# A STOCHASTIC BAYESIAN UPDATE AND LOGISTIC GROWTH MAPPING OF TRAVEL-TIME FLOW RELATIONSHIP

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Mohammad Mofigul Islam Molla

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Program:
Civil Engineering

December 2016

Fargo, North Dakota

North Dakota State University
Graduate School

**Title**

A STOCHASTIC BAYESIAN UPDATE AND LOGISTIC GROWTH MAPPING OF TRAVEL-TIME FLOW RELATIONSHIP

**By**

Mohammad Mofigul Islam Molla

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

**DOCTOR OF PHILOSOPHY**

SUPERVISORY COMMITTEE:

Matthew L. Stone, Ph.D.
Chair

Ying Huang, Ph.D.

Chowdhury K. A. Siddiqui, Ph.D., P.E.

Yong Bai, Ph.D., P.E.

Approved:

January 10, 2017 — Date

Dinesh Katti, Ph.D., P.E. — Department Chair

**ABSTRACT**

The travel-time flow relationship is not always increasing in nature, it is very difficult to predict precisely. Traditional method fails to replicate this unique conditions. Until millennium, although various researchers and practitioners have given much attention to develop travel-time flow relationships, the advancement to improve travel-time flow relationships was not substantial. The knowledge about the travel-time flow relationship is not commensurate with or parallel to the advancement of new knowledge in other fields. After millennium, most investigators did not devote enough attention to create new knowledge, except for application and performance evaluation of the existing knowledge. Therefore, it is necessary to provide a new theoretical and methodological advancement in travel-time flow relationship.

Consequentially, this research proposes a new methodology, which considers stochastic behavior of travel-time flow relationship with probabilistic Bayesian statistics and logistic growth mapping techniques. This research moderately improves the travel-time flow relationship. The unique contribution of this research is that the proposed methods outperforms the existing traditional travel-time flow theory, assumptions, and modeling techniques. The results shows that the proposed model is considerably a good candidate for travel-time predictions. The proposed model performs 36 percent better and accurate travel-time predictions in compared to the existing models.

Furthermore, travel-time flow relationship need capacity and free-flow speed estimations. Traditionally, practice of capacity estimation is mostly practical, subjective, and not steady-state capacity. Therefore, a robust and stable capacity-estimation method was developed to eliminate the subjectivity of capacity estimation. The proposed model shows robust and capable of replicating steady-state capacity estimation. The free-flow speed estimation should relate to the traffic-flow speed model while the density is zero. Therefore, this research investigates the existing deterministic speed-density models and recommends a better methodology in free-flow speed estimation. This research presents how the undefined practice of free-flow speed selection can be sensitive.

Additionally, finding suitable concurrent travel-time data and traffic volume is crucial and very challenging. To collect concurrent data, this research investigates and develops several technologies such as crowdsource, web app, virtual sensor method, test vehicle, smartphone, global positioning system, and utilized several state and local agencies data collection efforts.

**Keywords:** Travel-Time Flow, Travel-Time Delay, Volume-Delay Function, Travel Time, Origin-Destination Survey, Travel Demand Model, Travel Data Collection, Transportation Survey, Internet Sensor, Crowdsourcing, Virtual Sensor Method, VSM, Transportation Planning, GPS, Smartphone, Loop Detector, Travel -Time Prediction, Travel-Speed Prediction, TDM, Bayesian Inference, Logistic Growth Function.

## ACKNOWLEDGEMENTS

First, I would like to thank Allah, the almighty, for giving me the blessing and for sanctioning me with the confidence to fulfill this task.

Next, I owe my deepest and profound gratitude to my adviser, Dr. Matthew L. Stone, who shepherded me through the bulk of my work with his sage guidance and shrewd cooperation. I would like to thank my adviser for his outstanding supervision, guidance, continuous support, advice, motivation, and funding opportunities to finish this research and Doctor of Philosophy (PhD). Dr. Stone has a great deal of experience in many areas of civil engineering; his expertise was beneficial for this work. He provided me with the opportunity to graduate study and conduct research at North Dakota State University.

I would also like to thanks a lot to the following committee members for being supportive and for their valuable comments, reviews, inputs, and time. Their invaluable cooperation helps me to think critically on my research and study. I would like to thank each of them for everything that they have done to help me complete my doctoral degree and dissertation.

i. Dr. Ying Huang, Assistant Professor, Department of Civil and Environmental Engineering, Fargo, North Dakota.

ii. Dr. Chowdhury K. A. Siddiqui, System Performance Management Engineer, South Carolina Department of Transportation.

iii. Dr. Yong Bai, Professor, Department of Civil, Construction and Environmental Engineering, Marquette University, Milwaukee, Wisconsin.

I want to thank Mr. Bradley Wentz and Dr. Diomo Motuba for their advice and generous funding support towards my graduate study. I would also like to thank to the Advanced Traffic

Analysis Center (ATAC) at Upper Great Plains Transportation Institute (UGPTI) for providing me funding for graduate study.

I want to express my gratitude to the following entities:

i. North Dakota State University (NDSU)
ii. Department of Construction Management and Engineering, NDSU
iii. Advanced Traffic Analysis Center, UGPTI
iv. Department of Civil and Environmental Engineering, NDSU
v. North Dakota Department of Transportation (NDDOT)
vi. California Department of Transportation (Caltrans)
vii. Mid-Region Council of Governments, New Mexico
viii. Florida Department of Transportation (FDOT)
ix. Fargo-Moorhead Metro Council of Governments (FM Metro COG)
x. Grand Forks-East Grand Forks Metropolitan Planning Organization
xi. Bismarck-Mandan Metropolitan Planning Organization
xii. OpenStreetMap Services
xiii. R: The R Project for Statistical Computing
xiv. 128 References
xv. Families, Friends, and Colleagues

## DEDICATION

To my lovely parents, angel sisters, and supportive brothers.

**TABLE OF CONTENTS**

## LIST OF TABLES

## LIST OF FIGURES

# LIST OF EQUATIONS

## LIST OF ABBREVIATIONS

AJAX....................................Asynchronous JavaScript and Extensible Markup Language

AM.........................................Ante Meridiem

API.........................................Application Programming Interface

ARIMA..................................Auto Regressive Integrated Moving Average

ATAC....................................Advanced Traffic Analysis Center

ATIS......................................Advanced Traveler Information Systems

ATMS....................................Advanced Traffic Management System

AVI........................................Automatic Vehicle Identification

AVL.......................................Automatic Vehicle Location

BPR........................................Bureau of Public Roads

DLM......................................Dynamic Linear Model

DMI.......................................Distance Measuring Instrument

DOT.......................................Department of Transportation

DynaMIT................................Dynamic Network Assignment for the Management of Information of Travelers

EES........................................Extended Exponential Smoothing

ETC.......................................Electronic Toll Collection

FFS........................................Free-Flow Speed

FFT........................................Free-Flow Travel-Time

FHWA...................................Federal Highway Administration

FM Metro COG......................Fargo-Moorhead Metropolitan Council of Governments

FMMPA.................................Fargo-Moorhead Metropolitan Planning Area

GBM......................................Gradient Boosting Method

GIS.........................................Geographic Information Systems

GPS........................................Global Positioning Systems

HCM......................................Highway Capacity Manual

HTTP.....................................Hypertext Transfer Protocol

IOS........................................ Internetwork Operating System

ISTEA....................................Intermodal Surface Transportation Efficiency Act

ITS.........................................Intelligent Transportation Systems

JSON.....................................JavaScript Object Notation

KF..........................................Kalman Filter

KML......................................Keyhole Markup Language

KMZ......................................Keyhole Markup Language Zipped

K-NN.....................................K-Nearest Neighbor

LM.........................................Linear Model

LOKRR.................................Local Online Kernel Ridge Regression

LOS.......................................Level of Service

LPR........................................License Plate Recognition

MAC......................................Median Access Control

MPA......................................Metropolitan Planning Area

MPH…………………..……Miles per Hour

MPO......................................Metropolitan Planning Organization

NCHRP.................................National Cooperative Highway Research Program

NDSU....................................North Dakota State University

NEPA....................................National Environmental Policy Act

NHTS....................................National Household Travel Survey

NLS.......................................Non-Linear Least Square

NN.........................................Neural Network

NPMRDS..............................National Performance Measuring Research Data Set

O-D.......................................Origin-Destination

PDA.......................................Personal Digital Assistant

PEMS....................................Performance Measurement System

PHV.......................................Proportion of Heavy Vehicle

PM.........................................Post Meridiem

PTFM....................................Passive Target Flow Management

REST.....................................Representational State Transfer

RFID......................................Radio-Frequency Identification Transponders

RMSE....................................Root Mean Squared Error

RSE.......................................Residual Standard Error

SAFETEA-LU........................Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users

SAS....................................... SAS Institute Inc.

SDK.......................................Software Development Kit

SNN.......................................Spectral Neural Network

SSNN.....................................State Space Neural Network

STD.......................................Standard Deviation

ST-D......................................Space-Time Diurnal Method

SVR.......................................Support Vector Regression

TAZ.......................................Traffic Analysis Zone

TDM.......................................Travel Demand Model

TRB........................................Transportation Research Board

TTP.........................................Travel Time Prediction

VMT.......................................Vehicle Miles Travelled

VPH......................................Vehicles per Hour

VSM.......................................Virtual Sensor Methodology

WIM.......................................Weigh-In-Motion

WLT.......................................Wireless Location Technology

WPF........................................Windows Presentation Foundation

XML.......................................Extensible Markup Language

## LIST OF SYMBOLS

$\tilde{P}_{PERT}$……………………….PERT Likelihood Shape Parameter

$\tilde{Q}_{PERT}$……………………….PERT Likelihood Shape Parameter

$\sigma^2$…………………………....Statistical Variance

$\Delta x$………………………......Time Required to Reach 10-90 Percent of *l*

μ……………………………..Statistical Mean

$a$……………………………..Modeling Parameter

b……………………………..Modeling Parameter

$c$……………………………..Link Capacity

$c_n$…………………………….Cumulative Flow

$c_p$…………………………….Link practical Capacity

$c_s$…………………………… Link Stable Capacity

$c_u$…………………………..….Link Ultimate Capacity

$D$ ....………………………....Dataset

$f$…………………………….....Lower Boundary of Volume/Capacity Ratio

*FFS*………………………….Free-Flow Speed

$f_{LC}$…………………………..…...Adjustment of Right Side Lane Clearance

$f_{LW}$………………………….....Adjustment of Lane Width

$g$……………………………..Upper Boundary of Volume/Capacity Ratio

$j$…….….... ………………......Modeling Calibration Parameter

$k^{(p)}$ ……………………..…….Logistic Function Steepness of the Curve at p-th percentile

$k$…………………….....…….Logistic Function Steepness of the Curve

$l^{(p)}$ …….……………………….Logistic Function Maximum Saturation at p-th percentile

$L^*$……………………………Maximum Equivalent Likelihood

$l$...........................................Logistic Function Maximum Saturation Parameter

$M$....………………...………Most Likely Travel Time

$m$……………………………Number of Datasets

$\boldsymbol{N}$……………………………Natural Number

$p(D)$ ....……………………...Marginal Probability

$p(D|\theta)$ ....……………………Conditional Probability

$p(\theta)$ ....………………………Prior Probability

$p(\theta|D)$ ....……………………Posterior Probability

$P$ ....……………………………Beta Function Shape Parameter

$p$-$value$………………………Statistical P-Value

$Q$ ....……………………....…Beta Function Shape Parameter

$q$……………………………..Statistical Quantile

$Q$-$Q$…………………………Quantile-Quantile Plot

$\boldsymbol{R}$……………………………Real Number

$R$-$Squared$…………………...Statistical Coefficient of Determination

$t/t_o$....…………………………..Travel Time Delay Ratio

$t$……………………………...Travel Time at Traffic Flow $v$

$t_0$………………………….....Initial Travel Time at Zero Traffic Flow

$t_1$……………………….…….Travel Time at Different Region

$t_2$……………………………..Travel Time at Different Region

$TRD$…………………………Total Ramp Density (Ramps per Mile)

$t$-$value$………………….……Statistical t-value

$U$......................................Beta Function Lower Boundary

Ʊ......................................Modeling Calibration Parameter

$v/c$......................................Volume/Capacity Ratio

$V$......................................Beta Function Upper Boundary

$v$......................................Traffic Flow

$v_i$......................................Mean Hourly Flow at i-th Hour

$v_{max}$......................................Maximum Mean Hourly Flow

$v_{predicted}$......................................Predicted Traffic Flow

$x$......................................Random Variable

$x_m$......................................Midpoint in Time

$x_o^{(p)}$......................................Logistic Function Sigmoid Midpoint at p-th Percentile

$x_o$...................................... Logistic Function Sigmoid Midpoint

$y$......................................Random Variable

$y_i$......................................Normalized Flow at i-th Hour

$\alpha$......................................Modeling Calibration Parameter

$\beta$......................................Modeling Calibration Parameter

$\gamma$......................................Modeling Calibration Parameter

$\delta x$ ......................................Increment of $x$

$\theta$ ......................................Probabilistic Parameter

$\tau$...................................... .Modeling Calibration Parameters

$\epsilon$......................................Random Error

## LIST OF APPENDIX FIGURES

# 1. INTRODUCTION

## 1.1. General

The four-step travel demand model (TDM) is widely used, well known, and recognized in the transportation-planning industry. This four-step travel demand prediction model requires trip generation, trip distribution, modal split, and traffic assignment. The first step, trip generation, estimates the number of trips; the second step, trip distribution estimates the number of trips for given origins-destinations; the third step, modal split distributes the trips' origin-destination (O-D) probability for different modes; and the final step, highway assignment, assigns different route levels for the trips.

The efficacy of travel demand modeling substantially depends on how accurate the model performs when forecasting travel. The TDM involves accurate estimations of existing and future demand with proper alternatives. It is understood that the sequential TDM requires several input datasets and numerous input parameters. Because a TDM is a sequential model with several sub-models, errors of any stage could be propagated and accumulated into the entire model and final outcomes.

Several past studies have estimated the TDM's inaccuracies. For example, 50 percent of the traffic predictions had estimation errors that were greater than 20 percent (Yang et al., 2013). Rasouli and Timmermans (2012) presented an extensive literature review about the uncertainty of travel demand forecasting models. Flyvbjerg et al. (2005) found that, for the past 30 years, TDMs underperformed when predicting traffic.

Yang et al. (2013) studied the sensitivity-based uncertainty analysis of a combined TDM. They stated that the forecasting uncertainties were generated from socioeconomic datasets, input parameters, model methodology, and assumptions. Zhao and Kockelman (2002) studied the propagation of uncertainty for TDMs. Their studies inferred that transportation systems' operations are subjected to significant uncertainty due to the input parameters.

There are possibilities for errors to occur with the TDMs due to the highway-assignment procedure. Inaccuracy during the highway-assignment step could propagate and accumulate downstream into the entire model, eventually leading to erroneous models. Travel-time delay *($t/t_o$)* function is one of the key input for the highway-assignment step. Regardless of the highway-assignment algorithm, the $t/t_o$ functions that are being using with the highway assignment may have the biggest influence on TDMs inability to replicate the observed traffic. Therefore, to improve the existing TDM theory and methodology, it is important to obtain new knowledge that can capture travel-time uncertainties and stochastic behavior for the highway-assignment stage of TDM prediction.

In order to understand the travel-time uncertainty, knowledge about the traffic-flow theory is necessary. Therefore, the next section discusses the existing traffic-flow theory.

## 1.2. Background

"Traffic-flow theory" is an inaccurate term because traffic flow does not happen in theory, rather it occurs on real roadways (Roess et al., 2011). Roess et al. (2011) stated that there is one common characteristic among all speed-flow historical models and Greenshields' modern linear speed-flow model. The commonality is that speed always decreases with more flow (Roess et al., 2011). The two-dimensional speed-flow relationship has evolved through the three-dimensional, speed-flow-density relationship (Hall, 1975). Most of the modern, two-dimensional, speed-flow models are, thus, based on three special regions of this curve: 1) uncongested condition, 2) queue discharge, and 3) congested condition (Akcelik, 1991; Roess et al., 2011).

The first region explains the uncongested condition where speed reduces very slowly or is relatively constant until capacity (Roess et al., 2011). The second region establishes the queue-discharge theory. At this region, the queue represents a vertical variation at the capacity. This segment is relatively very small (Roess et al., 2011). The third region suggests that the traffic congestion start forming at this stage. The last two regions are empirical.

In contrast, ideally, travel time always increases with higher traffic flow. This relationship is recognized by the transportation industry as the travel-time flow, volume-delay, or delay functions. Until millennium, although various researchers and practitioners devoted much attention to developing the travel-time flow relationship, advances to improve the travel-time flow relationship is not substantial. Knowledge about the travel-time flow relationship is not commensurate with or parallel to the advancement of knowledge in other fields. After millennium, most researchers did not devote enough attention to investigate new knowledge, except for application and performance evaluation of the existing knowledge. One major issue is the way that the traffic theory can explain the speed-flow relationship; speed-flow model theory do not explain the travel-time flow theory in a similar way.

Spiess (1990) has defined following conditions that need to be satisfied for a given $t/t_o$ function to meet equilibrium assignment algorithm:

i. The function should be strictly increasing in order to make the assignment convergent.
ii. The function should generate one when the volume is zero and generate two when volume is equal to the capacity.
iii. The derivative of the function should exist and be strictly increasing.

iv. One of the model calibration parameter should be observed at the derivative's solution when the volume is equal to the capacity.

v. The derivative of the function should be a finite and positive constant.

vi. The derivative of the function should be positive when the volume is zero, i.e., at a free-flow travel time (FFT).

vii. The function should be computationally less time consuming than the Bureau of Public Roads' (BPR) method.

So far, the overall background of the $t/t_o$ function characteristics has been presented in this section. In a later section, the existing methodologies that might reveal the theoretical and methodological aspects of the $t/t_o$ functions has been presented. These characteristics will guide to find the current conditions, industry needs, and gaps in the literature.

### 1.3. Existing Methodologies

The review of literature shows that there are number of $t/t_o$ functions that are being used by researchers and practitioners. Some of the functions are presented in Table 1. Table 1 has been summarized and presented from the findings of Branston (1975), Gan et al. (2003), Ayad (1967), Gesalem and Fillone (2016), Mtoi and Moses (2014), Smith et al. (1999), and Kalaee (2010). Majority of the volume-delay functions presented in Table 1 were utilized from the works of Gan et al. (2003) and Branston (1975).

Branston (1975) presented a brief literature review (up to 1974) about different $t/t_o$ functions. He reviewed the works of Irwin et al. (1961), Irwin and Von Cube (1962), Smock (1962), Soltman (1965), Overgaard (1967), Mosher (1963), Bureau of Public Roads ([BPR], 1964), Steenbrink (1974), and Traffic Research Corporation (1966). Gan et al. (2003) presented the works of Campbell (1959), Irwin et al. (1961), Smock (1962), Mosher (1963), Soltman (1965), Overgaard (1967), Davidson (1966), and BPR (1974) models.

Table 1. List of Travel-Time Delay Functions[1]

| Author | Model | | Sources |
|---|---|---|---|
| Campbell (1959) | $t = t_0$<br>$t = t_0 + \alpha(v/c_u - 0.60)$ | where $v/c_u \leq 0.60$<br>where $v/c_u \geq 0.60$ | Gan et al. (2003), Branston (1975) |
| Irwin et al. (1961) | $t = t_1 + \alpha(v - c_p)$<br>$t = t_1 + \beta(v - c_p)$<br>$t_1 = t_0 + \alpha c_p$ | where $v < c_p$<br>where $v \geq c_p$ | Gan et al. (2003), Branston (1975) |
| Irwin and Van Cube (1962) | $t = t_1 + \alpha(v - c_p)$<br>$t = t_1 + \beta(v - c_p)$<br>$t = t_2 + \gamma(v - c_s)$<br>$t_1 = t_0 + \alpha c_p$<br>$t_2 = t_1 + \beta(c_s - c_p)$ | where $v < c_p$<br>where $c_p \leq v \leq c_s$<br>where $v \geq c_s$ | Gan et al. (2003), Branston (1975) |
| Smock (1962) | $t = t_0\, exp(v/c_u)$ | | Gan et al. (2003), Branston (1975) |
| Mosher (1963) | $t = t_0 + ln(\alpha) - ln(\alpha - v)$<br>$t = \beta - \alpha(t_0 - \beta)/(v - \alpha)$ | | Gan et al. (2003), Branston (1975) |
| BPR (1964) | $t = t_0\left(1 + \alpha\left(v/c_p\right)^{\beta}\right)$ | | Branston (1975), Gan et al. (2003) |
| Soltman (1966) | $t = t_0 2^{v/c_p}$ | where $v/c_p \leq 2$ | Gan et al. (2003), Branston (1975) |
| Traffic Research Corporation (1966) | $t = t_0 * \left(2 + \sqrt{\alpha^2(1 + v/c)^2 + \beta^2} - \alpha(1 - v/c) - \beta\right)$ | | Branston (1975) |
| Overgaard (1967) | $t = t_0 \alpha^{(v/c_p)^{\beta}}$ | | Branston (1975), Gan et al. (2003) |
| Ayad (1967) | $t = t_0\, exp(v/c - 1)$ | | Ayad (1967) |
| BPR (1974) | $t = t_0(1 + \alpha(v/c_u)^{\beta})$ | | Gesalem and Fillone (2016) |
| Steenbrink (1974) | $t = t_0\,(1 + \alpha(v/c_u)^{\beta})$ | | Gan et al. (2003), Branston (1975) |
| Davidson (1966), Davidson (1978) | $t = t_0\left[1 + \frac{j(v/c)}{(1-v/c)}\right]$<br>$t = t_0\left[1 + \frac{j\mho}{(1-\mho)} + \frac{j(v/c-1)}{(1-\mho)^2}\right]$ | where $v/c \leq \mho$<br>where $v/c \geq \mho$ | Mtoi and Moses (2014) |
| Spiess (1990) | $t = t_0\left(2 + \sqrt{\alpha^2(1 + v/c)^2 + \beta^2} - \alpha\,(1 - v/c) - \beta\right)$ | | Smith et al. (1999), Mtoi and Moses (2014) |
| Akcelik (1991) | $t = t_0\left[1 + 0.25 t_0 + \left[(v/c - 1) + \sqrt{(v/c - 1)^2 + 8\tau \frac{v/c}{t_0 c}}\right]\right]$ | | Mtoi and Moses (2014) |

[1] Majority of the travel time delay function are presented in Table 1 are from the works Branston (1975), Gan et al. (2003), and Mtoi and Moses (2014).

The following notation is used for this entire section of this dissertation:

i. $v$: traffic flow

ii. $t$: travel time at traffic flow $v$

iii. $t_0$: initial travel time at zero traffic flow

iv. $t_1$, $t_2$: travel time at different region

v. $t_0$: initial travel time at zero traffic flow

vi. $c$: link capacity

vii. $c_u$: link ultimate capacity

viii. $c_p$: link practical capacity

ix. $c_s$: link stable capacity

x. $\alpha, \beta, \gamma, j, \mho, \tau$ : Modeling calibration parameters

Campbell (1959) proposed the following two functions (Equations 1 and 2) for use in the Chicago area (Gan et al., 2003). Campbell's (1959) methodology indicates that free-flow travel is constant until the flow reaches 60 percent of the capacity, but his methodology shows a linear relationship when the flow rate is equal to or greater than 60 percent of its ultimate capacity (Gan et al., 2003). From a mathematical standpoint, the first function of Campbell methodology lacks the capability to explain the travel-time variation for congested places or larger cities, where traffic-flow characteristics is non-linear in nature.

$$t = t_0 \quad \text{where } v/c_u \leq 0.60 \qquad \text{(Equation 1)}$$

$$t = t_0 + \alpha(v/c_u - 0.60) \quad \text{where } v/c_u \geq 0.60 \qquad \text{(Equation 2)}$$

According to Branston (1975) and Gan et al. (2003), one of the earliest assignment models was developed by Irwin et al. (1961). Irwin et al. (1961) proposed a simple travel-time flow relationship (Equations 3-5) based on two regions and practical capacity. Their proposed model was convex linear for both the feasible region and the overloaded region (Branston, 1975; Gan et al., 2003).

$$t = t_1 + \alpha(v - c_p) \quad \text{where } v < c_p \qquad \text{(Equation 3)}$$

$$t = t_1 + \beta(v - c_p) \quad \text{where } v \geq c_p \qquad \text{(Equation 4)}$$

$$t_1 = t_0 + \alpha c_p \qquad \text{(Equation 5)}$$

Later, Irwin and Van Cube (1962) revised the methodology (Equations 6-10) by adding practical capacity and steady-state capacity (Branston, 1975). This method include linear-regression fitting for the difference between a link's flow per lane and the practical capacity. Based on Gan et al. (2003) and Irwin et al. (1961), model can easily be applied to a majority of the highway-assignment procedure. From the results of Irwin and Van Cube (1962) and Irwin et al. (1961), the model fitted the actual condition (Branston, 1975; Gan et al., 2003). However, Branston (1975) pointed out that this methodology has difficulty creating a linear relationship to the link characteristics at the location of the discontinuity at practical capacity.

$$t = t_1 + \alpha(v - c_p) \quad \text{where } v < c_p \qquad \text{(Equation 6)}$$

$$t = t_1 + \beta(v - c_p) \quad \text{where } c_p \leq v \leq c_s \qquad \text{(Equation 7)}$$

$$t = t_2 + \gamma(v - c_s) \quad \text{where } v \geq c_s \qquad \text{(Equation 8)}$$

$$t_1 = t_0 + \alpha c_p \qquad \text{(Equation 9)}$$

$$t_2 = t_1 + \beta(c_s - c_p) \qquad \text{(Equation 10)}$$

Smock (1962) developed the earliest curvilinear exponential $t/t_o$ function (Equation 11) for the Detroit area (Branston, 1975; Gan et al., 2003). A heuristic, iterative, capacity-restraint assignment was tested by the author. The model predicted that significantly fewer assigned volumes exceeded the steady-state capacity (Branston, 1975).

$$t = t_0\, exp(v/c_u) \qquad \text{(Equation 11)}$$

Mosher (1963) suggested one logarithmic function (Equation 12) and one hyperbolic function (Equation 13) (Branston, 1975; Gan et al., 2003). From a mathematical standpoint, for both functions, if the steady-state capacity is less than $\alpha$, then the functions create a computational problem or infinite travel time (Gan et al., 2003). Branston (1975) pointed out that this model is not suitable for iterative assignment.

$$t = t_0 + ln(\alpha) - ln(\alpha - v) \qquad \text{(Equation 12)}$$

$$t = \beta - \alpha(t_0 - \beta)/(v - \alpha) \qquad \text{(Equation 13)}$$

The most widely used $t/t_o$ function is the BPR function (Branston, 1975). The standard BPR function is shown in Equation 14. Later, the BPR modified this function as shown in Equation 15.

$$t = t_0\left(1 + \alpha\left(v/c_p\right)^{\beta}\right) \qquad \text{(Equation 14)}$$

$$t = t_0\left(1 + \alpha(v/c_u)^{\beta}\right) \qquad \text{(Equation 15)}$$

There are numerous studies about the BPR's $t/t_o$ functions. The original BPR curve was developed by fitting a polynomial equation based on a highway's speed-flow relationship (Transportation Research Board [TRB], 1985; Mtoi and Moses, 2014). According to Spiess (1990), the BPR function is simple and very convenient. However, Spiess (1990) studies presented some inherent drawbacks of BPR functions: 1) a very high $\alpha$ value slows down the convergence for the highway-assignment procedure and can create numerical problems; 2) the functions are very sensitive to FFS; i.e., the slightest change for the FFS may shift the highway assignment to a different path; and 3) the process is computationally expensive.

Davidson (1966, 1978) developed a $t/t_o$ function (Equation 16) based on the queuing theory (Gan et al., 2003). This function has one serious flaw. The function is not capable of estimating travel time when the volume is greater than 1 (Mtoi and Moses, 2014). The inconsistency for this method is discussed by Golding (1978), Akcelik (1991), Kalaee (2010), and Mtoi and Moses (2014).

$$t = t_0\left[1 + \frac{j(v/c)}{(1 - v/c)}\right] \qquad \text{(Equation 16)}$$

Akcelik (1991) proposed a time-dependent form (Equation 17) of the Davidson function which encompass the intersection delay (Mtoi and Moses, 2014). This function may be applied to volume/capacity ($v/c$) ratios above and below 1. This function can be applied to any facility (Smith et al., 1999). Dowling et al. (1998) reported that Akcelik's functions provide more accurate and faster estimates than the BPR method. The Akcelik function is of interest because it is not followed by a smooth curve like other functions. However, this model generated scattered results than the other existing models (Dowling and Skabardonis, 2006).

$$t = t_0\left[1 + 0.25t_0 + \left[(v/c - 1) + \sqrt{(v/c - 1)^2 + 8\tau\frac{v/c}{t_0 c}}\right]\right] \qquad \text{(Equation 17)}$$

Spiess (1990) presented a conical congestion function (Equation 18). He defined certain conditions that need to be satisfied for given $t/t_o$ functions to meet equilibrium assignment algorithm (Horowitz, 1991); these conditions were discussed earlier in the chapter.

$$t = t_0\left(2 + \sqrt{\alpha^2(1 + v/c)^2 + \beta^2} - \alpha\,(1 - v/c) - \beta\right) \qquad \text{(Equation 18)}$$

The first four conditions can be observed for the BPR functions (Spiess, 1990). The last three conditions were applied by Spiess (1990) to overcome the three drawbacks of the BPR functions. Eventually, Spiess' (1990) conical $t/t_o$ function improved the assignment convergence better than the BPR function (Spiess, 1990). According to Horowitz (1991), the first six conditions are met with the BPR functions. He claims that the second condition should be revised. The functions should generate a realistic value at zero volume, and volume at capacity. He also claims that third and seventh conditions are not necessary. However, he finds that Spiess' (1990) conical $t/t_o$ function has better, consistent results than BPR. At the same time, Horowitz (1991) utilized a least-square method to approximate $\alpha$ and $\beta$ parameters based on a highway-capacity manual (Smith et al., 1999).

In this section, the existing methodologies and state-of-the-art best practices in $t/t_o$ prediction has been presented. In the following sections, overall gaps and issues in the existing methods and review of literature are presented.

## 1.4. Gaps in the Current Research

The Literature Review indicates that there are certain gaps in the existing information about the travel-time flow relationship. Broadly, these issues can be classified in following six categories, which has been discussed in details in below.

i. Always strictly increasing assumption issues.
ii. Stochasticity and non-linearity issues.
iii. Theoretical-aspect and binding-constraint issues.
iv. Computational inability issues.
v. Subjectivity of capacity estimation.
vi. Undefined free-flow speed (FFS) consideration.

### 1.4.1. Issues with Always Strictly Increasing Assumption

Based on current practice and the theoretical aspect, to make the highway-user equilibrium-assignment procedure convergent, the delay functions need to be strictly increasing. It was expected that the $t/t_o$ for a given *v/c* ratio is stochastic in nature. This expectation partially outperforms the existing travel-time flow theory. According to TRB (1975), existing theory states that traffic conditions for a given road segment are stationary and that drivers behave the same way, on average, with the same average conditions (Kalaee, 2010). In a real situation for the Albuquerque metropolitan area in New Mexico, as presented in Figure 1, when the $t/t_o$ is stochastic in nature for a given, discrete *v/c* ratio, a strictly increasing relationship may not be observed. This situation is described in the case of a system-wide model. Furthermore, the study area was narrowed down the problem with 50 freeway sections in Los Angeles, California. It is proven in Chapter 7 that $t/t_o$ predictions are stochastic in nature for freeway.

Most of the modeling functions are strictly increasing and significantly sensitive with the increased *v/c*. If someone captures the stochasticity of $t/t_o$ by likelihood or probabilistic approximation, then the expected mean of the $t/t_o$ for a given *v/c* ratio might not be always increasing. To have a convergent solution for an equilibrium highway-assignment algorithm, an increasing curve is always required; therefore, most of the existing $t/t_o$ models are sensitive, especially when the *v/c* is greater than 1, i.e., in a congested situation. Therefore, the first assumption for this research is that the $t/t_o$ function might not be strictly increasing to follow the user-equilibrium method. The function may increase or decrease, especially when the stochasticity of the $t/t_o$ is very sensitive for a given *v/c* ratio. Therefore, a system should be designed in such a way such that it can capture the natural rise and fall of the $t/t_o$ with respect to the *v/c* ratio.

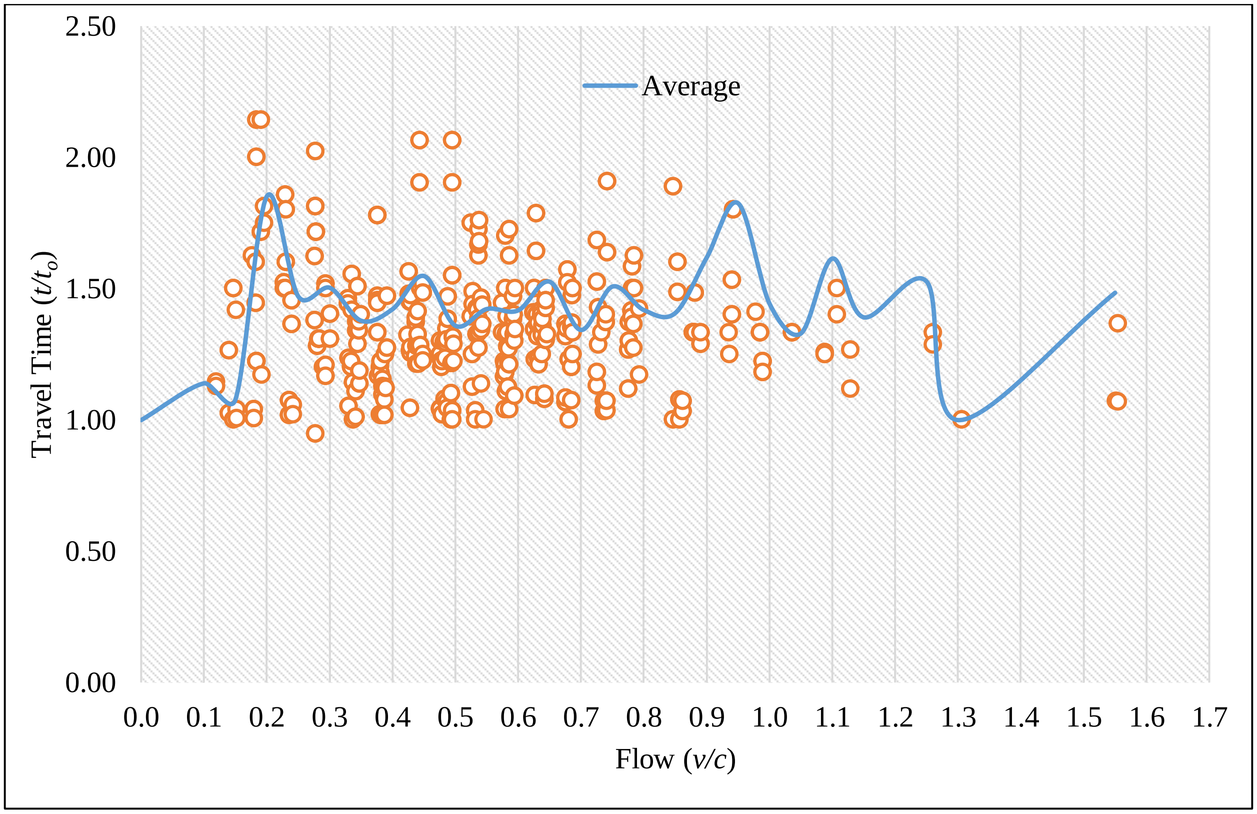


Figure 1. Realized Sensitivity of an Actual Condition

### 1.4.2. Stochasticity and Non-Linearity Issues

Earlier studies did not put enough attention to incorporate stochasticity and probability in their $t/t_o$ function development. The existing methods cannot suitably explain the probability for a $t/t_o$ with a given $v/c$ ratio. The existing models are fitted based on the least-square curve-fitting technique. The random errors generated with the existing models do not follow the assumption of a normal distribution where the random errors should be independent and identically distributed. The proof of this assumption is demonstrated in Section 7.1. Therefore, model may contains uncertainty. In the case of congested places such as Los Angeles, this uncertainty is extremely high. The least-square curve fitting technique may generate a very low value for the coefficient of determinations.

Therefore, a separate model is required to predict the $t/t_o$ for such a case. In order to reduce errors and the variability of the generally least-square fitted model, the mean and variance's chaotic nature need to be incorporated in the developed model.

### 1.4.3. Theoretical Aspect and Binding-Constraint Issues

Several existing $t/t_o$ functions are only limited when the $v/c$ ratio is less than 1, i.e., uncongested condition. Some of them can explain when the $v/c$ is greater than 1, but several existing models are based on an empirical or theoretical approach. The main reason might be the data unavailability. Several existing models are developed based on forcedly fitting the curve by constraining the $t/t_o$ function at zero volume and volume at capacity as presented in Spiess (1990) study. Horowitz (1991) claims to revisit Spiess' (1990) second condition. Horowitz (1991) supports that these constraints should be revisited. Therefore, one of the assumption for this research is that the $t/t_o$ function might not have any binding constraint at a FFS and volume at capacity. The proof of this assumption are presented in Section 3.5.

### 1.4.4. Computational Inability Issues

The literature shows that many models that are not perfectly suited for the highway-assignment procedure. Dividing by zero, generating infinite travel time, or inconsistency with the computation are frequently observed issues that are caused by different functions which outperform many of the developed functions. Therefore, a solution that can overcome this issue may be required.

### 1.4.5. Subjectivity of Capacity Estimation

The $t/t_o$ functions need capacity estimations for input parameters. When the capacity estimation is used to the transportation-planning stages, the estimate is mostly practical and subjective in nature (Branston, 1975). Second, highway-capacity manuals provide guidance to estimate capacity. This guidance is heavily concentrated on an operational aspect. When it comes to planning aspect, numerous assumptions are necessary to replicate the operational condition in the planning model. These assumptions come from various sources, which may include different error types. On the contrary, robust and stable capacity estimations might eliminate the subjectivity of capacity estimation. Therefore, a robust and stable capacity-estimation method may be necessary.

#### 1.4.6. Undefined Free-Flow Speed Consideration

FFS is the last input parameter for $t/t_o$ functions. The FFS has an important role. This FFS is used to calculate the initial FFT and the congested travel time for the highway-assignment steps. Wang and Huegy (2014) indicated that FFS is being used for regional travel demand by utilizing a look-up table. By definition and theoretically, the FFS formulation should relate to the traffic-flow speed while the density is zero. Therefore, the available speed-density model should be investigated, and a methodology should be proposed to find the FFS accordingly.

In the earlier section, different issues and problems, which need to be addressed in $t/t_o$ formulation have been presented. Considering these six specific issues, the research problem statement is defined. It is described in the following section.

### 1.5. Problem Statement

Until millennium, although various researchers and practitioners devoted much attention to developing the travel-time flow relationship, advances to improve the travel-time flow relationship is not substantial. One of the main reason might be the unavailability of concurrent travel time and traffic count data. Beside this, practitioners are widely using simple model. Because, even a simple model is computationally expensive in the highway assignment stage in travel demand model. Furthermore, $t/t_o$ functions are comprised of numerous factors. Therefore, after the millennium, most of the researchers and practitioners might be utilizing simple model and developing calibration/transferrable parameters based on the existing models. Therefore, a new $t/t_o$ model might be required which can move forward the scientific community.

The $t/t_o$ dynamics are a function of the $v/c$ ratio with two input parameters: FFS and capacity. There are specific problems with the current $t/t_o$ functions.

First, the current $t/t_o$ models with respect to $v/c$ are mostly represented by increasing function and, therefore, are significantly sensitive with the increased $v/c$. In practice, to achieve a convergent solution for the user-equilibrium method, the current models are forced to grow in a strictly increasing condition. In some cases, models are constrained at zero volume and capacity. To overcome the user-equilibrium assignment method drawbacks, the functions need to be strictly increasing. In reality, strictly increasing nature may not seem perfect. An always increasing function may not be possible. The $t/t_o$ dynamics shows stochastic in nature, implying that the $t/t_o$ growth with respect to the $v/c$ is not strictly increasing when the likelihood or probabilistic uncertainty is incorporated. The $t/t_o$ uncertainty may follow a logistic growth model which can explain the natural trends of $t/t_o$ function. If someone wants to incorporate the stochasticity for the $t/t_o$ functions' growth, then fitting a curve using the state-of-the-art best practices may not be suitable. Thus, to capture the stochastic nature of the data, a new function may be needed. Therefore, it is necessary to provide new theoretical and methodological advancement for $t/t_o$ dynamics so that, for a given transportation system's $t/t_o$ stochasticity can be eliminated.

Second, computing the $t/t_o$ by depending on $v/c$ is stochastic in nature. Several $t/t_o$ functions are fitted with the least-square curve technique where random errors portion of the model violates the assumption of normality. That way, the least-square curve fitting may produce larger errors and variances for a model. For example, if someone classify the data as an infinitely small, discrete *v/c* value instead of a continuous *v/c* value, then, for a given discrete *v/c* value, the $t/t_o$ is highly sensitive and stochastic in nature. For a given *v/c*, the distribution of the modeling errors may not normal. For a given transportation system of a network or a functional class, the $t/t_o$ may behave differently, especially when the stochasticity of the $t/t_o$ function for a given discrete *v/c* may have a significant influence. The $t/t_o$ functions are significantly sensitive for a given *v/c* ratio. An existing deterministic method may fail to represents this condition. Existing methods did not put suitable attention to incorporate this $t/t_o$ stochasticity based on the data's statistical distribution. Therefore, it is necessary to find knowledge which can resolve issues two and three, particularly performing a stochastic, probabilistic approximation based on the statistical distributions for the given data.

In several modeling functions, there are seven conditions of Spiess (1990) discussed earlier that might need to be met to become a good candidate as a $t/t_o$ function to meet equilibrium assignment algorithm. To meet the theoretical aspect, the $t/t_o$ function are forced to fit at zero volume and volume at capacity. The idea with the existing practices is that $t/t_o$ functions should generate one when the volume is zero and generate two when the volume is equal to the capacity. To remove these issues, Horowitz's (1991) constraint assumption was adopted.

Third, $t/t_o$ functions need an input parameter for capacity estimation. When capacity estimation is at the transportation-planning stage, capacity estimations are mostly practical capacity and are subjective in nature for a given link. Highway capacity manuals (HCM) provide guidance for estimating capacity. This guidance is heavily concentrated on an operational aspect. When it comes to planning, numerous assumption types are necessary to replicate the operational condition in the model. These assumptions come from various sources which may include different error types. On the other hand, robust and stable capacity estimations might eliminate the subjectivity of capacity estimation. Therefore, a robust and stable capacity-estimation method may be necessary.

Fourth, FFS is the last input parameter for $t/t_o$ functions. The FFS has an important role. This speed is used to calculate the initial FFT and the congested travel time for highway-assignment steps. Wang and Huegy (2014) indicated that FFS is being used for regional travel demand by utilizing a look-up table. By definition and theoretically, the FFS formulation should relate to the traffic-flow speed while the density is zero. Therefore, the available speed-density model should be investigated, and a methodology should be proposed to find the FFS accordingly.

There are two questions that need to be answered in order to obtain a solution for the previously discussed issues. First, how to get the prior information about a Bayesian model update for a given $t/t_o$? Second, how and where to get the data that include the concurrent travel time and volume for a given location? The next section details these two problems.

First, the Bayesian modeling technique requires prior information about a distribution. In order to estimate probability for the Bayesian model discussed with research issues two and three earlier, prior belief/information/knowledge about the data are necessary. Prior is the information about the $t/t_o$ parameters for a given $v/c$ ratio that needs to be learned before the Bayesian model update or experiment. Using historical data, prior information on the $t/t_o$ for a given $v/c$ can be estimated. The expected distribution of the prior information is stochastic, meaning that the random error is not normally distributed. From this random error, stochasticity may arise and should be accounted for in the prior estimation. Thereafter, question was how someone could estimate the prior belief when there was uncertainty associated with the mean and variance for the $t/t_o$ with a given $v/c$. On such an occasion, a stochastic method, the Program Evaluation and Review Technique (PERT), is well known and widely used in the field of operation research and project management. It was expected that PERT should have the capability to estimate the expected, most likely mean $t/t_o$ for a given $v/c$ based on the historical optimistic, pessimistic, and most likely mean $t/t_o$.

Second, finding suitable concurrent travel-time data and traffic volume is crucial for regional transportation planning. Many larger agencies are collecting real-time or near real-time travel-time or traffic-volume data, but the concurrent, real-time or near real-time travel-time and traffic-volume data are rare. In order to develop a $t/t_o$ function, researchers are always eager to have concurrent data. Because resources are constrained, it is necessary to develop a system such that, in the absence of data, it might be useful to represent concurrent data. If an agency is counting the traffic volume for a given highway's link segment, then the developed system should have the capabilities to collect travel time. Using this concept, a new technology called virtual sensor methodology (VSM) is a current interest among researchers. Morgul et al. (2014) used MapQuest and Bing Maps to collect travel-time data, but the authors did not include other crowdsource web services, such as Google, OpenStreetMap, and HERE. Some issues with VSM are MapQuest, Bing, Google, and HERE's are licensing and permission requirements. On the contrary, OpenStreetMap is the only free and open crowdsource service. OpenStreetMap does not have any limits to collect travel time data. Considering these issues, it may be necessary to extend the knowledge of VSM beside Morgul et al.'s (2014) work. A detail literature review is included at Chapter 2 to support this statement.

## 1.6. Assumptions

Based on the previous discussions, it was necessary to have some specific hypotheses or assumptions before obtaining a solution. In order to come up with a novel solution, there are certain assumptions that need to be considered:

i. The $t/t_o$ functions should not always be increasing. The $t/t_o$ function should follow natural rise and fall trends/growth which could be increasing or decreasing for a given $v/c$ ratio (Proof of this assumption is presented in Section 7.2 and 7.3).

ii. Adopted Horowitz's (1991) second assumption is that the functions should generate a realistic value at zero volume and capacity (Proof of this assumption is presented in Section 3.5).

iii. The first derivative of the function may exist and follow the delay curve's natural growth. This assumption is actually the complementary of the first assumption.

iv. The random error due the $t/t_o$ model for a given $v/c$ ratio is not normally distributed (Proof of this assumption is presented in Section 7.1.5).

## 1.7. Research Objectives

This dissertation's main goal was to propose a new methodology for a highway-link, $t/t_o$ function formulation based on a stochastic Bayesian model update and a market-adopted, logistic-growth curve-fitting technique. In order to attain this goal, the research was focused on six specific research objectives:

i. Approximate a new freeway travel-time congestion/highway assignment model based on knowledge borrowed from the logistic growth-mapping with eliminating stochasticity from the model.

ii. Provide a new approximation methodology for the traffic delay's prediction using the knowledge from Bayesian modeling technique.

iii. Provide a new, integrated logistics and quantile traffic-flow model for traffic flow prediction and capacity estimations.

iv. Investigate deterministic speed-density models and propose a methodology for FFS estimation.

v. Approximate of the informative prior distribution and its uncertainty of historical $t/t_o$ using a performance evaluation and review technique.

vi. Extend the knowledge about the VSM to collect travel-time data from crowdsource services.

### 1.8. Expected Significances and Contributions

There were six expected contribution for this research: 1) propose a new scientific methodology and theoretical foundation for the freeway $t/t_o$ stochastic approximation; 2) propose a new methodology for the $t/t_o$ dynamics/chaotic-behavior prediction; 3) investigate speed-density models, and propose a better methodology for FFS estimation and speed predictions; 4) develop a model that can approximate stable capacity; 5) approximate the stochasticity for an informative prior distribution of $t/t_o$ and its uncertainty; and 6) extend the VSM to collect travel-time data from crowdsource web services.

Two indirect, expected contributions for this research were generated: 1) reviewing the literature about crowdsourcing and VSM; and 2) developing a web app, several tools, and an automated workflow to collect and analyze travel-time data.

## 1.9. Expected Limitations

There were two expected limitations which could be raised while conducting this research: 1) concurrent travel time and volume might be collected from two different sources/methods, and 2) crowdsource data are contingent based on approving the crowdsource agencies' terms and conditions.

# 2. LITERATURE REVIEW

The Literature Review is divided into seven broad categories. The identified gaps, existing problems, research needs, and research motivation for the research are presented in seven different sections.

i. Existing travel-time flow relationship.
ii. Existing prediction methodology.
iii. Existing informative prior approximation methodology.
iv. Existing data-collection methodology.
v. Existing virtual sensor methodology.
vi. Existing capacity estimation.
vii. Existing FFS methodology.

First section, existing travel-time flow relationship was discussed in Chapter 1 (Section 1.3). Second, fourth, and fifth Sections are presented in Chapter 2. Third section, existing informative prior $t/t_o$ methodology is presented in Chapter 3 (Section 3.3). Sections sixth and seventh are presented in Chapters 6 and 7, respectively.

First section presents why there is a need of a new methodological and theoretical enhancement to predict the $t/t_o$. Second section presents why there is a need of a probabilistic Bayesian approach to predict the $t/t_o$ for a given $v/c$ ratio. Third section explains why the performance-evaluation and review technique to estimate the prior belief about $t/t_o$ was considered. Fourth and fifth sections are tied to the data needed to complete this dissertation. Sections sixth and seventh proposed methodologies for input-parameter estimations of capacity and FFS, respectively.

## 2.1. Existing Prediction Methodology

### 2.1.1. Definition of Travel-Time Prediction

Travel-time prediction (TTP) can be defined as the estimation and prediction of the travel time before a vehicle has traveled on a route of interest: the freeway or an arterial (Billings, 2006). Hao et al. (2009, p.189) defined TTP as "Travel time prediction refers to the calculation of the departure time for the future traffic conditions. Note that the principal difference between travel time estimation and travel-time prediction is that the former is based on the data available for current time instance. Thus, travel time estimation is mostly used for offline fashion. In addition, travel time estimation, by definition, is a technique used in cases where directly measured travel times are not available."

TTP is a critical element for releasing traffic information (Jiang et al., 2014). TTP has been an important research topic for the last two decades (Jiang et al., 2014; Billings, 2006). In order to make more informed, individual departure-time and route selections before trip planning or en-route navigation, advanced traveler information systems (ATIS) enable the inclusion of prevailing and/or predictive travel time (Fei et al., 2011; Zou et al., 2014; Zhang and Haghani, 2015).

Because travel time information is an important element for intelligent transportation systems (ITS), accurately predicting travel is crucial, specifically for the ATIS (Zou et al., 2014; Wu et al., 2004). Freeways travel time estimation and prediction are necessary to implement successful intelligent transportation system (Tak et al., 2014). The TTP is the major element to further the progress of the available of travel-time information application in both the transport and logistics fields (Lin et al., 2005).

The inclusion of accurate travel-time information through the ATIS provides travelers to make informed choices about the departure time, route, and mode (Zhang and Haghani, 2015). TTP is important element to proactively develop advanced traffic-management system (ATMS) strategies (Zhang and Haghani, 2015).

Traffic accidents, weather, and road conditions have different uncertainties, making the TTP very challenging to estimate accurately. It is a known problem (Billings, 2006). However, prediction techniques and mechanism play bigger role in prediction accuracy than any other factors (Ishak and Al-Deek, 2002).

Short-term delivery can be facilitated using the current travel-time information, but considering time-period perspective, long-term TTP is essential (Lin et al., 2005). The area with stable traffic conditions, a reasonable estimation can be utilized, but for an area with more traffic variations, a prediction model is required (Grol et al., 1999). Different factors, such as an incident, may affect the travel time with a higher sensitivity. Since the mid-1980s, short-term travel-time estimation and prediction has become more important (Park et al., 1999).

Research about freeway travel-time estimation is very rich, but the arterial travel time is quite limited (Billings, 2006). The arterial road TTP is more challenging compared to the freeway TTP because the arterial-road section contains a congestion and delay caused by entering vehicles from cross streets (Billings, 2006).

Mendes-Moreira et al. (2015, p.428) wrote- “Long-term bus travel time prediction (i.e., the prediction of the duration of bus trips several days ahead) is very important for the planning activities in freight and transport companies (e.g., definition of the schedules for trips and drivers). Long-term should be, in this context, understood as the prediction of a travel time for a single future trip that is expected to occur in a given future timestamp. Not much attention has been dedicated to this problem."

### 2.1.2. Factors Affecting Travel-Time Prediction

Wu et al. (2004, p.276) said, "Travel-time calculation depends on vehicle speed, traffic flow, and occupancy, which are highly sensitive to weather conditions and traffic incidents. These features make travel-time predictions very complex and difficult to reach optimal accuracy. Nonetheless, daily, weekly, and seasonal patterns can still be observed at a large scale."

Travel time may vary along the 1) individual locations, 2) short road sections, 3) long road sections, transit routes or trips, 4) corridors, 5) subareas, 6) regional networks, and 7) multimodal analysis (TRB, 2008) and may be based on different time periods, such as small time periods (peak or off-peak period), hourly, daily, and seasonal (TRB, 2008; May and Montgomery, 1984). Travel time may vary by area type such as urban and rural types (TRB, 2008). May and Montgomery (1984) have listed travel-time variations in several broad categories:

i. Inter-vehicle variation (car characteristics, driving style, traffic lanes chosen, and traffic conditions).
ii. Inter-period variation (traffic volume, traffic composition, specific incidents, weather, time of day, day of week, time of year, and secular trends).
iii. Inter-route variation (route or link).

TRB (2013) has defined following seven broad categories, which can affect the travel time.

i. Traffic-control devices

ii. Daily/seasonal variation

iii. Special events

iv. Physical capacity

v. Weather

vi. Incidents

vii. Work zones

**2.1.3. State-of-the-Art Best Practices for Travel-Time Predictions Method**

In this section, state-of-art best practices of TTP method are presented. Based on the findings of Lin et al. (2005), Chen et al. (2015), and Zhang and Rice (2003), most short-term TTPs can be listed as shown in Table 2.

Fangfang et al. (2008, p.3753) said, "Travel time can be estimated or predicted based on the data collected by fixed detectors (e.g. loop detectors, video cameras, radar detectors) or non-fixed detectors (e.g. AVI, AVL, GPS, Cellular Phone Tracking). At present, researchers have shown great interest in developing prediction models using data from probe vehicles equipped with GPS."

Billings (2006) studied the application of the auto regressive integrated moving average (ARIMA) models to an urban, arterial road's TTP with a case study of a section of Minnesota State Highway I-94. He used the global positioning system (GPS) probe-vehicle method to collect data and applied ARIMA because of the data's non-stationarity. His study indicated that the ARIMA model has potential and is efficient for short-term prediction; the model can be applied to other urban areas. He found that the predicted travel time replicates the observed condition. The research concluded that the ARIMA model is good for the higher speed limit. He also found that a lower speed limit, shorter link distance, and high cross traffic affected the prediction's performance. Zhang and Haghani (2015, p.309) said, "ARIMA models provide interpretable parameters with straightforward model structures. They can make very good predictions when traffic shows regular variations."

Table 2. List of TTP Methods

| Number | Acronym | Method |
|---|---|---|
| 1 | ATHENA | ATHENA |
| 2 | ARIMA | Auto Regressive Integrated Moving Average |
| 3 | DLM | Bayesian Inference-Based Dynamic Linear Model |
| 4 | Clustering | Clustering |
| 5 | DynaMIT | Dynamic Network Assignment for the Management of Information of Travelers |
| 6 | EES | Extended Exponential Smoothing |
| 7 | GBM | Gradient Boosting Method |
| 8 | Ensemble | Heterogeneous ensembles |
| 9 | KF | Kalman Filter |
| 10 | K-NN | K-Nearest Neighbor |
| 11 | LM | Linear Model |
| 12 | LOKRR | Local Online Kernel Ridge Regression |
| 13 | NN | Neural Network |
| 14 | - | Regression |
| 15 | - | Simple Statistical Model |
| 16 | ST-D | Space-Time Diurnal Method |
| 17 | SNN | Spectral Neural Network |
| 18 | SSNN | State Space Neural Network |
| 19 | SVR | Support Vector Regression |

Elhenawy et al. (2014) developed a data-clustering method using *K*-means and a genetic programming algorithm to predict the dynamic travel time for freeways. The authors' proposed algorithm attained a 25 percent reduction in the prediction error for the instantaneous average as well as a 76 percent prediction-error reduction for the historical average on a congested day for a given spatiotemporal variation in road-segment speeds. Elhenawy et al. (2014, p.87) said, "The instantaneous method is very simple where it assumes the segment speed does not change during the entire trip time." Their case study was based on INRIX traffic data.

Fei et al. (2011) presented a Bayesian dynamic linear model (DLM) approach for real-time, short-term freeway TTP. They used loop-detector data of an I-66 segment in northern Virginia. They concluded that the prediction is accurate and reliable for the travel time. Zhang and Haghani (2015) employed a gradient-boosting method (GBM) to predict the travel time for a freeway in Maryland. They used INRIX data.

Haworth et al. (2014) studied local, online kernel-ridge regression for forecasting the urban travel time with a case study of London's road network. Data were only collected in the daytime by using a camera. Hayworth et al. (2014) found that this method can replicate better forecasts compared to the historical average. They also revealed that this method cannot forecast an abnormal congestion scenario.

Jiang et al. (2014) studied the predictability for an urban road network based on multi-source data, which are floating-car data and fixed-detector data. The results were compared against the VISSIM simulated results, and the authors found that the absolute relative error of the travel time for both cases increases up to 13.4 percent. They established a prediction model based on KF using real-time floating-car and loop data. They found that multi-source prediction is better than single-source data. Huang et al. (2013) studied an urban expressway in Guangzhou for TTP based on a fuzzy, adaptive KF. They used real-time detection data. Their study's results indicated that the KF model had a better ability to estimate travel time than the conventional approach.

Zou et al. (2014) proposed a space-time diurnal (ST-D) method to predict the freeway travel time by considering spatial and temporal travel time. They collected data on a segment of US-290 in Houston, Texas. This segment displayed a traffic-flow pattern. Their research included six automatic vehicle identification (AVI) readers and vehicles with toll tags to collect the travel time. They found that the ST-D method is robust compared to the traditional vector-autoregressive models. Their results indicated that the short-term travel time can be reliably predicted with this method.

Ben-Akiva et al. (2001) presented the Dynamic Network Assignment for the Management of Information of Travelers (DynaMIT), a model-based, real-time system that can be used to guide travelers. DynaMIT has two functions: to estimate the travel time based on real-time, dynamic traffic assignments as well as to predict the current and future traffic conditions. Ben-Akiva et al. (2001) used historical and real-time travel-time data to develop the system. The prediction methodologies of them were based on the departure time, pre-trip route and mode choice, and en-route choice decision.

Mendes-Moreira et al. (2015) wanted to improve the accuracy of the long-term TTP for public transportation using heterogeneous ensembles method. Their model can predict the travel time in ahead of three days. They utilized random forest, projection-pursuit regression, and support vector machines method. They used the data from Sociedade de Transportes Colectivos do Porto, Portugal. The results showed that this method can produce higher accuracy and robustness compared to state-of-the-art learners. The authors acknowledged that, considering the complexity of this system, this method could be less attractive for some users although it enhances the results.

Park et al. (1999) utilized a spectral neural network (SNN) for real-time forecasting on US-290 in Houston, Texas. They predicted the travel time for one-to-five time periods ahead for the Houston area. They used AVI of the TranStart to collect the data. Their research discovered that an SNN can produce similar results, such as the modular NN, and outperformed the NN's conventional methods. Overall, it revealed that SNN performed better than KF the exponential smoothing model, the historical profile, and the real-time profile.

Wu et al. (2004) performed support vector regression (SVR) methods to predict travel time and substantially reduced the prediction error compared to baseline predictors. They proved the SVR method's feasibility and applicability with better performance. Fangfang et al. (2008) predicted link travel time using extended exponential smoothing (EES) and a KF in dynamic networks. The authors used the probe-vehicle data on an urban network. They integrated exponential smoothing with the KF method.

Lint et al. (2005) studied freeway, accurate travel-time forecasting using state-space neural networks (SSNN) with missing data. They showed that the SSNN method produced accurate and robust predictions for real and synthetic data.

Tak et al. (2014) proposed a multi-level K-NN algorithm and data fusion method to predict the real-time travel time. They used congested travel-time data that were collected from a Korean expressway. The proposed algorithm could predict travel time with an error of less than 5 minutes. Besides the aforementioned prediction methods, there are simple statistical techniques, such as a linear-regression model based on historical travel time or current travel time. Lin et al. (2005) said that the statistical technique has been used extensively in the field of travel-time estimation, prediction, and modeling. Tak et al. (2014, p.1861) stated, "The statistics-based methods have problems in accuracy, and some others are limited to predicting the travel time only during short time interval."

In this section, state-of-the-art best practices of TTP method has been presented. The summary of current state-of-the-art best practices is presented in Tables 3 and 4.

Table 3. Summary Findings of the TTP Techniques: Part I

| Acronym | Method | Pros-and-Cons | References |
|---|---|---|---|
| ARIMA | Auto Regressive Integrated Moving Average | • Urban Arterial<br>• Very good prediction for freeway | Billings (2006); Zhang and Haghani (2015) |
| DLM | Bayesian inference-based Dynamic Linear Model | • Freeway<br>• Accurate with recurrent and non-recurrent traffic conditions<br>• Reliable for recurrent and non-recurrent traffic conditions | Fei et al. (2011) |
| Clustering | | • Freeway<br>• Simplicity of the model and computationally efficient<br>• Interpretable and provides insight about the critical segments<br>• Lower prediction error compared to an instantaneous algorithm | Elhenawy et al. (2014) |
| GBM | Gradient Boosting Method | • Freeway<br>• Improved prediction accuracy compared to the base model<br>• Considerable advantages for freeway travel-time prediction | Zhang and Haghani (2015) |
| KF | Kalman Filter | • Urban road network<br>• Current state can be predicted from the past state<br>• Able to eliminate the random interference noise<br>• Approximates the real estimate<br>• Fuzzy adaptive KF produced better results than the conventional KF<br>• Good prediction accuracy | Huang et al. (2013) |
| LOKRR | Local Online Kernel Ridge Regression | • Can capture time-varying distribution<br>• Capable to replicate traffic patterns, seasonality, and heteroscedasticity<br>• New traffic data can be incorporated<br>• Accurate forecasts than historical average travel time<br>• Cannot forecast any abnormal congestion scenario<br>• Computationally efficient for a single location<br>• Good adaptability | Haworth et al. (2014) |

Table 4. Summary Findings of the TTP Techniques: Part II

| Acronym | Method | Pros-and-Cons | References |
|---|---|---|---|
| K-NN | K-Nearest Neighbor | • Higher computational efficiency<br>• Accurate prediction<br>• Data-fusion method | Tak et al. (2014) |
| EES | Extended Exponential Smoothing | • Urban arterial road network<br>• Produces acceptable results | Fangfang et al. (2008) |
| SSNN | State-Space Neural Network | • Freeway<br>• Accurate predictions<br>• Robust prediction | Lint et al. (2005) |
| SNN | Spectral Neural Network | • Freeway<br>• Predict one-to-five time periods ahead<br>• Link travel time<br>• Outperformed the conventional NN<br>• Similar results as the modular NN<br>• Overall best compared to KF, exponential smoothing, historical profile, and real-time profile<br>• Accurate and better function approximation | Park et al. (1999) |
| SVR | Support Vector Regression | • Freeway in Taiwan<br>• Substantially reduced the prediction error compared to baseline predictors<br>• Better performance | Wu et al. (2004) |
| ST-D | Space-Time Diurnal Method | • Freeway<br>• Considered spatial and temporal travel time, diurnal pattern, and the non-negativity of the travel time<br>• Robust compared to the traditional vector's autoregressive models | Zou et al. (2014) |
| DynaMIT | Dynamic Network Assignment for the Management of Information of Travelers | • Real time<br>• Modeling-assignment based | Ben-Akiva et al. (2001) |
| Ensemble | Heterogeneous Ensembles | • Public transportation<br>• Long-term travel-time predictions<br>• Higher accuracy<br>• Robust<br>• Mitigate seasonal data | Mendes-Moreira et al. (2015) |

### 2.1.4. Summary

There are numerous methodologies are available to predict the travel time. Each method has its own flaws and prediction errors. However, each method may produce satisfactory results with the tolerance error. Some methods are proven to be robust. Some methods are very complex to apply and are not feasible. It is evident that the probabilistic prediction method did not received proper attention to the researchers and practitioners.

Most existing methods generate single-value outcomes and cannot measure the prediction results' reliability (Fei et al., 2011). In contrast to these methods, according to Gelman et al. (2003), the Bayesian approach can update the knowledge systematically when new observations are available (Kim and Reinschmidt, 2009). The Bayesian approach has the capability to combine all related information in a systematic way (Kim and Reinschmidt, 2009). Bayesian process is a system that sequentially can update the prior knowledge, measure the uncertainty in a probabilistic way, and update the posterior estimates. Therefore, the Bayesian approach to predict the $t/t_o$ for a given $v/c$ was considered for this research.

There are certain advantages for the Bayesian approach:

i. Supports a natural and principled way of integrating prior beliefs (Moriarty, 2015; SAS Institute Inc. [SAS], 2016).
ii. Supports inferences based on the data's conditional treatment (Moriarty, 2015; SAS, 2016).
iii. Follows the likelihood principle (Moriarty, 2015; SAS, 2016).
iv. Generates interpretable answers (Moriarty, 2015; SAS, 2016).

There are certain disadvantages for the Bayesian approach:

i. Computation is expensive (Moriarty, 2015; SAS, 2016).
ii. Estimation of posterior is biased due to prior information (Moriarty, 2015; SAS, 2016).
iii. Produce subjective prior information (Moriarty, 2015; SAS, 2016).

## 2.2. Existing Data-Collection Methodology

### 2.2.1. General

There are many methods available for travel-time data collection. Every method has its own success stories, acceptability, errors, and flaws. Technology is changing, so are the data-collection methods, such as crowdsourcing, aerial surveys, weigh-in-motion, and laser scanning. Therefore, this section reviews recent literature and combines the knowledge about different methods. Studies show that there are five major divisions which may include a total of 24 sub-divisions of travel-time data-collection methodologies. Literature Review reveals a total of 44 factors which might be considered before starting any travel-time data collection.

### 2.2.2. Procedure for Travel-Time Data Collection

Travel time can be measured by traversing the route(s) that connect(s) any two or more points of interest (Federal Highway Administration [FHWA], 1998). According to the FHWA (1998, p.1-6), "Travel time can also be estimated in certain cases by assuming the average speed at a particular point (spot speed) is constant for a relatively short distance (typically less than 0.8 kilometer, or 0.5 mile). The assumption of consistent speeds over a short roadway segment is most applicable to uninterrupted flow facilities (e.g., freeways or expressways) with stable traffic flow patterns. The estimated travel time can be computed using the average spot speed, or time-mean speed, and the roadway segment length. Average or mean travel times are computed from individual travel times by using standard statistical formulas or computer software."

### 2.2.3. Current Practices

Travel time for given pairs of origins and destinations of a road link/route in the transportation network is the time required to travel between the origin and destination for a specific mode choice. In a broader context, the FHWA (1998) defined travel time between any desired two points of a route as the time required to traverse that route. According to that report, travel time is composed of running travel time while the mode of transport is in motion and stopped delay time while the mode of transport is not in motion or stopped. The definition of stopped or delayed travel has been characterized by a speed that is typically less than 5 miles per hour (mph) (FHWA, 1998).

Travel time is the most important measure in the transportation industry. The FHWA (1998, p.1-1) stated, “Travel time is a simple concept understood and communicated by a wide variety of audiences, including transportation engineers, planners, business persons, commuters, media representatives, administrators, and consumers. Engineers and planners have used travel time and delay studies since the late 1920s to evaluate transportation facilities and plan improvements.”

Soriguera et al. (2010, p.1242) said, “Travel time for a road trip is a drivers’ most appreciated traffic information. Measuring travel times on a real time basis is also a perfect indicator of the level of service in a road link, and therefore is a useful measurement for traffic managers in order to improve traffic operations on the network. In conclusion, accurate travel time measurement is one of the key factors in traffic management systems.”

Travel time is an important indicator to define the traffic quality (Jie et al., 2011). Travel time is the most influential parameter for the road user’s decision making. It serve as a bridge among the road users, engineers, planners, researchers, commuters, consumers, and business entities. There is abundant literature, such as Rasouli and Timmermans (2012), Wang and Xu (2011), Morgul et al. (2014), Cambridge Systematics (2013), FHWA (1998), and TRB (2015), that explains the travel time’s importance.

Urban travel time had a wider variation for a region (Jie et al., 2011). The authors concluded that a single or average travel time does not represent a meaningful measure because of the travel-time reliability. Their findings were further supported by Cambridge Systematics (2013) and Morgul et al. (2014). Cambridge Systematics revealed that the average travel time may vary from 0.80 to 1.5 times. Morgul et al. (2014, p.4) said, “Performance measures are defined as indicators of system efficiency. For example, in the context of transportation, travel time variability is an emerging performance measure increasingly used by decision-makers in making many transportation investment decisions. Information on how long it would take to travel between specific points is a vital information for all travelers. Accurate estimation of travel times reflects the system performance based on users’ point of view.”

Travel times are measured with two fundamental sources: 1) directly collected or measured, and 2) estimated using modeling (FHWA, 1998). Morgul et al. (2014) said that there are two distinct approaches for traffic-surveillance methods: 1) road-based technologies (in-road detectors and road-side detectors) and 2) vehicle-based technologies. The authors also grouped the methodology into three groups based on the detection capability: 1) single-point detection, 2) multi-point detection, and 3) area-wide detection.

Travel time is critically important for different perspectives among the road users, planners, transportation engineers, community, commuters, researchers, and business entities. Considering the transportation perspective, the 1991 Intermodal Surface Transportation Efficiency Act (ISTEA) and Moving Ahead for Progress (MAP 21) mandated the use of travel time to measure the traffic congestion, transportation analysis, and transportation decisions’ performance evaluation. There are several technologies for travel-time data-collection techniques; they are listed below from FHWA (1998) and Cambridge Systematics (2012). Each method has the capability of direct measurement or estimation; road-based or vehicle-based technologies; and single-point, multi-point, or area-wide detections. The Literature Review revealed that the travel-time data-collection techniques can be broadly distinguished with five categories: 1) probe-vehicle measure, 2) spot measure, 3) test vehicle, 4) license-plate matching, and 5) emerging technologies. These broad categories have several distinguishable sub-categories which are presented in Table 5. Details about these technologies are discussed in the following sections.

Table 5. Types of Data-Collection Techniques

| No. | Probe-Vehicle Measure | Spot Measure | Test Vehicle | License-Plate Matching | Emerging Technology |
|---|---|---|---|---|---|
| 1 | Automatic Vehicle Identification | Loop Detector | Distance-Measuring Instrument | Video with Manual Transcription | Aerial Surveys |
| 2 | Bluetooth | Magnetic Detector | Global Positioning System | Manual | Laser Scanning |
| 3 | Crowdsourcing | Radar | Manual | Portable Computer | Weigh-in-Motion |
| 4 | Global Positioning System | Video Imaging | | Video with Character Recognition | |
| 5 | Radio Frequency Identification Transponder | | | | |
| 6 | Radio Navigation | | | | |
| 7 | Signpost-Based Automatic Vehicle Location | | | | |
| 8 | Toll-Tag Readers | | | | |
| 9 | Wireless/Cellular | | | | |

It is evident that there are many methods available for travel-time data collection. Every method has its own success stories, acceptability, errors, and flaws. Researchers and practitioners experienced many advantages and disadvantages, depending on different methods. The combined articulation of scholarly articles for different methods needs to be properly addressed so that the new technology can be compared with each method. Therefore, this dissertation also aims to review the recent literature about travel-time data-collection methods, to identify the advantages and disadvantages for each method, and to find the factors that should be considered before collecting travel time.

### 2.2.4. Advantages and Disadvantages of Current Methods

This section lists the merits and demerits for each method found in the scholarly articles and reports. This section presents five major and 24 minor travel-time data-collection methods.

#### 2.2.4.1. Probe-Vehicle Measures

Probe-vehicle method (PVM) is an emerging technique for directly collecting travel time for a given route or a road segment. Cambridge Systematics (2012, p.1-1) explained, "Probe vehicle techniques involve direct measurement of travel time (along a route or point to point) using data from a portion of the vehicle stream." This technology does not require equipment installation and maintenance in the right-of-way (Young, 2007). Herrera et al. (2010, p.569) said, "In the era of mobile internet services, and with the shrinking costs and increased accuracy of GPS, probe based traffic monitoring has become one of the next arenas to conquer by industries working in the field of mobile sensing." There are different kinds of probe-vehicle technologies that may be utilized depending on the different perspectives. These techniques include the technology of automatic vehicle identification, Bluetooth, crowdsourcing, global positioning systems (GPS), radio-frequency identification, radio navigation, signpost-based AVL, toll-tag readers, and wireless/cellular. Young (2007) demonstrated the vehicle-probe technology's advantages and disadvantages.

There are nine major probe-vehicle measure technologies: AVI, Bluetooth, crowdsourcing, GPS, radio-frequency identification transponder, radio navigation, signpost-based AVL, toll-tag readers, and wireless/cellular. These methods' performance is presented in the following sections.

#### 2.2.4.1.1. Global Positioning System

Travel-time data collection using GPS is very rich. There are numerous studies described in the literature. Some of that work is presented here. Herrera et al. (2010) evaluated traffic data obtained with GPS-enable mobile phones. Their studies stated that a GPS-enabled smartphone can exploit the extensive coverage. The results showed that this methodology can accurately estimate the traffic flow's velocity. Sanwal and Walrand (1995), Zito et al. (1995), FHWA (1998), BRW (2000), Rahmani et al. (2010), and FHWA (1998) addressed the extensive advantages and disadvantages of the GPS technologies with travel-time data-collection techniques. The methodology's advantages are 1) fewer staff members needed (FHWA, 1998; BRW, 2000); 2) less human error (FHWA, 1998); 3) relatively portable (FHWA, 1998); 4) accurate (FHWA, 1998; BRW, 2000); 5) accuracy plus/minus 50 meters (Zito et al., 1995); 6) producing a larger dataset (FHWA, 1998); 7) cost effective (BRW, 2000); 8) providing a geospatial location (BRW, 2000); 9) almost continuous data (Rahmani et al., 2010); 10) extensive coverage (Herrera et al., 2010); 11) providing the best vehicle-position location; and 12) providing direct observations of vehicle speeds and travel directions (Zito et al., 1995). There are several issues with the GPS methods: 1) needing big storage (FHWA, 1998); 2) sometimes losing the signal (FHWA, 1998); 3) inherent errors, such as orbit errors, satellite-clock errors, receiver-noise errors, tropospheric and ionospheric errors, coordinate transformations, and selective availability (Zito et al., 1995); 4) multi-path errors (Zito et al., 1995); 5) data-processing errors (Zito et al., 1995); and 6) selective availability when the U.S. Department of Defense degrades the GPS' accuracies for real-time, non-U.S.-military users (Zito et al., 1995).

#### 2.2.4.1.2. Radio Frequency Identification Transponder

According to Herrera et al. (2010, p.569), "Readers located on the side of the road keep record of the time the transponder (i.e. the vehicle) crosses that location. Measurements from the same vehicle are matched between consecutive readers to obtain travel time." Wright and Dahlgren (2001), Herrera et al. (2010), Porter et al. (2011), and Mittal and Bhandari (2012) presented the advantages and disadvantages of Radio Frequency Identification (RFID) travel-time data-collection methodologies. The technique's advantages are 1) accurate (Porter et al., 2011), 2) cost effective (Mittal and Bhandari, 2012 ), 3) uninterrupted communication (Mittal and Bhandari, 2012 ), and 4) working in bad weather (Mittal and Bhandari, 2012 ). The method's disadvantages are 1) the individual vehicle's privacy issues (Porter et al., 2011), 2) limited coverage (Herrera et al., 2010), 3) installation costs for the reader's infrastructure (Herrera et al., 2010), and 4) only providing the travel time for two locations (Herrera et al., 2010), and 5) significantly lower quality than the loop-detector and video-image data (Wright and Dahlgren, 2001).

#### 2.2.4.1.3. Bluetooth

The methodology's advantages are 1) being cost effective (Bhaskar and Chung, 2013; Morgul et al., 2014); 2) provide true travel time (Haghani and Aliari, 2012); 3) having high quality and reliability (Morgul et al., 2014); 4) collecting continuous, larger data set (KMJ Consulting, 2011); 5) being easy to install and operate (KMJ Consulting, 2011), and 6) being a more realistic approach (KMJ Consulting, 2011). The methodology's disadvantages are 1) permanent/mobile/temporary installation (Cambridge Systematics, 2012), 2) unknown mode (Bhaskar and Chung, 2013), 3) limited zonal boundary (Bhaskar and Chung, 2013), 4) multiple matches causing noisy data (Bhaskar and Chung, 2013), 5) missed observations (Bhaskar and Chung, 2013), 6) a short range (Bhaskar and Chung, 2013), 7) expensive data collection (KMJ Consulting, 2011), 8) vulnerable to vandalism (KMJ Consulting, 2011), 9) disappointing data quality (Jie et al., 2011), 10) cannot eliminate outliers easily (Jie et al., 2011), and 11) uncertainty with the Bluetooth device's proper identification (Jie et al., 2011).

#### 2.2.4.1.4. Toll-Tag Readers

The methodology's advantage is the direct measurement of the travel time (Hass et al., 2009). The disadvantages are 1) require permanent, mobile, or temporary installation (Cambridge Systematics, 2012); 2) lower performance (Hass et al., 2009); 3) cannot read all tags (Hass et al., 2009); 4) cannot tag all vehicles (Hass et al., 2009); 5) occasionally failed to read (Hass et al., 2009); 6) problems caused by misaligned toll tags (Hass et al., 2009); and 7) duplicate readings for slow vehicles.

#### 2.2.4.1.5. Wireless/Cellular

The advantages of this methodology are 1) installation of equipment is not applicable (Cambridge Systematics, 2012) and 2) widely available (FHWA, 1998). The method's disadvantages are 1) difficulty distinguishing closely spaced facilities (Young, 2007), 2) no cell-phone technology was successfully assess the signalized arterial (Young, 2007), 3) questionable accuracy (FHWA, 1998), and 4) privacy concerns (Turner, 1995).

#### 2.2.4.1.6. Crowdsourcing

Because crowdsourcing is the main focus for this research, there is a separate section later in this chapter. However, the methodology's advantages and disadvantages are included in this section. The method's advantages are 1) installation is not necessary (Cambridge Systematics, 2012), 2) the only cost is to obtain the data (Cambridge Systematics, 2012), 3) there are significant sources of speed data (Cambridge Systematics, 2012), and 4) there is a lower price (Halder, 2014). The method's disadvantages are 1) mostly dependent on private industries (Rehan, 2015), 2) unknown algorithm (Rehan, 2015), 3) data quality not studied (Rehan, 2015), 4) sparse snapshot of the entire network (Rehan, 2015), 5) fairly accurate with freeway speed but arterials showed less-accurate speed data (Cambridge Systematics, 2012), 6) cannot provide volume data (Cambridge Systematics, 2012), 7) less credible output (Halder, 2014), and 8) privacy and protection of users' personal and sensitive data (Halder, 2014).

#### 2.2.4.1.7. Signpost-Based Automatic Vehicle Location

The methodology's advantages are 1) minimized or eliminated roadside infrastructure (Young, 2007), 2) real-time travel information (Turner, 1995), 3) eliminates human error (Turner, 1995), and 4) directly collects travel time (Hadi and Al-Deek, 2015). The method's disadvantages are 1) the quality depends on the monitoring-vehicle percentage (Young, 2007), 2) there is a higher initial cost (Turner, 1995), 3) the quality depends on the distribution of vehicle types (Young, 2007), and 4) there are errors with the exact location (Turner, 1995).

#### 2.2.4.1.8. Automatic Vehicle Identification

Wright and Dahlgren (2001, p.1) said, "Probe vehicles equipped with electronic tags/transponders communicate with overhead antennas/transceivers to recognize specific vehicles at successive locations. An individual vehicle's travel time is then calculated between these points. Generally, electronic tags are placed in vehicles as part of an electronic toll collection (ETC) system." The methodology's advantages are 1) continuously collecting large quantities of data (FHWA, 1998); 2) minimal human resources (FHWA, 1998); 3) highly accurate (FHWA, 1998); 4) a detection rate of 85 to 99 percent (FHWA, 1998); 5) collecting travel time for a specific lane (FHWA, 1998); 6) a proven technology for area-wide, real-time travel-time data (Tam and Lam, 2015); 7) low operating cost; and 8) good for annual, daily, and real-time data (Tam and Lam, 2015). The method's disadvantages are 1) restricted to a certain number of tags (FHWA, 1998), 2) dependent on infrastructure (FHWA, 1998), 3) increased maintenance costs with clock-drift problems (FHWA, 1998), 4) individual vehicle's privacy issues (FHWA, 1998), and 5) larger data storage needed (FHWA, 1998).

#### 2.2.4.1.9. Radio Navigation

Wright and Dahlgren (2001, p.1) said, “Probe vehicles (generally transit or other fleet) communicate with local radio tower infrastructure. There are no fixed detection points, so the system must calculate the position of the vehicle at different times to determine travel times.” The methodology’s advantage is good Accuracy (FHWA, 1998). The method’s disadvantage is a high initial or capital cost (FHWA, 1998).

### 2.2.4.2. License-Plate Recognition

#### 2.2.4.2.1. Manual

The methodology’s advantages are 1) a larger dataset (Porter et al., 2011; Turner, 1995), 2) accurate (Porter et al., 2011), 3) low initial cost (FHWA, 1998), 4) more representative of the driving population (Hamm, 1993), 5) a representative sample (Turner, 1995), 6) low operating cost (FHWA, 1998), and 7) more cost effective than the manual method (Turner, 1995). The method’s disadvantages are 1) privacy issues (Porter et al., 2011), 2) a high operating cost (FHWA, 1998), 3) questionable accuracy (FHWA, 1998), 4) time consuming (FHWA, 1998), 5) labor intensive (Hamm, 1993), 6) incorrect reading (Turner, 1995), and 7) less practical for high-speed traffic or long roadways (Turner, 1995).

#### 2.2.4.2.2. Video with Manual Transcription

The methodology’s advantage is 1) continuous data (FHWA, 1998). The method’s disadvantages are 1) time consuming (FHWA, 1998) and 2) limited geographic coverage (FHWA, 1998).

#### 2.2.4.3. Spot Measures

##### 2.2.4.3.1. Radar

The methodology's advantages are 1) lane accuracy (Siemens, 2014) and 2) inexpensive installation (Siemens, 2014). The method's disadvantages are 1) reusable (Siemens, 2014) and 2) not responsive to adverse conditions (Siemens, 2014).

##### 2.2.4.3.2. Loop Detector

The methodology's advantages are 1) no privacy issues (Porter et al., 2011), 2) spot measurement (Zhang, 2006), 3) most common (Zhang, 2006), and 4) accurate estimation (Petty et al., 1998). The method's disadvantages are 1) needing permanent installation (Cambridge Systematics, 2012) and 2) accuracy (Porter et al., 2011).

##### 2.2.4.3.3. Magnetic Detector

The methodology's advantages are 1) needing to install sensors into the road-like loop detector (Porter et al., 2011) and 2) being accurate (Porter et al., 2011). The method's disadvantages are 1) very costly (Porter et al., 2011) and 2) not suitable for a large scale (Porter et al., 2011).

##### 2.2.4.3.4. Video Imaging

The methodology's advantages are 1) real-time information (Washburn and Nihan, 1999), 2) a reasonable degree of accuracy (Washburn and Nihan, 1999), and 3) statistically validity (Washburn and Nihan, 1999). The method's disadvantages are 1) unknown accuracy (Washburn and Nihan, 1999), 2) indirect measurement (Washburn and Nihan, 1999), 3) sensitivity to color recognition and resolution (Washburn and Nihan, 1999), and 4) only overall travel times (Turner, 1995).

#### 2.2.4.4. Test Vehicle

##### 2.2.4.4.1. Manual

The methodology's advantages are 1) manually recording the data (Zhang et al., 1997), 2) having a person drive on the road (Zhang et al., 1997), and 3) no special equipment (Hamm, 1993). The method's disadvantages are 1) fewer samples (Zhang et al., 1997), 2) judgmental bias and random human error (Hamm, 1993), 3) a higher cost compared with the amount of data (Hamm, 1993), and 4) not suitable for real-time information (Hamm, 1993).

##### 2.2.4.4.2. Distance Measuring Instrument (DMI)

The methodology's advantages are 1) easy to collect data (Turner, 1995), 2) safer to collect data (Turner, 1995), 3) provides a detailed travel time (Turner, 1995), and 4) cost effective (Turner, 1995). The method's disadvantages are 1) few samples (Turner, 1995) and 2) biased results due to judgment (Turner, 1995).

##### 2.2.4.4.3. Global Positioning System

The methodology's advantages are 1) having lower equipment costs compared with using a direct measuring instrument (Dowling Associates, 1999) and 2) being more portable (Dowling Associates, 1999). The methodology's disadvantages are 1) needing specialized equipment (Dowling Associates, 1999); 2) losing the GPS signal in urban canyons, under trees, or around power lines (Dowling Associates, 1999); 3) frequently losing the signal, creating lost data (Dowling Associates, 1999); and 4) post processing for missing points (Dowling Associates, 1999).

#### 2.2.4.5. Miscellaneous

##### 2.2.4.5.1. Weigh-in-Motion

The methodology's advantage is high accuracy (FHWA, 1998). The method's disadvantages are 1) more suitable for larger vehicles and trucks (FHWA, 1998) and 2) limited applications (FHWA, 1998).

##### 2.2.4.5.2. Laser Scanning

The methodology's advantage is the capability to collect data for individual vehicles (FHWA, 1998). The method's disadvantage is the lack of literature.

##### 2.2.4.5.3. Aerial Surveys

The methodology's advantages are 1) good accuracy (Dowling Associates, 1999), 2) wide-area coverage (Dowling Associates, 1999), and 3) fair variability (Dowling Associates, 1999). The method's disadvantages are 1) being labor intensive (Dowling Associates, 1999) and 2) not being feasible without advance digital processing (Dowling Associates, 1999).

### 2.2.5. Summary

Every method has its own success stories, acceptability, errors, and flaws. A complete list of factors is presented in Table 6. Most methods were mainly tested on freeways and a few urban arterials. Researchers and practitioners compared their methodology based on accuracy, cost, maintenance, coverage, comparison to other methods, communication, staff requirements, direct/indirect measurements, human errors, portability, geospatial data availability, big data, continuous data, storage, inherent errors, data-processing errors, installation costs, real-time information, quality and reliability, mode availability, noise, privacy, vulnerability to vandalism, outliers, uncertainty, permanent, mobile, temporary, duplicities, performance failures, market availability, data types, credibility, roadside infrastructure, locations, operating costs, time, reusability, weather effects, area coverage, special equipment, bias, sample sizes, easiness to collect data, missing data, and individual vehicle perspectives.

Table 6. Influential Factors for Data-Collection Methodology

| Sl. | Factors | | Factors |
|---|---|---|---|
| 1 | Accuracy | 23 | Market availability |
| 2 | Area coverage | 24 | Methods |
| 3 | Bias | 25 | Missing data |
| 4 | Big data | 26 | Mode availability |
| 5 | Communication | 27 | Noise |
| 6 | Continuous data | 28 | Operating cost |
| 7 | Cost | 29 | Outliers |
| 8 | Coverage | 30 | Permanent/mobile/temporary |
| 9 | Credibility | 31 | Portability |
| 10 | Data-processing error | 32 | Privacy |
| 11 | Direct/indirect measurement | 33 | Quality and reliability |
| 12 | Duplicities | 34 | Real-time information |
| 13 | Easiness to collect data | 35 | Reusability |
| 14 | Weather effects | 36 | Roadside infrastructure |
| 15 | Performance failures | 37 | Sample size |
| 16 | Geospatial data availability | 38 | Special equipment |
| 17 | Human errors | 39 | Staff requirements |
| 18 | Individual vehicle perspective | 40 | Storage |
| 19 | Inherent errors | 41 | Time |
| 20 | Installation cost | 42 | Type of data |
| 21 | Location | 43 | Uncertainty |
| 22 | Maintenance | 44 | Vulnerable to vandalism |

Reviewing the existing methods leads to select a suitable methodology to collect travel time in a cheaper, free, and open way. Research reveals that the new crowdsourcing technology has the potential for travel-time data collection when considering economics as well as free, open, and big data. Therefore, review of literature related to crowdsourcing was further extended.

## 2.3. Crowdsourcing

### 2.3.1. General

The VSM is new technology for the transportation industry. This technique does not require any physical deployment of sensors, detectors, or equipment. Literature about the virtual sensor crowdsource technology in transportation is scarce. There is a strong need to rich the virtual sensor crowdsourcing literature. Therefore, this section describes the crowdsource literature by presenting the combined knowledge about crowdsourcing. This research reviews the literature about the virtual sensor crowdsource methodology and its potential for transportation planning. Studies indicate that the findings could be very helpful for the transportation agencies that have resource shortage for their transportation planning and operations.

The VSM is a new technology for the transportation industry. The term “virtual sensor” can be addressed as the “internet as sensor” type of crowdsourcing. A virtual sensor does not require any installations or any physical deployment of instruments, sensors, or equipment. There is not enough literature showing this methodology’s pros and cons. The VSM did not received suitable attention by the researchers or practitioners in the transportation industry. Dennies et al. (2015, p.2) said, “Any research project that includes a literature review leverages the combined intelligence of the authors of previous works; however, literature review is not generally thought of as crowdsourcing. Presumably, literature review and similar research tasks are not considered crowdsourcing because the contributors of knowledge are passive in the process. Nonetheless, many activities that are commonly considered crowdsourcing also involve extracting data from passive providers.”

According to Cambridge Systematics (2012, p.1-9), “Crowd-sourcing is the newest technology, which involves obtaining real-time traffic congestion information from drivers’

GPS-enabled mobile phones. Crowd-sourcing relies on a large number of users and is primarily to support personal navigation systems and mobile applications." The authors have listed several current applications of crowdsourcing, such as Google Maps, Google Latitude, TomTom Mapshare, Trapster, Waze, and INRIX. Crowdsourcing is gaining more popularity, is becoming well known, and has garnered more attention from researchers and practitioners in the last few years (Halder, 2014; Dennies et al., 2015; Misra et al., 2014). To obtain data, transportation agencies are often contract third-party vendors (Dennies et al., 2015).

### 2.3.2. Definition of Crowdsourcing

Misra et al. (2014, p.2) said, "Crowdsourcing is an example in which an organizer or an organization is able to use the network of collaborators to solve a problem that would otherwise be cost- or labor-intensive, or in which within a defined organization the expertise is unavailable or insufficient."

Dennies et al. (2015, p.V) defined crowdsourcing as follows: "Crowdsourcing involves leveraging the combined intelligence, knowledge, or experience of a group of people to answer a question, solve a problem, or manage a process."

### 2.3.3. Current Crowdsourcing Services

Before utilizing any VSM, understanding the crowdsource-service methodology, terms, and conditions, as well as the usage limit, is necessary. There are 23 crowdsource services and apps which are presented in Tables 7 and 8. The tables are self-explanatory, and they include the five major crowdsource services and vivid data elements.

Among the 23 listed services, Google Latitude and Waze are no longer available. Navteq and Nokia moved to HERE. Cellint, Garmin Viago, Graphhoper, INRIX, Sygic, TomTom, Tapster, and Trafficast do not provide free services. NextBus only provides transit services.

Scout is part of Telenav, which is essentially using OpenStreetMap services. To use ArcGIS, an organizational account with a license is required. Bing, Google, HERE, MapQuest, and OpenStreetMap have the potential for further evaluations. These five services allow access fully or partially to the public information through application programming interface (API). Except for OpenStreetMap, users need to abide by the foundation' terms and conditions. The five major services are discussed in the following sections.

#### 2.3.3.1. Bing Maps

Sinani (2015) said, "Bing Maps by Microsoft provide a collection of Application Programming Interfaces (APIs) that adds mapping capabilities to location-aware applications. The APIs, open to the public since their first beta release in 2005, can query for location or business by address or coordinates. They enable searching for routes, including traffic information; searching for geometrical shapes of geographical entities such as countries, regions, and other smaller administrative divisions. Consumers of Bing Maps APIs can geocode address, or reverse geocode coordinates via automated jobs."

Table 7. Existing Web Services' API

| No. | Services | Service Data Providers | Provide Travel Time | Provide Distance | Provide Incident Data | Limit (Free) | Modal Information | Time Unit | Distance Unit | Response Type | Scripting | Services | Require Key |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Bing | Microsoft | Yes | Yes | Yes | 10,000 monthly | Driving, Walking, Transit | Seconds | Kilometer | XML or JSON | AJAX v7, Silverlight, REST Services, WPF Control, .NET | Locations API, Elevation, Imagery, Routes, Traffic | Yes |
| 2 | Google | Google | Yes | Yes | Yes | 2,500 daily, 100 per query, 100 per second | Driving, Walking, Transit, Bicycling | Minutes | Kilometer | XML or JSON | JavaScript, Android SDK, iOS SDK | Directions, Elevation, Geocoding, Geolocation, Place, Roads, Time Zone | Yes |
| 3 | HERE | Nokia | Yes | Yes | Yes | 100,000 transactions per month for 3-month evaluations | Car, HOV, Pedestrian, Truck, Transit | Seconds | Meter | XML or JSON | JavaScript, REST API, Mobile SDK, Platform Extensions, Legacy | Map Tile API, Map Image API, Venue Maps API, Routing API, Geocoder API, Batch Geocoder API, Places API, Traffic API, Weather API, and Transit API | Yes |
| 4 | MapQuest | AOL | Yes | Yes | Yes | 15,000 transactions per month | Driving, Walking, Bicycling, Rail, Bus | Seconds | Mile | XML or JSON | JavaScript, iOS, Android | Geocoding, Mapping, Directions, Traffic, Search | Yes |
| 5 | OpenStreetMap | OpenStreetMap Foundation | Yes | Yes | No | No limit with request but not the excessive use | Driving, Biking, Walking | Seconds | Meter | JSON, GPX | Java, C/C++, C-Sharp, Scala, Ruby, Python | Routing, Geocoding | No |

Table 8. Existing Web Services' API

| No. | Services | Limit (Free) |
|---|---|---|
| 6 | Cellint | Private proprietorship |
| 7 | Garmin Viago | Private proprietorship |
| 8 | Google Latitude | No longer available |
| 9 | Graphhoper | Not free |
| 10 | INRIX | Not free |
| 11 | Navigon | Phone app |
| 12 | Navteq | Now HERE |
| 13 | NextBus | Transit |
| 14 | Nokia | Now here |
| 15 | Scout | Part of Telenav family |
| 16 | Sygic | App |
| 17 | Telenav | Following OpenStreetMap |
| 18 | TomTom | Private proprietorship |
| 19 | TrafficCast | Private proprietorship |
| 20 | Trapster | Private proprietorship |
| 21 | Waze | By Google earlier |
| 22 | Yahoo | Now here |
| 23 | ArcGIS | Need organizational account |

The Bing Maps Services API provides a Representational State Transfer (REST) interface to perform tasks such as creating a static map with pushpins, geocoding addresses, retrieving imagery metadata, or creating a route (Microsoft, 2015). Bing offers five APIs: Locations, Elevations, Imagery, Routes, and Traffic. The Routes API assists with finding a route. It provides a walking, driving, or transit route for given locations. The Traffic API provides traffic information along a route. Using this API, any incident details, incident severity, and the incident's location and type can be obtained.

The Bing API can communicate through several protocols, such as the Asynchronous JavaScript and extensible markup language (AJAX), Silverlight, REST services, Windows presentation foundation (WPF) control, .NET, and the simple object-access protocol (SOAP) based service. AJAX can be used for web applications and Windows store apps by using the

JavaScript language. The .NET approach is for Windows store apps. The SOAP-based service is for faster mobile applications.

Bing maps' Rest services is a collection of RESTful web services that can be accessed through a Uniform Resource Locator (URL) and the hypertext transfer protocol's (HTTP) GET and POST methods. The response format for Bing maps is extensible markup language (XML) or JavaScript object notation (JSON). Up to 25 waypoints can be added to a single route request. Between any pair of waypoints, 10 intermediate points can be added. Bing map services offers driving, walking, and transit routes. The number of transactions is unknown, but 10,000 free transactions per month can be used by the developer (DuVander, 2012).

Morgul et al. (2014) studied the VSM based on Bing maps' API and MapQuest API traffic data. The data were compared to the loop-detector and electronic tag-reader data. The authors investigated the travel-time reliability for the New Jersey Turnpike. Their results showed that Bing maps' data could be good for travel-time reliability analysis. They inferred that Bing map could be choice for an alternative traffic-monitoring method for transportation agencies with budget constraints.

#### 2.3.3.2. Google Maps

This API service can be run using JavaScript, Android Software Developer Kit (SDK), and Internet Operating System (iOS) SDK. Using the Google Maps API, driving, walking, bicycling, and transit travel time can be found. Google offers many APIs: Directions, Elevation, Distance Matrix, Geocoding, Geolocation, Place Web Services, Roads, and Time Zone.

Google has a distance-and-direction matrix API to calculate the travel time. The distance-and-direction matrix API can be accessed with the HTTP GET and POST methods. The response format for Google maps is Extensible Markup Language (XML) and JSON. The standard usage limit is 100 requests for a query, 100 requests for a second, or 2,500 daily. Wang and Xu (2011) have studied O-D implications, advantages, and implementation by using the Google Maps API. The authors have developed a desktop tool to complete the task by calling the Google Maps API. What they found is that, with various methods, the travel time leads to different accessibly patterns. They also said that using the Google Maps API is not free of concern because a user has no control over its quality and no editing rights.

According to Google (2015), "Use of the Google Maps Distance Matrix API must relate to the display of information on a Google Map; for example, to determine O-D pairs that fall within a specific driving time from one another, before requesting and displaying those destinations on a map. Use of the service in an application that doesn't display a Google map is prohibited." Google predicts travel time based on the historical time-of-day and day-of-week.

**2.3.3.3. HERE**

The HERE routing API calculates a route for a set of waypoints and is capable of calculating a matrix of routes between many start points and destinations (HERE, 2015). The HERE maps API uses JavaScript, the REST API, Mobile SDK, platform extensions, and legacy scripting to communicate. HERE REST API has the services of Map Tile, Map Image, Venue Maps, Routing, Geocoder, Batch Geocoder, Places, Traffic, Weather, and Transit.

The HERE Maps API can be evaluated with a 90-day trail of the entire platform, providing the opportunity for 100,000 transactions per month. With an authorized, valid application code and application identity, an HTTP GET or POST request's response format would be in the XML or JSON format. HERE traffic data are part of the FHWA's National Performance Measuring Research Data Set (NPMRDS) program which collects travel time continuously for national freeways and some arterials.

#### 2.3.3.4. MapQuest

MapQuest provides different services, such as geocoding, mapping, directions, traffic, and search. The MapQuest Directions API allows access to the patented routing algorithm via a simple HTTP request (MapQuest, 2015). The output-response format is XML or JSON. MapQuest offers 15,000 transactions per month even though it was unlimited access previously. The services include different travel modes, which includes driving, walking, bicycling, rail, and bus. MapQuest is the only organization which provides open and limited licensed data to the user simultaneously (Morgul et al., 2014).

#### 2.3.3.5. OpenStreetMap

OpenStreetMap router services is an open source which can be accessed with no limit on responses. OpenStreetMap has numerous web services, such as via-route, nearest, locate, table, match, and trip. An HTTP request can be made with JSON or global position system exchange format (GPX) output. OpenStreetMap services uses Java, C/C++, C Sharp, Scala, Ruby, and Python. OpenStreetMap services only includes driving, biking, and walking modes.

### 2.3.4. Crowdsourcing for Transportation

Halder (2014) presented the historical development of crowdsourcing in different fields. Dennies et al. (2015) grouped the crowdsource transportation-system data into four categories. Dennies et al. (2015) also provided the example of application, which are included below.

i. Third-party aggregated data (historical and real-time traffic conditions such as speed, vehicle counts, people movement, travel behavior, and mobility pattern).

ii. Social-media engagement (questions and concerns, communication systems for special events and disruptions).

iii. Internet as a sensor (speed estimation, mining social-media data, accident and prevention reporting, traffic-information dissemination, and predicting traffic spikes for special events).

iv. Dedicated platforms (automated vehicle location, pavement condition, parking management, and O-D data).

There are number of case study applications listed in Dennies et al. (2015) and Misra et al. (2014). A complete list can be reviewed in Dennies et al.'s (2015) work. Cambridge Systematics (2012), Dennies et al. (2015), and Misra et al. (2014) presented a brief overview of crowdsourcing and its use for transportation. They presented the current practices of crowdsourcing in the transportation industry. Misra et al. (2014) presented the feedback-based crowdsourcing systems, such as the SeeClickFix, FixMyStreet, PublicStuff, FixMyTransport, Shareabouts, StreetBump, and OneBusAway systems, and their case examples. Dennies et al. (2015) presented the City of Austin Social Networking and Planning Process (SNAPP); Twitter application during Hurricane Sandy in the New York Metropolitan Transportation Authority (MTA); and the Florida Department of Transportation (FDOT) and Waze Partnership project. Some of the case applications are presented in Table 9.

Most of the case-study applications are based on the social-media type of crowdsourcing. Morgul et al. (2014) presented some travel-time reliability applications using the VSM. Elhenawy et al. (2014) and Fei et al. (2011) analyzed the INRIX crowdsource data. Dailey and Cathey (2006) deployed a VSM based on transit probes in an operational traffic-management system (traveler information). Dailey and Cathey (2006) used a dedicated platform for the VSM when the AVL instrumentation was deployed. Their research considered vehicle-tracking data that were obtained on the transit vehicle. Their research indicated that the VSM can be utilized to capture recurring and non-recurring congestion. They stated that roadways which does not have any present speed traps can be instrumented without installing equipment on the roadway. They have conducted another research to estimate speed on arterial roads for the Washington Department of Transportation (WSDOT). Dailey and Cathey (2002) illustrated their past, successful implementation of VSMs for freeways and arterials. Their research recommended that judiciously selected virtual sensors on the freeways and arterials can extend the limited availability of speed traps. Kurkcu et al. (2015) extended Morgul et al.'s (2014) work with Twitter. They deployed the VSM for web-based, real-time transportation data collection and analysis for incident management. Their system collected incident data from Twitter. Most studies did not consider the VSM to collect travel-time data. Morgul et al. (2014) was the first to work with developing a VSM to collect the travel time with MapQuest and Bing Maps.

Table 9. Case Applications of Crowdsourcing

| Case | Purpose | Reference |
|---|---|---|
| SeeClickFix | To report issues and find information. Interactive website reporting for the citizens. | SeeClickFix (2015), Misra et al. (2014) |
| FixMyStreet | Report road maintenance issues. | Misra et al. (2014) |
| PublicStuff | Reporting public issues through mapping. | Misra et al. (2014), PublicStuff (2015) |
| FixMyTransport | Internet service could help people over the edge from grumbling about a problem. | Misra et al. (2014), FixMyTransport (2015) |
| Shareabouts | Uses maps to generate user feedback about the preferred location of facilities and amenities. | Misra et al. (2014) |
| Street Bump | To detect potholes and other street hazards as people drive around the city. | Misra et al. (2014) |
| OneBusAway | Created to address the reliability issues with the on-time performance of transit systems in Seattle, Washington | Misra et al. (2014) |
| City of Austin Social Networking and Planning Process (SNAPP) | Increase quality and quantity of public participation | Misra et al. (2014) |
| New York MTA Use of Twitter During Hurricane Sandy | Communicating with the public | Dennies et al. (2015) |
| Florida DOT-Waze Partnership project | Feed from FDOT's 511 system | Dennies et al. (2015) |

### 2.3.5. Summary

Finding suitable travel-time data is crucial for regional transportation planning. Suitable travel time data is unavailable for most of the smaller- or medium-size agencies. Therefore, this research presented the combined knowledge of crowdsource services that can be utilized for travel-time data collection. The literature indicated that crowdsource technology is going to be the next pioneer for travel-time data collection. By calling router, OpenStreetMap can be utilized flexibly compared to the other vendor's rigid terms and conditions. The Literature Review revealed that there are several crowdsourcing options that can be achieved with VSM. Current studies show that crowdsource data are considerably good in replicating the true travel-time estimation even though quality and noise are a bigger issue for the data.

# 3. RESEARCH METHODOLOGY

## 3.1. General

Overall, the research methodology has been divided into several steps which are presented in Figure 2. These steps are marked as 1-9 in Figure 2. Steps 1 and 2 are tied to research objective 6, which indicates the data-collection efforts. Steps 3 and 6 involve data generation for different objective analyses, displaying how the data will be utilized in this research. Immediately after this step, research objectives third and fourth have been completed. Step 4 is tied to research objective five. Step 5 is tied to research objective two. Steps 7-9 are tied to the first research objective. The reseach objectives was accomplished in descending order. Each step has been discussed thoroughly as follows.

At first and second step, VSM was developed to collect the travel time data from web services. Validations of VSM with the NPMRDS, manual crowdsourcing, a smartphone with a test vehicle, and a web-based crowdsource application was conducted. A geographic modeling tool for smartphone data conversion was developed. Automated workflow management for the NPMRDS data processing was created. After thorough investigation of the aforementioned sources and depending on its applicability, an alternative data sources was identified. At this stage, research objective sixth was completed. A detail description of this steps has been explained in Section 3.2 and Chapter 4.

Once the data was collected, then the input parameters: 1) steady-state capacity and 2) FFS was approximated. This two parameters are shown in curly bracket as an external input for this research as presented in Figure 2. At this step, a new integrated logistics and quantile traffic-flow model for traffic flow prediction was proposed and then steady-state capacity values were approximated. Capacity estimation merits a full separate Chapter. Therefore, Chapter 5 briefly described how the freeway steady-state capacity was observed for this research. Similarly, for the second input parameter, which is FFS are described separately in Chapter 6. In Chapter 6, an investigation about the deterministic speed-density model was conducted, a new methodology was proposed, and formulated guidelines for FFS approximation. At this stage, research objectives third and fourth were accomplished.

At Steps 3-5, evaluation of the performance of the historical $t/t_o$ was completed. Bayesian predication model was formulated at these steps. This three steps are an integrated approach of Bayesian modeling techniques. A brief discussion of this methodology are presented in Sections 3.3 and 3.4. Results and findings of this steps are presented in Chapter 7. Research objectives second and fourth were accomplished at this steps.

Steps 6-9 are tied to the logistic growth modeling techniques. At this stage, a new, link $t/t_o$ function based on the knowledge borrowed from the market-adopted, logistic-growth-curve technique and the Bayesian model update was established. A brief discussion of these steps are presented in Section 3.5. Results and findings are presented in Chapter 8. At this stage, first research objective was completed.

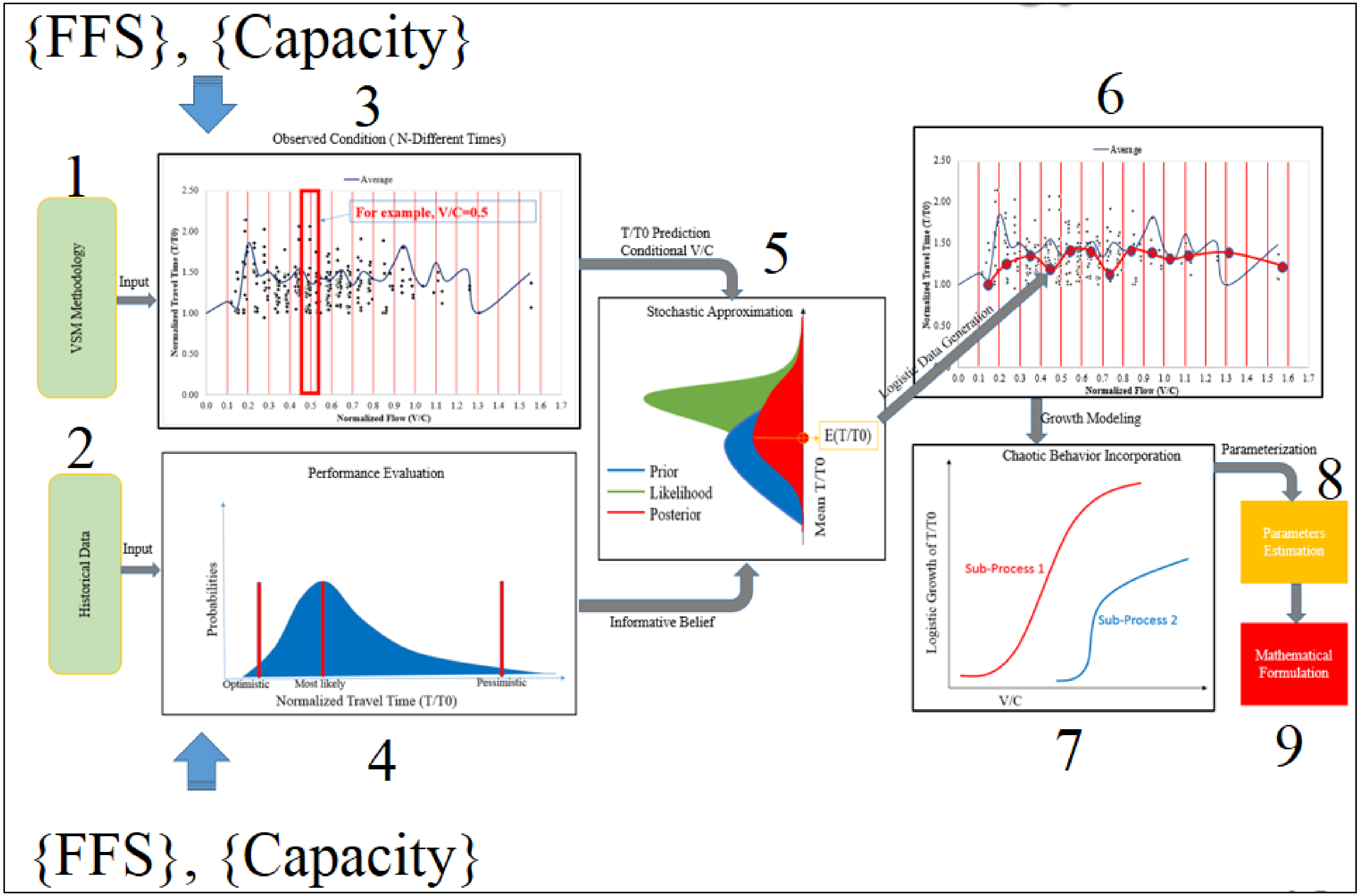


Figure 2. Research Methodology

### 3.2. Development of Virtual Sensor Methodology

Finding suitable, concurrent travel-time data and traffic-volume is crucial for regional transportation planning. However, they are not readily available for the transportation agencies. There are many larger agencies which are collecting real-time or near real-time travel-time or traffic-volume data. Concurrent real-time or near real-time travel-time and traffic-volume data are rare. In order to develop $t/t_o$ function, researcher are always eager to have concurrent data; those data are obviously necessary. Because of resource constraints, it is necessary to develop a system such that, in the absence of local data, it might be useful to collect concurrent data. For example, if an agency is counting the traffic volume for a highway's given link segment, then the developed system should have the capabilities to collect travel time. Using this concept, a new technology called the VSM is a current interest among researchers. One such work is that of Morgul et al. (2014). They have collected travel time using MapQuest and Bing Maps, but they did not include other crowdsource web services such as Google, OpenStreetMap, and HERE. Some issues with the VSM are the licensing and permission requirements for MapQuest, Bing, Google, and HERE crowdsource data. Only OpenStreetMap is a fully free and open crowdsource service. OpenStreetMap does not have any limits for each request to collect travel time. Considering these issues, it was necessary to extend the knowledge of VSM beside the work of Morgul et al. (2014).

A VSM is proposed in Figure 3. Figure 3 shows how these procedures will work overall. Each sections is described in the following sections.

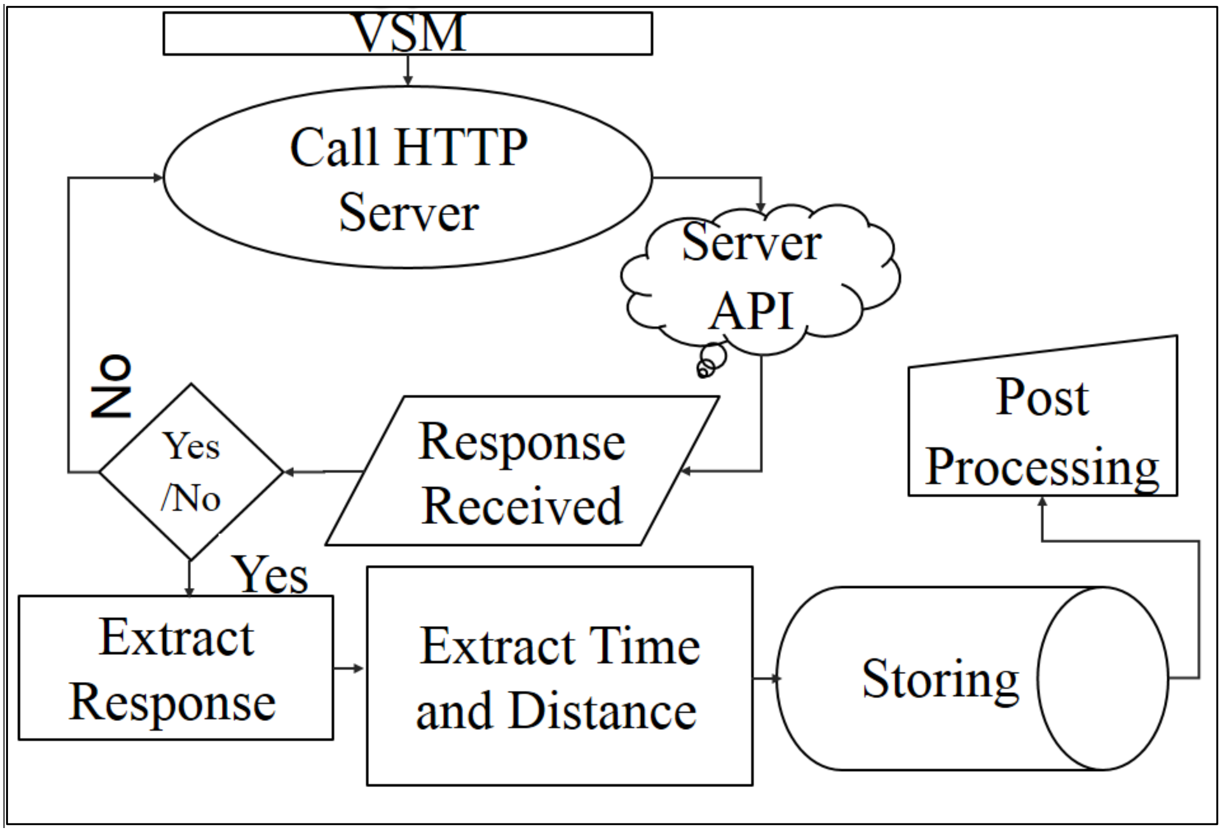


Figure 3. Virtual Sensor Methodology

### 3.2.1. Call the HTTP Web-Server API

Using the vendors' appropriate API calling, the respective HTTP protocol server can communicate the desired travel-time and distance response. In Google, the distance-and-direction matrix API can be accessed with the HTTP GET and POST methods. The Bing API can communicate with several protocols, such as AJAX, Silverlight, REST services, Windows Presentation Foundation (WPF) Control, .NET, and SOAP-based service. AJAX v7 can be used for web applications and Windows store apps using the JavaScript language. Bing Maps' Rest services is a collection of RESTful web services that can be accessed through a URL as well as the HTTP GET and POST methods. The HERE Maps API uses JavaScript, the REST API, Mobile SDK, platform extensions, and legacy scripting to communicate. The MapQuest Directions API allows users to access the patented routing algorithm with a simple HTTP request. With OpenStreetMap, an HTTP request can be made with the JSON- or GPX-formatted output. The HTTP server's communication protocol is different for different vendors. Each vendor has different platforms/language for this communication. Therefore, five different macros/programs using the SAS statistical software can be developed so that this macro can communicate with the respective server. Due to the lack of permission from the respective vendors, only the OpenStreetMap service was utilized.

### 3.2.2. Response Received

Each request's response is in XML or JSON format for all vendors. The proposed macro could store the JSON-format output in a temporary/permanent file location. For a given origin and destination, the macro generated following information.

i. Name of service (OpenStreetMap).
ii. Link's unique ID/corridor number.
iii. Date and time of data collection.
iv. Five minutes epoch number.
v. JSON output as a variable.

### 3.2.3. Extract Response

While extracting JSON responses, the macro searched for different keywords, such as "Travel Time" and "Travel Distance," depending on the scripting language for the JSON response.

### 3.2.4. Extract Travel Time and Distance

Once the extracted location for the travel-time and travel-time distance identification were confirmed, the travel time and distance were observed. Proper quality was maintained for accurate results.

### 3.2.5. Database Storage

For a given link and response, the extracted data were stored in a database by the macro. For cross checking, in addition to the master database, each response file was stored with a date and time stamp.

### 3.2.6. Post Processing

Post processing and additional cleaning was required after the data collection.

## 3.3. Performance Evaluation of Historical Prior

Most existing methods generate single-value outcomes and cannot measure the prediction results' reliability (Fei et al., 2011). In contrast to this existing method, according to Gelman et al. (2003), the Bayesian approach can update the knowledge systematically when new observations are available (Kim and Reinschmidt, 2009). The Bayesian approach has the capability to combine all related information in a systematic way (Kim and Reinschmidt, 2009). This process is where a system can sequentially update the prior knowledge, can measure the uncertainty in a probabilistic way, and can update the posterior estimates. The statistical measure of Bayesian inferences is based on Bayes' theorem which is expressed in Equation 19. In Bayesian method, the target is to measure the posterior probability of parameter $\theta$ for a given dataset or model $D$, where $p(D|\theta)$ is the conditional probability, or likelihood, of model $D$ for a given parameter, $\theta$; $p(\theta)$ is the knowledge before starting the experiment or Bayesian modeling; $p(D)$ is the marginal probability of the model or dataset; and $p(\theta|D)$ is the posterior probability. In this case, the goal is to find the probability of the $t/t_o$ for a given $v/c$. Replacing this equation by $\theta$ as $t/t_o$ and $D$ as $v/c$ becomes Equation 20,

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \qquad \text{(Equation 19)}$$

$$p(t/t_o|v/c\,) = \frac{p(v/c\,|t/t_o)p(t/t_o)}{p(v/c\,)} \quad \text{(Equation 20)}$$

To approximate the posterior distribution of the $t/t_o$ for a given *v/c*, one must first have knowledge about the prior probability *p(t/to)*. Bayesian prior should come from any earlier knowledge, beliefs, or thoughts before starting the Bayesian modeling or experient. The Bayesian modeling technique requires prior information about a distribution. In order to obtain the probability estimation with the Bayesian model, prior belief/information/knowledge about the data is necessary. Prior information is the belief about the $t/t_o$ parameters for a given discrete *v/c* ratio that needs to be known before the Bayesian model update or experiment. Using historical data, Bayesian prior about the $t/t_o$ for a given *v/c* can be estimated. The question is how to estimate Bayesian prior when there is uncertainty associated with the mean and variance for the $t/t_o$ for a given *v/c*; the expected distribution of the prior information is not normally distributed. Thus, uncertainty may arise and needs to be accounted during the prior estimation. On such an occasion, a stochastic method called the Program Evaluation and Review Technique (PERT) is well known and widely used in the field of operation research and project management. PERT might have the capability to estimate the expected, most likely mean $t/t_o$ for a given *v/c* based on the historical optimistic, pessimistic, and most likely mean $t/t_o$.

PERT's probability density function (PDF) originated from a beta distribution (Malcolm et al., 1959). The probability of the generalized beta function presented in Equation 21 for a given random, $t/t_o$ would be $(y=t/t_o)$, where $\alpha$ and $\beta$ are the shape parameters of the beta distribution and *(U, V)* is the domain of *y*.

$$f(y) = \frac{\Gamma(P+Q)(y-U)^{P-1}\,(V-\,y)^{Q-1}}{\Gamma(P)\Gamma(Q)\,(V-\,U)^{P-Q-1}};\; U \le y \le V;\; P, Q > 0 \quad \text{(Equation 21)}$$

This general form of the beta distribution is not a standard beta distribution which may need to be bound by *(U, V) = (0, 1)*. In this case, the standard beta distribution is shown in Equation 22.

$$f(y) = \frac{\Gamma(P+Q)y^{P-1}\,(1-\ y)^{Q-1}}{\Gamma(P)\Gamma(Q)}, \quad 0 \le y \le 1, \qquad P, Q > 0 \qquad \text{(Equation 22)}$$

If $t/t_o$, $y$ is a beta distribution with bounds $U$ and $V$, it can be transformed to a standard beta distribution with a variable, $Z$ by, normalization using Equation 23 (Stackexchange, 2015).

$$Z = \frac{y-U}{V-U} \qquad \text{(Equation 23)}$$

Based on Malcolm et al.'s (1959) pragmatic postulation, the mean and variance for the random $t/t_o(y)$ would be estimated as presented in Equations 24 and 25. Here $U$, $M$, and $V$ are the subjective "optimistic," "most likely," and "pessimistic" $t/t_o$ estimates, respectively, displayed in Figure 4. The mean and variance can further be related to Equations 26-27 in order to estimate parameters $P$ and $Q$.

$$\mu = \frac{(U+4M+V)}{6} \qquad \text{(Equation 24)}$$

$$\sigma^2 = \frac{(V-U)^2}{36} \qquad \text{(Equation 25)}$$

$$\mu = \frac{P}{P+Q} \qquad \text{(Equation 26)}$$

$$\sigma^2 = \frac{PQ}{(P+Q)^2(P+Q+1)} \qquad \text{(Equation 27)}$$

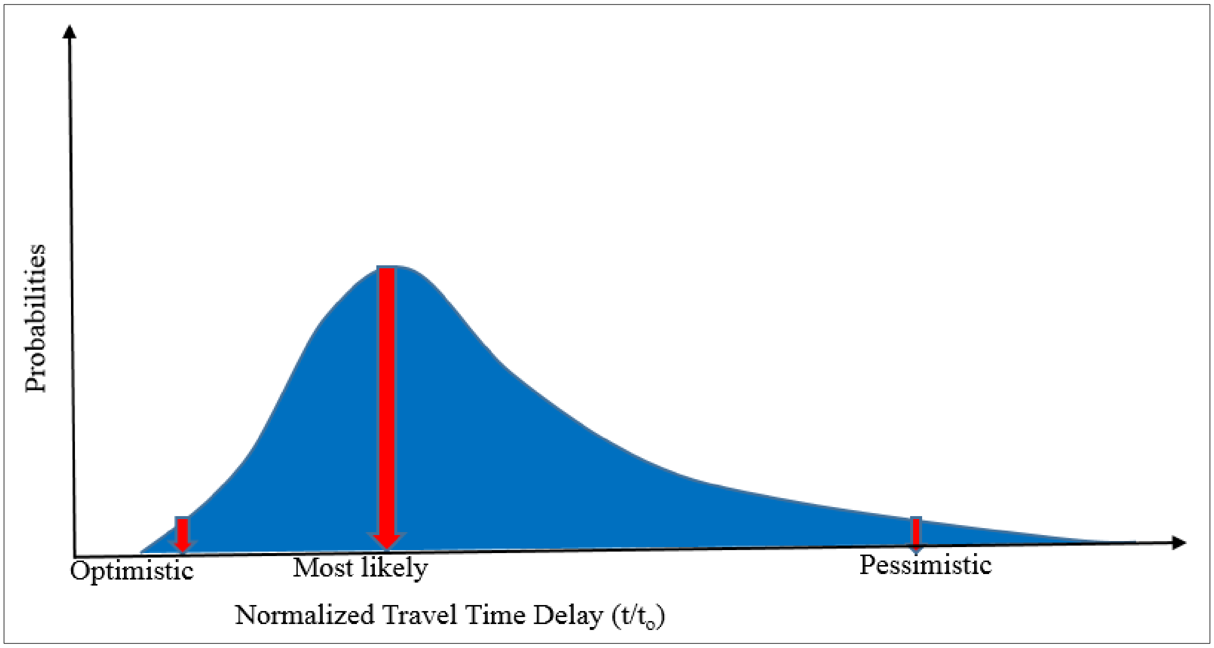


Figure 4. Distribution of the Prior Information Compiled from Malcolm et al. (1959)

For a generalized beta distribution defined on interval *[U,V]*, the mean and variance would be shown in Equations 28 and 29 (Stackexchange, 2011).

$$Mean, \mu = \frac{UQ + VP}{P + Q} \qquad \text{(Equation 28)}$$

$$Variance, \sigma^2 = \frac{PQ(V - U)^2}{(P + Q)^2(P + Q + 1)} \qquad \text{(Equation 29)}$$

By inverting Equations 28 and 29, the PDF-function shape parameters can be observed in Equations 30-32.

$$P = \gamma \frac{\mu - U}{V - U} \qquad \text{(Equation 30)}$$

$$Q = \gamma \frac{V - \mu}{V - U} \qquad \text{(Equation 31)}$$

$$\gamma = \frac{(\mu - U)(V - \mu)}{\sigma^2} - 1 \qquad \text{(Equation 32)}$$

The overall procedure for prior estimation can be followed with this proposed algorithm.

i. Step 1: Classify the historical larger dataset into the desired, discrete dataset by breaking down the *v/c*. *D* is a historical dataset, and *x* as *v/c* is a continuous random variable *(0, **R**)*, where ***R*** is a real number. Now, *x* needs to be stratified based on a desired increment of *x* (i.e., increment $\delta x=0.05$). To do further operation, the corresponding *v/c* value needs to rounded to the nearest increment of $\delta x$. Now, the continuous variable, *x*, becomes a real discrete set of $\{0, 0.05, 0.10, 0.15, 0.20, \ldots, 1, \ldots n\}$. The value of *n* could be observed from the given dataset. Figure 5 displays how each dataset can be generated for a given *x*. Thereafter, dataset *D* includes different a subset $\{D_0, D_{.05}, D_{0.10}, D_{0.15}, \ldots, D_{1.0}, \ldots, D_n)$. The increment should be selected in such a way that $\delta x$ gets smaller as presented in Equation 33, where *f* and *g* are the lower and upper boundaries of *x*. To obtain better results, a smaller increment is desired. It is obvious that choosing a smaller increment will be computationally expensive.

$$\int_f^g f(x)dx = \lim_{\delta X \to 0} \sum_{x=f}^{g} f(x)\,\delta x \qquad \text{(Equation 33)}$$

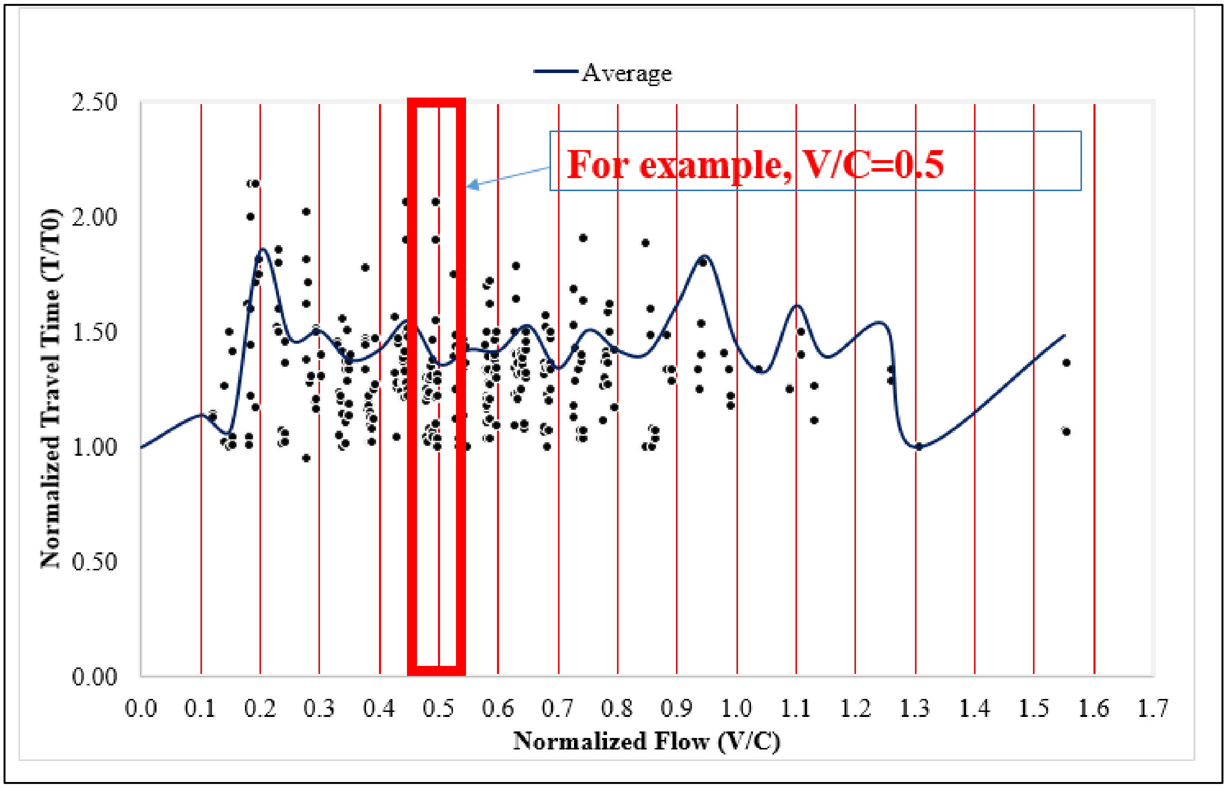


Figure 5. Conceptual Data Sampling

ii. Step 2: Normalize the $t/t_o$ by using Equation 23.

iii. Step 3: Test the type of distribution for $\{D_0, D_{0.05}, D_{0.10}, D_{0.15},...D_{1.0},..., D_n\}$.

iv. Step 4: Using Equations 24 and 25, estimate the mean and variance parameters based on each distribution for each $\{D_0, D_{0.05}, D_{0.10}, D_{0.15},...D_{1.0},..., D_n\}$.

v. Step 5: Generate a new dataset, $D$, with the mean parameters obtained from the previous step.

vi. Step 6: Test the distribution type for $D$. It is expected that the distribution would be a beta distribution.

vii. Step 7: Estimate the mean and variance using Equations 28 and 29.

viii. Step 8: Estimate the shape parameters, $P$ and $Q$, based on Equations 30 and 32.

ix. Step 9: Calculate the prior probability $p(y)$ using Equation 22 or 23.

## 3.4. Stochastic Approximation of Travel-Time Delay

This section presents the methodology involved with the likelihood as well as the marginal and posterior distribution presented in Equation 19. Section from 3.4.1 to 3.4.3 presented how the Bayesian method can be applied to predict $t/t_o$.

### 3.4.1. Likelihood Estimation

The likelihood function, $p(D|\theta)$, for the $t/t_o$, while it is a beta distribution, can be explained as shown in Equation 34.

$$p(D|\theta) = L(P, Q|\boldsymbol{y}) = \prod_{i=1}^{n} \frac{\Gamma(P+Q)}{\Gamma(P)\Gamma(Q)} y_i^{P-1} (1 - y_i)^{Q-1} \qquad \text{(Equation 34)}$$

Owen (2008) has approximated $P$ and $Q$ parameters of PERT, beta likelihood function using Equations 35-38. He has used the Newton-Raphson iterative procedure to find these parameters.

$$\tilde{\mu} = \frac{P}{P+Q} \qquad \text{(Equation 35)}$$

$$\tilde{\sigma}^2 = \frac{PQ}{(P+Q)^2(P+Q+1)} \qquad \text{(Equation 36)}$$

$$\tilde{P}_{PERT} = \tilde{\mu}\left(\frac{\tilde{\mu}(1-\tilde{\mu})}{\tilde{\sigma}^2} - 1\right) \qquad \text{(Equation 37)}$$

$$\tilde{Q}_{PERT} = (1-\tilde{\mu})\left(\frac{\tilde{\mu}\,(1-\tilde{\mu})}{\tilde{\sigma}^2} - 1\right) \qquad \text{(Equation 38)}$$

In order to calculate the likehood probability of the $t/t_o$ for a given $v/c$, following procedures were followed.

i. Step 1: Follow Steps 1-4 as described in Bayesian prior estimation for the larger dataset.

ii. Step 2: Calculate the parameters based on Equations 35 and 36 for dataset $\{D_0, D_{0.05}, D_{0.10}, D_{0.15},...D_{1.0},...,D_n\}$.

iii. Step 3: Calculate the shape parameters using Equations 37 and 38.

iv. Step 4: Iteratively follow Steps from one to three for *m* number of datasets. For example, $D_0$ dataset becomes, i.e., $\{D_{0,1}, D_{0,2}, D_{0,3}, \ldots, D_{0,m}\}$ datasets.

v. Step 5: Using Equation 34, approximate the likelihood estimate for all datasets $\{D_{0,1}, D_{0,2}, D_{0,3}, \ldots, D_{0,m}, D_{.05,1}, D_{.05,2}, D_{0.05,3}, \ldots, D_{.05,m}, \ldots, D_{n,1}, D_{n,2}, D_{n,3}, \ldots, D_{n,m}\}$.

### 3.4.2. Marginal Distribution

The marginal probability of Bayesian inferences was estimated using Equation 39.

$$p(D) = \int p(D, \theta) d\theta \qquad \text{(Equation 39)}$$

To find the marginal probability for Equation 39, following procedures were followed:

i. Step 1: Follow Steps one and two as described for the likelihood estimation.

*ii.* Step 2: Find the joint probability using $p(D|\theta)p(\theta)$.

iii. Step 3: The summation of all joint probabilities forms the number of datasets for a given *v/c*.

### 3.4.3. Posterior Distribution

i. Step 1: Using Equation 19, the posterior distribution was calculated.

ii. Step 2: Estimate the parameters' mean, variance, probability, bias, and accuracy.

## 3.5. Logistic Growth Modeling

The probabilistic $t/t_o$ dynamics with respect to the *v/c* ratio may show chaotic behavior, implying that the $t/t_o$ growth with respect to the *v/c* ratio is neither linear nor strictly increasing, i.e., chaotic in behavior. The chaotic behavior of the $t/t_o$ may follow a logistic growth model that can explain the natural phenomena of the $t/t_o$ functions' growth, which may have either increasing or decreasing characteristics. A system which is chaotic nature might not always generate the ever-increasing curve. To fit a given system like this, a cumulative, logistic growth modeling might need attention from the transportation industry. Cumulative-growth modeling, especially population-growth modeling, is widely applied in the social sciences.

To capture the chaotic nature of a given system's deterministic portion, modeling the cumulative growth might be mathematically explained with a logistic growth curve. Logistic growth modeling involves dealing with the logistic function; the logistic curve is a common "*S*" shape (sigmoid curve) that is presented in Equation 40. This Equation is the standard form of the sigmoid function. Sigmoid functions are common in statistics as cumulative distribution functions (Wikipedia, 2016). The first aspect of a logistic function is to produce a system so that it can capture the natural, cumulative rate of system growth that is chaotic in nature. The second aspect of a logistic function is to make a system such that the growth rate follows the system capacity. In that case, the more generalized form of a logistic function can be defined as shown in Equation 41. Here $k$ represents the curve's growth rate, $l$ is the saturated maximum capacity that can be sustained for a given system, and $x_o$ is a location parameter.

$$f(x) = \frac{1}{1 + e^{-x}}$$

(Equation 40)

$$y(x) = \frac{l}{1 + e^{-kx - x_0}}$$

(Equation 41)

It is obvious that the $t/t_o$ follows some laws with the $v/c$ ratio. All it was known about this system is that it is chaotic (increasing or decreasing) in nature. If someone fits the cumulative growth of a $t/t_o$, then the change for the $t/t_o$ might be directly proportional to the $v/c$, as explained by Equation 42. According to Fokas (2007), in 1798, Malthus suggested this relationship to estimate the population size. Here *dy/dx* represents the $t/t_o$ change, $k$ is the growth rate, and *y(x)* is the cumulative $t/t_o$.

$$\frac{dy}{dx} = ky(x) \qquad \text{(Equation 42)}$$

Verhulst added a correction item *(1-y(x)/k)* with Equation 42 when the system, *y(x)*, growth is exponential in nature as presented in Equation 43 (Meyer, 1994; Fokas, 2007). The growth change can be defined as presented in Equation 44 (Meyer, 1994; Fokas, 2007). A primer on logistic growth and substitution has been presented by Meyer (1994) and Fokas (2007). Interested readers are directed to read those articles. The proposed methodologies developed in this dissertation were formulated by utilizing the work of Meyer (1994) and Fokas (2007).

$$y(x) = x_o e^{kx} \qquad \text{(Equation 43)}$$

$$\frac{dy(x)}{dx} = ky(x)(1 - \frac{y(x)}{l}) \qquad \text{(Equation 44)}$$

A time-dependent solution, as presented in Equation 45, finds *y(x)* from Equation 44; this result has been reported in several studies (Fokas, 2007; Meyer, 1994).

$$y(x) = \frac{l}{1 + e^{\left[-\frac{ln(81)}{\Delta x}(x - x_m)\right]}} \qquad \text{(Equation 45)}$$

The function presented in Equation 45 has three parameters: 1) $l$ is the asymptotic maximum value where the $t/t_o$ growth is saturated; 2) $\Delta x$ is the time required to reach 10-90 percent of $l$; and 3) $x_m$ is the midpoint in time.

In the *n*-logistic function's theory, if the growth function, *y(x)*, shows a different pattern in another region with respect to *v/c*, then the *n*-logistic function can be established. In this case, the updated logistic function can be explained in Equation 46. In Figure 2 Part 7, this situation is visualized, indicating that, at some range, the growth change can be explained by sub-process one; in another region, two processes (a sub-process and sub-process two) are necessary.

$$y(x) = 1 + y_1(x) + y_2(x) + \ldots + y_n(x) \quad \text{(Equation 46)}$$

Model added a 1 to constrain the fitting when the traffic volume is zero. Later, it will be shown that adding this value, 1, to the function as a constraint does not affect the $t/t_o$ curve fitting, and not necessarily to add this value 1. In a more simplified form, the function in Equation 46 can be mathematically formulated as Equation 47.

$$y(x) = 1 + \sum_{i=1}^{n} y_i(x) \quad \text{(Equation 47)}$$

Furthermore, the function in Equation 46 is being utilized to predict the value for set *W= {1, 2, 3…w}*. Any predicted value for Equation 47 is the predicted cumulative $t/t_o$. For example, at point *w* in *W*, the cumulative $t/t_o$ can be expressed as the sum $t/t_o$ at point $w^{th}$ and $(w\text{-}1)^{th}$, which can be further expressed as Equation 48.

$$y_w(x) + y_{w-1}(x) = 1 + \sum_{i=1}^{n} \frac{l_i}{1 + e^{\left[-\frac{ln(81)}{\Delta x_i}(x_i - x_{mi})\right]}} \quad \text{(Equation 48)}$$

Again, for the example at the $w^{th}$ item in *W*, an individual $t/t_o$ can be computed by subtracting from the $w^{th}$ and $(w\text{-}1)^{th}$ item as expressed in Equations 49 and 50.

$$y_w(x) = y_w(x) - y_{w-1}(x) \quad \text{(Equation 49)}$$

$$y_j(x) = \sum_{i=1}^{n} \frac{l_i}{1 + e^{\left[-\frac{ln(81)}{\Delta x_i}(x_i - x_{mi})\right]}} - \sum_{i=1}^{n} \frac{l_i}{1 + e^{\left[-\frac{ln(81)}{\Delta x_i}(x_i - x_{mi} - 1)\right]}} \quad \text{(Equation 50)}$$

Here, it is proven that adding a constraint at volume zero into the $t/t_o$ function's formulation does not affect the model's dynamic function. Adding the constraint 1can neutralize the Equations 49 and 50. Equation 50 is the formal, proposed $t/t_o$ relationship which can be further expressed as Equation 51; Equations 52-54 are required to compute the $t/t_o$ in Equation 51. The solution of $k_i$ and $x_{mi}$ as presented in Equations 53 and 54 were utilized from the study of Meyer (1994) and Fokas (2007). Here $d$ is the initial $v/c$ value for the Bayesian prediction used earlier in the posterior estimation dataset, $\delta x$ is the $v/c$ increment used in for the Bayesian model prediction dataset in the posterior estimation dataset, $v$ is the traffic flow, and $c$ is the practical capacity.

$$t = t_0 \left[ \sum_{i=1}^{n} \frac{l_i}{1 + e^{\left[ -\frac{ln(81)}{\Delta x_i} (x_i - x_{mi}) \right]}} - \sum_{i=1}^{n} \frac{l_i}{1 + e^{\left[ -\frac{ln(81)}{\Delta x_i} ( x_i - x_{mi} - 1) \right]}} \right] \quad \text{(Equation 51)}$$

$$x = 1 + \frac{v/c - d}{\delta x} \quad \text{(Equation 52)}$$

$$k_i = \frac{ln(81)}{\Delta x_i} \quad \text{(Equation 53)}$$

$$x_{mi} = -\frac{x_{oi}}{k_i} \quad \text{(Equation 54)}$$

To find the relationship between the $t/t_o$ and $v/c$, the overall algorithm can be coded as follows.

i. Step 1: Observe the Bayesian-predicted, posterior, probabilistic mean $t/t_o$ for each $v/c$, i.e., *{E(t/t0)v/c=0, E(t/t0)v/c=0.05, E(t/t0)v/c=0.10, E(t/t0)v/c=0.15,…, E(t/t0)v/c=n}*, where $n$ is the largest-observed $v/c$ value rounded to the nearest 0.05. This portion is displayed in Figure 2 part 6, where each red dot shows the predicted, posterior probabilistic, mean $t/t_o$ for a given $v/c$.

ii. Step 2: Estimate the cumulative $t/t_o$ for a given $v/c$ with the previous $v/c$.

iii. Step 3: Fit the cumulative $t/t_o$ growth by using logistic growth modeling. For the $t/t_o$, it is expected that the exponentially always increasing, simple growth curve presented in in Equations 43-45 might be observed.

iv. Step 4: Fit the growth as an n-logistic function as shown in Equation 47.

v. Step 5: Use Equations 53 and 54 to estimate the parameters expressed in Equation 45 iteratively for Equation 47.

vi. Step 6: Formulate the formal $t/t_o$ relationship as Equations 51 and 52.

### 3.6. Tasks

There were certain tasks that needed to be done in order to accomplish the six research objectives. Some major tasks are presented in Figure 6; it displays the two aspects of the methodology. The left portion shows the data needs and a supplementary technique to collect the concurrent travel-time and traffic-counts data. The left side also shows how different technologies, such as smartphone and existing crowdsource technology, can be utilized. The right side of Figure 6 illustrates the six research objectives. The overall tasks are as follows.

i. Task 1: Conduct an extensive and rigorous Literature Review on the travel-time data-collection technique, crowdsourcing, VSM, the TTP methodology, and the travel-time functions for traffic assignment.

ii. Task 2: Establish a conceptual workflow, and develop the VSM's structural components. Develop VSM to collect the travel time from web services.

iii. Task 3: Test and validate the VSM with the NPMRDS, manual crowdsourcing, a smartphone with a test vehicle, and a web-based crowdsource application. Develop a geographic modeling tool for smartphone data conversion. Create automated workflow management for the NPMRDS data processing.

iv. Task 4: Collect concurrent travel-time and traffic-count data using VSM, NPMRDS, or varied transportation agencies.

v. Task 5: Develop a new, integrated logistic growth traffic-flow model. Then, estimate the capacity-input parameters. Validate the methodology.

vi. Task 6: Investigate the deterministic traffic speed-density model. Then, estimate the FFS' input parameters. Validate the methodology.

vii. Task 7: Evaluate the performance of the historical $t/t_o$. Approximation of Bayesian prior information for the historical $t/t_o$ was performed. The expected mean and probability for the $t/t_o$ conditioning the discrete $v/c$ ratio is computed during this step. Validate the methodology.

viii. Task 8: Provide a new, technical and scientific methodological approach to predict the $t/t_o$ using the stochastic Bayesian dynamic update and PERT techniques. Approximate the posterior distribution, and estimate the parameters. Validate the methodology.

ix. Task 9: Establish a new, link $t/t_o$ function based on the knowledge borrowed from the market-adopted, logistic-growth-curve technique and the Bayesian model update. Study the proposed method by comparing other suitable methods that are practiced by the transportation industry. Validate the methodology.

A well-defined and systematic methodology is crucial to achieve the research goals and tasks. The research workflow is presented in Figure 7. The entire workflow has four main phases.

At phase 1, this research started with a demanding topic. In the first phase, a thorough plan was developed after identifying the research interest. An extensive and rigorous Literature

Review on the development of the travel-time data-collection techniques, VSM, TTP methodologies, and the $t/t_o$ function was performed. From the Literature Review, the research needs, the literature gaps, and the industry needs, leading to specific research goals, hypotheses and assumptions, expected results, and a problem statement was identified. A pilot study was performed to check the data collection, data availability, assessing the research's possibility and feasibility, and finding the resource requirements.

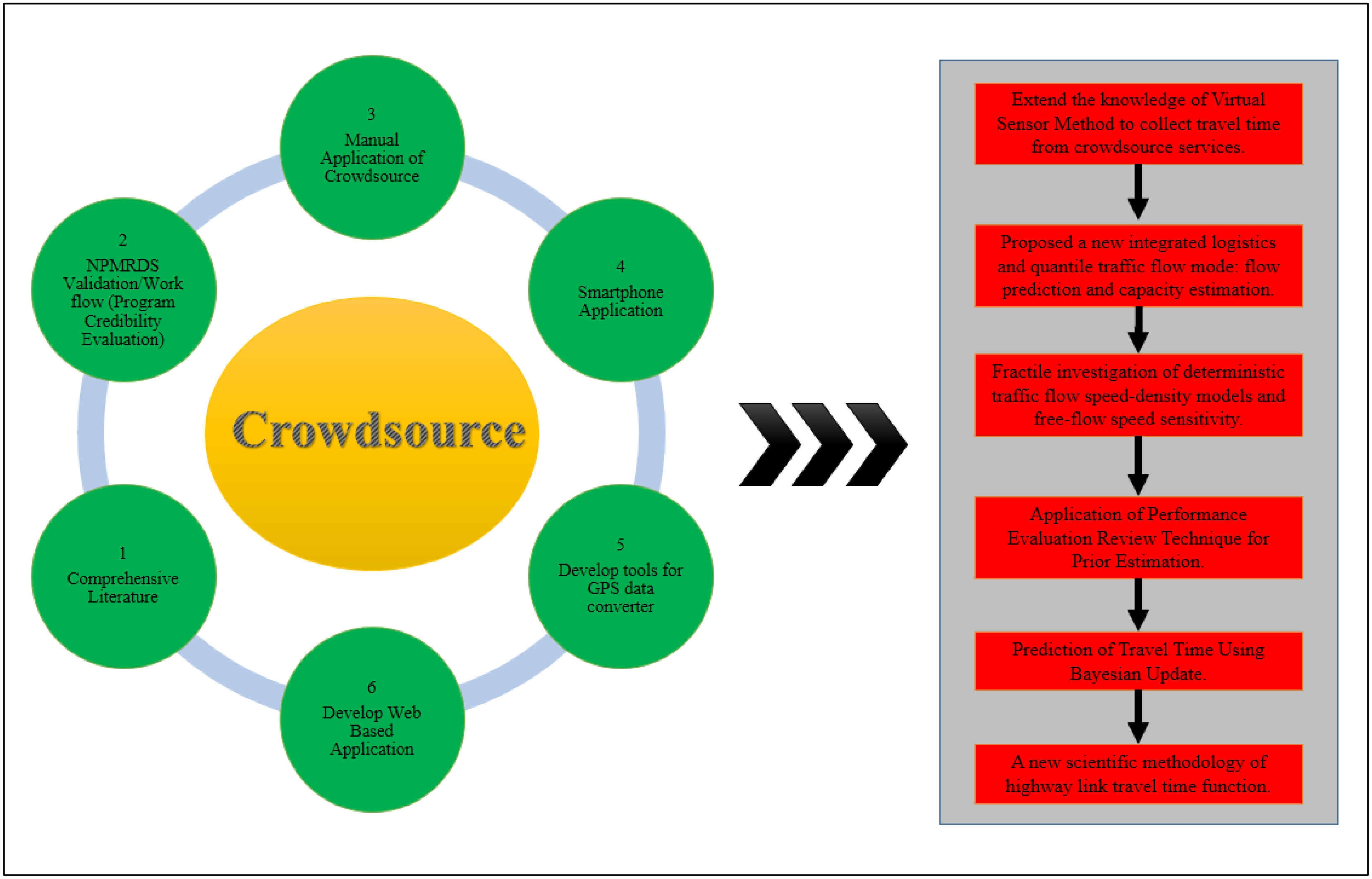


Figure 6. Research Objectives Flowchart

At phase 2, several sub-phases: extending the VSM as well as creating apps, macros, and programs to collect the travel time was completed. Framework for the VSM was developed. The VSM was then utilized to collect the travel time from web services based on OpenStreetMap. Testing and validating the VSM was also included in this phase.

To validate the VSM, the travel time was collected from different sources. Because of limited number of resources, this research considered several methodologies such as smartphone, test vehicle, and NPMRDS.

At phase three, immediately after testing the VSM, collecting the data using the new methodology was started. Necessary noise removal was performed. At phase four, the necessary analysis for my remaining research objectives was conducted. Finally, the results were presented, drew the overall summary and conclusions, found the research limitations, and provided the appropriate recommendations.

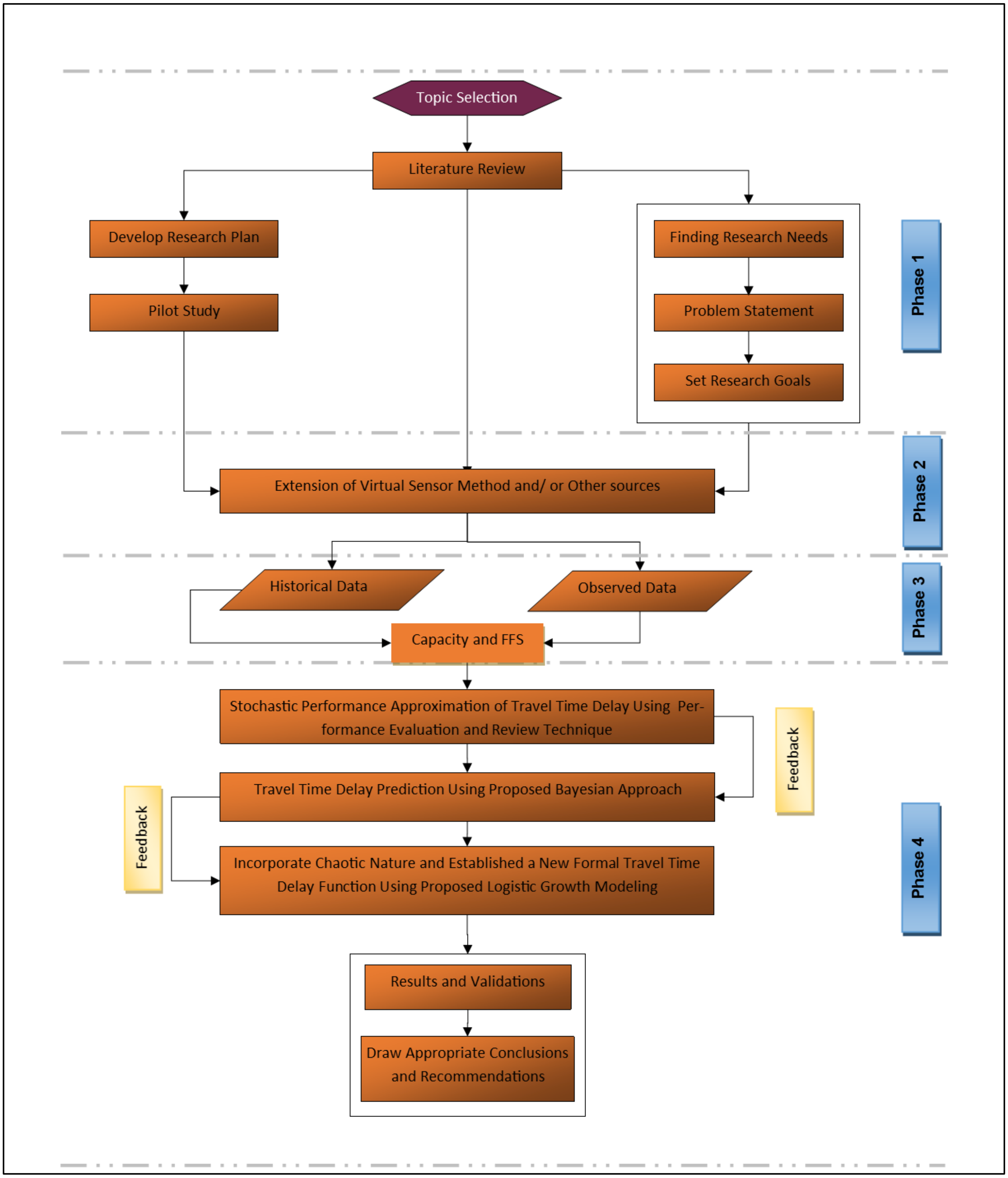


Figure 7. Dissertation Flowchart

# 4. DATA COLLECTION AND VIRTUAL SENSOR METHOD

This chapter discusses following items and investigates potential data sources.

i. Investigate VSM, Crowdsource, and NMPRDS

ii. Investigate Location Enabled Smartphone

iii. Research Location

In this Chapter, applicability of VSM, crowdsource, NPMRDS, and location enabled smartphone data are presented. The findings of these methodologies are presented in Sections 4.1 and 4.2. At Section 4.3, study location is presented.

## 4.1. Investigate VSM, Crowdsource, and NMPRDS

### 4.1.1. General

Travel time is an important parameter for transportation planning and operational decision making process especially for corridor, subarea, or any given study area. There are numerous methodologies such as Bluetooth, cellular phone, crowdsourcing, probe measures, and spot measures (radar, detector, image processing) involving travel time data collection. Literature review suggest that each of these methodologies require mobile, temporary, or permanent equipment installation and maintenance, or paid third party vendors, which will price a substantial amount of cost and time depending on the research goal. On the contrary, the OpenStreetMap crowdsource mapping application would leverage for an O-D travel time data collection technique. This methodology is probably lacking suitable attention to researchers and practitioners, which merits an investigation to see if this resource can be useful as an alternative data collection technique. Null hypotheses assumed that the collected travel time through crowdsource services may indicate real-time or near real-time travel time, eventually which can be useful in the absence of local data unavailability.

There is a strong needs for real-time or near real-time travel time data for the transportation planning due to the following reasons.

i. Travel time based on FFS following HCM.
ii. Congested assignments based on FFS may affect the models output accordingly (Motuba, 2012).
iii. In the absence of local data, the FFS based on national averages are being using to calibrate and validate the model outcomes in traditional practice (Motuba, 2012).
iv. National averages are based on limited numbers of model areas.
v. Need real-time or near real-time travel time data to replicate ground truth condition.

### 4.1.2. Problem Statement

Finding suitable travel time data is much crucial for regional transportation planning and considerably unavailable. Therefore, this research aimed to investigate whether a suitable travel time data collection technique can be used in model calibration and validation purpose for given matrices of origins and destinations.

### 4.1.3. Research Objective

The main objective of this research section was to investigate whether a crowdsource services could be an alternative resources for travel time data collection. To achieve the research goals, what lead to concentrate on this research is that how can someone use the best resources that is open, big, or freely available to some extent so that relate our everyday movement. Eventually, the idea of utilizing crowdsource data collection through online services.

There are many crowdsource vendors such as Bing, HERE, Google, OpenStreetMap, MapQuest, TomTom, INRIX, and AirSage. Cambridge Systematics (2012) stated that Google Map is a big player in crowdsourcing to collect the real-time traffic data collection. The study of

Wang and Xu (2011) evident that Google Map API can be deployed. Morgul et al. (2014) has studied Bing and MapQuest through VSM. Morgul et al. (2014) proposed that their methodology can be an attractive alternative for traffic surveillance method. In order to provide a novel solution, data collection using OpenStreetMap services would be a suitable crowdsource to collect travel time data. The main reason to select OpenStreetMap is that it is fully free, open and big. On the contrary, other services are not technically free and open except MapQuest. MapQuest services are free in some extent and with limited usage. In addition to that most of the crowdsource services requires explicit permission and have limits on uses before using the respective services.

#### 4.1.4. Data Collection and Analysis

A corridor containing nine segments of I-29 freeway (Northbound) within the Fargo-Moorhead Metropolitan Planning Area (FMMPA) was considered. The research corridor is presented in Figure 8. Figure 8 displayed two major freeways (I-29 and I-94) running over the research location. This Figure 8 also displayed the Traffic Message Channel (TMC) stations. This TMC points are coming from NPMRDS program of Federal Highway Administration. The NPMRDS data will be used for the validation purpose of the OpenStreetMap data[2]. Data was collected for different time period and day for a week from OpenStreetMap and is licensed under ODBL 1.0. Once the data collection was done, a workflow management for NPMRDS data collection was developed using ArcGIS, Python, and SAS. Travel time was then normalized as travel time per mile so that the selected nine segments can be compared.

[2] OpenStreetMap data was collected as a part of varied projects of Advanced Traffic Analysis Center at Upper Great Plains Transportation Institute.

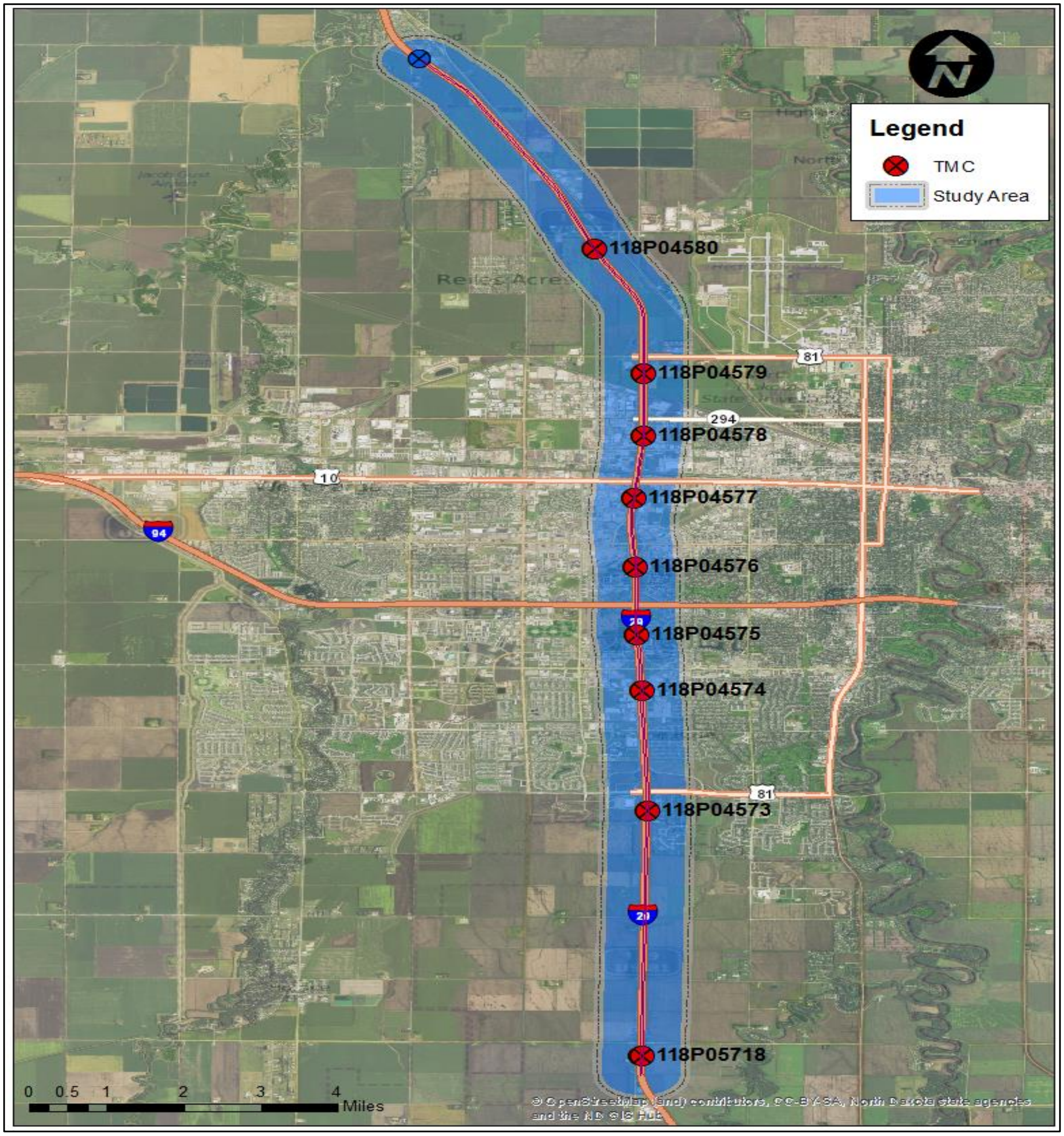


Figure 8. Research Area and TMC Locations

### 4.1.5. Results and Discussions

This section discussed some interesting findings and results that can be useful for the travel time data collections techniques and transportation planning. Null hypotheses was that travel time from the OpenStreetMap service is not different in compare to the observe data. In this regards, statistical paired t-test (NPMRDS-vs.-OpenStreetMap) was performed for the studied nine segments, which is presented in Table 10. The paired t-test for the OpenStreetMap services over NPMRDS is significantly different even at 90 percent confidence interval, which fails to accept the null hypotheses that the travel time from OpenStreetMap services are not different because the p-value found was 0.00316. The reason of these two are significantly different might be the update frequency of OpenStreetMap services. However, results indicate some interesting facts as well.

Table 10. OpenStreetMap Statistical Significance Test

| Items | NPMRDS | OpenStreetMap |
|---|---|---|
| Mean (Second per mile) | 62.65 | 77.53 |
| Variance (Second Square) | 46.98 | 110.59 |
| Standard Deviation (Second) | 6.85 | 10.52 |
| t-Stat | | 3.56 |
| P(T<=t) two-tail | | 0.00316 |
| t-Critical two-tail | | 2.14 |
| | | Significant |

Table 10 shows that the mean travel time per mile for the study freeway is 77.53 and 62.65 seconds for the OpenStreetMap and NPMRDS respectively. But the mean travel time varied 14.88 seconds per mile to replicate the observed condition. The standard deviation of the true data is 6.85 seconds but the OpenStreetMap is 10.52 seconds respectively. The difference in standard deviation found to be 3.67 second for the OpenStreetMap in compare to NPMRDS.

Estimated percent difference travel time of the OpenStreetMap against NPMRDS dataset is presented in Figure 9. In the horizontal axis in Figure 9 is the link segments from first to ninth. This corridor starts with rural area with speed limit 75 mph, in the middle urban area with speed limit 55 mph, and at the end again rural area with speed limit 75 mph. The vertical axis in Figure 9, indicate the percent of travel time varied by the OpenStreetMap compared to NPMRDS

The percent variation of travel time by the OpenStreetMap varied approximately 10-40 percent. In the urban area travel time variation is less compare to rural area. Probably the prediction power of OpenStreetMap has better algorithm for the urban area in compare to rural area.

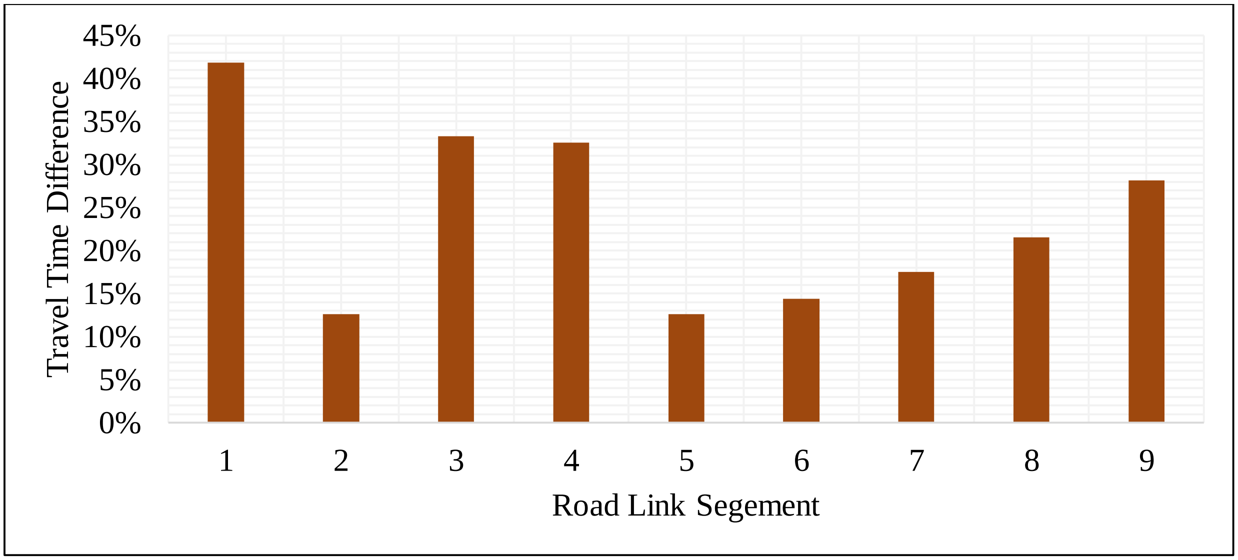


Figure 9. Percent Differences of Travel Time Plot over NPMRDS

Trends of mean travel time over each segment is presented in Figure 10. The vertical-axis of Figure 10 represents the mean travel time in seconds. Figure 10 also supports the previous discussions of Figure 9.

What can be seen is that the OpenStreetMap travel time is replicating the trends and variation of the observed data. Mean observed travel time ranged from 53.18-72.42 seconds for the nine segments. The OpenStreetMap travel time ranged from 62.28-96.54 seconds for the nine segments. The OpenStreetMap is always over predicting the travel time compare to the observed NPMRDS data. However, in the corridor level OpenStreetMap replicate the similar trends.

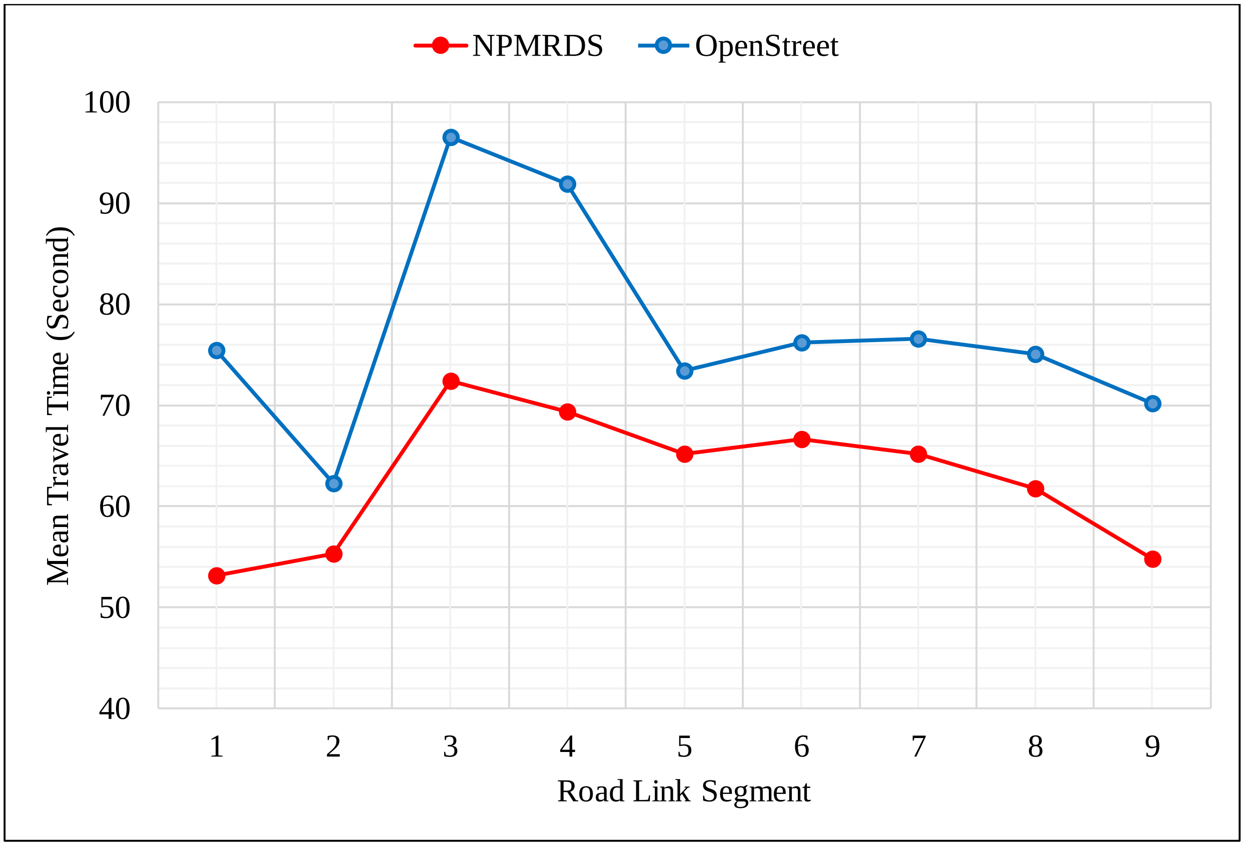


Figure 10. Trends of Corridor Travel Time

At last, cumulative distribution of the mean travel time per mile for the nine segments is presented in Figure 11. This plot reflects that travel time over the corridor does not varied much. The travel time over each section is considerably consistent by the OpenStreetMap.

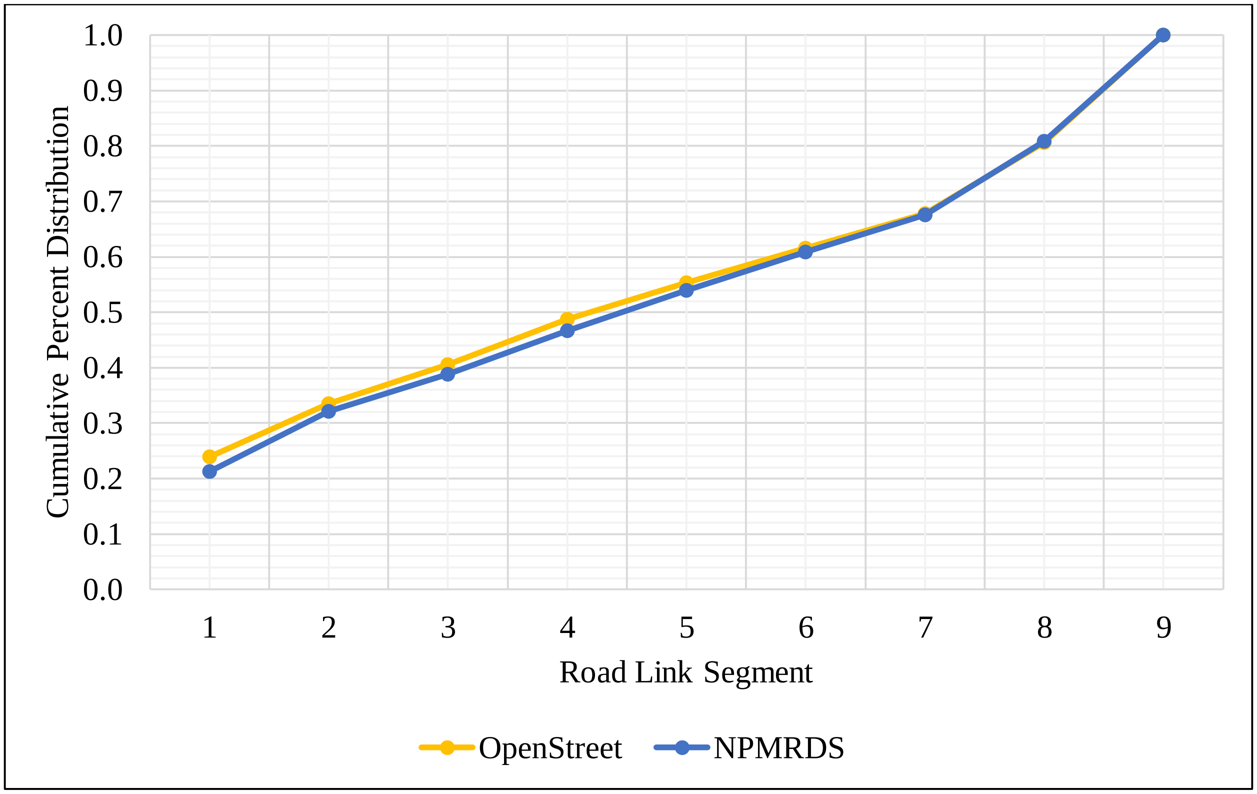


Figure 11. Cumulative Distribution of Travel Time

#### 4.1.6. Summary

OpenStreetMap might be a potential candidate for corridor level travel time study. The smaller and medium size transportation agencies with vivid resource constraint might get help substantially using this methodology especially for corridor level planning. Regardless of flaws, it is evident that this new technology is cheap; free with limited capability; no cost except obtaining the data cost; and finally installation is not required. However, literature of crowdsource is not rich. Therefore, this research will rich the literature of crowdsourcing data acquisition technology as well.

## 4.2. Investigate Location Enabled Smartphone

### 4.2.1. General

Travel time for a given pair of O-D or waypoints can be collected many ways such as broadly categorized into probe vehicle measure, spot measure, test vehicle, license plate matching, and emerging technology. Each method has its own advantages and disadvantages. However, crowdsourcing is a new technology falls under probe vehicle measure need suitable attention to the transportation industry. While the required resources for travel time collection is constrained heavily upon resources, the crowdsource travel time through web mapping vendors such as Bing, Google, MapQuest, OpenStreetMap, and Yahoo now named as HERE create a new milestone for the transportation industry. These web mapping services is pioneer and well known to the transportation user and industry especially for the travel navigation system. Therefore, this research aimed to investigate the proximity of the travel time data collection from the open and free web mapping services OpenStreetMap. Manual travel time collection from web mapping services was validated against global positioning system enabled smartphone and test vehicle. The outcomes of this research will contribute to the needs of the crowdsource literature, transportation industry, researcher, and practitioner significantly.

### 4.2.2. Introduction

Based on the literature findings, it is clearly indicative that there are some gaps of the study that need to enrich in order to fill the crowdsourcing literature. First, none of the earlier studies performed comprehensive comparative study such as Bing, Google, MapQuest, Open Street, Yahoo (HERE), and INRIX, etc. Second, none of the above study did investigate the big and completely free OpenStreetMap services though MapQuest is open with limited response. Third, collecting travel time through web-based virtual sensor or macro is programming oriented and advanced, where small or medium size cannot be benefited where technical expertise is limited. Therefore, considering the last two issues, the online travel time data may be collected manually through the open web mapping services OpenStreetMap. Of course, OpenStreetMap terms and conditions apply. Therefore, the research goal of this section was to investigate the real-time proximity of travel time data collection from web mapping services in compared to GPS enable smartphone. In order to validate the data, real-time travel time can be collected through GPS location enable smartphone by ignoring of its disadvantages. The reason of choosing GPS locations enable smartphone was based on the unavailability of other technology such as loop detector is most common methodology for travel time data collection methodology validation.

The research methodology incorporates manually collection of travel time with hyper linking in a simple excel data sheet from web mapping services; utilized publicly available free travel time data collection app that can be integrated with test vehicles. Post processing of the data dictated to develop a macro to analysis the data and a GIS tool to process the smartphone app generated data. However, the proposed methodology sometime experience several challenges, which has been describe in the methodology chapter. A case study of a freeway corridor within the medium size FMMPA is included for an illustrative example.

The section is organized as four sub-sections. Section 4.2.3 presents the methodology, develop the model and study the available apps. Section 4.2.4 discuss the results and presents the summary. Section 4.2.4 draws appropriate conclusion based on the research findings.

### 4.2.3. Methodology

To achieve this section research goal, a 5.3 miles' corridor of freeway I-29 northbound within the FMMPA was considered. The study area is portrayed in Figure 12 by the buffer line.

The origin was considered started closest to 52nd Avenue North (Latitude: 46.809985, Longitude: -96.838904) and destination was closest to 12th Avenue North (Latitude: 46.885929, Longitude: -96.839353). Travel time from OpenStreetMap was collected manually during AM and PM peak hours from July 24, 2015 to September 18, 2015. The AM peak hour are from 7:00 AM to 9:30 AM and the PM peak hour are from 4:00 PM to 6:00 PM. During each peak period, observations was recorded from the web services.

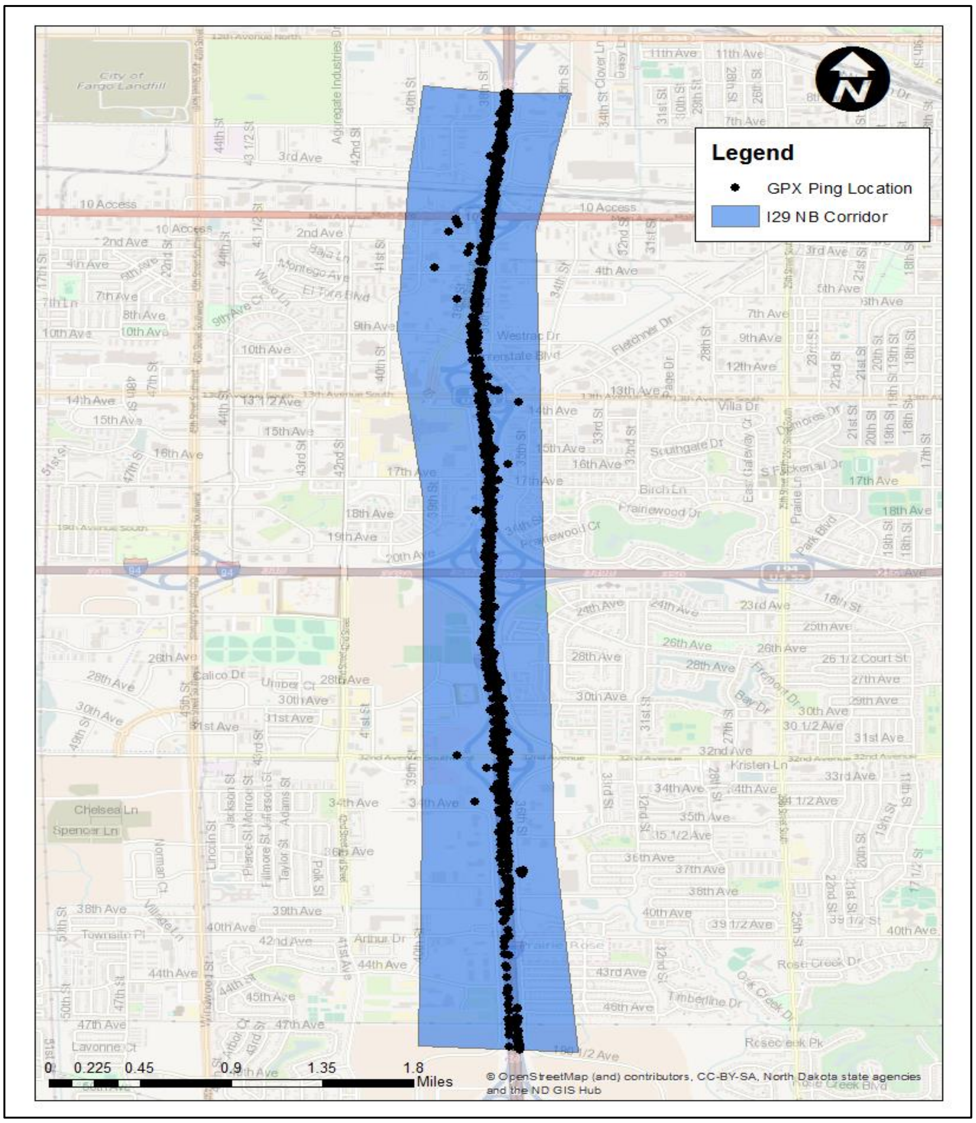


Figure 12. Study Area and Ping Points

The methodology of data collection from web mapping services was based on online (manual operation). To make the methods a little friendly, for the given origin and destination, a Hypertext Transfer Protocol (HTTP) link was generated. In an excel sheet, each link can be attached in a cell value. Clicking linked cell can directed to the web services and showed the travel time. However, this method indicated a little laborious job and limitation. First, for the same origin and destination, service showed different location and direction shifting of the actual location. Therefore, a further manual operation was needed to rectify the problem. Second, sometime, web service don’t allow to start routing in the middle of the freeway. It forced to select some closest facilities like a land, parking, business, housing, etc. The web service that create the second issues, was not considered in this study.

During the same time frame, a free smartphone app called “SPEEDVIEW” was utilized, which has the capability to enable current location coordinates of a test vehicle[3] using cell phone GPS; collect current time of the locations for a specific time, and store the data on the system; and finally would be able to create a track points line of the vehicle movement. Of course, every app has its own capability and limitation. Considering the SPEEDVIEW capability, the minimum time between points for pining of the GPS settings was considered one second and minimum accuracy 10 meter.

To remove the biasedness of the outcomes, the test vehicle was driving by following randomly selecting the closest vehicle with same relative speed and distance. If a selected vehicle, took an exit, then the next closest vehicle was followed. In the absence of no vehicle, posted speed limit was followed. However, data collection indicated some drawback of these methodologies. During the data collection, GPS signal lost found frequently. Data showed some

[3] Data was collected as part of varied projects of Advanced Traffic Analysis Center at Upper Great Plains Transportation Institute. Special thanks and acknowledge to Dr. Diomo Motuba for his permission to use this data.

noisy location and path of travel sometime. Therefore, the noisy ping location was eliminated using a buffer line. Sometime the app does not store the data or even store then don't allow to send the data.

To process the GPX data, a model tool was developed using ArcGIS software. The model is presented in Figure 13. The model initially collects the GPX file, convert from the GPX to feature, clipping the ping location within the buffer line, sorting, create suitable field and combined a dbf format file. At this step, the model just produced the combined GPS dataset. The major attributes in this step is unique sorted ID, date-time stamp, latitude, and longitude. A sample of the data is shown in Figure 14.

Since every peak hour have one GPX file but include many ping locations. To calculate the travel time for each GPX file in database created using the model tool earlier, the statistical SAS programming software was used utilized to consider the first and last location time stamp and calculate the difference of total travel time. After this process, travel time from the web services was aggregated into a single database. A sample of the final combined data and sample major attributes is presented in Figure 15.

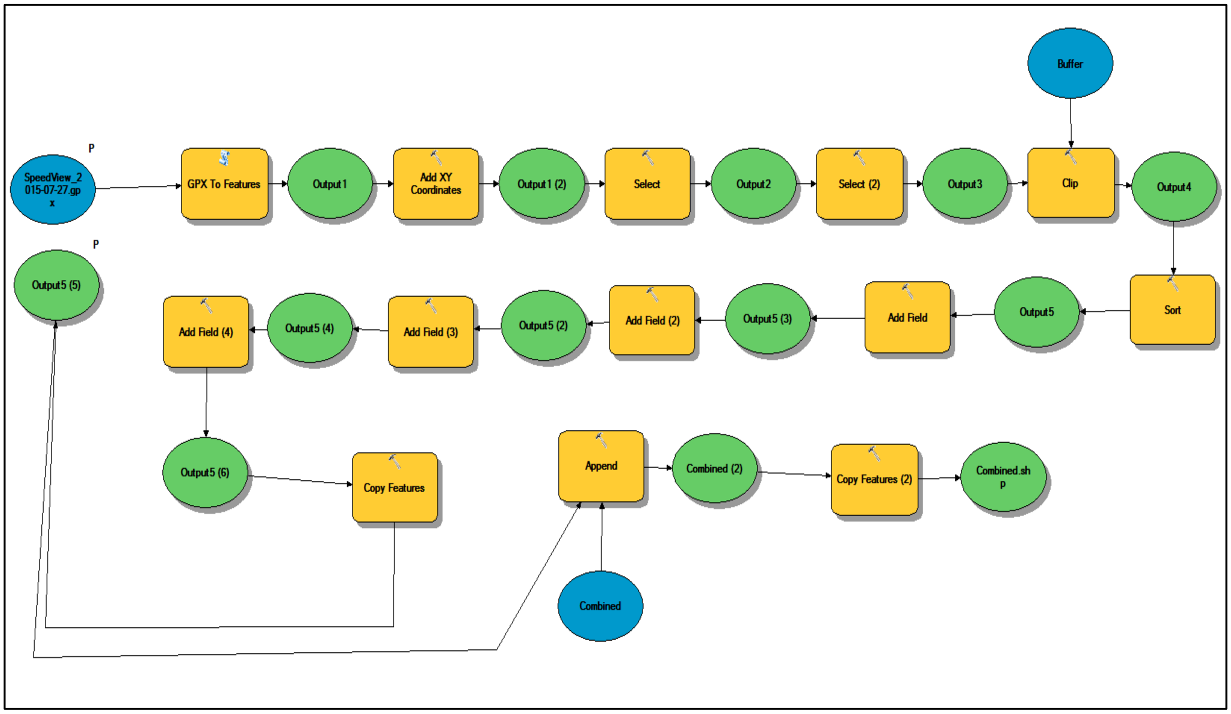


Figure 13. Model for Processing GPX Data

| OBJECTID * | DateTimeS | DateTime | POINT_X | POINT_Y |
|---|---|---|---|---|
| 1 | 2015-08-03T21:44:20Z | 8/3/2015 9:44:20 PM | -96.840208 | 46.883573 |

Figure 14. Sample Model Database

| DATETIME | Cat | Type | TravelTime |
|---|---|---|---|
| 30JUL2015 | AM | Openstreet | 6 |

Figure 15. Sample Final Database

### 4.2.4. Results and Discussions

This section presents the appropriate results and discussion based on analyzed data collected from OpenStreetMap web mapping service. The results are then validated and compared against the data collected from GPS enabled smartphone with test vehicle. Later this section will present the travel time reliability of these methodologies based on working day of month, workday of a week (Monday through Friday), and two peak hours (AM and PM).

First the analysis looked at the statistical mean travel time of working day of the month for the 5.3 miles' urban freeway corridor. It was found that the travel time was 6.08 and 6.00 minutes for Manual and OpenStreetMap respectively. The mean travel time of the OpenStreetMap services are considerably close to the manual method for the study period of July 24, 2015 to September 28, 2015. The difference of daily mean travel time with respect to manual method which is observed data found 4.8 seconds for OpenStreetMap. Even though the daily mean difference is ranged from 54.6-58.8 seconds for the 5.3-mile section, but the standard deviation of the daily mean travel time was observed 1.92 and 0 minutes for Manual and OpenStreetMap respectively. The standard deviation of OpenStreetMap services is zero indicating that the travel time updates is not so frequent. One of the interesting fact over the study period is that OpenStreetMap travel time did not show any variation. Therefore, standard deviation of OpenStreetMap web mapping services is zero. A complete list of mean statistics and standard deviation is presented in Table 11 based on daily, peak hour and working day of week.

Mean statistics showed that Monday and Friday showed more variation than other working days. Mean difference of web services compared to manual method range from 0.91-0.98 minutes from Monday to Friday. PM peak hour travel time shows higher than the AM peak hour travel time. The PM peak hour travel time indicate more variability than AM peak hour.

Table 11. Travel Time Mean Statistics

| | Work Day | |
|---|---|---|
| | Daily | |
| Type | Mean | Std |
| Manual | 6.09 | 1.93 |
| OpenStreetMap | 6 | 0 |

| | Peak Hour | | | |
|---|---|---|---|---|
| | AM | | PM | |
| Type | Mean | Std | Mean | Std |
| Manual | 5.93 | 1.96 | 6.52 | 1.9 |
| OpenStreetMap | 6 | 0 | 6 | 0 |

| | Day of Week | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Monday | | Tuesday | | Wednesday | | Thursday | | Friday | |
| Type | Mean | Std | Mean | Std | Mean | Std | Mean | Std | Mean | Std |
| Manual | 6.91 | 0.86 | 6.39 | 0.91 | 5.02 | 0.38 | 5.95 | 0.78 | 6.25 | 0.95 |
| OpenStreetMap | 6 | 0 | 6 | 0 | 6 | 0 | 6 | 0 | 6 | 0 |

Summary statistics only reveal each methods deviation from the ground truth condition. But the question, is whether or not the travel time is approximating the ground truth condition or the OpenStreetMap has its own complementarity. First test was done whether or not the daily travel time by web services is significantly different than the manual method. Least square means test showed in Table 12 that the mean travel time of OpenStreetMap is not significantly different than the manual method at 95 percent confidence interval.

Table 12. Daily Travel Time Significance Test

| Least Squares Means for Effect Type<br>t for H0: LSMean(i)=LSMean(j) / Pr > \|t\|<br>Dependent Variable: Travel Time | | |
|---|---|---|
| i/j | Manual | OpenStreet |
| Manual | | 0.48823<br>0.626 |
| OpenStreet | -0.48823<br>0.626 | |

Legend:

| | |
|---|---|
| (green) | Insignificant |
| (red) | Significant |

Later significance test was performed at 95 percent confidence interval for AM and PM peak hours. The test results is presented in Table 13. Test results indicate that OpenStreetMap AM peak hour travel time is not significantly different than the manual AM travel time at 95 percent confidence interval. Which clearly a good representation of near-time travels time estimation by the web services. Similarly, OpenStreetMap PM Peak hour travel time is not significantly different than the manual PM travel at 95 percent confidence interval. The manual method AM and PM travel time is not significantly different as well. Similarly, OpenStreetMap AM-vs.-OpenStreetMap PM is not significantly different though the mean statistics showed more variation in the PM peak than the AM peak hour. Since the case study area is a medium size metropolitan area and from the local experience, there is consistency in the peak hour travel time. Travel time does not vary much in the AM and PM peak hour which support the expected results. Therefore, it might be inferred that travel time collected from OpenStreetMap service are nearly-real-time for the AM and PM peak hour. It also showed that majority of the cases except few instances, the AM and PM peak hour travel time is complementary of each other.

Table 13. Peak Hour Significance Test

| | | | Least Squares Means for Effect Type*TimeCat<br>t for H0: LSMean(i)=LSMean(j) / Pr > \|t\|<br>Dependent Variable: Travel Time | | | |
|---|---|---|---|---|---|---|
| | | | Manual | | OpenStreet | |
| | | | AM | PM | AM | PM |
| | | i/j | 1 | 2 | 3 | 4 |
| Manual | AM | 1 | | -1.23<br>0.22 | -0.13<br>0.89 | -0.09<br>0.93 |
| | PM | 2 | 1.23<br>0.22 | | 0.87<br>0.39 | 0.67<br>0.51 |
| OpenStreetMap | AM | 3 | 0.13<br>0.89 | -0.87<br>0.39 | | 0.00<br>1.00 |
| | PM | 4 | 0.09<br>0.93 | -0.67<br>0.51 | 0.00<br>1.00 | |
| Legend: | | | (yellow) | AM Insignificant | | |
| | | | (blue) | PM Insignificant | | |
| | | | (red) | Significant | | |

## 4.3. Study Location

Five-minute interval traffic flow data captured with loop detectors were collected from Caltrans Program Evaluation and Monitoring System (PEMS) for the complete year 2011-2015. The study link location was a four lane segment of Interstate 5 south bound in Los Angeles County, California, which is was one of the 10 most congested places in the United States (USA Today, 2015). The geographic location of the loop detector was California post mile 11.3, which is station number 763980 presented in Figure 16. Consecutively, I-5 route was the highest congested place in Los Angeles since 2009 (CalTrans, 2012). However, according to Caltrans (CalTrans, 2012), the study location was the most congested bottleneck place at I-5 in 2012 during the afternoon/evening peak period.

The raw data processing was tedious and exhaustive in nature. After extracting the 5 minutes interval data for a complete year, it uses 60 Gigabytes of computer memory. The raw database contains almost 200 million records for the district seventh for each year. The agglomeration of data was performed using multiple statistical packages using SAS, R, and MS

Access. Once the data was agglomerated then the desired study location data was extracted from the main database. Raw data included bad weather and detector failure data. Therefore, only 100 percent quality data were utilized which does not have any bad weather effect and detector problem. Later, query out of any major events (accident, road construction, and so on) that negatively affect the traffic flow on a specific day and time was performed. For example, around 22 percent data for the year 2014 indicated reliable dataset which was then considered for this research.

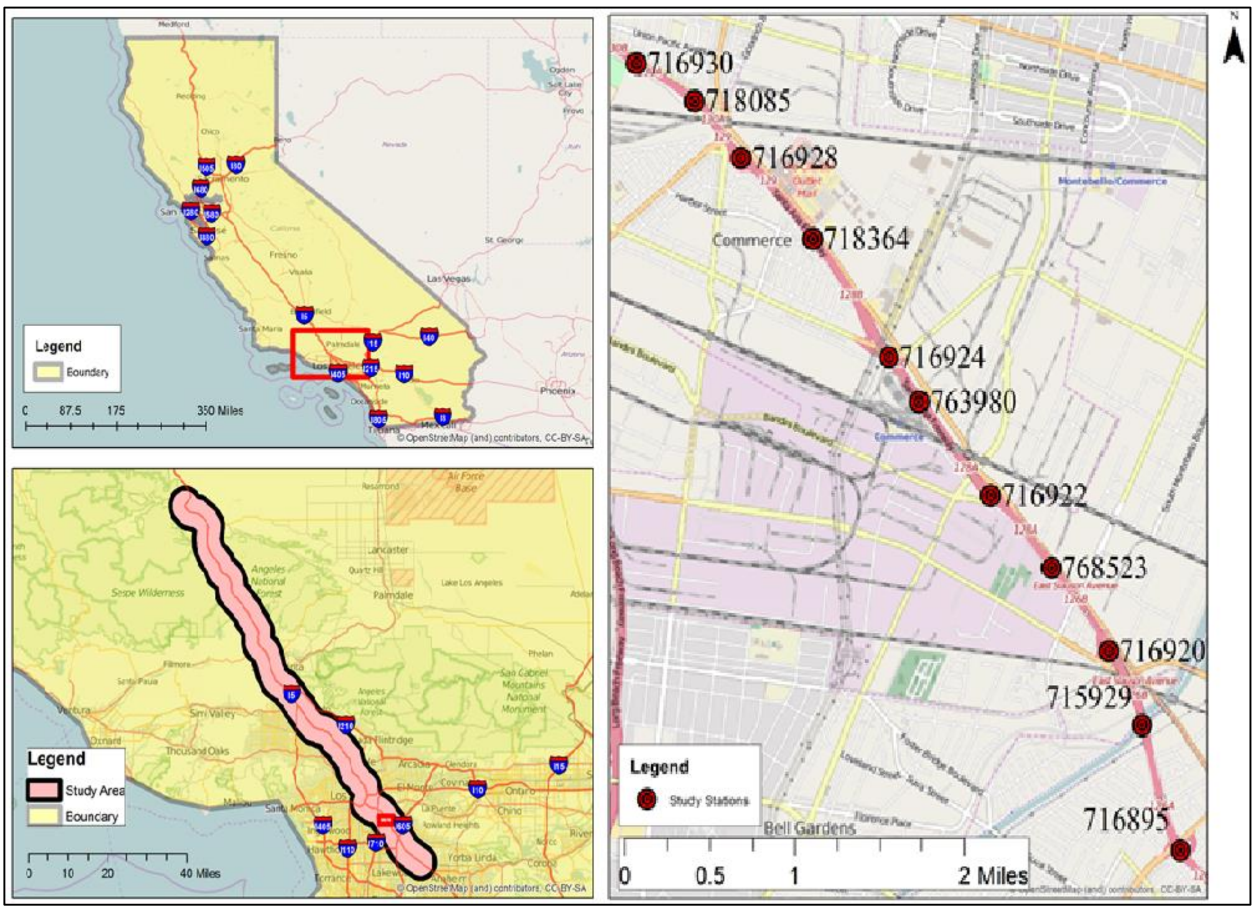


Figure 16. Study Location

# 5. MODELING OF STABLE CAPACITY[4]

## 5.1. Overview

Capacity approximation of highway link segments is critically essential, subjective, widely recognized by the transportation industry, and is a topic that generates a high level of research interest. It is obvious because transportation investment decisions are highly correlated to the accuracy of forecasting of the amount of traffic, vehicle hours travelled, vehicle miles travelled, $t/t_o$, LOS, and $v/c$ ratio measures. All of these measures are the direct outcomes of highway capacity oriented inputs. One of the example would be the last step of a TDM is highway assignment, which require calculation of accurate capacity estimation in the network data preparation stage. Therefore, this study aims to develop a new traffic flow prediction model for freeway so that it may capture the non-linear flow characteristics. Results shows that the proposed methodology are highly potential and considerably a good candidate for capacity approximation and flow prediction. This study brings a new knowledge of non-linear logistic quantile function into the traffic flow prediction and capacity estimation.

## 5.2. Introduction

Branston has defined the capacity of a link as the maximum steady-state flow on that link (Branston, 1975). This capacity is also known as the steady state capacity (Branston, 1975). Another capacity term is practical capacity. Branston (1975, p.226) has defined the practical capacity as, "the maximum number of vehicles that can pass a given point on a roadway or in a designated lane during one hour without the traffic density being so great as to cause unreasonable delay, hazard, or restriction to the driver's freedom to manoeuvers under prevailing

[4] This Chapter is ongoing manuscript for potential publication with supporting author Dr. Matthew L. Stone at North Dakota State University and Babak Mirzazadeh at Upper Great Plains Transportation Institute. Both the supporting authors helped in proof reading. Main contribution of Babak Mirzazadeh portion was a part of review of literature, which is not included in this Chapter.

roadway and traffic conditions". According to that paper, practical capacity is a single value capacity. A single value capacity cannot represent different operating condition and LOSs (Branston, 1975). Therefore, LOS concept was added in 1965 HCM and capacity was defined as service volume (Branston, 1975). Branston also stated that practical capacity is more often used in capacity formations than in steady-state capacity. He indicated that practical capacity and services is very subjective and confusing. According to his study, it is preferable to use steady-state capacity in the formulation in link capacity functions.

The HCM 2010 (TRB, 2010; TRB 2000) is the state-of-the-art of best practices for roadway capacity estimation. HCM 2010 includes numerous factors such as FFS, ramp density, percentage of heavy vehicles, peak hour factor, and *v/c* ratio that can affect the freeway capacity estimation. Estimating capacity from HCM is a subjective nature because it incorporates the LOS or *v/c* ratio. Molla (2016) has presented a case study of FMMPA area on implications of highway capacity manual on freeway measure of effectiveness. His study presented how the new HCM 2010 can be different than old HCM 2000 based on LOS or v/c ratio. Utilization of HCM has significant effect on estimation of performance measure of freeway (Molla, 2016). In volume delay function formulations, researchers are using various percentile flow to represent capacity. Mtoi and Moses (2014) used $99^{th}$ percentile flow as the practical capacity for volume delay function postulation.

There are several issues that need proper attention to the researchers, practitioners, and policy makers. First, the subjectivity of the practical capacity is well addressed by many studies. When capacity estimation comes to transportation planning stage, estimations of capacity are mostly practical capacity and subjective in nature for a given link. Second, highway capacity manuals provide very good guidance in estimating capacity. But this guidance is heavily concentrating on an operational aspect. When it comes to planning, numerous types of assumptions are necessary to replicate the operational condition into the planning model. These assumptions are coming from different sources. The variety of assumptions in capacity estimation is cumbersome, which may lead the practitioner to adopt a simple methodology.

The aforementioned issues are mostly general problems and there are more traffic flow characteristics that need to be included while approximating capacity. Traffic flow mean and variance of each hour are varying considerably making the traffic flow prediction erroneous if a LM is used because it might lead to higher errors in prediction. In this regard, a non-linear logistic growth model has been developed.

An ordinary non-linear model may exhibit various flow characteristics in different quantile of the observations. To find a more robust model that include comprehensive analysis of the prediction variability, quantile regression has been considered as well.

As this literature review reveals, earlier studies did not addressed suitably the logistic function or quantile regression knowledge in traffic flow studies. Therefore, the primary objective of this research is to provide an integrated model for capacity estimation and traffic flow prediction. In this regard, a mathematical non-linear logistic function was developed based on sigmoid curve fitting technique. Furthermore, a quantile regression analysis was performed for the proposed logistic model. The results were compared to different statistical measures, existing methodology, and capacity values found in the literature.

## 5.3. Methodology

The research methodology includes three major components: 1) start value selection; 2) logistic growth modeling; and 3) non-linear quantile regression.

### 5.3.1. Start Value Selection

The overall goal of this research section is to predict the traffic flow (say random variable $y$) using Verhulst (Wikipedia, 2016) logistic growth function conditioning hours of a day (say $x$-th hour of the day). In order to fit the logistic growth function, an initial start value with three parameters of logistic function including: 1) curve maximum saturation parameter ($l$), 2) steepness of the curve ($k$), and 3) the sigmoid midpoint ($x_o$) are required. A well-defined explanation of logistic growth modeling and its parameters estimation can be understood from Meyer (1994) study. The initial value was achieved in such a way that the cumulative distribution pattern of the observed data and the model data could be similar pattern.

It was required to normalize the observed data using a suitable method. Since it was aimed to predict the traffic flow ($v$), normalization with maximum mean hourly flow ($v_{max}$) using Equation 55 is acquired. Here $v_i$ is the mean hourly flow from random observations, $v_{max}$ is the maximum hourly flow, and $y_i$ is the normalized flow for an individual observation. This ratio indicates how much the weight of each observation is compared to the $v_{max}$. Cumulative summation of the random observation $y$ was then obtained. Given a sequence $y_1$ to $y_n$, of estimated elements can be summed as, such as presented in Equation 56 and 57. $c_n$ represents the cumulative summation of $n$-item, and $\boldsymbol{N}$ represents the natural number.

$$y_i = \frac{v_i}{v_{max}} - 1 \qquad \text{(Equation 55)}$$

$$c_n = y_1 + y_2 + \cdots y_n, \quad n \in \boldsymbol{N} \qquad \text{(Equation 56)}$$

$$c_{n\ items} = \sum_{i=1}^{n} y_n \qquad \text{(Equation 57)}$$

The initial cumulative growth pattern of sigmoid growth plot was then established over the observed data using the generalized logistic function presented in Equation 58. Root sum squared error was then optimized (minimized) using non-linear Generalized Reduced Gradient (GRG) algorithm. A corresponding statistical coefficient of determination (R-Squared) or other desired measures can be considered to select the trial value of the non-linear logistic function parameters. Then predicted observations were computed using Equation 59.

$$f(x) = \frac{l}{1 + e^{[-k(x-x_0)]}} \qquad \text{(Equation 58)}$$

$$y_{Predicted(Normalized)} = \frac{l}{1 + e^{[-k(x-x_0)]}} - \frac{l}{1 + e^{[-k(x-x_0-1)]}} \qquad \text{(Equation 59)}$$

In Equation 59, *l* is the curve maximum saturation parameter, *k* is the steepness of the curve, $x_o$ is the sigmoid midpoint, *x* is hour of the day, and *f(x)* is the cumulative prediction of an observation. Once the start values of the model parameters are estimated, the model functions as a responsive system which normalizes the predicted values. In this case, it is just mathematical reformulation of Equation 55, which is expressed in Equation 60.

$$v_{Predicted} = \frac{v_{max}}{y_{Predicted\ (Normalized)} + 1} \qquad \text{(Equation 60)}$$

### 5.3.2. Non-Linear Logistic Growth Modeling

Up to this point, the initial coefficient (start value) for the logistic function for the hourly mean observed data was estimated. With taking all random observations into consideration, the following formal flow relationship proposed in Equation 61 depending on hour of the day is then proposed and optimized to reduce the root sum squared error and then logistic parameters were estimated. $v_{predicted}$ is the predicted flow before the quantile analysis and $\epsilon$ is the random error portion of the model.

$$v_{Predicted} = \frac{v_{max}}{1 + \frac{l}{1 + e^{[-k(x-x_0)]}} - \frac{l}{1 + e^{[-k(x-x_0-1)]}}} + \epsilon \qquad \text{(Equation 61)}$$

### 5.3.3. Quantile Logistic Function

Theory of quantile functions are well defined by Koenker (2016). Moreover, proposed formal traffic flow model was then analyzed for different quantile or percentile functions. The final non-linear logistic quantile function is presented in Equation 62. Each parameters coefficient was presented by *p*-th (percentile) value and $\epsilon_i$ indicates the random error generated by each quantile functions.

$$v_{Predicted\ (i)} = \frac{v_{max}}{1 + \frac{l^{(p)}}{1 + e^{\left[-k^{(p)}\left(x_i - x_0^{(p)}\right)\right]}} - \frac{l^{(p)}}{1 + e^{\left[-k^{(p)}\left(x_i - x_0^{(p)} - 1\right)\right]}}} + \epsilon_i^{(p)} \qquad \text{(Equation 62)}$$

## 5.4. Traffic Flow Data Characteristics

Five-minute interval flows were converted to an hourly basis in order to find the capacity per hour per lane. The data for the study location has been presented over the hours of the day in the top left section of Figure 17. This Figure indicates the total hourly flow as vehicles per hour (vph) of the link for different hours. Figure 17 indicates non-linear relationship between flow and hours of a day.

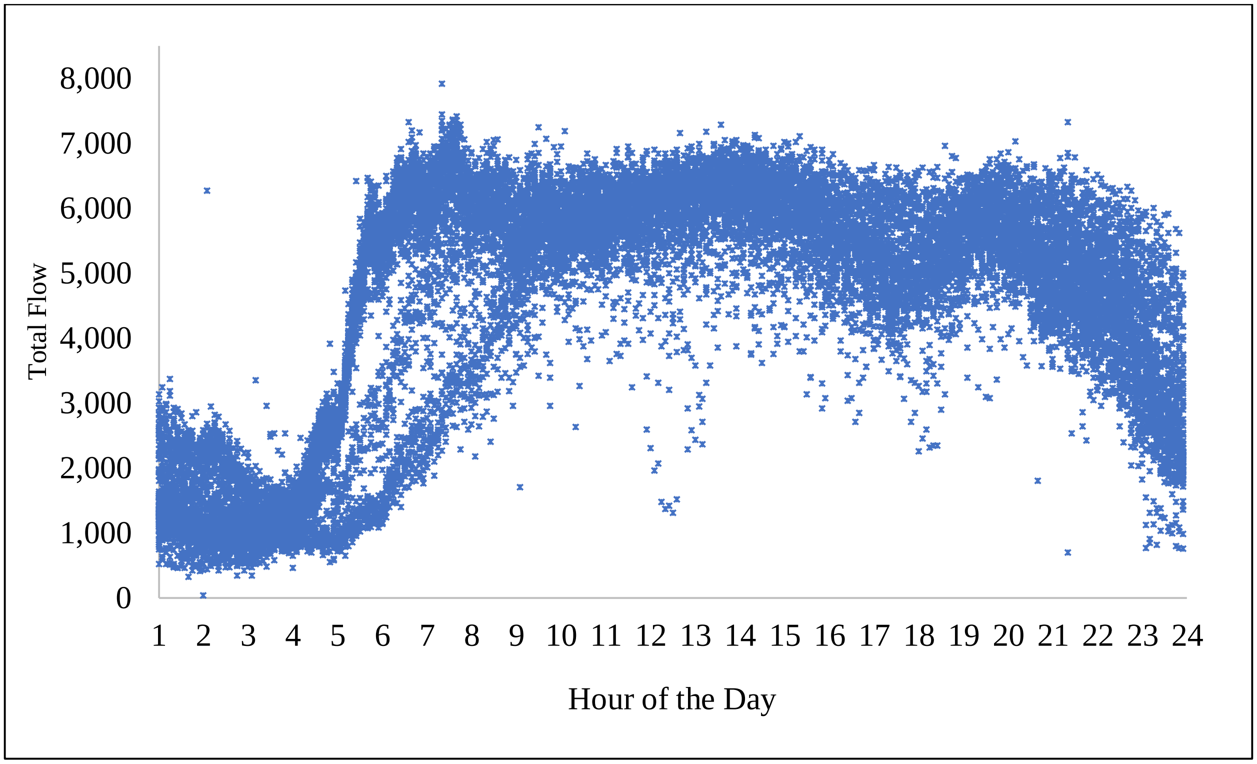


Figure 17. Hourly Flow Characteristics

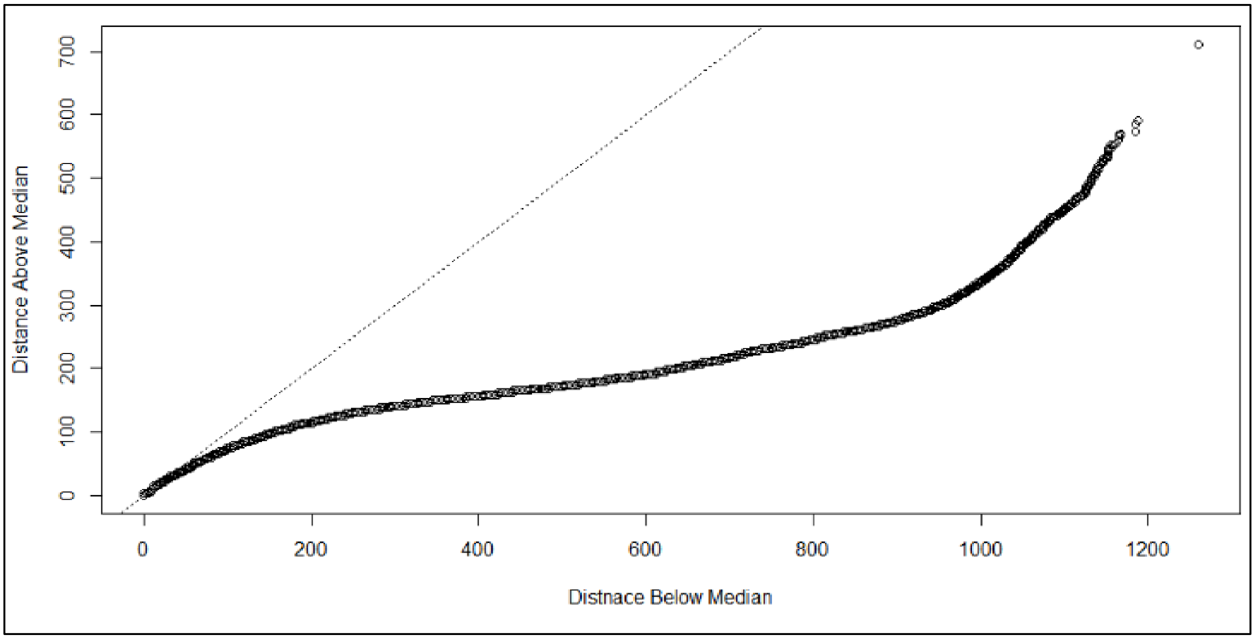


Figure 18. Sys Plot of Hourly Flow

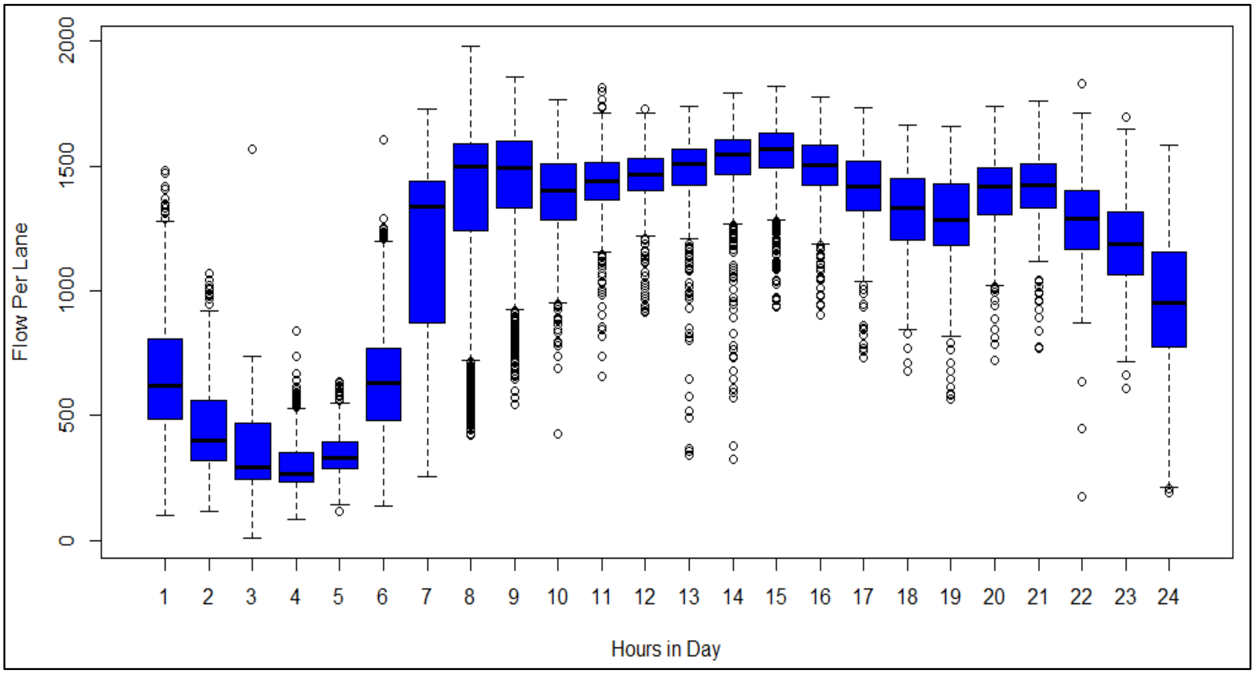


Figure 19. Box Plot of Hourly Flow

Figure 18 presents how the observations deviate from the $50^{th}$- percentile or median value. Furthermore, data was categorized by hourly basis, which has been presented in Figure 19. Statistical tests proved that the flow distribution, mean, and variance of each hour is significantly different from those of other hours.

### 5.5. Results and Discussions

#### 5.5.1. Integrated Quantile Logistic Growth Model

A suitable startup value for three parameters of the logistic growth function was selected based on root sum squared error and R-squared value. For the case study, the starting value of three parameters $(l = 20, k = 1, x_o = 5)$ were optimized to $l = 18.12$, $k = 1.07$, $x_o = 3.09$ with the root sum squared error of 4.55 and R-squared value of 95.08 percent. Figure 20 displays how the initial logistic fitting has been approximated compared to the observed cumulative condition. Figure 21 displays how the initial logistic fitting has been approximated compared to the observed condition. The R-squared values were considerably good and the pattern distribution was approximately similar to the observed data. It can be observed from the top right portion of Figure 21, after around 1 post meridiem (PM), the flow becomes stable. Figure 21 indicates that the freeway reached steady-state flow condition rapidly from 6 after meridiem (AM) to 1 PM

The quantile, *q*, was chosen from *q*= 0.05 to *q* = 1.00, by 0.05 increments. The overall fitted model has been presented in Figure 22. Figure 21 shows the predicted flow conditioning on hours of a day and each quantile. It also includes non-linear least square (NLS), median, 95 percent confidence interval band, and other quantile predicted values. From these Figures, it might be inferred that at 0.025, 0.05 and 0.10 quantile, the model predicted value is almost linear but above 0.10 quantile, the flow rate per lane follows the pattern of a non-linear logistic growth model. It might be inferred that at or below 0.10 quantile, the flow rate is different than other quantile. In addition, lowest flow rate has been observed from 12 AM to 6 AM After 6 AM it increases rapidly and becomes stable during the rest of the day. Plus, after 6 PM, the predicted change of flow for all the quantile seems to be similar.

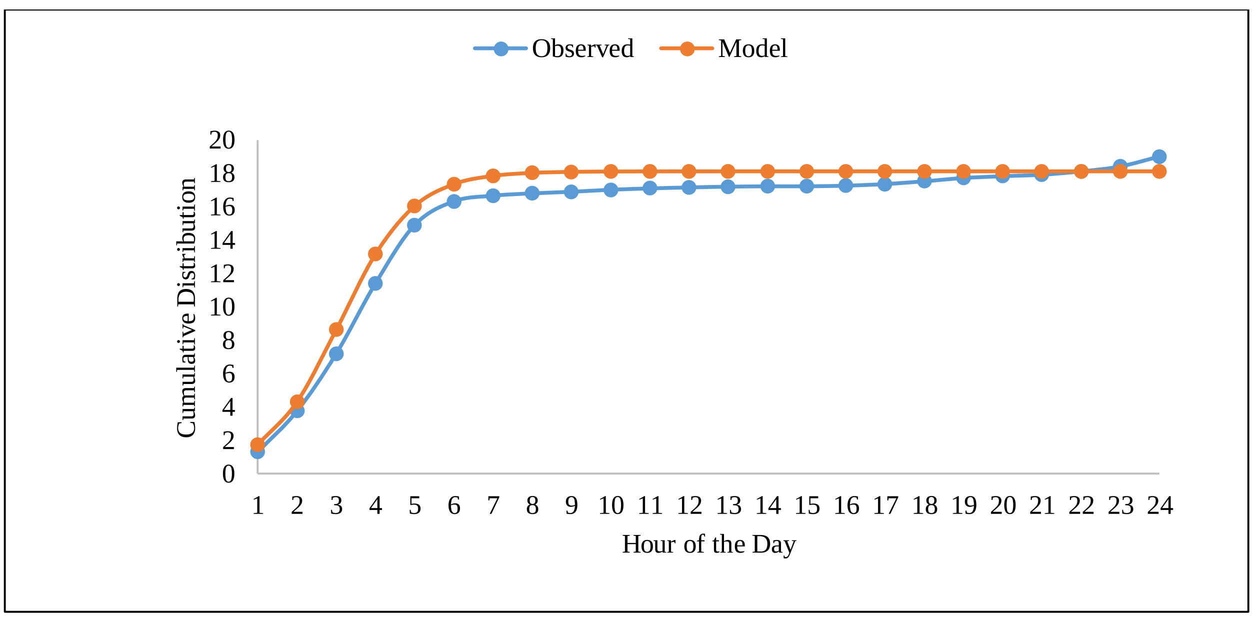


Figure 20. Cumulative Plot of Initial Value

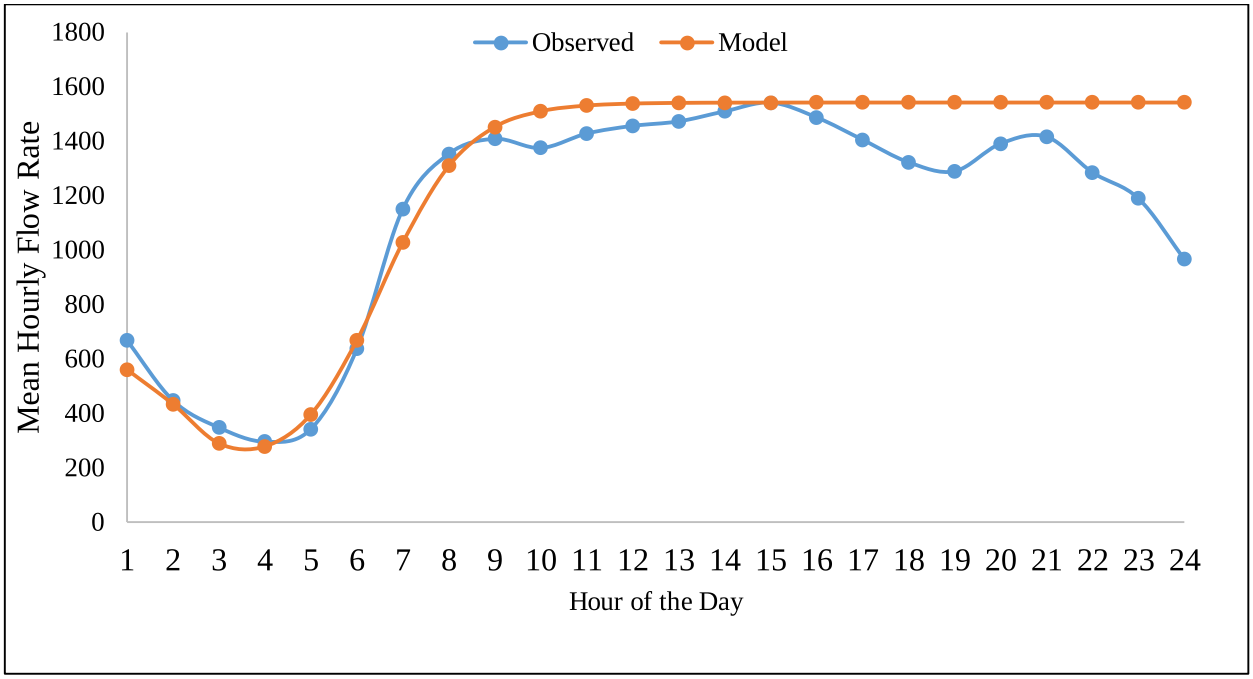


Figure 21. Mean Hourly Flow Rate of Initial Fit

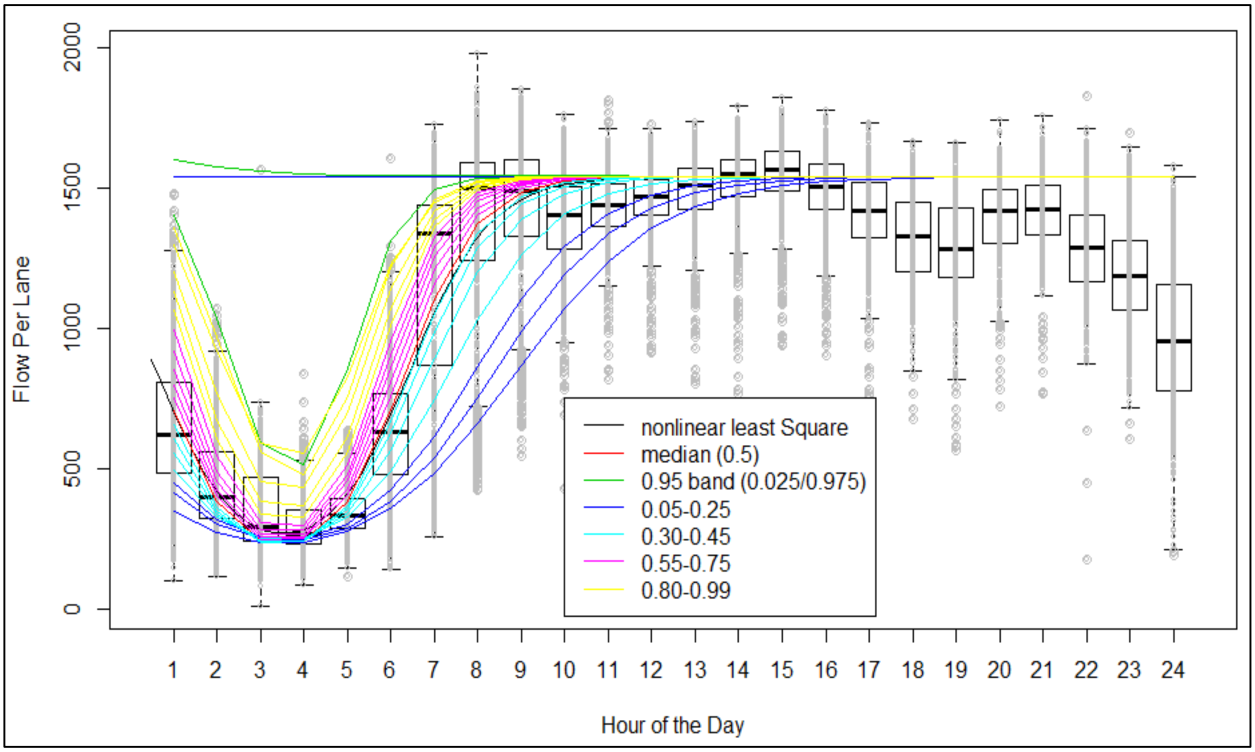


Figure 22. Quantile Model Function Characteristics

### 5.5.2. Parameters and Test Statistics

Various statistical tests were performed in order to identify the best suitable stable portion of the fitted model so that it could find the steady-state capacity. Three parameters *l*, *k*, $x_0$, and standard error, *t*-value and p-value corresponding to each quantile are presented in Table 14. It can be seen that at or below the 10$^{th}$ percentile, the saturation parameter *l* becomes so high compared to other quantile.

Table 14. Model Parameters and Test Statistics of Proposed Model

| | | Curve Maximum Saturation (*l*) | | | | | Steepness of the Curve (*k*) | | | | | Sigmoid Midpoint ($x_0$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Qu. | | Coeff. | | Std. Error | t-value | p-value | Coeff. | | Std. Error | t-value | p-value | Coeff. | | Std. Error | t-value | p-value |
| 5 | ** | 292.06 | * | 0 | - | 0 | -0.26 | | 72.284 | 0.0 | 0.997 | -54.95 | | 14351.293 | 0.0 | 0.997 |
| 10 | ** | 573.89 | * | 0 | - | 0 | -0.18 | | 21.361 | 0.0 | 0.993 | -78.77 | | 9069.201 | 0.0 | 0.993 |
| 15 | ** | 37.75 | * | 0.348 | 108.50 | 0 | 0.6 | * | 0.005 | 131.1 | 0 | 2.97 | * | 0.049 | 60.7 | 0 |
| 20 | ** | 32.48 | * | 0.252 | 128.66 | 0 | 0.67 | * | 0.004 | 155.0 | 0 | 3.11 | * | 0.028 | 111.1 | 0 |
| 25 | ** | 28.95 | * | 0.22 | 131.40 | 0 | 0.73 | * | 0.005 | 147.4 | 0 | 3.08 | * | 0.025 | 124.9 | 0 |
| 30 | ** | 25.86 | * | 0.215 | 120.02 | 0 | 0.84 | * | 0.008 | 101.5 | 0 | 3.01 | * | 0.021 | 145.4 | 0 |
| 35 | ** | 23.76 | * | 0.164 | 144.88 | 0 | 0.97 | * | 0.011 | 92.0 | 0 | 2.98 | * | 0.013 | 225.6 | 0 |
| 40 | ** | 22.48 | * | 0.166 | 135.39 | 0 | 1.06 | * | 0.01 | 103.4 | 0 | 2.98 | * | 0.015 | 196.7 | 0 |
| 45 | ** | 21.43 | * | 0.149 | 144.18 | 0 | 1.13 | * | 0.012 | 92.2 | 0 | 2.99 | * | 0.013 | 225.1 | 0 |
| 50 | ** | 20.27 | * | 0.158 | 128.43 | 0 | 1.18 | * | 0.011 | 103.1 | 0 | 3 | * | 0.014 | 219.0 | 0 |
| 55 | ** | 19.1 | * | 0.147 | 129.99 | 0 | 1.22 | * | 0.012 | 105.1 | 0 | 2.99 | * | 0.014 | 218.4 | 0 |
| 60 | ** | 17.92 | * | 0.154 | 116.23 | 0 | 1.27 | * | 0.015 | 85.7 | 0 | 2.99 | * | 0.016 | 182.3 | 0 |
| 65 | ** | 16.55 | * | 0.147 | 112.77 | 0 | 1.31 | * | 0.012 | 113.6 | 0 | 3 | * | 0.013 | 223.2 | 0 |
| 70 | ** | 15.26 | * | 0.147 | 103.52 | 0 | 1.35 | * | 0.013 | 104.9 | 0 | 3.02 | * | 0.015 | 201.2 | 0 |
| 75 | ** | 13.68 | * | 0.164 | 83.23 | 0 | 1.4 | * | 0.014 | 102.6 | 0 | 3.04 | * | 0.015 | 204.3 | 0 |
| 80 | ** | 11.9 | * | 0.179 | 66.34 | 0 | 1.43 | * | 0.014 | 100.9 | 0 | 3.06 | * | 0.014 | 218.2 | 0 |
| 85 | ** | 10.09 | * | 0.176 | 57.25 | 0 | 1.44 | * | 0.015 | 97.2 | 0 | 3.07 | * | 0.015 | 201.3 | 0 |
| 90 | ** | 8.01 | * | 0.163 | 49.25 | 0 | 1.45 | * | 0.016 | 90.7 | 0 | 3.08 | * | 0.02 | 151.9 | 0 |
| 95 | ** | 6.13 | * | 0.071 | 86.41 | 0 | 1.6 | * | 0.02 | 80.0 | 0 | 3.22 | * | 0.016 | 195.9 | 0 |
| 99 | ** | 5.52 | * | 0.193 | 28.55 | 0 | 1.44 | * | 0.067 | 21.5 | 0 | 3.11 | * | 0.041 | 75.9 | 0 |

* Significantly different from zero (5 percent significance)

** Significantly different from non-linear least square method (5 percent significance)

Above 10$^{th}$ percentile all three parameters coefficients and *t*-values are positively significant. There were two hypotheses tested against the model. The first hypothesis was tested to examine the significance of each quantile coefficient against zero. This test results are indicated by single asterisk (*) symbol. Test results also indicate each quantile parameter coefficient has significant effect on flow prediction. Second hypothesis was that whether each quantile analysis was significantly different than NLS regression model. The test results were presented by double asterisk (**) symbol. Test results suggest that all the quantile model is significantly different than the NLS model. Therefore, it might be inferred that non-linear model is robust compared to the non-linear least square method.

### 5.5.3. Model Sensitivity

Furthermore, each parameter sensitivity or response pattern against each quantile was simulated. Figure 23 presented sensitivity of three statistical coefficient parameters over different quantile or as a function of *q*.

The rate of change of curve maximum saturation point *l* has been centered approximately at its 60$^{th}$ percentile of the flow, therefore, it estimates the quantile function of traffic flow depending on 60$^{th}$ percentile flow. Because at this percentile, saturation parameters are not significantly different than the NLS saturation parameters. However, steepness of the quantile function does not vary significantly from the NLS method at around 40$^{th}$ percentile value. Then, to find the best model either 40$^{th}$ percentile model or 60$^{th}$ percentile model should be selected. This is why the capacity value should indicate its confidence level or range. But Figure 23 shows that the rate of change of the third parameters is robust for different quantile. One may consider a different quantile function to identify the tolerance value of capacity. Considering this case, the capacity value might be cut off and approximated from 40$^{th}$ percentile to 60$^{th}$ percentile of the traffic flow observed for the case study location.

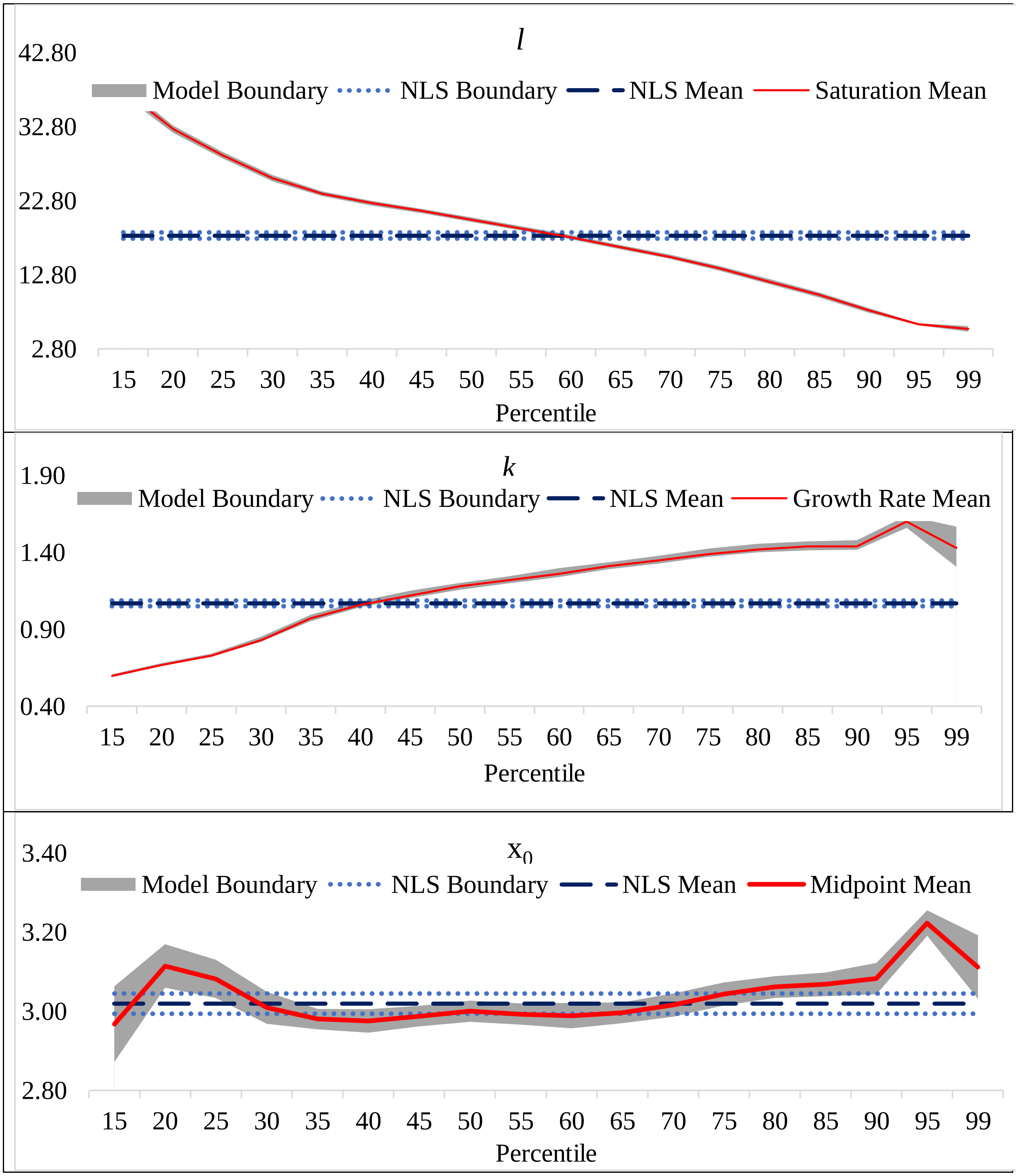


Figure 23. Rate of Change of Quantile Function Parameters

### 5.5.4. Capacity Estimation

After approximation and establishment of a non-linear quantile logistic function for the flow prediction conditioning hours of a day and τ, the next step is to approximate the capacity from this model. Figure 24 shows the first and second derivative for the 40th and 60th percentile quantile function. It is evident that the model first derivative converges to zero around hour 3:47 of the day which is a local maximum for both 40th and 60th percentile function. First derivative cannot converge to zero after the hour 3:47 of the day for both cases. But using the second derivative, global maxima and main point of inflation can be observed at hour 6:59 of the day for 40th percentile and hour 6.08 of the day for 60th percentile. The stable capacity might be observed after hour 13:00 of the day. For example, both 40th and 60th percentile of the non-linear quantile logistic function at hour 13 would generate around 1,550 vph per lane, which might be identify as the steady-stated capacity.

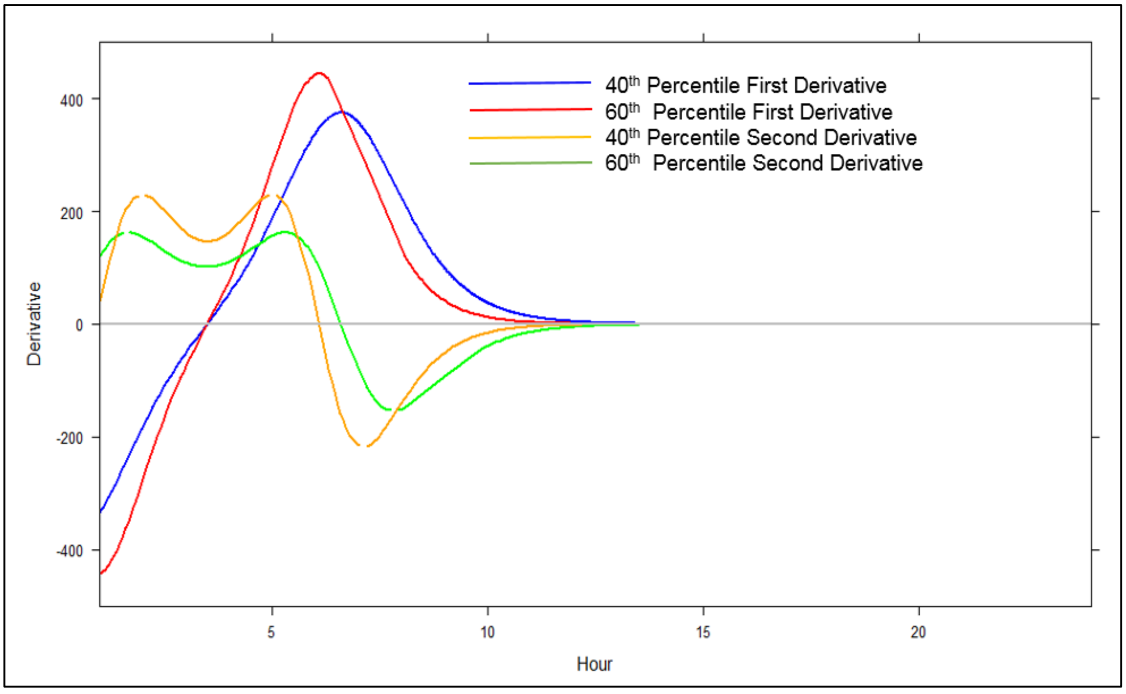


Figure 24. Quantile Simulation of Proposed Model

### 5.5.5. Validation

The results have been compared to several smaller to larger public agencies freeway capacity values which are presented in Figure 25. It can be seen that the proposed model predicts low capacity on freeways. Freeway capacity has been observed in other public agencies ranged from 1,700 to 2,250 vph per lane (Figure 25). The used capacity of various organizations is collected from the published report of each organization. But the proposed model in this research predicts 1,550 vph. The proposed model generates significantly different capacity values compared to the than mean capacity of the reported organizations (with p-value of 0.00012).

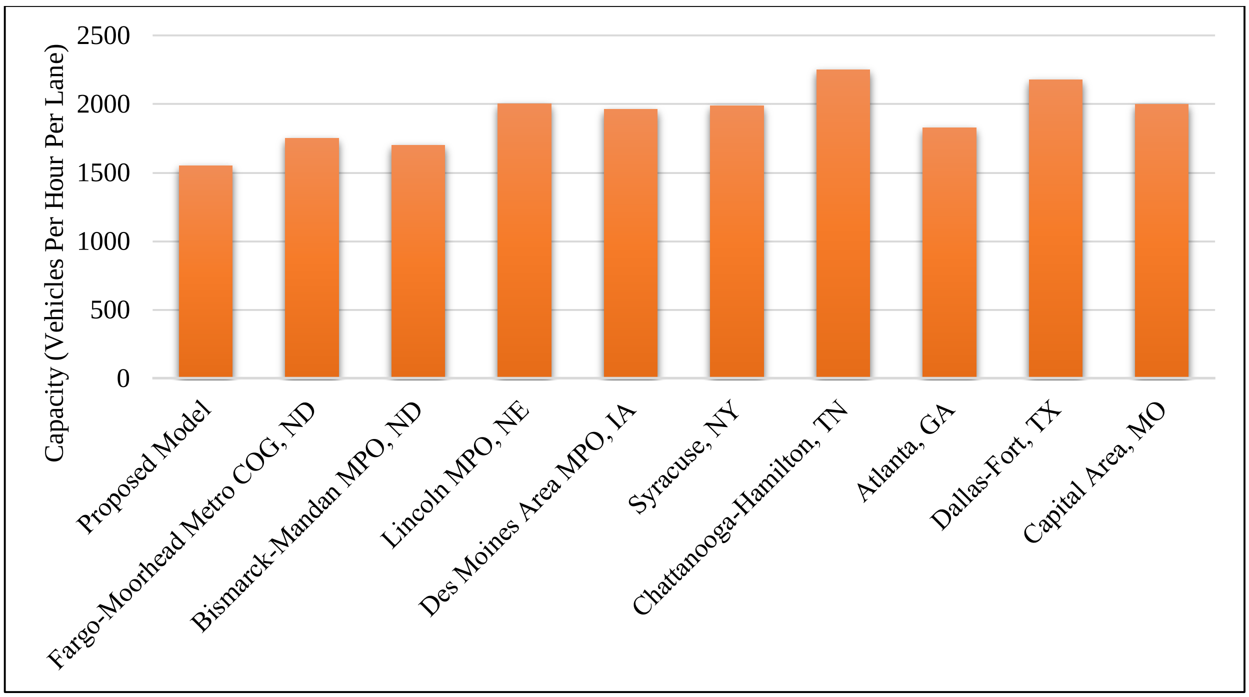


Figure 25. Capacity among MPOs

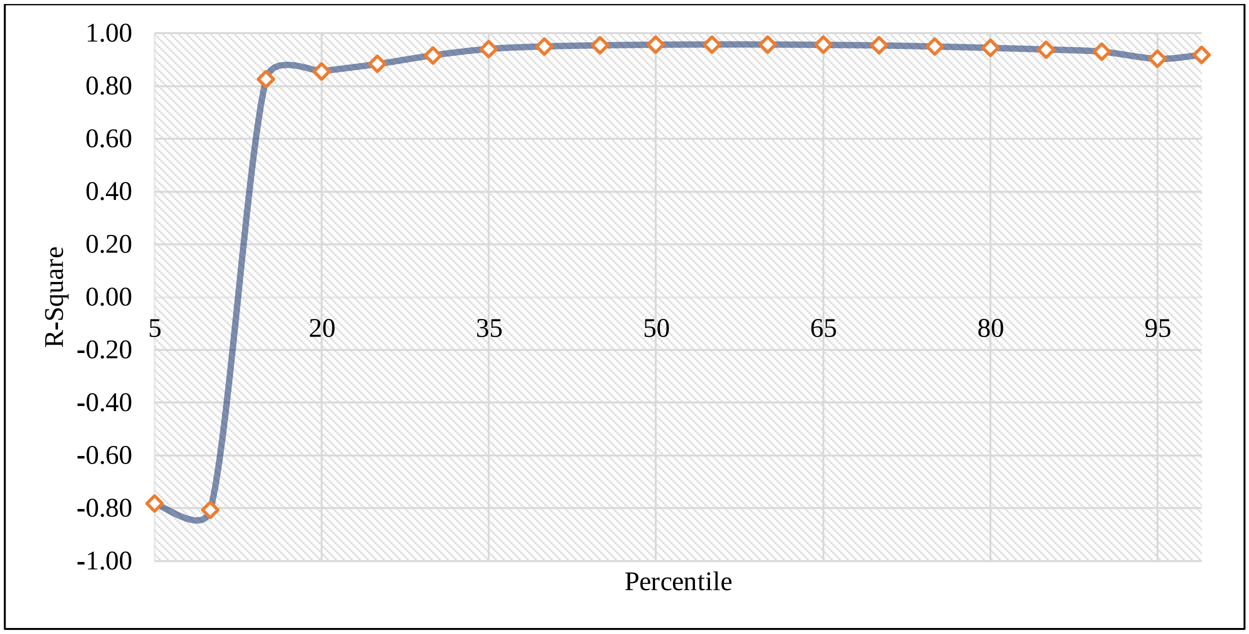


Figure 26. R-Square by Percentiles

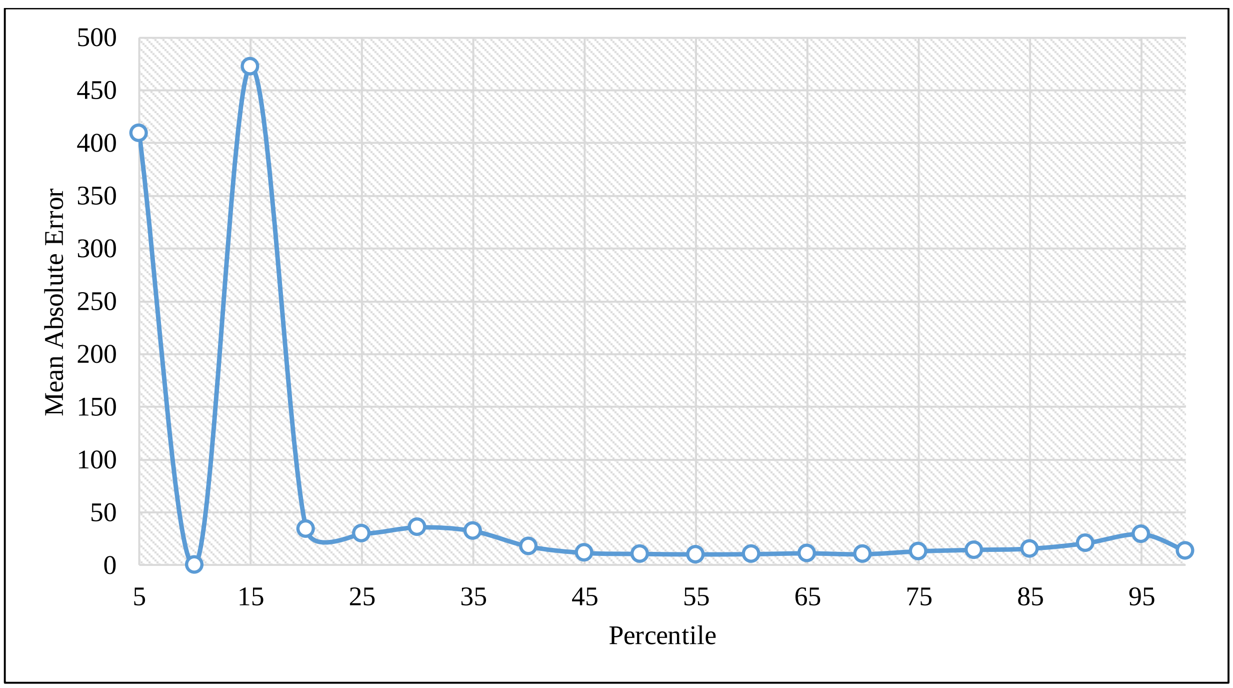


Figure 27. Simulated Error of Modeling Errors

In addition, simulation for the R-squared value and mean absolute error (MAE) were conducted. Results are presented in Figures 26-27 for each quantile function. These figures reflect how the R-squared value has been changed over each quantile. At quantile above 0.10, the R-squared values are strong and positively correlated with a range of 82.62 to 95.78 percent. At quantile below 0.10, the R-squared values indicate strong negative correlation. Negative R-square indicates that at or below 0.10 quantile, the model does not follow the trend of the observed data. A horizontal fitted line will be more reasonable in that case. It can be inferred that even though the model is generating significantly different results in compared to the reported capacity values of different organizations, this model is more robust in predicting traffic flow and capacity in each quantile function. It was observed that at or below the 15$^{th}$ percentile quantile function, model MAE is high; however, above the 15$^{th}$ percentile function, the error is not so significant and always remains below 36. It can be inferred that traffic flow has very distinguished characteristics at or below 15$^{th}$ percentile function. It may also be inferred that model is more robust and strong in predicting traffic flow above 15$^{th}$ percentile. It could be stated that model has a good capability of predicting any kind of non-linearity formation of traffic flow characteristics.

# 6. INVESTIGATION OF FREE-FLOW SPEED MODEL

## 6.1. General

TRB (2000, p.5-6) defined FFS as “(1) The theoretical speed of traffic, in kilometers per hour, when density is zero, that is, when no vehicles are present; (2) the average speed of vehicles over an urban street segment without signalized intersections, under conditions of low volume; (3) the average speed of passenger cars over a basic freeway or multilane highway segment under conditions of low volume.” Wang and Huegy (2014, p.3) said, “free-flow speed is the theoretical speed of traffic as density approaches zero, so it is the basis for speed estimation and should also be determined carefully.”

## 6.2. Introduction

With travel demand modeling, FFS has an important role. This speed is used to calculate the initial FFT and the congested travel time for highway-assignment steps. Motuba (2012) stated that the wrong FFS may affect the TDM’s output. In the absence of local data, the FFS are calibrated based on national averages, area types, and available data for similar roadways.

Based on the current HCM 2010 (TRB, 2010), the freeway’s FFS can be determined with Equation 63. Here *FFS* is the free-flow speed (mph), $f_{LW}$ is the adjustment for lane width, $f_{LC}$ is the adjustment for the right-side clearance, and *TRD* is the total ramp density (ramps per mile). Wang and Huegy (2014) indicated that FFS can be reasonably obtained from the HCM with calibration for the regional TDM. Their literature review indicated how FFS is being utilized for regional travel demand, mostly by using a look-up table.

$$FFS = 75.4 - f_{LW} - f_{LC} - 3.22\, TRD^{0.84} \qquad \text{(Equation 63)}$$

Using Equation 63, it is easier to calculate the *FFS*. The question is whether this generalized formula can be applied to congested places such as Los Angeles. HCM are widely used, recognized, applicable, acceptable, and recommended guidelines for FFS estimation. By definition and theoretically, the FFS formulation should relate to the traffic-flow speed while the density is zero. The transportation industry may address this issue with FFS estimation. There are many speed-density models are available. Wang et al. (2009) presented and investigated some deterministic approaches and developed a stochastic process to model speed-density relationship. They included how different deterministic models are criticized over one another. Moreover, each model had its own success or failure story. The speed-density model can be managed with different perspectives (single-regime or multi-regime models). Multi-regime models were based on traffic-flow characteristics for the free-flow portion and the congested portion of the speed-density curve.  The modeling partition of multi-regime speed-density models depended on density. Wang et al. (2009) observed that that Edie's two regime models' cut-off density was less than or equal to 50, the modified Greenberg model's cut-off density is less than or equal to 35, and May's two-regime model is less than or equal to 65. It might be inferred that free-flow regimes are addressed by different threshold values of density, which ranges from 35-65. Using such a high value showed that the LOS for FFS is from *E* to *F* (TRB, 2010). In that sense, estimating FFS based on the current multi-regime model would be questionable. Free-flow conditions should represent the LOS *A*, where density is less than or equal to 11 (TRB, 2000), which is a free-flow operation. Mtoi and Moses (2014) used 10 passenger cars per hour per mile per lane for the uninterrupted flow facilities' free-flow estimation. Moses et al. (2013) used density 5 as the cut-off level. They were not certain which percentiles' ($50^{th}$, $85^{th}$, and $90^{th}$) speed would be useable to compute the FFS. Their studies addressed how speed limit, access-point density, median type, curve presence, segment length, the number of lanes, and area type had a significant influence on the FFS estimation.

Considering these issues, there are certain aspects that need to be incorporated into the speed-density model.

- The free-flow regime cut-off or threshed-density value for FFS estimation should incorporated LOS *A*.
- The existing deterministic speed-density model should merit an investigation which overcomes the aforementioned first issue.
- A statistically sound and reasonable quantile model might be established.

Therefore, this study aimed to investigate the current deterministic speed-density model and proposed a modified model to explain speed-density model with different quantile and to compute the FFS accordingly.

**6.3. Methodology**

The research methodology included three major components: 1) investigating the single-regime model, 2) investigating the multi-regime model, and 3) proposing the most suitable, quantile speed-density model. All three sections are discussed in this section.

At first, a comparative investigation was performed using Greenshields, Greenberg, Underwood, Northwestern, and Drake's models. The dataset was utilized to develop those five models. To find the FFS, the formal relationship was modified in a simpler form (Table 15). A more formal definition for each relationship is presented in Wang et al. (2009) study.

Table 15. Single-Regime Models

| Model | Relationship | Source |
|---|---|---|
| Greenshields | $y = b - ax$ | Wang et al. (2009) |
| Underwood | $y = exp(b + ax)$ | Wang et al. (2009) |
| Greenberg | $y = a\ log\left(\frac{1}{x} + b\right)$ | Wang et al. (2009) |
| Northwestern | $y = exp\left(b - \frac{a}{2}x^2\right)$ | Wang et al. (2009) |
| Drake | $y = exp\left(b + \frac{a}{2}x^2\right)$ | Wang et al. (2009) |

where *y* is the speed, *x* is the density, and *a* and *b* are parameters.

At the second stage, two regime models were investigated and compared among the Greenshields, Greenberg, Underwood, Northwestern, Drake, Edie's two-regime, May's two-regime, and the modified Greenburg models which are presented in Wang et al. (2009). Interested readers are referred to Wang et al.'s (2009) paper. Two regimes (the free-flow and congested regimes) considered the LOS *A* which indicates that the density is less than or equal to 10 vehicles per hour per mile per lane. To find the FFS, the formal relationship was modified in a simpler form and presented in Table 16. More formal definitions for each relationship is presented in Wang et al. (2009). In this study, Edie, two-regime, and modified Greenberg models were considered based on the cut-off level at density 10 instead of the formal cut-off levels.

Best candidate model using the single-regime or multi-regime modeling approach was identified. Later, speed-density model was then analyzed for different quantile or percentile for the best candidate model. Finally, based on statistical measures, a suitable model was proposed and FFS was computed.

Table 16. Multi-Regime Models

| Model | Free-Flow Regime | Congested Regime | Source |
|---|---|---|---|
| Greenshields | $y = b - ax$ *(where x≤10)* | $y = b - ax$ *(where x >10)* | Proposed |
| Underwood | $y = exp(b + ax)$ *(where x≤10)* | $y = exp(b + ax)$ *(where x >10)* | Proposed |
| Greenberg | $y = a\ log\left(\frac{1}{x} + b\right)$ *(where x≤10)* | $y = a\ log\left(\frac{1}{x} + b\right)$ *(where x >10)* | Proposed |
| Northwestern | $y = exp\left(b - \frac{a}{2}x^2\right)$ *(where x≤10)* | $y = exp\left(b - \frac{a}{2}x^2\right)$ *(where x >10)* | Proposed |
| Drake | $y = exp\left(b + \frac{a}{2}x^2\right)$ *(where x≤10)* | $y = exp\left(b + \frac{a}{2}x^2\right)$ *(where x >10)* | Proposed |
| Edie | $y = 54.9\ exp\left(\frac{-x}{163.9}\right)$ *(where x≤50)* | $y = 26.8\ ln\left(\frac{162.5}{x}\right)$ *(where ≥50)* | Wang et al. (2009) |
| Two-regime | $y = 60.90 - 0.515x$ *(where x≤65)* | $y = 40 - 0.265x$ *(where ≥65)* | Wang et al. (2009) |
| Mod. Greenberg | $y = 48$ *(where x≤35)* | $y = 32\ ln\left(\frac{145.5}{x}\right)$ *(where ≥35)* | Wang et al. (2009) |

where *y* is the speed, *x* is the density, and *a* and *b* are parameters.

## 6.4. Results and Discussions

### 6.4.1. Data Characteristics

Five-minute interval speeds were analyzed in order to see how the speed data are distributed daily. The study location's speed data are presented for the day's hours in Figure 28. This figure indicates the link's total hourly speed (mph) for different hours. Figure 28 illustrates the non-linear relationship between speed and the day's hours. Statistical tests proved that the speed distribution, mean, and variance for each hour is significantly different from those data for other hours. From midnight to 2:00 PM, a normal day may contain many outliers which are outside the inter-quartile range. Until 6 AM in the morning, the speed's mean and variance seem to be constant, but later, these two parameters' variation is significant. The quantile-quantile *(Q-Q)* plot in Figure 29 represents how the speed distribution over the day's hours deviated from a straight line. The figure basically shows that the distribution is skewed. This statement can be supported with the bi-modal distribution of speed that is presented in Figure 30. Figure 31 illustrates how the observations deviate from the 50th percentile, the median value. The figure shows how the speed distribution is multiplicative in nature.

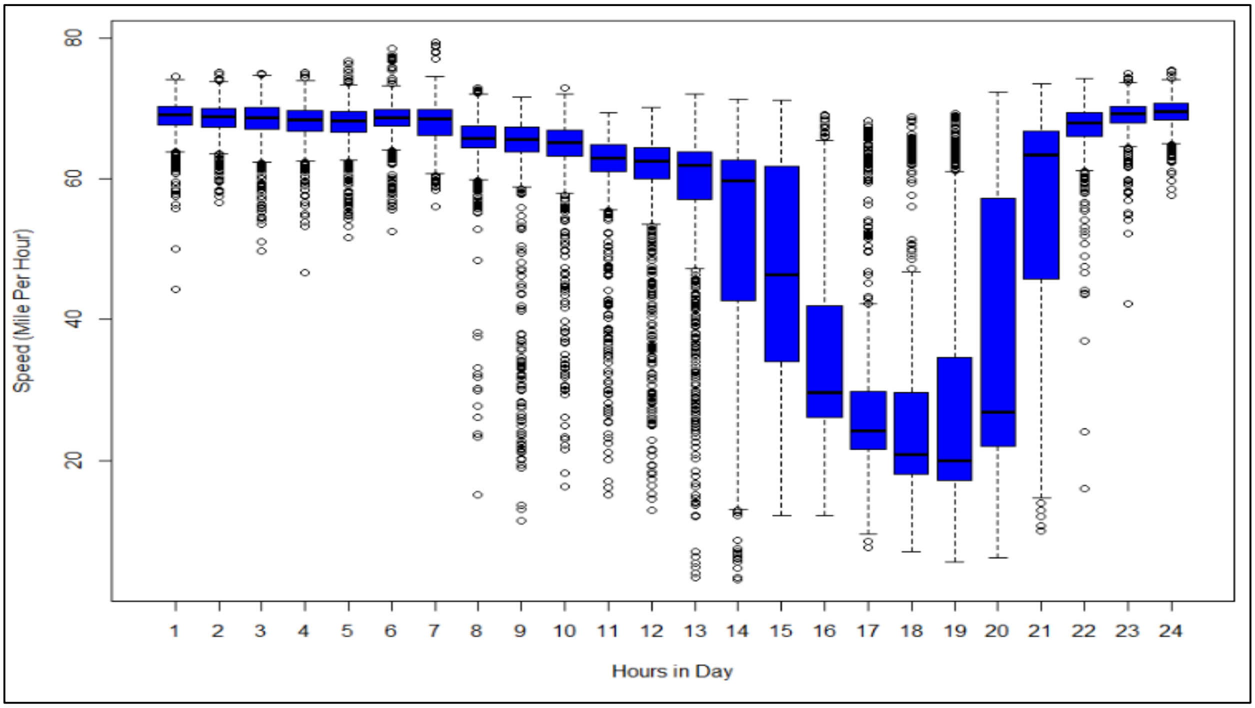


Figure 28. Box Plot of the Hourly Speed Distribution

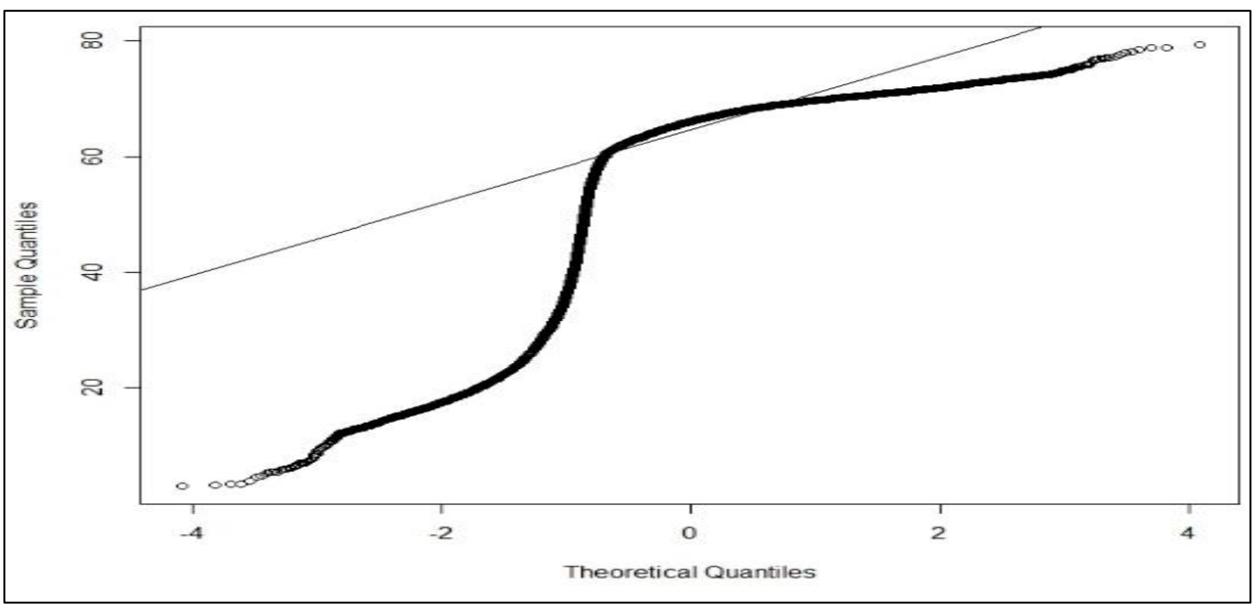


Figure 29. Q-Q Plot of the Hourly Speed

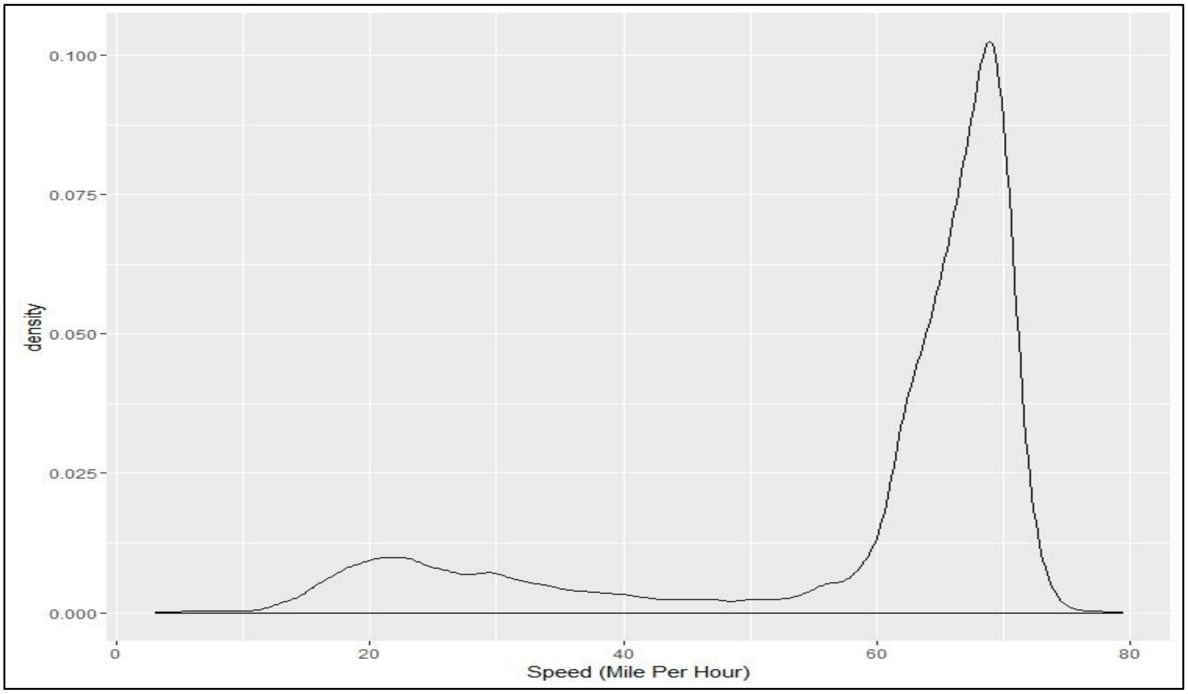


Figure 30. Density of the Hourly Speed

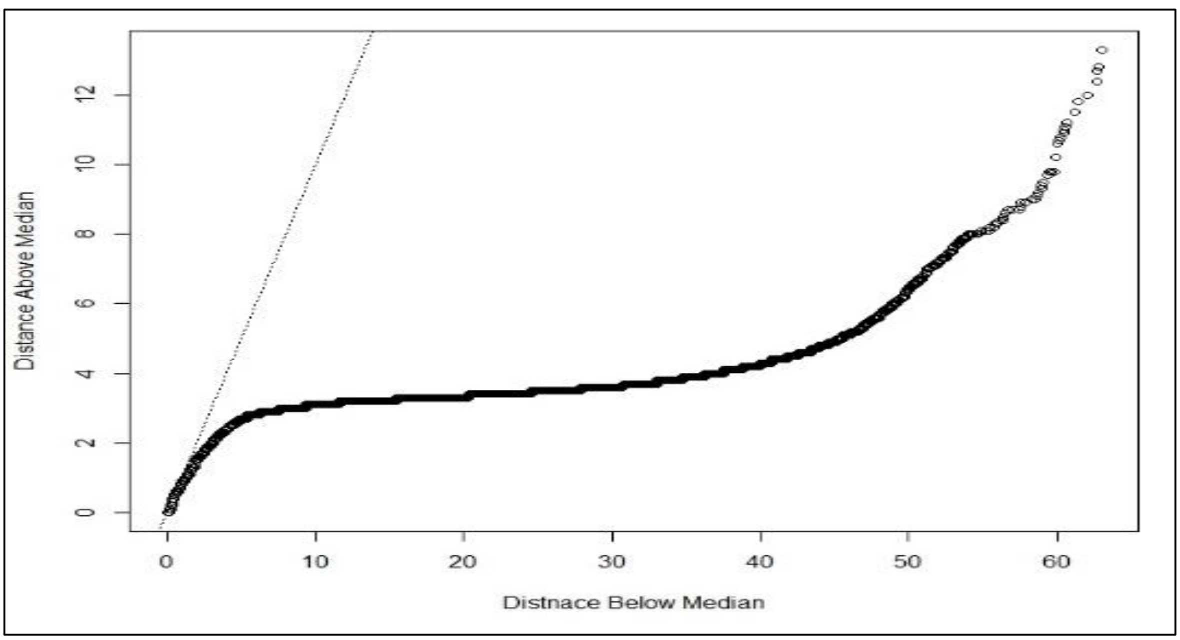


Figure 31. Speed-Data Characteristics

Later, the speed (mph) over density (vehicles per hour per mile per lane) data were analyzed. Figure 32 shows how the relationship would look. With a deterministic approach, these relationships were developed in different settings: 1) single regime and 2) multi regime. Regardless of the regime, the relationship could be linear, log-normal, log-transformed, exponential, etc. The Greenshields, Greenberg, Underwood, Northwestern, Drake, Edie's two-regime, May's two-regime, and modified Greenburg model are representations of this relationship.

The first goal is to see which model better replicates the overall data if there is a single-regime model. The second goal is to see whether the two-regime model is better. To create a two-regime model, speed data were extracted from the study area where the density was less than or equal to 10. These criteria were chosen based on the LOS *A*. In Figure 32, the bottom-left side shows the free-flow regime, and the bottom-right side shows the congested regime.

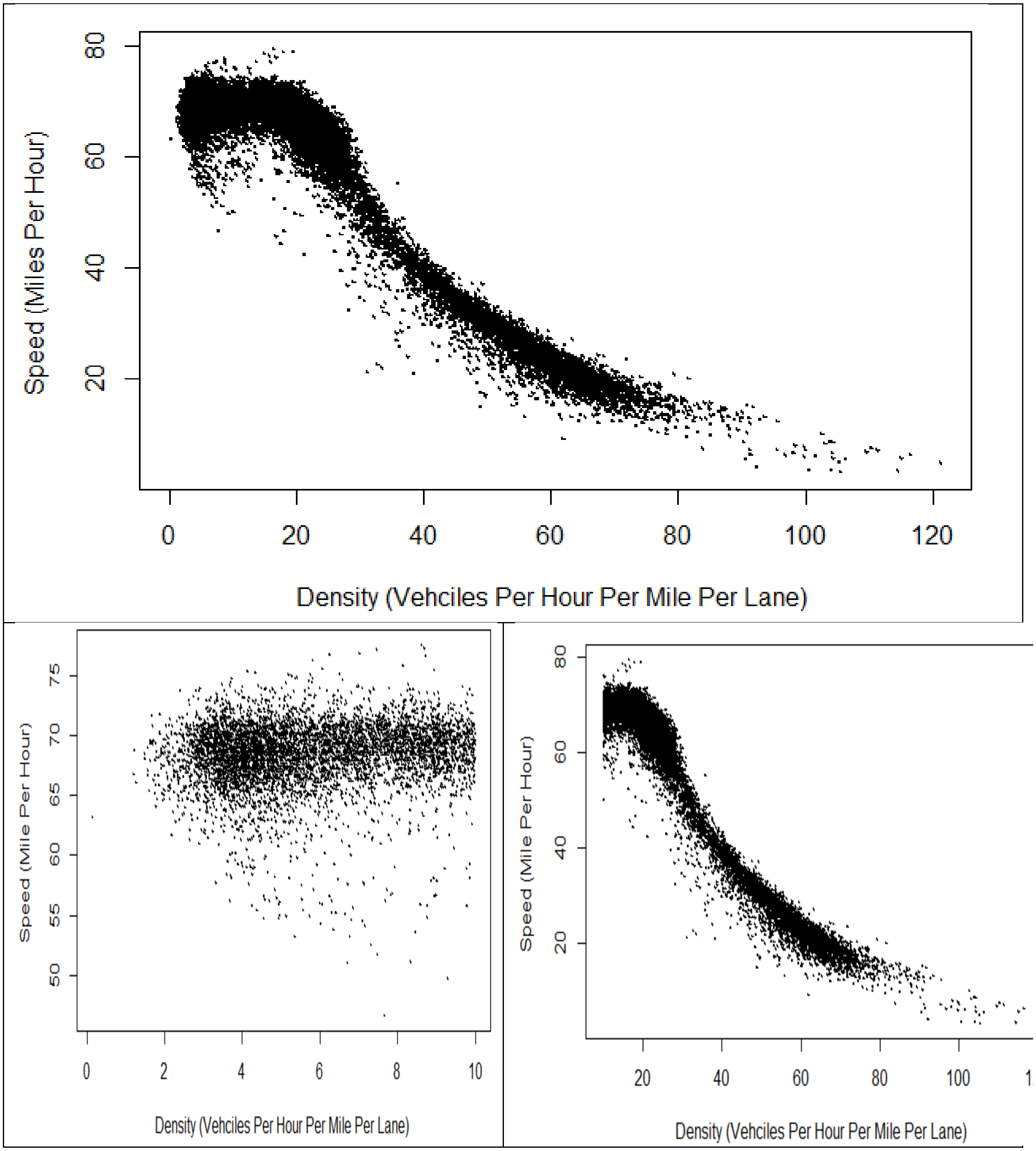


Figure 32. Speed-Density Scatter Plot

### 6.4.2. Investigate Single-Regime Model

The single-regime model was investigated; it is reported in Table 17. The results indicate that, for each model, both parameters are statistically significant. The expected sign for each of the model's coefficient parameters is positive, which happens in every model except the Underwood model. The Underwood model shows anomaly for the coefficient estimate's sign. The fitted speed-density model is presented in Figure 33. The Northwestern and Drake models are complement of each other.

Table 17. Single-Regime Model Comparisons

| Model | Coefficient | Estimate | Std. Error | t-value | Pr(>\|t\|) |
|---|---|---|---|---|---|
| Underwood | a | -0.01 | 0.00006 | -234.2 | <2e-16 *** |
| | b | 4.39 | 0.00131 | 3358.8 | <2e-16 *** |
| Greenshields | a | 0.85 | 0.00213 | 397.8 | <2e-16 *** |
| | b | 78.37 | 0.06245 | 1255 | <2e-16 *** |
| Greenberg | a | 14.08 | 0.08639 | 163 | <2e-16 *** |
| | b | 98.42 | 0.25387 | 387.7 | <2e-16 *** |
| Northwestern | a | 0.00 | 0.00000 | 389.3 | <2e-16 *** |
| | b | 4.28 | 0.00053 | 8089 | <2e-16 *** |
| Drake | a | 0.00 | 0.00000 | -389.3 | <2e-16 *** |
| | b | 4.28 | 0.00053 | 8089 | <2e-16 *** |
| Significance level: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' and 1 | | | | | |

The R-square value and the residual standard error (RSE) are included in Table 18. The Northwestern and Drakes models show a maximum R-square value of 93.97 percent with the lowest RSE at 4.034. On the other hand, the Greenberg model has the lowest R-square value of 54.02 percent with the highest standard error at 11.14.

Table 18. R-Square Value for a Single-Regime Models

| Model | R-Square | RSE |
|---|---|---|
| Underwood | 79.63 | 7.414 |
| Greenshields | 87.5 | 5.808 |
| Greenberg | 54.02 | 11.14 |
| Northwestern | 93.97 | 4.034 |
| Drake | 93.97 | 4.034 |

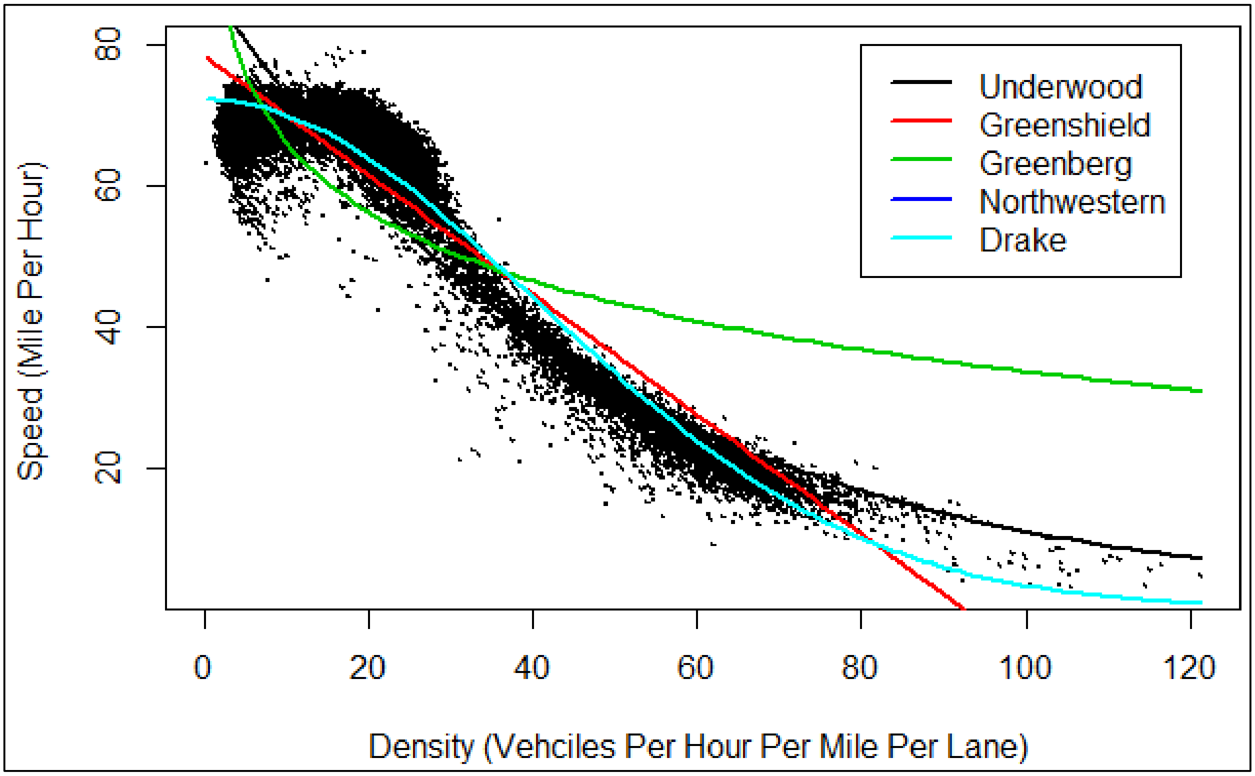


Figure 33. Single-Regime Model Fittings

### 6.4.3. Investigating Multi-Regime Model

The multi-regime model, which includes the free-flow model and the congested regime model, was developed and reported in Table 19. The results indicated that, for each model, both parameters are statistically significant.

In the free-flow model, the sign of the coefficient estimates is positive for every case except the Greenshields, Greenburg, Northwestern, and Edie models. However, with the congested regime, all parameters are positive, except Underwood model parameters. The fitted speed-density model is presented in Figures 34 and 35.

Table 19. Multi-Regime Model Comparisons

| | | Free-flow Regime | | | | | Congested Regime | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | Coef | Estimate | Std. Error | t-value | Pr(>\|t\|) | | Estimate | Std. Error | t-value | Pr(>\|t\|) | |
| Underwood | a | 0.002021 | 0.0002326 | 8.692 | <2e-16 | *** | -0.02233 | 6.93E-05 | -322.3 | <2e-16 | *** |
| | b | 4.213833 | 0.0014139 | 2980.275 | <2e-16 | *** | 4.61700 | 1.66E-03 | 2778.7 | <2e-16 | *** |
| Greenshields | a | -0.138320 | 0.01593 | -8.683 | <2e-16 | *** | 1.01950 | 0.002027 | 503 | <2e-16 | *** |
| | b | 67.610970 | 0.09657 | 700.102 | <2e-16 | *** | 85.50897 | 0.06964 | 1228 | <2e-16 | *** |
| Greenberg | a | -0.720300 | 0.08611 | -8.364 | <2e-16 | *** | 33.25553 | 0.09766 | 340.5 | <2e-16 | *** |
| | b | 67.196480 | 0.14749 | 455.594 | <2e-16 | *** | 163.44622 | 0.32193 | 507.7 | <2e-16 | *** |
| Northwestern | a | -0.000329 | 0.0000378 | -8.689 | <2e-16 | *** | 0.00069 | 1.68E-06 | 413.8 | <2e-16 | *** |
| | b | 4.219308 | 0.0008502 | 4962.808 | <2e-16 | *** | 4.32700 | 6.65E-04 | 6507.3 | <2e-16 | *** |
| Drake | a | 0.000329 | 0.0000378 | 8.689 | <2e-16 | *** | -0.00069 | 1.68E-06 | -413.8 | <2e-16 | *** |
| | b | 4.219308 | 0.0008502 | 4962.809 | <2e-16 | *** | 4.32700 | 6.65E-04 | 6507.3 | <2e-16 | *** |
| | a | 0.002021 | 0.0002326 | 8.692 | <2e-16 | *** | 33.25553 | 0.09766 | 340.5 | <2e-16 | *** |
| Edie | b | 4.213833 | 0.0014139 | 2980.275 | <2e-16 | *** | 163.44622 | 0.32193 | 507.7 | <2e-16 | *** |
| | a | -0.138320 | 0.01593 | -8.683 | <2e-16 | *** | 1.01950 | 0.002027 | 503 | <2e-16 | *** |
| Two Regime | b | 67.610960 | 0.09657 | 700.102 | <2e-16 | *** | 85.50897 | 0.06964 | 1228 | <2e-16 | *** |
| | | | | | | | 33.25553 | 0.09766 | 340.5 | <2e-16 | *** |
| Modified Greenburg | | | | | | | 163.44622 | 0.32193 | 507.7 | <2e-16 | *** |

Significance level: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' and 1

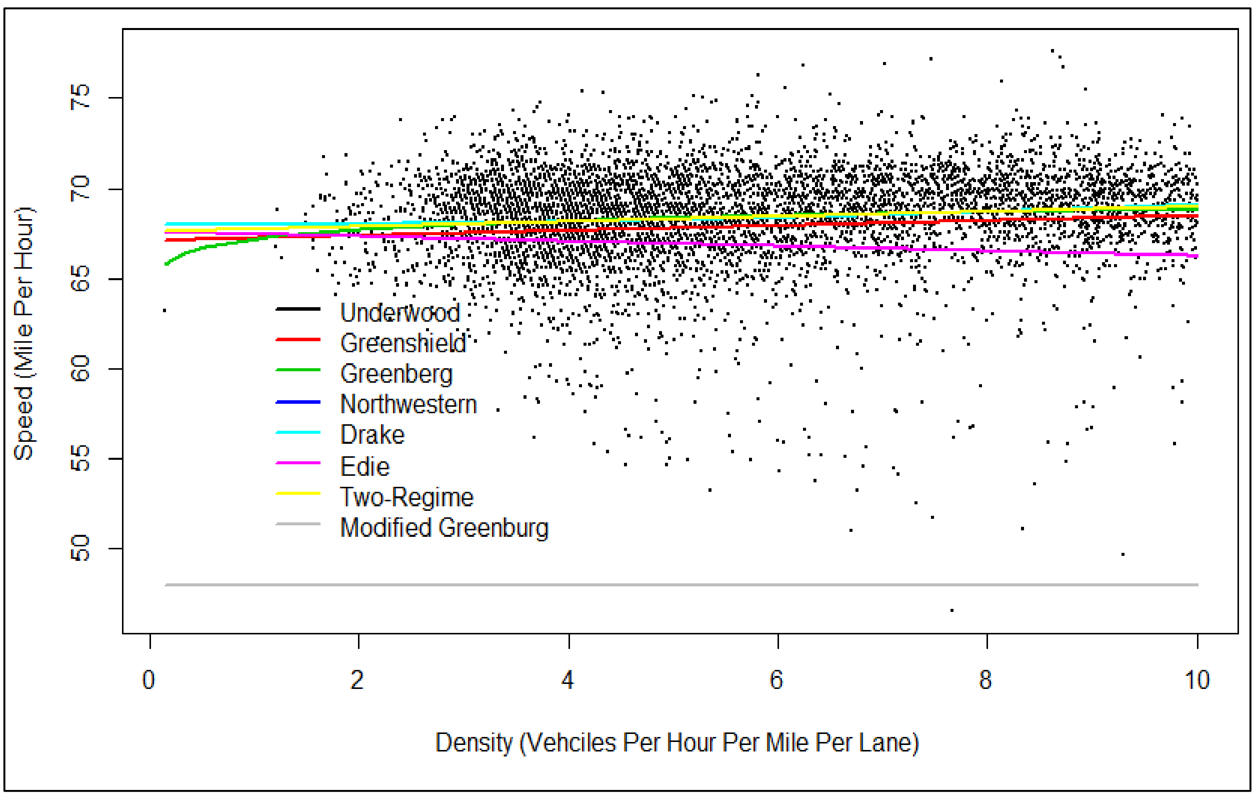


Figure 34. Free-Flow Regime Model

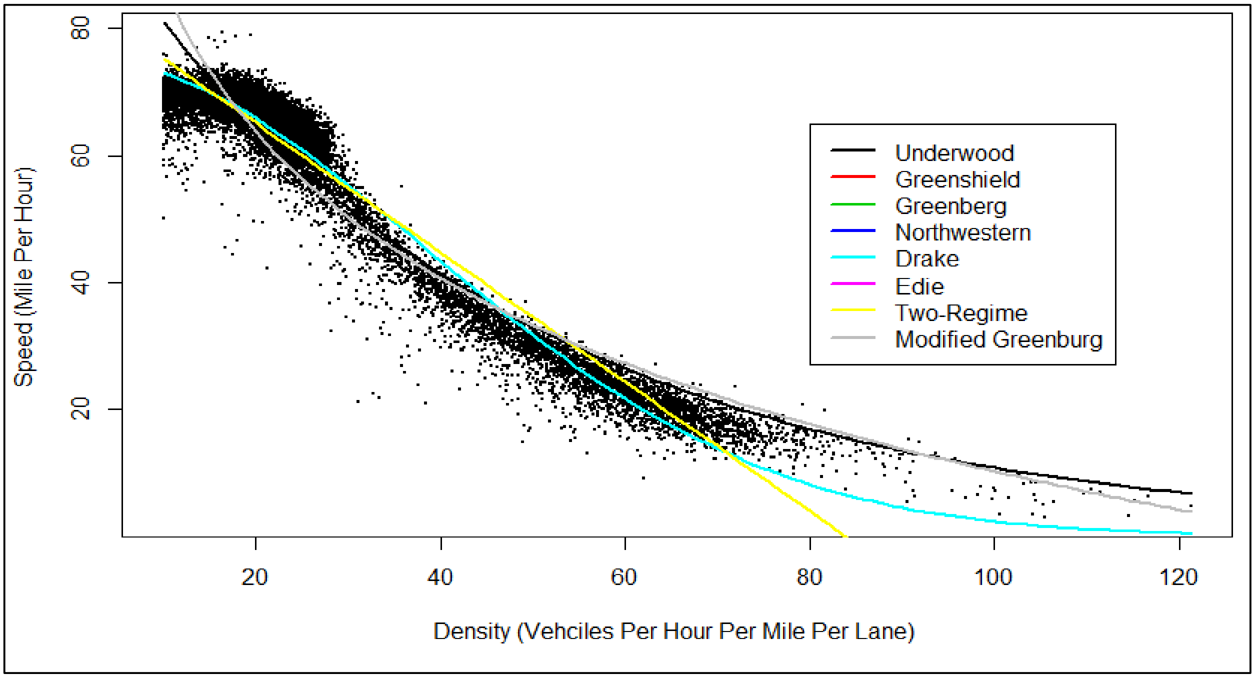


Figure 35. Congested Flow Regime

The R-square value and the RSE are included in Table 20. The Northwestern and Drake models show a maximum R-square value of 96.17 percent with the lowest RSE at 3.519 in the congested regime model. However, in the free-flow regime model, the R-square exhibit very low as 1.17 percent, which clearly indicates that the free-flow regime model is not a good fit for the aforementioned models.

Table 20. R-Square Value for the Multi-Regime Model

| | Free-flow Regime | | Congested Regime | |
|---|---|---|---|---|
| Model | R-Square | RSE | R-Square | RSE |
| Underwood | 1.17 | 2.641 | 92.18 | 5.025 |
| Greenshields | 1.17 | 2.641 | 93.96 | 4.416 |
| Greenberg | 1.08 | 2.642 | 87.71 | 6.302 |
| Northwestern | 1.17 | 2.641 | 96.17 | 3.519 |
| Drake | 1.17 | 2.641 | 96.17 | 3.519 |
| Edie | 1.17 | 2.641 | 87.71 | 6.302 |
| Two-Regime | 1.17 | 2.641 | 93.97 | 4.416 |
| Modified Greenburg | - | - | 87.71 | 6.302 |

### 6.4.4. Quantile Formulation of Free-Flow Regime Model

With a single-regime model, the model considers the data's overall fit. With a multi-regime model, different regions are considered based on the cut-off level's subjective nature. Because the multi-regime model depends on the cut-off level, the results are theoretically biased or subjective. In a broader perspective, integrating some established threshold/cut-off level, which is widely accepted or applied, such as a LOS, would leverage researchers to select the multi-regime model as the better option. The speed-density model clearly shows two distinguished characteristics: free-flow and congested. The free-flow regime is linear and is decreasing very slowly in nature, or stable; on the other hand, the congested regime may exhibit either linear or non-linear relationship, rapidly decreasing in nature. In every case, the Northwestern and Drake models perform better for both the single regime and multi-regime (congested regime part). Based on these discussions, the Northwestern model is chosen for further quantile analysis.

The quantile analyses for the single-regime and multi-regime models (free-flow and congested regime) are portrayed in Figures 36-38. The figures illustrate how the speed can be observed, depending on the traffic density and quantile functions.

The R-square value was simulated for different quantile for both model types based on the Northwestern model presented in Figure 39. The multi-regime model, especially in the congested portion, was better for predicting speed depending on the density. The multi-regime model generated a higher R-square value than single regime model. Results indicated that the Northwestern congested-regime and single-regime models were robust and accurate. However, the free-flow model was very sensitive, depending on each quantile. This sensitivity is presented in Figure 39.

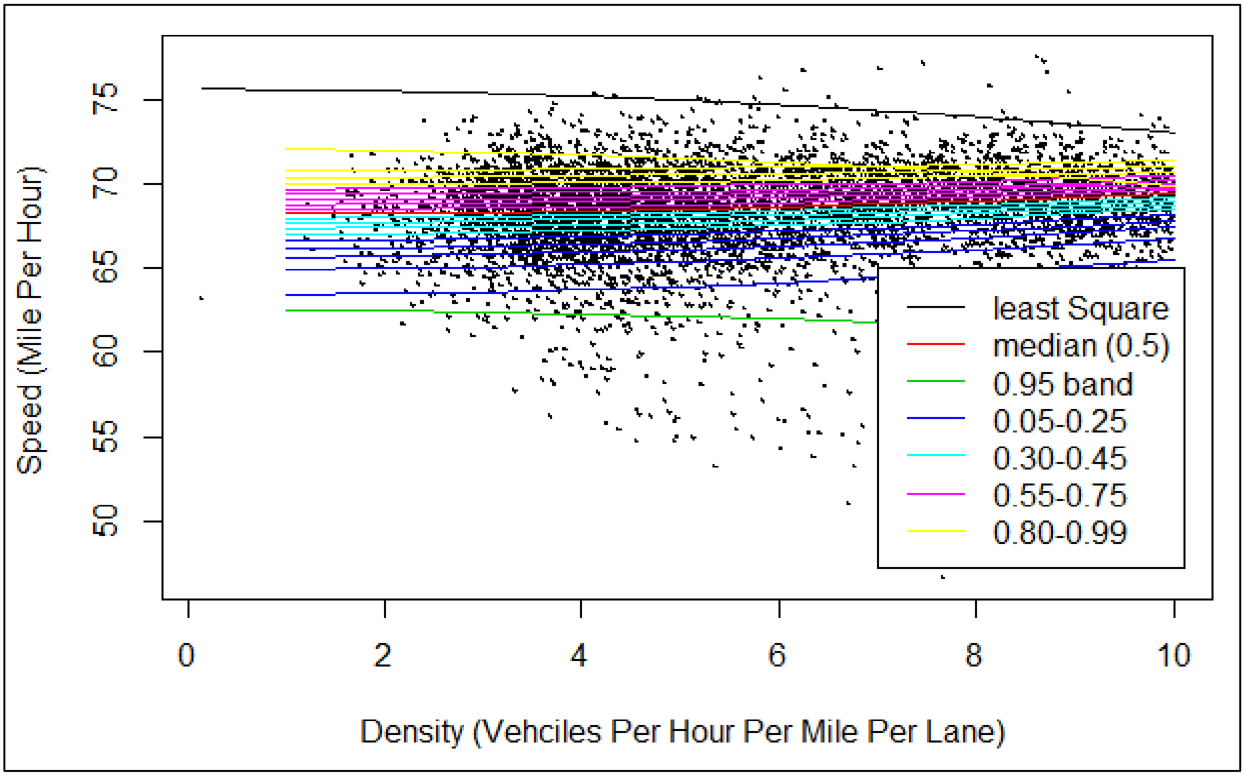


Figure 36. Free-Flow Regime Quantile

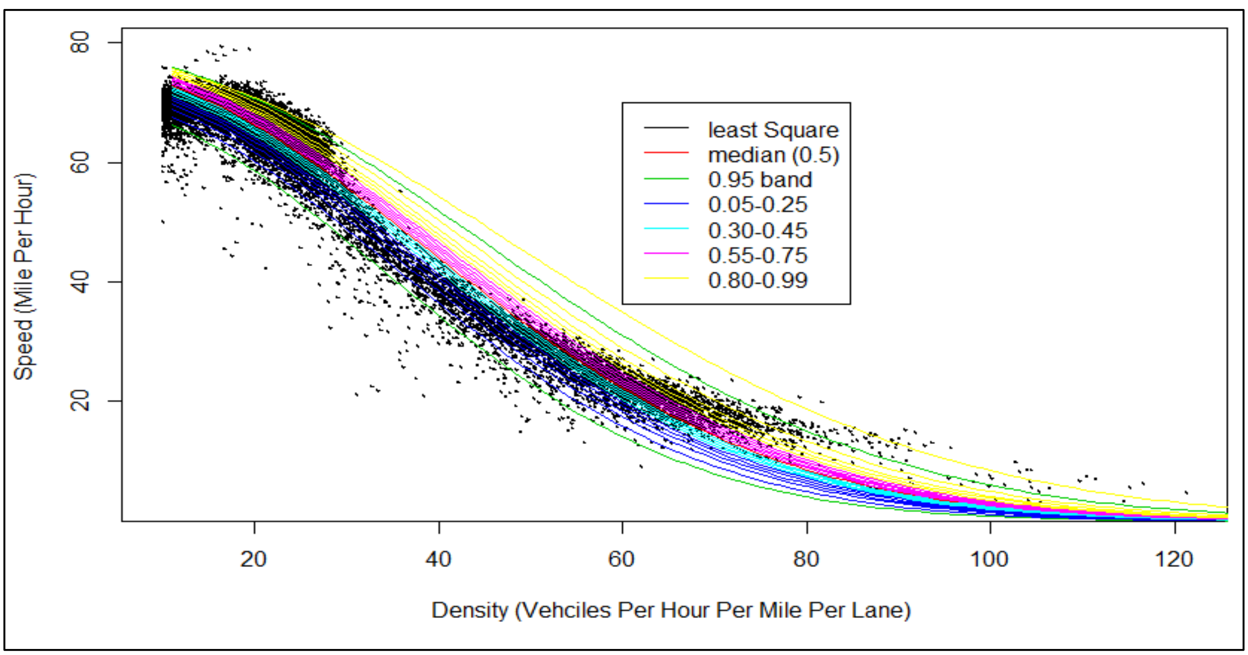


Figure 37. Congested Regime Quantile

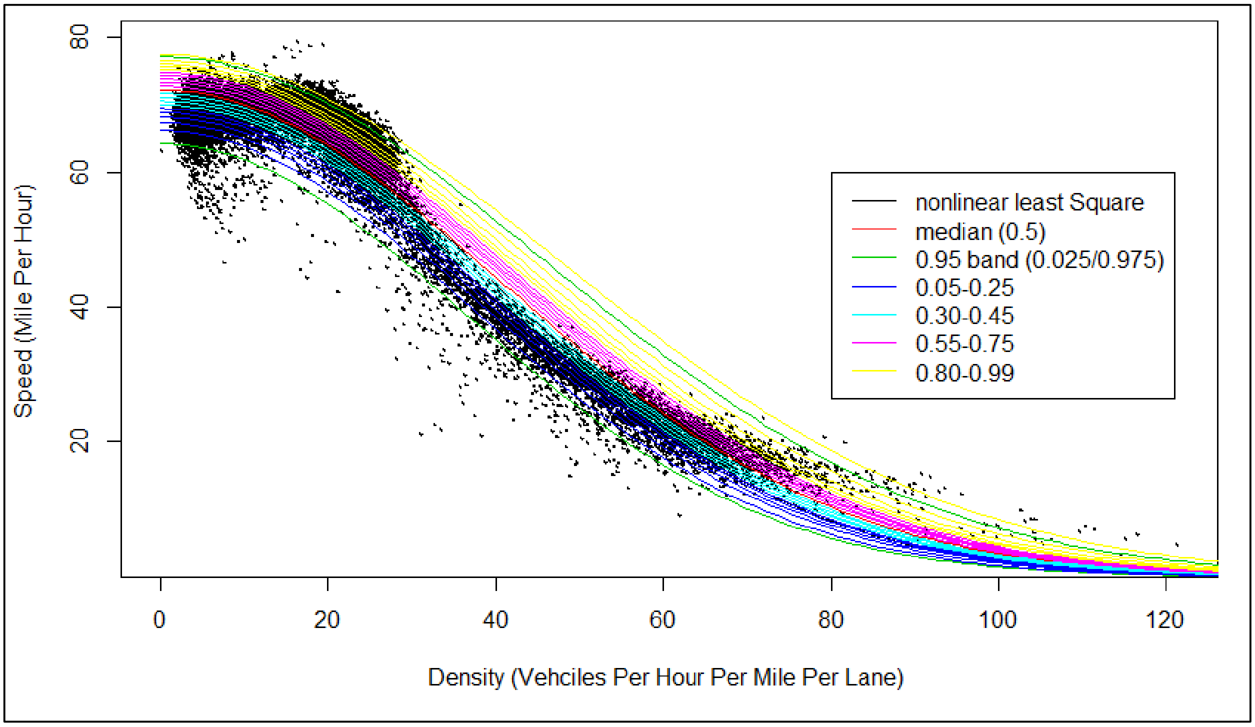


Figure 38. Single-Regime Quantile

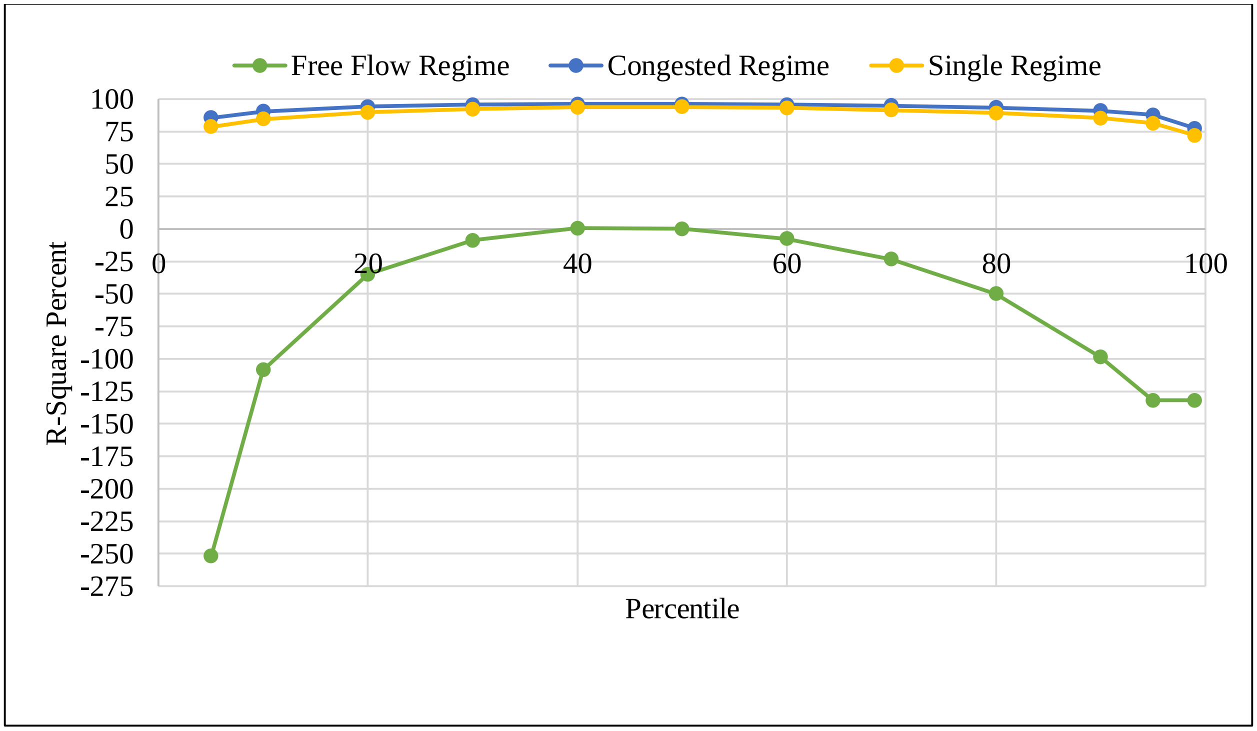


Figure 39. Simulated R-Square by Percentile

### 6.4.5. Sensitivity and Selection of Free-Flow Speed

Furthermore, the model parameters' sensitivity was investigated by each quantile (Figure 40). In the 45th percentile of the free-flow regime model, the FFS-coefficient estimation parameter (*a*) for Northwestern model was not significantly different than the least-square method. Northwestern model was also significantly different than zero.

Based on existing knowledge, it was not possible to find any suitable guidelines to choose the best quantile congested-regime model. Estimated FFS based on the single-regime and multi-regime models presented in Figure 41. It can be suggested that the FFS is increasing in nature with the conditioning percentile. For a given percentile, the single-regime model generated a higher FFS value than the multi-regime model.

Based on the previous discussion in this chapter, 45th quantile on Northwestern model can be suggested to find the FFS value. This study showed that for 45th quantile on Northwestern model, FFS can be observed 68 and 72 mph for multi-regime and single regime model respectively. FFS value of 68 mph was chosen for this study. In order to make conservative estimation of $t/t_o$ prediction, the lower value of FFS was used for this study. This value has been used for all over the analysis for this research. The question is why this research incorporated 45th percentile or why not 85th percentile because at 45th percentile, the FFS prediction is not significantly different than the NLS method. At 45th percentile, the traffic flow represents the average conditions of traffic flow. In practice, 85th percentile is being widely used by the practitioner. However, in this case at 85th percentile, the FFS prediction would be significantly different than the NLS method or average conditions. In addition to this, based on the study of Mtoi and Moses (2014), it is not certain that which percentiles' (50th, 85th, and 90th) speed would be useable to compute the FFS.

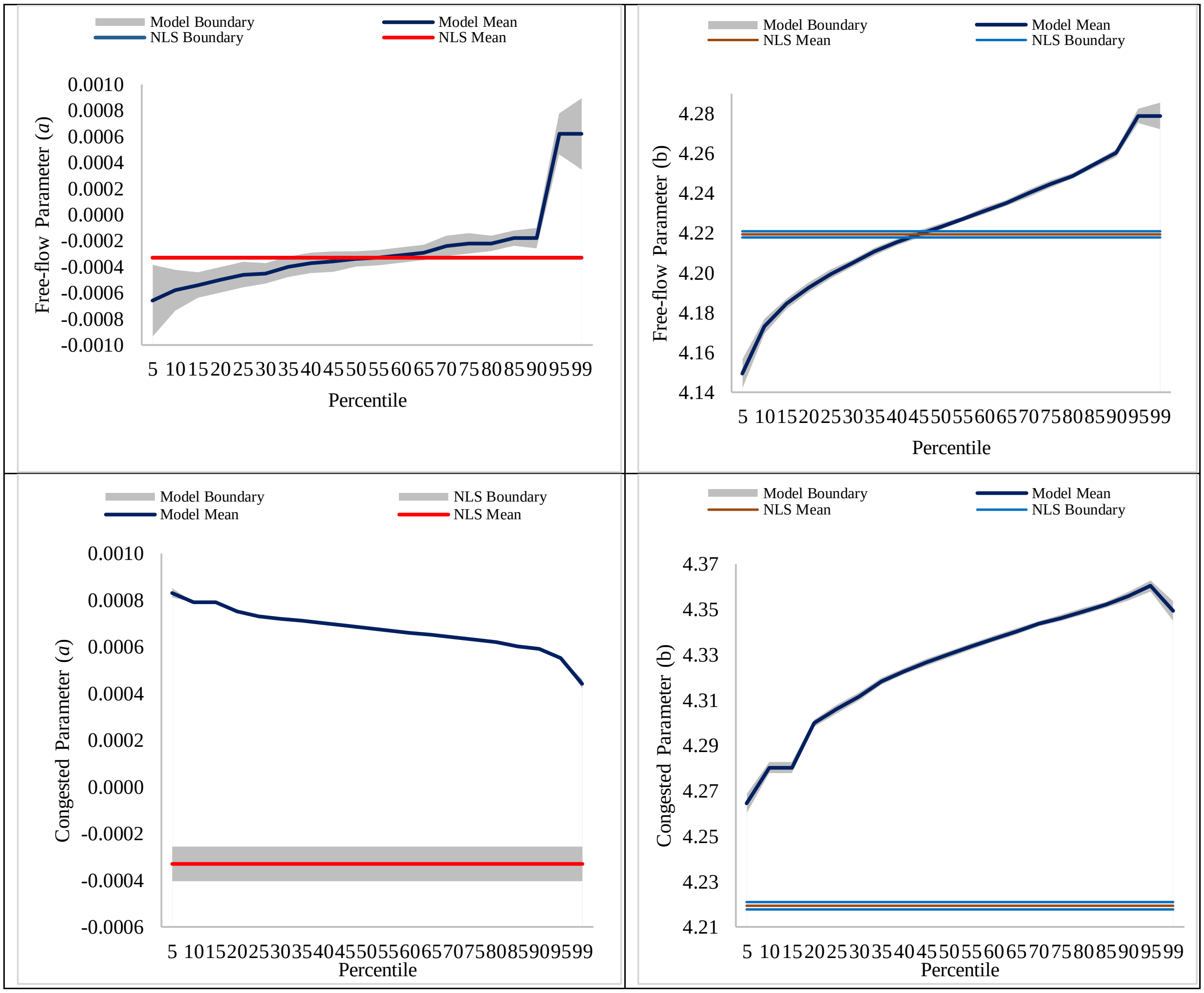


Figure 40. Multi-Regime Quantile Sensitivity

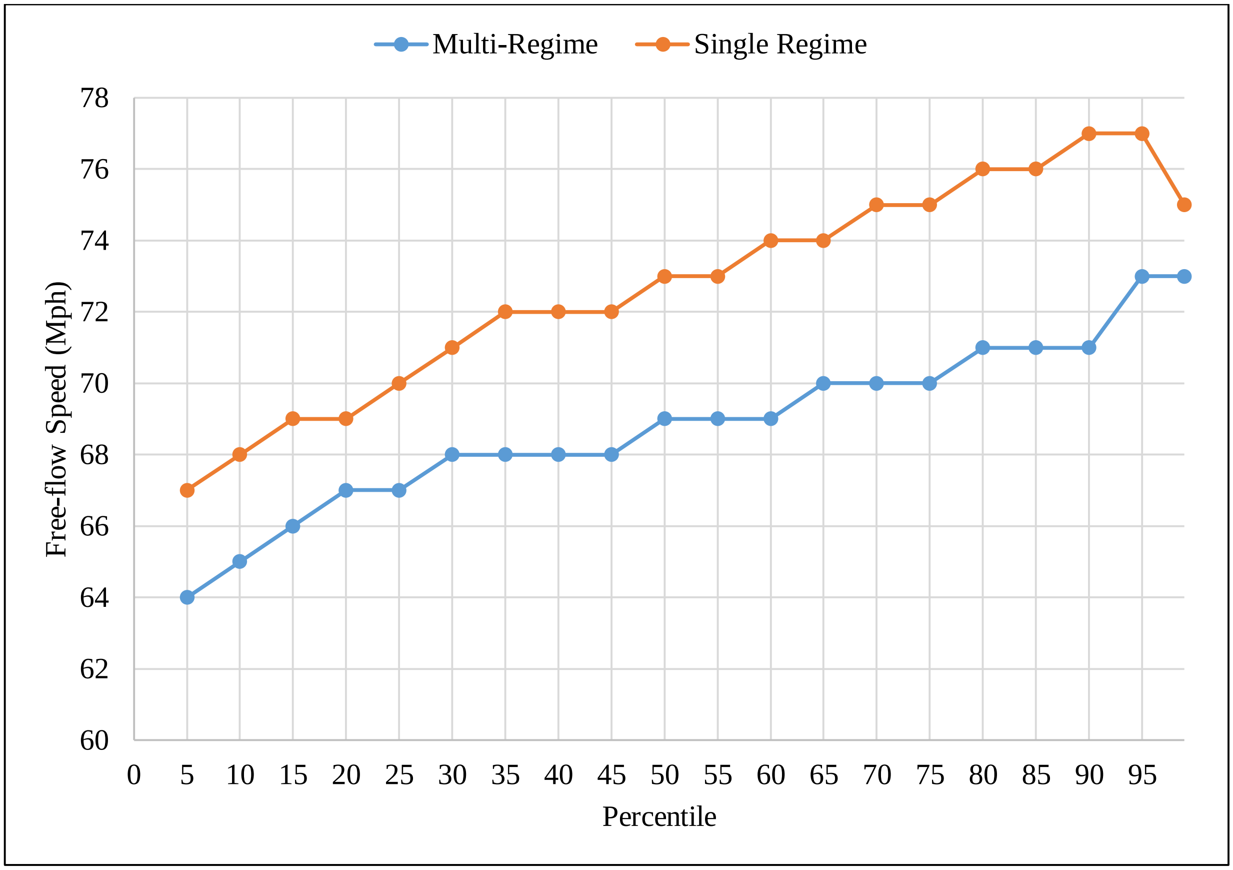


Figure 41. FFS Computation

# 7. BAYESIAN MODELING UPDATE

This chapter discusses several sections of the research:

i. Stochasticity of Traffic-Flow Characteristics

ii. Prior-Information Parameters' Estimation

iii. Likelihood and Posterior Parameters' Estimation

## 7.1. Stochasticity of Traffic-Flow Characteristics

Before explaining the stochasticity of the traffic-flow characteristics, the overall characteristics of the data was examined. The characteristics of traffic flow, speed, $t/t_o$ ratio, and $v/c$ ratio by different category are described. Later, variables correlation is presented.

### 7.1.1. Daily and Hourly Flow Characteristics

In Chapter 5, it was shown that the traffic-flow distribution, mean, and variance for each hour were significantly different than the values for other hours. Five-minute interval flows were converted to an hourly basis in order to find the capacity per hour per lane. The data for the study location are presented for the day's hours in Figure 17. This figure indicates the link's total hourly flow for different hours. Figure 17 illustrates the non-linear relationship between the flow and the day's hours. Figure 18 presents how the observations deviate from the 50$^{th}$ percentile, the median value. Furthermore, data are categorized by hour, which is presented in Figure 19. Statistical tests proved that the flow distribution, mean, and variance for each hour were significantly different than the values for other hours.

### 7.1.2. Daily and Hourly Speed Characteristics

In Chapter 6, five-minute interval speeds were analyzed to see how the speed data are distributed daily. The speed distribution indicates a non-linear relationship between the speed and the day's hours. Statistical tests proved that each hour's speed distribution, mean, and variance were significantly different than the values for other hours. From midnight to 2:00 PM, a normal day may contain many outliers which are outside the inter-quartile range. Until 6 AM in the morning, the speed's mean and variance seems to be constant, but later, these two parameters' variation is significant. The *Q-Q* plot in Figure 29 represents how the speed's distribution for the day's hours deviate from a straight line. The Figure basically shows how the distribution are skewed. Figure 31 illustrates how the observations deviate from the 50$^{th}$ percentile, the median value. The figure shows how the speed's distribution is multiplicative in nature.

### 7.1.3. Daily and Peak-Hour Characteristics

First, let examine how the overall data appear in scatter plot (Figure 42) and a *Q-Q* plot (Figure 43). The scatter plot shows the delay ratio on vertical axis and the *v/c* ratio on the horizontal axis. In normal traffic conditions, the scatter plot indicates how the delay's mean and variation would look and their trends. Neither the scatter plot nor any suitable mathematical transformations would generate the specific trends for this functions. It is not certain whether the function would be increasing and decreasing. Therefore, this uncertainty should incorporate the probabilistic method to capture these scenarios. The *Q-Q* plot in Figure 43 displays how the delay has a relationship with the *v/c* ratio. Figure 43 clearly indicates that the delay does not have a linear relationship with the *v/c* ratio. Figures 44-46 present box plots for the delay variations that occur daily, and during the AM and PM peak hours.

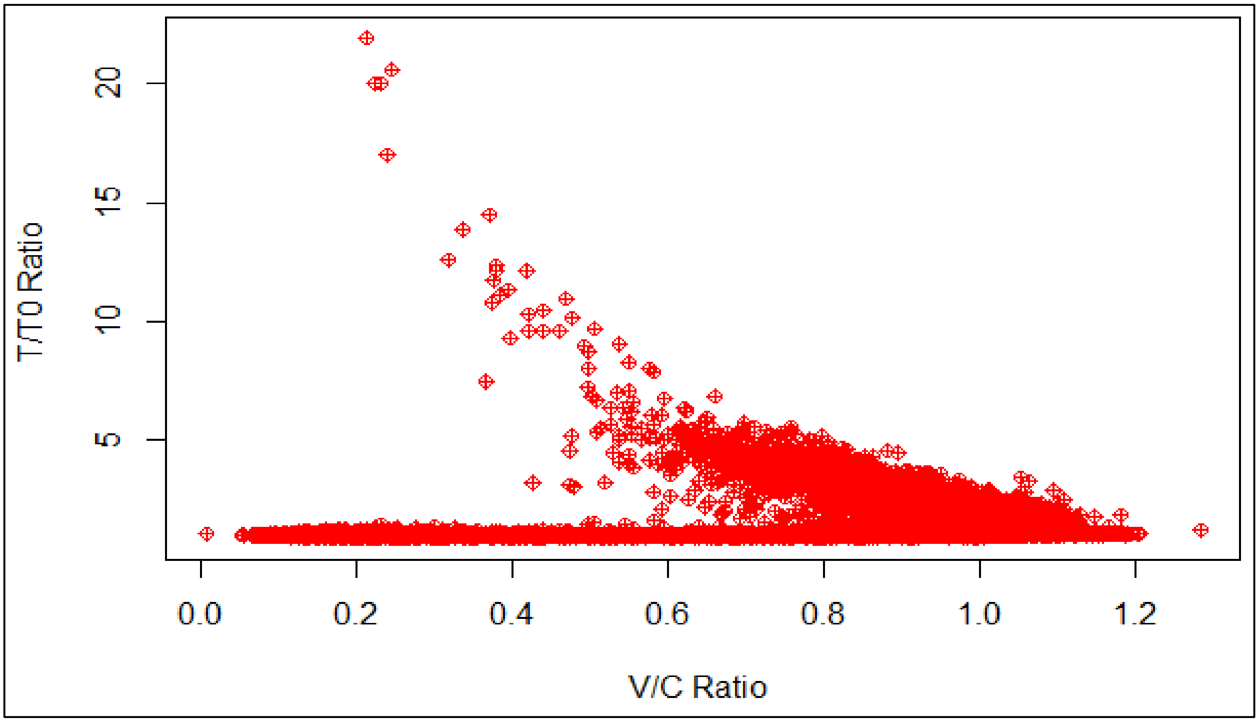


Figure 42. Scatter Plot of the Travel-Time Delay

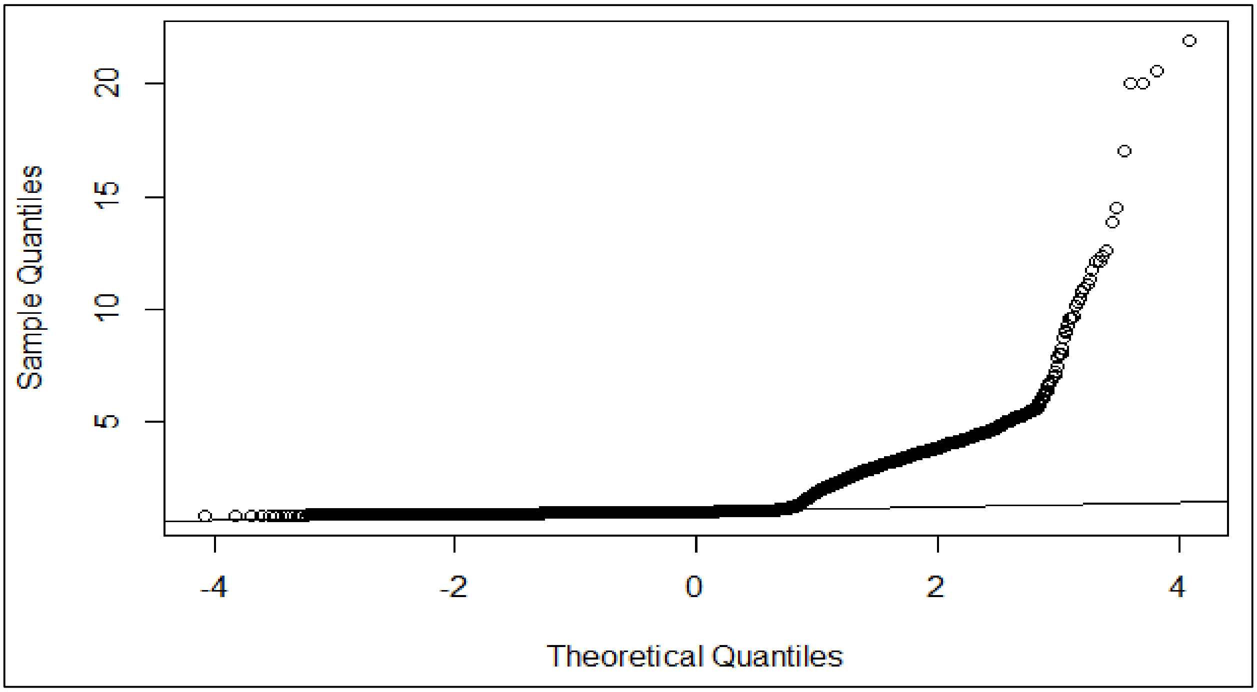


Figure 43. Q-Q Plot of the Travel-Time Delay

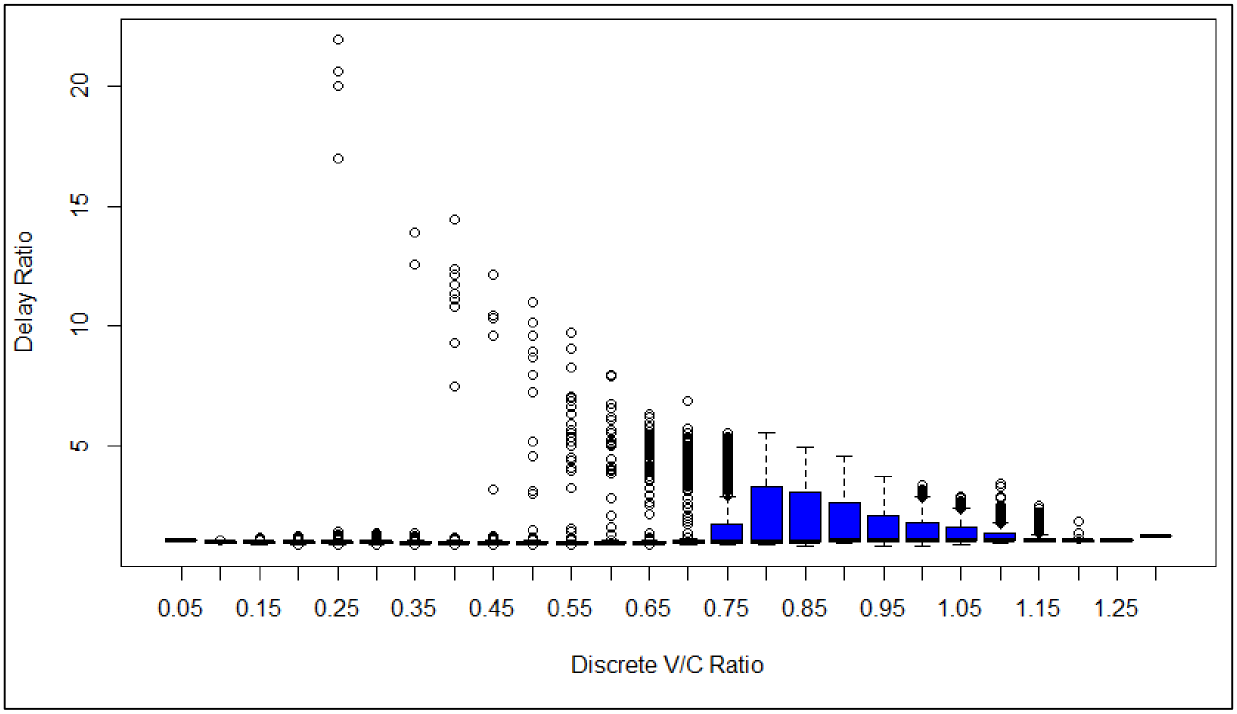


Figure 44. Box Plot of the Daily Delay Ratio

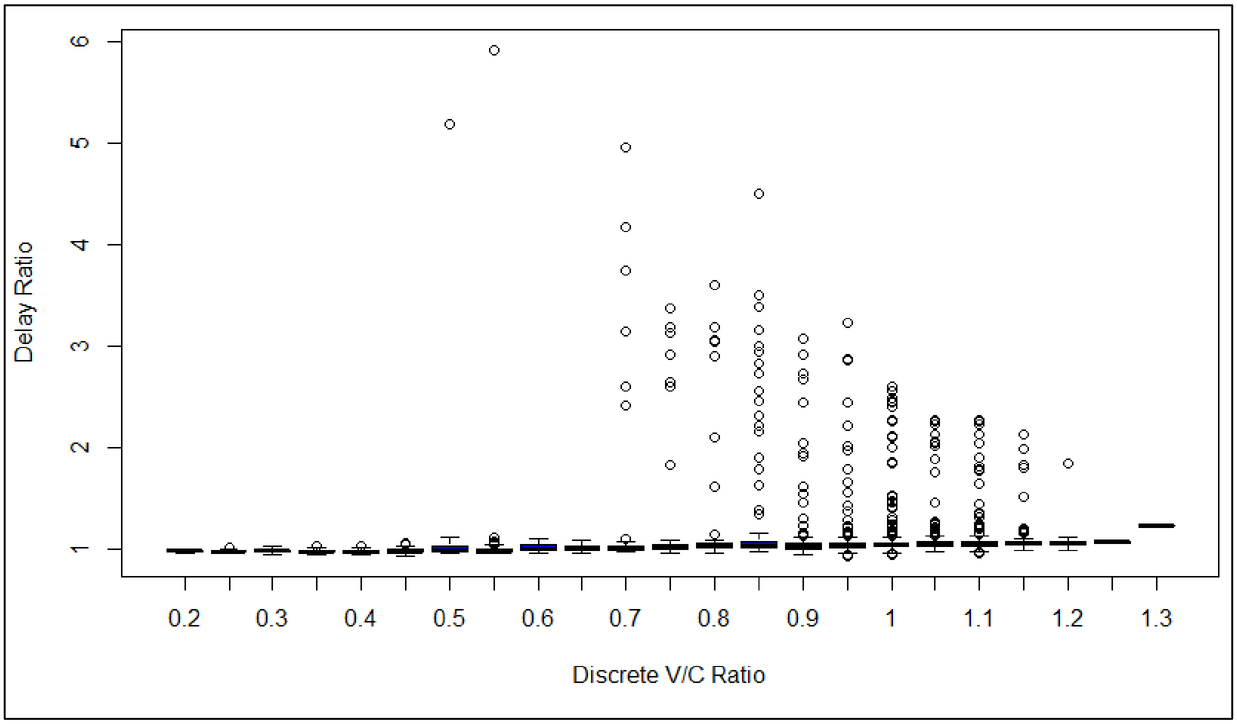


Figure 45. Box Plot of the AM Peak-Hour Delay

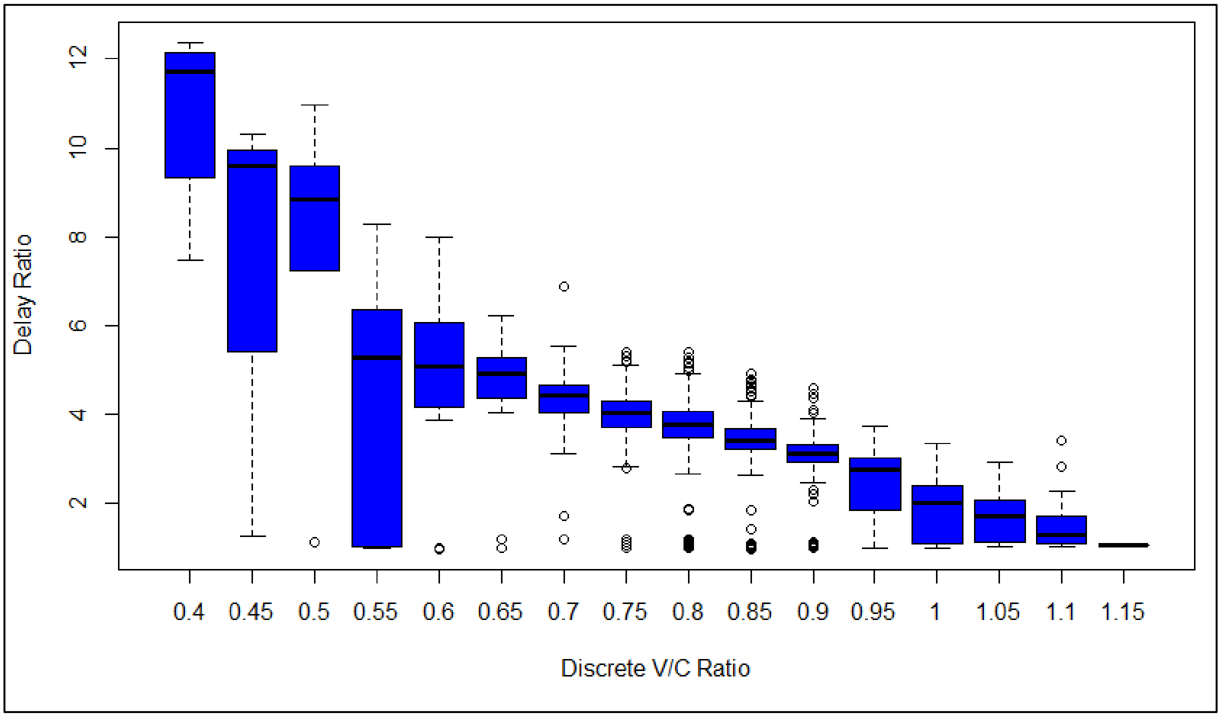


Figure 46. Box Plot of the PM Peak-Hour Delay

Overall, the statistics presented in Table 21 indicate that the minimum-observed *v/c* ratio is 0.006 while the maximum *v/c* ratio is 1.28. Within this range, the $t/t_o$ ratio is observed as 0.8564, and the maximum delay ratio is 21.93 for a normal day. The maximum delay ratio ($t/t_o$) is 21.93 indicates that the travel time might be reached 21.93 times faster than the free-flow conditions.

Table 21. Summary Statistics for the Overall Data

| Item | $t/t_o$ | *v/c* |
|---|---|---|
| Minimum | 0.8564 | 0.0060 |
| 1st Quartile | 0.9869 | 0.3890 |
| Median | 1.0287 | 0.8230 |
| Mean | 1.3746 | 0.6981 |
| 3rd Quartile | 1.1258 | 0.9590 |
| Maximum | 21.9355 | 1.2840 |

Later, the overall delay's normality was tested and reported in Table 22. Normality tests included Anderson-Darling, Cramer-von Mises, Pearson Chi-Square, and the Shapiro-Francia test for the daily as well as the AM and PM peak-hour $t/t_o$. All four methods proved that the overall time delay is not normality distributed for all three cases (daily, AM peak hour, and PM peak hour). The AM peak hour was considered to be 6:00 AM to 9:00 AM, and the PM peak hour was considered to be 4:00 PM to 7:00 PM.

Table 22. Overall Data Normality Test

| Normality Test by Overall Data | | | |
|---|---|---|---|
| | Daily | AM Peak Hour | PM Peak Hour |
| Anderson-Darling | A = 4215.9, p-value < 2.2e-16 | A = 718.87, p-value < 2.2e-16 | A = 30.999, p-value < 2.2e-16 |
| Cramer-von Mises | W = 880.24, p-value = 7.37e-10 | W = 147.6, p-value = 7.37e-10 | W = 4.4371, p-value = 7.37e-10 |
| Pearson Chi-Square | P = 183320, p-value < 2.2e-16 | P = 12541, p-value < 2.2e-16 | P = 1655.9, p-value < 2.2e-16 |
| Shapiro-Francia | Sample size too high | W = 0.2752, p-value < 2.2e-16 | W = 0.9056, p-value < 2.2e-16 |

The Pearson Chi-Square Normality test was checked for each *v/c* category and reported in Table 23. The test results showed that the daily $t/t_o$ was not normally distributed from *v/c* category 0.15 to 1.25. During the AM peak hour, data from *v/c* category 0.50 to 1.25 were not normally distributed. In the PM peak hour, data from *v/c* category 0.70 to 1.10 were not normally distributed. It might be inferred that, for a majority of the cases, the travel-time ratio was not normally distributed. Because the delay was not normally distributed, a formulation for $t/t_o$ functions with the least-square curve-fitting technique would include uncertainty.

Because the $t/t_o$ was not normally distributed in most cases, the test was investigated to find each category's exact distribution. To confirm the distribution, Cullen and Frey's graph was plot and analyzed to see the exact distribution. Figures 47-49 show that the daily delay distribution might fall between the log-normal and gamma distribution. The AM peak-hour distribution clearly showed a beta distribution, and the PM peak hour distribution was log-normal distribution. Because the beta distribution can explain all kinds of distributions with its shape parameters, considering the beta distribution would be most reasonable for this analysis.

Table 23. Pearson Chi-Square Normality Test by *v/c* Category

| Pearson Chi-Square Normality Test by *v/c* | | | |
|---|---|---|---|
| *v/c* | Daily | AM Peak Hour | PM Peak Hour |
| 0.00 | P = 0, p-value = NA | P = 0, p-value = NA | P = 0, p-value = NA |
| 0.05 | P = 1, p-value = NA | P = 0, p-value = NA | P = 0, p-value = NA |
| 0.10 | P = 6.3048, p-value = 0.789 | P = 0, p-value = NA | P = 0, p-value = NA |
| 0.15 | P = 100.53, p-value = 5.113e-11 | P = 0, p-value = NA | P = 0, p-value = NA |
| 0.20 | P = 329.97, p-value < 2.2e-16 | P = 2, p-value = 0.1573 | P = 0, p-value = NA |
| 0.25 | P = 16440, p-value < 2.2e-16 | P = 3.087, p-value = 0.6866 | P = 0, p-value = NA |
| 0.30 | P = 257.75, p-value < 2.2e-16 | P = 3.65, p-value = 0.7239 | P = 0, p-value = NA |
| 0.35 | P = 9242.2, p-value < 2.2e-16 | P = 14.557, p-value = 0.06835 | P = 0, p-value = NA |
| 0.40 | P = 12493, p-value < 2.2e-16 | P = 9.7966, p-value = 0.2796 | P = 0.6, p-value = 0.4386 |
| 0.45 | P = 11248, p-value < 2.2e-16 | P = 13.825, p-value = 0.08645 | P = 3.6667, p-value = 0.05551 |
| 0.50 | P = 8179.1, p-value < 2.2e-16 | P = 292.49, p-value < 2.2e-16 | P = 4, p-value = 0.1353 |
| 0.55 | P = 3474.6, p-value < 2.2e-16 | P = 366.33, p-value < 2.2e-16 | P = 8, p-value = 0.04601 |
| 0.60 | P = 4258.8, p-value < 2.2e-16 | P = 13.8, p-value = 0.1296 | P = 4.5556, p-value = 0.336 |
| 0.65 | P = 7369.6, p-value < 2.2e-16 | P = 18.373, p-value = 0.0186 | P = 12.571, p-value = 0.02774 |
| 0.70 | P = 9436.1, p-value < 2.2e-16 | P = 398.34, p-value < 2.2e-16 | P = 23.822, p-value = 0.008087 |
| 0.75 | P = 11427, p-value < 2.2e-16 | P = 692.49, p-value < 2.2e-16 | P = 58, p-value = 1.194e-07 |
| 0.80 | P = 11137, p-value < 2.2e-16 | P = 252.59, p-value < 2.2e-16 | P = 116.81, p-value < 2.2e-16 |
| 0.85 | P = 9897.1, p-value < 2.2e-16 | P = 937.09, p-value < 2.2e-16 | P = 210.94, p-value < 2.2e-16 |
| 0.90 | P = 11298, p-value < 2.2e-16 | P = 1011.9, p-value < 2.2e-16 | P = 222.14, p-value < 2.2e-16 |
| 0.95 | P = 12736, p-value < 2.2e-16 | P = 1463, p-value < 2.2e-16 | P = 308.98, p-value < 2.2e-16 |
| 1.00 | P = 11006, p-value < 2.2e-16 | P = 2004.7, p-value < 2.2e-16 | P = 434.7, p-value < 2.2e-16 |
| 1.05 | P = 6205.8, p-value < 2.2e-16 | P = 1111.2, p-value < 2.2e-16 | P = 434.7, p-value < 2.2e-16 |
| 1.10 | P = 2247.8, p-value < 2.2e-16 | P = 872.29, p-value < 2.2e-16 | P = 31.391, p-value = 5.265e-05 |
| 1.15 | P = 566.85, p-value < 2.2e-16 | P = 306.67, p-value < 2.2e-16 | P = 1, p-value = 0.3173 |
| 1.20 | P = 93.273, p-value < 2.2e-16 | P = 82, p-value = 1.379e-15 | P = 0, p-value = NA |
| 1.25 | P = 1, p-value < 2.2e-16 | P = 1, p-value < 2.2e-16 | P = 0, p-value = NA |
| 1.30 | P = 1, p-value = NA | P = 1, p-value = NA | P = 0, p-value = NA |

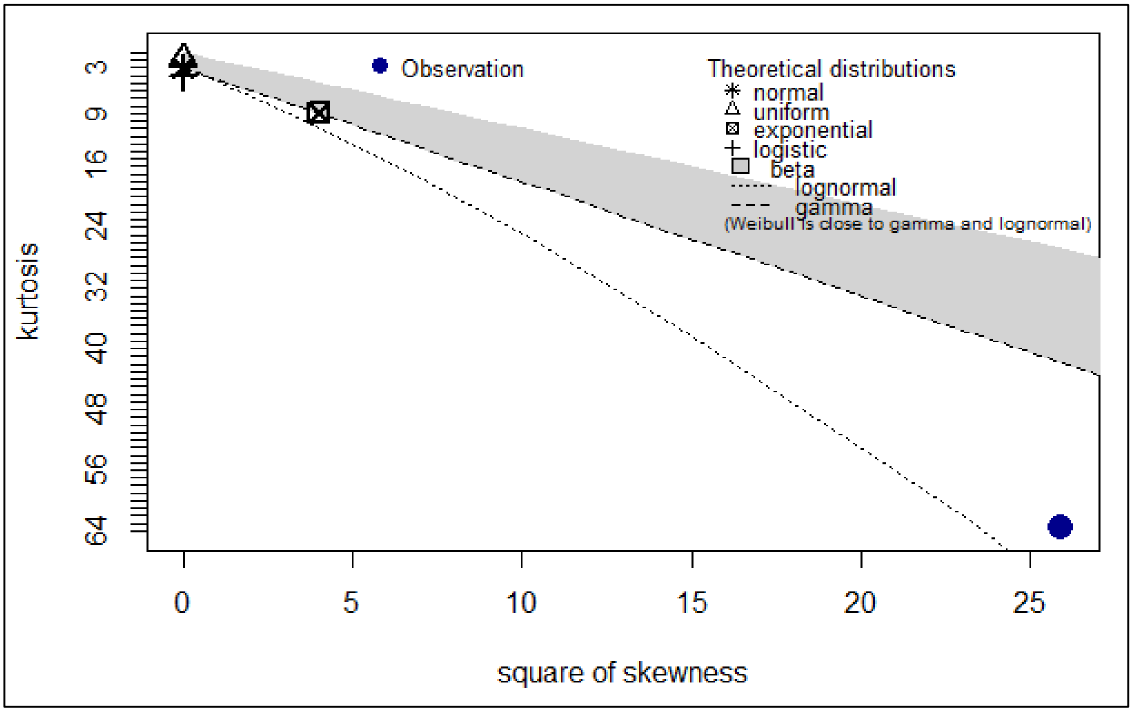


Figure 47. Cullen and Frey's Diagram of Daily Data

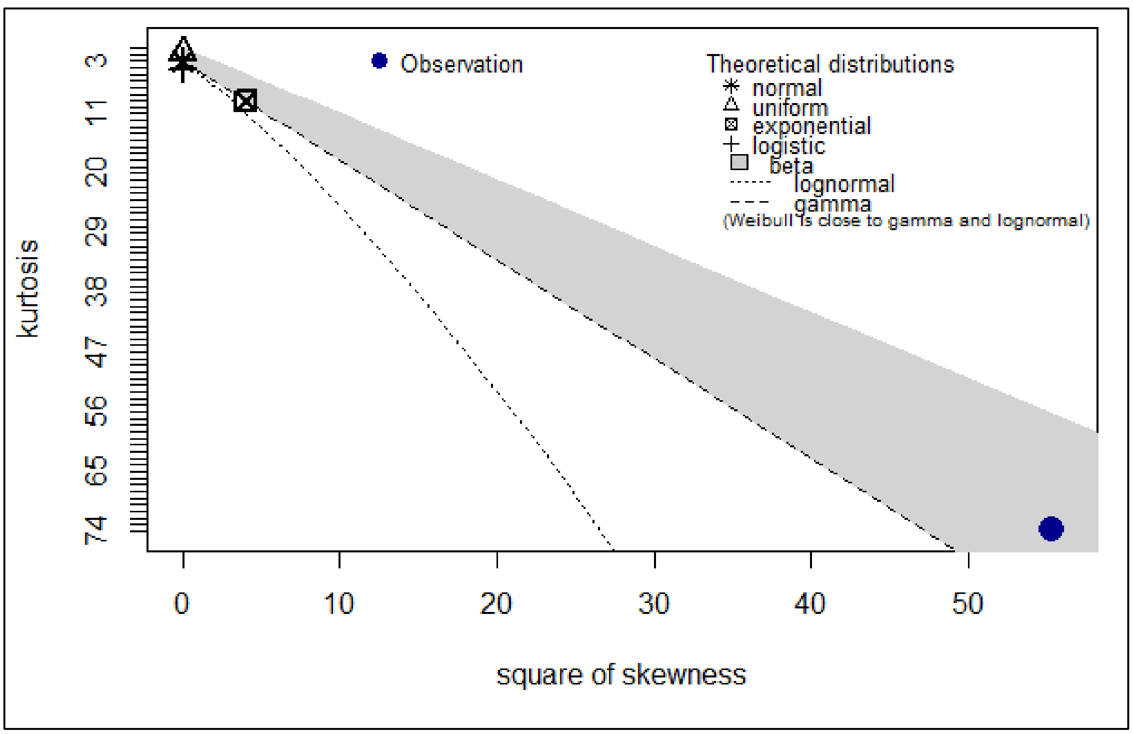


Figure 48. Cullen and Frey's Diagram of the AM Peak-Hour Data

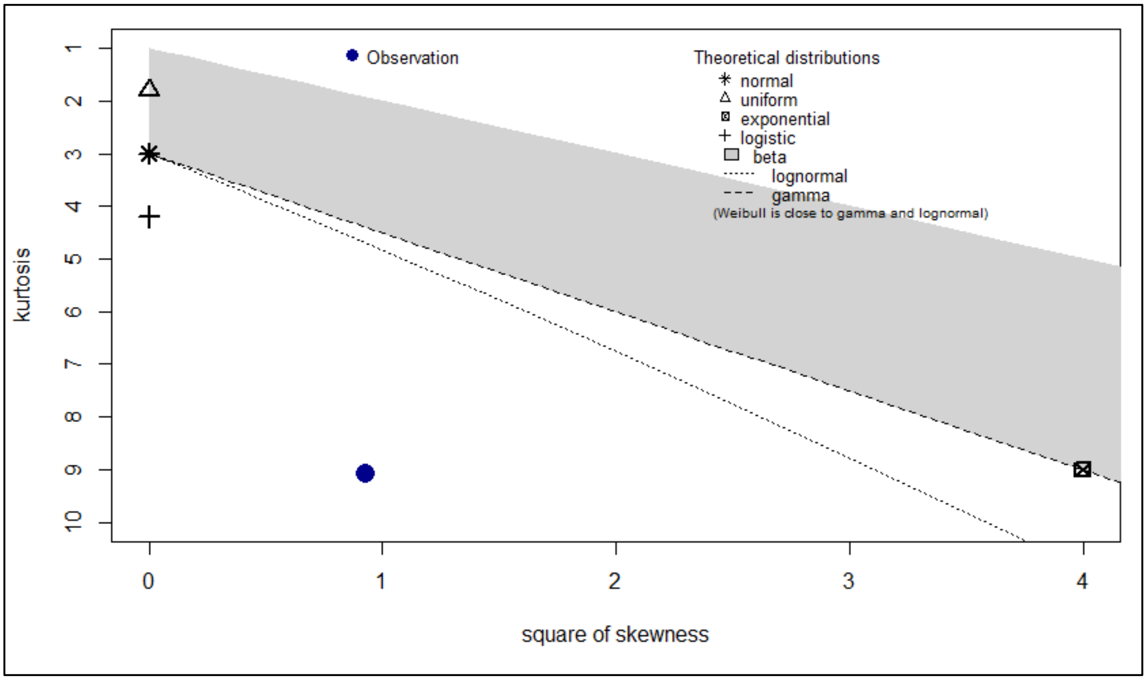


Figure 49. Cullen and Frey's Diagram of the PM Peak-Hour Data

### 7.1.4. Stochasticity

The stochastic process deals with uncertainty using a probabilistic model. This uncertainty can originate from the developed model. Any statistical model has two parts: 1) structural portion and 2) random-error portion, which can be further designated with Equation 64.

$$Predictive\ Mode = Structural\ Portion + Random\ Errors \qquad \text{(Equation 64)}$$

Random errors might be influenced by several factors, such as driving behavior, road conditions, traffic conditions, area type, weather conditions, and any major events. A fitted model would generate this random error. For better predictions, the random errors must be normally distributed, and must be independent and identically distributed. Violating these assumptions will create a stochastic process. In this regard, several existing, popular volume-delay models (BPR and Spiess) were utilized by using the least-square method with a minimizing residual-sum square error; further random errors were tested to see whether the random error violated the normality assumption for the overall data.

The normality test was performed using the Anderson-Darling, Craver-von Mises, and Pearson Chi-Square method for the proposed method, BPR method, and Spiess functions for the overall data. In every case (Table 24), the residual errors were not normally distributed, violating the assumption of normality. It can be inferred that the generated errors were uncertain with these three methods and that uncertainty needed to be incorporated into the modeling formulations. Normality test results are presented in Figures 50-52.

Table 24. Normality Test by Model

| Testing Method | Proposed | BPR | Spiess |
|---|---|---|---|
| Anderson-Darling | A = 1685.4, p-value < 2.2e-16 | A = 3205.2, p-value < 2.2e-16 | A = 1075.4, p-value < 2.2e-16 |
| Cramer-von Mises | W = 291.76, p-value = 7.37e-10 | W = 644.31, p-value = 7.37e-10 | W = 183.72, p-value = 7.37e-10 |
| Pearson Chi-Square | P = 39405, p-value < 2.2e-16 | P = 49950, p-value < 2.2e-16 | P = 19700, p-value < 2.2e-16 |

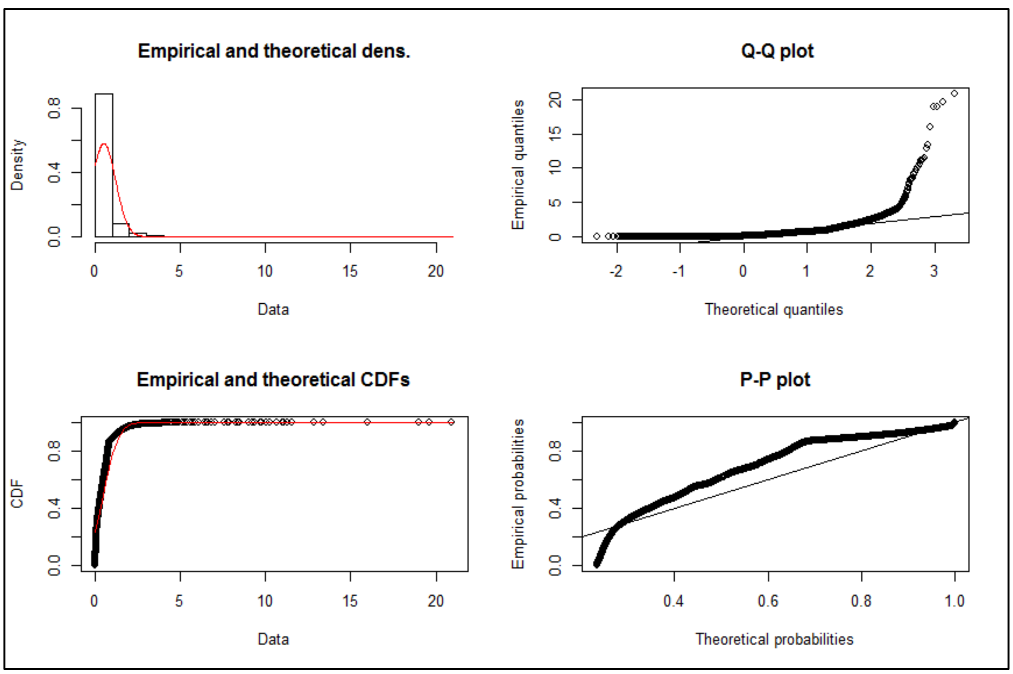


Figure 50. Proposed Method's Diagnostic Results

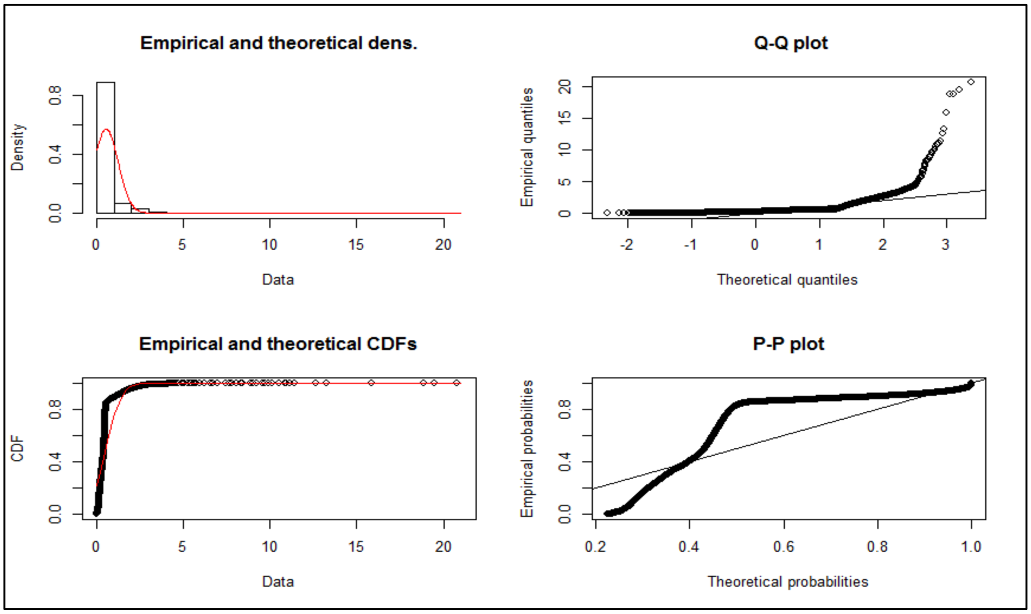


Figure 51. Modified BPR Method's Diagnostic Results

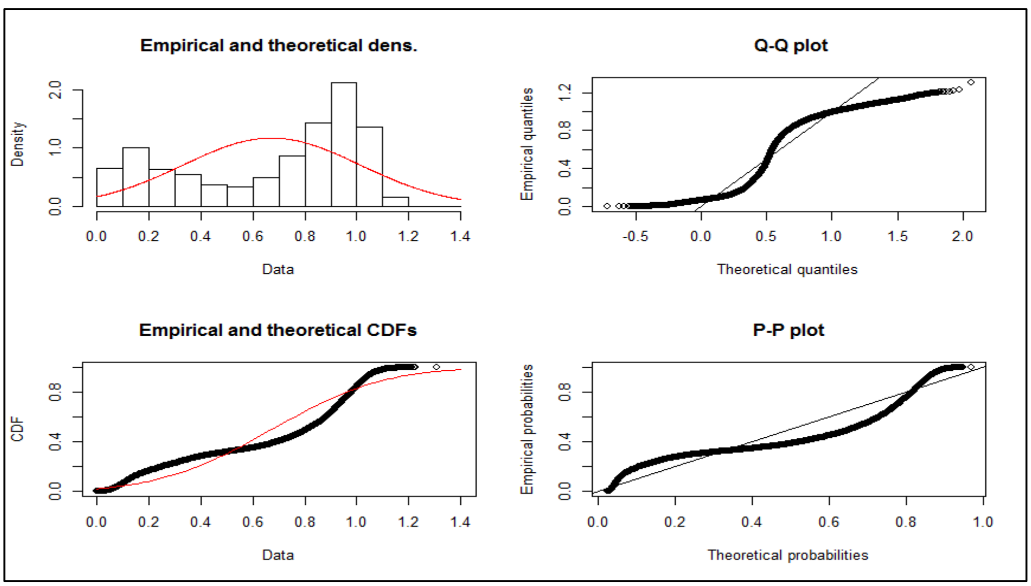


Figure 52. Spiess Method's Diagnostic Results

### 7.1.5. Correlation and Dependency of Variables

The overall Pearson correlation coefficient was observed 19.92 percent between the $t/t_o$ ratio and the *v/c* ratio. The Pearson correlation test with a *p-value* < 2.2e-16 suggested that, at a 95 percent confidence interval, these two variables were not correlated. The Pearson coefficient supported that these two variables were not linearly related. Because the coefficient was very low, the linear relationship would be very weak.

The Pearson correlation coefficient for each *v/c* group category was evaluated (Table 25). The results showed that, in a *v/c* group, $t/t_o$ was weakly, linearly related. This coefficient did not represent any strong, positive correlation. Therefore, it might be inferred that a non-linear relationship exists.

Table 25. Pearson Correlation Coefficient

| Pearson Correlation Coefficient | | | |
|---|---|---|---|
| *v/c* | Overall | AM Peak Hour | PM Peak Hour |
| 0.00 | NA | NA | NA |
| 0.05 | NA | NA | NA |
| 0.10 | -0.0092 | NA | NA |
| 0.15 | -0.0560 | NA | NA |
| 0.20 | -0.0501 | -0.0799 | NA |
| 0.25 | 0.0235 | 0.1201 | NA |
| 0.30 | -0.0445 | 0.0484 | NA |
| 0.35 | 0.0037 | -0.0310 | NA |
| 0.40 | 0.0340 | -0.4279 | 0.1354 |
| 0.45 | 0.0225 | 0.0916 | -0.5611 |
| 0.50 | 0.0599 | -0.0008 | 0.2586 |
| 0.55 | 0.0682 | 0.1475 | 0.3817 |
| 0.60 | -0.0049 | 0.0639 | -0.2749 |
| 0.65 | 0.0326 | -0.1033 | -0.0846 |
| 0.70 | 0.0152 | 0.0303 | -0.1397 |
| 0.75 | 0.0692 | 0.0477 | -0.1330 |
| 0.80 | -0.0347 | 0.0451 | -0.1827 |
| 0.85 | -0.0257 | -0.1061 | -0.1364 |
| 0.90 | -0.0772 | -0.0310 | -0.1657 |
| 0.95 | -0.0253 | -0.0228 | -0.2186 |
| 1.00 | -0.0509 | -0.0169 | -0.1690 |
| 1.05 | -0.0566 | 0.0905 | 0.0066 |
| 1.10 | -0.1512 | -0.1074 | -0.1199 |
| 1.15 | -0.1690 | 0.0032 | -0.6244 |
| 1.20 | 0.1391 | 0.2043 | NA |

## 7.2. Prior-Information Parameter Estimation

Prior information about the $t/t_o$ parameters for a given *v/c* ratio needed to be learned before the Bayesian model update or experiment was conducted. Using historical data for the completed year 2013, Bayesian prior was estimated. The PERT method was applied to estimate the prior information. It was expected that PERT would have the capability to estimate most likely mean travel time for a given *v/c* based on the historical optimistic, pessimistic, and most likely mean $t/t_o$. In this study, the maximum $t/t_o$ was considered as optimistic; the minimum $t/t_o$ was considered to be the pessimistic $t/t_o$ parameters' estimation.

Uncertainity about the prior information was estimated in a sequential and systematic way. As a probabilistic method, PERT can captured the uncertainity with the prior information. The Bayesian modeling technique required the distribution of prior information. The results indicated that the PDF for a given, random $t/t_o$ conditioning *v/c* are beta distribution. This study reported the cumulative density function (CDF) of the $t/t_o$ for a given *v/c* ratio as presented in Figure 53.

Results showed that, for a *v/c* from zero to 0.40 and from 1 to 1.25, both shape parameters are greater than 1, which represents a uni-modal distribution. For a *v/c* from 0.45 to 0.95, parameter $\alpha$ was less than 1, and parameter $\beta$ was greater than or equal to 1, indicating a reversed J-shaped distribution. The only exceptions were observed while the *v/c* was 0.85 and 1.10, where both beta-distribution parameters were equal, representing a systematic uni-modal distribution.

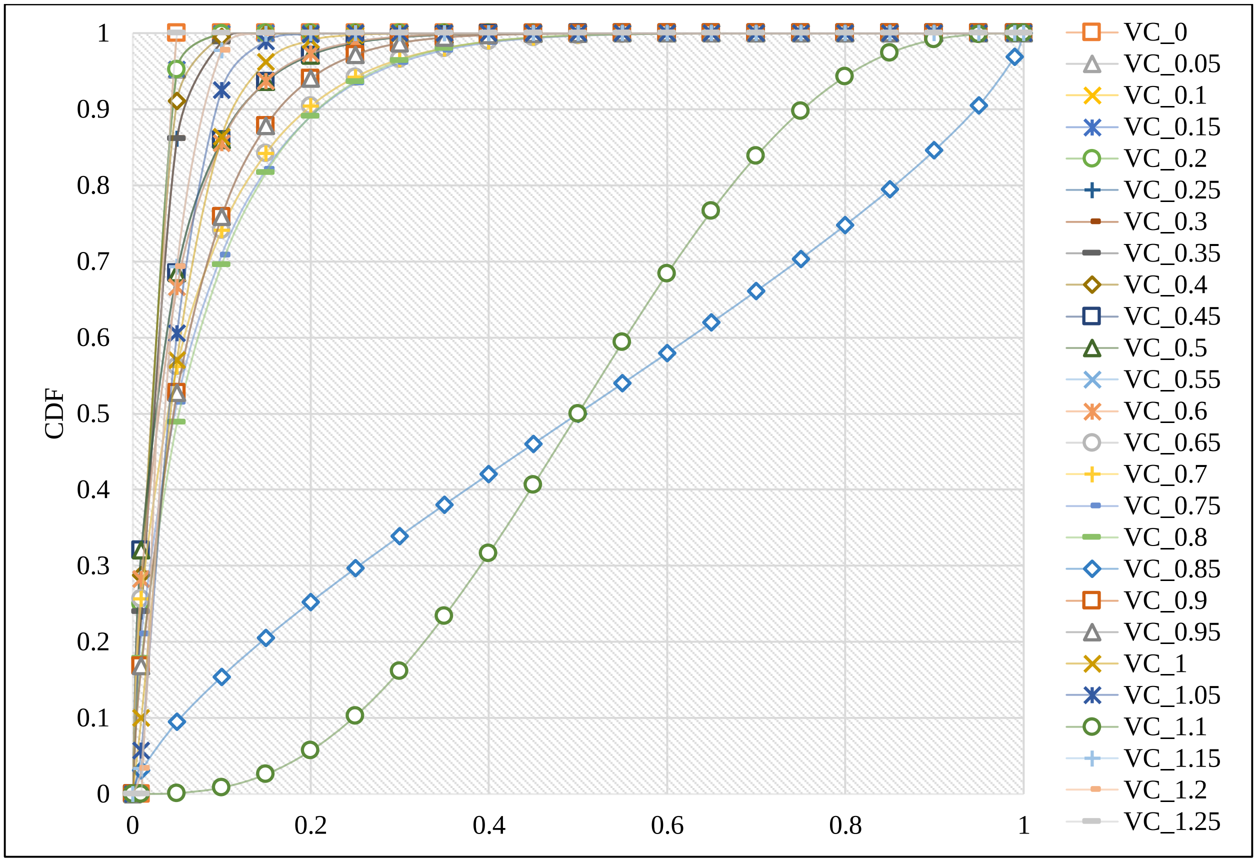


Figure 53. Prior-Information Distribution

Using the above distribution and the proposed methodology, the expected mean and variance for the $t/t_o$, conditioning $v/c$ were estimated. The results are presented in Figure 54. The PERT method's mean was significantly different than observed mean of $t/t_o$ with a p-value of 0.006196878. The PERT mean is not different than the beta mean's theoretical distribution. The PERT method captured the most likelihood of the delay ratio. The observed mean lacks to replicate this modal peaks (Figure 54). Most likely, each $v/c$'s delay is higher than observed in most cases.

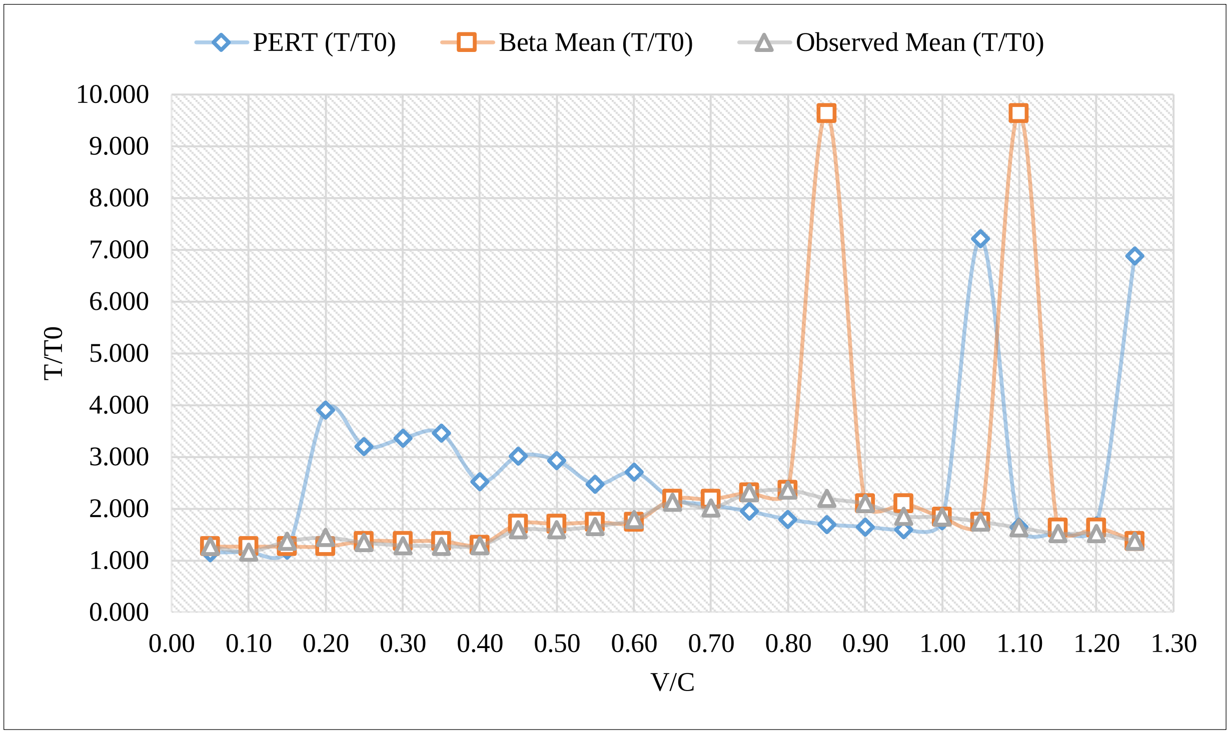


Figure 54. Prior-Information Characteristic's Curve

Prior information suggests that the expected, most likely $t/t_o$ decreases in nature until *v/c* is 1 (See Figure 54). At *v/c* of 1.05 and 1.25, the maximum expected, most likely $t/t_o$ was observed. The simple statistical mean showed the opposite characteristic, which is increasing in nature until the *v/c* is 0.80. Later, the trend was decreasing in nature. The $t/t_o$ cannot always be increasing in nature which does not support a congested place such as Los Angeles. When the most likely travel-time delay matters, the assumption of the always-increasing characteristic for the $t/t_o$ functions should be outperformed and needs to be revised.

**7.3. Likelihood and Posterior Parameters' Estimations**

The likelihood and posterior distribution, as well as its parameters about the $t/t_o$ for a given *v/c* ratio, was approximated using the method described in Chapter 3. The historical data for the complete 2014 year were considered for this purpose. For each *v/c* ratio, prior information was considered as a uniform beta distribution for the likelihood and posterior estimation.

The likelihood function (likelihood) for the $t/t_o$ conditioning *v/c* is the function of the model's parameters with given data. The likelihood is proportional to the Bayesian prior. The Bayesian posterior is proportional to the prior * likelihood function. Therefore, the assumption about the prior distribution leads to the posterior distribution's outcomes. The following Figure 55 with *v/c* at 1.0 would exemplify for further.

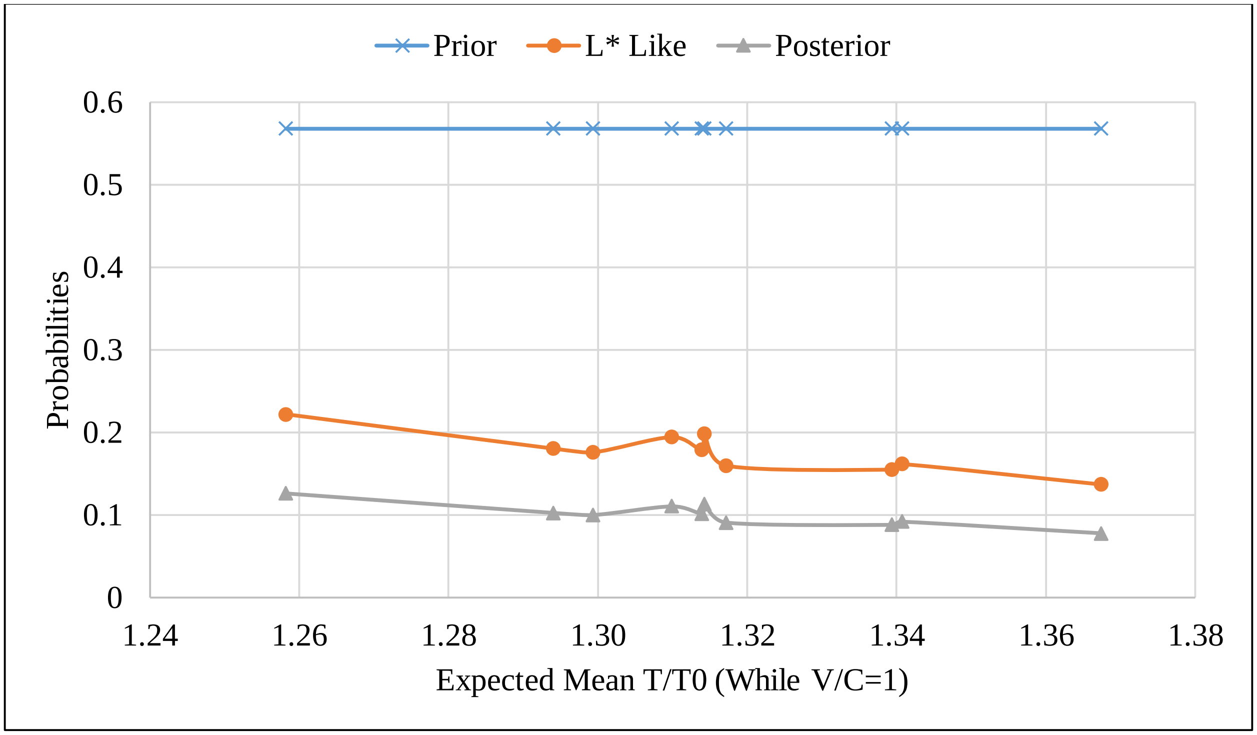


Figure 55. Bayesian Inference when *v/c=1*

Figure 55 represents the expected, mean $t/t_o$ for different models on the horizontal axis and model mean parameters' probabilities on the vertical axis. In Figure 55, solid line with cross represents the prior distribution; a solid circle represents equivalent likelihood, and a solid triangle represents the posterior probabilities. While *v/c* at 1, the prior probabilities of a given, expected mean ($t/t_o$) are always higher than the likelihood and posterior probabilities. The posterior probabilities are higher than the likelihood probabilities. Because the prior distribution was assumed to be a uniform beta distribution, the likelihood and posterior distribution follow the same distribution.

Figure 55 illustrates that the maximum equivalent likelihood ($L^*$) for a mean delay ratio of 1.26 is 0.222; the prior probability of 0.5677 for this mean increases, by a factor of about 0.222, to about 0.12. A similar analysis has been done for all other *v/c* ratio.

Furthermore, using the proposed method with the given *v/c* and posterior distributions observed above, the expected mean was computed. Figure 56 portrays this estimation. The top portion of Figure 56 includes the observed points. The bottom portion of Figure 56 does not show the observation.

It might be inferred that the Bayesian posterior trends and most likelihood trends followed a similar pattern which opposed the delay's mean observed trends with respect to the *v/c* ratio. The $t/t_o$ functions might not be always increasing in nature. With a p-value of 0.9877, the likelihood and posterior estimates are not significantly different. With a p-value of 0.015 at the 95 percent confidence interval, the posterior distribution is statistically different than the most frequently observed mean.

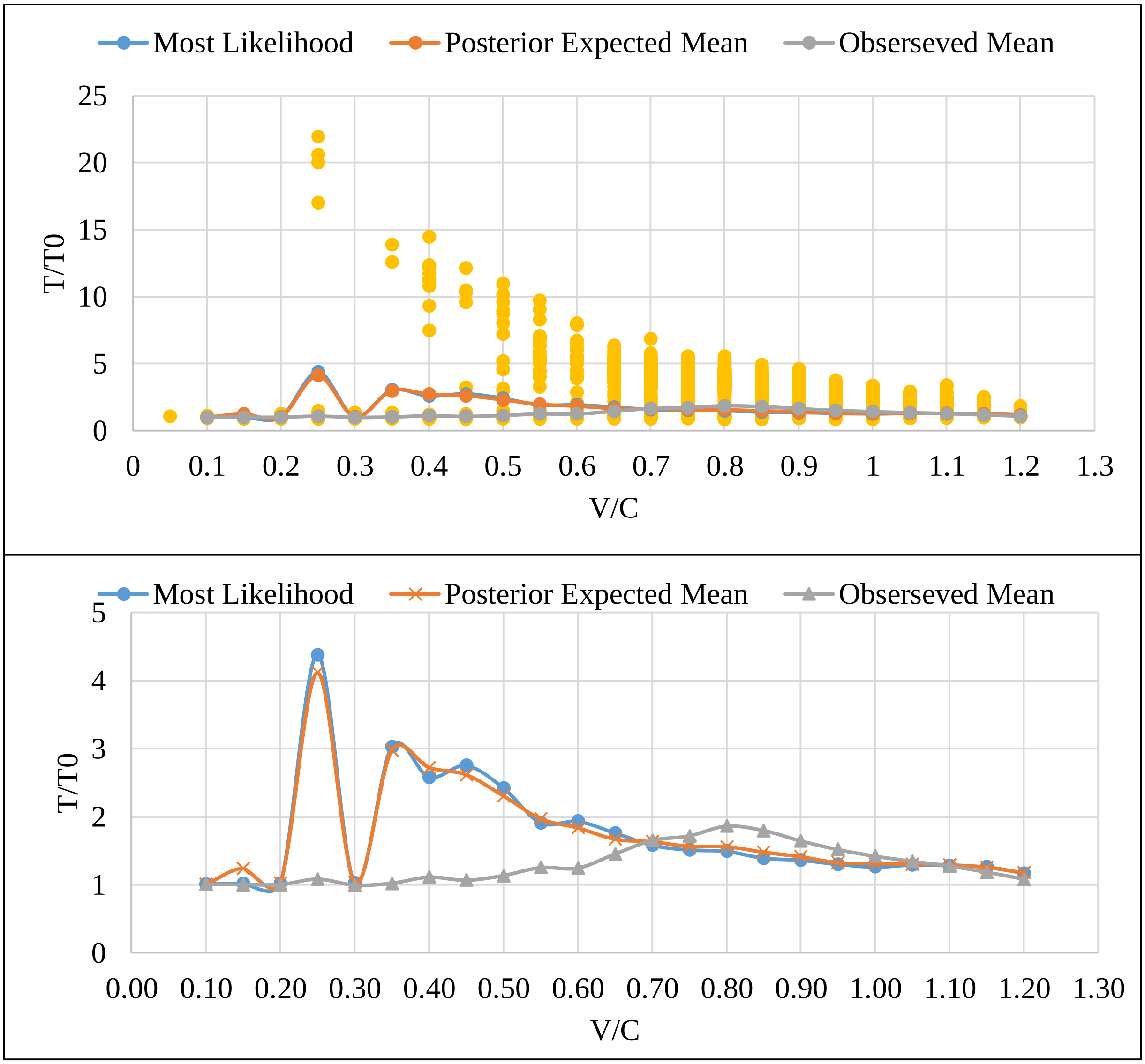


Figure 56. Predicated Trends of Different Approximation

# 8. LOGISTIC GROWTH MODELING

This chapter describes the proposed logistic growth model, explains its three parameters, simulates the parameters' sensitivity, and validates the model. The primary objective of this dissertation are discussed in this chapter.

## 8.1. Logistic *S*-Curve Formation

The logistic *S*-curve was formed from the Bayesian, predicted, expected, cumulative mean $t/t_o$ for each given $v/c$ ratio and the most frequently observed, cumulative mean $t/t_o$ for each given $v/c$ ratio. Figure 57 shows variation for the growth of the cumulative delay functions. In addition, Figure 58 shows the cumulative percentage of growth.

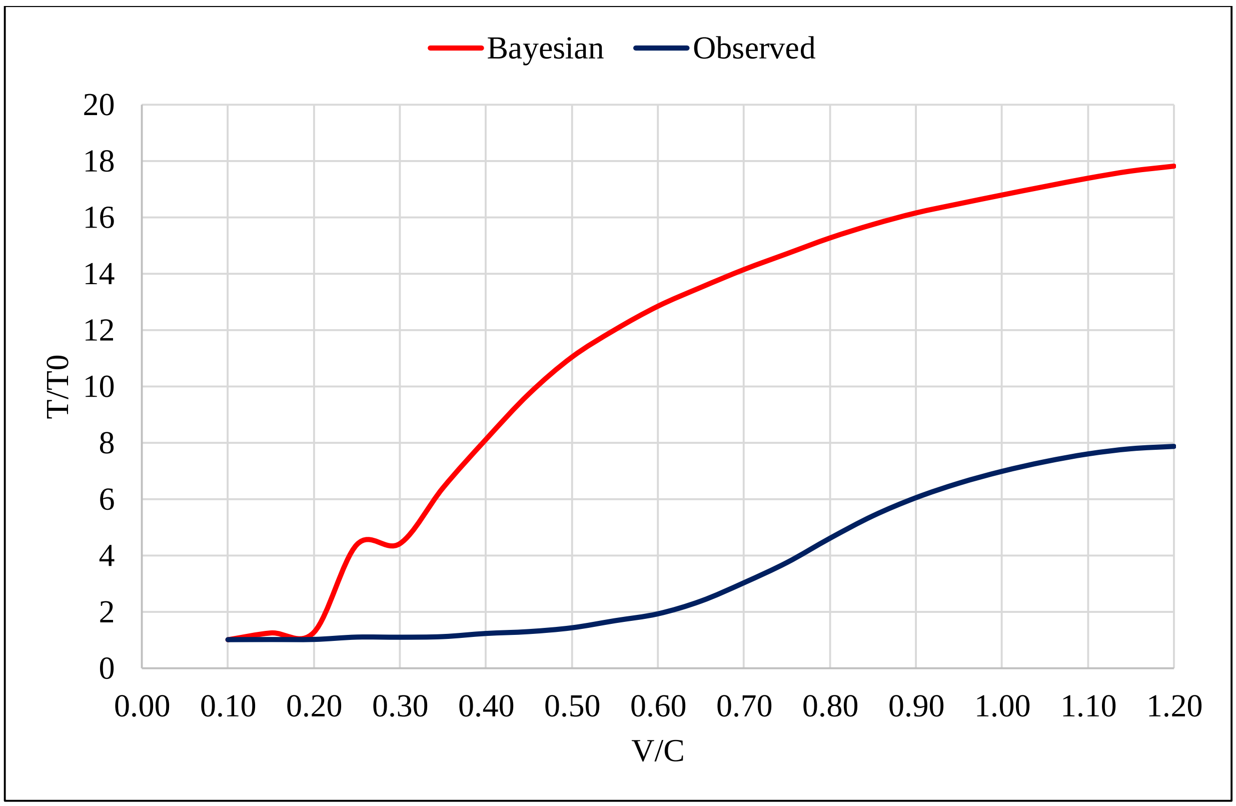


Figure 57. Logistic *S*-Curve Formation

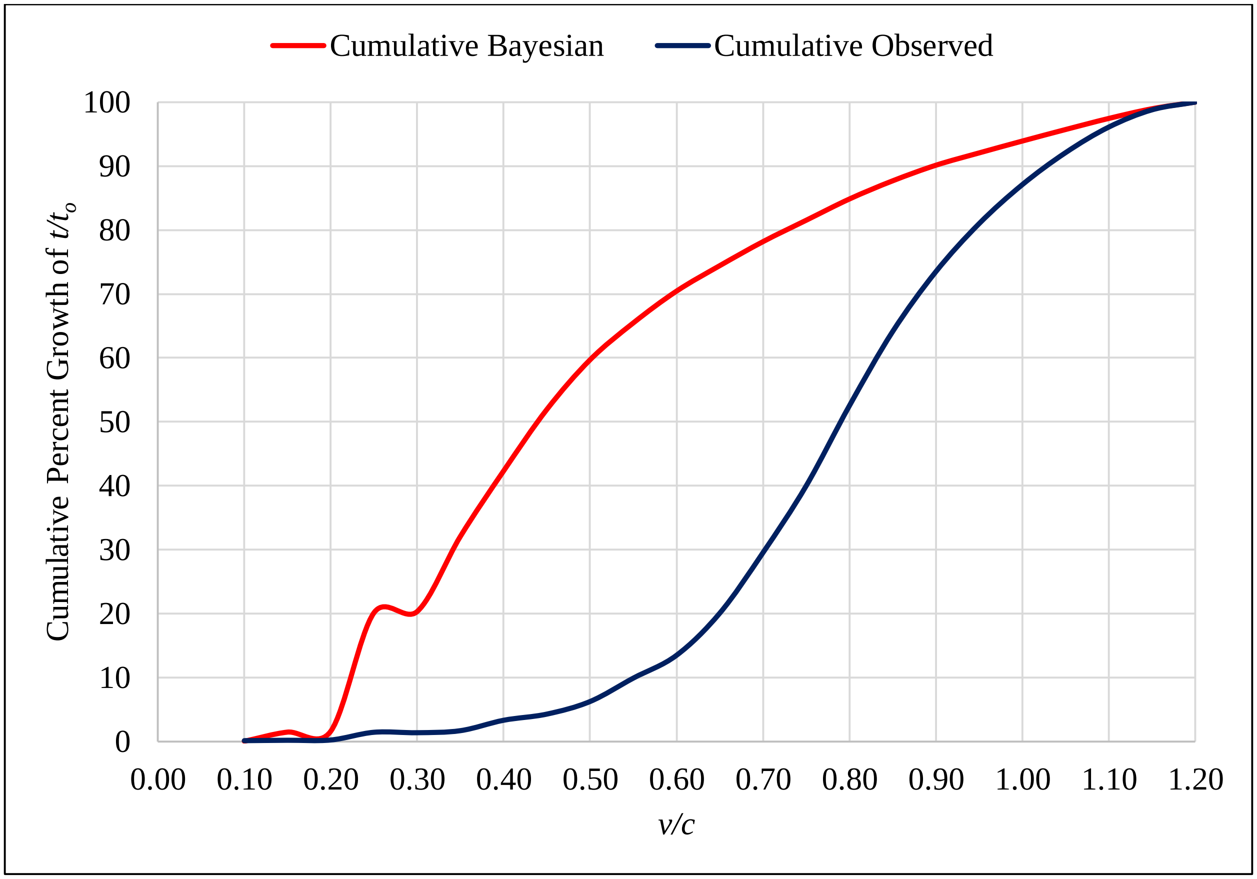


Figure 58. Cumulative Percentage of Growth for the Travel-Time Delay

Figures 58 and 59 show that, when using Bayesian predicted data, the cumulative growth is higher at the beginning then reduces slowly. The growth rate is faster and steep at first, and then slows down. Using the Bayesian prediction model, the maximum growth of 18.53 percent happens while the *v/c* ratio is between 0.20 and 0.25. In contrast, using the most-frequent method, the maximum growth of 12.53 percent happens while the *v/c* ratio is between 0.75 and 0.80. The growth-pattern trends after *v/c= 0.80* are similar for both the Bayesian and simple statistical mean methods. The most frequent method indicates that the growth rate is increasing until the *v/c* ratio reaches 0.80, and then, the growth rate decreases. The Bayesian methods indicate that the growth rate is heavily fluctuating until the *v/c* ratio is 0.35; immediately after the *v/c* ratio is 0.35, the growth rate decreases.

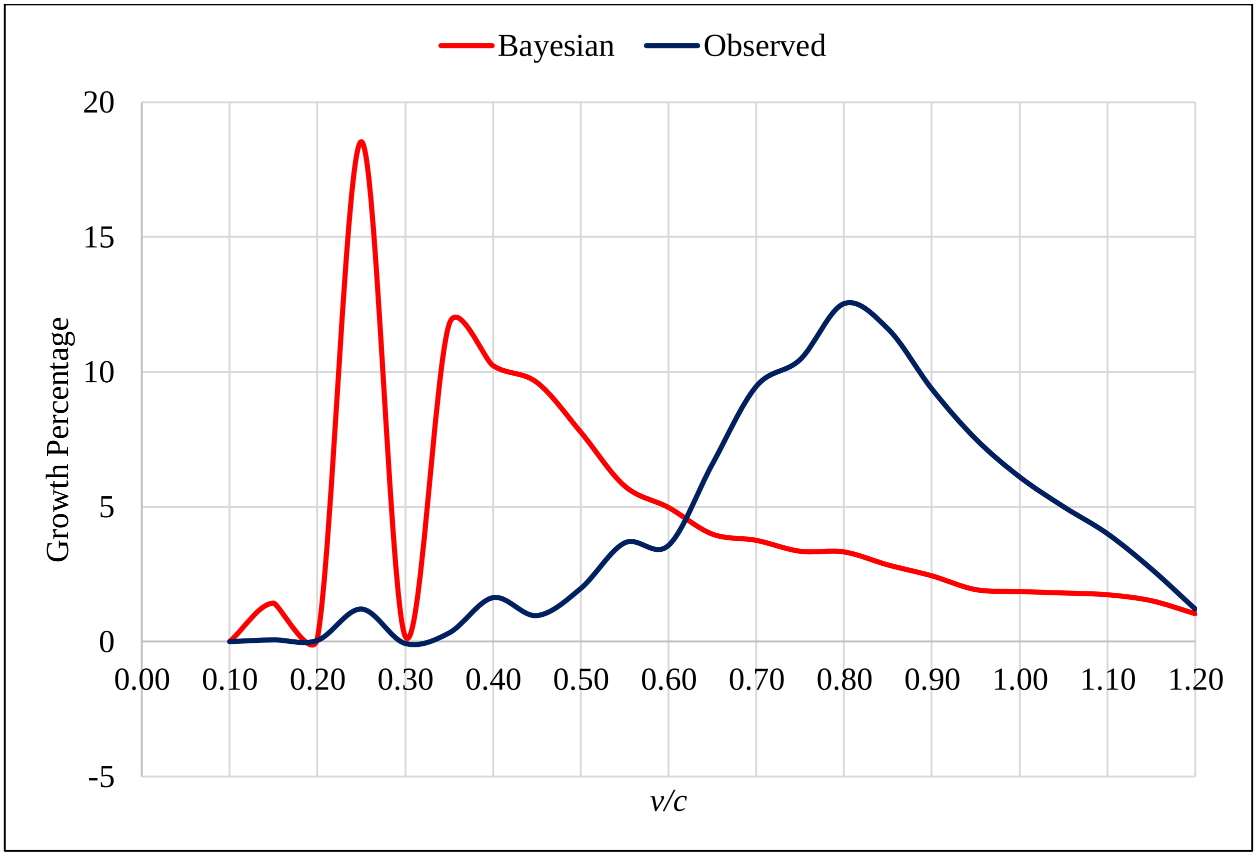


Figure 59. Change for the Growth Percentage in the S-Curve

**8.2. Logistic Model Parameters' Estimation**

To fit this type of system, a cumulative, logistic growth modeling that used the proposed method, as described in Chapter 3, was applied. The first aspect of the logistic function was to produce a system so that it could capture the system's natural, cumulative growth rate of growth. The second aspect of the logistic function was to make a system such that the growth rate follows the system's capacity. Using the simple logistic method and Loglet Lab software, three parameters were approximated: 1) $k$ is the curve's growth rate; 2) $l$ is the saturated maximum capacity that a given system can sustain; and 3) $x_o$ is a location parameter.

The fitted model showed a 43.34 percent correlation with the observed data. The cumulative fitted models are shown in Figures 60-62. The results showed that the expected value of the saturation parameter was 16 ($t/t_o$) with a 95 percent confidence interval of (15.3, 16.4).

These parameters confidence interval were observed based on Bootstrap random sampling with replacement. The second parameter's growth time was 11.4 with a 95 percent confidence interval (9.6, 12.9), indicating that the characteristic growth of the function would lie between 9.6 and 12.9. The third parameter midpoint, or location parameter, was 8, which was equivalent to the *v/c* ratio at 0.45, indicating that the inflation point observed at *v/c* equals to 0.45. At the 95 percent confidence interval, the mid-point would lie when the *v/c* ratio was from 0.43 to 0.475.

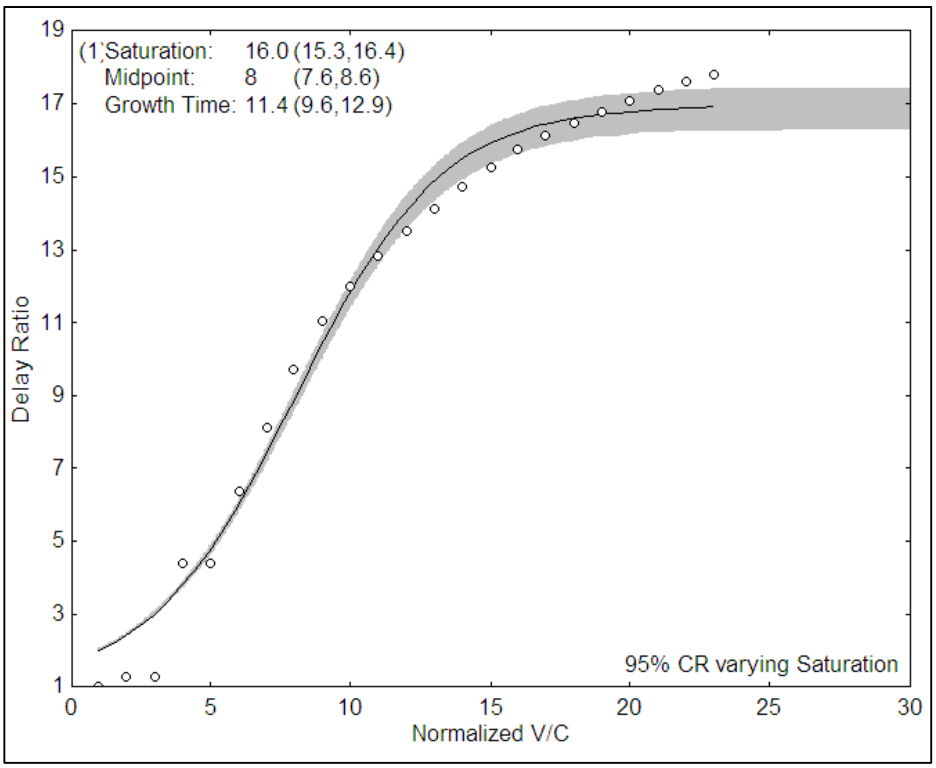


Figure 60. Saturation Parameters at 95 percent Confidence Interval

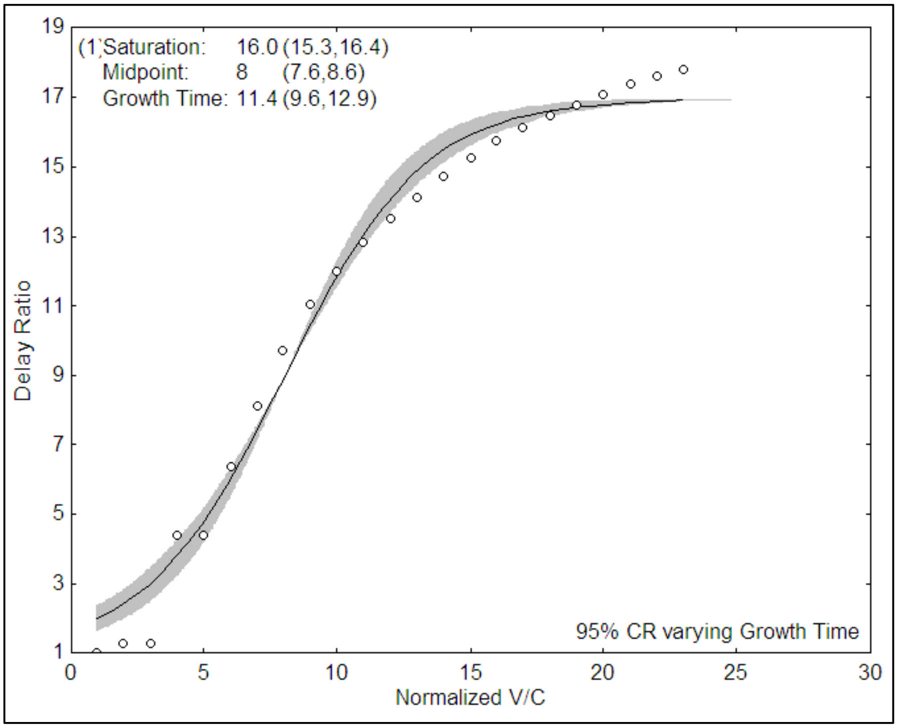


Figure 61. Growth Time Parameters at 95 Percent Confidence Interval

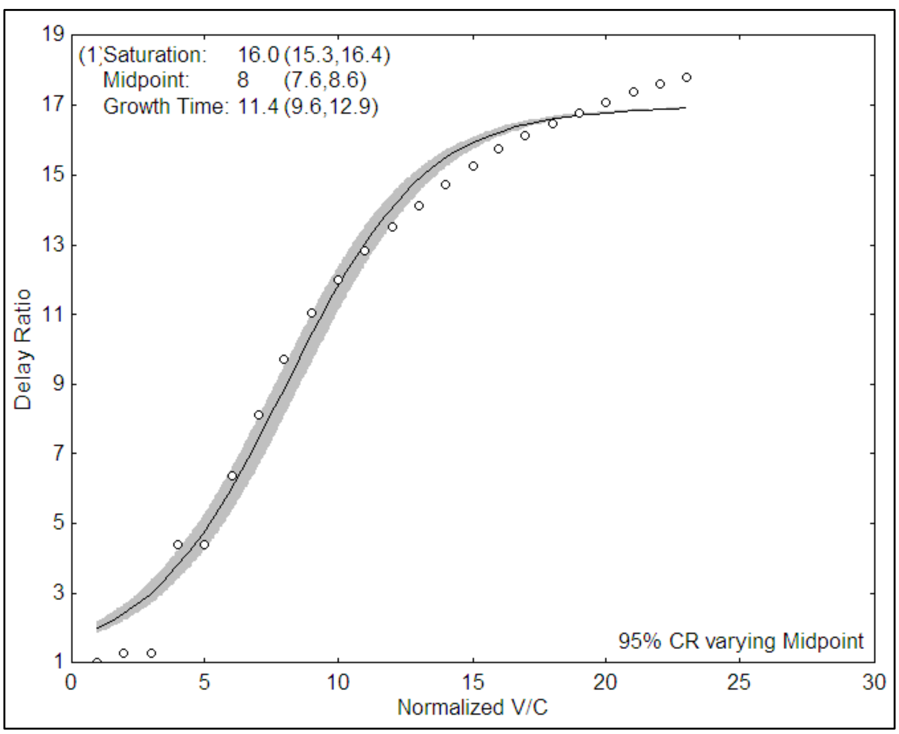


Figure 62. Midpoint Parameter at 95 Percent Confidence Interval

### 8.3. Validations of Proposed Model

The cumulative-growth function was fed back into the system to compute the $t/t_o$ with respect to the *v/c* ratio. Figure 63 shows how the predicted model was fitted with the observed conditions. The proposed model generated an R-squared value of 41.69 percent, indicating that the proposed model can explain approximately 41.69 percent of the variation. Validation of the proposed model with the modified BPR, Spiess Conical function, or modified Davidson model was performed.

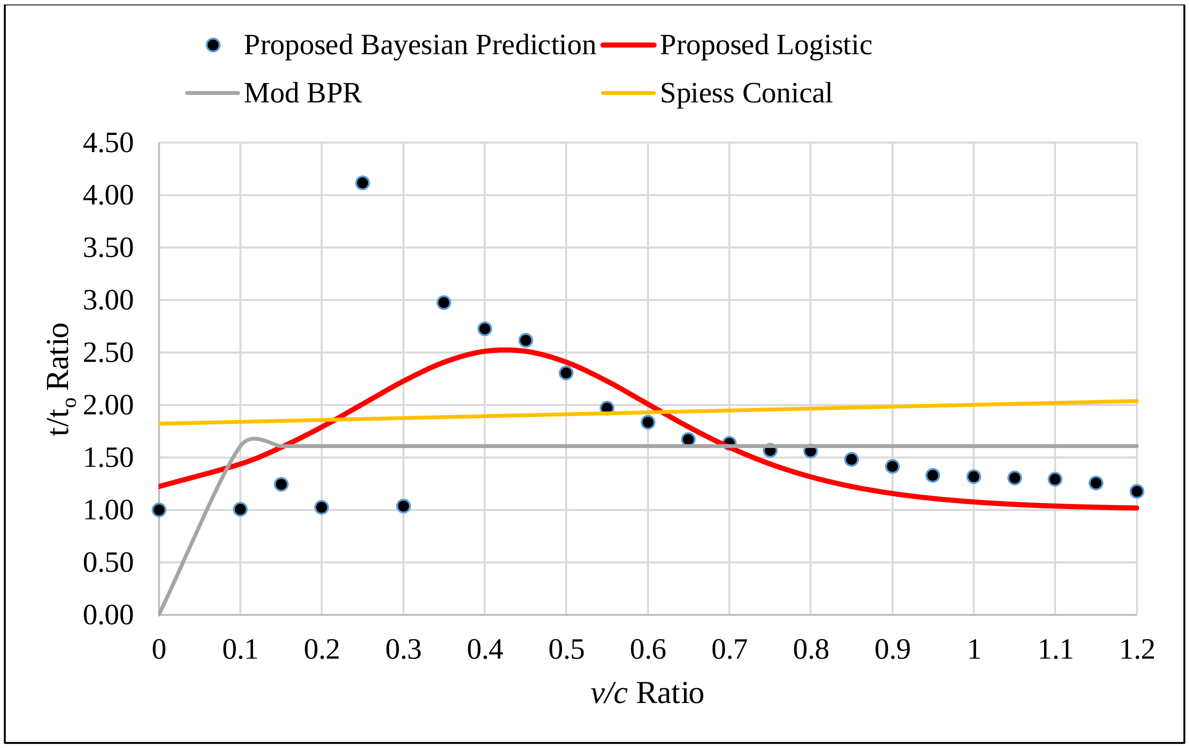


Figure 63. Predicted Value for the Different Fitted Models

The proposed model shows a better approximation for this case-study area. Existing models generate a very low R-square value, less than 7 percent in every case. Therefore, based on this case-study area, it might be inferred that the proposed, integrated Bayesian and logistic growth models performed better in $t/t_o$ predictions.

## 8.4. Sensitivity of Logistic Parameters

The three parameters' sensitivity was simulated. Each parameter's sensitivity was measured by holding the remaining parameters constant. The results showed that the first parameter (saturation parameter) was sensitive, i.e., $t/t_o$ higher with an increase for this parameter. Visualization of the results is presented in Figures 64-66. The simulation was done for parameter values of 5, 10, 50, 100, and 150. When the saturation parameter was 5, then the corresponding simulation results were presented by the Fitted5 line. The Fitted1 line illustrated the results obtained by the logistic function for the case-study area or the proposed model. For example, if the saturation parameter was 150 when holding the other two parameters constant, then the $t/t_o$ would be 16 times higher than the free-flow condition.

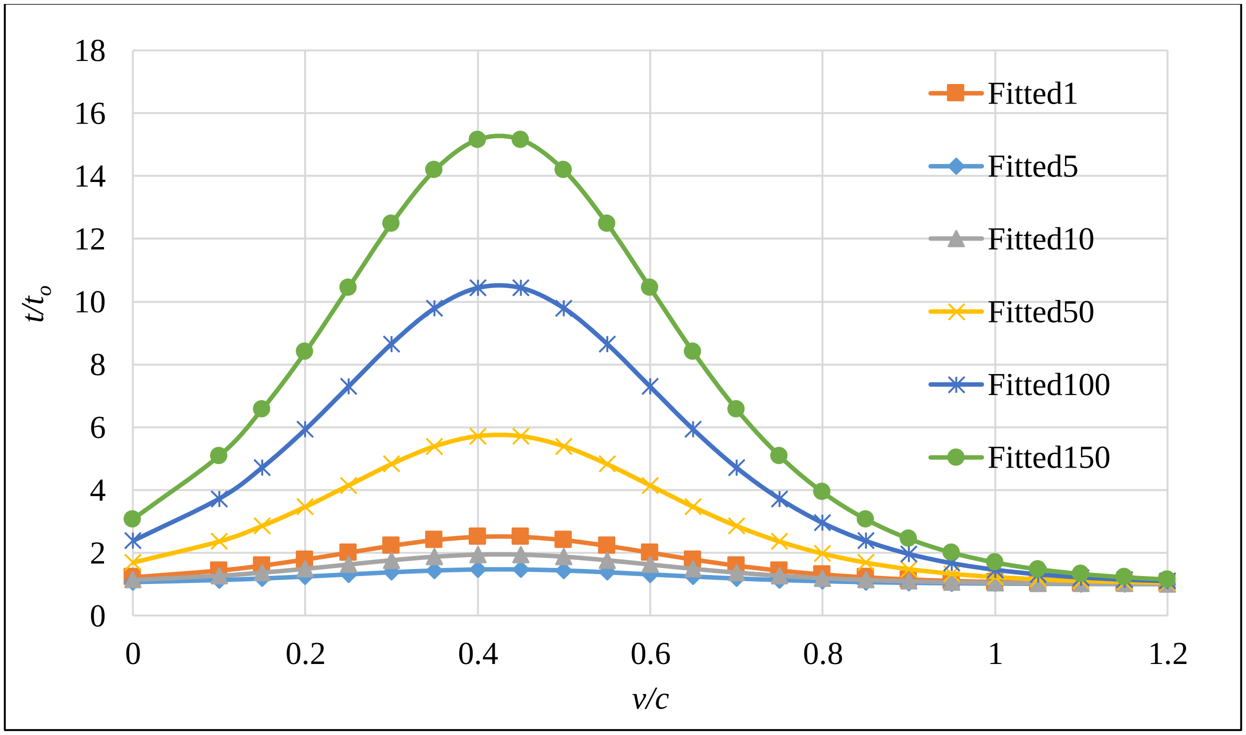


Figure 64. Saturation Parameter Sensitivity

A similar analysis was performed for the other two parameters. The model was not very sensitive with increased midpoint parameter. However, above a midpoint value of 40, which indicated an equivalent *v/c* ratio at or higher than 2.05, was straight line. In this case, the observed condition did not have sufficient data to justify this case, and practically, observation with the *v/c* ratio equal to or higher than 2.05 was impossible. Similarly, the growth-rate parameter was not highly sensitive compared to the saturation parameters.

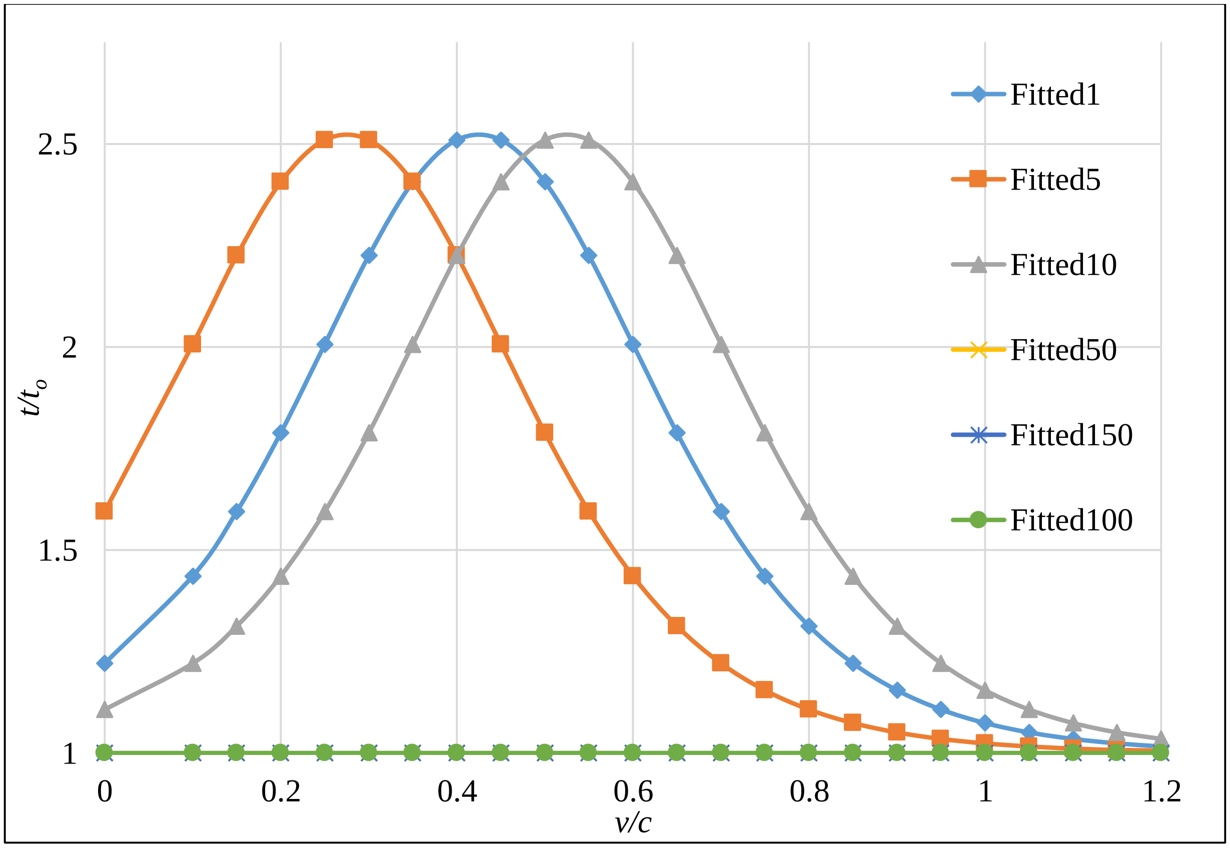


Figure 65. Midpoint Parameter Sensitivity

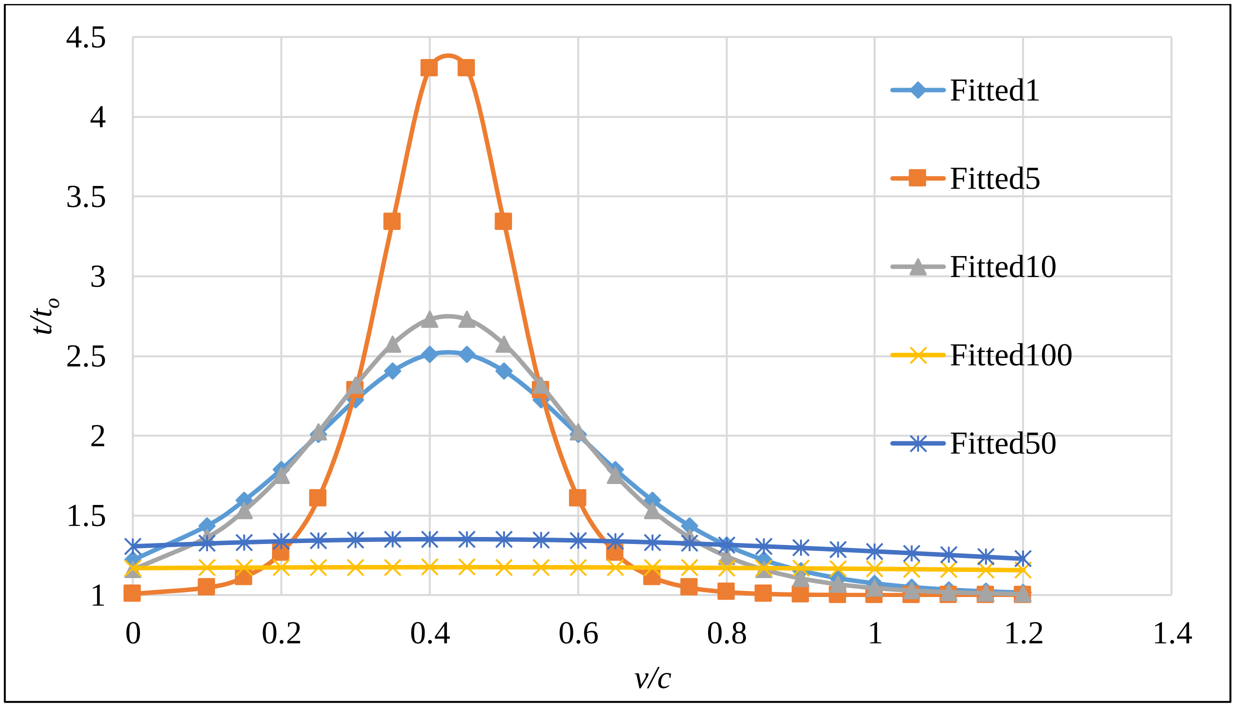


Figure 66. Growth Time Parameter Sensitivity

# 9. CONCLUSIONS AND RECOMMENDATIONS

This chapter summarizes and describes the research conclusions, the research's significance and contributions, recommendations for the transportation industry, recommendations for further study, and research limitations.

## 9.1. Summary of Virtual Sensor Method and Crowdsourcing

The VSM and crowdsource applications using OpenStreetMap services for smaller to heavier congested place in different State in the United States was tested. The results showed that this method is not suitable to achieve the research goals. In addition to this, NPMRDS travel time data was investigated for the Fargo-Moorhead Metropolitan area in the state of North Dakota and Minnesota. Validation using NPMRDS and VSM data with smartphone and test vehicle was conducted. But results showed that NPMRDS is very noisy. Since it's a new program through FHWA, this data source has never been evaluated comprehensively about the usability of this source for this purpose. Then Florida DOT's STEWARD programs concurrent data which includes traffic counts and travel time was utilized. But data was not available for consecutive years, which was required for this study. FDOT's Sample size was not sufficient enough to conduct this research as well.

Later, concurrent data in Chicago through (TravelMidWest program) was tested. Suitable format of data was not available. Then the data was collected from Mid Region of Council of Governments in New Mexico. One of the problem with this data is that sample size was not sufficient enough. Extracting data was tedious and not a suitable method to use this purpose. Subsequently, numerous efforts was made to collect concurrent data.

Finally, collecting concurrent data through CalTrans PEMS program was performed for this research. The case study area for this research included Los Angeles in California.

This research investigated the proximity of OpenStreetMap Services to collect the real-time or near-real-time travel time. The study validated the methodology against the NPMRDS datasets which are probe-vehicle information programed by the Federal Highway Administration.

Travel-time data using OpenStreetMap services indicated a great potential for the transportation industry. This study suggested that the crowdsource travel time through OpenStreetMap was significantly different than the real-time or near-time travel time. The reason for these significant difference might be the OpenStreetMap services' update frequency. However, it can be inferred that OpenStreetMap had the ability to replicate the observed trends in a corridor-level analysis.

The research's outcomes and developed tools will enrich the transportation industry's national transportation issues, contribute to the literature's needs, provide support for the national demand mandated by the Federal 21st Century Act Moving Ahead for Progress (MAP) 21. The smaller- and medium-size transportation agencies with vivid resource constraints might get substantial help when using OpenStreetMap.

Based on the findings and results, OpenStreetMap could be recommended as an alternative choice for the travel-time data-collection techniques, especially for agencies with resource constraints. Link, corridor, or O-D analysis can be performed very well with this method. There is a strong, potential opportunity to apply crowdsourcing technology, which needs due consideration by the engineers, policy makers, planners, researchers, practitioners, and the appropriate government or business entities, in the transportation field.

## 9.2. Summary of Location-Enabled Smartphone Applications

Travel-time data using web-mapping services indicated great potential for the transportation industry. This research investigated the proximity of replicating the real-time or near-real-time travel time by utilizing OpenStreetMap services. This study validated the proximity using a GPS-enabled smartphone and a test vehicle. During the research period, in addition to investigating the manually collected travel-time data with web-mapping service, the research investigate the suitability of a free app that can be used to collect the travel time. This research developed a geoprocessing tool which can automatically store and process the GPX file in a database with a single workflow.

The study revealed that collecting the travel time from web-mapping service had some issues. First, for the same origin and destination, the service may show different locations and direction shifting than the actual location. Therefore, a manual operation is needed to solve the problem. Second, sometimes, web service do not allow routing to start in the middle of the freeway. Because, the web router services forced us to select the closest facilities, such as land, parking, businesses, housing, etc. Data collection using a GPS-location-enabled smartphone illustrated some drawbacks for these methodologies. During the data collection, the GPS signal may be lost frequently. Data may show a noisy location and travel path. Sometimes, the app may not allow people to send or store the system's GPX data. Even if someone follows a random vehicle selection, the results may still have bias issues.

Regardless of each methodology's drawbacks, the results provided some interesting findings. At a 95 percent confidence interval, the probability of significance was greater than 0.05 for many cases, indicating that the travel time obtained with a web-mapping service may indicate the real-time or near-real-time travel time. This finding will be helpful, especially for smaller- or medium-size agencies where the travel time does not vary much for stable traffic area. The study suggested that the travel time is not significantly different than the observed condition for the average daily or peak-hour travel time. The OpenStreetMap services are complementary in replicating observed daily and peak-hour travel time. The study suggested that the AM and PM peak hours did not vary significantly on the freeway of smaller- or medium-size agencies. The study suggested OpenStreetMap as a potential candidate for the evolving crowdsource technology.

### 9.3. Summary of Capacity Estimation Methodology

It was very challenging to determine the highway capacity using existing knowledge. From local experience, there might be a question or doubt among the stakeholders about the capacity value borrowed from similar region and function class, or nationally default parameters. Therefore, different procedures were considered to utilize the existing knowledge for capacity estimation.

Literature review was performed about the highway capacity values from smaller to larger transportation agencies travel demand modeling documents. Several agencies has documented default capacity value. Some of the agencies has adopted different factors and formulas to calculate capacity. This knowledge-based literature review indicated that there are remarkable anomaly in capacity value utilization from smaller to larger agencies. Therefore, later an investigation with capacity enhancement was performed, which is out of scope of this dissertation. Based on 18 smaller to larger transportation agencies specially MPOs, a combined knowledge based system with incorporating HCM 2010 was proposed. This method incorporated several factors to be considered for capacity value. However, from local expertise and discussion, this method would not be suitable especially for smaller/medium size agencies because of the data unavailability. The definition of smaller to larger size agencies can be found in national co-operative highway research program (NCHRP) 716 report (TRB, 2012). Therefore, the enhanced method was further narrowed down by limiting influential factors which can be easily observed in online by the transportation agencies. But using this method, validation was challenging for capacity value. One of the biggest concern was that the collector road capacity value was less than the local road. The reason was the unavailability of capacity influential factors. Furthermore, simulation using varied softwares such as Synchro and VISSIM was performed for different classification of highway with 21 types of intersections based on number of lanes. Left-turn and right turn lane capacity was simulated using Synchro. This simulated results were further modeled to develop BPR speed-flow curve and find the capacity value. But one of the biggest problem with the simulation results was the shockwave of the traffic flow. It was evident that there were several reasons for not to using any existing capacity value.

Therefore, this research proposed a new methodology for the capacity estimation. This dissertation developed a new traffic-flow prediction model with an integrated approach as well as knowledge about the non-linear logistic growth and quantile-regression theorems. Later, this study included guidelines or framework to approximate the capacity from the observed flow characteristics. This dissertation established a new formal relationship between traffic flow and hour of a day. Later, the developed model provides guidelines estimating the steady-state capacity of freeway. Furthermore, the results were validated using several statistical measures and the existing capacity values from several public agencies.

Subjectivity for a region’s capacity estimation is very common. This study should eliminate the subjectivity of the freeway’s practical-capacity estimation. Even though this study suggests tolerance values for capacity estimation, the proposed model works in such a way that a better, stable region for the flow characteristics can be observed in a more advanced way.

Due to resource limitations, this research was not compared to the existing models, which could be considered as grounds for further research. It might be concluded that the model was advanced, strong, robust, and capable of replicating the natural trends of the traffic-flow characteristics for any condition. At the stable region of traffic flow, the proposed model converged to the maximum mean-hourly traffic-flow rate. Therefore, in future studies, the logistic growth model's chaotic behavior could be considered, which might improve the proposed methodology.

### 9.4. Summary of Free-Flow Speed-Estimation Methodology

This Chapter investigated and reported various deterministic, speed-density traffic-flow characteristic models. This study also included multi-regime and single-regime models. The study proposed and modified the existing, deterministic speed-density model. It incorporated LOS *A* in a FFS computation and speed prediction. Furthermore, a detailed investigation with simulated quantile characteristics and FFS sensitivity was also included. This study concluded that FFS estimation based on the speed-density model should incorporate the LOS and quantile function. The research also suggested the multi-regime model as the better one, especially in congested portions, for predicting speed depending on the density.

### 9.5. Summary of Bayesian Prediction

In Chapter 7, how the stochasticity of the traffic-flow characteristics are incorporated has been demonstrated. Later, using the historical data and PERT, prior knowledge about the $t/t_o$ condition's $v/c$ ratio was established. PERT was capable to represents the non-traditional characteristics of traffic flow.  Furthermore, likelihood and posterior, its parameters estimates significantly capture the existing condition and demonstrated that strictly increasing function should not be considered in travel time delay functions.

### 9.6. Summary of Logistic Growth Modeling

Literature review showed that several issues with the existing $t/t_o$ functions. First, existing methods that are in practice are always strictly increasing functions. This research proposed a new model, which can replicate the natural trends of the $t/t_o$ behavior.

Second, some of the models are linear in nature. The proposed model are non-linear and approximate the stochasticity of the $t/t_o$. Statistically, heavy congested place like Los Angeles cannot be well understood by the linear model. Therefore, considering the non-linearity issues, the proposed model would be considerably a better model instead of linear model.

Third, some models have issues when traffic volume is above capacity respectively. Therefore, the proposed model is a single regime, which does not have this issues.

Fourth, some models are very sensitive to the increase of its parameter. In the contrary, the proposed model parameters are not so sensitive. In a case where saturation parameter value is 150, then $t/t_o$ ratio could be less than 16. In reality, such a high value of saturation parameter might not be possible although study area's maximum delay ratio was observed approximately 22.

Fifth, there are binding constraint issues with different models. The proposed model showed that the binding constraint should not be consider while utilizing any model.

Sixth, the existing models, which produce scatter results. Study area cannot be well explained by existing model, which generates very low R-square value of less than seven percent. On the contrary, this model generates around 42 percent R-square value. On the contrary, the proposed model was 35 percent better for performing accurate predictions in compared to the existing methods.

Finally, this research incorporated better approximated steady-state capacity value and FFS value. In this way, the subjectivity of capacity selection and undefined FFS selection in practice can be eliminated.

In future studies, a better methodology might be considered to improve the coefficient of the determinations. However, this study incorporated the stochasticity for the $t/t_o$ function by integrating the Bayesian and logistic methods. The results showed that the proposed model was considerably a good candidate for $t/t_o$ predictions. The integrated model showed better in delay prediction than a single model using the overall data. It can be suggest that further improvement in stochasticity of $t/t_o$ functions is necessary. Because this study suggested that the travel-time delay was very unlikely and uncertain, it was very difficult to predict precisely. This research moderately improved the travel-time delay functions; moreover, a chaos study about the stochastic uncertainty of this modeling error might be formulated and investigated with further study.

### 9.7. Significances and Contributions

There are certain contributions and significance for this study based on two perspectives: 1) modeling perspective, and 2) data-collection perspective. First, significance and contribution on the modeling perspective are discussed. Second, significance and contributions with the data-collection efforts are included.

The research's unique contribution was that it proposed a new, stochastic approximation for the freeway travel-time congestion function based on knowledge borrowed from logistic growth mapping's market-adoption curve. This research provided a new scientific methodology and theoretical foundation for the $t/t_o$ prediction. This study developed an integrated approach to predict the travel time with Bayesian statistics and logistic growth mapping. The proposed methods were considerably a good candidates for $t/t_o$ estimations. This method can be used to estimate the impact of travel time to the travel demand modeling and transportation planning and operation. Incorporating methods that can be used to verify how well models perform when forecasting future traffic in metropolitan TDM processes will help improve the TDMs.

This study outperformed the traditional, strictly increasing functions' (theoretical aspect) consideration for $t/t_o$ predictions. A natural, multi-peak growth included delay rises as well as falls with respect to the $v/c$ ratio and was been incorporated with delay predictions. The probabilistic method was incorporated to remove partially the random error generated by the existing model.

One of the question is whether benefits worth the cost of this modeling or not. This proposed methodologies and models are computationally expensive and heavy resource intensive. However, in a heavy congested place, $t/t_o$ are stochastic and uncertain. Existing models showed incapable of replicating this stochastic and uncertain nature of $t/t_o$ prediction. The proposed method generates 35 percent better results than modified BPR and Spiess model. In addition to this, the proposed model overcome several issues, which has been discussed in Chapter 1. Therefore, considering the accurate and better results, the proposed methodologies and models benefits worth the costs.

This study revisited the assumptions of forcedly fitting the curve by constraining the travel-time function at zero volume or the FFS and volume at ultimate capacity. This study also supported revising the binding constraint at the ultimate capacity and FFS.

This study proposed a new, robust and stable, capacity-estimations method that can eliminate the capacity estimation's subjectivity. Investigation of different speed-density models by incorporating the LOS in speed density, modified the existing methods, compute FFS, and investigate sensitivity of FFS were completed. The capacity and FFS input parameters for developing the $t/t_o$ function are undefined, subjective, and biased. Therefore, a formal procedure was formulated to develop a delay model. The outcomes for these two items were unique and should address quantile-investigation scenarios, a new area for the transportation industry.

For the data-collection efforts, this study investigated different technologies, such as crowdsourcing, VSM, web applications, smartphone applications, test vehicles, and NPMRDS. The study also presented nationwide data needs for the $t/t_o$ function's development. Major contributions for data-collection efforts were extending knowledge about the VSM to collect travel-time data from crowdsource services. This study developed a web app, several tools, and an automated workflow to collect and analyze travel time; these items can useful in the transportation industry. The developed macro can be used for travel-time data collection and performance analysis of a transportation systems between unlimited number of origins and destinations (link level, route level, corridor level, O-D pair level, path level, and network level). Many critical scripts were written with R and SAS; the scripts can be utilized for data analysis and to develop a noise-cleaning algorithm from the data.

### 9.8. Limitations and Future Study

There were many challenges to overcome while conducting this study. The biggest challenge was to develop tools, macros, and writing codes as well as to collect data. The second challenge was to process data, including millions of records, overnight and to allocate a larger portion of the computer's memory. The data-analysis process considered numerous scripts in the

R and SAS programming languages at different analysis stages. However, certain limitations and future directions were addressed in this study.

This research only investigated the OpenStreetMap services. Other services can be utilized for further research. Therefore, this study welcomes other crowdsource services, such as Google, HERE, Bing, and MapQuest.

Due to the resource limitations, the research's capacity estimation was not compared to existing models, which could be considered grounds for further research. At the stable region of traffic flow, the proposed model converged to the maximum mean-hourly traffic-flow rate. In future studies, the growth model's chaotic behavior could be considered, which might improve the proposed capacity-estimation methodology.

This study included FFS estimation based on the existing speed-density models. For future study, a speed-density model can be developed, and the FFS sensitivity can be measured in a similar way.

Prior-information inference includes the PERT techniques. Other techniques can be utilized for future study. The minimum and maximum observed delay were used as the optimistic and pessimistic delay respectively. Considering various methods may produce different results, further study might include other techniques.

Validation methods for the Bayesian predictions are rare. Therefore, in future study, it is aimed to see if it is possible to find a validation testing methods for the Bayesian predictions of the $t/t_o$.

Further improvements with the stochasticity of the $t/t_o$ functions might be recommended. Because this study suggested that the $t/t_o$ is random, non-linear, and stochastic in nature, the delay is very difficult to predict precisely. This research improved the $t/t_o$ functions moderately but better compared to the existing models. Moreover, a chaos study about the stochastic uncertainty of this modeling error might be formulated and investigated with further study.

The model is sensitive to the maximum saturation parameter. Thus, chaos mapping might be necessary to incorporate this variation into the model. The future research intends to include several variable to evaluate the proposed model's behavior. In a future research, it is aimed to see the effects of other variables, such as traffic incident, major events, construction detours, and weather, in this model.

This method is computationally expensive. Therefore, it was aimed to develop software/tools that can be used by the transportation industry. Real-time ATIS and ATMS systems might be developed using this method which needs future consideration by the policy makers and decision-makers. Different highway-assignment algorithms might be investigated.

Subsets of dataset interval was considered 0.05. Better model might be developed if the dataset interval was 0.01. Interval 0.01 was considered due to several reasons. First of all, it will be computationally expensive for this dissertation. Second, sample size becomes lower for each subset. It was expected to have at least 30 random samples, so that a normal distribution requirement might be checked. For example, see Figure A.1 and A.2. These Figures shows that if the subset were 0.01, then the sample size requirement for prior information using the year 2013 data does not fulfill. However, in future study, it is aimed to see the difference of the results and model outcomes by comparing the subsets interval.

This research mainly focused on freeway functional class of road network and congested place like Los Angeles in California. This method can be tested for different time periods, functional classes, or roadway elements. Since the proposed model was investigated utilizing one case study, the observed parameters should be different for different regions or other highway functional class. In order to transfer or use the results and findings, the parameters needs to be calibrated before using it. Complete model needs to be tested before applying to other region or functional class. Therefore, transferability of this model outcomes would be considered for future depending on the suitability of data availability.

## REFERENCES


Akcelik, R. 1991. "Travel Time Functions for Transport Planning Purposes: Davidson's Functions, its Time-Dependent Form and an Alternative Travel Time Function." *Australian Road Research* 21 (3): 49-59.

Ayad, Hassan. 1967. "System Evaluation by the Simplified Proportional Assignment Technique." Purdue University, Lafayette, IN.

Ben-Akiva, Moshe, Michel Bierlaire, Didier Burton, Haris N. Koutspoulos, and Rabi Mishalani. 2001. "Network State Estimation and Prediction for Real-Time Traffic Management." *Network and Spatial Economics* 1: 293-318.

Bhaskar, Ashish, and Edward Chung. 2013. "Fundamental understanding on the use of Bluetooth scanner as a complementary transport data." *Transportation Research Part C* 37: 42-72. doi:http://dx.doi.org/10.1016/j.trc.2013.09.013.

Billings, Daniel. 2006. "Application of the ARIMA Models to Urban Roadway Travel Time Prediction-A Case Study." *2006 IEEE International Conference on Systemens, Man, and Cybernetics.* Taipei, Taiwan.

Branston, David. 1975. "Link Capacity Functions: A Review." *Transportation Research* 10: 223-236.

BRW. 2000. *Travel Time Data Collection Using Global Positioning System Technology.* Puget Sound Regional Council. Accessed 10 28, 2015. http://www.psrc.org/assets/1724/gps-ttreport8-00.pdf?processed=true.

Bureau of Public Roads. 1964. *Traffic Assignment Manual.* Washington, D.C.: Department of Commerce.

Caltrans. 2012. "District 07 Mobility Performance Report 2012."

Cambridge Systematics. 2013. "Incorporating Reliability Performance Measures into the Transportation Planning and Programming Processes." Accessed 10 20, 2015. http://www.camsys.com/pubs/SHRP2prepubL05Report.pdf.

Cambridge Systematics. 2012. *Travel Time Data Collection.* Tallahassee, FL: Cambridge Systematics.

Campbell, Robert B. 1959. *Design Hour Voulme Relationships.* Ontario Department of Highways. Planning Division.

Chen, Hao, Hesham A. Rakha, and Catherine C. McGhee. 2015. *Dynamic Travel Time Prediction using Pattern Recognition.* Accessed 11 23, 2015. http://www.citg.tudelft.nl/fileadmin/Faculteit/CiTG/Over_de_faculteit/Afdelingen/Afdeling_Transport_en_Planning/002Onderzoek/Chen__Rakha_and_McGhee_2013.pd.

Dailey, D. J., and F. W. Cathey. 2002. "Virtual Speed Sensors using Transit Vehicles as Traffic Probes." *The IEEE 5th International Conference on Intelligent Transportation Systems.* Singapore. 560-565.

Dailey, Daniel J., and Frederick W. Cathey. 2006. *Deployment of a Virtual Sensor System, based on Transit Probes, in an Operational Traffic Management System.* Washington State Transportation Center (TRAC), University of Washington, Seattle, Washington: Washington State Department of Transportation. Accessed 11 28, 2015. http://www.wsdot.wa.gov/research/reports/fullreports/660.1.pdf.

Davidson, K B. 1966. "A flow travel time relationship for use in transportation planning." *3rd Australian Road Research Board (AARB) Conference.* Sydney, Australia. 183-194.

Davidson, K B. 1978. "The theoretical basis of a flow–travel time relationship for use in transportation planning." *Australian Road Research* 8 (1): 32-35.

Dennies, Eric Paul, Richard Wallace, and Brian Reed. 2015. *Crowdsourcing Transportation Systems Data.* Lansing, MI: Michigan Department of Transportation (MDOT).

Dowling Associates. 1999. *Metropolitan Transportation Commission Travel Time Data Collection Pilot Project.* Metropolitan Transportation Commission. Accessed 11 3, 2015. http://www.mtc.ca.gov/library/TTPP.pdf.

Dowling, Richard, and Alexander Skabardonis. 2006. "Urban Arterial Speed-Flow Equaitons for Travel Demand Models." *Innovations in Travel Modeling 2006 A Transportation Research Board Conference.* Austin, Texas. Accessed 5 3, 2016. http://onlinepubs.trb.org/onlinepubs/archive/Conferences/TDM/papers/BS2C%20-%20Urban%20Arterial%20Speed-Flow.pdf.

Dowling, Richard, Rupinder Singh, and Willis Cheng. 1998. "Accuracy and Performance of Improved Speed-Flow Curves." *Journal of the Transportation Research Board* (1646): 9-17.

DuVander, Adam. 2012. *7 Free Geocoding APIs: Google, Bing, Yahoo and MapQuest.* 6 21. Accessed 11 21, 2015. http://www.programmableweb.com/news/7-free-geocoding-apis-google-bing-yahoo-and-mapquest/2012/06/21.

Elhenawy, Mohammed, Hao Chen, and Hesham A. Rakha. 2014. "Dynamic travel time prediction using data clustering and genetic programming." *Transportation Research Part C* 42: 82-98. doi:http://dx.doi.org/10.1016/j.trc.2014.02.016.

Fangfang, Zheng, WanYu, and Wu Pingheng. 2008. "Link Travel_Tiem Prediction Using Extended Exponential Smoothing and Kalman Filter in Dynamic Networks." *The Eighth International Conference of Chinese Logistics and Transportation Professionals Logistics.* ASCE. 3753-3759.

Fei, Xiang, Chung-Cheng Lu, and Ke Liu. 2011. "A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction." *Transportation Research Part C* 19: 1306-1318.

FHWA. 1998. "Travel Time Data Collection Handbook." Accessed 10 15, 2015. https://www.fhwa.dot.gov/ohim/tvtw/natmec/00020.pdf.

FixMyTransport. n.d. *Looking for FixMyTransport?* Accessed 11 28, 2015. http://www.fixmytransport.com/.

Flyvbjerg, Bent, Mette K.S. Holm, and Soren L. Buhl. 2005. "How (In)accurate are Demand Forecasts in Public Works Projects?" *Journal of the American Planning Association* 71(2).

Fokas, Nikos. 2007. "Growth Functions, Social Diffusion, and Social Change." *Review of Sociology* 13: 5-30.

Gan, Albert, Rax Jung, Min-Tang Li, and Chang-Jen Lan. 2003. "Incorporating Variable Peak-to-Daily Ratios into FSUTMS to Reduce Assignmetn Errors." Accessed 5 13, 2016. http://www.fsutmsonline.net/images/uploads/reports/FDOT_BC791_v2_rpt.pdf.

Gelman, Andrew, John B Carline, and Hal S. Stern. 2003. *BAyesian Data Analysis.* Boca Raton, FL: CRC Press.

Gesalem, Jiaan Regis G., and Alexis M. Fillone. 2016. Accessed 5 3, 2016. http://www.dynamicglobalsoft.com/easts2015/program/pdf_files/1576.pdf.

Golding, S. 1978. "On Davidson’s Flow-Travel Time Relationship for Use in Transportation Planning." *Australian Road Research* 8 (3): 36-37.

Google. 2015. Accessed 7 15, 2015. https://developers.google.com/maps/documentation/distancematrix/intro.

Grol, Rik Van, Karel Lindveld, Simonetta Manfredi, and Mehdi Danech-Pajouh. 1999. *DACCORD: On-Line Travel Time Estimation/Prediction Results.* Nedarlands. Accessed 11 27, 2015. http://www.significance.nl/papers/1999-ITS-DACCORD-Online-Travel-Time-Estimation.pdf.

Hadi, Mohammed, and Haitham Al-Deek. 2015. *NCTSPM.* Accessed 10 28, 2015. http://nctspm.gatech.edu/pi/performance-measurements-transportation-systems-based-fine-grained-data-collected-avi-and-avl.

Haghani, Ali, and Yashar Aliari. 2012. "Using Bluetooth Sensor Data for Ground-truth Testing of Reported Travel Times." *Transportation Research Board 91st Annual.* Washington, D.C.

Halder , Buddhadeb. 2014. "Evolution of Crowdsourcing: Potential Data Protection, Privacy and Security Concerns under the New Media Age." *Democracia Digital e Governo Eletrônico* 10: 377-393.

Hall, Fred L. 1975. *Traffic Stream Characteristics.* Department of Civil Engineering and Department of Geography, McMaster University, Ontario, Canada: Federal Highway Administration. Accessed 12 31, 2016. http://www.fhwa.dot.gov/publications/research/operations/tft/chap2.pdf.

Hamm, Robert A. 1993. "An Evaluation of Travel Time Estimation Methodologies." College Staion, TX. Accessed 11 3, 2015. http://d2dtl5nnlpfr0r.cloudfront.net/swutc.tamu.edu/publications/technicalreports/712410-4D.pdf.

Hao, Liu, Zhang Ke, Wang Zilei, and Niu Shuyun. 2009. "A Comparison of Existing Algorithms for Travle Time Estimation." *International Conference on Transportation Engieering 2009.* Chengdu, China. 189-194.

Hass, Robert, Mark Carter, Eric Perry, Jeff Trombly, Elisabeth Bedsole, and Rich Margiotta. 2009. *iFlorida Model Deployment Final Evaluation Report.* National Transportation Library. Accessed 10 28, 2015. http://ntl.bts.gov/lib/31000/31000/31051/14480_files/iflorida.pdf.

Haworth, James, John Shawe-Taylor, Tao Cheng, and Jiaqui Wang. 2014. "Local online kernel ridge regression for forecasting of urban travel times." *Transportation Research Part C* 46: 151-178.

HERE. 2015. Accessed 11 21, 2015. https://developer.here.com/rest-apis/documentation/routing/topics/introduction.html.

Herrera, Juan C, Daniel B Work, Ryan Herring, Xuegang Jeff Ban, and Quinn Jacobson. 2010. "Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment." *Transportation Research Part C: Emerging Technologies* 18 (4): 568-583. doi:10.1016/j.trc.2009.10.006.

Horowitz, Alan J. 1991. *Delay-Volume Relations for Travel Forecasting: Based on the 1985 Highway Capacity Manual.* Federal Highway Administation. Accessed 4 23, 2016. https://www.fhwa.dot.gov/planning/tmip/publications/other_reports/delay_volume_relations/dvrt.pdf.

Huang, Yanguo, Lunhui Xu, Qiang Luo, and Xianyan Kuang. 2013. "Urban Expressway Travel Time Prediction Method Based on Fuzzy Adaptive Kalman Filter." *Applied Mathematics & Information Sciences* 7 (2L): 625-630.

Irwin, N A, and H G Van Cube. 1962. "Capacity restraint in multi-travel mode assignment programs." *Transportation Research Board* (347): 258-289.

Irwin, N A, Norman Dodd, and H G Von Cube. 1961. "Capacity Restraint in Assignment Programs." *Transportation Research Board* 297: 109-127.

Ishak, Sherif, and Haitham Al-Deek. 2002. "Performance Evaluation of Short-Term Time-Series Traffic Prediciton Model." *Journal of Transportation Engineering* 128 (6): 490-498.

Jiang, Zhou, Cunbao Zhang, and Yinxia Xia. 2014. "Travel Time Prediction Model for Urban Road Network Based on Multi-source Data." *The 9th International Conference on Traffic & Transportation Studies (ICTTS'2014).* Procedia - Social and Behavioral Sciences. 811-818.

Jie, Li, Henk Van Zuylen, Liu Chunhua, and Lu Shoufeng. 2011. "Monitoring travel times in an urban network using video, GPS and Bluetooth." *Procedia Social and Behavioral Sciences* 20: 630-637. doi:10.1016/j.sbspro.2011.08.070.

Kalaee, Meead Saberi. 2010. *A MS Thesis: Investigating Freeway Speed-Flow Relationships for Traffic Assignment Applications.* Portland State University.

Kim, Byung-cheol, and Kenneth F. Reinschmidt. 2009. "Probabilistic Forecasting of Project Duration Using Baysian Inference and the Beta Distribution." *Journal of Construction Engineering and Management* (ASCE) 135: 178-186.

KMJ Consulting. 2011. "Using Bluetooth to Measure Travel Tiem Along Arterial Corridors." Ardmore, PA. Accessed 10 27, 2015. http://www.kmjinc.com/wp-content/uploads/BT-white-paper-11_29_2011.pdf.

Koenker, Roger. 2016. "Quantile Regression Package 'Quantreg'." Accessed 7 27, 2016. https://cran.r-project.org/web/packages/quantreg/quantreg.pdf.

Kurkcu, Abdullah, Ender Faruk Morgul, and Kaan Ozbay. 2015. "Extended Implementation Methodology for Virtual Sensors: Web-based Real Time Transportation Data Collection and Analysis for Incident Management." *Paper submitted for Publication in the Transportation Research Record, Journal of Transportation Research Board.*

Lin, Hong-En, Rocco Zito, and Michael A P Taylor. 2005. "A Review of Travel-Time Prediciton in Transport and Logistics." *Proceedings of the Eastern Asia Society for Transportation Studies.* 1433-1448.

Lint, J.W.C. Van, S.P. Hoogendoorm, and H.J. van Zuylen. 2005. "Accurate freeway travel time prediction with state-space neural networks under missing data." *Transportation Research Part C* 13: 347-369.

Malcolm, D. G., J. H. Roseboom, C. E. Clark, and W. Fazar. 1959. "Application of a Technique for Research and Development Program Evaluation." *Operations Research* 7: 646-69.

MapQuest. 2015. *Directions API.* Accessed 11 21, 2015. https://developer.mapquest.com/products/directions/.

May, A. D., and F. O. Montgomery. 1984. "Factors Affecting Travle Times on Urban Radial Routes." Institute of Transport Studies, University of Leeds, Leeds, UK. Accessed 11 26, 2015. http://eprints.whiterose.ac.uk/2362/1/ITS366A_WP177_uploadable.pdf.

Mendes-Moreira, Joao, Aliipio Mario Jorge, Freire de Sousa, and Carlos Soares. 2015. "Improvingtheaccuracyoflong-termtraveltimepredictionusing." *Neurocomputing* 150: 428-439.

Meyer, Perrin. 1994. "Bi-Logistic Growth." *Technological Forecasting and Social Change* 47: 89-102. Accessed 5 21, 2016. http://phe.rockefeller.edu/Bi-Logistic/.

Microsoft. 2015. Accessed 11 21, 2015. https://www.microsoft.com/en-us/download/confirmation.aspx?id=19768.

Misra, Aditi, Aaron Gooze, Kari Watkins, Mariam Asad, and Christopher A. Le Dantec. 2014. "Crowdsourcing and Its Applicaiton to Transportation Data Collection and Management." *Journal of the Transportation Research Board* 2414: 1-8.

Mittal, Ayush Kr., and Deepika Bhandari. 2012 . "A Novel Approach to Implement Green Wave system and Detection of Stolen Vehicles." *IEEE.* Accessed 10 28, 2015. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6514372.

Molla, Mohammad M. 2016. "Implications of Highway Capacity Manual on Freeway Measure of Effectiveness: A Case Study." *IJRET: International Journal of Research in Engineering and Technology* 5 (1): 72-78.

Morgul, Ender Faruk, Hong Yang, Abdullah Kurkcu, Kann Ozbay, Bekir Bartin, Camille Kamga, and Richard Salloum. 2014. "Virtual Sensors: A Web-based Real-Time Data Collection Methodology for Transportation Operation Performance Analysis." *TRB 2014 Annual Meeting.* Washington, D.C. Accessed 10 15, 2015. http://engineering.nyu.edu/urbanmits/trbpaper/14-4119.pdf.

Moriarty, David J. 2015. 5 14. Accessed 5 23, 2016. https://www.cpp.edu/~djmoriarty/wed/bayes_handout.pdf.

Moriasi, D. N., J. G. Arnold, M. W. V. Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. 2007. "Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations." *Transactions of the ASABE* 50 (3): 885-900.

Moses, Ren, Enock Mtoi, Steve Ruegg, and Heinrich McBean. 2013. "Development of Speed Models for Improving Travel Forecasting and Highway Performance Evaluation." Project

No. BDK83 Task Work Order No. 977-14, Prepared for Florida Department of Transportation. Accessed 9 3, 2016. http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_PL/FDOT-BDK83-977-14-rpt.pdf.

Mosher Jr, Walter W. 1963. "A Capacity-Restriant Algorithm for Assigning Flow to a Transport Network." *Transporation Research Board* (6): 41-70.

Motuba, Diomo. 2012. *FM Metro COG 2010 Travel Demand Model Update: Proposed Models and Methods.* Fargo, ND: Fargo-Moorhead Metropolitan Council of Goverments. Accessed 9 3, 2016. http://www.fmmetrocog.org/new/assets/documents/TTC%20packets/2012/420%20Packet%20November%208%202012.pdf.

Mtoi, Enock T., and Ren Moses. 2014. "Calibration and Evaluation of Link Congestion Functions: Applying Intrinsic Sensitivity of Link Speed as a Practical Consideration to Heterogeneous Facility Types within Urban Network." *Journal of Transportation Technologies* 4: 141-149.

Overgaard, K R. 1967. "Urban Transportation Planning' Traffic Estimation." *Eno Transportation Foundation* 21 (5).

Owen, Claire Elayne Bangerter. 2008. *Parameter Estimation for the Beta Distribution.* All Theses and Dissertations. Paper 1614.

Park, Dongjoo, Laurence R. Rillet, and Gunhee Han. 1999. "Spectral Basis Neural Networsk for Real-Time Travel Time Forecasting." *Journal of Transportation Engineering* 125 (6).

Petty, Karl F., Peter Bickel, Michael Ostland, John Rice, Frederic Schoenberg, Jiming Jiang, and Ya'Acov Ritov. 1998. "Accurate Estimation of Travel Times from Single-Loop Detectors." *Transporation Research Part: A* 32 (1): 1-18.

Porter, J. David, David S. Kim, and Mario E. Magana. 2011. "Wireless Data Collection System for Real-time Arterial Travel Time Estimates." Portland, OR. Accessed 10 28, 2015. http://www.oregon.gov/odot/td/tp_res/docs/reports/2011/wirelessdata.pdf.

Publicstuff. n.d. Accessed 11 28, 2015. http://www.publicstuff.com/submit.

Rahmani, Mahmood, Haris N. Koutsopoulos, and Anand Ranganathan. 2010. "Requirements and Potential of GPS-based Floating Car Data for Traffic Management: Stockholm Case Study." *13th International IEEE Annual Conference on Intelligent Transportation Systems.* Madeira Island, Portugal. 730-735. Accessed 10 28, 2015. https://people.kth.se/~mahmoodr/Publications_files/Rahmani,%20Koutsopoulos,%20Ranganathan%20-%202010%20-%20Requirements%20and%20potential%20of%20GPS-based%20floating%20car%20data%20for%20traffic%20management%20Stockholm%20c.pdf.

Rasouli, Soora, and Harry Timmermans. 2012. "Uncertainity in travel demand forecasting models: Literature review and research agenda." Accessed 7 30, 2015. http://www.uncertweb.org/uploads/papers/fb8c49bb0ef9c11d2aa4890220beeeb4488e510e.pdf.

Rehan, Asif. 2015. "Crowdsourcing Real-Time Traveler Information Systems." *University of Connecticut.* 8 6. Accessed 10 27, 2015. http://digitalcommons.uconn.edu/cgi/viewcontent.cgi?article=1893&context=gs_theses.

Roess, Roger P., Elena S. Prassas, and William R. McShane. 2011. *Traffic Engineering.* Vol. 4th Edition. Pearson Prentice Hall.

Sanwal, Kumud K., and Jean Walrand. 1995. "Vehicle as Probes." Institute of Transportation Studies, University of California, Barkeley, CA. Accessed 11 29, 2015. http://www.path.berkeley.edu/sites/default/files/publications/PWP-95-11.pdf.

SAS Institute Inc. 2016. *SAS/STAT(R) 9.2 User's Guide, Second Edition.* Accessed 5 23, 2016. https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_introbayes_sect006.htm.

SeeClickFix. 2015. Accessed 11 28, 2015. http://gov.seeclickfix.com/products.

Siemens AG. 2014. "Siemens traffic detectors." Munich, Germany. Accessed 11 3, 2015. http://www.mobility.siemens.com/mobility/global/sitecollectiondocuments/en/road-solutions/urban/infrastructure/verkehrsdetektoren-en.pdf.

Sinani, Artan. 2015. *Learning Bing Maps API.* Accessed 11 21, 2015. https://www.packtpub.com/sites/default/files/9781783550371_Chapter-01.pdf.

Smith, Reynolds, Alan Hiriwitz Hills, William McShane, and Deakin Harvey Skabardonis. 1999. "NCHRP 3-55(2)A: Pllaning Applications for the Year 2000 Highway Capacity Manual."

Smock, Robert. 1962. "An Iterative Assignment Approach to Capacity Restraint on Arterial Networks." *Transportation Research Board* 347: 60-66.

Soltman, Theodore J. 1966. "Effects of Alternate Loading Sequences on Results from Chicago Trip Distribution and Assignment Model." *Transportation Research Board* 114: 122-140.

Soriguera, F., D. Rosas, and F. Robuste. 2010. "Travel time measurement in closed toll highways." *Transportation Research Part B* 44: 1242-1267.

Spiess, Heinz. 1990. "Technical Note-Conical Volume-Delay Functions." *Transportation Science* 24 (2): 153-158.

Stackexchange. 2011. *Calculating the parameters of a Beta distribution using the mean and variance.* 6 22. Accessed 5 14, 2016. http://stats.stackexchange.com/questions/12232/calculating-the-parameters-of-a-beta-distribution-using-the-mean-and-variance.

Stackexchange. 2015. *Difference between standard beta and unstandard beta distributions?* 12 12. Accessed 5 14, 2016. http://stats.stackexchange.com/questions/186465/difference-between-standard-beta-and-unstandard-beta-distributions?answertab=active#tab-top.

Steenbrink, PA. 1974. *Optimization of Transport Networks.* John Wiley & Sons Ltd.

Tak, Sehyun, Sunghoon Kim, Kiate Jang, and Hwasoo Yeo. 2014. "Real-Time Travel Time Prediction Using Multi-level k-Nearest Neighbor Algorithm and Data Fusion Method." *Computing in Civil and Building Engineering* 1861-1868.

Tam, Mei Lam, and William H.K. Lam. 2015. "Using Automatic Vehicle Idenification Data for Travel Time Estimation in Hong Kong." *Transportmetrica* 179-194. Accessed 10 27, 2015. doi:10.1080/18128600808685688.

Tam, Mei Lam, and William H.K. Lam. 2015. "Using Automatic Vehicle Identification Data for Travel Time Estimation in Hong Kong." *Transportmetrica* 179-194. Accessed 10 27, 2015. doi:10.1080/18128600808685688.

Traffic Research Corporation. 1966. "Winnipeg Area Transportation Study. Report Prepared for the Streets and Transit Division of the Metropolitan Corporation of Greater Winnipeg."

TRB. 2010. *HCM 2010: Highway Capacity Manual* . National Academy of Sciences.

TRB. 1985. *Special Report 209: Highway Capacity Manual.* Washington, D.C.: National Research Council.

TRB. 2013. "Analytical Procedures for Determining the Impacts of Reliability Mitigation Strategies." SHRP2, Washington, D.C.

TRB. 2008. "Cost-Effective Performance Measures for Travel Time Delay, Variation, and Reliability." Washington, D.C.

TRB. 2000. "Highway Capacity Manual 2000." Washington, D.C.

TRB. 2012. "NCHRP Report 716: Travel Demand Forecasting: Parameters and Techniques." Washington, D.C.

TRB. 2015. "The On-Line Travel Survey Manual: A Dynamic Document for Transportation Professionals." *Travel Survey Manual Update.* Accessed 7 15, 2015. http://www.travelsurveymanual.org/Chapter-1-1.html.

TRB. 1975. "Traffic Flow Theory A Monograph, Special Report 165." Washington, D.C.

Turner, Shawn M. 1995. "Advanced Techniques for Travel Time Data Collection." *IEEE* .

USA Today. 2015. "Los Angeles Top Worst Cities for Traffic in USA." Accessed 7 28, 2016.

Wang, Chao, and Joseph B. Huegy. 2014. "Determining the Free-Flow Speeds in a Regional Travel Demand Model Based on the Highway Capacity Manual." *Submitted for presentation at the 93rd TRB Annual Meeting and publication in the Transportation Research Record.* Washington, D.C.

Wang, Fahui, and Yanqing Xu. 2011. "Estimating O-D travel time matrix by Google Maps API: implementation, advantages, and implications." (Annals of GIS) 14:4: 199-209.

Wang, Haizhong, Jia LI, Qian-Yong Chen, and Daiheng Bi. 2009. "Speed-Density Relationship: From Deterministic to Stochastic." *Resubmitted to TRB 88th Annual Meeting* . Washington, D.C.

Washburn, Scott S., and Nancy L. Nihan. 1999. "Estimating Link Travel Time with the Mobilizer Video Image Tracking System." *Journal of Transportation Engineeering* 125 (1): 15-20.

Wikipedia. 2016. "https://en.wikipedia.org/wiki/Logistic_function." Accessed 7 28, 2016.

Wright, John, and Joy Dahlgren. 2001. "Using Vehicles Equipped with Toll Tags as Probes for Providing Travel Times." California Partners for Advanced Transportation Technology UC Berkely. Accessed 10 28, 2015. http://escholarship.org/uc/item/9f17h2j0.

Wu, Chun-Hsin, Jan-Ming Ho, and D. T. Lee. 2004. "ravel-Time Prediction With Support Vector Regression." *IEEE Transactions on Intelligent Transportation Systems* 5 (4).

Yang, Chao, Anthony Chen, Xiangdong Xu, and S.C. Wong. 2013. "Sensitivity-based uncertainty analysis of a combined." *Transportation Research Part B* 57: 225-244.

Young, Stan. 2007. "Real-Time Traffic Operations Data Using Vehicle Probe Technology." *Proceedings of the 2007 Mid-Continent Transportation Research Symposium.* Ames, IA. Accessed 10 27, 2015. http://www.ctre.iastate.edu/pubs/midcon2007/youngvehicleprobe.pdf.

Zhang, Michael, Tong Qiang Wu, and Eil Kwon. 1997. *Arterial Link Travel Time Estimation Using Loop Detector Data.* Iowa City, IA: The Univeristy of Iowa. Accessed 10 29, 2015. http://ir.uiowa.edu/cgi/viewcontent.cgi?article=1006&context=ppc_transportation.

Zhang, Wang. 2006. "Freeway Travel Time Estimation Based on Spot Speed Measurmentas." Virginia Polytechnic Institute and State University, Blacksburg, Virginia. Accessed 10 29, 2015. http://scholar.lib.vt.edu/theses/available/etd-06292006-000823/unrestricted/Wang_Zhang_Dissertation_ETD_Copy_Ver_08_18_After_ETD_4th_Review.pdf.

Zhang, Xiaoyan, and John A. Rice. 2003. "Short-term travel time prediciton." *Trasnsportation Research Part C* 11: 187-210.

Zhang, Yanru, and Ali Haghani. 2015. "A gradient boosting method to improve travel time prediction." *Transportation Research Part C* 58: 308-324. doi:http://dx.doi.org/10.1016/j.trc.2015.02.019.

Zhao, Yong, and Kara Maria Kockelman. 2002. "The propagation of uncertainty through travel demand models: An exploratory analysis." *The Annals of Regional Science* 36 (1): 145-163.

Zito, R., G. D'este, and M. A. P. Taylor. 1995. "Global Positioning Systems in the Time Domain: How Useful a toll for intellgient vehicle-highway systems." *Transporation Research Part C* 3 (4): 193-209.

Zou, Yajie, Xinsin Zhu, Yunlong Zhang, and Xiaosi Zeng. 2014. "A space–time diurnal method for short-term freeway travel time prediciton." *Transportation Research Part C* 43: 33-49. doi:http://dx.doi.org/10.1016/j.trc.2013.10.007.

# APPENDIX

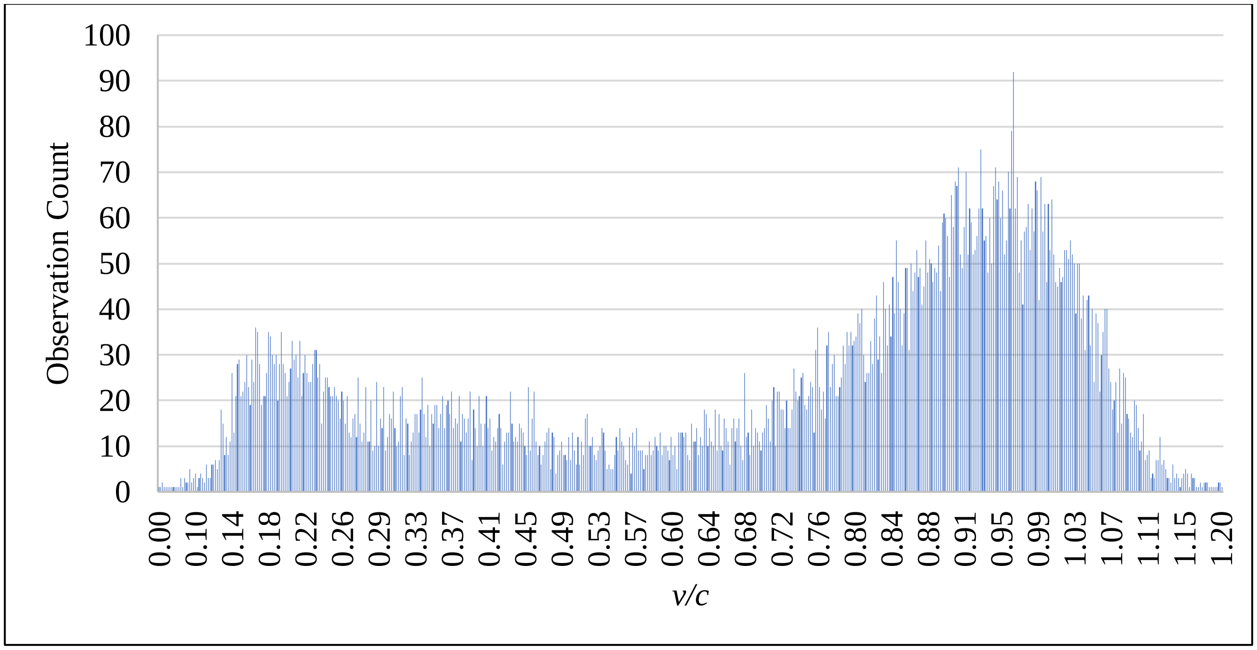


Figure A1. Number of Observation Rounded at *v/c=0.01*

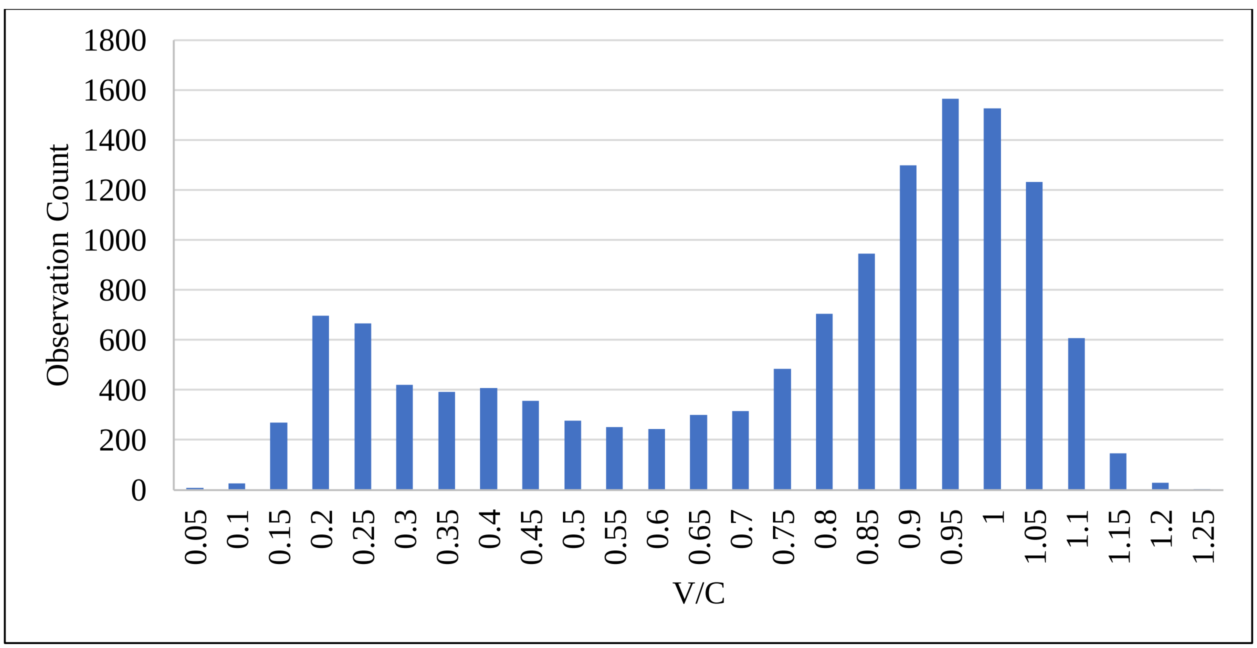


Figure A2. Number of Observation Rounded at *v/c=0.05*