ABSTRACT

LIU, CHENHAO. Temporal Stability of Macroscopic Traffic Stream Models. (Under the direction of Dr. Billy M. Williams.)

A quantitative analysis regarding the long-term temporal stability of freeway macroscopic traffic stream models is developed and presented. A macroscopic traffic stream model typically includes several parameters, such as free-flow speed, jam density and so on. Previous research similar to this topic has focused on the variation pattern of traffic flow rate only, therefore none of them has accounted for the temporal stability of other model parameters that quantify the characteristics of freeway traffic flow. The analysis presented herein utilizes 15-year freeway traffic data series recorded by MIDAS database from London, UK. Results indicate that the parameters of typical freeway traffic stream models remain relatively stable during the 15-year analysis period; therefore it is favorably positive to use the model parameters’ average value over the entire 15 years to represent future traffic conditions.
Temporal Stability of Macroscopic Traffic Stream Models

by
Chenhao Liu

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Civil Engineering

Raleigh, North Carolina
2010

APPROVED BY:

Dr. Billy M. Williams  Dr. Nagui M. Rouphail
Committee Chair

Dr. Peter Bloomfield
DEDICATION

To my parents.
BIOGRAPHY

Chenhao Liu comes from Binzhou, a small but peaceful town in Shandong Province, P.R.China. He received a B.S. in Civil Engineering in July, 2008 from Tsinghua University, Beijing, P.R.China. He started his graduate school at the Civil Engineering program of North Carolina State University in August, 2008 and expects to graduate with an M.S. in August, 2010.

During his time at N.C. State, Chenhao served as a graduate research assistant on the NCDOT Project “Assessing Operational, Pricing, and Intelligent Transportation System Strategies for the I-40 Corridor Using DYNASMART-P” under advisor Dr. Billy M. Williams. He maintained a 3.94 GPA upon graduation. He is an active member of TRB and ITE. After graduation, he plans to start his Ph.D. at Georgia Institute of Technology, Atlanta, GA.
ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Billy M. Williams, whose guidance, intellect and profound passion have been invaluably inspirational over the past two years. His ideas, advice and help made this research possible.

I would also like to extend my thanks to my thesis committee members, Dr. Nagui M. Roupail and Dr. Peter Bloomfield. Dr. Roupail’s profound academic knowledge and research experience in transportation engineering makes him an invaluable resource. Dr. Bloomfield’s broad vision in statistics helps to further develop and refine the statistical analysis in the research of this thesis.

I would also like to thank my colleagues in the transportation engineering program at ITRE, including Anxi Jia, William Reynolds, Hyejung Hu and Hassan Swidan. They have offered constructive ideas and enthusiastic encouragement throughout the processes of this research.

During the last two years of life in the U.S., I have acquainted myself with a lot of friends in and out of NCSU. I would like to express my thanks to them for making my life wonderful and memorable. They are: Jing Zhao, Zhuoning Liu, Yan Wang, Yingfang Yao, Qin Sun, Gongyi Liu, Zhuolun Zhang, Ziran Gu…

Finally, I want to express a sincere thank you to my ex-girlfriend, Lichao Zhang, for being supportive and inspiring when we were together. Without that precious memory none of this would have come true.
# TABLE OF CONTENTS

LIST OF TABLES ......................................................................................................................................... vii

LIST OF FIGURES ...................................................................................................................................... viii

Chapter 1. INTRODUCTION ........................................................................................................................ 1
   1.1 Motivation and Problem Statement ................................................................................................. 1
   1.2 Thesis Contribution .......................................................................................................................... 8
   1.3 Thesis Outline ................................................................................................................................ 9

Chapter 2. LITERATURE REVIEW .......................................................................................................... 10
   2.1 Freeway Macroscopic Traffic Stream Models ............................................................................... 10
   2.2 Model Calibration .......................................................................................................................... 21

Chapter 3. STUDY SCOPE AND DATA DESCRIPTION ........................................................................... 25
   3.1 Introduction to the MIDAS database ............................................................................................. 25
   3.2 Descriptions of Study Sites ......................................................................................................... 27
   3.3 Description of Traffic Data ........................................................................................................... 30

Chapter 4. DATA PROCESSING AND MODEL CALIBRATION ............................................................... 35
   4.1 Data Manipulation and Filtration ................................................................................................. 35
   4.2 Model Calibration .......................................................................................................................... 41

Chapter 5. ANALYSIS OF MODEL CALIBRATION RESULTS ............................................................... 60
   5.1 Model Calibration Results of Detector 4737A ............................................................................ 62
   5.2 Model Calibration Results for Detector 4792B ......................................................................... 82

Chapter 6. CONCLUSIONS & OUTLOOK .............................................................................................. 100
   6.1 Conclusions ................................................................................................................................ 100
   6.2 Recommendations ......................................................................................................................... 100

Chapter 7. REFERENCES ............................................................................................................................ 102

APPENDICES .............................................................................................................................................. 108
APPENDIX A. SAS CODE FOR CALIBRATING MODEL PARAMETERS .......... 109
LIST OF TABLES

Table 1 - Types of Information in MIDAS Database ............................................................. 30
Table 2 - Criterions of Filtration Step I .................................................................................. 36
Table 3 - Parameter Settings of Three-Regime Model ........................................................... 44
Table 4 - Parameter Settings of Modified Dual-Regime Greenshields Model ....................... 44
Table 5 - Summary of Missing Dates from 1996-2010 .......................................................... 47
Table 6 - The Classification of 13-week Data Samples (Detector 4737A) ............................ 49
Table 7 - The Classification of 13-week Data Samples (Detector 4792B) ............................. 50
Table 8 - Variables by Report Type of Each Calibrated Model ............................................. 52
Table 9 - Mathematical Functions of Isolation Lines ............................................................. 57
Table 10 - Model-fitting Results and Related Statistics Based on 15-year Data .................... 60
Table 11 - GOF Statistics of Model Calibration Results Based on 15-year Data ................. 61
LIST OF FIGURES

Figure 1 - Trend of Congestion Growth in the U.S. (Schrank and Lomax, 2009) ............... 1
Figure 2 - The Constituent Structure of Computer-based Simulators .................................. 5
Figure 3 - Single-Regime Greenshields Speed-Density Relationship (Mathew, 2007) .......... 11
Figure 4 - Single-Regime Greenshields Flow-Density Relationship (Mathew, 2007)......... 12
Figure 5 - Single-Regime Greenshields Speed-Flow Relationship (Mathew, 2007)........... 12
Figure 6 - The Modified Single-regime Greenshields Model ............................................ 15
Figure 7 - The Modified Dual-regime Greenshields Model ................................................ 16
Figure 8 - Graphical Illustration of Transitional-Flow Regime ........................................ 17
Figure 9 - Illustration of three-phase traffic theory and phase transitions ....................... 19
Figure 10 - A Model Calibration Procedure (Chiu et al., 2009) ....................................... 22
Figure 11 - Illustration of Speed Influence Region (SIR) (Chiu et al., 2009) ...................... 23
Figure 12 - Graphical User Interface of MIDAS Online Database .................................... 25
Figure 13 - Geographical Illustration of RCCs in MIDAS Database ................................. 26
Figure 14 - Freeway Network in South East RCC and Station Locations ......................... 27
Figure 15 - Geographical Location of Detector 4737A ..................................................... 28
Figure 16 - Geographical Location of Detector 4792B ..................................................... 28
Figure 17 - Geometrics of Detector 4737A ...................................................................... 29
Figure 18 - Geometrics of Detector 4792B ...................................................................... 29
Figure 19 - Plot of Detector 4737A’s Raw 15-year Data in Speed-Density Diagram .......... 31
Figure 20 - Plot of Detector 4737A’s Raw 15-year Data in Speed-Flow Diagram ............. 32
Figure 21 - Plot of Detector 4737A’s Raw 15-year Data in Flow-Density Diagram .......... 32
Figure 22 - Plot of Detector 4792B’s Raw 15-year Data in Speed-Density Diagram......... 33
Figure 23 - Plot of Detector 4792B’s Raw 15-year Data in Speed-Flow Diagram .......... 33
Figure 24 - Plot of Detector 4792B’s Raw 15-year Data in Flow-Density Diagram ........ 34
Figure 25 - Filtration Step I in Speed-Density Diagram of 4737A’s 15-year Data .......... 37
Figure 26 - Filtration Step I in Speed-Density Diagram of 4792B’s 15-year Data .......... 37
Figure 27 - Data Manipulation, Filtration and Model Calibration Procedure ............... 39
Figure 28 - Illustrations of Three-regime Model’s Parameters .................................. 42
Figure 29 - Illustration of Minimum Uncongested Speed ......................................... 43
Figure 30 - Loop of Model Calibration Procedure ..................................................... 45
Figure 31 - # of Outliers of Detector 4737A’s Data Series ....................................... 48
Figure 32 - # of Outliers of Detector 4792B’s Data Series ....................................... 48
Figure 33 - Bar Plot of Table 4.5 .............................................................................. 51
Figure 34 - Bar Plot of Table 4.6 .............................................................................. 51
Figure 35 - Actual Data Points & Model Results in Speed-Density Diagram............ 53
Figure 36 - Actual Data Points & Model Results in Flow-Density Diagram ............ 53
Figure 37 - Actual Data Points & Model Results in Speed-Flow Diagram ............ 54
Figure 38 - Illustration of the Clustered Points in Speed-Density Diagram ............. 55
Figure 39 - Illustration of the Clustered Points in Speed-Flow Diagram ............. 56
Figure 40 - Illustration of the Clustered Points in Flow-Density Diagram .......... 57
Figure 41 - Illustration of Clustered Points Isolation ................................................. 58
Figure 42 - The Histogram of Clustered Points ............................................................ 59
Figure 65 - Model Estimates of Free-flow Speed .............................................................. 81
Figure 66 - Model Estimates of Vint2 .............................................................................. 82
Figure 67 - Model Estimates of Vint ................................................................................ 82
Figure 68 - Model Estimates of Kb .................................................................................. 83
Figure 69 - Model Estimates of Kint2 ............................................................................... 83
Figure 70 - Model Estimates of Kj ................................................................................... 84
Figure 71 - Model Estimates of $a_2$ ................................................................................ 85
Figure 72 - Model Estimates of $a$ ................................................................................... 85
Figure 73 - Plot of Minimum Uncongested Speed ............................................................ 85
Figure 74 - Correlation Analysis of Three-Regime Model’s Free-flow Speed .................... 86
Figure 75 - Correlation Analysis of Three-Regime Model’s Kb ......................................... 87
Figure 76 - Correlation Analysis of Three-Regime Model’s Vint ....................................... 88
Figure 77 - Correlation Analysis of Three-Regime Model’s Vint2 ................................... 89
Figure 78 - Correlation Analysis of Three-Regime Model’s Kint2 ..................................... 90
Figure 79 - Correlation Analysis of Three-Regime Model’s Kj ......................................... 91
Figure 80 - Correlation Analysis of Three-Regime Model’s $a$ ......................................... 92
Figure 81 - Correlation Analysis of Three-Regime Model’s $a_2$ ....................................... 93
Figure 82 - Correlation Analysis of Dual-regime Model’s Free-flow Speed ....................... 94
Figure 83 - Correlation Analysis of Dual-regime Model’s Kb .......................................... 95
Figure 84 - Correlation Analysis of Dual-regime Model’s Vint .......................................... 96
Figure 85 - Correlation Analysis of Dual-regime Model’s Kj ............................................ 97
Figure 86 - Correlation Analysis of Dual-regime Model’s $a$ ............................................ 98
1.1 Motivation and Problem Statement

Traffic congestion has grown drastically across the U.S. over the past decades, especially in the metropolitan areas (FHWA, 2005). The increasing congestion has resulted in considerably negative economic, environmental and social impacts. According to the 2009 Urban Mobility Report (Schrank and Lomax, 2009), traffic congestion caused urban Americans to travel 4.2 billion hours more in 2007, which resulted in an extra 2.8 billion gallons of gasoline consumption for the total cost of $87.2 billion.

![Bar Chart showing trend of congestion growth in the U.S.](image)

**Figure 1.1 Trend of Congestion Growth in the U.S.** (Source: Schrank and Lomax, 2009)

It has been widely recognized that the traffic congestion issue cannot be emolliated by merely building more infrastructures due to political, financial and environmental constraints. Therefore, Intelligent Transportation Systems (ITS), Advanced Traffic Management Systems (ATMS), and Advanced Traffic Information Systems (ATIS), which aim to make the best use of current transportation infrastructures to relief congestion with the advances in information technology, are increasingly drawing attention.
To evaluate the management strategies powered by the aforementioned systems without interrupting the real-world traffic, computer-based traffic simulation programs are conceptualized and developed. With the help of rapid technological advancements in computer science, the various state-of-the-art traffic simulators are capable of generating and evaluating large-scale, ITS-related planning scenarios and forecasting network operational performance under these scenarios.

The simulation-based Dynamic Traffic Assignment (DTA) model, which captures the dynamic nature of both travel demand and traffic flow, is a critical opponent of the computer traffic simulation programs. Rather than the conventional capacity-based Static Traffic Assignment model, which is only capable of roughly estimating the traffic impact of major infrastructure changes, the simulation-based DTA models can accommodate the sophisticated computational analysis of large-scale urban traffic network with heavy stochastic OD load by iterative computer simulations, including the evaluation of various management and control strategies (e.g. congestion pricing, ramp metering, HOT lanes, etc).

Depending on the level of details with which the model represents the transportation system, the simulation-based DTA models can be classified as three distinguished types, ranging from Microscopic, Mesoscopic to Macroscopic.

The microscopic models keep track of individual simulated vehicles with high level of detail. Each vehicle’s maneuver within every simulation step is derived from the driving behavior models, such as car-following, lane-changing and merging/diverging models, etc. Examples of microscopic models include VISSIM (PTV, 2005), AIMSUN/2 (Barcelo and Casas, 2003) and Paramics (Smith and Druitt, 1994), etc. The macroscopic models, on the
other hand, treat the aggregate traffic uniformly as a homogeneous flow without considering the interactions between its constituent individual vehicles. These models propagate the aggregate traffic flow throughout the network using low-fidelity analytical approaches. For example, the LWR model (Lighthill and Whitham, 1955) utilized sets of partial differential Equations analogous to one-dimensional Navier-Stokes wave Equations in hydrodynamics to represent traffic evolution over time and space.

The mesoscopic models, in particular, represent each individual vehicle with high level of detail, but model the vehicles’ activities and interactions with low level of detail. In fact, the mesoscopic simulation-based DTA models incorporate the advantages of both dynamic disaggregate traffic modeling and easy-to-calibrate macroscopic traffic stream relationships. They are capable of modeling disaggregate route-choice for each individual vehicle, which is essential in modeling en-route route choice and the impact of traffic information for drivers’ travel plans. Therefore, the mesoscopic simulation-based DTA models have been employed by several state-of-the-art mesoscopic traffic simulators, such as DynaMIT (Ben-Akiva et al., 1997), DTASQ (Mahut, 2001), DYNASMART (Mahmassani, 2001) and DynusT (Chiu et al., 2008).

The mesoscopic simulation-based DTA models have various approaches to simulate traffic. One approach is to aggregate vehicles into packets, and route the packets throughout the network (CONTRAM, Leonard et al., 1989). Each packet of vehicles is treated as an individual entity, sharing the behaviors defined by the macroscopic traffic stream models. Specifically, the speed of the vehicle packet is derived from the macroscopic speed-density relationship with the density value of the travel link at the packet’s entry moment. Another
approach is derived from queuing theory. In this approach, the roadway segments are modeled as two separate parts: the queuing part and the moving part. The operation of the moving part is almost identical with that of the first approach: vehicles travel through the moving part at a speed calculated from the macroscopic speed-density relationship. The queuing part, on the other hand, employs queue-servers to transfer vehicles from one segment to another, or form queues on the segments to represent congestion. The constituent relationships of the various models discussed above are illustrated in Figure 1.2.

Both of the aforementioned modeling approaches of mesoscopic simulation-based DTA models inevitably call for a deep understanding of macroscopic traffic stream models, which provide mathematical representations of the fundamental relationships of macroscopic traffic stream characteristics (flow, density and space-mean-speed) that researchers had observed from plots of realistic traffic data (A. D. May, 1991). Most important among them is the speed-density relationship. In light of this, the “macroscopic traffic stream model” in the remainder of thesis actually refers to the model of the speed-density relationship unless specified otherwise.
Figure 1.2 The Constituent Structure of Computer-based Simulators

A great deal of research has been done in investigating the speed-density relationships of general traffic streams. At the heart of all the research efforts is the essential proposition that density is the explanatory independent variable, and the behavior of aggregate traffic speed changes with respect to density. Deterministic mathematical model is then formulated to model the causality between speed and density. Given this model, researchers can derive the speed-flow and flow-density relationships by applying the fundamental state equation in traffic flow theory (Equation 1.1):

\[ q = k \times u \]  

where \( q \) is flow rate (number of vehicles crossing a point per unit of time), \( k \) is density (number of vehicles per unit length of roadway) and \( u \) is space-mean-speed (the average speed of a traffic stream computed as the length of roadway segment divided by the total time required to travel the segment).
Years’ research yielded an array of widely-acknowledged models. Some of the models are derived based on the assumption that a unique speed-density relationship is valid for the entire range of densities observed in traffic streams. Therefore, these models are called single-regime models, as exemplified by the classical Greenshields linear model, Greenberg’s logarithmic model; Underwood’s transposed exponential model and Edie’s discontinuous exponential model, etc.

While the above single-regime models might be adequate for representing the dominant trend in the speed-density relationship of arterial and highway traffic flow, their applications in modeling freeway traffic stream have been subject to scrutiny due to the fact that field observations of freeway traffic have shown different trends of relations at different range of densities (Edie, 1961). Specifically, researchers observed that a moderate increase in density of freeway traffic under the low-density regime would not significantly deteriorate the aggregative flow speed. In light of this, multi-regime models, which use separate equations to represent the speed-density relationship of freeway traffic flow at uncongested (low-density) and congested (high-density) traffic regimes, such as the two-regime models (e.g. modified dual-regime Greenshields model) and three-regime models, are proposed and developed.

The merits and dis-merits of the aforementioned macroscopic traffic stream models have long been debated by various literatures (e.g. Chen et al., 2004). Unfortunately, none of them so far has probed into the models’ temporal stability, i.e. the correlation between a given roadway segment’s traffic stream model and time. In fact, researchers’ interest of the traffic stream models’ temporal stability is naturally triggered by the research of the
aforementioned mesoscopic simulation-based DTA models. As discussed above, the macroscopic traffic stream model plays a key role in the mesoscopic simulation-based DTA models. Therefore, if the mesoscopic simulation-based DTA models are to be used to simulate and forecast traffic network’s future performance at the operational level, the first hurdle that researchers have to bypass is to correctly design the macroscopic traffic flow models associated with the desirable planning horizon.

Since there is no field data to directly generate and calibrate any of the aforementioned single-regime or multi-regime models, the only option available is to investigate the historical tracks of the given segment’s macroscopic traffic flow model to see if the models are independent from the time evolution. Specifically, researchers could collect and group historical traffic data of the given segment every fixed interval over a long time period, and fit an optimal model for each data group respectively. Then the correlations within this family of fitted models are quantitatively analyzed using statistical methods.

If analysis shows that the models are not statistically different from each other, then it indicates that the optimal macroscopic traffic flow model of the given segment is temporally stable. In light of this, researchers could derive a model using available historical data, and apply it into the simulator directly as the desirable future traffic flow model. If, on the other hand, significant variations are observed in the models’ comparison, then the variation structure will be studied using statistical methods so that researchers could “predict” the future model based on a credible current or historical model derived from available data sets.

In sum, a rigorous research in the temporal stability of macroscopic traffic stream models is well-motivated. There have been attempts to deal with similar research topics, such
as using time series models to solve the short-term traffic forecasting problem (Gazis, D. and C. Knapp, 1971; Ahmed, M. and A. Cook, 1982; Williams, B.M. et al., 1998; Lee, S. and D. Fambro, 1999; Williams and Hoel, 2003), but most of these attempts had focused on the variation pattern of the flow term only and barely studied the temporal periodicity of the macroscopic relationships between speed, flow and density. This has promoted the attempt in the thesis to better understand the temporal stability of the macroscopic freeway traffic stream models.

This thesis will make use of the Motorway Incident Detection and Automatic Signaling (MIDAS) database, which consistently records and stores freeway traffic data collected in London, United Kingdom. The data collection period of MIDAS database started from 1996. Therefore, the 15-year-long traffic data series provide a perfect platform for the thesis’ research.

1.2 Thesis Contribution

The main contribution of this thesis is a rigorous justification that basic freeway segments’ optimal macroscopic traffic stream models are temporally stable during the 15-year study period. In addition, SAS scripts are programmed to calibrate the optimal model parameters based on MIDAS data sets.

Although the analysis and conclusions in the research of this thesis are by no means specific to basic freeway segments, the methodologies are generally applicable to other types of freeway segments, such as the ones close to on/off-ramps and interchanges.
1.3 Thesis Outline

The remainder of the thesis is organized as follows. Chapter 2 presents a detailed review of
the freeway macroscopic traffic stream models and the calibration procedures. The
description of the MIDAS database and data collection locations are collectively summarized
in Chapter 3. The detailed data processing, model generation and optimization procedures are
presented in Chapter 4. Chapter 5 applies statistical method to analyze the temporal stability
of the models generated in Chapter 4. Finally, conclusions and future research direction are
discussed in Chapter 6.
Chapter 2 Literature Review

This chapter reviews the freeway macroscopic traffic stream models, including the modified dual-regime Greenshields model and three-regime model. The model calibration procedures are also discussed.

2.1 Freeway Macroscopic Traffic Stream Models

As discussed in Chapter 1, multi-regime models are typically used to model the freeway traffic stream, as particularly exemplified by the modified dual-regime Greenshields model and the three regime model.

2.1.1 The Modified Dual-Regime Greenshields Model

The modified Greenshields Model is firstly derived on the basis of the classical single-regime Greenshields model brought forward by Greenshields in 1935 (Greenshields, 1935). The original single-regime Greenshields model assumed a linear speed-density relationship as illustrated in Figure 2.1. The model equation of this linear relationship is as follows:

\[ v = v_f - \frac{v_f^2}{k_j} \times k \]  

(2.1)

where: \( v \) is speed, \( k \) is density, \( v_f \) is free-flow speed and \( k_j \) is jam density.
The flow-density and speed-flow relationships can be easily derived as Equation 2.2 and 2.3 respectively:

\[ q = v_f \cdot k - \left( \frac{v_f}{k_j} \right) \times k^2 \quad (2.2) \]

where \( q \) is flow, while \( k, v_f \) and \( k_j \) are similarly defined as in Equation 2.1.

\[ q = k_j \cdot v - \left[ \frac{k_j}{v_f} \right] \times v^2 \quad (2.3) \]

where \( q, v, v_f \) and \( k_j \) are similarly defined as in Equation 2.1 and 2.2.

The diagrams of the flow-density and speed-flow relationships are illustrated in Figure 2.2 and 2.3 respectively.
The classical single-regime Greenshields model explicitly characterized the general mathematical nature between traffic stream parameters. In reality, however, researchers can hardly find a linear relationship between speed and density from plots of empirical traffic data (Mathew, 2007). Hence the validity of this model had long been questioned (e.g. Chen et al., 2004).
In addition, Gazis et al. (1959, 1961) and May et al. (1967, 1968) discovered the connections between macroscopic traffic stream models and microscopic car-following models. Their findings demonstrated that the family of single-regime models could be generalized as Equation 2.4:

\[ v = v_f \times \left( 1 - \frac{k}{k_f} \right)^{\frac{1}{1-m}} \]  

(2.4)

where \( v, v_f, k \) and \( k_f \) are similarly defined as in Equation 2.1, 2.2 and 2.3, \( l \) and \( m \) are parameters associated with the corresponding microscopic car-following model (\( 0 \leq m < 1, l > 1 \)).

The original single-regime Greenshields model is equivalent with Equation 2.4 when \( l \) and \( m \) are both specified as 2. In fact, the modified single-regime Greenshields model (as expressed in Equation 2.5) also keeps \( l \) as 2 and only supersedes the term \( \frac{1}{1-m} \) by an individual power term \( \alpha \) by comparison with the original model version:

\[ v = v_f \times \left( 1 - \frac{k}{k_f} \right)^\alpha \]  

(2.5)

where \( v, v_f, k \) and \( k_f \) are similarly defined as in Equation 2.4.

The extra degree of freedom (power term \( \alpha \)) in Equation 2.5 introduced a curvilinear feature into the Greenshields model and greatly improved its performance in terms of reproducing empirical traffic conditions in the context of arterial and highway traffic.

In order to correctly model freeway traffic, the modified dual-regime Greenshields model is developed based on Equation 2.5 while incorporating the proposition that the aggregative speed of freeway traffic under free-flow regime is constant regardless of density,
since freeways provide more capacity than arterials and can accommodate very dense traffic at near free-flow speeds.

The general form of the modified dual-regime Greenshields model is expressed as Equation 2.6:

\[
\begin{cases}
  v = u_f, & 0 \leq k \leq k_b \\
  v = v_f \times (1 - \frac{k}{k_j})^\alpha, & k_b \leq k \leq k_j
\end{cases}
\]

(2.6)

where \( u_f \) is free-flow speed, \( v_f \) is speed intercept, \( k_b \) is break-point density. \( v, k, k_j \) and \( \alpha \) are similarly defined as in Equation 2.5.

Additional tune-ups have been introduced to Equation 2.5 and 2.6 at researchers’ discretion in order to meet practical requirements. For example, Mahmassani et al., (2004) and Chiu et al., (2008) created an extra minimum speed artifact \( v_0 \) in both the single-regime and dual-regime modified Greenshields models when implementing these models into the mesoscopic DTA-based simulators DYNASMArt and Dynus-T, respectively. Their propositions are expressed in Equation 2.7 (single-regime) and 2.8 (dual-regime):

\[
v - v_0 = (v_f - v_0) \times (1 - \frac{k}{k_j})^\alpha
\]

(2.7)

where all other parameters are similarly defined as in Equation 2.6.

Figure 2.4 depicts the illustration of Equation 2.7:
Figure 2.4 The Modified Single-regime Greenshields Model with Minimum Speed Artifact

(Source: Mahmassani et al., 2009)

\[
\begin{align*}
&v = u_f, \quad 0 \leq k \leq k_b \\
&v - v_0 = (v_f - v_0) \times \left(1 - \frac{k}{k_j}\right)^\alpha, \quad k_b \leq k \leq k_j
\end{align*}
\]  \hspace{1cm} (2.8)

where all parameters are similarly defined as in Equation 2.6.

Figure 2.5 depicts the illustration of Equation 2.8:
Figure 2.5 The Modified Dual-regime Greenshields Model with Minimum Speed Artifact

(Source: Mahmassani et al., 2009)

Other than including the minimum speed artifact, Ben-Akiva et al. (1997) suggested keep the “l-1” term of Equation 2.4 as an individual parameter as well and denoted this term as β instead of keeping it a fixed value. His proposition is expressed in Equation 2.9:

\[
\begin{align*}
  v &= u_f, \\
  v - v_0 &= (v_f - v_0) \times \left(1 - \left(\frac{k}{k_f}\right)^\beta\right)^\alpha, \\
  k_b &\leq k \leq k_f
\end{align*}
\]

(2.9)

where all parameters are similarly defined as in Equation 2.6.

Equation 2.9 is implemented in another mesoscopic DTA-based simulator DynaMIT.

2.1.2 The Three-Regime Model

Drake et al. (1967) suggested the first three-regime model with the proposition of an additional transitional-flow regime on the basis of the regular free-flow regime and
congested-flow regime, all of which are represented by linear Greenshields formulations (Equation 2.10):

\[
u = \begin{cases} 
50 - 0.098k & \text{for } k \leq 40 \\
81.4 - 0.913k & \text{for } 40 \leq k \leq 65 \\
40 - 0.265k & \text{for } k \geq 65
\end{cases}
\]  

(2.10)

Figure 2.6 depicts the graphical illustration of the transitional-flow regime (circled area) in the context of a speed-flow plot of field data:

![Graphical Illustration of Transitional-Flow Regime](image)

Figure 2.6 Graphical Illustration of Transitional-Flow Regime

The transitional-flow regime has been defined as a stochastic transition from the uncongested traffic state to the congested state in recent research advancements (e.g. Mahnke, 2004). The traffic behavior under this regime is claimed to be highly stochastic and unstable, which should be distinguished from that of other regimes and modeled separately using appropriate measures. Research has also shown, however, that conventional deterministic models that are statistically sound lack the capability of fully reproducing the
stochastic nature of traffic flow under transitional-flow regime, while the outputs of probabilistic approaches are normally analytical procedures rather than parametric mathematics functions.

Kerner (2009) described three phases of traffic from the perspective of interpreting the physics of traffic breakdown and consequent traffic congestion. The three phases are Free flow (F), Synchronized flow (S) and Wide moving jam (J) as illustrated in Figure 2.7. Figure interpretation: “(a) Qualitative presentation of free flow states $F$, a 2D-region of steady states of synchronized flow $S$, and the line $J$ in the flow–density plane. (b) Qualitative explanation of the speed adaptation effect. (c) Qualitative explanation of a competition between speed adaptation and over-acceleration in the space-gap–speed plane. (d) Qualitative Z-shaped function of probabilities of passing and over-acceleration on density. (e) Qualitative shape of the Z-characteristic for traffic breakdown in the speed–flow plane. (f) Qualitative shape of the double Z-characteristic for $F \rightarrow S \rightarrow J$ transitions in the speed–density plane. In (b–f), dashed regions $S$ are parts of the 2D-region for synchronized flow in (a)” (Source: Kerner, 2009)
Figure 2.7 Illustration of three-phase traffic theory and phase transitions (Source: Kerner, 2009)
Although Kerner’s three-phase traffic theory does not exactly coincide with the conventional perspective of three-regime models, his rigorous analysis has greatly solidified the theoretical foundation of decomposing the freeway traffic stream into three distinct models.

2.1.3 Determination of Break-Point Density

A major difficulty in developing multi-regime models is on how to determine break-point density between regimes. Segmentation of uncongested and congested regimes has been frequently performed exogenously based on the best guess of model developers (Kockelman, 2001). Such ad-hoc treatments greatly rely on modelers’ subjective judgments and therefore lack theoretical substantiations.

Maddala (1983) and Bhat (1997) proposed a binary logit-type segmentation model (Equation 2.11) used to probabilistically estimate observations’ memberships in the two regimes, and an iterative search (Equation 2.12) was conducted for the likelihood-maximizing estimators.

\[
\Pr(\text{nth observation } \in \text{congested regime}) = \frac{e^{x'\beta}}{1 + e^{x'\beta}}
\]  

(2.11)

where \( \beta \) is model parameter estimated from empirical data sets.

\[
Lik_n = \Pr(\text{Flow}_n = q_n)
\]

\[
= \Pr(\text{Flow}_n = q_n \mid \text{Uncong.}) \Pr(\text{Uncong.}) + \Pr(\text{Flow}_n = q_n \mid \text{Cong.}) \Pr(\text{Cong.})
\]

(2.12)

\[
= \varphi \left[ \frac{q_n - \sum v_{free,i} p_i k}{\sigma_u} \right] \frac{1}{1 + e^{x'\beta}} + \varphi \left[ \frac{q_n - \sum p_i a_i + \sum p_i b_i v}{\sigma_c} \right] \frac{e^{x'\beta}}{1 + e^{x'\beta}}
\]

where \( a_i, b_i \) and \( v_{free,i} \) are model parameters estimated from empirical data sets.
Sun and Zhou (2005) used $K$-means clustering to divide traffic flow into several regimes, and then applied conventional regression analysis to develop traffic stream models. The benefit of cluster analysis is that traffic data with highest similarity patterns are assigned to the same cluster. After clustering, data in one cluster is more homogeneous and the degree of similarity is relatively high. As a result, the cluster analysis provides an endogenous way of partitioning traffic stream data into multiple regimes.

### 2.2 Model Calibration

Generally, the calibration of macroscopic traffic stream model is defined as the adjustment of model parameters so that the model could more accurately mimic field conditions (Park et al., 2005). Since the macroscopic traffic stream model is a critical component of the mesoscopic DTA-based simulators, the simulators’ credibility would fall short if the traffic stream model’s parameters are not properly tuned based on field traffic data.

Chiu et al. (2009) calibrated Equation 2.8 and 2.9 respectively using traffic data from Next Generation SIMulation (NGSIM) program, and compared the models’ calibration results. Figure 2.8 depicted the calibration procedure:
Figure 2.8 A Model Calibration Procedure (Source Chiu et al., 2009)

As depicted in Figure 2.8, the concept of Speed Influence Region (SIR) is defined and applied in the derivation of density value to be used in the speed-density function. The calculated and observed speed values constituted the residual column matrix:

\[
X = [(v_{1}^{cal} - v_{1}^{obs}), ..., (v_{i}^{cal} - v_{i}^{obs}), ...]^{T}
\]

(2.10)

where \(v_{i}^{cal}\) is the speed calculated according to the modified dual-regime Greenshields model (M1 or M2) and \(v_{i}^{obs}\) is the \(i\)th observed speed in NGSIM dataset.
The objective of optimizing model parameters is achieved through minimizing the sum of squares of the deviations between the observed and calculated speeds for all data points, i.e. solving the least-square problem in Equation 2.11:

$$\min_{\alpha, \beta, \frac{1}{2}X^TX} f = \frac{1}{2}X^TX$$

subject to $\alpha > 0; \beta > 0; v_f > 0; 0 < k_b < k_j$

Many nonlinear optimization algorithms, such as Newton’s Method, Gauss – Newton’s Method and Levenberg – Marquadt Method are available to solve Equation 2.11.

Figure 2.9 Illustration of Speed Influence Region (SIR) [Source Chiu et al., 2009]

Mahmassani et al., (2009) specifically discussed the calibration procedure of the M1 model. The speed-density function’s curvilinear congested-regime part (as shown in
Equation 2.8) is transformed into a linear form (Equation 2.12) by taking logarithm on both sides of the function:

\[
\ln(v - v_0) = \ln(v_f - v_0) + \alpha \ln \left(1 - \frac{k}{k_j}\right), \quad k_b \leq k \leq k_j \tag{2.12}
\]

Equation 2.12 is a typical form of “\(Y = aX + b\)”. Therefore, a simple linear regression analysis would be more than sufficient to optimize the relevant model parameters. The MOEs of the calibration procedure are the goodness-of-fit R square value and the root mean square error for speed.
3.1 Introduction to the MIDAS database

The freeway traffic data to be used in this thesis is collected from the MIDAS database (Figure 3.1), which is operated and maintained by the Mott Macdonald’s company located in London, United Kingdom. The MIDAS database has been recording and storing traffic data of the entire freeway system in London using paired inductive loop detectors since 1996. Figure 3.2 depicts the locations and index of the MIDAS database’s Regional Control Centre (RCC) areas. The scope of the research in this thesis focuses on outer loop of South East RCC of M25 motorway.

![Image of MIDAS Online Database](image)

**Figure 3.1 Graphical User Interface of MIDAS Online Database**
Figure 3.2 Geographical Illustration of RCCs in MIDAS Database
3.2 Descriptions of Study Sites

The study sites to be used in the research of this thesis are chosen among the regular mainline freeway segments whose traffic flow characteristics are not impacted by the turbulent lane-changing and merging/diverging behaviors due to on/off ramps and interchanges.

Site 1 is a four-lane freeway segment associated with detector # 4737A. This segment is located half way between the interchange of Portsmouth RD/M25 and Samway RD, west bound.

Site 2 is also a four-lane freeway segment. Its corresponding detector is # 4792B. This site is located between Liberty Lane and Spinney Hill, east bound. The distance between the two locations is 5.5 kilometers.

The two sites are labeled in the sketchy freeway network within SE RCC as depicted in Figure 3.3. Figure 3.4-3.7 are aerial photos taken from Google Maps to demonstrate the geographical locations and geometrics of the study sites.

Figure 3.3 Freeway Network in South East RCC and Station Locations
Figure 3.4 Geographical Location of Detector 4737A

Figure 3.5 Geographical Location of Detector 4792B
Figure 3.6 Geometrics of Detector 4737A

Figure 3.7 Geometrics of Detector 4792B
### 3.3 Description of Traffic Data

Table 3.1 lists the types of information (of the four-lane study sites) contained in the raw MIDAS traffic data sets:

**Table 3.1 Types of Information in MIDAS Database**

<table>
<thead>
<tr>
<th>#</th>
<th>Variable</th>
<th>Type</th>
<th>Len</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TIMESTAMP</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>TR_ADDRESS</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>LCC_ADDRESS</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>CO_ADDRESS</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>DEV_ADDRESS</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>LANE1_FLOW</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>13</td>
<td>LANE1_HEADWAY</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>12</td>
<td>LANE1_OCCUPANCY</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>LANE1_SPEED</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>15</td>
<td>LANE2_FLOW</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>17</td>
<td>LANE2_HEADWAY</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>16</td>
<td>LANE2_OCCUPANCY</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>14</td>
<td>LANE2_SPEED</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>19</td>
<td>LANE3_FLOW</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>21</td>
<td>LANE3_HEADWAY</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>20</td>
<td>LANE3_OCCUPANCY</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>18</td>
<td>LANE3_SPEED</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>23</td>
<td>LANE4_FLOW</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>25</td>
<td>LANE4_HEADWAY</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>24</td>
<td>LANE4_OCCUPANCY</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>22</td>
<td>LANE4_SPEED</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>CATGORY1_FLOW</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>CATGORY2_FLOW</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>CATGORY3_FLOW</td>
<td>Num</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>CATGORY4_FLOW</td>
<td>Num</td>
<td>8</td>
</tr>
</tbody>
</table>
The raw MIDAS data sets are averaged over discrete 1-minute intervals per lane per direction. They are suppressed on a daily basis into .tcd format files. Researchers could track and extract specific freeway segments’ raw traffic data by specifying the site information, including Control Office (CO), Local Control Center (LCC), Transponder Address (TR) and Device Address (Dev). Shekhar (2002) originally programmed SAS scripts to perform the above tracking and data extraction tasks. Williams (2004) modified Shekhar’s code to embrace a minor format change in the raw data sets. Both of their contributions are used in this thesis as the foundation to obtain the desirable traffic data series.

The available MIDAS traffic data starts from January 1st, 1996 up to May 31st, 2010. Figure 3.8 - 3.13 plot the 15-year raw data points:

![Figure 3.8 Plot of Detector 4737A’s Raw 15-year Data in Speed-Density Diagram](image-url)
Figure 3.9 Plot of Detector 4737A’s Raw 15-year Data in Speed-Flow Diagram

Figure 3.10 Plot of Detector 4737A’s Raw 15-year Data in Flow-Density Diagram
Figure 3.11 Plot of Detector 4792B’s Raw 15-year Data in Speed-Density Diagram

Figure 3.12 Plot of Detector 4792B’s Raw 15-year Data in Speed-Flow Diagram
Figure 3.12 Plot of Detector 4792B’s Raw 15-year Data in Flow-Density Diagram
4.1 Data Manipulation and Filtration

The raw MIDAS traffic data sets collected by detectors need to go through a data manipulation and filtration procedure before being used to calibrate the desirable models. There are several reasons to substantiate this proposition:

- A large number of data are missing from the raw data sets due to unknown reasons, and the SAS scripts used “.” to fill in the missing numeric values when reading the raw data sets into the program. Since the raw traffic data are the foundation of further calculation and analysis, the “.” marks should be replaced by a reasonable substitution accordingly.

- The raw traffic data is aggregated over discrete 1 minute intervals per lane per direction. To be consistent with the conventional perspective of traffic engineering, the 15-minute aggregative data averaged across all lanes are desired.

- After the 15-minute data sets are correctly assembled, part of the data points should be filtered out before feeding the data sets into the model calibration procedure because these data points are outliers. The filtration mechanism is specifically designed to rule out scattered outliers to make sure that the calibrated model would fit the dominant trend of the traffic data to the highest degree. This mechanism is carried out in two steps: Step I loosely filters out the visible outliers by applying filters as summarized in Table 4.1 and illustrated in both detector 4737A and 4792B’s speed-density plots of the 15-year raw data series.(Figure 4.1
A preliminary three-regime model with modified Greenshields formulations (specification of the model will be introduced in the follow-ups) is then calibrated based on the refined data sets. According to the outputs of calibrated model, Step II further refines the filtered data sets by ruling out the data points that are visibly implicit but still significantly different from model estimates in statistical sense (i.e. $\frac{|Actual \ Speed - Model \ Estimated \ Speed|}{Model \ Standard \ Error} \geq 3$).

Table 4.1 Criterions of Filtration Step I

<table>
<thead>
<tr>
<th>Filtration Criterion #</th>
<th>Speed</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt; 100</td>
<td>&lt; 10</td>
</tr>
<tr>
<td>2</td>
<td>&lt; 80</td>
<td>&lt; 15</td>
</tr>
<tr>
<td>3</td>
<td>&lt; 40</td>
<td>&lt; 28</td>
</tr>
<tr>
<td>4</td>
<td>&lt; 20</td>
<td>&lt; 40</td>
</tr>
<tr>
<td>5</td>
<td>NA</td>
<td>=0</td>
</tr>
</tbody>
</table>
Figure 4.1 Filtration Step I in Speed-Density Diagram of 4737A’s 15-year Data

Figure 4.2 Filtration Step I in Speed-Density Diagram of 4792B’s 15-year Data
Therefore, a well-organized data manipulation, filtration and model calibration procedure is essential to the quality and credibility of the ultimate model outputs. Figure 4.3 presents the flowchart of the procedure accepted in this thesis.

Figure 4.3 Data Manipulation, Filtration and Model Calibration Procedure

- Find representative freeway segment from map
- Retrieve the CO LCC TR DEV info of the segment
- Extract daily raw data sets of the segment

**Flowchart Explanation:**

1. **Find representative freeway segment from map**
2. **Retrieve the CO LCC TR DEV info of the segment**
3. **Extract daily raw data sets of the segment**

**Flow Calculation:**

\[
\text{Flow}_t = \text{Lane}_1\_\text{flow} + \text{Lane}_2\_\text{flow} + \ldots
\]

\[
\text{Speed}_{1\ min} = \frac{\text{Flow}_t}{(\text{Lane}_1\_\text{flow} \times \text{Lane}_1\_\text{pace} + \text{Lane}_2\_\text{flow} \times \text{Lane}_2\_\text{pace} + \ldots)}
\]

\[
\text{Flow}_{1\ min} = \frac{((\text{Cat}_1\_\text{flow} + 1.5 \times (\text{Cat}_2\_\text{flow} + \text{Cat}_3\_\text{flow} + \text{Cat}_4\_\text{flow})) \times 60)}{(\text{# of lanes})}
\]
If Speed_1 min = 0?

- Yes: Let Density_1 min = 0
- No: Density_1 min = Flow_1 min / Speed_1 min

Mathematically Average the 1-min data over every 15-min

Obtain the aggregative Flow_15 min data and Density_15 min data

Speed_15 min = Flow_15 min / Density_15 min

Gather the 15 min data within the proposed time interval

Filtration Step I

Fit a preliminary three-regime model

Apply Filtration Step II based on the above model fitting results

Re-fit the three-regime model

Fit the dual-regime model
4.2 Model Calibration

4.2.1 Model Specifications

The research in this thesis proposes to calibrate a modified dual-regime Greenshields model and a three-regime model with modified Greenshields’ formulations for all three regimes. The models’ specifications are summarized in Equation 4.1 and 4.2, respectively:

\[
\begin{align*}
    v &= u_f = v_f \times \left(1 - \frac{k_b}{k_j}\right)^{\alpha}, \quad 0 \leq k \leq k_b \\
    v &= v_f \times \left(1 - \frac{k}{k_j}\right)^{\alpha}, \quad k_b \leq k \leq k_j
\end{align*}
\] (4.1)

where all parameters are similarly defined as in Equation 2.6.

\[
\begin{align*}
    v &= u_f = v_{f2} \times \left(1 - \frac{k_b}{k_{int2}}\right)^{\alpha_2}, \quad 0 \leq k \leq k_b \text{ and } v_a \geq v_{muc} \\
    v &= v_{f2} \times \left(1 - \frac{k}{k_{int2}}\right)^{\alpha_2}, \quad k_b \leq k \text{ and } v_a \geq v_{muc} \quad (4.2) \\
    v &= v_f \times \left(1 - \frac{k}{k_j}\right)^{\alpha}, \quad v_a < v_{muc}
\end{align*}
\]

where \(v_a\) is actual speed and \(v_{muc}\) is minimum uncongested speed (defined later). \(v_{f2}\) and \(k_{int2}\) are illustrated in Figure 4.4. The rest of the variables are similarly defined as in Equation 4.1. Note that the jam density \(k_j\) is defined to be the density value at the intercept of the extension of the fitted 3rd regime model and the horizontal density axis.
The minimum uncongested speed \( v_{muc} \) is defined to be the minimum speed of the data that meet the criterion in Equation 4.3 as follows:

\[
q \geq q_{0.95} \& v \geq 80 \text{ (mi/hr)}
\]  

(4.3)

where \( q_{0.95} \) is the 95 percentile flow rate.
4.2.2 Parameter Optimization Settings

The research in this thesis mainly applies SAS command “PROC MODEL” to execute the task of model calibration. The optimal values of model parameters are generated based on iterative optimization processes. The initial values of associated parameters and the corresponding optimization bounds are summarized in Table 4.2 and 4.3. The maximum number of iterations for the optimization process is 500.

Table 4.2 Parameter Settings of Three-Regime Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial Value</th>
<th>Optimization Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>(v_f)</td>
<td>135</td>
<td>[80,200]</td>
</tr>
</tbody>
</table>
Table 4.2 Continued

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial Value</th>
<th>Optimization Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{int2}$</td>
<td>65</td>
<td>[22,200]</td>
</tr>
<tr>
<td>$k_b$</td>
<td>12</td>
<td>[5,22]</td>
</tr>
<tr>
<td>$v_f$</td>
<td>175</td>
<td>[80,300]</td>
</tr>
<tr>
<td>$k_j$</td>
<td>110</td>
<td>[90,130]</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>6</td>
<td>$\geq 0$</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.9</td>
<td>$\geq 0$</td>
</tr>
</tbody>
</table>

Table 4.3 Parameter Settings of Modified Dual-Regime Greenshields Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial Value</th>
<th>Optimization Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_f$</td>
<td>160</td>
<td>[80,300]</td>
</tr>
<tr>
<td>$k_b$</td>
<td>13</td>
<td>[9,25]</td>
</tr>
<tr>
<td>$k_j$</td>
<td>110</td>
<td>[90,130]</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>2</td>
<td>$&gt;0$</td>
</tr>
</tbody>
</table>

The parameters’ optimization bounds are loosely set up (except for the bounds of the Jam Density term ($K_j$)) so that the PROC MODEL procedure will be free to systematically
adjust the value of each parameter to minimize the Sum of Squared Errors (SSE) related to model fitting. The bounds of \( K_j \), however, are basically driven by the commonsense of engineering experience and results of previous research. For example, A. D. May (1991) pointed out that the jam density usually lies in the range of 185-250 pc/mi/lane, which is approximately equivalent to 115-155 pc/km/lane. Meanwhile, the far end of the speed-density plot of the detectors’ 15-year raw data visually indicates that the actual jam density herein is likely to fall in a lower range by comparison with normal. Considering both factors, the optimization bound of jam density in the context of this research is proposed to be 90-130 (pc/km/lane).

4.2.3 Model Calibration Settings and Outputs

In order to fully capture the nature of macroscopic traffic stream models’ temporal stability, the model calibration procedure used in this thesis is delicately designed to loop through the entire 15-year filtered data series using 13-week (1/4 year) samples that are iterating in one day intervals, starting from January 1\(^{st}\), 1996. For example, data points belonging to the very first 13 weeks (i.e. from Jan 1\(^{st}\), 1996 to April 1\(^{st}\), 1996) yield the first group of calibrated models. Similarly, data sets of the next 13 weeks with one day lag (i.e. from Jan 2\(^{nd}\), 1996 to April 2\(^{nd}\), 1996) generate the second group of model parameters’ estimates, and so on. Figure 4.6 illustrated the loop of model calibration procedure as discussed above:
Figure 4.6 Loop of Model Calibration Procedure

Therefore, the model calibration loop yields in total 5175 groups of models based on the available 15-year data sets provided by each detector.

Quite a few days (in total 87 days out of the entire 15 years) of data are missing, possibly due to device malfunctions or system upgrades. The specific missing dates of each year are summarized in Table 4.4. Considering this issue, the SAS scripts programmed to perform the above calibration loop are designed to be robust enough to bypass the possible hurdles. Specifically, suppose the data of day A are missing from one of the 13-week samples. When the loop of model calibration procedure reaches this incomplete data sample, the program will automatically ignore day A and skip to the next consecutive available day. In other words, the SAS script is designed to use all available data within each 13-week time window and simply ignores gaps between traffic condition data observations.

The research in this thesis will also calibrate each of the proposed models based on the entire 15-year data sets provided by each detector. The reason is to compare the calibrated model parameters based on 13-week data samples and the 15-year time window to better sense the model parameters’ temporal fluctuation pattern.
Table 4.4 Summary of Missing Dates from 1996-2010

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>18-Jan</td>
<td>2-Jan</td>
<td>20-Jun</td>
<td>16-Aug</td>
<td>28-Aug</td>
<td>16-Jun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>19-Jan</td>
<td>3-Jan</td>
<td>17-Dec</td>
<td>17-Aug</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20-Jan</td>
<td>4-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>21-Jan</td>
<td>5-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>14-Feb</td>
<td>6-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1-Dec</td>
<td>7-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2-Dec</td>
<td>8-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3-Dec</td>
<td>9-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4-Dec</td>
<td>10-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>5-Dec</td>
<td>11-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>6-Dec</td>
<td>12-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>7-Dec</td>
<td>13-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>8-Dec</td>
<td>14-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>9-Dec</td>
<td>15-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>10-Dec</td>
<td>16-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>11-Dec</td>
<td>17-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>12-Dec</td>
<td>18-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>13-Dec</td>
<td>19-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>14-Dec</td>
<td>20-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>15-Dec</td>
<td>21-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>16-Dec</td>
<td>22-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>17-Dec</td>
<td>23-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>18-Dec</td>
<td>24-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>19-Dec</td>
<td>25-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>20-Dec</td>
<td>26-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>21-Dec</td>
<td>27-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>22-Dec</td>
<td>28-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>23-Dec</td>
<td>29-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>24-Dec</td>
<td>30-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>25-Dec</td>
<td>31-Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>26-Dec</td>
<td>12-Nov</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>27-Dec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>28-Dec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>29-Dec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>30-Dec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>31-Dec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Each complete 13-week aggregation interval originally contains \(8736 \times (60 \div 15) \times 24 \times (13 \times 7)\) data points. However, not all data points are used for calibrating the desirable traffic stream models because of the application of filtration procedures discussed previously. The number of filtered outliers corresponding to each data aggregation interval will be reported in the outputs of the model calibration procedure. Figure 4.7 and Figure 4.8 plot the number of outliers out of each 13-week sample for each detector’s data series.
In fact, the number of outliers is a key criterion in assessing the calibrated models’ validity. The reason is that if the majority of the original 8832 observations of a certain 13-week sample are outliers, then there may not be enough righteous data points to credibly
substantiate any of the proposed models. Therefore, the report of outliers is an essential constituent of the model analysis and comparison in Chapter 5. For example, the red circles in Figure 4.7 and 4.8 indicate that abnormally many data points have been filtered out from each of the corresponding 13-week samples and the credibility of the associated model estimates fall short accordingly.

Table 4.5 and 4.6 summarize the # of 13-week samples based on the percentage of acceptable data points within each 13-week sample for detector 4737A and 4792B, respectively.

Table 4.5 The Classification of 13-week Data Samples (Detector 4737A)

<table>
<thead>
<tr>
<th>Percent Good</th>
<th># of Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10%</td>
<td>23</td>
</tr>
<tr>
<td>10%-20%</td>
<td>20</td>
</tr>
<tr>
<td>20%-30%</td>
<td>22</td>
</tr>
<tr>
<td>30%-40%</td>
<td>111</td>
</tr>
<tr>
<td>40%-50%</td>
<td>56</td>
</tr>
<tr>
<td>50%-60%</td>
<td>56</td>
</tr>
<tr>
<td>60%-70%</td>
<td>58</td>
</tr>
<tr>
<td>70%-80%</td>
<td>165</td>
</tr>
<tr>
<td>80%-90%</td>
<td>374</td>
</tr>
<tr>
<td>90%-100%</td>
<td>4290</td>
</tr>
</tbody>
</table>
Table 4.6 The Classification of 13-week Data Samples (Detector 4792B)

<table>
<thead>
<tr>
<th>Percent Good</th>
<th># of Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10%</td>
<td>319</td>
</tr>
<tr>
<td>10%-20%</td>
<td>37</td>
</tr>
<tr>
<td>20%-30%</td>
<td>40</td>
</tr>
<tr>
<td>30%-40%</td>
<td>36</td>
</tr>
<tr>
<td>40%-50%</td>
<td>39</td>
</tr>
<tr>
<td>50%-60%</td>
<td>37</td>
</tr>
<tr>
<td>60%-70%</td>
<td>38</td>
</tr>
<tr>
<td>70%-80%</td>
<td>129</td>
</tr>
<tr>
<td>80%-90%</td>
<td>319</td>
</tr>
<tr>
<td>90%-100%</td>
<td>4181</td>
</tr>
</tbody>
</table>

Table 4.5 and 4.6 are easy to interpret. For example, the 1st row of Table 4.5 (in the blue square) means that there are 23 13-week data samples whose acceptable data points are less than 10%. Similarly, the 2nd row means that the number of 13-week samples whose acceptable data points are between 10% and 20% is 20.

Figure 4.9 and 4.10 explicitly demonstrate Table 4.5 and 4.6 respectively in the form of bar plot.
Figure 4.9 Bar Plot of Table 4.5

Figure 4.10 Bar Plot of Table 4.6
As discussed previously, if the majority of any 13-week sample’s data points are outliers and filtered out, then the model fitting results based on the rest of the data points within this sample may not be credible. Therefore, in the context of this research, only the data samples whose percentage of outliers is less than 30% will be used for estimating model parameters and further analysis. Based on this criterion, detector 4737A’s 15-year data series provide 4829 eligible 13-week data samples (out of 5175 in total) while detector 4792B yield 4629 samples.

Other than the reports of outliers’ information, the model calibration procedure also creates informative reports about model parameter estimates, goodness-of-fit statistics and model-fitted speed results of actual density observations.

Table 4.7 summarizes the list of available variables by report type of each calibrated model.

<table>
<thead>
<tr>
<th>Report Type</th>
<th>Variable</th>
<th>Modified dual-regime Greenshields model</th>
<th>Three-regime model after Step I filtration</th>
<th>Three-regime model after Step II filtration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter Estimate</td>
<td>FFS</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>KJ</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>KJ SE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>KI</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>KI SE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Kint2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Kint2 SE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Vt</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Vt SE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Vf</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Vf SE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>α</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>α SE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>α2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>α2 SE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>C adjR2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>C_adj2R</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Goodness-of-fit Statistic</td>
<td>C_R2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>C_RMSE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>C_RMSE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>UC AdjR2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>UC_R2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>UC_RMSE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>UC_RMSE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>all AdjR2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>all R2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>all RMSE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>all RMSE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

51
As an illustration, Figure 4.11 – 4.13 plot the model-fitting results and actual data observations corresponding to detector 4737A (time interval reflected: 4\textsuperscript{th} quarter, 2007):

Figure 4.11 Actual Data Points & Model Results in Speed-Density Diagram

Figure 4.12 Actual Data Points & Model Results in Flow-Density Diagram
Figure 4.13 Actual Data Points & Model Results in Speed-Flow Diagram

The direct comparison between actual data sets and model-fitting results as illustrated in Figure 4.11-4.13 explicitly demonstrate the similarity between optimal macroscopic traffic stream models and field data. Researchers could also make preliminary judgments between the two fitted models in terms of reproducing actual traffic streams to the highest degree.

4.2.4 Analysis of Abnormally Clustered Points of Detector 4792B

From Figure 4.2, researchers could observe an abnormal cluster of data points, as illustrated in Figure 4.14 – 4.16 (roughly within the blue circle). Clearly, these clustered points are driven away by certain reasons from the dominant trend line.

In order to analyze the reason that resulted in the clustered data points, the research in this thesis proposes to roughly isolate these points as illustrated in Figure 4.17. The mathematical functions of the isolation lines are summarized in Table 4.8.
A histogram plot (Figure 4.18) is then formulated to demonstrate the temporal distribution of the isolated data points. It is quite obvious that the majority of the abnormally clustered data points appeared in 1) most of 2001, 2) about 2/3 of 2002 and 3) approximately June & July of 2006, which indicate that certain events related with freeway traffic must have occurred.

Figure 4.14 Illustration of the Clustered Points in Speed-Density Diagram
Figure 4.15 Illustration of the Clustered Points in Speed-Flow Diagram

Figure 4.16 Illustration of the Clustered Points in Flow-Density Diagram
Figure 4.17 Illustration of Clustered Points Isolation

Table 4.8 Mathematical Functions of Isolation Lines

<table>
<thead>
<tr>
<th># of Isolation Line</th>
<th>Mathematical Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( u=40 )</td>
</tr>
<tr>
<td>2</td>
<td>( u=70 )</td>
</tr>
<tr>
<td>3</td>
<td>( u=112-2.5*k )</td>
</tr>
<tr>
<td>4</td>
<td>( u=128-2.5*k )</td>
</tr>
</tbody>
</table>
Figure 4.18 The Histogram of Clustered Points
Chapter 5 Analysis of Model Calibration Results

The model calibration procedure programmed by SAS is applied for the data series collected from detector 4737A and 4792B, respectively. This chapter mainly analyzes the macroscopic traffic stream models’ temporal stability based on the model calibration results.

As discussed previously, each 13-week data sample will yield a set of calibrated model parameters. Considering the factor of # of outliers within each 13-week sample, the 15-year data series of detector 4737A generate 4829 sets of models, and detector 4792B yields 4629 sets. Each proposed model prototype is also calibrated based on the entire 15-year data set other than going through the loop of calibration procedure, and the fitted results of model parameters and related statistics are summarized in Table 5.1 (3RM: the estimates of three-regime model based on the data sets after filtration step1 & 2; 2RMGS: the estimates of modified dual-regime Greenshields model based on the data sets after filtration step1 & 2;). The goodness-of-fit result is reported in Table 5.2 (“UC” refers to the uncongested-regime, and “C” refers to the congested-regime). Note that the break-point density ($K_b$) in modified dual-regime Greenshields model is fixed to be 13 for detector 4737A and 11 for detector 4792B, respectively. The reason is that the iterative optimization procedure of the modified dual-regime Greenshields model based on the entire 15-year data sets with all parameters being completely free could not converge in 500 iterations.

For the purpose of checking the validity of the model-fitting results, the research in this thesis calculated the free-flow speed by following the instructions of HCM 2000. Specifically, Chapter 13 of HCM 2000 claims that in a speed-flow diagram, speed is constant for flows up to 1300 (pc/hr/ln). Therefore, the free-flow speed based on HCM 2000 could be
calculated after removing “Step I” outliers and taking the average of all the speed values with flow rate $\leq 1300$ (pc/hr/ln). As a result, the HCM-based free-flow speed of detector 4737A and detector 4792B are 108.75 (km/hr) and 107.13 (km/hr), respectively. By comparison, the free-flow speed values estimated by the proposed model prototypes based on each detector’s 15-year data sets are fairly close to the HCM-based free-flow speeds.

Table 5.1 Model-fitting Results and Related Statistics Based on 15-year Data

<table>
<thead>
<tr>
<th>Detector</th>
<th>Model</th>
<th>Estimate &amp; Statistics</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Vint2</td>
<td>Kb</td>
</tr>
<tr>
<td>4737A</td>
<td>3RM</td>
<td>Estimate</td>
<td>133.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard Error</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95% Confidence Interval Upper Bound</td>
<td>134.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95% Confidence Interval Lower Bound</td>
<td>132.97</td>
</tr>
<tr>
<td></td>
<td>2RMGS</td>
<td>Estimate</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard Error</td>
<td>NA (Fixed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95% Confidence Interval Upper Bound</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95% Confidence Interval Lower Bound</td>
<td>NA</td>
</tr>
<tr>
<td>4792B</td>
<td>3RM</td>
<td>Estimate</td>
<td>132.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard Error</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95% Confidence Interval Upper Bound</td>
<td>133.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95% Confidence Interval Lower Bound</td>
<td>132.35</td>
</tr>
<tr>
<td></td>
<td>2RMGS</td>
<td>Estimate</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard Error</td>
<td>NA (Fixed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95% Confidence Interval Upper Bound</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95% Confidence Interval Lower Bound</td>
<td>NA</td>
</tr>
</tbody>
</table>
Table 5.2 Goodness-of-fit Statistics of Model Calibration Results Based on 15-year Data Sets

<table>
<thead>
<tr>
<th>Detector</th>
<th>Model</th>
<th>Aspect</th>
<th>FREQ</th>
<th>SST</th>
<th>SSE</th>
<th>R2</th>
<th>AdjR2</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4737A</td>
<td>3RM</td>
<td>All</td>
<td>429031</td>
<td>94731077</td>
<td>6762732</td>
<td>0.93</td>
<td>0.93</td>
<td>3.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UC</td>
<td>401376</td>
<td>24521787</td>
<td>5201075</td>
<td>0.79</td>
<td>0.79</td>
<td>3.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>27655</td>
<td>8379156</td>
<td>1561657</td>
<td>0.81</td>
<td>0.81</td>
<td>7.52</td>
</tr>
<tr>
<td></td>
<td>2RMGS</td>
<td>All</td>
<td>429031</td>
<td>94731077</td>
<td>9266637</td>
<td>0.90</td>
<td>0.90</td>
<td>4.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UC</td>
<td>401376</td>
<td>24521787</td>
<td>5201075</td>
<td>0.79</td>
<td>0.79</td>
<td>3.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>27655</td>
<td>8379156</td>
<td>2985933</td>
<td>0.64</td>
<td>0.64</td>
<td>10.39</td>
</tr>
<tr>
<td>4792B</td>
<td>3RM</td>
<td>All</td>
<td>451747</td>
<td>154987345</td>
<td>9717270</td>
<td>0.94</td>
<td>0.94</td>
<td>4.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UC</td>
<td>413960</td>
<td>17282603</td>
<td>8024279</td>
<td>0.54</td>
<td>0.54</td>
<td>4.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>37787</td>
<td>9716527</td>
<td>1692991</td>
<td>0.83</td>
<td>0.83</td>
<td>6.69</td>
</tr>
<tr>
<td></td>
<td>2RMGS</td>
<td>All</td>
<td>451747</td>
<td>154987345</td>
<td>12225783</td>
<td>0.92</td>
<td>0.92</td>
<td>5.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UC</td>
<td>413960</td>
<td>17282603</td>
<td>9940018</td>
<td>0.42</td>
<td>0.42</td>
<td>4.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>37787</td>
<td>9716527</td>
<td>2285765</td>
<td>0.76</td>
<td>0.76</td>
<td>7.78</td>
</tr>
</tbody>
</table>

5.1 Model Calibration Results of Detector 4737A

5.1.1 Demonstration of Model Parameters’ Estimates

Figure 5.1 - 5.8 are the plots of the temporal variation patterns of the two fitted models’ parameter estimates together with the confidence interval excerpted from the model-fitting results of the entire 15-year data sets (UB3RM: the upper bound of the 95% CI of the 15-year 3RM model fitting results; LB3RM: the lower bound of the 95% CI of the 15-year 3RM model fitting results; UB2RM and LB2RM are defined in the same manner for the 2RM model). Figure 5.9 is the plot of Minimum Uncongested Speed (Vmu) associated with each 13-week sample.
Figure 5.1 Model Estimates of Free-flow Speed (FFS)

Figure 5.2 Model Estimates of Vint2
Figure 5.3 Model Estimates of $V_{int}$

Figure 5.4 Model Estimates of $K_b$
Figure 5.5 Model Estimates of Kint2

Figure 5.6 Model Estimates of Kj
Figure 5.7 Model Estimates of $\alpha$

Figure 5.8 Model Estimates of $\alpha_2$
By qualitatively analyzing the above plots, it is justifiable to conclude that none of the series of model parameter estimates show a major increasing or decreasing trend during the 15-year analysis period. For example, the Free-Flow Speed ($FFS$) parameter is fairly stable over the entire 15 years period. The plots of $V_{\text{intercept}}$ ($V_{\text{int}}$), $V_{\text{intercept2}}$ ($V_{\text{int2}}$), Breakpoint Density ($Kb$) and $\alpha$ also display a stable temporal pattern despite some moderate fluctuations.

In particular, the plot of Jam Density ($Kj$) is visibly different from other parameters’ due to the fact that the optimization bounds of $Kj$ are greatly restricted, as discussed previously. Consequently, the model estimate for the $Kj$ parameter hits the lower bound (i.e. 90) many times, especially for the dual-regime MGS model. But again, there is no dominant
increasing or decreasing trend in $K_j$’s serial estimates, which is consistent with other aforementioned parameters.

On the other hand, the plots of $K_{\text{intercept2}}$ ($K_{\text{int2}}$) and $\alpha_2$ do demonstrate a relatively strong periodical fluctuation pattern. One possible reason that could have contributed to the visually periodical pattern in the plots is that $K_{\text{int2}}$ and $\alpha_2$ are more “sensitive” than the rest of the parameters. Specifically, the parameters that characterize the 2nd regime of macroscopic speed-density relationship, including $K_{\text{int2}}$ and $\alpha_2$, are based on relatively small amount of data points because of the regime classification rules defined in the model calibration procedures. Therefore, a slight difference in the number and/or distribution of the data points that belong to the 2nd regime would greatly impact the estimates of the associated model parameters, such as $K_{\text{int2}}$ and $\alpha_2$.

5.1.3 Correlation Analysis of Model Parameters’ Estimates

The research in this thesis applies SAS Procedure “PROC ARIMA” to analyze the autocorrelation property of each proposed model parameter. The lag is set to be 1 week.

Figure 5.10 – 5.17 demonstrate the autocorrelation analysis results of three-regime model’s parameter estimates.
Figure 5.10 Correlation Analysis of Three-Regime Model’s Free-flow Speed (FFS)
Figure 5.11 Correlation Analysis of Three-Regime Model’s Break-point Density ($K_b$)
Figure 5.12 Correlation Analysis of Three-Regime Model’s K_intercept2 (Kint2)
Figure 5.13 Correlation Analysis of Three-Regime Model’s V_intercept2 (Vint2)
Figure 5.14 Correlation Analysis of Three-Regime Model’s $\alpha_2$
Figure 5.15 Correlation Analysis of Three-Regime Model’s V_intercept (Vint)
Figure 5.16 Correlation Analysis of Three-Regime Model’s Jam Density ($K_j$)
Figure 5.17 Correlation Analysis of Three-Regime Model’s $\alpha$

Figure 5.18 – 5.22 demonstrate the correlation analysis of the dual-regime modified Greenshields model’s parameter estimates.
Figure 5.18 Correlation Analysis of Dual-regime Model’s Free-flow Speed ($FFS$)
Figure 5.19 Correlation Analysis of Dual-regime Model’s Break-point Density (Kb)
Figure 5.20 Correlation Analysis of Dual-regime Model’s V_intercept (V_{int})
Figure 5.21 Correlation Analysis of Dual-regime Model’s Jam Density ($K_j$)
The above autocorrelation bar plots clearly indicate strong serial correlation for all of the involved model parameters. In fact, the significant correlation within each series of parameter estimates is to be expected, because consecutive 13-week samples used to calibrate model parameters share a large portion of data points (recall Figure 4.6). Therefore, the consecutive parameter estimates are certainly correlated.
In particular, the autocorrelation plots of parameters $K_{int2}$ (Figure 5.12) and $\alpha_2$ (Figure 5.14) also suggest that the correlation property of parameter estimates with 1-year lag is more pronounced than the estimates with half-year lag. This is also consistent with common experience of traffic engineering in the sense that the traffic condition at a certain time point in a year is most similar with the traffic condition of the same time next year.

5.2 Model Calibration Results for Detector 4792B

5.2.1 Demonstration of Model Parameters’ Estimates

![Graph](image)

Figure 5.23 Model Estimates of Free-flow Speed ($FFS$)
Figure 5.24 Model Estimates of $Vint_2$

Figure 5.25 Model Estimates of $Vint$
Figure 5.26 Model Estimates of $K_b$

Figure 5.27 Model Estimates of $K_{int2}$
Figure 5.28 Model Estimates of $K_j$

Figure 5.29 Model Estimates of $\alpha_2$
Figure 5.30 Model Estimates of $\alpha$

### 5.2.2 Qualitative Analysis of Model Parameters’ Estimates

Qualitative analysis of the above plots of model parameter estimates based on detector 4792B’s data sets clearly leads to similar conclusions with previous ones. Specifically, none of the estimated parameters demonstrate a dominant increasing or decreasing trend, although some normal fluctuations do occur occasionally. The plot of $K_j$ shows essentially identical pattern with that of detector 4737A’s $K_j$. To sum it up, the traffic flow characteristics of the two locations are visually consistent in terms of comparing the key model parameters.

### 5.2.3 Correlation Analysis of Model Parameters’ Estimates

Figure 5.31 – 5.38 demonstrate the correlation analysis results of three-regime model’s parameter estimates.
Figure 5.31 Correlation Analysis of Three-regime Model’s Free-flow Speed (FFS)
Figure 5.32 Correlation Analysis of Three-regime Model’s Break-point Density ($K_b$)
Figure 5.33 Correlation Analysis of Three-regime Model’s $V_{\text{intercept}}$ ($V_{\text{int}}$)
Figure 5.34 Correlation Analysis of Three-regime Model’s V_intercept2 ($V_{int2}$)
Figure 5.35 Correlation Analysis of Three-regime Model’s $K_{\text{intercept2}}$ ($K_{\text{int2}}$)
Figure 5.36 Correlation Analysis of Three-regime Model’s Jam Density ($K_j$)
Figure 5.37 Correlation Analysis of Three-regime Model’s $\alpha$
Figure 5.38 Correlation Analysis of Three-regime Model’s $\alpha_2$

Figure 5.39 – 5.43 demonstrate the correlation analysis results of modified dual-regime Greenshields model’s parameter estimates.
Figure 5.39 Correlation Analysis of Dual-Regime model’s Free-flow Speed (FFS)
Figure 5.40 Correlation Analysis of Dual-Regime model’s Break-point Density ($Kb$)
Figure 5.41 Correlation Analysis of Dual-Regime model’s V_intercept (Vint)
Figure 5.42 Correlation Analysis of Dual-Regime model’s Jam Density ($K_j$)
The autocorrelation plots of detector 4792B’s serial parameter estimates are quite similar with the ones of detector 4737A. Note that the estimates of three-regime model’s $K_j$ term also displays a pattern that correlation between parameter estimates with 1-year lag is more pronounced than estimates with other lags. Again, this is because traffic conditions are more likely to be reproducible at the same time in a year.
Chapter 6 Conclusions & Outlook

6.1 Conclusions

By making use of the 15-year long traffic data series of MIDAS database in London, UK, the research in this thesis successfully shows that the key parameters of three-regime model and dual-regime modified Greenshields model estimated by SAS iterative optimization procedures do not indicate any significantly dominant increasing or decreasing trend lines within the 15-year analysis period. Additional correlation analysis shows that the serial model parameter estimates based on both data resources are highly correlated, as expected given the rolling windows on which they are based.

6.2 Recommendations

For the purposes of further developing the research topic in this thesis, the following list provides a brief description of some of the possible aspects that are not included in this thesis:

- Freeway Segment Type: The analysis in this research is only applied on basic freeway segments. It is recommended that the analysis procedure be validated on other types of typical freeway segments, such as on/off ramps and interchanges, to generalize the conclusions in this thesis.

- Minimum Uncongested Speed ($V_{mu}$): Current model parameters being fitted by the SAS PROC MODEL does not include the minimum uncongested speed term. Rather, it is empirically determined by experience-oriented approaches. It is recommended that the minimum uncongested speed be included in the model fitting procedure to be consistent with other parameters.
• Data source from United States: The data source being used in this research is from UK. It is possible that the driver behavior in United Kingdom is different from here in the United States. Therefore, it is recommended that follow-up research explore this research topic using American traffic data sets.

Despite the limitations of the research in its current progress, it clearly provides a theoretically rigorous foundation for researchers and engineers to justify their efforts of correctly designing macroscopic traffic stream models.
REFERENCES


APPENDIX A. SAS CODE FOR CALIBRATING MODEL PARAMETERS

/*

NAME: sdfit

AUTHORS: Billy Williams and Chenhao Liu

VERSION DATE: Aug 03, 2010

PURPOSE: Iterate through a long 15-min traffic condition data set and fit
speed-density models to 13 week samples

INPUT: %sdfit(site,lib,st_date,end_date,dir);

where site - Site descriptor AND name of SAS data set to process, e.g. M254737a
st_date - Date from which model fitting starts
end_date - Date through which model fitting is done
dir - Directory path where SAS data set resides and output files will be written

The st_date and end_date values should be in ddmmyyyy format.
For example, a valid date is 01jan1996

Enclose in the directory (dir) in single quotes e.g 'c:\Output'
or 'E:\Project Files\Model Fit Results'

NOTES:

SAS data set must have SAS datetime variable named TIMESTAMP plus density, speed, and flow.

See code for details on permanent output data sets created.

Make sure that site data set exists in the given SAS library.

*/

%macro sdfit(site,st_date,end_date,dir);

/* Set up library */
libname tmp &dir;

/* Set up macro variables for dates */

%let begin=%str(%'&st_date%'d);
%let end=%str(%'&end_date%'d);

data null;
start=%sysevalf(&begin)*24*3600;
last=%sysevalf(&end)+1)*24*3600;
call symput('start',start);
call symput('last',last);
/* Loop through entire data set using 13 week samples iterating in one day intervals */

%let day1=&start;
%let day92=%eval(&start+7862400);

%do %while(%eval(&day92)<=%eval(&last));

/* Assign current sample to data set temp */

data temp;
set tmp.&site;
where TIMESTAMP>=&day1 and TIMESTAMP<&day92;
run;

/* Determine the 95th percentile flow and assign to macro variable */

proc means data=temp noprint;
output out=stats P95(flow)=U95 MAX(flow)=max mean(flow)=avg;
quit;
data null;
set stats;
call symput('u95',u95);
run;

/* Extract observations >= 95th percentile flow and speed >= 80 */

data temp2;
set temp;
where flow>=&u95 and speed >= 80;
run;

/* Determine minimum uncongested speed and assign to macro variable */

proc means data=temp2 noprint;
output out=stats min(speed)=mus;
quit;
data null;
set stats;
call symput('mus',mus);
run;

/* Remove low speed,low density outliers -- type 1 */

data temp;
set temp;
where ^(((speed<100 & density<10) or (speed<80 & density<15) or (speed<40 & density<28) or (speed<20 & density<40) or flow=0));
call symput('obs_f1',_n_);
run;
%let n_ol_1=%eval(8736-&obs_f1);
/* Fit three-regime model to set with only type 1 outliers removed */
/* Regime 1 - FFS, Regime 2 - Uncongested with speed drop, Regime 3 - Congested */
proc model data=temp outparms=Full3Rparms; * noprint;
parms Vint2=135 Kb=12 Kint2=65 Vint3=175 Kj=110 a2=0.9 a3=6;
bounds 80 <= Vint2 <= 200,
  5  <= Kb  <= 22,
  22  <= Kint2 <= 200,
  80  <= Vint3 <= 300,
  90  <= Kj  <= 130,
a2  >= 0,
  a3  >= 0;
if speed >= &mus and density <= Kb then
  speed = Vint2*(1-Kb/Kint2)**a2;
if  speed >= &mus and density > Kb then
  speed = Vint2*(1-density/Kint2)**a2;
if  speed < &mus then
  speed = Vint3*(1-density/Kj)**a3;
fit speed / fiml outactual outpredict outresid outcov out=Full3RFit outest=Full3Rest ginv=g4 maxiter=500;
quit;
/* Assign model parameters and standard errors to macro variables */
data null;
set Full3Rest;
if _n_ =1 then do;
call symput('Full3R_Vint2',Vint2);
call symput('Full3R_Kb',Kb);
call symput('Full3R_Kint2',Kint2);
call symput('Full3R_Vint3',Vint3);
call symput('Full3R_Kj',Kj);
call symput('Full3R_a2',a2);
call symput('Full3R_a3',a3);
end;
if _NAME_='Vint2' then call symput('Full3R_Vint2SE',Vint2**(0.5));
if _NAME_='Kb' then call symput('Full3R_KbSE',Kb**(0.5));
if _NAME_='Kint2' then call symput('Full3R_Kint2SE',Kint2**(0.5));
if _NAME_='Vint3' then call symput('Full3R_Vint3SE',Vint3**(0.5));
if _NAME_='Kj' then call symput('Full3R_KjSE',Kj**(0.5));
if _NAME_='a2' then call symput('Full3R_a2SE',a2**(0.5));
if _NAME_='a3' then call symput('Full3R_a3SE',a3**(0.5));
run;
/* Assign FFS to macro variable */
data null;
set full3rparms;
ffs=Vint2*(1-Kb/Kint2)**a2;
call symput('Full3R_ffs',ffs);
run;
/* Extract 3-regime model fit statistics */

data actual;
set Full3RFit;
where _TYPE_="ACTUAL";
rename speed=ActSpd;
keep Density Speed;
data predict;
set Full3RFit;
where _TYPE_="PREDICT";
rename speed=EstSpd;
keep Speed;
data residual;
set Full3RFit;
where _TYPE_="RESIDUAL";
rename speed=ErrSpd;
keep Speed;
data ModelAppl;
merge actual predict residual;
run;
proc means data=modelappl noprint;
output out=GOFall css(actspd)=sst uss(errspd)=sse;
data GOFall;
set GOFall;
R2=1-sse/sst;
AdjR2=1-(((_FREQ_-1)/(_FREQ_-7))*(sse/sst));
RMSE=(sse/(_FREQ_-7)**0.5);
call symput('all_R2',R2);
call symput('all_AdjR2',AdjR2);
call symput('all_obs',_FREQ_);
call symput('all_RMSE',RMSE);
run;
data ModelUC;
set ModelAppl;
where actspd>=&mus;
data ModelC;
set ModelAppl;
where actspd<&mus;
run;
proc means data=modelUC noprint;
output out=GOFUC css(actspd)=sst uss(errspd)=sse std(errspd)=sderr;
data GOFUC;
set GOFUC;
R2=1-sse/sst;
AdjR2=1-(((_FREQ_-1)/(_FREQ_-7))*(sse/sst));
RMSE=(sse/(_FREQ_-7)**0.5);
call symput('UC_R2',R2);
call symput('UC_AdjR2',AdjR2);
call symput('UC_sderr',sderr);
call symput('UC_obs',_FREQ_);
call symput('UC_RMSE',RMSE);
run;
proc means data=modelC noprint;
output out=GOFC css(actspd)=sst uss(errspd)=sse std(errspd)=sderr;
data GOFC;
set GOFC;
R2=1-sse/sst;
AdjR2=1-(((_FREQ_-1)/(_FREQ_-7))*(sse/sst));
RMSE=(sse/(_FREQ_-7))**0.5;
call symput('C_R2',R2);
call symput('C_AdjR2',AdjR2);
call symput('C_sderr',sderr);
call symput('C_obs',_FREQ_);
call symput('C_RMSE',RMSE);
run;

/* Put model parameters and standard errors in iteration data set */
data full3rmodel;
TIMESTAMP=&day1;
format TIMESTAMP datetime9.;
Vint2=&Full3R_Vint2;
Vint2_SE=&Full3R_Vint2SE;
Kb=&Full3R_Kb;
Kb_SE=&Full3R_KbSE;
Kint2=&Full3R_Kint2;
Kint2_SE=&Full3R_Kint2SE;
Vint3=&Full3R_Vint3;
Vint3_SE=&Full3R_Vint3SE;
Kj=&Full3R_Kj;
Kj_SE=&Full3R_KjSE;
a2=&Full3R_a2;
a2_SE=&Full3R_a2SE;
a3=&Full3R_a3;
a3_SE=&Full3R_a3SE;
FFS=&Full3R_ffs;
Vmu=&mus;
run;

/* Populate permanent parameter data set for all iterations */
%if %eval(&day1)=%eval(&start) %then %do;
%let f3rm=%str(&site)f3Rm;
data tmp.&f3rm;
set full3rmodel;
run;
%end;
%else %do;
proc append base=tmp.&f3rm data=full3rmodel;
quit;
%end;

/* Put model goodness of fit stats in iteration data set */
data full3rGOF;
TIMESTAMP=&day1;
format TIMESTAMP datetime9.;
all_R2=&all_R2;
all_AdjR2=&all_AdjR2;
all_RMSE=&all_RMSE;
all_obs=&all_obs;
UC_R2=&UC_R2;
UC_AdjR2=&UC_AdjR2;
UC_RMSE=&UC_RMSE;
UC_obs=&UC_obs;
C_R2=&C_R2;
C_AdjR2=&C_AdjR2;
C_RMSE=&C_RMSE;
C_obs=&C_obs;
run;

/* Populate permanent goodness of fit data set for all iterations */
%if %eval(&day1)=%eval(&start) %then %do;
%let f3rgof=%str(&site)f3Rgof;
data tmp.&f3rgof;
set full3rgof;
run;
%end;
%else %do;
proc append base=tmp.&f3rgof data=full3rgof;
quit;
%end;

/* Use statistics to flag model-based outliers - type 2 */
data Filter;
set Modelappl;
if ActSpd>=&mus then
  UC_crit=abs(ErrSpd)/&UC_sderr;
else UC_crit=0;
if ActSpd<&mus then
  C_crit=abs(ErrSpd)/&C_sderr;
else C_crit=0;
UC_flag=0;
C_flag=0;
data Filter;
set Filter;
if UC_crit>=3 then
  UC_flag=1;
if C_crit>=3 then
  C_flag=1;
proc means data=Filter noprint;
output out=outlier sum(UC_flag C_flag)=UC_outliers C_outliers;
data null;
set outlier;
call symput('UC_outliers',UC_outliers);
call symput('C_outliers',C_outliers);
data Filtered;
set Filter;
where ~(UC_flag=1 or C_flag=1);
keep ActSpd Density;
rename ActSpd=Speed;
run;

/* Put number of outliers by type in iteration data set */
data outliers;
TIMESTAMP=&day1;
format TIMESTAMP datetime9.;
apriori=&n_ol_1;
uncongested=&UC_outliers;
congested=&C_outliers;
run;

/* Populate permanent outlier data set for all iterations */
%if %eval(&day1)=%eval(&start) %then %do;
%let f3rout=%str(&site)outliers;
data tmp.&f3rout;
set outliers;
run;
%end;
%else %do;
proc append base=tmp.&f3rout data=outliers;
quit;
%end;

/* Fit three-regime model to set with type 1 & type 2 outliers removed */
/* Regime 1 - FFS, Regime 2 - Uncongested with speed drop, Regime 3 - Congested */
proc model data=Filtered outparms=Fltd3Rparms outmodel=Fltd3Rmodel noprint;
parms Vint2=135 Kb=12 Kint2=65 Vint3=175 Kj=110 a2=0.9 a3=6;
bounds 80 <= Vint2 <= 200,
5 <= Kb <= 22,
22 <= Kint2 <= 200,
80 <= Vint3 <= 300,
90 <= Kj <= 130,
a2 >= 0,
a3 >= 0;
if speed >= &mus and density <= Kb then
  speed = Vint2*(1-Kb/Kint2)**a2;
if speed >= &mus and density > Kb then
  speed = Vint2*(1-density/Kint2)**a2;
if speed < &mus then
  speed = Vint3*(1-density/Kj)**a3;
fit speed / fiml outactual outpredict outresid out=Fltd3RFit outcov outtest=Fltd3Rest ginv=g4 maxiter=500;
quit;

/* Assign model parameters and standard errors to macro variables */
data null;
set Fltd3Rest;
if _n_ = 1 then do;
call symput('Fltd3R_Vint2', Vint2);
call symput('Fltd3R_Kb', Kb);
call symput('Fltd3R_Kint2', Kint2);
call symput('Fltd3R_Vint3', Vint3);
call symput('Fltd3R_Kj', Kj);
call symput('Fltd3R_a2', a2);
call symput('Fltd3R_a3', a3);
end;
if _NAME_='Vint2' then call symput('Fltd3R_Vint2SE', Vint2**0.5);
if _NAME_='Kb' then call symput('Fltd3R_KbSE', Kb**0.5);
if _NAME_='Kint2' then call symput('Fltd3R_Kint2SE', Kint2**0.5);
if _NAME_='Vint3' then call symput('Fltd3R_Vint3SE', Vint3**0.5);
if _NAME_='Kj' then call symput('Fltd3R_KjSE', Kj**0.5);
if _NAME_='a2' then call symput('Fltd3R_a2SE', a2**0.5);
if _NAME_='a3' then call symput('Fltd3R_a3SE', a3**0.5);
run;

/* Assign FFS to macro variable */
data null;
set fltd3rparms;
ffs=Vint2*(1-Kb/Kint2)**a2;
call symput('Fltd3R_ffs', ffs);
run;

/* Extract 3-regime model fit statistics */
data actual;
set Fltd3RFit;
where _TYPE_='ACTUAL';
rename speed=ActSpd;
keep Density Speed;
data predict;
set Fltd3RFit;
where _TYPE_='PREDICT';
rename speed=EstSpd;
keep Speed;
data residual;
set Fltd3RFit;
where _TYPE_='RESIDUAL';
rename speed=ErrSpd;
keep Speed;
data ModelApplFltd;
merge actual predict residual;
run;
proc means data=modelapplfltd noprint;
output out=GOFallfltd css(actspd)=sst uss(errspd)=sse;
data GOFallfltd;
set GOFallfltd;
R2=1-sse/sst;
AdjR2=1-(((_FREQ_-1)/(_FREQ_-7))*(sse/sst));
RMSE=(sse/(_FREQ_-7))**0.5;
call symput('all_R2fltd',R2);
call symput('all_AdjR2fltd',AdjR2);
call symput('all_obsfltd',_FREQ_);
call symput('all_RMSEfltd',RMSE);
run;
data ModelUCfltd;
set ModelApplfltd;
where actspd>=&mus;
data ModelCfltd;
set ModelApplfltd;
where actspd<&mus;
run;
proc means data=modelUCfltd noprint;
output out=GOFUCfltd css(actspd)=sst uss(errspd)=sse;
data GOFUCfltd;
set GOFUCfltd;
R2=1-sse/sst;
AdjR2=1-(((_FREQ_-1)/(_FREQ_-7))*(sse/sst));
RMSE=(sse/(_FREQ_-7))**0.5;
call symput('UC_R2fltd',R2);
call symput('UC_AdjR2fltd',AdjR2);
call symput('UC_obsfltd',_FREQ_);
call symput('UC_RMSEfltd',RMSE);
run;
proc means data=modelCfltd noprint;
output out=GOFCfltd css(actspd)=sst uss(errspd)=sse;
data GOFCfltd;
set GOFCfltd;
R2=1-sse/sst;
AdjR2=1-(((_FREQ_-1)/(_FREQ_-7))*(sse/sst));
RMSE=(sse/(_FREQ_-7))**0.5;
call symput('C_R2fltd',R2);
call symput('C_AdjR2fltd',AdjR2);
call symput('C_obsfltd',_FREQ_);
call symput('C_RMSEfltd',RMSE);
run;
/* Put model parameters and standard errors in iteration data set */
data fld3rmodel;
TIMESTAMP=&day1;
format TIMESTAMP datetime9.;
Vint2=&fltd3R_Vint2;
Vint2_SE=&fltd3R_Vint2SE;
Kb=&fltd3R_Kb;
Kb_SE=&fltd3R_KbSE;
Kint2=&fltd3R_Kint2;
Kint2_SE=&fltd3R_Kint2SE;
Vint3=&fltd3R_Vint3;
Vint3_SE=&fltd3R_Vint3SE;
Kj=&fltd3R_Kj;
Kj_SE=&fltd3R_KjSE;
a2=&fltd3R_a2;
a2_SE=&fltd3R_a2SE;
a3=&fltd3R_a3;
a3_SE=&fltd3R_a3SE;
FFS=&fltd3R_ffs;
run;

/* Populate permanent parameter data set for all iterations */
%if %eval(&day1)=%eval(&start) %then %do;
%let f3rfm=%str(&site)f3Rfm;
data tmp.&f3rfm;
set fltd3rmodel;
run;
%end;
%else %do;
proc append base=tmp.&f3rfm data=fltd3rmodel;
quit;
%end;

/* Put model goodness of fit stats in iteration data set */
data fltd3rGOF;
TIMESTAMP=&day1;
format TIMESTAMP datetime9.;
all_R2=&all_R2fltd;
all_AdjR2=&all_AdjR2fltd;
all_RMSE=&all_RMSEfltd;
all_obs=&all_obsfltd;
UC_R2=&UC_R2fltd;
UC_AdjR2=&UC_AdjR2fltd;
UC_RMSE=&UC_RMSEfltd;
UC_obs=&UC_obsfltd;
C_R2=&C_R2fltd;
C_AdjR2=&C_AdjR2fltd;
C_RMSE=&C_RMSEfltd;
C_obs=&C_obsfltd;
run;

/* Populate permanent goodness of fit data set for all iterations */
%if %eval(&day1)=%eval(&start) %then %do;
%let f3rfgof=%str(&site)f3Rfgof;
data tmp.&f3rfgof;
set fltd3rfgof;

/* Fit two-regime modified Greenshield's to set with type 1 and type 2 outliers removed */

proc model data=filtered outparms=MGS2Rparms outmodel=MGS2Rmodel ;
  *noprint;
  parms Vint=160 Kb=13 Kj=90 a=2;
  bounds 80 <= Vint <= 300,
         9 <= Kb <= 25,
         90 <= Kj <= 130,
         a > 0;
  if density < Kb then
    speed = Vint*(1-Kb/Kj)**a;
  else speed = Vint*(1-density/Kj)**a;
  fit speed / fiml outactual outpredict outresid out=MGS2RFit outcov outest=MGS2Rest ginv=g4 maxiter=500;
quit;

/* Assign model parameters and standard errors to macro variables */

data null;
  set MGS2Rest;
  if _n_=1 then do;
    call symput('MGS2R_Vint',Vint);
    call symput('MGS2R_Kb',Kb);
    call symput('MGS2R_Kj',Kj);
    call symput('MGS2R_a',a);
  end;
  if _NAME_='Vint' then call symput('MGS2R_VintSE',Vint**0.5);
  if _NAME_='Kb' then call symput('MGS2R_KbSE',Kb**0.5);
  if _NAME_='Kj' then call symput('MGS2R_KjSE',Kj**0.5);
  if _NAME_='a' then call symput('MGS2R_aSE',a**0.5);
  run;

/* Assign FFS to macro variable */

data null;
  set mgs2rparms;
  ffs=Vint*(1-Kb/Kj)**a;
  call symput('MGS2R_ffs',ffs);
  run;

/* Extract 2-regime model fit statistics */

data actual;
  set MGS2RFit;
  where _TYPE_="ACTUAL";
  rename speed=ActSpd;
  keep Density Speed;
data predict;
set MGS2RFit;
where _TYPE_="PREDICT";
rename speed=EstSpd;
keep Speed;
data residual;
set MGS2RFit;
where _TYPE_="RESIDUAL";
rename speed=ErrSpd;
keep Speed;
data ModelApplMGS;
merge actual predict residual;
r
proc means data=modelapplMGS noprint;
output out=GOFallMGS css(actspd)=sst uss(errspd)=sse;
data GOFallMGS;
set GOFallMGS;
R2=1-sse/sst;
AdjR2=1-(((FREQ-1)/(FREQ-4))*(sse/sst));
RMSE=(sse/(FREQ-4))**0.5;
call symput('all_R2MGS',R2);
call symput('all_AdjR2MGS',AdjR2);
call symput('all_obsMGS',FREQ);
call symput('all_RMSEMGS',RMSE);
r
data ModelUCMGS;
set ModelApplMGS;
where actspd>=&mus;
data ModelCMGS;
set ModelApplMGS;
where actspd<&mus;
r
proc means data=modelUCMGS noprint;
output out=GOFUCMGS css(actspd)=sst uss(errspd)=sse;
data GOFUCMGS;
set GOFUCMGS;
R2=1-sse/sst;
AdjR2=1-(((FREQ-1)/(FREQ-4))*(sse/sst));
RMSE=(sse/(FREQ-4))**0.5;
call symput('UC_R2MGS',R2);
call symput('UC_AdjR2MGS',AdjR2);
call symput('UC_obsMGS',FREQ);
call symput('UC_RMSEMGS',RMSE);
r
proc means data=modelCMGS noprint;
output out=GOFCMGS css(actspd)=sst uss(errspd)=sse;
data GOFCMGS;
set GOFCMGS;
R2=1-sse/sst;
AdjR2=1-(((FREQ-1)/(FREQ-4))*(sse/sst));
RMSE=(sse/(FREQ-4))**0.5;
call symput('C_R2MGS',R2);
call symput('C_AdjR2MGS', AdjR2);
call symput('C_obsMGS', _FREQ_);
call symput('C_RMSEMGS', RMSE);
run;

/* Put model parameters and standard errors in iteration data set */
data MGS2rmodel;
TIMESTAMP=&day1;
format TIMESTAMP datetime9.;
Vint=&MGS2R_Vint;
Vint_SE=&MGS2R_VintSE;
Kb=&MGS2R_Kb;
Kb_SE=&MGS2R_KbSE;
Kj=&MGS2R_Kj;
Kj_SE=&MGS2R_KjSE;
a=&MGS2R_a;
a_SE=&MGS2R_aSE;
FFS=&MGS2R_ffs;
run;

/* Populate permanent parameter data set for all iterations */
%if %eval(&day1)=%eval(&start) %then %do;
%let MGS2Rm=%str(&site)MGS2Rm;
data tmp.&MGS2Rm;
set MGS2rmodel;
run;
%end;
%else %do;
proc append base=tmp.&MGS2Rm data=MGS2rmodel;
quit;
%end;

/* Put model goodness of fit stats in iteration data set */
data MGS2RGOF;
TIMESTAMP=&day1;
format TIMESTAMP datetime9.;
all_R2=&all_R2MGS;
all_AdjR2=&all_AdjR2MGS;
all_RMSE=&all_RMSEMGS;
all_obs=&all_obsMGS;
UC_R2=&UC_R2MGS;
UC_AdjR2=&UC_AdjR2MGS;
UC_RMSE=&UC_RMSEMGS;
UC_obs=&UC_obsMGS;
C_R2=&C_R2MGS;
C_AdjR2=&C_AdjR2MGS;
C_RMSE=&C_RMSEMGS;
C_obs=&C_obsMGS;
run;
/* Populate permanent goodness of fit data set for all iterations */

%if %eval(&day1)=%eval(&start) %then %do;
%let MGS2Rgof=%str(&site)MGS2Rgof;
data tmp.&MGS2Rgof;set MGS2Rgof;
run;
%end;
%else %do;
proc append base=tmp.&MGS2Rgof data=MGS2Rgof;
quit;
%end;

data null;
date=datepart(&day1);
format date date7.;
day=day(date);
month=month(date);
year=year(date);
call symputx('day',day);
call symputx('month',month);
call symputx('year',year);
run;

%if (%eval(&day)=1)&(%eval(&month)=1 or %eval(&month)=4 or %eval(&month)=7 or %eval(&month)=10) %then %do;
%let sep=_;
%let qtr=&site&sep&month&sep&year;
data density;
set filtered;
keep density;
data flt3r;
set modelapplfltd;
rename EstSpd=EstSpd3R;
data flt3r;
set flt3r;
keep ActSpd EstSpd3R;
data mgs;
set modelapplmgs;
rename EstSpd=EstSpdMGS;
data mgs;
set mgs;
keep EstSpdMGS;
run;
data tmp.&qtr;
merge density flt3r mgs;
run;
%end;

%let day1=%eval(&day1+86400);
%let day92=%eval(&day92+86400);
%end;
%mend;

%sdfit(FileName,StartingDate,EndDate,'File Folder');

quit;}