamic Bayesian Network (DBN) or Hidden Markov Model (HMM) for predicting traffic information [KM00, PN11] or raffic conditions in arterial networks using probe data [AB12]. The overlapping assumptions of these studies are 1) a hidden state at each station that takes Q possible values (free flow or congestion), 2) the observed velocity is a noisy representation of the current state. Additionally, the distributions of flow and density are assumed to be Gaussian as many papers that successfully applied machine learning technique to predict short-term traffic flow of PeMS data assumed [KM00, PN11].

4.1 Shock Wave Inference

The underlying assumption of the Hidden Markov Model (HMM) model for shock wave prediction is the traffic status of the current timestamp at a certain station will affect the traffic status of the station at the next timestamp. In Figure 4.1, Round nodes represent discrete random variables with binomial distribution, i.e., shock wave or non-shock wave of a lane at a certain station. On the other hand, rectangular nodes are the hidden state variable nodes and represent continuous random variable with Gaussian distributions, i.e., occupancy (Please refer to Section 3.3). We can see that

the traffic status of a lane $L1$ at time t is affecting the traffic status of not only the same lane but also the neighboring lane $L2$ at time $t+1$.

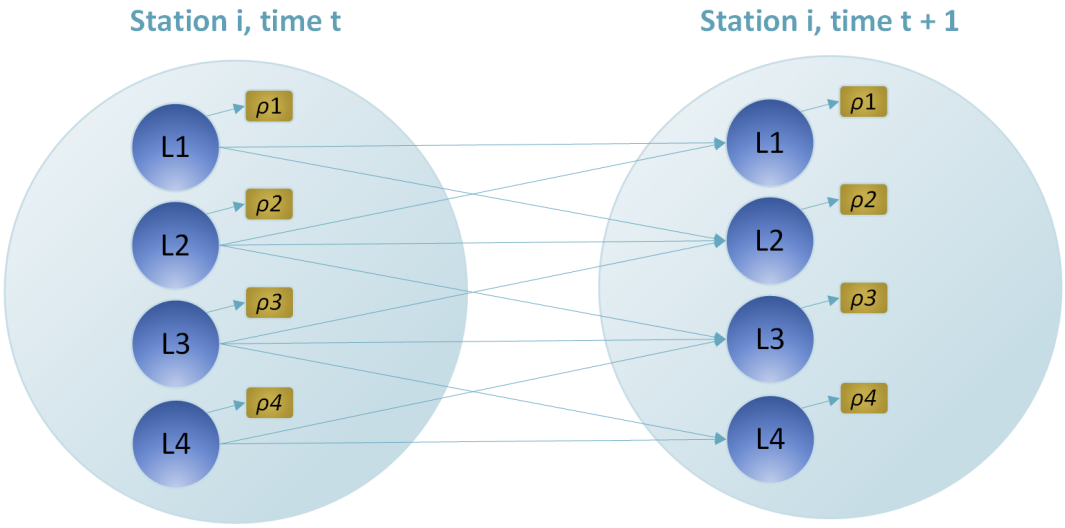

Figure 4.1: Hidden Markov Model represented as a Dynamic Bayesian Network (One station)

Another assumption to the model is the traffic status of the next station towards downstream is affecting the traffic status of the current station at the next timestamp. In Figure 4.2, lane $L1$ of station $i+1$ at time t is affecting lane $L1$ of station i at time $t+1$.

These assumptions can be described by a formula as follows: Let's denote the hidden state variable $x_{l,m,t} \in S = \{s_1, ..., s_Q\}$ where l is the number of lane, m is the station number, t is the timestamp, Q is the number of possible states. Also, let's assume that the hidden process of $x_t = (x_{1,m,t}, ..., x_{L,M,t}) \in S^{M \otimes L}$ is Markovian and its transition probability is decomposable.

$$\begin{aligned} Pr(x_{t+1}|x_t) &= Pr(x_{l,m,t+1}|x_t) \\ &= Pr(x_{l,m,t+1}|x_{l-1,m,t}, x_{l,m,t}, x_{l+1,m,t}, x_{l-1,m+1,t}, x_{l,m+1,t}, x_{l+1,m+1,t}) \end{aligned} \tag{4.1}$$

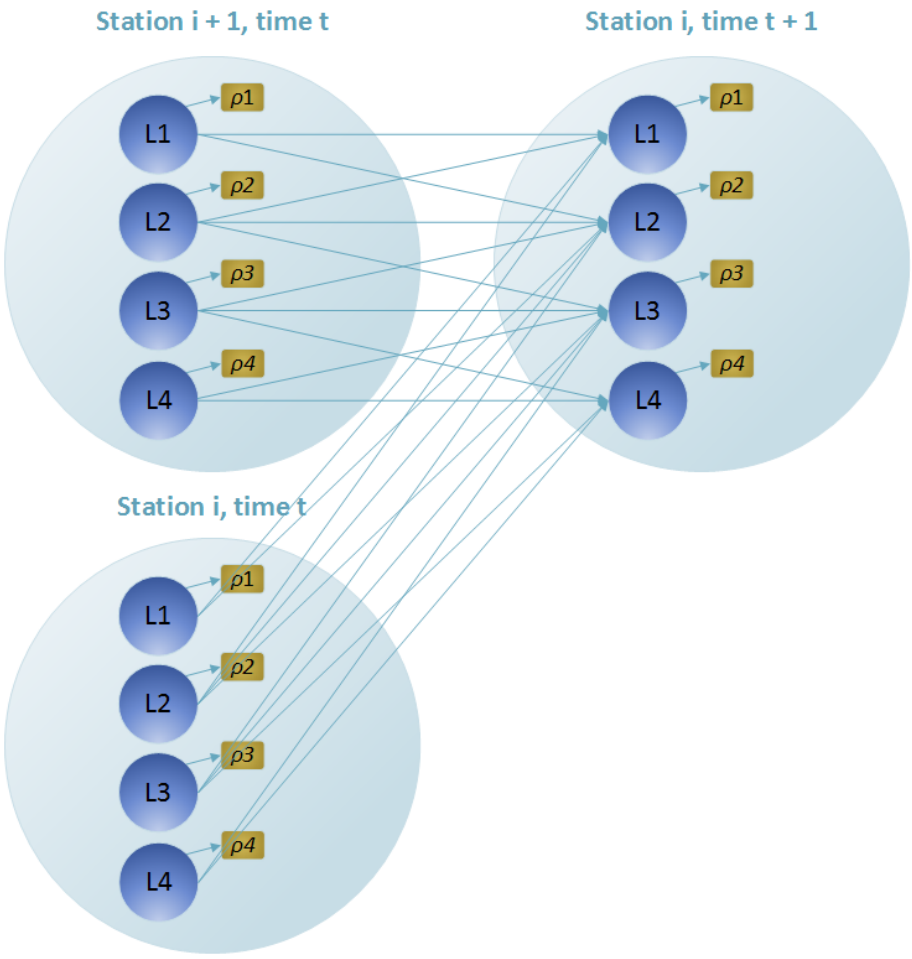

Figure 4.2: Hidden Markov Model represented as a Dynamic Bayesian Network (Multi stations)

4.2 Lane Change Prediction

4.2.1 The Model

The underlying assumption of the Hidden Markov Model (HMM) model is the traffic status of a lane, at the next timestamp, at the same station, is affected by itself and the neighboring lanes of the previous timestamp. Formally, at time t, at station i, on lane j affects the traffic status at time $t+1$, at station i, on lane $j-1$, j, and $j+1$. Unlike previous studies on predicting short-term traffic information, such

as average velocity of vehicles at a certain station, there is a hidden state at each station that has Q possible values, i.e., stay or switch lane. The value of Q differs from lane to lane. If a lane has one neighboring lane, such as carpool lane, left lane, or lane on the ramp that is merged to the freeway, the value of Q is two since it has two options, i.e., stay or switch to the neighboring left or right lane. Similarly, assume there are three lanes on a freeway, the value of Q is three since it has three options, i.e., stay, switch to the left lane, or switch to the right lane (Please refer to Figure 4.1).

Let us denote the hidden state variable $x_{l,m,t} \in S = \{s_1, ..., s_Q\}$ where l is the number of lane, m is the number of stations, t is the timestamp, and Q is the number of possible states, respectively. In the aforementioned example, Q is either two or three. Let us assume that the hidden process of $x_t = (x_{1,m,t}, ..., x_{L,M,t}) \in S^{M \otimes L}$ is Markovian and its transition probability is decomposable. For a carpool lane or left lane, the transition probability is as follows:

$$Pr(x_{t+1}|x_t) = Pr(x_{l,m,t+1}|x_t) = Pr(x_{l,m,t+1}|x_{l,m,t}, x_{l+1,m,t}) \tag{4.2}$$

where $l = 1$. For merging lanes on the ramp, the transition probability is given as below:

$$Pr(x_{t+1}|x_t) = Pr(x_{l,m,t+1}|x_t) = Pr(x_{l,m,t+1}|x_{l-1,m,t}, x_{l,m,t}) \tag{4.3}$$

where $l = N$ and N the number of total lanes on a freeway at a certain station. Finally, for other lanes on a freeway at a certain station, the transition probability is given as follows:

$$Pr(x_{t+1}|x_t) = Pr(x_{l,m,t+1}|x_t) = Pr(x_{l,m,t+1}|x_{l-1,m,t}, x_{l,m,t}, x_{l+1,m,t}) \tag{4.4}$$

where $1 < l < N$.

Since the finest granularity of the PeMS datasets is the one that has aggregated the traffic information collected every 30 seconds, it is important to have an assumption

on the distribution of flow and density of each lane of the monitored freeway. Flow and density of each lane are collected every 30 seconds and can be assumed as Gaussian since both of them aggregated.

4.2.2 Baum-Welch Algorithm

Baum-Welch algorithm has been very successful in the speech recognition, genomic sequence modeling field. Leonard E. Baum proposed Hidden Markov Models and Baum-Welch algorithm in the late 1960s [BE, Rab]. The Baum-Welch algorithm utilizes Expectation-Maximization (EM) algorithm in order to find the Maximum Likelihood Estimate (MLE) of the parameters of the Hidden Markov Model (HMM). This is the most important step to predict the average probability of lane change per lane since PeMS data sets only provide aggregated traffic information (Please refer to table3.1).

Equations 4.2, 4.3, 4.4 are called the *Markov Property* of the chain [Rab89]. Due to the *Markov Property* the complete probability distribution of the states of a Markov chain is defined by the initial distribution $\pi_i = Pr(x_{l,1} = i)$ and the state transition probability which is given as follows:

$$a_{ij} = Pr(x_{l,t+1} = j | x_{l,t} = i), \quad 1 \leq i, j \leq Q \tag{4.5}$$

where Q is the number of possible states.

Let's denote $\boldsymbol{\pi} = \{\pi_i\}$ and $\boldsymbol{A} = \{a_{ij}\}$. Additionally, let's assume the transition probabilities are time-independent. Since the observation values are the density or occupancy values of a lane, the Markov Model has continuous observations. A Hidden Markov Model that has continuous observations is called Continuous Density Hidden Markov Model (CDHMM). Unlike HMM with categorical observations, the model parameter $\boldsymbol{B}$ of CDHMM is not described as matrix of point probabilities, but rather as a complete Probabilistic Density Function (PDF) over the continuous observation

space for each state. Therefore, the probability of a certain observation at time t for state i is given by

$$b_i(y(t)) = Pr(y(t)|x_t = i), \quad \forall y(t), i \tag{4.6}$$

where $y(t)$ is the pdf of the continuous observations. Let us denote $\boldsymbol{B} = \{b_i(y(t))\}$. The Hidden Markov chain is described as $\theta = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$, and the goal of Baum-Welch algorithm is to find the local maxima of $\theta^* = \boldsymbol{argmax_\theta} Pr(y(t)|\theta)$. Baum-Welch algorithm is formally described as follows:

a) Initialization

Set $\theta = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$ with random values if prior distributions of parameters are not available.

b) Forward Procedure

Let us denote $\alpha_i(t) = Pr(y(t), x_t = i|\theta), \quad 1 \leq t \leq \boldsymbol{t}$ which is the probability of being state i at time t with the observation pdf $y(t)$ and often called as *forward probability*. We first initialize the forward probability as given:

$$\alpha_i(1) = \pi_i b_i(y(1)) \tag{4.7}$$

then recursively find

$$\alpha_j(t+1) = b_j(y(t+1)) \sum_{i=1}^{Q} \alpha_i(t) a_{ij} \tag{4.8}$$

until it converges.

c) Backward Procedure

Let us denote $\beta_i(t) = Pr(y(t)|x_t = i, \theta), \quad \boldsymbol{t} \leq t \leq T$ which is the probability viewing pdf of partial sequence from t to T given state i at time t, and often called as *backward probability*. We first initialize the forward probability as given:

$$\beta_i(T) = 1 \tag{4.9}$$

then recursively find

$$\beta_i(t) = \sum_{i=1}^{Q} \beta_j(t+1) a_{ij} b_j(y(t)) \tag{4.10}$$

until it converges.

d) Update Parameters

After operating Forward Procedure and Backward Procedure, we need to update HMM parameters. Let us denote two temporary variables $\gamma_i(t)$ and $\xi_{ij}(t)$. $\gamma_i(t)$ is the probability of being in state i at time t given the observation sequence $y(t)$ and the parameters θ. Formally,

$$\gamma_i(t) = Pr(x_{l,t} = i | y(t), \theta) = \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^{Q} \alpha_j(t)\beta_j(t)} \tag{4.11}$$

$\xi_{ij}(t)$ is the probability of being in state i and j at times t and $t+1$ respectively given the observed sequence $y(t)$ and parameters θ. Formally it is given as follows:

$$\xi_i j(t) = Pr(x_{l,t} = i, x_{l,t+1} = j | y(t), \theta) = \frac{\alpha_i(t) a_{ij} \beta_j(t+1) b_j(y(t+1))}{\sum_{i=1}^{Q} \sum_{j=1}^{Q} \alpha_i(t) a_{ij} \beta_j(t+1) b_j(y(t+1))} \tag{4.12}$$

After the calculation of two temporary variables $\gamma_i(t)$ and $\xi_{ij}(t)$, we are able to update θ. According to our assumption the observation is a univariate Gaussian. We direct our readers to [Rab89] for more information on updating θ and Baum-Welch re-estimation for means and covariances of a CDHMM with Gaussian pdf.

4.3 Prediction Results

4.3.1 Shock Wave Prediction Accuracy

Since we know the threshold of the occupancy that tells you whether a shock wave occurred or not, we can categorize the state of a node (lane) of the DBN representation

of the HMM. We compared the accuracy of the categorization of the two states, shock wave or non-shock wave, performed by HMM. The following table demonstrates the accuracy of the prediction. SW denotes shock wave which one of the possible state variable values. NSW denotes non-shock wave.

Lane1	SW	NSW
SW	0.903	0.085
NSW	0.097	0.915

Table 4.1: Confusion Matrix of Lane 1 of I-405

The above matrix (table) is the confusion matrix of the prediction. We can see from this confusion matrix that accuracy is 90.9%. The average accuracy of lanes are high, however, the accuracy of the prediction of each lane differed from each other. Refer to the following tables.

Lane2	SW	NSW
SW	0.868	0.049
NSW	0.132	0.851

Table 4.2: Confusion Matrix of Lane 2 of I-405

Lane3	SW	NSW
SW	0.887	0.098
NSW	0.113	0.902

Table 4.3: Confusion Matrix of Lane 3 of I-405

Lane4	SW	NSW
SW	0.923	0.093
NSW	0.077	0.907

Table 4.4: Confusion Matrix of Lane 4 of I-405

4.3.2 Lane Switch Prediction

In this section, we present the results of the prediction of parameters, i.e., probability of a vehicle staying on a certain lane to switch to the neighboring lane(s) or stay on the same lane. Since the data has aggregated occupancy values, we can consider the prediction as the probability that majority of cars remain on lane or switch lanes.

The code for Hidden Markov Model and Baum-Welch algorithm was written in Python. The data sets used for the prediction are the 30-second PeMS data sets collected from January 1, 2016 through June 30, 2016. Since there is no baseline model for the prediction, we consider this prediction as a result of an unsupervised learning. The following tables are the average results of the aforementioned predictions during rush hours. Please be aware the probabilities are all normalized.

	L1	L2
L1	0.685	0.315
L2	0.287	0.713

Table 4.5: Final Transition Matrix of Lane 1 of I-405

One thing we realized is the probability that the occupancy of L2 will change is lower than other lanes. The reason we can think of is the existence of the carpool lane.

	L1	L2	L3
L1	0.685	0.315	0
L2	0.287	0.478	0.235
L3	0	0.613	0.387

Table 4.6: Final Transition Matrix of Lane 2 of I-405

	L2	L3	L4
L2	0.672	0.328	0
L3	0.431	0.483	0.086
L4	0	0.704	0.296

Table 4.7: Final Transition Matrix of Lane 3 of I-405

4.4 Summary

In this section, we proposed a model that predicts the upcoming shock wave occurrences based on the neighboring lanes and stations. Additionally, using Baum-Welch algorithm we could predict the lane switch probability of each lane at a certain station. However, this does not mean each vehicle will switch the lane by a certain probability value but rather the occupancy of a lane will change by that value.

CHAPTER 5

Deep Learning

Deep Learning (DL) has been very popular in various areas such as image recognition, speech recognition, natural language processing, and so forth [Da15]. Although Deep Learning architecture consists of multiple layers of Neural Networks (NN), it is different from multiple Artificial Neural Network (ANN) because it shares weights between NN layers unlike ANN. There is a study on traffic flow prediction using Deep Learning [LDK15]. By its black box nature, Deep Learning has been successful not only in unsupervised learning but also in supervised learning. Since Artificial Neural Network (ANN) has been successful in predicting short-term traffic [KPK13], Deep Learning has also been successful in predicting traffic information [LDK15].

Deep Learning algorithms adopt either deep architecture or multilayer architecture to extract complex data abstractions. These algorithms construct a hierarchical architecture of learning and data representations where higher level features are derived from lower level features [LDK15]. Due to the hierarchical nature of the Deep Learning algorithms, they can learn multiple levels of data representations that correspond to different levels of abstractions. Traffic data involves timeseries, and timeseries prediction involves analyzing non-linearity aspect of the timeseries. Therefore, Deep Learning is suitable for predicting vehicular shockwave without prior knowledge. In this chapter, specifically Stacked Auto Encoder (SAE) and Deep Belief Network (DBN) are introduced among Deep Learning algorithms.

5.1 Encoders

5.1.1 Autoencoder

An autoencoder is a Neural Network that attempts to reproduce its input, i.e., the output layer has the same number nodes as the input layer of the model [HOT06a]. An autoencoder has one input layer, an output layer, and one ore more hidden layers connecting them. Given a set of training samples $\mathbf{x} = \{x^{(1)}, ..., x^{(n)}\}$, where $x^{(i)} \in R^d$, an autoencoder first encodes the input to a hidden representation $\mathbf{y} = y(\mathbf{x})$ via a deterministic mapping as follows:

$$y(x) = s(\boldsymbol{W}x + \mathbf{b}) \quad or \quad \mathbf{y} = s(\boldsymbol{W}\mathbf{x} + \mathbf{b}) \tag{5.1}$$

where s is the non-linearity function such as sigmoid function, i.e., $\frac{1}{(1+\exp^{-x})}$, $\boldsymbol{W}$ is the weight matrix, and $\boldsymbol{b}$ is the encoding bias vector, respectively. The latent representation $\mathbf{y}$ is decoded into a reconstruction $\mathbf{z} = z(\mathbf{x})$ via a similar deterministic mapping:

$$z(x) = s(\boldsymbol{W}'x + \mathbf{c}) \quad or \quad \mathbf{z} = s(\boldsymbol{W}'\mathbf{x} + \mathbf{c}) \tag{5.2}$$

where s is the non-linearity function, $\boldsymbol{W}'$ is the weight matrix, and $\boldsymbol{c}$ is the decoding bias vector, respectively. Given a code $\mathbf{y}$, $\mathbf{z}$ is considered as the prediction of input vector $\mathbf{x}$. Minimizing the reconstruction error $L(\mathbf{x}, \mathbf{z})$ is crucial in order to obtain the best model parameters which are denoted as θ. The formula of reconstruction error is given as below:

$$L(\mathbf{x}, \mathbf{z}) = \| \mathbf{x} - \mathbf{z} \|^2 \tag{5.3}$$

which is the *squared error*, and the model parameters as follows:

$$\theta = arg\, min_\theta L(\mathbf{x}, \mathbf{z}) = arg\, min_\theta \| \mathbf{x} - \mathbf{z} \|^2 \tag{5.4}$$

5.1.2 Stacked Autoencoder

Stacked Autoencoder is formed by stacking the aforementioned autoencoders in order to build deep neural network. Architecture-wise the input is the at the lowest layer of the network, and the encoded hidden layer produces the output layer. The output of the kth hidden layer becomes the input of the $(k+1)$th hidden layer. In this way, the hierarchy of the deep neural network is formed. The top layer is a predictor. In Figure 5.1, the bottom layer of SAE in the left most layer and the top layer is the right most layer.

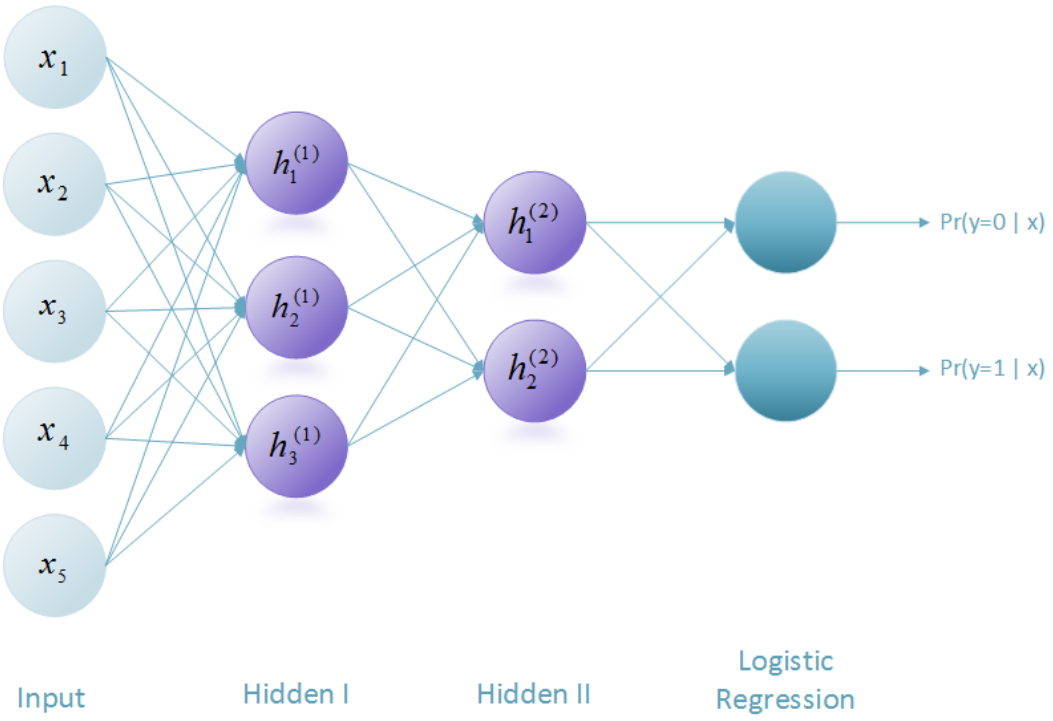

Figure 5.1: Stacked Autoencoder with a binary Logistic Regression Predictor

Training stacked autoencoders is performed by utilizing greedy layerwise unsupervised learning algorithm [LDK15, HOT06a]. The training procedure is summarized as follows:

Algorithm 1. Training SAEs

Given training samples X and the desired number of hidden layers l,

Step 1) Pretrain

— Set weight parameters and initialize weight matrices and bias vectors randomly.

— Greedy layerwise training hidden layers

— Use the output of the kth layer as the input of the $k+1$th layer. The first layer's input is the training data set.

— Find the encoding parameters $\{W_1^{k+1}, b_1^{k+1}\}_{k=1}^{l}$ for $k+1$th hidden layer by minimizing the objective function.

Step 2) Fine tuning the whole network

— Initialize $\{W_1^{k+1}, b_1^{k+1}\}$ or by supervised training.

— Use the Back Propagation with the gradient-based optimization technique to update the whole network's parameters.

Table 5.1: Trainign SAE Algorithm

5.2 Boltzmann Machine

5.2.1 Restricted Boltzmann Machine

Restricted Boltzmann Machine (RBM) is a variant of Boltzmann machines with the restriction of the usage of bipartite graph. The two disjoint sets are called hidden layer and visible layer. The nodes of each layer has a connection with nodes in the other layer, however, the nodes within the same layer cannot have a connection between them. Formally, RBM has binary-valued hidden and visible units, and consists of a weight matrix $\boldsymbol{W} = \{w_{i,j}\}$. $\boldsymbol{W}$ is associated with the connection between the hidden unit h_j and the visible unit v_i (See Figure 5.2). Additionally, it is associated with the bias weight for the visible units, a_i, and the bias weight for the hidden units,

b_i [HOT06a].

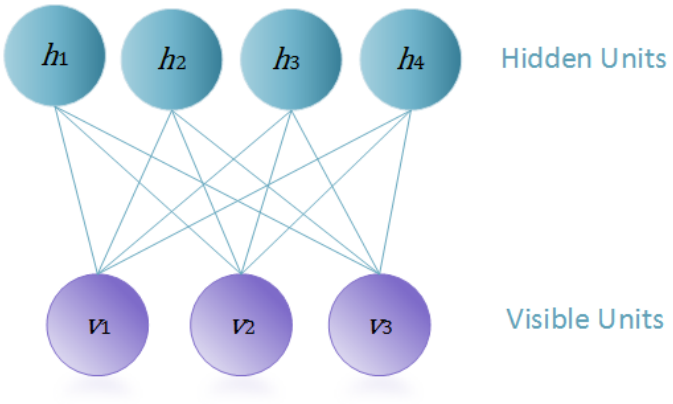

Figure 5.2: Restricted Boltzmann Machine

The energy of a configuration (v, h) is defined as follows:

$$E(v,h) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_i \sum_j v_i w_{i,j} h_j \tag{5.5}$$

The conditional probability of visible units $\boldsymbol{v}$ given hidden units $\boldsymbol{h}$ is

$$Pr(\boldsymbol{v}|\boldsymbol{h}) = \prod_i^m Pr(v_i|\boldsymbol{h}) \tag{5.6}$$

Conversely, the conditional probability of hidden units $\boldsymbol{h}$ given visible units $\boldsymbol{v}$ is

$$Pr(\boldsymbol{h}|\boldsymbol{v}) = \prod_i^m Pr(h_i|\boldsymbol{v}) \tag{5.7}$$

The individual activation probabilities are

$$Pr(h_j = 1|\boldsymbol{v}) = s(b_j + \sum_{i=1}^m w_{i,j} v_i) \tag{5.8}$$

and

$$Pr(v_i = 1|\boldsymbol{h}) = s(a_i + \sum_{j=1}^n w_{i,j} h_j) \tag{5.9}$$

where s is the sigmoid function.

The goal of RBM is to maximize the product of probabilities assigned to the training data set V. Formally,

$$arg\,max_W \prod_{v \in V} Pr(v) \tag{5.10}$$

The training algorithm of RBM is summarized as follows [HOT06a].

Algorithm 2. Training RBM
Given training samples X, select a training sample v. — Compute the probabilities of the hidden units and sample a hidden activation vector h from this probability distribution. — Computer the outer product of v and h (positive gradient) — From h, sample a reconstruction $v^{'}$ of the visible units, then resample the hidden activations $h^{'}$ from h. (Gibbs sampling) — Computer the outer product of $v^{'}$ and $v^{'}$ (negative gradient) — Update weight matrix ($\boldsymbol{W}$): $\Delta \boldsymbol{W} = \epsilon(vh^T - v'h'^T)$ where ϵ is the learning rate.

Table 5.2: Trainign RBM Algorithm

5.2.2 Deep Belief Network

Deep Belief Network is formed by composing simple and unsupervised networks such as Restricted Boltzmann machines or autoencoders [Hin09]. In this dissertation, we will use RBM to construct a Deep Belief Network. Since Restricted Boltzmann Machine consists of a hidden layer and a visible layer (See Figure 5.2), when we construct a Deep Belief Network the hidden layer of the sub-RBM serves as the visible layer of the current RBM (See Figure 5.3).

The training algorithm of RBM is summarized as follows [HOT06b]:

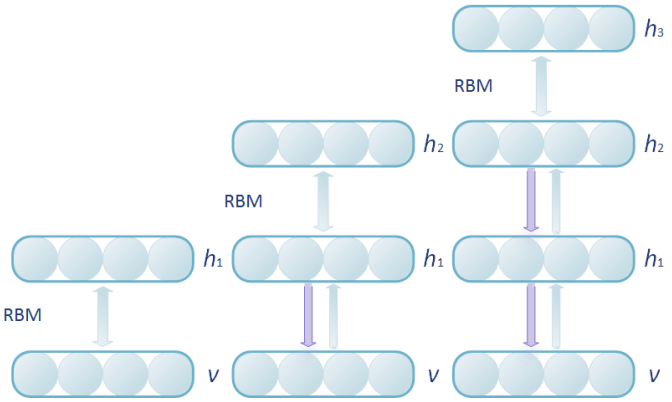

Figure 5.3: Deep Belief Network composed of Restricted Boltzmann Machine

Algorithm 3. Training Deep Belief Network

Given training samples X,

— Train a RBM to obtain weight matrix $\boldsymbol{W}$ and utilize it as the weight matrix for the lowest two layers of the network.

— Transform X by sampling or by computing the mean activation of the hidden units to obtain new data X'.

— Repeat the above two steps with the new input X' until the top two layers are reached.

— Fine tune the parameters.

Table 5.3: Trainign Deep Belief Network

5.3 Inference Results Comparison

In this section, we compare the shock wave prediction results obtain from SAE, Deep Belief Network, and HMM. Originally, we attempted to compare the lane switch prediction probability generated by HMM. The transition matrix created by Baum-Welch algorithm indicates the likelihood of the occupancy of a lane will change, but not the likelihood of an individual vehicle switching lanes based on the occupancy.

We were anticipating Deep Learning algorithms could help us to obtain an insight by performing several feature engineering methods, however, since the data itself is an aggregated macroscopic traffic data, we couldn't find a way to verify the probability of a vehicle to switch lanes without the help of microscopic traffic data. Hence, the lane switch prediction is excluded.

5.3.1 Shock Wave Prediction

Since we know that under certain threshold value of occupancy, shock wave occurs, we could label the data set with two categories, SW (shock wave) and NSW (non-shock wave). We took into account of the neighboring lanes' traffic status as considered in Section 4.

5.3.1.1 SAE

The following is the confusion matrix of each lane of I-405 during rush hours.

Lane1	SW	NSW
SW	0.931	0.096
NSW	0.078	0.922

Table 5.4: Confusion Matrix of Lane 1 (SAE)

Lane2	SW	NSW
SW	0.923	0.077
NSW	0.076	0.924

Table 5.5: Confusion Matrix of Lane 2 (SAE)

Lane3	SW	NSW
SW	0.932	0.068
NSW	0.091	0.909

Table 5.6: Confusion Matrix of Lane 3 (SAE)

Lane4	SW	NSW
SW	0.935	0.065
NSW	0.070	0.930

Table 5.7: Confusion Matrix of Lane 4 (SAE)

5.3.1.2 Deep Belief Network

The following is the confusion matrix of each lane of I-405 during rush hours.

Lane1	SW	NSW
SW	0.928	0.072
NSW	0.073	0.927

Table 5.8: Confusion Matrix of Lane 1 (DBN)

Lane2	SW	NSW
SW	0.926	0.064
NSW	0.068	0.932

Table 5.9: Confusion Matrix of Lane 2 (DBN)

Lane3	SW	NSW
SW	0.940	0.060
NSW	0.081	0.919

Table 5.10: Confusion Matrix of Lane 3 (DBN)

Lane4	SW	NSW
SW	0.927	0.063
NSW	0.064	0.936

Table 5.11: Confusion Matrix of Lane 4 (DBN)

5.3.1.3 Comparison with HMM

The following is the prediction accuracy comparison chart of the shock wave occurrence of SAE, Deep Belief Network (DBN), and HMM.

Lane	SAE	DBN	HMM
Lane1	92.7 (%)	92.8 (%)	90.9 (%)
Lane2	92.4 (%)	92.9 (%)	86 (%)
Lane3	92.1 (%)	93.0 (%)	89.5 (%)
Lane4	93.3 (%)	93.2 (%)	91.5 (%)

Table 5.12: Prediction Accuracy Comparison between SAE, DBN, and HMM

The Deep Belief Network slightly outperformed SAE, however, in a very negligible amount.

5.4 Summary

In this chapter, we adopted two Deep Learning algorithms, Stacked Autoencoder (SAE) and Deep Belief Network (DBN) in order to predict the occurrences of shock waves on freeway I-405. We compared the results of the prediction from the aforementioned two DL algorithms with that of of HMM. All three models have high accuracy in terms of predicting shock wave. Lane change prediction has been excluded since there is no ground truth to verify the prediction. In order to verify the lane change prediction it is crucial to have a microscopic traffic data that has individual vehicle information.

CHAPTER 6

Conclusion

In this dissertation, the main goal of the research is to predict the occurrences of shock waves on highways given an aggregated macroscopic traffic data set. We first attempted to obtain the vehicular propagation speed by plotting a heatmap for each lane of a freeway. We utilized Linear Regression with cross-validation in order to obtain the best slope of the heatmap which is shock wave propagation speed. Since the finest granularity PeMS data set has only occupancy (density) and flow information of a lane, we attempted to infer the average speed of vehicles passing by the station from either flow or occupancy. After investigating the correlation between speed and flow, and speed and occupancy, we realized there was a stronger correlation between speed and occupancy than speed and flow. We then utilized Polynomial Regression to infer speed from occupancy, however, the prediction was inaccurate. We changed our strategy to see if a heatmap of occupancy can be used to calculate shock wave propagation speed. Fortunately, the heatmaps of occupancy and the heatmaps of speed were almost the same. Therefore, the finest granularity data set was used to infer shock wave.

We built a model based on HMM in the assumption of the traffic status of a lane is affected by itself and the neighboring lanes at current station and a station ahead at the previous timestamp. From the HMM, we could categorize whether a data point will experience shock wave based on the lane occupancy. We evaluated the model performance by comparing the prediction accuracy. Additionally, from the Baum-Welch algorithm, we obtained the probability of lane switch. Note that this is

not the probability of an individual car to switch lane, but more like the occupancy of lane will change.

We utilized Stacked Autoencoder and Deep Belief Network that are Deep Learning algorithms in order to predict shock wave occurrences on a freeway. We compared the prediction results with those of HMM, and obtained high accuracy rate.

For future work, we will evaluate our models that are based on macroscopic traffic data if they perform well on mixed microscopic traffic data and macroscopic data.

References

[AB12] Aude Hofleitnerand Ryan Herringand Pieter Abbeeland and Alexandre Bayen. "Learning the Dynamics of Arterial Traffic From Probe Data Using a Dynamic Bayesian Network." **13**(4), 2012.

[BE] Leonard E. Baum and J. A. Eagon. "An Inequality with Applications to Statistical Estimation for Probabilistic Functions of Markov Processes and to a Model for Ecology." **73**(3):360–363.

[BMS12] Carolina Tripp Barba, Miguel ngel Mateos, Pablo Regaas Soto, Ahmad Mohamad Mezher, and Mnica Aguilar Igartua. "Smart city for VANETs using warning messages, traffic statistics and intelligent traffic lights." Intelligent Vehicles Symposium (IV), IEEE, 2012.

[Bri06] Economist Intelligence Unit (Great Britain). "Driving Change: How Policymakers are Using Road Charging to Tackle Congestion.", 2006.

[BSS98] R. Barlovic, L. Santen, A. Schadschneider, and M. Schreckenberg. "Metastable States in Cellular Automata for Traffic Flow." *The European Physical Journal B-Condensed Matter and Complex Systems*, **5**(3):793–800, 1998.

[Da15] George Edward Da. "Deep learning approaches to problems in speech recognition, computational chemistry, and natural language text proc." 2015.

[DK82] Pierre A. Devijver and Josef Kittler. *Pattern Recognition: A Statistical Approach*. Prentice-Hall, 1982.

[FFG14] Markus Forster, Raphael Frank, Mario Gerla, and Thomas Engel. "A Cooperative Advanced Driver Assistance System to Mitigate Vehicular Traffic Shock Waves." IEEE INFOCOM, 2014.

[Gei93] Seymour Geisser. *Predictive Inference*. Chapman and Hall, 1993.

[Hin09] Geoffrey E. Hinton. "Deep Belief Networks." **4**(5):5947, 2009.

[HOT06a] G. E. Hinton, S. Osindero, and Y. Teh. "A fast learning algorithm for deep belief nets." **18**:1527–1554, 2006.

[HOT06b] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets." 2006.

[HPM90] A. Huerre, Patrick, and P. Monkewitz. "Local and global instabilities in spatially developing flows." **22**:473 537, 1990.

[KM00] Jaimyoung Kwon and Kevin Murphy. "Modeling Freeway Traffic with Coupled HMMs." 2000.

[Koh95] Ron Kohavi. "A study of cross-validation and bootstrap for accuracy estimation and model selection." **2**:1137 1143, 1995.

[KPK13] Kranti Kumar, M. Parida, and V.K. Katiyar. "Short Term Traffic Flow Prediction for a Non Urban Highway Using Artificial Neural Network." **104**:755–764, 2013.

[Kra98] S. Krauss. "Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics." 1998.

[KWG97] S. Krauss, P. Wagner, and C. Gawron. "Metastable states in a microscopic model of traffic flow." **55**(304):55–97, 1997.

[LDK15] Yisheng Lv, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang. "Traffic Flow Prediction With Big Data: A Deep Learning Approach." **16**:865 – 873, 2015.

[LJ99] David R. Legates and Gregory J. McCabe Jr. "Evaluating the use of goodness-of-fit Measures in hydrologic and hydroclimatic model validation." **35**(1):233–241, 1999.

[LW55] M. Lighthill and G. Whitham. "On kinematic waves. II. A theory of traffic flow on long crowded roads." **229**(1178):317 345, 1955.

[NS92] K. Nagel and M. Schreckenberg. "A cellular automaton model for freeway traffic." **2**(12):2221 2229, 1992.

[PN11] A. Pascale and M. Nicoli. "Adaptive Bayesian Network for Traffic Flow Prediction." 2011.

[Rab] Lawrence Rabiner. "First Hand: The Hidden Markov Model.".

[Rab89] Lawrence R. Rabiner. "A tutorial on hidden Markov models and selected applications in speech recognition." **77**(2):257–286, 1989.

[Ric] P. Richards. "Shock Waves on the Highway." **4**(1):4251.

[RKP10] H. Rehborn, S. Klenov, and J. Palmer. "Common Traffic Congestion Features studied in USA, UK, and Germany employing Kerners Three-Phase Traffic Theory." 2010.

[Tra15] Califorina Department of Transportation. "Caltrans Performance Measurement(PeMS) data set.", 2015.

[Wil08] R. E. Wilson. "Mechanisms for spatio-temporal pattern formation in highway traffic models." **366**(1872), 2008.

[Wir78] S. Chandana Wirasinghe. "Determination of Traffic Delays from Shockwave Analysis." *Transportation Research*, **12**(5):343–348, 1978.