

7-13-2012

# Improving Seasonal Factor Estimates for Adjustment of Annual Average Daily Traffic

Shanshan Yang

Florida International University, syang003@fiu.edu

**DOI:** 10.25148/etd.FI12080803

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

---

## Recommended Citation

Yang, Shanshan, "Improving Seasonal Factor Estimates for Adjustment of Annual Average Daily Traffic" (2012). *FIU Electronic Theses and Dissertations*. 709.

<https://digitalcommons.fiu.edu/etd/709>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact [dcc@fiu.edu](mailto:dcc@fiu.edu).

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

IMPROVING SEASONAL FACTOR ESTIMATES FOR ADJUSTMENT OF  
ANNUAL AVERAGE DAILY TRAFFIC

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CIVIL ENGINEERING

by

Shanshan Yang

2012

To: Dean Amir Mirmiran  
College of Engineering and Computing

This dissertation, written by Shanshan Yang, and entitled Improving Seasonal Factor Estimates for Adjustment of Annual Average Daily Traffic, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

---

L. David Shen

---

Mohammed Hadi

---

Zhenmin Chen

---

Albert Gan, Major Professor

Date of Defense: July 13, 2012

The dissertation of Shanshan Yang is approved.

---

Dean Amir Mirmiran  
College of Engineering and Computing

---

Dean Lakshmi N. Reddi  
University Graduate School

Florida International University, 2012

## DEDICATION

I dedicate this dissertation to my family for their unconditional and endless love.



## ACKNOWLEDGMENTS

I wish to express my deepest gratitude and appreciation to my former major advisor, Dr. Fang Zhao, who passed away during this dissertation research. Dr. Zhao was a great mentor and will always have a special place in my heart.

I am equally grateful to Dr. Albert Gan for his guidance, support, and confidence in me. Dr. Gan started as a member of my committee and I am thankful for his willingness to serve as my major advisor and guide me to the completion of this dissertation. Dr. Gan exemplifies a responsible professional in our society.

I would like to thank Dr. David Shen, Dr. Mohammed Hadi, and Dr. Zhenmin Chen for serving on my committee. I took many major core courses from both Dr. Shen and Dr. Hadi. They helped me prepare for this research and build a solid foundation for my future career. As a statistics expert, Dr. Chen provided me with very useful guidance in my research.

I would also like to thank the following colleagues at Florida International University (FIU) for collaborating with me all of these years: Dr. Chenxi Lu, Dr. Soon Chung, Dr. Lee-Fang Chow, Dr. Yan Xiao, Ms. Meng Ma, Dr. Xuemei Liu, Dr. Zhen Ding, and Dr. Jing Qi. Additionally, Ms. Laura Osorno, Ms. Amy Diaz, and Ms. Haydee Cadena provided many helpful services to the students in the department, and I am grateful for their help in my overseas study.

I would like to acknowledge financial support for this research from the Florida Department of Transportation (FDOT) and the FIU Department of Civil and Environmental Engineering.

Finally, I would like to thank my family for their constant encouragement and love. Without their patience, understanding, and support, the completion of this work would not have been possible.

ABSTRACT OF THE DISSERTATION

IMPROVING SEASONAL FACTOR ESTIMATES FOR ADJUSTMENT OF

ANNUAL AVERAGE DAILY TRAFFIC

by

Shanshan Yang

Florida International University, 2012

Miami, Florida

Professor Albert Gan, Major Professor

Traffic volume data are input to many transportation analyses including planning, roadway design, pavement design, air quality, roadway maintenance, funding allocation, etc. Annual Average Daily Traffic (AADT) is one of the most often used measures of traffic volume. Acquiring the actual AADT data requires the collection of traffic counts continuously throughout a year, which is expensive, thus, can only be conducted at a very limited number of locations. Typically, AADTs are estimated by applying seasonal factors (SFs) to short-term counts collected at portable traffic monitoring sites (PTMSs).

Statewide in Florida, the Florida Department of Transportation (FDOT) operates about 300 permanent traffic monitoring sites (TTMSs) to collect traffic counts at these sites continuously. TTMSs are first manually classified into different groups (known as seasonal factor categories) based on both engineering judgment and similarities in the traffic and roadway characteristics. A seasonal factor category is then assigned to each PTMS according to the site's functional classification and geographical location. The SFs of the assigned category are then used to adjust traffic counts collected at PTMSs to

estimate the final AADTs. This dissertation research aims to develop a more objective and data-driven method to improve the accuracy of SFs for adjusting PTMSs.

A statewide investigation was first conducted to identify potential influential factors that contribute to seasonal fluctuations in traffic volumes in both urban and rural areas in Florida. The influential factors considered include roadway functional classification, demographic, socioeconomic, land use, etc. Based on these factors, a methodology was developed for assigning seasonal factors from one or more TTMSs to each PTMS.

The assigned seasonal factors were validated with data from existing TTMSs. The results show that the average errors of the estimated seasonal factors are, on average, about 4 percent. Nearly 95 percent of the estimated monthly SFs contain errors of no more than 10 percent. It was concluded that the method could be applied to improve the accuracy in AADT estimation for both urban and rural areas in Florida.

## TABLE OF CONTENTS

CHAPTER	PAGE
1. INTRODUCTION .....	1
1.1 Background .....	1
1.2 Problem Statement .....	2
1.3 Research Goals and Objectives .....	6
1.4 Organization .....	6
2. LITERATURE REVIEW .....	8
2.1 Current Practice in Florida .....	8
2.2 Cluster Analysis .....	11
2.2.1 Nonparametric Agglomerative Hierarchical Clustering Methods .....	11
2.2.2 Nonhierarchical Clustering Methods .....	14
2.2.3 Model-Based Gaussian Cluster Analysis .....	15
2.3 Geographic/Functional Assignment .....	24
2.4 Regression Analysis .....	27
2.5 Artificial Neural Networks .....	31
2.5.1 Supervised Learning .....	32
2.5.2 Unsupervised Learning .....	37
2.6 Genetic Algorithms .....	41
2.7 Assignment of Count Sites .....	44
2.8 Summary .....	46
3. METHODOLOGY .....	48
3.1 Methodology Framework .....	48
3.2 Data Preparation .....	50
3.2.1 Data Imputation for Dependent Variables .....	50
3.2.2 Data Acquisition for Independent Variables .....	50
3.3 Model Development .....	54
3.4 Monthly Seasonal Factor Assignment .....	56
3.5 Summary .....	57
4. DATA ACQUISITION AND PROCESSING .....	58
4.1 Imputation of Monthly Seasonal Factor Data .....	58
4.1.1 Data Imputation Procedure .....	58
4.1.2 Data Imputation Results .....	63
4.2 Definition of Independent Variables .....	64
4.2.1 Roadway Characteristic Variables .....	64
4.2.2 Demographic and Socioeconomic Variables .....	65
4.2.3 Special Land Use Variables for Urban Models .....	67
4.2.4 Location Variables for Rural Models .....	69
4.3 Summary .....	70

5. MODEL DEVELOPMENT .....	71
5.1 Modeling Influential Variables of Seasonal Factors in Urban Areas .....	71
5.1.1 Delineation of Model Areas for Urban Areas .....	71
5.1.2 North Florida Model Results .....	75
5.1.3 Central Florida Model Results .....	78
5.1.4 South Florida Model Results .....	81
5.2 Modeling Influential Variables of Seasonal Factors in Rural Areas .....	83
5.2.1 Classification of Hourly Traffic Pattern for Rural TTMSs .....	84
5.2.2 Single-Peak Models .....	89
5.2.3 Double-Peak Models .....	93
5.3 Summary .....	96
6. SEASONAL FACTOR ASSIGNMENT .....	97
6.1 Methodology for Measuring Similarity between Two Count Sites .....	97
6.2 Urban TTMS Assignment within Model Regions .....	99
6.2.1 Urban TTMS Assignment Based on Full Variable Set .....	99
6.2.2 Urban TTMS Assignment Based on a Reduced Variable Set .....	100
6.3 Rural TTMS Assignment within Hourly Traffic Pattern Groups .....	101
6.3.1 Rural TTMS Assignment Based on Full Variable Set .....	101
6.3.2 Rural TTMS Assignment Based on a Reduced Variable Set .....	102
6.4 Summary .....	102
7. EVALUATION OF THE ASSIGNMENT RESULTS .....	104
7.1 Evaluation of the Assignment Results .....	104
7.1.1 Evaluation of the Assignment Results for Urban Areas .....	108
7.1.2 Evaluation of the Assignment Results for Rural Area .....	110
7.2 Evaluation of the Assignment Methods .....	112
7.3 Evaluation of the Current TTMSs .....	117
7.3.1 Evaluation of the MSFs Pattern for Current TTMSs .....	117
7.3.2 Evaluation of the Coverage for Current TTMSs in Urban Area .....	121
7.4 Summary .....	127
8. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS .....	128
8.1 Summary and Conclusions .....	128
8.2 Recommendations for Future Research .....	132
REFERENCES .....	134
APPENDICES .....	139
VITA .....	157

## LIST OF TABLES

TABLE	PAGE
Table 2.1 Available Parameterizations of Covariance Matrix .....	19
Table 4.1 Imputation of Site 930099 .....	60
Table 4.2 Imputation of Site 930174 .....	61
Table 4.3 Imputation of Site 740047 .....	62
Table 4.4 List of TTMS with Missing Data after Imputation .....	63
Table 4.5 Roadway Characteristic Variables for Urban Roads .....	65
Table 4.6 Variables for Urban and Rural Areas .....	66
Table 4.7 Age Group Variables for Rural Roads .....	67
Table 4.8 Location Characteristic Variables for Urban Roads .....	68
Table 4.9 Land use Dummy Variables for Urban Road .....	69
Table 4.10 Special Location Variables .....	69
Table 5.1 List of Counties within Each of the Sub-region (Analysis Area) .....	74
Table 5.2 Regression Models for North Florida (NFL) .....	75
Table 5.3 Variables from NFL Model Sorted by Month and Partial $R^2$ Value .....	76
Table 5.4 Variables from NFL Model Sorted by Name and Partial $R^2$ Value .....	76
Table 5.5 Variables from NFL Model Sorted by Partial $R^2$ Value .....	77
Table 5.6 Regression Models for Central Florida (CFL) .....	79
Table 5.7 Variables from CFL Model Sorted by Month and Partial $R^2$ Value .....	80
Table 5.8 Variables from CFL Model Sorted by Name and Partial $R^2$ Value .....	81
Table 5.9 Variables from CFL Model Sorted by Partial $R^2$ Value .....	81
Table 5.10 Regression Models for South Florida (SFL) .....	82
Table 5.11 Variables from SFL Model Sorted by Month and Partial $R^2$ Value .....	82

Table 5.12	Variables from SFL Model Sorted by Name and Partial $R^2$ Value.....	83
Table 5.13	Variables from SFL Model Sorted by Partial $R^2$ Value .....	83
Table 5.14	Comparison of Adjusted R-square for Different Monthly SF Models.....	88
Table 5.15	Regression Models for the Single-Peak Group for Rural Areas .....	91
Table 5.16	Model Variables for Rural SP Group Sorted by Month and Partial $R^2$ Value .....	92
Table 5.17	Model Variables for Rural SP Group Sorted by Name and Partial $R^2$ Value .....	92
Table 5.18	Model Variables for Rural SP Group Sorted by Partial $R^2$ Value.....	93
Table 5.19	Regression Models for Rural DP Group .....	94
Table 5.20	Model Variables for Rural DP Group Sorted by Month and Partial $R^2$ value .....	95
Table 5.21	Model Variables for Rural DP Group Sorted by Name and Partial $R^2$ Value .....	95
Table 5.22	Model Variables for Rural DP Group Sorted by Partial $R^2$ Value.....	96
Table 6.1	Variable Sets Used for Assignment for Model Regions .....	99
Table 6.2	Sample Assignment Results for TTMSs in North Florida with Full Variable Set .....	100
Table 6.3	Reduced Variable Sets Used for Assignment for Three Model Regions ....	101
Table 6.4	Variable Sets Used for Assignment for Model Groups.....	102
Table 6.5	Variable Sets Used for Assignment for Model Groups.....	102
Table 7.1	Seasonal Factors for Sample Site 899921 and First Five Best Matched Sites .....	105
Table 7.2	Average Percentage Errors of the Assignment Results for Urban Area. ....	107
Table 7.3	Average Errors of the Assignment Results for Rural Area. ....	107
Table 7.4	Distribution for Percentage Differences in Urban Areas .....	109
Table 7.5	Distribution of Percentage Differences for TTMSs in Rural Areas .....	111
Table A.1	List of TTMSs with Missing Data in Urban Areas .....	139



Table A.2	List of TTMSs with Missing Data in Rural Areas .....	140
Table B.1	Assignment Results for TTMSs in North Florida with Full Variable Set...	141
Table B.2	Assignment Results for TTMSs in Central Florida with Full Variable Set	143
Table B.3	Assignment Results for TTMSs in South Florida with Full Variable Set...	144
Table B.4	Assignment Results for TTMSs in North Florida with Reduced Variable Set.....	146
Table B.5	Assignment Results for TTMSs in Central Florida with Reduced Variable Set.....	148
Table B.6	Assignment Results for TTMSs in South Florida with Reduced Variable Set.....	149
Table B.7	Assignment Results for Rural TTMSs in Single Peak Group with Full Variable set.....	151
Table B.8	Assignment Results for Rural TTMSs in Double Peak Group with Full Variable Set .....	152
Table B.9	Assignment Results for Rural TTMSs in Single Peak Group with Reduced Variable Set .....	154
Table B.10	Assignment Results for Rural TTMSs in Double Peak Group with Reduced Variable Set .....	155

## LIST OF FIGURES

FIGURE	PAGE
Figure 1.1 TTMS and PTMS Locations in Florida.....	5
Figure 2.1 Florida's Traffic Monitoring Sites Used in 2000.....	9
Figure 2.2 Process Used to Estimate AADT, K, D, and T Factors.....	10
Figure 2.3 EM Algorithm for Clustering via Gaussian Mixture Models.....	23
Figure 2.4 Neural Network Architecture .....	33
Figure 3.1 Framework of the Methodology.....	49
Figure 3.2 Buffer Areas Around TTMS in Urban Areas .....	52
Figure 3.3 Buffer Areas Around TTMS in Rural Areas .....	54
Figure 4.1 Historical Data Plot for Site 930099 .....	60
Figure 4.2 Historical Data plot of Site 930174.....	61
Figure 4.3 Historical Data Plot for Site 740047 .....	62
Figure 4.4 Universities with large enrollment in Florida.....	68
Figure 5.1 Classification of Sub-model Groups for Urban Areas .....	72
Figure 5.2 Classification of Sub-model Regions for North and Central Florida.....	73
Figure 5.3 Boundaries of Study Areas.....	74
Figure 5.4 Classification of Sub-model Groups for Rural Areas.....	84
Figure 5.5 Single-peak Pattern and Variables Describing Peaking Characteristics .....	85
Figure 5.6 Double-peak Pattern and Variables Describing Peaking Characteristics.....	86
Figure 5.7 Hourly Traffic Variations for Selected TTMSs.....	87
Figure 7.1 MSFs for Test Site 899921 and First Five Closest Matching Sites.....	106
Figure 7.2 Percentage Difference in the MSFs for Site 899921 and the Best Matches.	106
Figure 7.3 12-Month Average Percentage Differences in Urban Areas .....	109
Figure 7.4 Distributions of Variances for TTMSs in Urban Areas.....	110

Figure 7.5	Accumulated Percentage Distributions of Variances Urban Area TTMSs ..	110
Figure 7.6	12-Month Average Variances Plot for Rural Areas .....	111
Figure 7.7	Distributions of Variances for TTMSs in Rural Areas.....	112
Figure 7.8	Accumulated Percentage Distributions of Variances Rural Area TTMSs ...	112
Figure 7.9	$\bar{S}_i$ and $\bar{e}_i$ for Sites in North Florida .....	114
Figure 7.10	$\bar{S}_i$ and $\bar{e}_i$ for Sites in Central Florida.....	115
Figure 7.11	$\bar{S}_i$ and $\bar{e}_i$ for Sites in South Florida .....	115
Figure 7.13	$\bar{S}_i$ and $\bar{e}_i$ for Sites in Single Peak Model.....	116
Figure 7.12	$\bar{S}_i$ and $\bar{e}_i$ for Sites in Double Peak Model .....	116
Figure 7.14	Distribution of $e_{ij}$ for Sites in North Florida.....	118
Figure 7.15	Distribution of $e_{ij}$ for Sites in Central Florida .....	118
Figure 7.16	Distribution of $e_{ij}$ for Sites in South Florida.....	119
Figure 7.17	Distribution of $e_{ij}$ for Sites in Single-Peak Group .....	119
Figure 7.18	Distribution of $e_{ij}$ for Sites in Double-Peak Group .....	120
Figure 7.19	$\bar{e}_i$ Value for all TTMSs.....	121
Figure 7.20	Census Block Groups Not Covered by any TTMS Buffer.....	122
Figure 7.21	Census Block Groups with Special Demographic Characteristics in D1 & D4 .....	123
Figure 7.22	Census Block Groups with Special Demographic Characteristics in D2 & D3 .....	124
Figure 7.23	Census Block Groups with Special Demographic Characteristics in D5 & D7 .....	125
Figure 7.24	Census Block Groups with Special Demographic Characteristics in D6...	126

# CHAPTER I

## INTRODUCTION

### 1.1 Background

State Departments of Transportation (DOTs) routinely collect traffic data for use as inputs to different types of analyses, including planning, roadway design, pavement design, air quality assessment, roadway maintenance, funding allocation, etc. One of the most important types of data is traffic volume, which is often measured in terms of Annual Average Daily Traffic (AADT). AADT is defined for a roadway section as the total number of vehicle trips in both directions in one year, divided by the number of days in the year. Acquiring the actual AADT data, however, requires the collection of traffic counts continuously throughout a year, which is expensive. As such, DOTs could collect continuous traffic data for only a limited number of sites. For many other locations where traffic count data are required, 24- to 72-hour traffic counts are usually collected. Such traffic counts are referred to as short-term counts or coverage counts, from which average daily traffic counts (ADTs) are computed. These ADTs are then adjusted by seasonal factors (SFs) and axle correction factors (Axles) to estimate the AADT, as follows:

$$AADT = ADT \times SF \times Axle \quad (1-1)$$

where,

$AADT$  = estimate of typical daily traffic on a road segment for all days of the week, Sunday through Saturday, over the period of one year;

$ADT$  = average daily traffic, typically the average value of a 24- to 72-hour traffic count collected from Tuesday to Thursday;

$SF$  = seasonal factor that reflects traffic seasonal fluctuation; and

$Axle$  = axle correction factor that converts the counted number of axles to the number of vehicles.

Seasonal factors may be expressed as weekly and monthly factors. In Florida, weekly seasonal factors are used to account for traffic volume variations during a week, which are determined by interpolating between the monthly seasonal factors (MSFs) for two consecutive months. The 12 MSFs for a specific month at a particular location is derived from dividing the monthly average daily traffic (MADT) at a given location with its AADT.

## **1.2 Problem Statement**

It can be seen from Equation (1-1) that the accuracy of SFs has a direct impact on the AADT estimates. For instance, a 10% error in a SF will translate into a 10% error in AADT. Currently in Florida, seasonal factor groups are first derived from around 300 Telemetry Traffic Monitoring Sites (TTMSs) across the state and then assigned to the coverage count sites, commonly referred to as the Portable Traffic Monitoring Sites (PTMSs), based on the geographic location of a coverage count site and its functional classification. One major challenge with this approach is the difficulty in identifying a set of definable characteristics that are indicative of seasonal traffic patterns and can serve as the basis for objective assignment of SFs to PTMSs.

Roadway functional classifications (such as rural, urban, interstate, collector, and recreational) and locations have been recognized as possible influential factors (Sharma 1983; Sharma et al. 1986; Capparuccini 2008). There are other factors, such as demographics, socioeconomics, land use types, etc., that could also provide an

explanation for seasonal traffic variations. These factors, if understood and quantified, may potentially be exploited. They can aid in the assignment of seasonal factors from one or more TTMSs to a coverage count site or PTMS, which will reduce the data collection effort and improve the accuracy of AADT estimations. Thus far, however, the assignment of SFs still remains a research topic because of the challenges it poses regarding accuracy, transferability, and interpretability. This is because the different influential factors could affect the SFs of different months in different ways. As such, the influence of one factor on the SFs for different months needs to be quantified individually.

In a previous study in 2004, Zhao et al conducted a linear regression analysis with demographic, socioeconomic, and land use types, etc., data to identify possible explanatory variables for seasonal traffic fluctuations. The study was limited in geographical coverage. Only the data from TTMSs in the urban areas in Miami-Dade, Broward, and Palm Beach counties and the rural areas in North Florida have been analyzed. Four variables were identified as significant indicators for seasonal traffic fluctuations for the urban roads in southeast Florida. They are seasonal residents, tourists, retired people aged 65 to 75 with high income, and retail employment. For rural roads variables such as the functional classification for highways, seasonal households, agricultural employment, and truck factor were identified as potential explanatory variables. Li et al. (2006) developed a fuzzy decision tree to classify a count site based on the value of selected variables that were identified in regression analyses. However, to validate the assignment results requires the considerable effort of collecting additional monthly short-term counts.

As shown in Figure 1.1, the TTMS sites (about 300) across the state are represented by red diamond dots, while the PTMSs (over 6,000 of them) are represented by blue round dots. The inset clearly shows that many areas are left with no TTMSs that are remotely close, as is the case with the large area north of the urban region across the river. Under such a condition, the existing TTMSs may not adequately reflect the entire range of variation in demographics, socioeconomics, land use, and roadway characteristics of all locations. Therefore, the accuracy of the assignment method cannot be assured when the values of influential variables for a PTMS fall outside the range of those for TTMSs.

It can be concluded from the foregoing discussion that the successful development of a method to more accurately assign MSFs to PTMSs for both rural and urban areas statewide will require the following:

- An understanding of the factors affecting seasonal traffic patterns (due to the noticeable changes in climate from the north to the south, and the different types of local economies—these factors may differ depending on the area);
- Adequate TTMS data to ensure valid statistical analyses;
- Identification and quantification of the underlying factors as variables, and the statistical verification of those variables that have a link to MSFs;
- Successful development of a methodology for assigning a set of MSFs (obtained from TTMSs) to PTMSs. This methodology must be applicable to both urban and rural areas, which depends on the success of identifying variables that adequately explain the seasonal variations in traffic;

- Validation of results from the developed methodologies to evaluate the accuracy of the methodology;
- Identification of the best locations to install additional TTMSs that will maximize the accuracy of SFs.

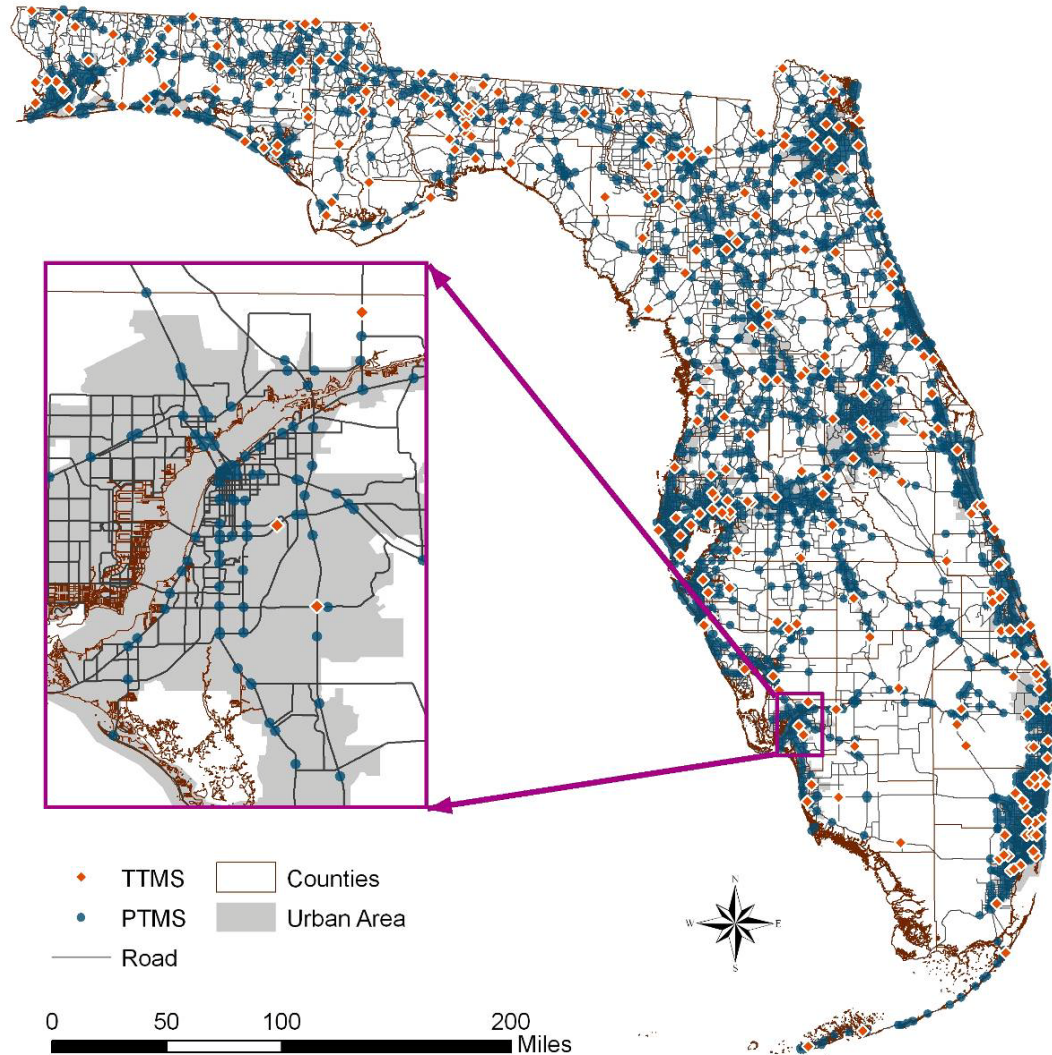


Figure 1.1 TTMS and PTMS Locations in Florida



### **1.3 Research Goals and Objectives**

The goal of this dissertation research is to improve the accuracy of existing seasonal factor estimation methods. This is accomplished through the following four objectives:

1. Identify possible explanatory variables that allow more accurate assignment of short-count sites to a given seasonal factor group.
2. Improve the accuracy of SF estimation by considering traffic measures (e.g., hourly volume patterns) as potential predictors of SFs.
3. Develop a methodology to assign established seasonal factor groups to short count sites based on the explanatory variables.
4. Identify the areas in which SFs can be more confidently estimated through better understanding of the spatial patterns of the TTMS influence areas.

### **1.4 Organization**

This dissertation consists of seven chapters. Chapter I introduces the background of the research, describes the research needs, and sets forth the research's goals and objectives.

Chapter II summarizes various methods for incorporating seasonal variations in the calculation of AADT, including conventional approaches such as statistical cluster analysis, geographic/functional assignment, and regression analysis, as well as machine learning techniques such as neural networks and genetic algorithms. The existing literature on assigning a short count site to a seasonal group is also reviewed.

Chapter III introduces the methodology framework of this research, including data imputation, variable selection, SF assignment for urban and rural areas, and describes the technique to be employed to evaluate the coverage of the TTMS sites.

Chapter IV describes the data preparation for both SFs and potential influence variables. A data imputation effort is employed to maximize the number of TTMSs that may be used in this study.

The possible identification of influential variables of traffic seasonality is discussed in Chapter V. For the urban areas, regression models are developed for northern Florida, central Florida, and southern Florida respectively. Rural area TTMSs are separately modeled based on their daily traffic patterns.

Chapter VI describes a preliminary investigation of a potential assignment procedure. A unique score is computed based on the influential variables to determine similarities between a PTMS and TTMSs.

The evaluation for the proposed assignment methodology is provided in Chapter VII. The evaluation of the distribution of current TTMSs is also discussed in this chapter.

Finally, Chapter VIII summarizes the major research results, draws conclusions, and offers recommendations for future research.

## **CHAPTER II**

### **LITERATURE REVIEW**

This chapter reviews past research efforts to enhance the factoring process for accuracy improvement in the estimation of traffic volume at a short-count station. As aforementioned, the factoring process consists of two parts: defining the TTMS groups and assigning the PTMS sites to a seasonal group. Up to the present, most research efforts have focused on seasonal factor (SF) grouping, and numerous approaches have been proposed to obtain better groupings in the factoring process. These approaches differ mainly in their data and processes, and include conventional approaches, such as statistical cluster analysis, geographic/functional assignment, and regression analysis, as well as machine learning techniques, such as neural networks and genetic algorithms. The existing literature on assigning a short-count site to a seasonal group is also reviewed. In the following sections, the current practice in Florida of estimating the SF at a given short-count station is first described. This is followed by a discussion of numerous modeling techniques relevant to the factoring process, including grouping and assignment procedures.

#### **2.1 Current Practice in Florida**

As aforementioned, the Florida Department of Transportation (FDOT) collects traffic count data from about 300 telemetry traffic monitoring sites (TTMSs) located throughout the state. These TTMSs are continuous counters, and their true Annual Average Daily Traffic (AADT) and SFs are available. There are more than 7,000 portable traffic monitoring sites (PTMSs), where a short-term traffic count (e.g., 24- to

72-hour) is conducted only a few days in a year. Figure 2.1 shows the type and amount of permanent and portable TMSs used in Florida in 2000.

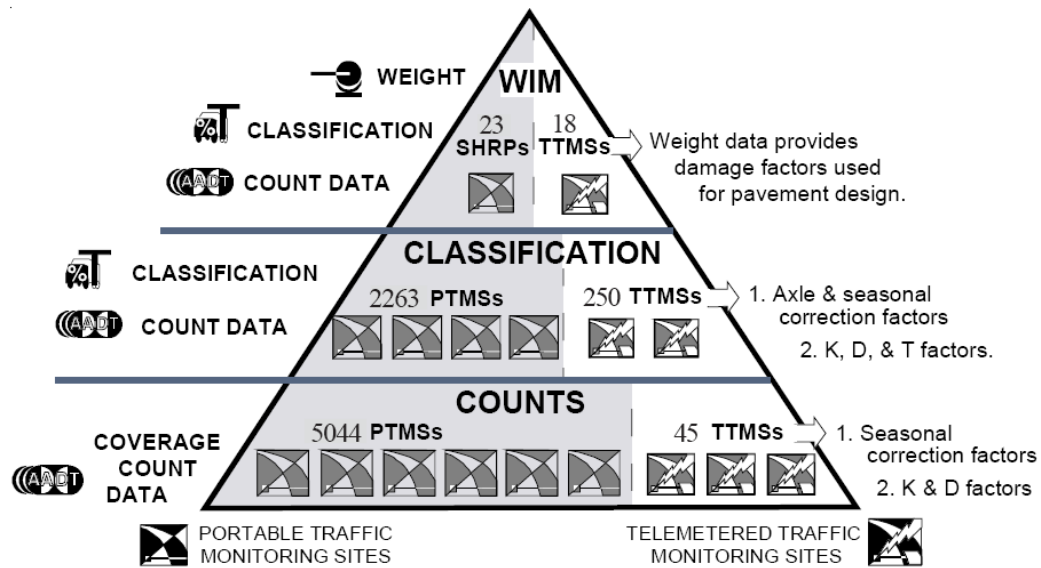


Figure 2.1 Florida's Traffic Monitoring Sites Used in 2000

To estimate AADT from average daily traffic (ADT), FDOT applies the following equation (FDOT 2002):

$$AADT = ADT \times SF \times Axle \quad (2-1)$$

where,

*AADT* = estimate of typical daily traffic on a road segment for all days of the week, Sunday through Saturday, over the period of one year;

*ADT* = average daily traffic, typically the average value of a 24- to 72-hour traffic count collected from Tuesday to Thursday;

*SF* = seasonal factor that reflects traffic seasonal fluctuation; and

*Axle* = axle correction factor that converts the counted number of axles to the number of vehicles.

The monthly seasonal factors are the main object of this study. They are obtained from approximately 300 TTMSs in Florida. These TTMSs are grouped into 178

SF categories based on similarities in monthly variation patterns. The SFs for a specific category are obtained by averaging the SFs from the TTMSs in the group. Next, each PTMS is assigned to one of these factor groups based on consideration of roadway function classification, spatial proximity between the PTMS and nearby TTMSs, and the judgment of the analyst. The AADT for a given coverage count location is then estimated using the SFs of the assigned factor group. Figure 2.2 shows the procedure employed by FDOT for the estimation of AADT and conversion factors, such as  $K$  factor, directional factor ( $D$ ), and truck factor ( $T$ ). The FDOT, however, desires a more data-driven, objective method in order to improve the accuracy of the SF estimation for short-term count sites.

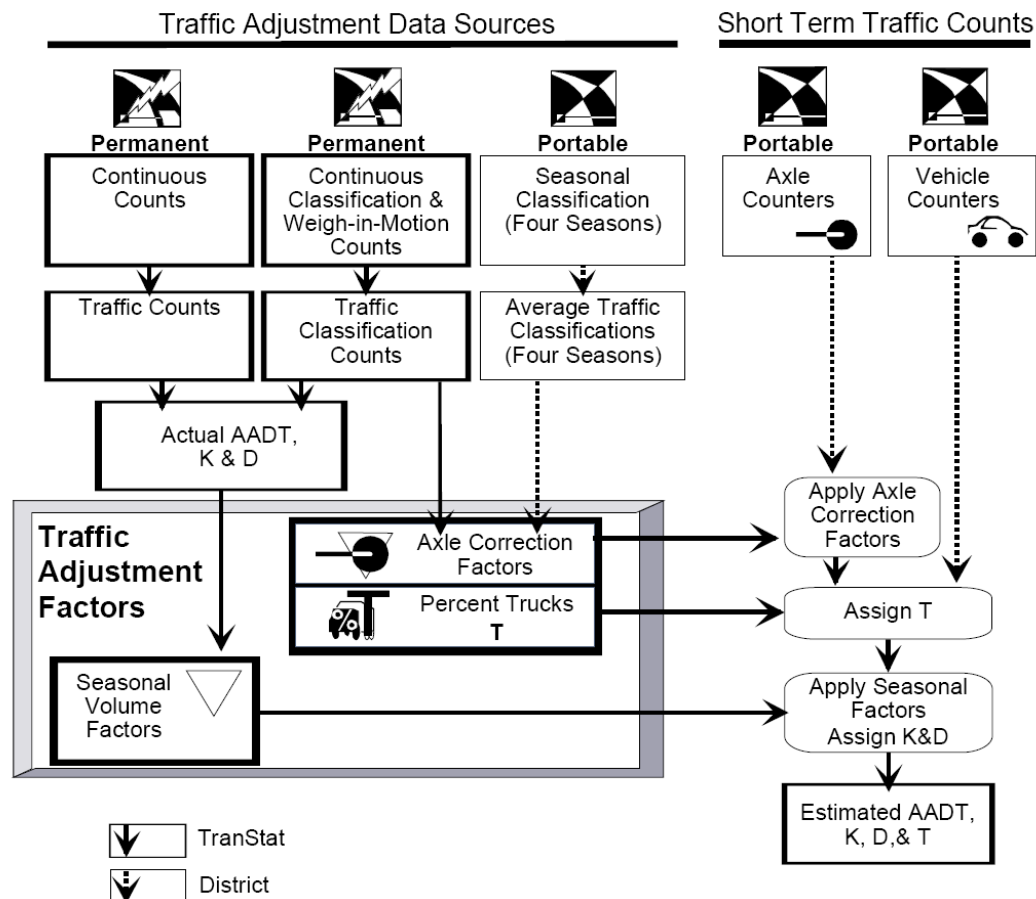


Figure 2.2 Process Used to Estimate AADT, K, D, and T Factors

## 2.2 Cluster Analysis

The purpose of cluster analysis is to help classify TTMSs into categories, such that the sites within the same category share a similar seasonal pattern. Categories are defined based on the measure of similarities, which are usually designated by some sort of distance; thus, the variables selected to support cluster analysis are usually data from which similarities may be measured. In the context of SF grouping, input to cluster analysis is usually 12 MSFs for each TTMS.

There are several types of statistical clustering methods for the grouping of objects. Among these methods, nonparametric methods, including agglomerative hierarchical clustering and nonhierarchical clustering, have been typically used to determine SF groups. The parametric model-based clustering approach, however, has now become popular for the determination of cluster membership in a variety of disciplines. The following sections describe these applications.

### 2.2.1 *Nonparametric Agglomerative Hierarchical Clustering Methods*

Nonparametric clustering classifies objects into categories based on a measure of similarity between clusters. The basis of nonparametric clustering is that groups correspond to modes of an unknown distribution function. Consequently, the goal is to estimate the modes and assign each observation to the domain of attraction for a mode. The nonparametric agglomerative hierarchical cluster analysis process (referred to hereafter as hierarchical clustering analysis) begins by treating each observation as a cluster unto itself. The two closest clusters, determined by a specific similarity measure, are merged to form a new cluster that replaces the two older clusters. Merging of the two closest clusters is then repeated until only a single cluster remains. The hierarchical

clustering analysis method organizes objects so that one cluster may be entirely contained within another, and ensures that no overlap between the clusters is allowed.

In a cluster analysis, similarity or closeness between two p-dimensional observations is usually measured by Euclidean distance, as shown in Equation (2-2):

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} = \sqrt{(x - y)^T (x - y)} \quad (2-2)$$

where  $x = [x_1, x_2, \dots, x_p]^T$  and  $y = [y_1, y_2, \dots, y_p]^T$ .

Hierarchical clustering analysis is used to determine the groupings of the data. In the case of the SF process, the seasonality observed at each TTMS from month to month is considered in the grouping process. The basic intent of the clustering analysis is to identify patterns of variation in order to provide the analyst the knowledge and insight to develop grouping criteria for the conversion of short counts to AADT. To this end, multiple clustering methods have been employed in SF analysis. Sharma and Werner (1981; 1983) applied a hierarchical clustering method to group 45 TTMSs in Alberta, Canada based on their 12 monthly factors. The Scheffe's S-method of multiple comparisons of group means was used to determine the optimal number of groups, ranging from 6 to 10 (obtained from the hierarchical process), each containing more than two counters. The results showed that eight to nine groups were desirable. A subsequent study conducted by Sharma and Allipuram (1993) applied the same method to group 61 TTMSs in Alberta, by using the data collected in 1989, and obtained a total of seven cluster groups. In their later work, Sharma et al. (1996) concluded that the accuracy of the AADT estimation is more sensitive to the correctness of PTMS site assignment than the duration of short-term counts.

Aunet (2000) used cluster analysis to examine the variation in Wisconsin's traffic data. The results revealed that seasonal patterns remained stable over time. Additionally, although significant variations existed in the MSFs for the permanent count stations classified in the same group, SF groups could be generally defined according to roadway functional classifications (i.e., urban, rural, and recreational).

The process of cluster analysis is completely driven by the variability in the MSFs. Two apparent advantages of cluster analysis are: 1) it allows for the independent determination of "similarity" between groups, thus making the groups less subject to bias; and 2) it is able to identify travel patterns that may not be intuitively obvious to the analyst (TMG 2001). As such, this method helps agency staff investigate road groupings that might not otherwise be examined, which in turn may lead to more efficient and accurate grouping of factors, thereby providing new insights into travel patterns.

Hierarchical clustering analysis also has its shortcomings in that it provides no definable characteristics or criteria upon which to form groups. Consequently, although well-adopted in the practice, this type of clustering application suffers from the following two major weaknesses (TMG 2001):

1. *Lack of theoretical guidelines for establishing the optimal number of groups.* It is often difficult to determine how many groups should be formed. The difficulty is in determining at what point the sequential merging process should stop. Unfortunately, the "optimal" number of groups cannot be determined mathematically. Consequently, the results of the cluster analysis may not be the ultimate answer; modifications are to be expected. Statistical models may be used to better understand the variation of data by identifying the seasonal fluctuation



patterns and eliminating stations with extreme variations. However, the development of the final factor groups must account for variability, and also include characteristics that define the groups, to allow for the assignment of short counts to the groups in the subsequent process. The knowledge of other criteria, e.g., functional class, geography, topography, degree of urbanization, etc., as well as the use of analytical judgment, are still necessary in interpreting the results.

2. *Lack of theoretical guidelines for group assignment.* Often, the groups formed cannot be adequately defined because the clustering procedure only considers the traffic variability at TTMSs, which are not directly applicable to the short counts. Although plotting the sites that fall within a specific cluster group on a map is sometimes helpful when attempting to define a given group output, in some cases, the purely mathematical nature of the clustering process simply does not lend itself to easily identifiable groups. When no criteria for the assignment of short counts to the groups are defined by the hierarchical clustering analysis, the use of descriptive analysis and of functional class, geography, or topography is necessary in order to provide additional criteria for assignment formation.

### 2.2.2 *Nonhierarchical Clustering Methods*

Nonhierarchical clustering refers to methods that are commonly known as  $k$ -mean methods or partitioning clustering methods that place each object in only one cluster. The methods usually begin by randomly partitioning individual items into  $k$  groups to avoid any overt biases. Items are then assigned to clusters with the nearest medium or mean. The number of clusters ( $k$ ) may be given either as a priori or determined by the algorithm. When  $k$  is unknown, nonhierarchical methods are generally repeated for

several values of  $k$ . The optimal value evaluated by the criterion associated with each nonhierarchical method is then selected as the desired number of groups. Since the  $k$  clusters are generated simultaneously, the resulting classification is non-hierarchical.

Flaherty (1993) used the hierarchical clustering method and the  $k$ -means method to analyze the monthly factor data collected over a five-year period from 28 permanent traffic counters installed in Arizona. The  $k$ -means algorithm was used to produce clusters of prescribed numbers, varying from two to nine, by maximizing the ratio of between-cluster variation to within-cluster variation. This approach was analogous to a one-way ANOVA seeking the largest  $F$ -value by reassigning objects.

Flaherty concluded that similarity in the patterns of the monthly factors was more a function of geography and topography than the functional classification of the highways where the count stations were located, and that the population of the surrounding area did not appear to be an explanatory factor for the factor groups. Flaherty also found that four clusters were the best and most stable of all the variations used in the analysis. Similar to hierarchical clustering methods, difficulties were encountered as to how to appropriately interpret the groups resulting from nonhierarchical clustering methods, as well as how to conduct short-count site assignments.

### *2.2.3 Model-Based Gaussian Cluster Analysis*

Model-based clustering assumes that each seasonal factor group may be represented by a density function that is a member of some parametric family (e.g., the multivariate normal [Gaussian] family), and that the associated parameters may be estimated from observations (Fraley 1998). The fundamental concept of model-based clustering analysis is to determine the probabilistic density function for the  $k^{\text{th}}$  seasonal

factor group by estimating the first two orders of statistics, i.e., the  $p$ -dimensional mean vector ( $\boldsymbol{\mu}_k$ ) and the  $p \times p$  covariance matrix ( $\boldsymbol{\Sigma}_k$ ). If  $\boldsymbol{\Sigma}_k$  is expressed in terms of its eigenvalue decomposition, i.e.,  $\boldsymbol{\Sigma}_k = \lambda_k \mathbf{D}_k \mathbf{A}_k \mathbf{D}_k^T$ , where superscript T denotes matrix transpose and  $\mathbf{D}_k$ ,  $\lambda_k$ , and  $\mathbf{A}_k$  govern the orientation, the volume, and the shape for the  $k^{\text{th}}$  seasonal factor group, a systematic analysis may be performed by treating these geometric features as different parameters. Examples of models include  $\lambda \mathbf{I}$ ,  $\lambda_k \mathbf{I}$ ,  $\lambda \mathbf{DAD}^T$ , etc.

In model-based methods, a maximum-likelihood criterion is used to merge groups. Two approaches are commonly applied in model-based clustering analysis: the classification approach and the mixture approach (Dundar 2002). The classification approach aims at maximizing the likelihood over the mixture parameters and identifying the group to which each sample belongs, while the mixture approach merely aims at maximizing the likelihood over the mixture parameters. Different from a discrete value indicating the cluster in the classification approach, a probability is obtained for a given observation that is classified to a specific group in the mixture approach, and the sum of the probabilities is equal to 1. Compared to non-parametric clustering methods, the ability to estimate the number of groups is an important advantage of the model-based approach. Fraley and Raftery (1998) employed the Bayesian Information Criterion (BIC) with a penalty for the complexity of the model subtracted from the mixture log likelihood to find the optimal number of clusters. The BIC may be used to systematically compare models with different parameterizations, different numbers of seasonal factor groups, or both.

The background of the model-based cluster analysis for SF grouping is briefly described as follows (Tantrum 2002). First, assume that there are  $G$  seasonal factor groups in a given study area. For each permanent count station  $i$ , the MSF for every month in a year (or a linear combination of these factors) and other characteristics form a  $p$ -dimensional vector,  $\mathbf{x}_i$ . Given  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ , where  $n$  is the number of TTMSs, the density function for the  $i^{\text{th}}$  TTMS from the  $k^{\text{th}}$  seasonal factor group is  $f_k(\mathbf{x}_i | \boldsymbol{\theta}_k)$ , with some unknown vector for parameters  $\boldsymbol{\theta}_k$ , where  $\boldsymbol{\theta}_k$  consists of a mean vector  $\boldsymbol{\mu}_k$  of length  $p$  for the mean in each dimension and a  $p \times p$  covariance matrix  $\boldsymbol{\Sigma}_k$ . Assuming  $f_k(\mathbf{x}_i | \boldsymbol{\theta}_k)$  is multivariate normal (Gaussian), the probability density function has the following form:

$$f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k)\right\}}{(2\pi)^{p/2} |\boldsymbol{\Sigma}_k|^{1/2}} \quad (2-3)$$

Each SF group forms an ellipsoid that is centered at its mean  $\boldsymbol{\mu}_k$  with its geometric characteristics determined by the covariance matrix  $\boldsymbol{\Sigma}_k$ . The covariance matrix may be expressed in terms of its eigenvalue decomposition as follows (Banfield 1993):

$$\boldsymbol{\Sigma}_k = \mathbf{D}_k \boldsymbol{\Lambda}_k \mathbf{D}_k^T \quad (2-4)$$

where,

$\mathbf{D}_k$  = orthogonal matrix of eigenvectors, which determines the orientation of  $\boldsymbol{\Sigma}_k$ ;

and

$\boldsymbol{\Lambda}_k$  = a diagonal matrix with the eigenvalues of  $\boldsymbol{\Sigma}_k$  on the diagonal, which specifies the size and shape of the density contours.

$\mathbf{\Lambda}_k$  may be furthered decomposed as follows:

$$\mathbf{\Lambda}_k = \lambda_k \mathbf{A}_k \quad (2-5)$$

where,

$\lambda_k$  = the first eigenvalue of  $\mathbf{\Sigma}_k$ , which specified the volume of the  $k^{\text{th}}$  seasonal factor group; and

$$\mathbf{A}_k = \text{diag}[\alpha_{1k}, \dots, \alpha_{pk}]^T, \quad 1 = \alpha_{1k} \geq \alpha_{2k} \geq \dots \geq \alpha_{pk} > 0.$$

Consequently, Equation (2-4) becomes:

$$\mathbf{\Sigma}_k = \lambda_k \mathbf{D}_k \mathbf{A}_k \mathbf{D}_k^T \quad (2-6)$$

$\mathbf{D}_k$ ,  $\lambda_k$ , and  $\mathbf{A}_k$  govern the orientation, the volume occupied by the cluster in  $p$ -space, and the shape for the  $k^{\text{th}}$  seasonal factor group, respectively. By treating these geometric features as independent sets of parameters, a systematic analysis may be carried out by constructing models with different parameters. Table 2.1 shows the models proposed in the context of cluster analysis for covariance matrices (Fraley 2002). In this table, the coded geometric characteristics of the model are identified. For example, EVI denotes a model in which the volumes of all clusters are equal (E), the shapes of the clusters may vary (V), and the orientation is the identity (I). Clusters in this model consist of diagonal covariances with orientation parallel to the coordinate axes. Parameters that are associated with characteristics designated by E or V may be determined from the data.

It is claimed by Kamvar (2002) that the common heuristic agglomerative clustering algorithms, e.g., average linkage, single linkage, complete linkage, and Ward's method, are equivalent to a model-based method. More specifically, under the assumption that every  $\mathbf{\Sigma}_k$  is an independently and identically distributed (IID) normal

variant, i.e.,  $\Sigma_k = \lambda I$  (the EII model in Table 2.1), every SF group would have the same shape, volume, and orientation since  $\Sigma_k = \Sigma = \lambda \mathbf{DAD}^T$ , which is commonly known as the Ward's method of the conventional clustering approach (Kamvar 2002). The model for the composite of the clusters is usually formulated by the classification likelihood approach or the mixture likelihood approach. The following sections describe the background of these two model-based approaches.

Table 2.1 Available Parameterizations of Covariance Matrix

Model	Identifier	Distribution	Volume	Shape	Orientation
$\lambda \mathbf{I}$	EII	Spherical	Equal	Equal	NA
$\lambda_k \mathbf{I}$	VII	Spherical	Variable	Equal	NA
$\lambda \mathbf{A}$	EEI	Diagonal	Equal	Equal	Coordinate Axes
$\lambda_k \mathbf{A}$	VEI	Diagonal	Variable	Equal	Coordinate Axes
$\lambda \mathbf{A}_k$	EVI	Diagonal	Equal	Variable	Coordinate Axes
$\lambda_k \mathbf{A}_k$	VVI	Diagonal	Variable	Variable	Coordinate Axes
$\lambda \mathbf{DAD}^T$	EEE	Ellipsoidal	Equal	Equal	Equal
$\lambda_k \mathbf{D}_k \mathbf{A}_k \mathbf{D}_k^T$	VVV	Ellipsoidal	Variable	Variable	Variable
$\lambda \mathbf{D}_k \mathbf{AD}_k^T$	EEV	Ellipsoidal	Equal	Equal	Variable
$\lambda_k \mathbf{D}_k \mathbf{AD}_k^T$	VEV	Ellipsoidal	Variable	Equal	Variable

### 2.2.3.1 Classification Likelihood Approach

In the classification likelihood approach, the objective is to identify the parameters  $\theta$  and labels  $\gamma$  that maximize the following likelihood function:

$$L_C(\theta_1, \dots, \theta_n; \gamma_1, \dots, \gamma_n | \mathbf{x}) = \prod_{i=1}^n f_{\gamma_i}(\mathbf{x}_i | \theta_{\gamma_i}) \quad (2-7)$$

where  $\gamma = (\gamma_1, \dots, \gamma_n)^T$  denotes the identifying labels for the classification, i.e.,  $\gamma_i = k$  for the  $i^{\text{th}}$  TTMS that is classified to the  $k^{\text{th}}$  seasonal factor group. The presence of the class labels in the classification likelihood introduces a combinatorial aspect that makes exact maximization impractical (Fraley 2002). Consequently, model-based hierarchical clustering methods are commonly implemented since they usually provide a good

approximation of the optimal grouping and are relatively easy to compute (Fraley 1996). The process is to successively merge a pair of clusters that yields the greatest increase in maximum likelihood, as is expressed in Equation (2-7). The resulting partitions are suboptimal since the final results may not be globally optimal.

### 2.2.3.2 Mixture Likelihood Approach

The objective function in the mixture likelihood clustering approach is to identify the parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\tau}$  that maximize the following likelihood function:

$$L_M(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_G; \tau_1, \dots, \tau_G | \mathbf{x}) = \prod_{i=1}^n \sum_{k=1}^G \tau_k f_k(\mathbf{x}_i | \boldsymbol{\theta}_k) \quad (2-8)$$

where  $\tau_k$  is the probability that a TTMS belongs to the  $k$ th seasonal factor group that meets the following constraints:

$$\tau_k \geq 0 \quad (2-9)$$

$$\sum_{k=1}^G \tau_k = 1 \quad (2-10)$$

In the mixture likelihood approach, it is assumed that there exists a finite set of  $G$  seasonal factor groups, and each TTMS is associated with an indicator vector  $\mathbf{z}_i$  of length  $G$ , whose components are all zero, except for one indicating the classification. The key difference between the classification and mixture approaches is that, in the former, each TTMS is assigned to a unique cluster, while in the latter, each TTMS is assigned with a probability of originating from each SF group. Moreover, the mixture approach allows the uncertainties associated with the class membership of the observations to be estimated. The equivalent log-likelihood function of Equation (2-8) is shown below:

$$l_M(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_G; \tau_1, \dots, \tau_G | \mathbf{x}) = \sum_{i=1}^n \ln \left( \sum_{k=1}^G \tau_k f_k(\mathbf{x}_i | \boldsymbol{\theta}_k) \right) \quad (2-11)$$

Equation (2-11) may be optimized over  $\tau_k$ ,  $\boldsymbol{\mu}_k$ , and  $\boldsymbol{\Sigma}_k$  using the expectation-maximization (EM) algorithm. The EM algorithm is a general approach to maximum-likelihood estimation (MLE) in the presence of incomplete data. The complete data are considered to be  $\mathbf{y}_i = (\mathbf{x}_i, \mathbf{z}_i)$ , where  $\mathbf{z}_i = (z_{i1}, \dots, z_{iG})$  constitutes the “missing” data, and  $z_{ik}$  is equal to one for a TTMS ( $\mathbf{x}_i$ ) belonging to SF group  $k$  and is zero otherwise. Equation (2-11) is thus considered the log-likelihood function from the observed data,  $\mathbf{x}_i$ . Assuming that each  $\mathbf{z}_i$  is independent and identically distributed according to a multinomial distribution of one drawn from  $G$  seasonal factor groups with unknown probabilities  $\tau_1, \dots, \tau_G$ , the probability mass function for the  $i^{\text{th}}$  TTMS (i.e.,  $\mathbf{x}_i$ ) belonging to SF group  $k$  may be expressed as follows (Dundar 2002):

$$f(\mathbf{z}_i) = \frac{1!}{0! \dots 1! \dots 0!} \tau_1^0 \dots \tau_{k-1}^0 \tau_k^1 \tau_{k+1}^0 \dots \tau_G^0 = \tau_k \quad (2-12)$$

Assuming the probability density function for  $\mathbf{x}_i | \mathbf{z}_i$  (i.e.,  $\mathbf{x}_i$  given  $\mathbf{z}_i$ ) as

$$f(\mathbf{x}_i | \mathbf{z}_i) = \prod_{k=1}^G f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_{ik}} \quad (2-13)$$

The probability density function for  $\mathbf{y}_i$  can be obtained by combining Equations (2-12) and (2-13) as follows:

$$f(\mathbf{y}_i) = f(\mathbf{x}_i | \mathbf{z}_i) \times f(\mathbf{z}_i) = \prod_{k=1}^G f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_{ik}} \tau_k \quad (2-14)$$

Under the condition that  $z_{ik}$  is equal to one for  $\mathbf{x}_i$  belonging to SF group  $k$  and zero otherwise, Equation (2-14) may be generalized as follows:



$$f(\mathbf{y}_i) = \prod_{k=1}^G \left( f_k(\mathbf{x}_i | \boldsymbol{\theta}_k, \tau_k) \tau_k \right)^{z_{ik}} \quad (2-15)$$

For a total of  $n$  TTMSs, Equation (2-15) may be written as

$$f(\mathbf{y}) = \prod_{i=1}^n \prod_{k=1}^G \left( f_k(\mathbf{x}_i | \boldsymbol{\theta}_k, \tau_k) \tau_k \right)^{z_{ik}} \quad (2-16)$$

The resulting complete-data log-likelihood is as follows:

$$l(\boldsymbol{\theta}_k, \tau_k, z_{ik} | \mathbf{x}) = \sum_{i=1}^n \sum_{k=1}^G z_{ik} \times \ln \left( \tau_k f_k(\mathbf{x}_i | \boldsymbol{\theta}_k) \right) \quad (2-17)$$

Let  $\hat{z}_{ik}$  denote the condition expectation of  $z_{ik}$  given  $\mathbf{x}_i$  and associated parameter values, i.e.,  $\hat{z}_{ik} = E[z_{ik} | \mathbf{x}_i, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_G]$ , and  $z_{ik}^*$  denote the value of  $\hat{z}_{ik}$  at a maximum of Equation (2-13), which is the conditional probability that the  $i^{\text{th}}$  TTMS belongs to group  $k$ . Figure 2.3 illustrates the EM algorithm for clustering via Gaussian mixture models (Fraley 1998). The EM algorithm alternates between two steps: an “E-step” and an “M-step.” During the E-step, values of  $\hat{z}_{ik}$  are computed from the data with the current parameter estimates. At the M-step, the complete likelihood for Equation (2-17) with each  $z_{ik}$  replaced by its current conditional expectation  $\hat{z}_{ik}$  is maximized with respect to the parameters.

The EM algorithm contains the following limitations (Fraley 1998):

- Unless starting with reasonable initial values, the rate of convergence may be slow.
- The number of conditional probabilities associated with each TTMS equals the number of components in the mixture. As a result, it is not practical for models with a large number of SF groups.

- EM breaks down when the covariance matrix corresponding to one or more SF groups is ill-conditioned, i.e., singular or near singular.

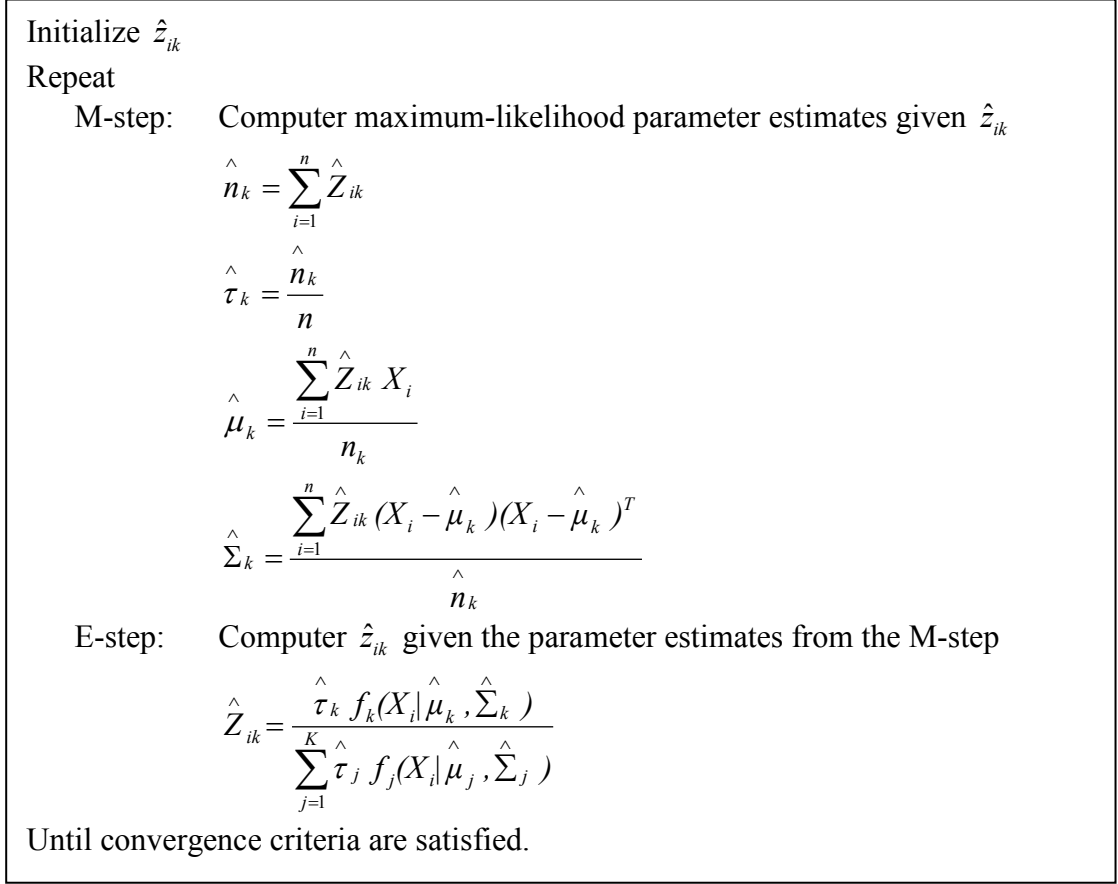


Figure 2.3 EM Algorithm for Clustering via Gaussian Mixture Models

There are two key issues in the clustering analysis: the selection of the clustering method, such as those presented in Table 2.1, and the determination of the number of clusters. To this end, the BIC, defined as follows, is applicable to finding the maximum mixture likelihood (Fraley 1998):

$$BIC = 2L - r \log(n) \quad (2-18)$$

where,

$L$  = log-likelihood of the model,

$r$  = total number of parameters to be estimated in the model, and

$n$  = number of TTMSs.

The number of clusters is not considered an independent parameter for the purpose of computing the BIC. The “likelihood” value cannot be used directly to evaluate a model since the fit of a mixture model to a given data improves as more terms are added to the model. In the expression of BIC, a term is added to the log-likelihood to penalize the complexity of the model. Consequently, the BIC allows for a smaller number of groups than that of the log-likelihood.

### **2.3 Geographic/Functional Assignment**

The method documented in the *Guide for Traffic Volume Counting Manual* (Bureau of Public Roads 1965) involves a manual ranking system. Using this method, monthly traffic factors of permanent count stations and the ratio of the AADT to the average weekday traffic of the month are sorted in ascending order. For each month, a group of counters is determined so that the difference between the smallest and the largest factors does not exceed 0.2. The final grouping of counters is manually examined to ensure as many counters as possible fall into the same group each month.

Bellamy (1978) described a subjective classification system for determining the grouping of a site. Four classes were identified: urban-commuter, low flow (< 1000 vehicle/day) non-recreational rural, rural long-distance, and recreational. Sharma (1983) proposed a method to classify rural roads based on trip purpose and trip length information collected from past origin-destination (OD) surveys by the Ministry of Transportation in Alberta, Canada. Based on the daily traffic patterns collected in 1978 or 1977 from a total of 45 counter sites, five predominant road uses were identified: commuter, commuter-recreational, commuter-recreational-tourist, tourist, and highly

recreational. Three typical hourly traffic patterns were also identified: commuter, partially commuter, and non-commuter. Trip purpose data, as well as trip length data, were obtained from external station surveys. Trip purpose data were then categorized into two groups, work-business and social-recreation, to verify the temporal volume variations. Cumulative trip length distribution information was used to classify roads for serving mainly regional, interregional, or long-distance travel. OD surveys, however, are often economically impractical, except in cases of corridor or inter-city travel studies. The same procedure was used to examine the data from 52 sites in Alberta, and the grouping was tested on the data from 28 sites in Saskatchewan, Canada (Sharma et al. 1986). Eight road classes were subsequently defined: regional commuter, regional recreational and commuter, interregional, long distance, long distance and recreational, highly recreational, rural commuter and business, and special.

Roadway functional classification and locations have also been recognized as possible influential factors of seasonality. Faghri and Hua (1995) classified roads as urban/rural, recreation/non-recreational, and recreational–arterial/otherwise, based on their physical and functional characteristics. Such classification of traffic characteristics, however, is difficult to obtain for large urban areas due to the dispersion and mixing of different types of activity centers, making it unlikely that a particular type of trip will be dominant on a given road.

Ritchie (1986) proposed a statistical framework to analyze statewide traffic count data. This approach incorporated the seasonal effect on traffic volumes by first stratifying the highway system according to geographic region and functional classification. The strata with similar seasonal patterns were then combined. Using the

data collected from 1980 to 1984 in Washington, seven groups were obtained: rural interstates, urban roads, other rural roads in the northeastern, southeastern, northwestern, and southwestern parts of Washington, and central mountain passes. The following regression model was then calibrated to estimate SFs for each group:

$$\text{Seasonal Factor}_i = \frac{\text{AADT}}{\text{VOL}_i} = \beta + \varepsilon \quad (2-19)$$

where,

$\text{VOL}_i$  = average 24-hour short-count volumes calculated from the 72-hour

Tuesday through Thursday counts for month  $i$ ;

$\varepsilon$  = error term whose variance was considered a constant; and

$\beta$  = regression coefficient, which was interpreted as the estimated SF for a specific factor group for a given month.

The Delaware Department of Transportation (DelDOT) utilized the procedure suggested in the FHWA Traffic Monitoring Guide (TMG) and categorized permanent count stations according to their monthly coefficient of variations (MCVs) into the following four groups (Faghri et al. 1986): urban ( $\text{MCV} < 10\%$ ), rural ( $10\% \leq \text{MCV} < 25\%$ ), recreational ( $25\% \leq \text{MCV} < 35\%$ ), and predominantly recreational ( $\text{MCV} > 35\%$ ). The MCVs were determined using the following formula, where MADT is the monthly AADT:

$$\text{MCV} = \frac{\sqrt{\frac{1}{2} \sum_{i=1}^{12} (\text{MADT} - \text{AADT})^2}}{\text{AADT}} \quad (2-20)$$

Another method by Kentucky State applied an approach to factor short-term vehicle classification counts by simultaneously considering both roadway types and

vehicle classifications (Stamatiadis et al. 1997). Eighty-four combined-month-and-day-of-week (CMDW) factors, which were developed for each day of the week and month of the year (i.e., 7 days  $\times$  12 months), were developed for four different types of roadways: rural interstates and parkways; urban interstates and parkways, rural non-interstates and non-parkways, and urban non-interstates and non-parkways (for each of the 15 vehicle types). The preliminary validation showed that more accurate AADT estimates were obtained when each vehicle type was factored alone. The estimates for different vehicle types were then added to obtain the overall AADT for a given roadway segment.

## 2.4 Regression Analysis

Regression techniques may be used as a tool to analyze the relationship between seasonal variations in traffic volume and several predictors. The variables used generally include those that represent the physical and functional characteristics or their combinations. Dummy regressors are used to represent these characteristics as the “yes/no” type of variable. The regression model is commonly defined in a linear form as follows:

$$f_{sm} = \alpha_{0m} + \alpha_{1m}x_1 + \alpha_{2m}x_2 + \dots \quad (2-21)$$

where,

$f_{sm}$  = seasonal factor in month  $m$ , and

$x_i$  = dummy variable that takes the value of 0 and 1 ( $i = 1, 2, \dots$ ).

Faghri and Hua (1995) concluded that urban/rural, recreation/non-recreational, and recreation-arterial/otherwise variables were statistically significant and could provide better results than cluster analysis.

Regression analysis was also often used to estimate AADT directly to avoid the use of SFs. Lam *et al.* (2000) and Erhunmwunsee (1991) selected the independent variables for a short-period count at designated stations, during a specific time frame. Erhunmwunsee concluded that the period with its midpoint centered at 3:00 PM was the best period in a day to begin a short-term count, and that the best month to conduct short-term counts was April, followed by June and October.

Seaver et al. (2000) proposed a statistical procedure utilizing principal component analysis, multivariate regression, regression clustering, and multiple regression analysis to model ADT on rural local roads. Data collected from 80 randomly selected counties in Georgia were utilized in the model's development. The procedure consisted of the following steps:

- Apply principal component analysis to identify  $p$  principal components ( $y_1, y_2, \dots, y_p$ ) from  $n$  initial independent variables ( $x_1, x_2, \dots, x_n$ ) for each paved (Road Type 4) and unpaved (Road Type 5) rural road in the metropolitan statistical area (MSA) and non-MSA.
- Apply multivariate regression to find the principal variables from the  $n$  initial independent variables ( $x_1, x_2, \dots, x_n$ ) that correlate with the principal components ( $y_1, y_2, \dots, y_p$ ) identified in the first step.
- Apply regression clustering to determine strata for each road type in both MSA and non-MSA by using the ADT in a county as the dependent variable and the principal variables identified in Step 2 as the regressors.
- Perform a multiple regression on the data within each cluster.

The advantage of this method is that all independent variables used in the procedure for developing the models were obtained from the U.S. census. The time and cost for obtaining the data were subsequently reduced. On the other hand, the disadvantage is that the census may not be up-to-date, and data verification is needed. The statistical procedure proposed by Seaver et al. may be applicable to grouping TTMSs into seasonal clusters for AADT estimates, provided the data for the independent variables at the TTMS level are available.

Zhao and Chung (2001) performed various multiple linear regression analyses to investigate factors affecting AADT estimates in Broward County, Florida. Geographic Information System (GIS) technology was utilized to compile intensive land use and accessibility measures. After outliers were removed, four models were calibrated. Two variables, functional classification and number of lanes, were found to be the most significant predictors for estimating AADTs. Other land use variables, including direct access to expressway, employment size in the buffer area surrounding a given count station, distance to spatial mean centers of population, and regional accessibility to employment centers, were also found to be significant.

Davis (1997) applied the weighted least-squares regression to calibrate the following model for traffic counts:

$$y_t = \mu + \sum_{i=1}^{12} \Delta_{t,i} m_{k,i} + \sum_{j=1}^7 \delta_{t,j} w_{k,j} + \varepsilon_t \quad (2-22)$$

where,

$y_t$  = natural logarithm of the traffic count on day  $t$ ;

$\mu$  = expected log traffic count on a typical day;

$\Delta_{t,i} = 1$ , if the count was made during month  $i$  ( $i = 1, \dots, 12$ ), and 0 otherwise;



$m_{k,i}$  = correction term for month  $i$ , characteristic of factor group  $k$  ( $k = 1, 2$ , or  $3$ );

$\delta_{t,j} = 1$ , if the count was made on day-of-week  $j$  ( $j = 1, \dots, 7$ ), and  $0$  otherwise;

$w_{k,j}$  = correction term for day-of-week  $j$ , characteristic of factor group  $k$ ; and

$\varepsilon_t$  = random error (residual).

After eliminating missing and imputed data from the traffic data collected in 1992, a total of 50 TTMSs classified into three factor groups by the Minnesota Department of Transportation (MnDOT) were included in the model's development. The mean-value ( $\mu$ ), monthly ( $m_{k,i}$ ), and day-of-week ( $w_{k,j}$ ) terms were estimated using re-weighted least squares in MINITAB with the following procedures, iteratively:

- Estimate mean, monthly, and day-of-week parameters via the GLM (General Linear Model) procedure in MINITAB,
- Compute the residual variance for each ATR in a group given the current regression parameter estimates, and
- Use the variance estimates to compute separate weighting vectors for each ATR.

The monthly and day-of-week terms were constants for all ATRs within a specific factor group  $k$ , but each ATR in the factor group was allowed to have its own mean-value parameter  $\mu$ . The weighted least squares approach in MINITAB's GLM procedure was chosen due to the heteroscedasticity caused by the ATRs' different day-to-day variances. The residuals, i.e.,  $\varepsilon_t$ 's, were further validated by their temporal dependency, and the following seasonal multiplicative autoregressive model was obtained:

$$\varepsilon_t = \Phi_1 \varepsilon_{t-1} + \Phi_7 \varepsilon_{t-7} - \Phi_1 \Phi_7 \varepsilon_{t-8} + a_t \quad (2-23)$$

where,

$a_t$  = independently and identically distributed normal random variables with mean equal to 0 and common variance, and

$\Phi_1, \Phi_7$  = autoregressive coefficients.

Once the parameters in the above autoregressive model were estimated, i.e.,  $\hat{\Phi}_1$  and  $\hat{\Phi}_7$ , and the residuals, i.e.,  $\varepsilon_t - (\hat{\Phi}_1 \varepsilon_{t-1} + \hat{\Phi}_7 \varepsilon_{t-7} - \hat{\Phi}_1 \hat{\Phi}_7 \varepsilon_{t-8})$ , were validated, it was confirmed that the residuals were not significantly auto-correlated and would pass the goodness-of-fit test for being normally distributed.

## 2.5 Artificial Neural Networks

Artificial neural networks (ANNs) are computing techniques that attempt to simulate the workings of the human brain. It is known that ANNs are superior to traditional computing techniques in solving pattern classification problems due to their unique properties (Faghri et al. 1995), such as:

- Ability to deal with incomplete input information,
- Ability to deal with noisy input data, and
- Ability to learn and associate patterns from historical data.

The ANN models consist of many simple processing elements, i.e., neurons, with dense parallel interconnections. They may be classified according to various criteria, such as their learning methods (supervised versus unsupervised), architectures (feed-forward versus recurrent), output types (binary versus continuous), node types (uniform versus hybrid), implementations (software versus hardware), connection weights (adjustable versus hardwired), operations (biologically motivated versus psychologically motivated), etc. (Jang et al. 1997). In ANNs, feed-forward means the output of each

processing element generally propagates from the input side to the output side. If there is a feedback link that forms a circular path in a network, the network is called “recurrent.”

Training and testing are the two stages in the development of an ANN model. During the training stage, an inductive learning principle is used to learn from a set of examples called a “training set.” Several neural network learning schemes, including supervised learning and unsupervised learning, are developed. A supervised learning ANN is first trained by a selected algorithm to learn from the AADTs collected at permanent count stations. The trained ANN may then be used to estimate AADTs at short-count stations. Consequently, unlike the traditional method of estimating AADT from sample volume counts, determining TTMS factor groups according to similarities in their temporal traffic variations, and then assigning each short-count station to one of the established factor groups, is no longer given a priori. Unsupervised learning ANNs may be trained without any information of the desired output to determine factor groups after frequently occurring traffic patterns are recognized. The following sections provide a brief introduction for supervised and unsupervised learning methods and their application to grouping traffic patterns and/or AADT estimation.

#### *2.5.1 Supervised Learning*

Supervised learning involves providing an ANN with “examples” that consist of inputs and the corresponding outputs. The learning algorithm attempts to adjust the weights of the connections between neurons to produce the desired output. As a result, such networks are also referred to as mapping networks. During the mapping process, the error in the output is propagated back to the previous neurons by adjusting the weights of the connections. This is called the back-propagation (BP) method for

propagating the error, and is also known as the generalized delta rule (GDR). Figure 2.4 illustrates the architecture for supervised learning and back-propagation neural networks where the target is the desired output. The process begins by assigning weights with small random values and terminates when either the maximum number of iterations is reached or the sum of absolute error (SAE) is reduced to an acceptable value.

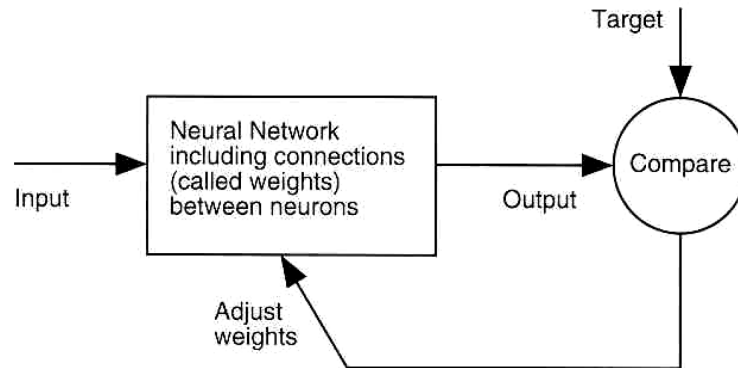


Figure 2.4 Neural Network Architecture

The multi-layered feed-forward network is probably the most commonly used model for estimating AADT. Sharma et al. (1999) investigated the traffic volume data from 63 ATR sites located on the regional and rural roads in Minnesota using a multi-layered, feed-forward, back-propagation, and supervised learning approach. The data were collected between May and August of 1993. The model consisted of three layers of neurons, i.e., one input layer, one output layer, and one hidden layer for feeding data from the input layer to the output layer. The input was the hourly volumes of vehicles included in a sample counting program, divided by the sample average daily traffic (SADT), which was simply the total volume for one or more short-period traffic counts in the sample divided by the number of sample days. Therefore, the number of neurons in the input layer was equal to the total number of hourly volumes. The hidden layer had half of the number of neurons in the input layer, and the output layer only contained one

neuron, which gave the estimated value for an AADT factor. The actual AADT factor for the output layer was defined as follows:

$$\text{Actual AADT factor} = \frac{0.25 \times AADT}{SADT} \quad (2-24)$$

The estimated AADT was calculated using the following equation:

$$\text{Estimated AADT} = 4 \times (\text{SADT} \times \text{output factor from ANNs}) \quad (2-25)$$

The learning cycles were set at 25,000. The results from the neural networks were compared with those from the traditional hierarchical grouping method proposed by Sharma and Werner (1981). Although the comparison from their study indicated that the errors from the neural network model were larger than those from a traditional grouping method, the authors argued that short-period count sites could not be assigned to one of the factor groups with 100 percent accuracy in practice. As a result, the neural network approach would be a better alternative to estimate AADT since it would not require classifying permanent count stations to groups and then assigning sample count sites to their associated TTMS groups.

Sharma et al. (2000; 2001) reached similar conclusions regarding the accuracy of the AADT estimates on low-volume roads using the traditional factor approach and ANNs. The traffic volume data collected from 55 ATR sites located on the rural roads in Alberta, Canada in 1996 were investigated. The low-volume roads referred to those for which the AADT volumes were between 120 and 999 vehicles. Sharma et al. concluded, once again, that the factor approach produced better AADT estimates than ANNs if the ATR sites were grouped appropriately and the sample sites were correctly assigned to their associated groups.

Lam and Xu (2000) implemented a multi-layered feed-forward, back-propagation neural network that consisted of one input layer, one output layer, and one hidden layer to group the traffic flow data collected in 1991 from 13 count sites in Hong Kong. Different lengths of counts (i.e., 4 hours, 6 hours, 8 hours, 10 hours, 12 hours, 14 hours, and 16 hours), each associated with several starting times in a day, were investigated. For the 13 count stations, the sum of absolute percentage errors (SAE) from the TTMSs included in the study was calculated using the following equation for both methods:

$$SAE = \sum_i^{13} |Error_i(\%)| \quad (2-26)$$

where  $Error_i(\%)$  is the percentage error between the estimated and actual AADT at the  $i$ th TTMS. An effectiveness index ( $Eff$ ) was defined to measure the effect of the extra counting time under the assumption that the cost of traffic counts was proportional to the count duration:

$$Eff = \frac{RSAE}{ETLC} \quad (2-27)$$

where,

$RSAE$  = reduction in SAE, and

$ETLC$  = amount of extra time length of count.

By comparing SAEs, Lam and Xu (2000) concluded that the neural network approach consistently performed better than the regression analysis approach in estimating AADT. The 12-hour count period was found to be the most accurate period for AADT estimation because of the minimum SAE. However, the 8-hour count was the most effective period with the highest  $Eff$  value of 5.41.

Lingras et al. (2000) applied a time-delay neural network (TDNN) and an autoregression (AR) model to forecast daily traffic volumes at 78 TTMS sites in Alberta, Canada. To simplify the analysis, the TTMS sites were first classified into the following five types of road groups:

1. Highly recreational
2. Regional recreational
3. Long distance
4. Urban commuter
5. Regional commuter

The method suggested by Sharma and Werner (1981) was used to determine different groups of road classifications based on the traffic data collected in 1993. After road types were determined, one TTMS from groups 1, 2, and 3, and three TTMSs from groups 4 and 5, were selected. Only TTMSs with continuous traffic data from 1989 to 1993, inclusively, were selected. The traffic data collected from 1989 through 1992 were then used to train the TDNN and calibrate the autoregression (AR) model for each classification group. These models were subsequently tested using the data collected from the TTMSs selected in 1993. Daily traffic volumes for the previous 13 days (i.e.,  $x_1, x_2, \dots, x_{13}$ ,) were defined as the independent variables or input variables to predict traffic volume for the following day ( $x_{14}$ ). The AR equation is shown below.

$$x_{14} = \sum_{i=1}^{13} a_i x_{14-i} + e_{14} \quad (2-28)$$

The TDNN had 13 input nodes corresponding to the previous 13 daily traffic volumes, and one output to predict the traffic volume. The average and maximum percentage errors (between the predicted and the actual traffic volumes), as well as the

50<sup>th</sup>, 85<sup>th</sup>, and 95<sup>th</sup> percentile errors from the cumulative frequency distributions, were used as model performance measures. Lingras et al. (2000) concluded that TDNN models produced better predictions than AR models for all five road groups, since all of the error measures were smaller with the neural network approach.

Theoretically, if the number of neurons in the hidden layer is large enough, supervised learning ANNs will be able to closely approximate any complicated non-linear function. Current practices, however, utilize the ANN paradigm designed with one hidden layer to reduce the intensive computing efforts required by the training process. Consequently, the performance of supervised learning neural networks for the estimation of AADT has not been adequately explored.

### *2.5.2 Unsupervised Learning*

Unsupervised learning is an approach that extracts features or regularities from presented patterns without any information for the desired output (Jang 1997). ANNs with unsupervised learning update weights solely base on the input patterns and are trained to respond to frequently occurring patterns. The following sections describe the unsupervised learning paradigms for competitive learning and the Kohonen self-organizing feature map.

#### *2.5.2.1 Competitive Learning and ART1*

In competitive learning ANNs, the number of output units is equal to the number of clusters into which the data are divided. The weights of the neural connections are updated according to the competitive, or winner-take-all, learning rule. Competitive learning ANNs have two disadvantages. One is that the number of classification clusters must be specified before the learning proceeds, and the model lacks the capability to add



new clusters when necessary. In other words, competitive learning classifies a given pattern into exactly one of the mutually exclusive predetermined classes. The other disadvantage is that responses to the same input pattern may differ on each successive presentation of that input pattern, and the winning unit that responds to a particular pattern may continue to change during training. This is usually referred to as the stability-plasticity dilemma. Such unstable learning in response to prescribed input is due to the learning that occurs with other intervening inputs. Consequently, the network adaptability, or plasticity, enables prior learning to be replaced by more recent learning, in response to a wide variety of input environments. Carpenter and Grossberg (1988) proposed the ART1 architecture as a solution to the dilemma, which was capable of recognizing patterns from arbitrary binary input patterns. The ART1 neural network is a paradigm of adaptive resonance theory (ART) that processes binary patterns, in which each element of input vector takes only a value of 0 or 1. The ART1 learning scheme is also capable of creating new clusters when needed.

Faghri and Hua (1995) applied the ART1 neural network to group 29 ATR stations in Delaware with traffic data collected from 1985 through 1989. ART1 had only one layer of processing units. The ART1 ANNs set up certain categories for the input and classify the input pattern into a proper category. If an input pattern did not match any existing categories, the network would create a new category for it. The ratio of a MADT to the corresponding AADT, i.e.,  $V_0$ , for a given TTMS was first converted using the following formula:

$$V_n = \frac{V_0 - V_0^{\min}}{V_0^{\max} - V_0^{\min}} \quad (2-29)$$

where,

$$\begin{aligned}
V_n &= \text{conversion result for the ratio of MADT to AADT,} \\
V_0 &= \text{ratio of MADT to AADT,} \\
V_0^{max} &= \text{maximum value of the MADT to AADT ratio, and} \\
V_0^{min} &= \text{minimum value of the MADT to AADT ratio.}
\end{aligned}$$

The 12 new ratios corresponding to the 12 months in a year were then converted to binary numbers and entered into each column of a  $10 \times 12$  matrix. This matrix was used as an input to the ART1 ANNs for the traffic pattern obtained from a given TTMS. Some accuracy was lost due to rounding because each MSF was represented by a  $10 \times 1$  vector. This loss of accuracy was considered insignificant and ignored in the study. A value of 0.83 was determined the vigilance factor after a few pre-designated count sites were correctly classified into proper categories. The results from the ART1 method were compared with those obtained from both cluster and regression analyses. While four seasonal categories were produced by all three methods, they differed in the way that the TTMSs were grouped. Cluster and regression analyses created categories of urban, rural, recreational-arterial, and recreational collector, while the ANN created categories of urban or interstate, rural-arterial, rural collector, and recreation. There were only two stations for which categories were not determined by the ART1 method. As for at least five TTMSs, the groups changed from year to year and from method to method.

The following equation was used to measure the comprehensive performance of the three methods for estimating SFs:

$$average_{type} = \frac{1}{12} \sum_{\forall j} err_{type}(j) = \frac{1}{12} \sum_{\forall j} \sum_{\forall i} [sf_{type}(i, j) - sf_{act}(i, j)]^2 \quad (2-30)$$

where,

$$average_{type} = \text{average error for method } type,$$

$err_{type}(j)$  = dissimilarity between estimated and actual SFs in month  $j$  for method  $type$ ,

$sf_{type}(i, j)$  = estimated seasonal factor for method  $type$  at  $i$  ART in  $j$  month, and

$sf_{act}(i, j)$  = actual SF at  $i$  ART in  $j$  month.

By comparing the average errors of the three grouping methods, Faghri and Hua (1995) concluded that the neural network method outperformed the cluster and regression methods. The results indicated that ART1 networks had the ability to organize inputs into their natural groups, as well as the capability of weeding out random seasonal fluctuations in the input patterns.

#### 2.5.2.2 Kohonen Self-Organizing Feature Map

Kohonen self-organizing networks, also known as Kohonen feature maps or topology-preserving maps, are another competition-based network paradigm for data grouping (Lingras 1995). The learning procedure of Kohonen feature maps is similar to that of competitive learning ANNs. However, in addition to updating the weights for the winning units, all of the weights in a neighborhood surrounding the winning units are updated as well. The network consists of two layers: input and Kohonen layers. The network receives the input vector as a given pattern. If the pattern belongs to the  $k^{\text{th}}$  group, the  $k^{\text{th}}$  unit in the Kohonen layers will have an output value of one, while the other neurons will have a value of zero.

Lingras (1995) compared the classification groups from Kohonen unsupervised learning ANNs with those from a hierarchical grouping method, using data collected from 72 PTC sites in Alberta, Canada. Five seasonal categories were specified for Kohonen ANNs. The number of iterations was set to 100, since grouping stabilized after

presenting the training set to the ANN 100 times. The findings concluded that the Kohonen ANNs produced results that were similar to the hierarchical grouping method. As a result, ANNs could be used to substitute the statistical techniques for the grouping of traffic patterns. Moreover, Kohonen ANNs updated the weights on the connections only when complete patterns were presented. For incomplete patterns, the ANNs could find the categories using the least mean-square error or other similarity measures. This feature enabled Kohonen ANNs to classify incomplete monthly traffic patterns.

## **2.6 Genetic Algorithms**

Genetic algorithms (GAs), originally called genetic plans, have received a great deal of attention because of their potential to solve optimization problems (Sakawa 2002). The GA technique is a stochastic search process based on the mechanism of natural selection and genetics. In a GA, problem solutions are represented as chromosomes, which are made up of genes. Starting with an initial population of individuals, or chromosomes, genetic operators are applied to evolve the population by producing successively new populations with improved individual “fitness.” Each iteration produces a new generation of solutions. For any given generation, each individual in the population is evaluated using some measure of fitness, usually the objective function in an optimization problem. Genetic operators, such as selection, reproduction, crossover, and mutation, are then used to create the next generation of the population. Individuals are subsequently selected from the current population based on their fitness values. Reproduction involves applying crossover and mutation operators to some of the selected individuals to produce a new generation with improved fitness. The crossover operator selects individuals from the population at random and exchanges

portions of genes to produce new individuals, while the mutation operator randomly alters one or more genes of a selected individual. The process continues until the termination condition is satisfied, which occurs either when the best fitness value of the population stops improving, or a prescribed number of iterations is exceeded.

GA-based methods have several advantages:

- GA formulations do not require the calculation of gradient matrices or other higher-order derivative matrices and their approximations.
- A GA-based solution directly operates its search process, such as transformation, through genetic operators and selection based on fitness. Therefore, there is no need to formulate a system of governing equations that mathematically represents or simulates the relationship between various parameters and unknowns. This is particularly attractive for practical applications where the difficulty lies in establishing mathematical formulations to accurately and effectively simulate complex problems.
- Constraint conditions are relatively easy to incorporate into a GA-based process. Constraint conditions may be simply defined as a part of the environmental conditions, or by assigning large penalty numbers to individuals that violate certain constraints, thus reducing the possibility of survival in the selection process. This may be especially suitable for problems where constraints are complicated and cannot be properly defined.

Research into the field of GAs has shown increased activity over the past several decades, the results of which have been widely used in a variety of applications. However, GAs also have two main disadvantages:

- GAs are stochastic algorithms whose search methods are based on the natural evolution principle. Although a sufficiently large number of “individuals” may result in a near-optimal solution to an optimization problem, the GA technique does not guarantee globally optimal solutions.
- GAs may require an extremely large amount of computational CPU time when dealing with large-scale problems.

Lingras (2001) utilized a GA to group TTMSs and compared the classifications with those from the traditional hierarchical grouping method developed by Sharma and Werner (1981, 1983). The monthly traffic patterns collected between 1987 and 1991 from TTMSs on Alberta’s highways were used. The number of genes in a chromosome was set to equal the number of seasonal patterns that needed to be classified. Each chromosome corresponded to a classification scheme. A gene was then randomly assigned with an initial value between 1 and  $m$ , where  $m$  is the desired number of groups. Solutions of 2 to 15 factor groups with the following object function were investigated:

$$\frac{\Delta_1}{\Delta_m} = \frac{\sum_{i=1}^n \sum_{j=1}^n d(P_i, P_j)}{\sum_{i=1}^m \sum_{x_j, x_k \in X_i} d(x_j, x_k)} \quad (2-31)$$

where,

$\Delta_1$  = maximum possible within-group error,

$\Delta_m$  = sum of within-group error for  $m$  groups of seasonal patterns,

$P_i$  = seasonal traffic pattern  $i$ ,

$d()$  = a distance function to measure the dissimilarity between patterns,

$x_j$  = seasonal traffic pattern  $j$  in factor group  $X_i$ , and

$n$  = total number of seasonal patterns.

The behaviors of both GAs and hierarchical methods were also compared for 20, 30, 40, and 50 groups. The classification schemes for different numbers of groups with the highest values of  $\Delta_l/\Delta_m$  from 1,000 generations of evolution were compared with those from the traditional hierarchical clustering approach. The results indicated that the hierarchical grouping method performed better when the number of groups was greater than 14. However, GAs performed better when the number of groups was less than nine. Since the initial grouping patterns were randomly assigned, the results were verified by repeating the experiment in five factor groups for a total number of 22 times. The within-groups errors varied between 680 and 730, which were consistently and significantly lower than the hierarchical grouping error of 861. The genetic approach was also applied with different numbers of generations, ranging from 100 to 1,000, with an interval of 100 for five factor groups. The results showed that the GAs' errors were less than the hierarchical grouping error after 400 generations.

## **2.7 Assignment of Count Sites**

There is considerable vagueness in the current practice of assigning count sites to SF groups. Currently, assigning short-count sites to factor groups and determining the precision of short-count estimates is generally accomplished by considering the physical proximity of short-count sites to a TTMS, based on engineering judgment (TMG 2001). If the true factor group for a site is known, it was reported that traditional short-counts could provide estimates of mean daily traffic with the PI95 (precision achievable with 95% confidence) between 10 and 23 percent (Davis 1996). Inappropriately assigning a site to a factor group may result in a drastic decline in precision.

The North Carolina Department of Transportation (NCDOT) implemented a data management system developed with GIS to assign each short-count site to one of the seven seasonal groups based on the most recent data at that site (McDonald 1999). Short-count stations that had at least three or four 48-hour sampling traffic counts available were used to identify the seasonal group that was highly correlated over these short count stations' day and month variations. In other words, statistical correlation and their associated  $p$ -values were used to determine the best seasonal group for a given short count site.

Davis and Guan (1996) employed the Bayesian theorem to assign a given site to a known SF group with the highest posterior probability. The probability was defined as follows:

$$\text{Prob}[\text{site} \in G_k | z_1, \dots, z_N] = \frac{f(z_1, \dots, z_N | G_k) \alpha_k}{\sum_{l=1}^n f(z_1, \dots, z_N | G_l) \alpha_l} \quad (2-32)$$

where,

$f(z_1, \dots, z_N)$  = a likelihood function measuring the probability of obtaining the count sample had the site belonged to a given SF group,

$z_1, \dots, z_N$  = a sequence of  $N$  daily traffic counts at a short-count site,

$G_1, \dots, G_n$  = a total of  $n$  different factor groups, and

$\alpha_k$  = probability that the given site belongs to  $G_k$  prior classification.

The prior classification probability, or  $\alpha_k$ , was assumed to equal  $1/n$ , indicating complete prior uncertainty as to which group a short-count site belonged. The linear regression model described in Section 2.4 was used as the likelihood function in the posterior classification probabilities, as shown below:



$$y_t = \mu + \sum_{i=1}^{12} \Delta_{t,i} m_{k,i} + \sum_{j=1}^7 \delta_{t,j} w_{k,j} + \varepsilon_t \quad (2-33)$$

It was further assumed that  $\varepsilon_1, \dots, \varepsilon_N$ , were normally distributed random errors with a mean value of 0 and a covariance matrix  $\sigma^2 \mathbf{V}$ , where  $\sigma^2$  was the common unconditional variance of  $y_t$ , and  $\mathbf{V}$  was a  $N \times N$  matrix of correlation coefficients such that the element in row  $s$  and column  $t$ ,  $V_{s,t}$ , was the correlation coefficient for  $y_s$  and  $y_t$ . The approach was developed based on the assumption that short-term count sites should be assigned to one of the SF groups that had a similar monthly and daily variation pattern. The model was validated using data from 48 TTMS stations for 1991 and 50 for 1992. This data-driven approach was shown to produce mean daily traffic estimates that were near  $\pm 20$  percent of actual values, based on 14 well-selected sampling days from particular months and days of the week. Although the method did not provide significant improvement in precision over what may be achieved when the appropriate SFs were known, the reliance on subjective judgment was reduced by this process. A potential problem with the Bayesian assignment approach proposed by Davis and Guan, however, is that a longer period of data collection at short-count sites is needed. The approach is also complicated and time-consuming.

## 2.8 Summary

Several approaches that were developed to incorporate seasonal effects in the calculation of total traffic volumes on a given roadway segment were discussed in this chapter. For models developed based on artificial intelligence technologies, such as neural networks and GAs, it may be difficult to interpret the resulting seasonal patterns, especially when not in agreement with engineering judgment. Although ANNs have

been shown to be effective at representing complex nonlinear relationships, it is difficult to determine the relationships between variables. It is also possible to over-train a network, resulting in memorization of the training data rather than a generalization of the relationship. Consequently, recommendations were made to use a larger database for training purposes, as well as to use proper judgment regarding when to cease training (Smith 1997). However, such requirements are generally difficult to meet since installing and maintaining a large number of TTMS sites is unlikely. Moreover, without formulating the human judgment, the process of determining seasonal groups cannot be automated.

The theoretical backgrounds for the nonparametric hierarchical clustering methods described in Section 2.2.1 are relatively easy to understand. These models have been generally implemented for the grouping of TTMSs via popular commercial statistical software packages, such as SAS. The parametric model-based clustering models described in Section 2.2.3 are complicated and require a greater knowledge of statistics, but also allow the parameters measured in different scales, such as the geographic locations of the TTMSs, to be simultaneously considered in the grouping process without additional transformation.

Seasonal variations in traffic are the result of patterns in human activities, which are commonly influenced by land use patterns. The land use and travel behavior aspects of SFs, however, have not been adequately studied in the existing literature. By considering and incorporating the functional classifications of roadways, land use, and other factors relevant to data collection and processing, it is possible to reduce the data collection effort while improving the accuracy of SF estimations.

## **CHAPTER III**

### **METHODOLOGY**

This chapter describes the methodology of this research. The components of the methodology framework include data preparation for dependent and independent variables, multiple regression analysis for variable selection, and the assignment of MSF for continuous count sites to coverage sites. Several of the steps include the same methodology applicable to both urban and rural areas, such as the data imputation for the monthly seasonal factors (MSFs). However, different sets of influential variables are compiled for urban and rural areas to better reflect the fluctuation of seasonal traffic, and distinguishing approaches are employed to further divide the TTMSs into sub-regions/groups in urban and rural areas. The methodology framework is first introduced in this chapter, and techniques developed for each step are then described in the subsequent sections.

#### **3.1 Methodology Framework**

Figure 3.1 shows an overview of the research methodology that consists of four major steps. Step 1 consists of preparation for dependent and independent variables in regression analysis. The dependent variables are the 12 MSFs of continuous traffic count sites for the year 2000. A data imputation technique was employed for the compilation of the MSFs in order to maximize the number of TTMSs that may be used in this study. The independent variables reflect land use, demographics, socioeconomics, roadway characteristics, and other variables. They are extracted via GIS techniques, such as buffer analysis and spatial analysis.

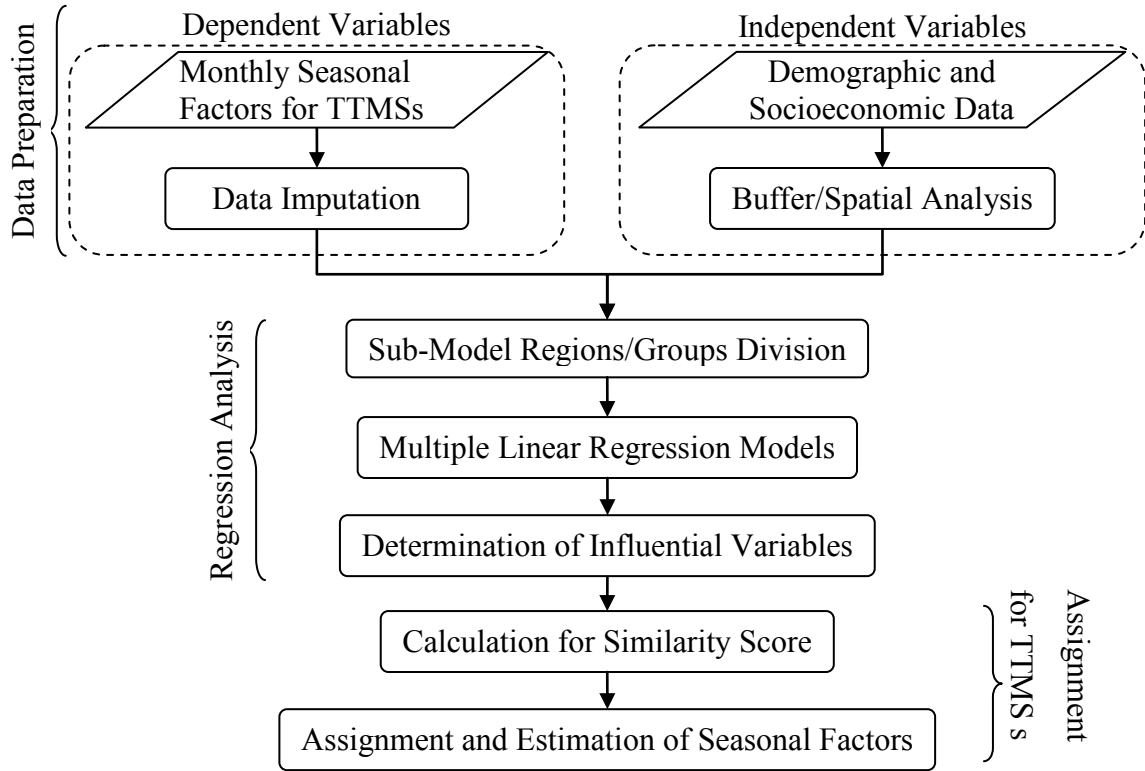


Figure 3.1 Framework of the Methodology

In Step 2, multiple linear regression analyses are used to identify the potentially influential factors that contribute variations to the SFs. For the urban areas, the entire state is divided into three analysis regions to account for the significant differences in climate across the state; for the rural areas, TTMSs are divided into two groups based on their typical weekday hourly traffic patterns. Separate models are then developed for each region/group to associate the 12 monthly SFs with potentially influential variables. In Step 3, the variables identified in the regression analysis are used to measure the similarity between the TTMSs and a short-count site, and estimation of the MSFs for the short-count site are based on those of the TTMSs that are determined to be similar. Finally, in Step 4, the coverage area for a TTMS site is determined and, subsequently, the

coverage for current TTMS sites evaluated. The following sections describe each of the steps in further detail.

## **3.2 Data Preparation**

### *3.2.1 Data Imputation for Dependent Variables*

The data used in this research are from the year 2000. The decision to use these data is based on census data availability for that year. Hence, the data on demographics are likely to be more accurate.

Traffic volume data were continuously collected from nearly 285 TTMSs located in 68 counties in Florida in the year 2000. Due to the operational environment of the devices, missing and erroneous data are unavoidable. As a result, 60 TTMSs are missing a portion or all of the seasonal factors. These 60 TTMSs comprise more than 20 percent of all TTMSs.

A common practice when treating incomplete data is the removal of records with missing values. However, because of the limited number of TTMSs and the large area they need to serve, it is important that as many TTMSs as possible be used for this analysis. This will ensure that 1) the largest possible geographic coverage is achieved, and 2) statistical results are valid.

Historical data for the TTMSs with missing MSFs are examined. The missing data are then estimated based on techniques in trend analysis and averaging. This procedure is described in Chapter IV.

### *3.2.2 Data Acquisition for Independent Variables*

Potential independent variables used in regression analysis are those considered likely to have a causal relationship with MSFs. They describe the roadway, demographic,

and socioeconomic characteristics of an area where a TTMS is located. Their inclusion is based on two major considerations: 1) whether the source data are readily available or can be collected easily and economically for both base and forecast years, and 2) whether variables can be quantified. The independent variables can generally be classified into the following categories:

- Roadway characteristics.
- Aggregate demographic and socioeconomic variables in the area surrounding count stations.
- Special land-use variables for the urban area.
- Location variables for the rural area.

Roadway characteristic variables are naturally come with the TTMS locations. However, other area-based variables are obtained via GIS techniques.

Aggregate demographic and socioeconomic variables are compiled using the buffer analysis method. For urban areas, a circular buffer around each count station is created and used as the basis for the estimation of the variable values. The use of buffer method is based on the assumption that traffic at a count station is affected by trips generated in or attracted to the area within a certain distance of the count station. The buffer radii vary according to the functional classification of the roadway segment where a TTMS is located (Zhao and Chung, 2001). This variation reflects the size of the service area for different types of roads. As shown in Figure 3.2, the buffer radii are 5 miles for freeway and principal arterials, 0.5 mile for minor arterials, and 0.25 mile for collectors. These radii are based on the common spacing of roads for different functional classes. A

larger buffer zone implies that the MSFs for a count station are impacted by the characteristics of a larger surrounding area.

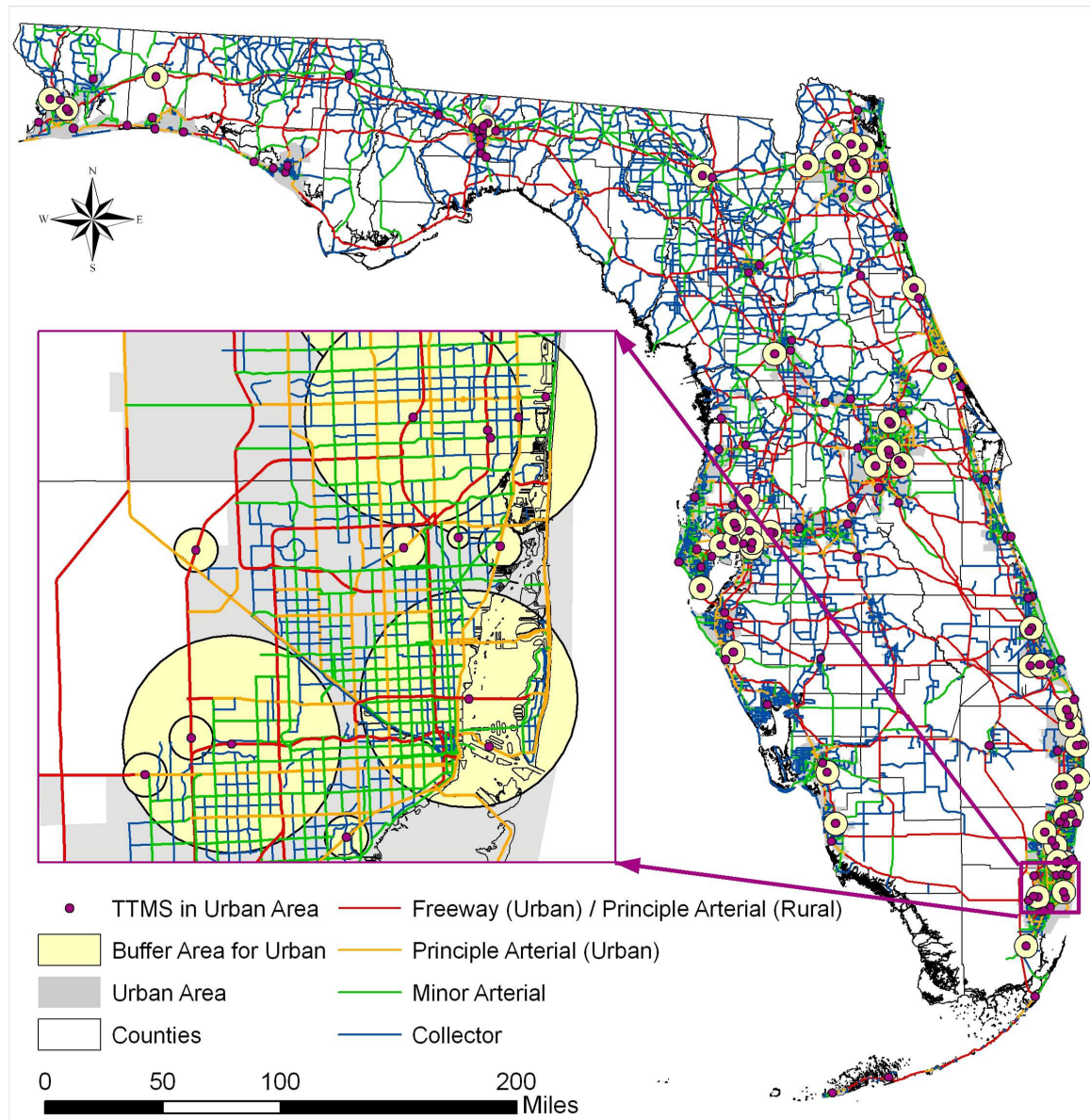


Figure 3.2 Buffer Areas Around TTMS in Urban Areas

However, because roadway spacing in rural areas is irregular, a uniform buffer size is inappropriate even for TTMSs on roads of the same functional classification. Therefore, a variable buffer method is used. Under this method, GIS is utilized to calculate the distance between the road where a TTMS is located and the closest road

with the same functional classification. A fixed percentage is then applied to this distance to determine the buffer size. Three percentages are tested with regression analysis: 25%, 50%, and 75%. Because 50% yields the best regression models, it is selected as the percentage used to compute buffer size. For instance, if the distance between a TTMS and the next road with the same functional classification is eight miles, applying the 50% will yield a buffer size of four miles. However, if this distance exceeds ten miles, then a five-mile upper limit of the buffer size is applied. In addition to the buffer size limit, the buffer area may also be modified if it overlaps with any of the urban areas. The overlapping urban areas are removed from a buffer area to arrive at the final impact area, which is then used to compile independent variables. The buffer area around TTMS in rural areas is shown in Figure 3.3.

Special land use variables for urban areas, including universities, tourist attractions, and recreational sites are designed to reflect the locations that tend to generate or attract seasonal traffic. Dummy variables are used to determine whether a TTMS is located within the boundary of a hot spot or any portion of the buffer area around the TTMS that is occupied by a certain kind of recreational land use.

For rural areas, the large percentage of through traffic information is hard to capture. Location variables are measured, including the distance between a TTMS and an urban area, beach, or interstate highway. The population size of the urban area is also taken into consideration, as a larger urban area may have a greater impact on a nearby TTMS. The detailed description and the method for calculation for each variable are described in the next chapter.



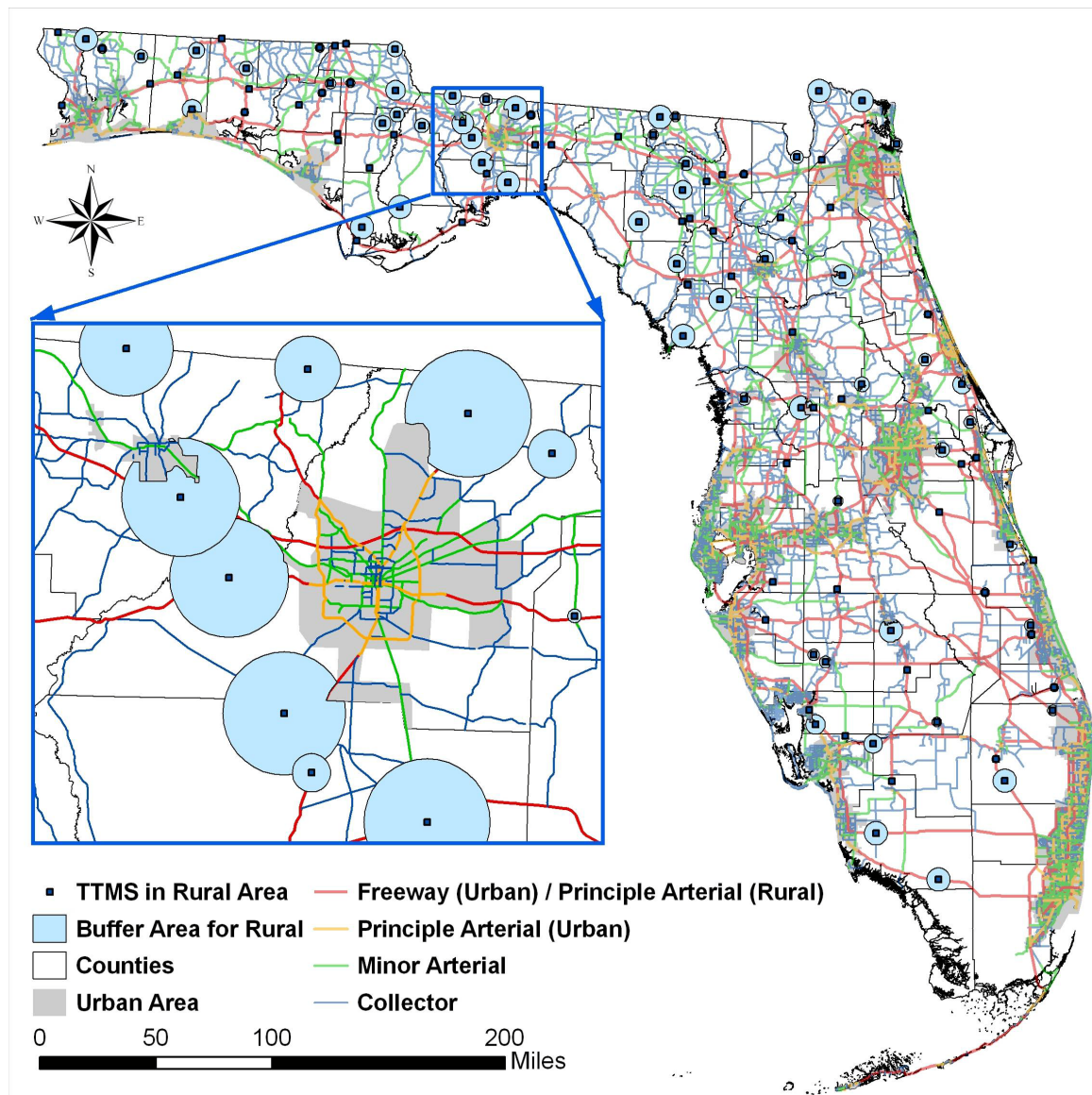


Figure 3.3 Buffer Areas Around TTMS in Rural Areas

### 3.3 Model Development

Multiple regression analyses are conducted to associate the seasonal factors with potentially influential variables. The dependent variables are the 12 MSFs. The stepwise selection method is applied, with the significance level set at 0.05 for a variable entering and staying in the model. The  $t$ -statistics and variance inflation factors (VIFs) are also checked in each variable to ensure that the variables are significant, as well as to remove

multicollinearity in the models. The regression analyses attempt to establish the relationships between the MSFs and potentially influential variables as linear equations.

These equations are formatted as follows:

$$MSF_k = \beta_{0k} + \beta_{1k} x_1 + \dots + \beta_{ik} x_i + \dots + \beta_{pk} x_p \quad (3-1)$$

where,

$MSF_k$  = a monthly seasonal factor for month  $k$ ,

$\beta_{ik}$  = the regression coefficient for the  $i$ th independent variable for month  $k$ , and

$x_i$  = the  $i$ th independent variable.

Since Florida stretches over several climate zones, from temperate in the north to subtropical in the south, the seasonal effects of the same variables may be different depending on the latitude. South Florida, for example, attracts many visitors and welcomes the return of large numbers of seasonal residents in the winter months due to its warm temperatures. In contrast, summer in northern Florida is the season for tourists and for outdoor recreation. Therefore, the same variables may impact traffic in a similar manner, but during different seasons in northern and southern Florida. For this reason and for modeling purposes, the state is divided into three regions representative of three climate zones: North Florida, Central Florida, and South Florida. Therefore, separate models are developed for each region in the urban areas.

As for rural areas modeled by different regions, results still show a weak relationship between MSFs and the independent variables describing demographic, socioeconomic, and roadway characteristics. One reason for this may be that the monthly variation in traffic is more significant on rural roads than for urban and commuter routes (HCM 2000). In addition, another cause of the poor model results may be that urban

traffic is dominated by commuting. The rural areas often lack commuters, and most traffic may be generated from other activities such as agriculture, mining, fishing, and recreational travel. The hourly traffic pattern for TTMSs in the rural area was carefully studied and analyzed in order to reflect the characteristics for commuting traffic and traffic for other purposes. A strategy was developed to model TTMSs separately, based on their hourly traffic patterns, so as to improve model results (Lu et al. 2012). Section 5.2.1 describes the method used to classify the rural TTMSs into two groups: one with daily traffic patterns characterized by a single peak, and the other with patterns characterized by double peaks.

### **3.4 Monthly Seasonal Factor Assignment**

The regression models for each region or group indicate a relationship between the monthly seasonal factors and land use variables. Even though the models cannot be used to directly predict the monthly seasonal factors, they provide likely connections between the seasonal factors and the various variables modeled. These variables may be used to develop a metric to determine which TTMS(s) may be used for the assignment of seasonal factors to a coverage count site. This metric is based on the similarity between land use and other characteristics. The underlying assumption is that the MSFs of two count stations are similar if they share similar characteristics, which are defined in terms of the influential variables. This provides the basis for the development of an assignment method that relies on a similarity score,  $S$ , to match a TTMS, or the known MSFs to a PTMS (the MFSs that will be estimated). This score is calculated based on a set of selected variables that have been identified in the regression analyses. The method about how the similarity score,  $S$ , is defined and calculated is discussed in Chapter VI.

### 3.5 Summary

This chapter proposed a complete methodology framework to solve the stated problems. The framework is composed of four parts: data preparation, model development, MSF assignment for continuous count sites to coverage sites, and the assignment results evaluation.

To maximize the data that may be used in the analysis, the missing MSFs are replaced with the historical TTMSs data in the data preparation phase. GIS techniques, such as buffer analysis and spatial analysis, are applied to compile the roadway characteristics, socioeconomic and demographic data, and land use information for the regions surrounding the TTMSs.

Multiple regression analyses aim to associate the seasonal factors with potentially influential variables. To account for the significant differences in climate across the state, the urban areas are divided into three analysis regions: North Florida, Central Florida, and South Florida. The results for the rural TTMS models are then improved by dividing the TTMSs into two groups based on their typical weekday hourly traffic patterns.

The variables identified in the regression analysis should indicate that there is a relationship between the monthly seasonal factors and land use variables. The assignment method for this research is to develop a metric to determine which TTMS(s) may be used for the assignment of seasonal factors to a coverage count site. This metric is based on the similarity between land use and other characteristics.

## CHAPTER IV

### DATA ACQUISITION AND PROCESSING

#### 4.1 Imputation of Monthly Seasonal Factor Data

The MSF data used in this research are from the year 2000. During the year, a total of 285 TTMSs existed statewide, among which 187 of the sites were located in the urban areas and 149 sites in the rural areas. For the urban areas, MSFs are obtained based on the traffic counts throughout the entire year; while for the rural areas, only the weekday data are used to calculate the MSFs. The decision to reduce the problem's complexity was made based on the effect of atypical traffic data due to weekends, holidays, special events, and other non-recurring events, which may be more pronounced in the rural areas than the urban areas because of the relatively light traffic on rural roads. Due to the operational environment of the devices, missing and erroneous data are unavoidable. As a result, 69 TTMSs are missing one or more MSFs; these missing data result in a loss of 24.2 percent of the useable data for the year 2000. Of the 69 TTMSs, 36 sites were located in the urban areas and 33 in the rural areas. TTMSs with missing MSFs in the urban and rural areas are listed in the tables in Appendix A.

##### *4.1.1 Data Imputation Procedure*

In order to impute the missing data from the year 2000 data for the urban areas, all of the available historical MSFs data for each site from 1997 to 2005 were checked. However, for the rural areas, the MSFs were imputed based on the historical data from 1998 to 2005. This is because the data structure used for 1997 is different from that of the other years. The data imputation procedure follows the rules described below:

- If only one or two MSFs are missing from the 2000 data, use the data from 1999 or another year for the months with missing data.
- If more than two MSFs are missing, check the 1999 data. If data are available and the seasonal pattern is consistent with those from the other years, use the 1999 data for the year 2000.
- If the data from 1999 have MSFs that are significantly different from those of the other years, use the average values from all years, but exclude the 1999 MSF(s).
- If the data from 1999 are also missing, look for the next closest year that has complete MSFs.

Following the above rules, three examples are presented below to illustrate the procedure for data imputation.

**Example 1:** The data from the year 2000 borrowed from the year 1999

For site #930099, the MSFs are missing for six months. Figure 4.1 plots the historical data. Note that the annual seasonal patterns are quite similar, and that the 1999 pattern is consistent with those of other years. This means that the 1999 data can be borrowed for the year 2000. Table 4.1 shows that the 1999 MSF data are complete. Therefore, the 1999 data were adopted.

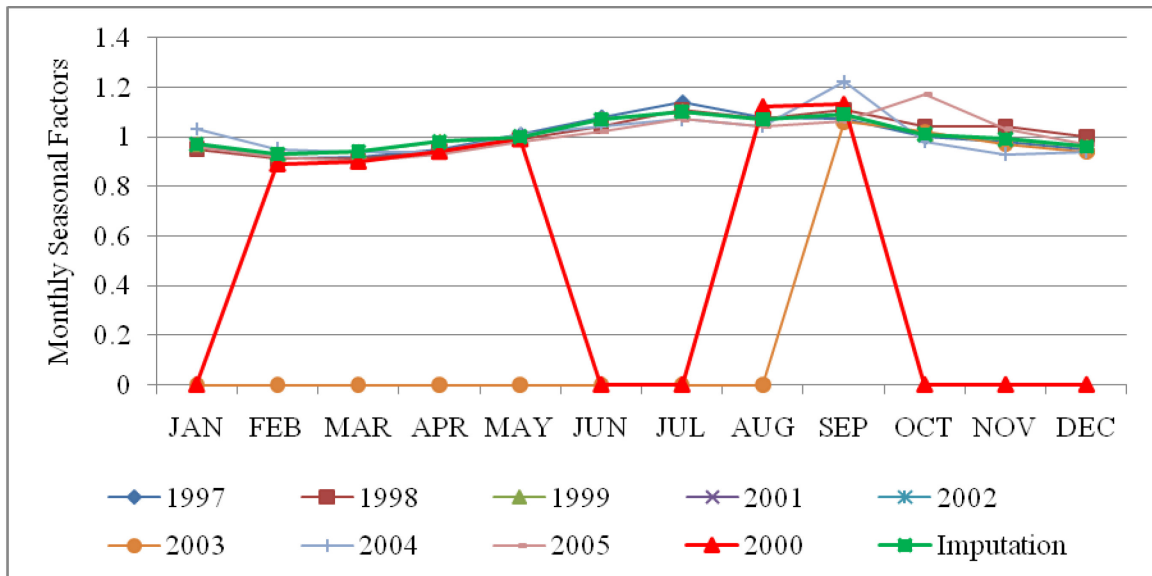


Figure 4.1 Historical Data Plot for Site 930099

Table 4.1 Imputation of Site 930099

YEAR	MSFs											
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1997	0.97	0.91	0.92	0.95	1.01	1.08	1.14	1.08	1.07	1.00	0.98	0.95
1998	0.95	0.91	0.91	0.94	0.99	1.04	1.11	1.07	1.11	1.04	1.04	1.00
1999	0.97	0.93	0.94	0.98	1.00	1.07	1.10	1.07	1.09	1.01	0.99	0.96
2001												
2002												
2003	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.06	1.02	0.97	0.94
2004	1.03	0.95	0.94	0.94	1.01	1.04	1.07	1.04	1.22	0.98	0.93	0.94
2005	0.96	0.92	0.9	0.93	0.98	1.02	1.07	1.04	1.06	1.17	1.03	0.97
2000	0.00	0.89	0.9	0.94	0.99	0.00	0.00	1.12	1.13	0.00	0.00	0.00
Imputed	0.97	0.93	0.94	0.98	1	1.07	1.1	1.07	1.09	1.01	0.99	0.96
Source year	1999	1999	1999	1999	1999	1999	1999	1999	1999	1999	1999	1999

**Example 2:** The October MSF borrowed from the 1999 data based on the average of all other years.

For site 930174, all of the 12 MSFs for 1999 are available. However, the MSF for October 1999 is different from those from all other years. Therefore, the average value of all other years was computed as the imputed value. Figure 4.2 shows the historical data plot. Table 4.2 provides the data used and the imputed values.

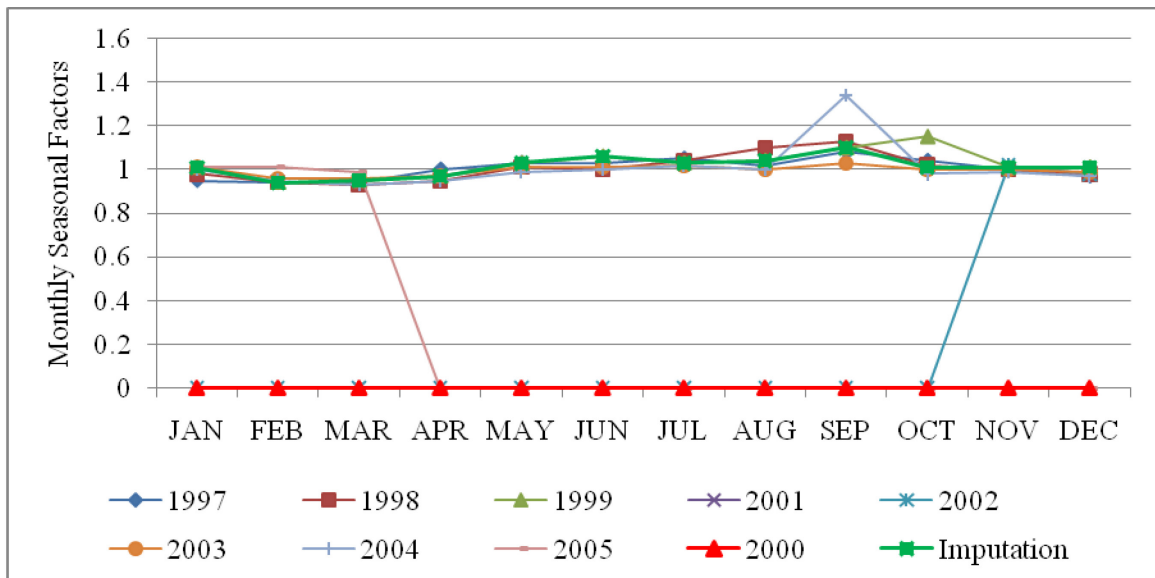


Figure 4.2 Historical Data plot of Site 930174

Table 4.2 Imputation of Site 930174

YEAR	MSF											
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1997	0.95	0.94	0.94	1.00	1.03	1.03	1.05	1.02	1.08	1.04	1.00	0.97
1998	0.98	0.94	0.93	0.95	1.01	1.00	1.04	1.10	1.13	1.02	1.00	0.98
1999	1.01	0.94	0.95	0.97	1.03	1.06	1.03	1.04	1.10	1.15	1.01	1.01
2001												
2002	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.02	0.98
2003	1.01	0.96	0.96	0.97	1.01	1.01	1.02	1.00	1.03	1.00	1.00	0.99
2004	1.00	0.95	0.93	0.95	0.99	1.00	1.02	1.00	1.34	0.98	0.99	0.97
2005	1.01	1.01	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Imputed	1.01	0.94	0.95	0.97	1.03	1.06	1.03	1.04	1.1	1.01	1.01	1.01
Source year	1999	1999	1999	1999	1999	1999	1999	1999	1999	Avg.	1999	1999



**Example 3:** Data borrowed from the year 2001.

For site 740047, data for four months in 2000 are missing. The 1999 data are also missing for several months and cannot be used. As a result, the 2001 data were borrowed. Figure 4.3 shows that seasonal patterns are similar for the period between 1997 and 2005. Table 4.3 displays the imputation results.

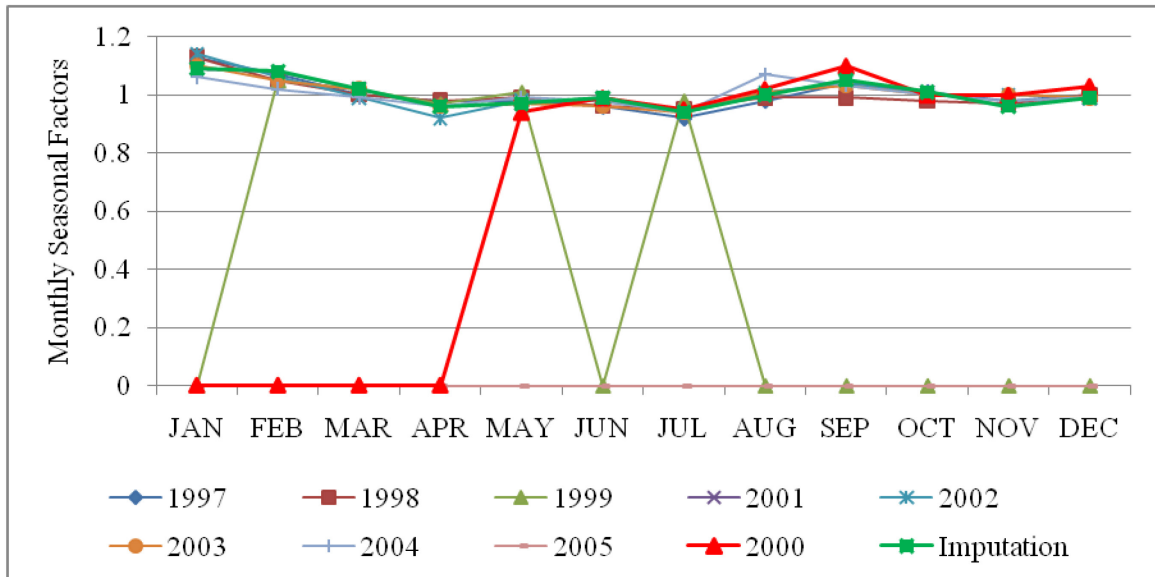


Figure 4.3 Historical Data Plot for Site 740047

Table 4.3 Imputation of Site 740047

YEAR	MSF											
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1997	1.13	1.07	1.00	0.98	0.98	0.96	0.92	0.98	1.04	1.01	0.98	1.00
1998	1.13	1.05	1.00	0.98	0.99	0.96	0.95	0.99	0.99	0.98	0.97	1.00
1999	0.00	1.05	1.02	0.97	1.01	0.00	0.98	0.00	0.00	0.00	0.00	0.00
2001	1.09	1.08	1.02	0.96	0.97	0.99	0.94	1.00	1.05	1.01	0.96	0.99
2002	1.14	1.06	0.99	0.92	0.98	0.98	0.94	1.01	1.04	1.00	1.00	0.99
2003	1.10	1.05	1.02	0.96	0.97	0.96	0.94	1.01	1.03	1.00	1.00	0.99
2004	1.06	1.02	0.99	0.96	0.99	0.98	0.94	1.07	1.03	1.00	0.98	0.99
2005	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2000	0.00	0.00	0.00	0.00	0.94	0.99	0.95	1.02	1.10	1.00	1.00	1.03
Imputed	1.09	1.08	1.02	0.96	0.97	0.99	0.94	1.00	1.05	1.01	0.96	0.99
Source year	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001

#### 4.1.2 Data Imputation Results

For the 69 TTMSs with missing data in the year 2000, the MSFs for 54 sites were imputed successfully. The remaining 15 sites cannot be imputed for the following reasons:

1. Three sites do not have reliable historical data.
2. Three sites show inconsistent patterns in the historical data.
3. Two sites are co-located with other sites.
4. Seven sites have no MSF data for all of the years.

These 15 sites were excluded from the datasets used for analyses. Table 4.4 lists these sites and provides the reasons for unsuccessful imputation. In this table, “C” indicates that a TTMS is co-located with another TTMS (therefore, it is not needed), “U” indicates unreliable historical data, “M” indicates no data are available, and “V” indicates that there are large variations in the historical data.

Table 4.4 List of TTMS with Missing Data after Imputation

Index	SITE	Description	Reason
1	019917	US41, 4.8 MI N OF LEE CO (NEAR R 14, 1000 & 117)	C
2	100338	SR583 (56TH ST), 1216 FT S OF SLIGH AVE - HILLS#03	U
3	100339	SR60 (CC CSWY), 1996 FT W ROCKY PT DR - HILLS#18	U
4	100341	SR674-COLLEGE AV, 285 FT W CYPRESS V BLVD-HILLS#53	U
5	269904	I-75/SR-93, 3 MILES NORTH OF MARION COUNTY LINE	V
6	550201	US-319(CAPITAL CIRCLE), 0.3 MI. EAST OF SR-61	C
7	609928	I-10/SR-8, APPROX. 1.3 MI. WEST OF BOY SCOUT ROAD	V
8	799906	I-4, 0.4 MI E ENTERPRISE RD OP -- REPL TTMS 0179	V
9	079918	SR 25/80, US 27 1.6 MI EAST OF SR 80 R-160	M
10	100342	SR45/US41, 574 FT N OF TRENTON ST - HILLS#58	M
11	549901	I10 JEFFERSON CO, APPROX 1.0 MI E OF SR257, WIM#1	M
12	720157	I-295,3.0 MI N OF I-10,WIM#14 -- UC 9/94	M
13	860255	SR 834/SAMPLE RD. 0.14 MI.W OF NW 14TH AVE. TTMS	M
14	860256	SR 818/GRIFFIN RD, 112' WEST OF SW 70TH AVE. TTMS	M
15	920303	I-4/SR-400, APPROX. 0.4 MI. SW OF ORANGE CTY. LINE	M

“C”: Co-located with another site.

“U”: Historical data are unreliable.

“M”: Not included in MSF dataset.

“V”: Cannot be imputed as large variation in historical data.

## 4.2 Definition of Independent Variables

The data used to compile these variables include the following:

- Population and number of occupied hotel/motel rooms at the Traffic Analysis Zone (TAZ) level. These data are estimated by county planning departments or metropolitan planning organizations (MPOs) for their 1999 or 2000 transportation models.
- Population by different age groups, number of retired households by different income groups, number of seasonal households, number of total households, and number of total housing units from the 2000 Census at the census tract level.
- Employment data for the year 2000 from the InfoUSA database purchased by FDOT. For each business establishment, these data include the business name, address, location, business type (identified by a Standard Industrial Classification (SIC) code), number of employees, etc.
- Street network with federal functional classification.
- Land use/cover features categorized according to the Florida Land use and Cover Classification System (FLUCCS).

The independent variables are described in the following subsections.

### 4.2.1 Roadway Characteristic Variables

Variables in this category are summarized in Table 4.5. The data are from the 2000 FDOT Traffic Information and the Roadway Characteristics Inventory (RCI) database. Four variables, *FR*, *PA*, *MA*, and *CO*, are dummy variables that take a value of 0 or 1 and indicate the type of road where the TTMS is located. These variables are used

in the models for both urban and rural areas. Variable *TF* is the value for truck factor, which is only employed by the model for rural areas.

Table 4.5 Roadway Characteristic Variables for Urban Roads

Variable	Description
<i>FR</i>	Equals 0 if TTMS is not located on an urban freeway; 1 otherwise
<i>PA</i>	Equals 0 if TTMS is not located on an urban/rural principal arterial; 1 otherwise
<i>MA</i>	Equals 0 if TTMS is not located on an urban/rural minor arterial; 1 otherwise
<i>CO</i>	Equals 0 if TTMS is not located on an urban/rural collector; 1 otherwise
<i>TF</i>	Truck factor (exclusive to rural areas)

#### 4.2.2 Demographic and Socioeconomic Variables

It is well known that socioeconomic conditions affect the travel behavior of trip makers. The variables in this category are designed to reflect the socioeconomic characteristics of the population in the area surrounding a count station. The variables are compiled using the buffer analysis method.

There are many common variables shared by the models for urban and rural areas. They include:

- Percentage of student population by different age groups.
- Percentage of retired households by different income levels.
- Seasonal household percentage.
- Median household income.
- Employment variables.

Table 4.6 summarizes the description, range, source, and the updating frequency of the variables that are employed by both urban and rural areas. Additional age group variables, shown in Table 4.7, are also tested for rural areas.

Table 4.6 Variables for Urban and Rural Areas

Variable	Description	Range	Source	Update Frequency
<i>Rt Low</i>	Percentage of retired HHs with low income out of total households	0-48.9618	Census	10 years
<i>Rt High</i>	Percentage of retired HHs with high income out of total households	0-38.7589	Census	10 years
<i>RETIRE</i>	Percentage of retired HHs out of total households	1.5672-65.7629	Census	10 years
<i>AgriP</i>	Agriculture workers as a percentage of total workers	0-56.7071	INFO USA	every year
<i>FishP</i>	Fishing & Hunting workers as a percentage of total workers	0-2.25	INFO USA	every year
<i>TranP</i>	Transportation workers as a percentage of total workers	0-52.1255	INFO USA	every year
<i>WholP</i>	Wholesale workers as a percentage of total workers	0-73.2546	INFO USA	every year
<i>RestP</i>	Restaurant workers as a percentage of total workers	0-59.5177	INFO USA	every year
<i>EdP</i>	Education workers as a percentage of total workers	0-76.0180	INFO USA	every year
<i>RecServP</i>	Amusement & Recreation Service workers as a percentage of total workers	0-53.9735	INFO USA	every year
<i>MineP</i>	Mining workers as a percentage of total workers	0-47.5	INFO USA	every year
<i>ManuP</i>	Manufacturing workers as a percentage of total workers	0-95.7	INFO USA	every year
<i>ServP</i>	Service workers as a percentage of total workers	0-104	INFO USA	every year
<i>OffP</i>	Office workers as a percentage of total workers	0-23.67	INFO USA	every year
<i>ST1</i>	Population percentage under 4 years old	0.0001-0.0497	Census	10 years
<i>ST2</i>	Population percentage of ages 5 to 17	0.0005-0.2327	Census	10 years
<i>STU21</i>	Population percentage of ages 5 to 10	0.0002-0.0874	Census	10 years
<i>STU22</i>	Population percentage of ages 11 to 13	0.0001-0.0529	Census	10 years
<i>STU23</i>	Population percentage of ages 14 to 17	0.0001-0.0925	Census	10 years
<i>MInc</i>	Median household income	19235-70939	Census	10 years

Table 4.7 Age Group Variables for Rural Roads

Variable	Description
<i>PPA5</i>	Population aged 5 and under as a percentage of total population
<i>PPA6 17</i>	Population aged between 6 and 17 as a percentage of total population
<i>PPA22 64</i>	Population aged between 22 and 64 as a percentage of total population
<i>PPA18 64</i>	Population aged between 18 and 64 as a percentage of total population
<i>PPA6 21</i>	Population aged between 6 and 21 as a percentage of total population
<i>PPA18 21</i>	Population aged between 18 and 21 as a percentage of total population
<i>PPA65up</i>	Population aged 65 and over as a percentage of total population
<i>PDA5</i>	Population density aged 5 and under
<i>PDA6 17</i>	Population density aged between 6 and 17
<i>PDA22 64</i>	Population density aged between 22 and 64
<i>PDA18 64</i>	Population density aged between 18 and 64
<i>PDA6 21</i>	Population density aged between 6 and 21
<i>PDA18 21</i>	Population density aged between 18 and 21
<i>PDA65up</i>	Population density aged 65 and over

#### 4.2.3 Special Land Use Variables for Urban Models

Variables in this category are designed to account for the effects of special land use types, including universities, tourist attractions, and recreational sites.

##### 4.2.3.1 University Variables

Variables in this category are summarized in Table 4.8. *DLEG* is the variable that represents the impact of legislative sessions on the TTMSs located in Leon County. Since the 13 state universities, as shown in Figure 4.4, have large enrollments and can potentially affect the seasonality of travel, two dummy variables, *SU* and *FU*, are created to distinguish mostly residential universities (University of Florida, Florida State University, Florida A&M University, and University of Miami) from universities that have a significant commuting student body. Because Gainesville and Tallahassee are college towns, the universities' impacts are considered to be county-wide. As for the University of Miami, which is located in a large urban area, the radii for impact area is assumed to be three miles.

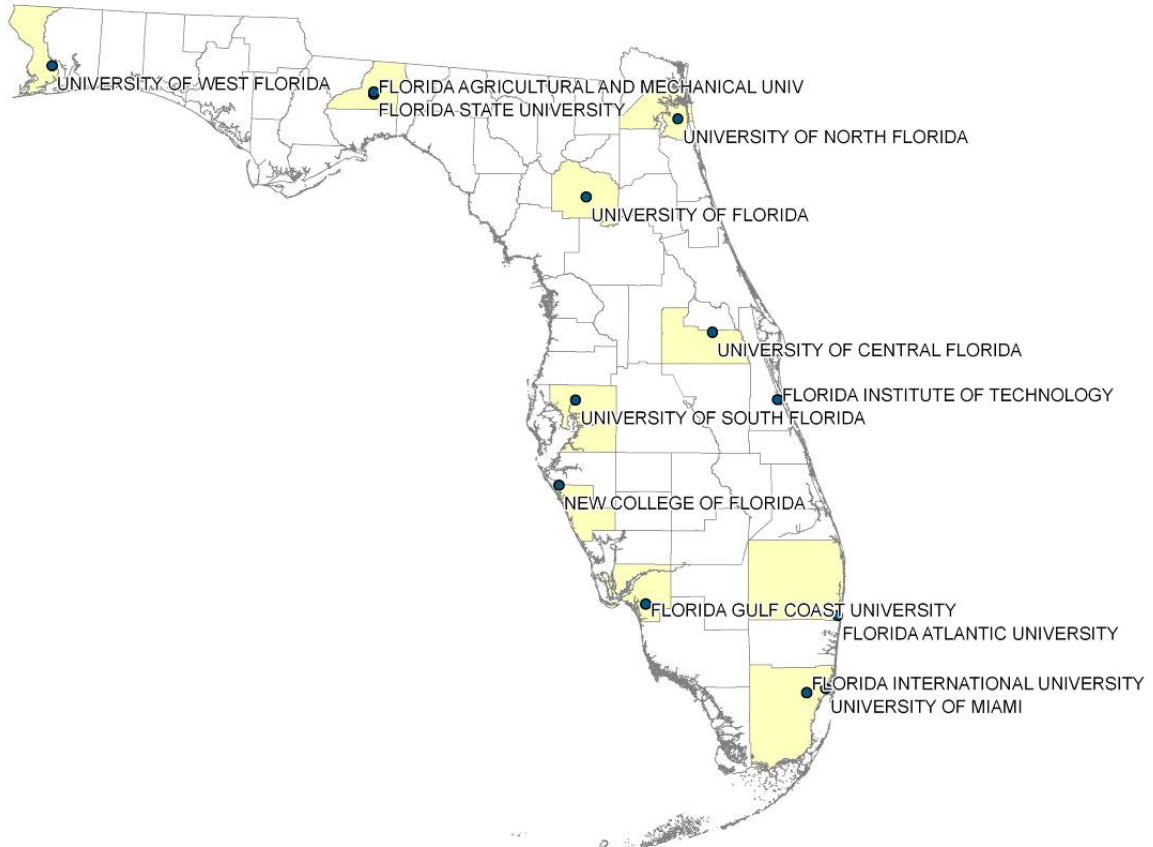


Figure 4.4 Universities with large enrollment in Florida

Table 4.8 Location Characteristic Variables for Urban Roads

Variable	Description
<i>DLEG</i>	Equals 1 if TTMS is located in Leon County; 0 otherwise
<i>SU</i>	Equals 1 if TTMS is in the county of UF, FSU, and FAMU, or within three miles of UM or FIU; 0 otherwise
<i>FU</i>	Equals 1 if TTMS is located within three miles of other state universities; 0 otherwise

#### 4.2.3.2 Tourist Attraction and Recreational Site Variables

The Disney parks and other amusement parks located in Osceola County attract a significant number of tourists. These tourists often generate seasonal traffic. Therefore, the variable *DISNEY* is created to represent the effects of tourism in Osceola County. The variable assumes a value of 1 if a TTMS is located in Osceola County, and 0 otherwise.

Four dummy land use variables are also created. The variables and the land use types they represent are summarized in Table 4.9. The values of these variables for a given TTMS are determined based on whether any portion of the buffer area of the TTMS is one of the four land use types. If a part of the buffer area is one of the four land use types, the corresponding variable assumes a value of 1. Otherwise, the variable for that TTMS is 0.

Table 4.9 Land use Dummy Variables for Urban Road

Variable	FLUCCS Code	Definition
<i>LU1</i>	1810	Swimming Beach
<i>LU2</i>	1820	Golf Course
<i>LU3</i>	1840	Marinas and Fishing Camps
<i>LU4</i>	1850	Parks and Zoos

#### 4.2.4 Location Variables for Rural Models

Rural areas typically have low land use intensity and a higher portion of through traffic. This traffic is not generated locally and cannot be captured by the buffer method. Since the amount of through traffic may be affected by the location of a road in relation to a nearby urban area, beach, or interstate highway, special dummy variables are created to account for such impacts. The distance between a TTMS and an urban area, beach, or interstate highway is subsequently measured. The population size of the urban area is also taken into consideration, as a larger urban area may have a greater impact on a nearby TTMS. These variables are defined in Table 4.10.

Table 4.10 Special Location Variables

Variable	Description
<i>Dist1</i>	Max of ratio of population of a metropolitan area to the distance from the TTMS to the metropolitan area (person/mile)
<i>Indextdist2</i>	$(\sum \frac{\text{Metropolitan population}}{\text{Distance from the TTMS to the metropolitan area}})^{-1}$ ( $10^{-5}$ mile/person)
<i>Interdist</i>	Distance from a TTMS to the closest highway interchange (meter)
<i>Beachdist</i>	Distance from a TTMS to the closest beach site (mile)



### **4.3 Summary**

This chapter describes the data preparation for both dependent and independent variables. The dependent variables are 12 MSFs of TTMSs from the year 2000. Due to the operational environment of the devices, 69 TTMSs are missing a few or all SFs, which comprise more than 20% of all TTMSs. Data imputation based on techniques in trend analysis and averaging was employed for the compilation of the missing MSFs. A total of 54 sites were successfully imputed with MSFs obtained from historical data from the years 1997-2005.

The independent variables reflect land use, demographics, socioeconomics, roadway characteristics, and other variables. They were extracted via GIS techniques, such as buffer analysis and spatial analysis. Several special land use variables were designed to account for the effects of special land use types in urban areas. Location variables, such as the relation between a road location to a nearby urban area, beach, or interstate highway were created to capture the influence of through traffic that is not locally generated.

## CHAPTER V

### MODEL DEVELOPMENT

The seasonal factors were associated with potentially influential variables using multiple regression analyses. The dependent variables are the 12 MSFs. The stepwise selection method was applied, with the significance level set at 0.05 for a variable entering and staying in the model. The  $t$ -statistics and variance inflation factors (VIFs) were also checked for each of the variables used to ensure their significance, as well as to remove multicollinearity within the models. The regression analyses attempted to establish the relationships between the MSFs and potentially influential variables as linear equations. These equations are formulated as follows:

$$MSF_k = \beta_{0k} + \beta_{1k} x_1 + \dots + \beta_{ik} x_i + \dots + \beta_{pk} x_p \quad (5-1)$$

where,

$MSF_k$  = a monthly seasonal factor for month  $k$ ,

$\beta_{ik}$  = the regression coefficient for the  $i$ th independent variable for month  $k$ , and

$x_i$  = the  $i$ th independent variable.

### 5.1 Modeling Influential Variables of Seasonal Factors in Urban Areas

#### 5.1.1 Delineation of Model Areas for Urban Areas

TTMSs in the urban areas are further divided into sub-regions. This was done because Florida stretches over several climate zones, from temperate in the north to subtropical in the south. The seasonal effects of the same variables may vary depending on the latitude.

As shown in Figure 5.1, multi-regression analyses were adopted to further divide the urban areas into three sub-regions, North Florida, Central Florida, and South Florida. Counties in each region are divided into groups to investigate whether counties within the same region may differ in terms of climate, land use, and demographics. If the counties do differ with regard to these variables, the result may include different seasonal traffic patterns. Regression models were estimated first for one group, then for an expanded group with one or more groups of counties added. This was repeated until all groups within the same region are included in the models. After each step, model results were carefully examined. This ensures that the models do not change significantly in terms of  $R^2$  values, influential variables, and coefficients. When such a change is observed, it may indicate that the newly added group of counties may not belong to the region.

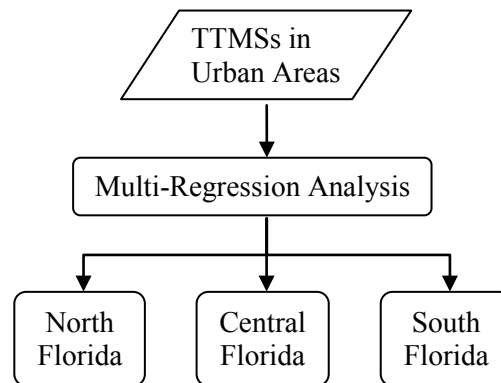
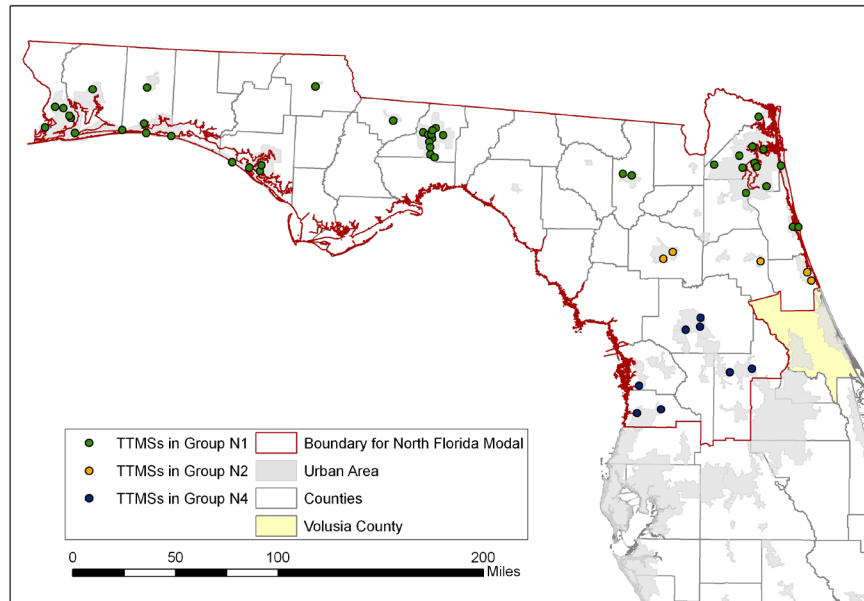


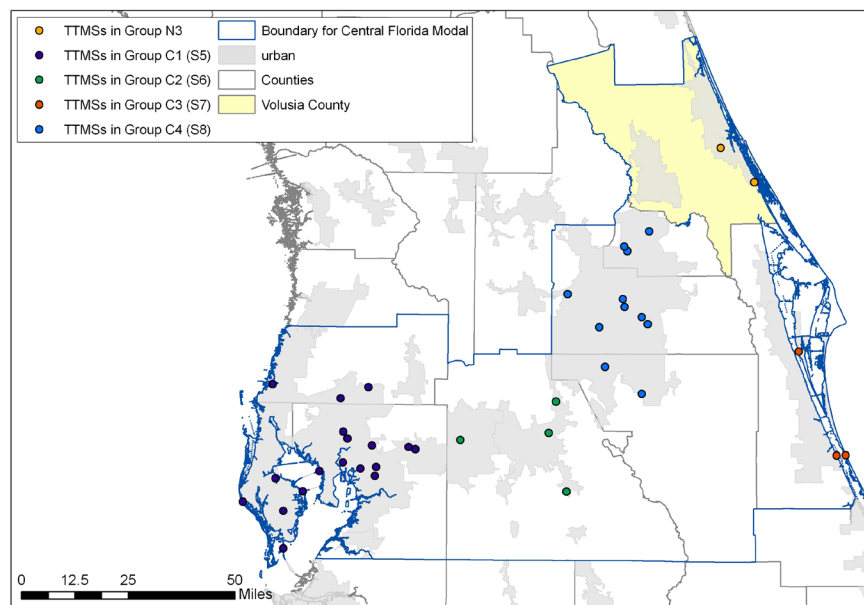
Figure 5.1 Classification of Sub-model Groups for Urban Areas

As an example shown in Figure 5.2, North Florida was originally divided into four groups of counties roughly based on latitude and urban boundaries. Three groups are described in Table 5.1. The fourth group (N3) is Volusia County, which is also highlighted in the maps in Figure 5.2. Modeling results achieved a higher  $R^2$  for both North and Central Florida models by combining the N3 area with the Central Florida

region, instead of the North Florida region. As a result, Volusia County was removed from the North Florida region and included in the Central Florida region.



(a) Boundary of North Florida Model



(b) Boundary of Central Florida Model

Figure 5.2 Classification of Sub-model Regions for North and Central Florida

Based on the regression results, 57 TTMSs are used for the North Florida models, 38 for Central Florida, and 56 for South Florida. The list of the counties for each

model region is summarized in Table 5.1, and the boundaries of the three regions and the TTMS locations are shown in Figure 5.3.

Table 5.1 List of Counties within Each of the Sub-region (Analysis Area)

Model Region	Group	County	# of TTMSs
North Florida	<i>N1</i>	Escambia, Santa Rosa, Okaloosa, Walton, Bay, Jackson, Gadsden, Leon, Columbia, Nassau, Duval, St Johns	44
	<i>N2</i>	Alachua, Putnam, Flagler	5
	<i>N4</i>	Lake, Marion, Citrus, Hernando	8
Central Florida	<i>N3</i>	Volusia	2
	<i>C1 (S5)</i>	Pasco, Hillsborough, Pinellas	18
	<i>C2 (S6)</i>	Polk	4
	<i>C3 (S7)</i>	Brevard	3
	<i>C4 (S8)</i>	Orange, Seminole, Osceola	11
South Florida	<i>S1</i>	Broward, Miami-Dade, Palm Beach	37
	<i>S2</i>	Lee, Collier	4
	<i>S3</i>	Martin, St Lucie, Indian River	9
	<i>S4</i>	Charlotte, Sarasota, Manatee, Desoto	6

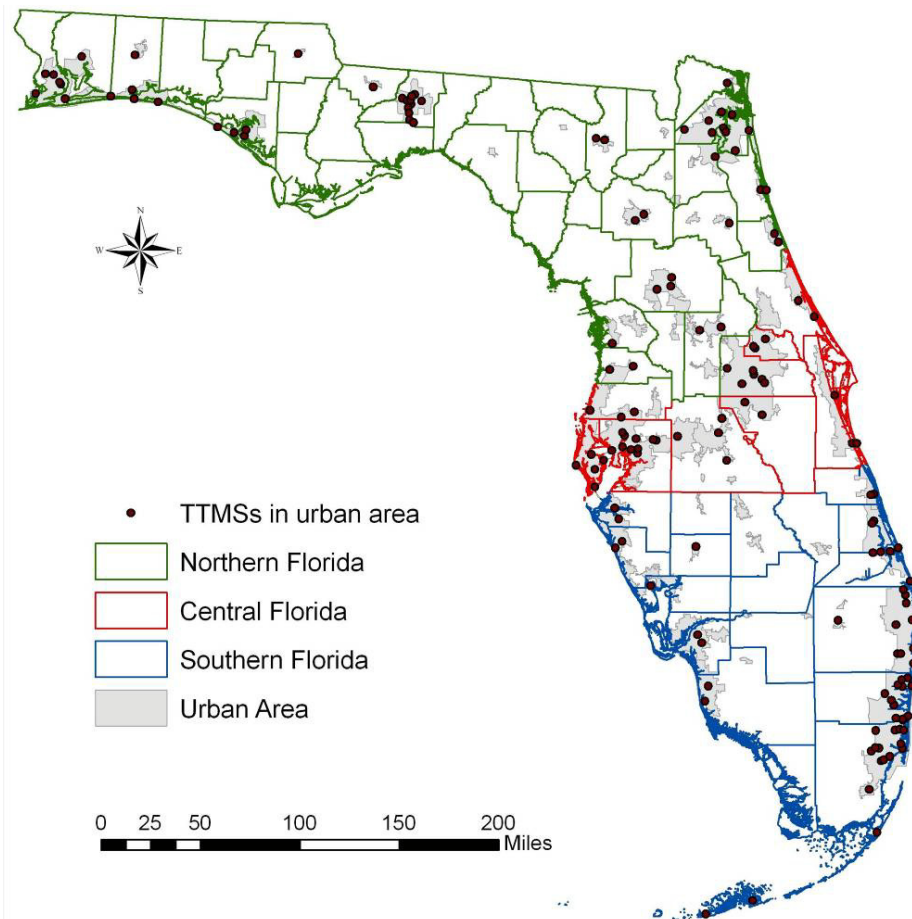


Figure 5.3 Boundaries of Study Areas

### 5.1.2 North Florida Model Results

The regression models of the 12 MSFs for North Florida are given in Table 5.2. A total of 57 TTMSs are included in the models. The variables included in these models above are listed in Table 5.3, along with their partial  $R^2$  values and the months for which they are significant. Tables 5.4 and 5.5 list only those variables that have a partial  $R^2$  greater than 0.05 by name and by partial  $R^2$  value, respectively.

Table 5.2 Regression Models for North Florida (NFL)

Month	Seasonal Factor Equation	$R^2$	Adj. $R^2$	RMSE
JAN	$JAN\_SF = 0.94385 + 1.99490 \times ST22 + 0.00307 \times SHP + 1.12005 \times FishP + 0.01004 \times HotIP + 0.84707 \times MseumP$	0.6581	0.6246	0.0512
FEB	$FEB\_SF = 0.99449 - 0.03588 \times SU - 0.00221 \times RETIRE + 0.76741 \times ST23 + 0.00154 \times SHP + 0.46882 \times MseumP$	0.5477	0.5034	0.0386
MAR	$MAR\_SF = 1.00727 - 0.00260 \times Rt\_Low - 0.00083117 \times SHP - 0.00550 \times HotIP$	0.5227	0.4957	0.0324
APR	$APR\_SF = 0.96401 + 0.04544 \times FR - 0.02808 \times MA - 0.00463 \times HotIP$	0.4087	0.3752	0.0306
MAY	$MAY\_SF = 1.04120 - 0.02183 \times LEG - 0.92003 \times ST23 - 0.00230 \times SHP - 0.85560 \times FishP + 0.00130 \times RestP - 0.00544 \times HotIP - 0.42097 \times MseumP + 0.00191 \times OffP$	0.6934	0.6423	0.0274
JUN	$JUN\_SF = 1.08684 - 2.00258 \times ST22 - 0.00207 \times SHP - 1.00131 \times FishP - 0.00522 \times HotIP - 0.73832 \times MseumP$	0.6506	0.6163	0.0383
JUL	$JUL\_SF = 1.00827 + 0.05153 \times SU + 0.00322 \times Rt\_Low - 1.65684 \times ST22 - 0.00226 \times SHP - 0.00911 \times HotIP + 0.00134 \times EdP - 0.67131 \times MseumP$	0.6269	0.5736	0.0548
AUG	$AUG\_SF = 0.97445 + 0.01612 \times LEG + 0.00275 \times Rt\_Low - 0.47684 \times ST1 - 0.00073571 \times SHP + 0.00327 \times WholP + 0.00161 \times RcServP - 0.27203 \times MseumP$	0.7776	0.7458	0.0180
SEP	$SEP\_SF = 1.01290 - 0.03285 \times SU + 0.00150 \times RETIRE$	0.2879	0.2615	0.0445
OCT	$OCT\_SF = 0.96692 - 0.02716 \times SU + 0.85153 \times ST23 + 0.00201 \times SHP + 0.00781 \times HotIP - 0.00108 \times EdP$	0.5776	0.5362	0.0365
NOV	$NOV\_SF = 1.00212 + 0.00276 \times SHP + 0.00916 \times HotIP + 1.00144 \times MseumP$	0.6020	0.5795	0.0591
DEC	$DEC\_SF = 0.99280 + 0.00348 \times SHP + 0.00544 \times TranP + 0.01648 \times HotIP + 1.37402 \times MseumP$	0.6164	0.5869	0.0851

Table 5.3 Variables from NFL Model Sorted by Month and Partial R<sup>2</sup> Value

Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month
<i>SHP</i>	0.2964	JAN	<i>HotlP</i>	0.0395	MAY	<i>SHP</i>	0.0619	AUG
<i>MseumP</i>	0.1659	JAN	<i>RestP</i>	0.0407	MAY	<i>RcServP</i>	0.0647	AUG
<i>ST22</i>	0.0775	JAN	<i>FishP</i>	0.0368	MAY	<i>WholP</i>	0.0344	AUG
<i>HotlP</i>	0.0709	JAN	<i>OffP</i>	0.0617	MAY	<i>ST1</i>	0.0603	AUG
<i>FishP</i>	0.0473	JAN	<i>LEG</i>	0.0301	MAY	<i>LEG</i>	0.0271	AUG
<i>MseumP</i>	0.1572	FEB	<i>MseumP</i>	0.2701	JUN	<i>RETIRE</i>	0.2256	SEP
<i>RETIRE</i>	0.144	FEB	<i>SHP</i>	0.1298	JUN	<i>SU</i>	0.0623	SEP
<i>SU</i>	0.1046	FEB	<i>ST22</i>	0.1487	JUN	<i>SHP</i>	0.3353	OCT
<i>SHP</i>	0.0851	FEB	<i>FishP</i>	0.0642	JUN	<i>HotlP</i>	0.0767	OCT
<i>ST23</i>	0.0569	FEB	<i>HotlP</i>	0.0377	JUN	<i>SU</i>	0.0673	OCT
<i>Rt Low</i>	0.3359	MAR	<i>MseumP</i>	0.1851	JUL	<i>ST23</i>	0.046	OCT
<i>HotlP</i>	0.1367	MAR	<i>SHP</i>	0.1178	JUL	<i>EdP</i>	0.0522	OCT
<i>SHP</i>	0.0501	MAR	<i>ST22</i>	0.1102	JUL	<i>MseumP</i>	0.3448	NOV
<i>FR</i>	0.1973	APR	<i>Rt Low</i>	0.073	JUL	<i>SHP</i>	0.2035	NOV
<i>HotlP</i>	0.1291	APR	<i>SU</i>	0.0678	JUL	<i>HotlP</i>	0.0537	NOV
<i>MA</i>	0.0822	APR	<i>HotlP</i>	0.0404	JUL	<i>MseumP</i>	0.2973	DEC
<i>SHP</i>	0.2662	MAY	<i>EdP</i>	0.0325	JUL	<i>SHP</i>	0.1648	DEC
<i>ST23</i>	0.1439	MAY	<i>Rt Low</i>	0.3854	AUG	<i>HotlP</i>	0.0742	DEC
<i>MseumP</i>	0.0745	MAY	<i>MseumP</i>	0.1438	AUG	<i>TranP</i>	0.08	DEC

Table 5.4 Variables from NFL Model Sorted by Name and Partial R<sup>2</sup> Value

Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month
<i>EdP</i>	0.0522	OCT	<i>MseumP</i>	0.1438	AUG	<i>ST1</i>	0.0603	AUG
<i>FishP</i>	0.0642	JUN	<i>MseumP</i>	0.0745	MAY	<i>ST22</i>	0.1487	JUN
<i>FR</i>	0.1973	APR	<i>OffP</i>	0.0617	MAY	<i>ST22</i>	0.1102	JUL
<i>HotlP</i>	0.1367	MAR	<i>RcServP</i>	0.0647	AUG	<i>ST22</i>	0.0775	JAN
<i>HotlP</i>	0.1291	APR	<i>RETIRE</i>	0.2256	SEP	<i>ST23</i>	0.1439	MAY
<i>HotlP</i>	0.0767	OCT	<i>RETIRE</i>	0.144	FEB	<i>ST23</i>	0.0569	FEB
<i>HotlP</i>	0.0742	DEC	<i>SHP</i>	0.3353	OCT	<i>Rt Low</i>	0.3854	AUG
<i>HotlP</i>	0.0709	JAN	<i>SHP</i>	0.2964	JAN	<i>Rt Low</i>	0.3359	MAR
<i>HotlP</i>	0.0537	NOV	<i>SHP</i>	0.2662	MAY	<i>Rt Low</i>	0.073	JUL
<i>MA</i>	0.0822	APR	<i>SHP</i>	0.2035	NOV	<i>SU</i>	0.1046	FEB
<i>MseumP</i>	0.3448	NOV	<i>SHP</i>	0.1648	DEC	<i>SU</i>	0.0678	JUL
<i>MseumP</i>	0.2973	DEC	<i>SHP</i>	0.1298	JUN	<i>SU</i>	0.0673	OCT
<i>MseumP</i>	0.2701	JUN	<i>SHP</i>	0.1178	JUL	<i>SU</i>	0.0623	SEP
<i>MseumP</i>	0.1851	JUL	<i>SHP</i>	0.0851	FEB	<i>TranP</i>	0.08	DEC
<i>MseumP</i>	0.1659	JAN	<i>SHP</i>	0.0619	AUG	<i>WholP</i>	0.0344	AUG
<i>MseumP</i>	0.1572	FEB	<i>SHP</i>	0.0501	MAR			

Table 5.5 Variables from NFL Model Sorted by Partial  $R^2$  Value

Variable	Partial $R^2$	Month	Variable	Partial $R^2$	Month	Variable	Partial $R^2$	Month
<i>Rt_Low</i>	0.3854	AUG	<i>RETIRE</i>	0.144	FEB	<i>Rt_Low</i>	0.073	JUL
<i>MseumP</i>	0.3448	NOV	<i>ST23</i>	0.1439	MAY	<i>HotlP</i>	0.0709	JAN
<i>Rt_Low</i>	0.3359	MAR	<i>MseumP</i>	0.1438	AUG	<i>SU</i>	0.0678	JUL
<i>SHP</i>	0.3353	OCT	<i>HotlP</i>	0.1367	MAR	<i>SU</i>	0.0673	OCT
<i>MseumP</i>	0.2973	DEC	<i>SHP</i>	0.1298	JUN	<i>RcServP</i>	0.0647	AUG
<i>SHP</i>	0.2964	JAN	<i>HotlP</i>	0.1291	APR	<i>FishP</i>	0.0642	JUN
<i>MseumP</i>	0.2701	JUN	<i>SHP</i>	0.1178	JUL	<i>SU</i>	0.0623	SEP
<i>SHP</i>	0.2662	MAY	<i>ST22</i>	0.1102	JUL	<i>SHP</i>	0.0619	AUG
<i>RETIRE</i>	0.2256	SEP	<i>SU</i>	0.1046	FEB	<i>OffP</i>	0.0617	MAY
<i>SHP</i>	0.2035	NOV	<i>SHP</i>	0.0851	FEB	<i>ST1</i>	0.0603	AUG
<i>FR</i>	0.1973	APR	<i>MA</i>	0.0822	APR	<i>ST23</i>	0.0569	FEB
<i>MseumP</i>	0.1851	JUL	<i>TranP</i>	0.08	DEC	<i>HotlP</i>	0.0537	NOV
<i>MseumP</i>	0.1659	JAN	<i>ST22</i>	0.0775	JAN	<i>EdP</i>	0.0522	OCT
<i>SHP</i>	0.1648	DEC	<i>HotlP</i>	0.0767	OCT	<i>SHP</i>	0.0501	MAR
<i>MseumP</i>	0.1572	FEB	<i>MseumP</i>	0.0745	MAY			
<i>ST22</i>	0.1487	JUN	<i>HotlP</i>	0.0742	DEC			

The models show that the *SHP* (percentage of seasonal households), *MseumP* (percentage of museums/art/galleries/gardens workers), and *HotlP* (percentage of hotel & camp workers) variables appear in most of the models with relatively large partial  $R^2$  values. The *Rt\_Low* (percentage of retired households with low income), *RETIRE* (percentage of retired households), *FR* (freeway), *ST22* (percentage of population ages 11-13), and *ST23* (percentage of population ages 14-17) are some of the variables that appear infrequently. However, these variables have noticeably partial  $R^2$  values when they do appear in the models.

In general, variables representing tourist-related activities, such as fishing, visiting a museum, and hotel-related employment, tend to have a positive coefficient in the winter months and a negative coefficient in the summer months. This suggests that the tourist season is summer in North Florida.

The variable that represents residential universities, *SU*, appears in the February, September, and October models with a negative coefficient. In contrast, it appears in the



July model with a positive coefficient. This is possibly due to an increase in traffic related to the beginning of the academic year and a decrease in traffic caused by holidays during the summer.

The variable *SHP* appears in all models, except for April and September. Note that there is a positive coefficient during the colder months (from October to February) and a negative coefficient during the warmer months (March, and from May to August). This shows that there may be more seasonal households in North Florida during the warmer months than the colder months.

Variables representing the student population ages 11-17 (*ST22* and *ST23*) tend to be associated with more traffic during the summer vacation season (May, June, and July), but less traffic during January and February.

The variable that represents low income retired households, *Rt\_Low*, is included in the March model with a negative coefficient, and the August model with a positive coefficient. Similarly, the variable *RETIRE* displays a negative coefficient in the February model, while a positive coefficient is displayed in the October model. This suggests that low-income retired households tend to have greater activity during February and March than in August and September.

### 5.1.3 Central Florida Model Results

The regression models of the MSFs in Central Florida (CFL) are shown in Table 5.6. A total of 38 TTMSs are included in the models. For the Central Florida models, significant variables are *AgriP* (percentage of agriculture employment) and *RETIRE* (percentage of retired households). The *AgriP* variable appears most often and

contributes a large portion of partial  $R^2$  to the June, August, September, and October models with a positive coefficient.

Table 5.6 Regression Models for Central Florida (CFL)

Month	Seasonal Factor Equation	$R^2$	Adj. $R^2$	RMSE
JAN	$JAN\_SF = 1.05296 + 0.04354 \times LU2 + 0.08528 \times LU3 - 0.00253 \times Rt\_Low$	0.4966	0.4522	0.0386
FEB	$FEB\_SF = 0.99220 - 0.00192 \times RETIRE$	0.3266	0.3079	0.0329
MAR	$MAR\_SF = 0.95455 + 0.04189 \times PA - 0.00197 \times SHP - 0.03780 \times AgriP - 0.00106 \times RtlP$	0.5923	0.5429	0.0328
APR	$APR\_SF = 0.96193 - 0.03752 \times MineP$	0.1130	0.0884	0.0258
MAY	$MAY\_SF = 1.04696 + 0.05252 \times CO - 9.35253E-7 \times Minc - 0.00052936 \times ServP$	0.3670	0.3111	0.0202
JUN	$JUN\_SF = 0.98370 + 0.00130 \times RETIRE + 0.05039 \times AgriP$	0.5481	0.5223	0.0230
JUL	$JUL\_SF = 1.14257 - 0.08904 \times DISN - 0.05963 \times LU4 - 0.00000203 \times Minc$	0.3377	0.2793	0.0470
AUG	$AUG\_SF = 1.03812 - 0.02298 \times PA - 0.06810 \times DISN - 0.03790 \times LU2 + 0.00111 \times SHP + 0.09428 \times AgriP - 0.02329 \times OffP$	0.7671	0.7220	0.0292
SEP	$SEP\_SF = 1.14004 - 0.04809 \times PA - 1.59261 \times ST22 + 0.06048 \times AgriP$	0.5357	0.4947	0.0335
OCT	$OCT\_SF = 1.04013 - 0.09017 \times CO + 0.00326 \times Rt\_High - 1.05551 \times ST23 + 0.04152 \times AgriP$	0.5600	0.5066	0.0229
NOV	$NOV\_SF = 1.01727 + 0.06101 \times MA + 0.00409 \times TranP - 0.00308 \times EdP$	0.4189	0.3676	0.0251
DEC	$DEC\_SF = 1.00861 + 0.08639 \times MA$	0.1336	0.1096	0.0505

Of the variables that describe tourist attractions and recreational sites,  $LU2$  (the golf course variable) and  $LU3$  (the marinas/fishing camp variable) are included in the January model with a positive coefficient, indicating a decreased level of travel related to these activities. The  $LU4$  (i.e., parks and zoos variable) variable is selected by the July model with a negative coefficient, indicating an increase in travel to parks and zoos. The variable  $DISNEY$  shows up in models for July and August with a negative coefficient. This means traffic around Disney parks tends to increase during these two months, which coincide with summer vacation time for schools.

Roadway characteristic variables ( $PA$ ,  $MA$ , and  $CO$ ) are selected by two models. The  $PA$  (principal arterial) variable enters the March model with a positive coefficient

and the September model with a negative coefficient. This suggests that principal arterials tend to have more traffic in September, but lower traffic in March in the rural areas. The *MA* (minor arterial) variable is selected by the November and December models. Their coefficients are both positive, suggesting that minor arterials carry less traffic during the last two months of a year. The *CO* (collector) variable appears in the May model with a positive coefficient, and in the October model with a negative coefficient. This indicates that traffic on collectors tends to increase during October and decrease during May. The model results also show that retired households seem to contribute to the increase in traffic during February and the decrease during June. The *MInc* (median household income) variable appears in the May and July models with a negative coefficient. This suggests that during these two months in Central Florida, the higher the median income households generate more traffic.

Table 5.7 lists all the variables included in the models above, along with their partial  $R^2$  values and the months for which they are significant. Tables 5.8 and 5.9 list all of the variables that have a partial  $R^2$  larger than 0.05 by name and partial  $R^2$  value, respectively.

Table 5.7 Variables from CFL Model Sorted by Month and Partial  $R^2$  Value

Variable	Partial $R^2$	Month	Variable	Partial $R^2$	Month	Variable	Partial $R^2$	Month
<i>Rt Low</i>	0.2177	JAN	<i>RETIRE</i>	0.2019	JUN	<i>ST23</i>	0.1912	OCT
<i>LU3</i>	0.1914	JAN	<i>MInc</i>	0.1334	JUL	<i>AgriP</i>	0.1494	OCT
<i>LU2</i>	0.0875	JAN	<i>DISN</i>	0.1193	JUL	<i>Rt High</i>	0.0832	OCT
<i>RETIRE</i>	0.3266	FEB	<i>LU4</i>	0.0851	JUL	<i>CO</i>	0.1363	OCT
<i>SHP</i>	0.2821	MAR	<i>AgriP</i>	0.4459	AUG	<i>MA</i>	0.1797	NOV
<i>PA</i>	0.1863	MAR	<i>SHP</i>	0.1321	AUG	<i>EdP</i>	0.1144	NOV
<i>AgriP</i>	0.0639	MAR	<i>DISN</i>	0.0584	AUG	<i>TranP</i>	0.1248	NOV
<i>RtlP</i>	0.0601	MAR	<i>OffP</i>	0.0572	AUG	<i>MA</i>	0.1336	DEC
<i>MineP</i>	0.113	APR	<i>LU2</i>	0.0367	AUG			
<i>CO</i>	0.1625	MAY	<i>PA</i>	0.0369	AUG			
<i>MInc</i>	0.1198	MAY	<i>PA</i>	0.2403	SEP			
<i>ServP</i>	0.0846	MAY	<i>ST22</i>	0.1238	SEP			
<i>AgriP</i>	0.3462	JUN	<i>AgriP</i>	0.1716	SEP			

Table 5.8 Variables from CFL Model Sorted by Name and Partial  $R^2$  Value

Variable	Partial $R^2$	Month	Variable	Partial $R^2$	Month	Variable	Partial $R^2$	Month
<i>TranP</i>	0.1248	NOV	<i>PA</i>	0.1863	MAR	<i>DISN</i>	0.0584	AUG
<i>Rt Low</i>	0.2177	JAN	<i>OffP</i>	0.0572	AUG	<i>CO</i>	0.1625	MAY
<i>Rt High</i>	0.0832	OCT	<i>MineP</i>	0.113	APR	<i>CO</i>	0.1363	OCT
<i>ST23</i>	0.1912	OCT	<i>MInc</i>	0.1334	JUL	<i>AgriP</i>	0.4459	AUG
<i>ST22</i>	0.1238	SEP	<i>MInc</i>	0.1198	MAY	<i>AgriP</i>	0.3462	JUN
<i>SHP</i>	0.2821	MAR	<i>MA</i>	0.1797	NOV	<i>AgriP</i>	0.1716	SEP
<i>SHP</i>	0.1321	AUG	<i>MA</i>	0.1336	DEC	<i>AgriP</i>	0.1494	OCT
<i>ServP</i>	0.0846	MAY	<i>LU4</i>	0.0851	JUL	<i>AgriP</i>	0.0639	MAR
<i>RtlP</i>	0.0601	MAR	<i>LU3</i>	0.1914	JAN	<i>TranP</i>	0.1248	NOV
<i>RETIRE</i>	0.3266	FEB	<i>LU2</i>	0.0875	JAN	<i>Rt Low</i>	0.2177	JAN
<i>RETIRE</i>	0.2019	JUN	<i>EdP</i>	0.1144	NOV			
<i>PA</i>	0.2403	SEP	<i>DISN</i>	0.1193	JUL			

Table 5.9 Variables from CFL Model Sorted by Partial  $R^2$  Value

Variable	Partial $R^2$	Month	Variable	Partial $R^2$	Month	Variable	Partial $R^2$	Month
<i>AgriP</i>	0.4459	AUG	<i>AgriP</i>	0.1716	SEP	<i>EdP</i>	0.1144	NOV
<i>AgriP</i>	0.3462	JUN	<i>CO</i>	0.1625	MAY	<i>MineP</i>	0.113	APR
<i>RETIRE</i>	0.3266	FEB	<i>AgriP</i>	0.1494	OCT	<i>LU2</i>	0.0875	JAN
<i>SHP</i>	0.2821	MAR	<i>CO</i>	0.1363	OCT	<i>LU4</i>	0.0851	JUL
<i>PA</i>	0.2403	SEP	<i>MA</i>	0.1336	DEC	<i>ServP</i>	0.0846	MAY
<i>Rt Low</i>	0.2177	JAN	<i>MInc</i>	0.1334	JUL	<i>Rt High</i>	0.0832	OCT
<i>RETIRE</i>	0.2019	JUN	<i>SHP</i>	0.1321	AUG	<i>AgriP</i>	0.0639	MAR
<i>LU3</i>	0.1914	JAN	<i>TranP</i>	0.1248	NOV	<i>RtlP</i>	0.0601	MAR
<i>ST23</i>	0.1912	OCT	<i>ST22</i>	0.1238	SEP	<i>DISN</i>	0.0584	AUG
<i>PA</i>	0.1863	MAR	<i>MInc</i>	0.1198	MAY	<i>OffP</i>	0.0572	AUG
<i>MA</i>	0.1797	NOV	<i>DISN</i>	0.1193	JUL			

#### 5.1.4 South Florida Model Results

Regression models for the MSFs in South Florida (SFL) are shown in Table 5.10. A total of 56 TTMSs are included in the models. The variables in the models are listed by month and sorted by their partial  $R^2$  value in Table 5.11. The variables with a partial  $R^2$  larger than 0.05 are sorted by name in Table 5.12, and by partial  $R^2$  value in Table 5.13.

The most significant variable for South Florida is *SHP* (percentage of seasonal households). This appears in eight models and also contributes the largest portion of  $R^2$ . The *SHP* variable appears in the first four models (from January to April) with a negative coefficient, and in the next four models (from May to August) with a positive coefficient.

This indicates that seasonal households tend to reside during the winter months in South Florida and leave in the summer months. The *RETIRE* (percentage of retired households) variable is only included in the May and November models, but the partial  $R^2$  contributed by this variable is noticeable. The negative coefficient for this variable in the November model and the positive coefficient in the May model suggest that retired households are inclined to increase activities during winter and decrease activities in May.

Table 5.10 Regression Models for South Florida (SFL)

Month	Seasonal Factor Equation (S1234 Area)	$R^2$	Adj. $R^2$	RMSE
JAN	$JAN\_SF = 0.99986 - 0.00174 \times SHP + 0.00311 \times ManuP$	0.4451	0.4242	0.0405
FEB	$FEB\_SF = 0.95036 - 0.00260 \times SHP + 0.00271 \times ManuP$	0.5490	0.5319	0.0431
MAR	$MAR\_SF = 0.95697 - 0.00246 \times SHP$	0.5224	0.5135	0.0375
APR	$APR\_SF = 0.98516 - 0.00126 \times SHP$	0.4226	0.4119	0.0234
MAY	$MAY\_SF = 0.97891 + 0.00114 \times RETIRE - 0.00931 \times AgriP$	0.3284	0.3030	0.0207
JUN	$JUN\_SF = 1.01870 + 0.00242 \times SHP$	0.5719	0.5640	0.0334
JUL	$JUL\_SF = 1.06287 + 0.00271 \times SHP - 0.00624 \times HotIP - 0.00284 \times ManuP$	0.4763	0.4460	0.0438
AUG	$AUG\_SF = 1.01573 + 0.00253 \times SHP$	0.4649	0.4550	0.0433
SEP	$SEP\_SF = 1.04084 + 0.00226 \times SHP + 0.00730 \times RcServP$	0.4448	0.4238	0.0489
OCT	$OCT\_SF = 1.01530 + 0.00530 \times HotIP$	0.2247	0.2103	0.0355
NOV	$NOV\_SF = 1.02076 + 0.05257 \times MA - 0.00106 \times RETIRE + 0.73320 \times ST1 - 1.09422 \times ST23$	0.3956	0.3482	0.0241
DEC	$DEC\_SF = 0.96605 + 0.06337 \times SU + 0.07313 \times MseumP$	0.2483	0.2199	0.0332

Table 5.11 Variables from SFL Model Sorted by Month and Partial  $R^2$  Value

Variable	Partial $R^2$	Month	Variable	Partial $R^2$	Month	Variable	Partial $R^2$	Month
<i>SHP</i>	0.3394	JAN	<i>SHP</i>	0.5719	JUN	<i>MA</i>	0.1637	NOV
<i>ManuP</i>	0.1057	JAN	<i>SHP</i>	0.3305	JUL	<i>RETIRE</i>	0.1102	NOV
<i>SHP</i>	0.4916	FEB	<i>HotIP</i>	0.0751	JUL	<i>ST23</i>	0.0659	NOV
<i>ManuP</i>	0.0573	FEB	<i>ManuP</i>	0.0707	JUL	<i>ST1</i>	0.0558	NOV
<i>SHP</i>	0.5224	MAR	<i>SHP</i>	0.4649	AUG	<i>MseumP</i>	0.1488	DEC
<i>SHP</i>	0.4226	APR	<i>SHP</i>	0.3681	SEP	<i>SU</i>	0.0995	DEC
<i>RETIRE</i>	0.2599	MAY	<i>RcServP</i>	0.0766	SEP			
<i>AgriP</i>	0.0685	MAY	<i>HotIP</i>	0.2247	OCT			

Table 5.12 Variables from SFL Model Sorted by Name and Partial R<sup>2</sup> Value

Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month
<i>AgriP</i>	0.0685	MAY	<i>RcServP</i>	0.0766	SEP	<i>SHP</i>	0.3681	SEP
<i>HotlP</i>	0.2247	OCT	<i>RETIRE</i>	0.2599	MAY	<i>SHP</i>	0.3394	JAN
<i>HotlP</i>	0.0751	JUL	<i>RETIRE</i>	0.1102	NOV	<i>SHP</i>	0.3305	JUL
<i>MA</i>	0.1637	NOV	<i>SHP</i>	0.5719	JUN	<i>ST1</i>	0.0558	NOV
<i>ManuP</i>	0.1057	JAN	<i>SHP</i>	0.5224	MAR	<i>ST23</i>	0.0659	NOV
<i>ManuP</i>	0.0707	JUL	<i>SHP</i>	0.4916	FEB	<i>SU</i>	0.0995	DEC
<i>ManuP</i>	0.0573	FEB	<i>SHP</i>	0.4649	AUG			
<i>MseumP</i>	0.1488	DEC	<i>SHP</i>	0.4226	APR			

Table 5.13 Variables from SFL Model Sorted by Partial R<sup>2</sup> Value

Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month
<i>SHP</i>	0.5719	JUN	<i>RETIRE</i>	0.2599	MAY	<i>HotlP</i>	0.0751	JUL
<i>SHP</i>	0.5224	MAR	<i>HotlP</i>	0.2247	OCT	<i>ManuP</i>	0.0707	JUL
<i>SHP</i>	0.4916	FEB	<i>MA</i>	0.1637	NOV	<i>AgriP</i>	0.0685	MAY
<i>SHP</i>	0.4649	AUG	<i>MseumP</i>	0.1488	DEC	<i>ST23</i>	0.0659	NOV
<i>SHP</i>	0.4226	APR	<i>RETIRE</i>	0.1102	NOV	<i>ManuP</i>	0.0573	FEB
<i>SHP</i>	0.3681	SEP	<i>ManuP</i>	0.1057	JAN	<i>ST1</i>	0.0558	NOV
<i>SHP</i>	0.3394	JAN	<i>SU</i>	0.0995	DEC			
<i>SHP</i>	0.3305	JUL	<i>RcServP</i>	0.0766	SEP			

## 5.2 Modeling Influential Variables of Seasonal Factors in Rural Areas

Preliminary regression analyses of the MSFs for rural TTMSs indicate that employing similar variables will result in poor regression models. The link between the MSFs and the independent variables describing demographic, socioeconomic, and roadway characteristics is weak. One reason for this may be that the monthly variation in traffic is more significant on rural roads than on urban and commuter routes (HCM 2000). In addition, another cause of poor model results may be that urban traffic is dominated by commuting. In the rural areas, there is often lack of commuters, and most traffic may be generated from other activities such as agriculture, mining, fishing, recreational travel, etc. Due to low land use intensity, irregular road networks, and longer travel distances, the generators of such activities are difficult to capture for a given TTMS.

A strategy was developed to separately model TTMSs based on their daily traffic patterns in order to improve model results (Lu et al. 2012). Section 5.2.1 describes the method used to classify the rural TTMSs into two groups: one with daily traffic patterns characterized by a single peak, and the other with patterns characterized by double peaks. Figure 5.4 below summarizes the methods for classifying the TTMSs into sub-model groups.

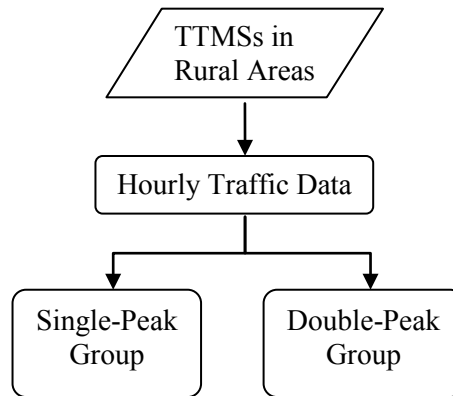


Figure 5.4 Classification of Sub-model Groups for Rural Areas

#### 5.2.1 Classification of Hourly Traffic Pattern for Rural TTMSs

In this section, a method proposed by Lu et al. (2012) is introduced to classify roads based on their daily traffic patterns. This approach separates roads that have a significant portion of commuter traffic from those that do not. Based on their location, the rural TTMSs are then classified into two groups, with each group modeled separately. The purpose of this division is to reduce the variability in the data within the same group and to improve model results. This also helps to identify the independent variables that are most relevant to each group of TTMSs.

The classification of TTMSs based on whether commuting traffic is noticeable or not is achieved by examining the hourly traffic pattern at a TTMS. A traffic pattern dominated by commuter travel usually shows two peaks, one in the morning and one in

the afternoon. The traffic pattern on a road that is used by fewer commuters, but by more people for recreational purposes, typically exhibits a single peak around mid-day. Therefore, a method is used to determine whether the given hourly traffic exhibits a single-peak (SP) or a double-peak (DP) pattern.

There were 116 TTMSs in the rural areas of Florida. Their hourly traffic patterns are determined based on the data collected during a typical weekday. The representative weekday is chosen as Wednesday in the year 2000. The hourly traffic volumes for all Wednesdays were extracted for each TTMS, and were then averaged to arrive at their annual average weekday hourly volumes.

The maximum and minimum hourly volumes were examined to determine if the hourly traffic pattern of a site exhibits a single peak (SP) or double peaks (DPs). Figures 5.5 and 5.6 display a single-peak and a double-peak traffic pattern, respectively. For both SP and DP patterns,  $Max_1$  is defined as the maximum hourly traffic volume in the morning from hour 0 (0:00) to hour 10 (10:00). The  $Max_2$  variable is the maximum hourly traffic volume in the afternoon from hour 15 (15:00) to hour 24 (24:00). The  $Min\_midday$  variable is the minimum hourly volume between the hour 10 (10:00) and hour 15 (15:00).

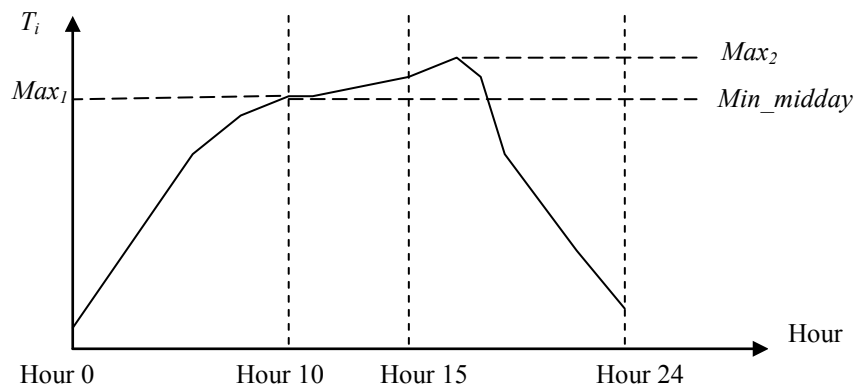


Figure 5.5 Single-peak Pattern and Variables Describing Peaking Characteristics



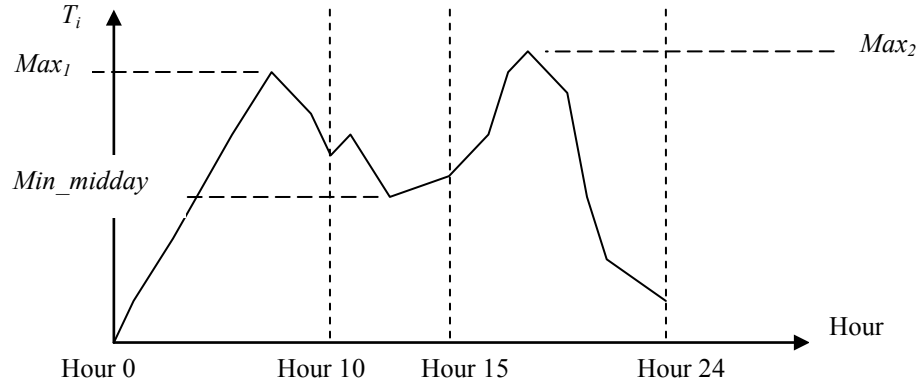


Figure 5.6 Double-peak Pattern and Variables Describing Peaking Characteristics

Determining the presence of double peaks involves checking if both the morning peak traffic volume  $Max_1$  and afternoon peak traffic volume  $Max_2$  are larger than the minimum traffic volume between hour 10 and hour 15 (i.e.,  $Min\_midday$ ). The smaller of the morning peak volume and the afternoon peak volume is defined as follows:

$$Min\_peak = \min\{Max_1, Max_2\} \quad (5-2)$$

The difference between  $Min\_peak$  and  $Min\_midday$  indicates the magnitude of the variation in midday traffic, which is defined below:

$$D = Min\_peak - Min\_midday \quad (5-3)$$

$Max$  is defined as the maximum hourly traffic volume for an entire day, such that:

$$Max = \max\{T_i\} \quad (5-4)$$

where  $T_i$  is the traffic for hour  $i$  ( $i = 1, 2, \dots, 24$ ). The  $D$  variable can be normalized by dividing it by  $Max$ . This determines the difference between the smaller of the peak traffic volumes and the minimum midday traffic volume as a percentage of the maximum daily hourly traffic:

$$PD = \frac{D}{Max} \quad (5-5)$$

An hourly traffic pattern is classified based on the value of  $PD$ . If the  $Min\_peak$  is not greater than the  $Min\_midday$ , it means that at least one of  $Max_1$  and  $Max_2$  is equal to the minimum hourly traffic between hours 10 and 15 (see Figure 5.5). In this case,  $PD$  is 0, which suggests that the traffic pattern has a single peak either at noon or in the early afternoon (seldom in the morning). Theoretically, it is possible to have two peaks: one morning or afternoon peak, and one that may appear between hours 10 and 15. This would result in a TTMS being wrongly classified as having a single peak; however, this was not observed within the data from the 116 TTMSs.

If  $PD$  is greater than 0, there exists at least two peaks. The value of  $PD$  indicates how great the difference is between the minimum peak traffic volume and the minimum midday traffic volume. If  $PD$  is small enough, the pattern is considered single-peaked. The hourly traffic patterns of three TTMSs with a single peak are plotted in Figure 5.7(a) for illustration purposes. The criterion applied to classify these TTMSs into the SP group is  $PD = 0$ . The traffic patterns of three other TTMSs with double peaks are shown in Figure 5.7(b). From these two figures, it can be seen that this criterion has worked reasonably well.

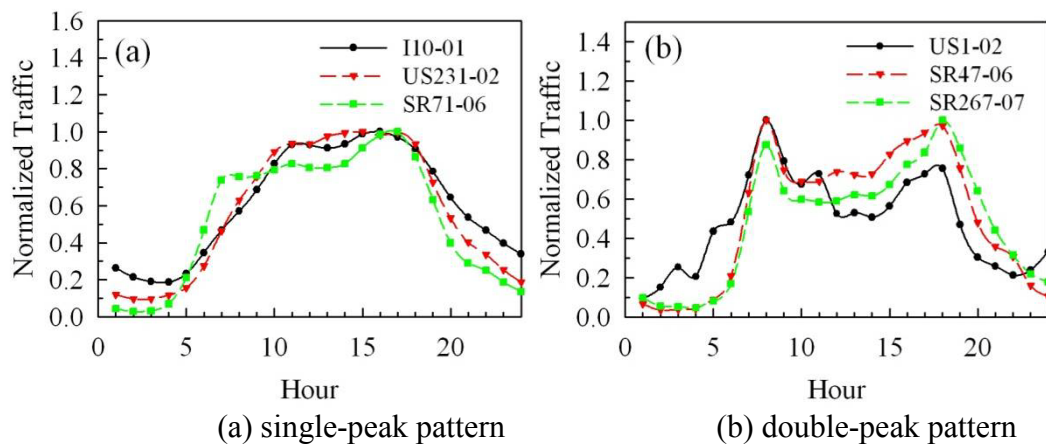


Figure 5.7 Hourly Traffic Variations for Selected TTMSs

Four different scenarios were tested for cases of  $PD = 0.00, 0.05, 0.10$ , and  $0.20$ , to investigate which scenarios resulted in better models. Each of the four criterion values of  $PD$  ( $0, 0.05, 0.1$ , and  $0.2$ ) led to two sets of models, one for the SP group and one for the DP group. A total of eight sets of models were therefore developed. Each set included 12 monthly models. Table 5.14 lists the adjusted  $R^2$  values for the 12 MSF models for the four criterion values of  $PD$ . The first column indicates the month. The second column provides the  $R^2$  values for models that were calibrated with all 116 TTMSs, without separating them into SP and DP groups. Columns 3 through 6 list the  $R^2$  values for the SP group models corresponding to the four criterion values of  $PD$ . Columns 7 through 10 show the  $R^2$  values for DP group models corresponding to the different criterion values of  $PD$ . The number of TTMSs used to develop the models is shown in the second row. The third row provides the criterion values of  $PD$  used to develop the models. It is possible that there may be a third peak during the midday period, but this rarely happens. Only two cases of a third peak have been observed, and are still considered similar to the double-peak TTMSs.

Table 5.14 Comparison of Adjusted R-square for Different Monthly SF Models.

	ALL	SP Models				DP Models			
No. of TTMSs	116	33	44	55	73	83	72	61	43
$PD$ Value	NA	0.00	0.05	0.10	0.20	0.00	0.05	0.10	0.20
JAN	0.484	0.934	0.602	0.489	0.484	0.573	0.541	0.587	0.600
FEB	0.657	0.820	0.688	0.647	0.613	0.606	0.654	0.621	0.593
MAR	0.589	0.782	0.438	0.490	0.489	0.552	0.514	0.680	0.777
APR	0.297	0.635	0.175	0.179	0.180	0.230	0.224	0.489	0.629
MAY	0.218	0.489	0.491	0.488	0.420	0.256	0.254	0.401	0.126
JUN	0.452	0.621	0.531	0.562	0.447	0.491	0.410	0.436	0.446
JUL	0.501	0.900	0.511	0.484	0.465	0.646	0.610	0.552	0.577
AUG	0.520	0.641	0.472	0.606	0.571	0.541	0.526	0.688	0.724
SEP	0.531	0.812	0.625	0.453	0.549	0.445	0.533	0.537	0.662
OCT	0.238	0.370	0.381	0.439	0.469	0.241	0.204	0.255	0.463
NOV	0.347	0.764	0.463	0.359	0.465	0.323	0.387	0.408	0.302
DEC	0.408	0.607	0.596	0.490	0.471	0.358	0.310	0.365	0.474

The model  $R^2$  values suggest that modeling TTMSs separately with SP or DP patterns improves the explanatory power of the models. Although the improvement in model  $R^2$  values for the DP models is not significant, the SP models have much higher  $R^2$  values than the models without the SP and DP classifications. By comparing models based on the different criterion values of  $PD$ , the SP models with criteria  $PD = 0$  have overall higher adjusted  $R^2$  values. This suggests that the classification of TTMS hourly traffic patterns may take the cutoff criterion  $PD = 0$ .

After the TTMSs were classified into the single-peak and double-peak groups, regression models were developed to relate the SFs with variables that describe land use, accessibility, and roadway characteristics. Four variables  $SHP$ ,  $HotIP$ ,  $RtlP$  and  $MseumP$ , are defined for each of the climate zones by the prefix  $N$ ,  $C$ , or  $S$  to indicate whether a TTMS is located in North, Central, or South Florida. These variables, therefore, become  $NSHP$ ,  $CSHP$ ,  $SSHP$ ,  $NHotIP$ ,  $CHotIP$ ,  $SHotIP$ ,  $NRtlP$ ,  $CRtlP$ ,  $SRtlP$ ,  $NMseumP$ ,  $CMseumP$ , and  $SMseumP$ , respectively. The modeling results are presented in the next two sections.

### 5.2.2 Single-Peak Models

A total of 33 TTMSs belonged to the SP group. The regression models for the SP group are shown in Table 5.15. For the SP group models, the  $R^2$  values are between 0.5529 and 0.9468. The only exception is the model for October with an  $R^2$  value of 0.4093. Overall, these  $R^2$  values are much higher than those of the models generated by considering all the TTMSs in the rural areas as one group.

The most significant variables are location variables, i.e.,  $DistI$ ,  $SrtlP$ , and  $SSHP$ . Variables  $SrtlP$  and  $SSHP$  indicate that climate is an important factor. Moreover, they

indicate that the same types of employment do not necessarily affect traffic seasonality during the same months in different climate zones.

The *Dist1* variable contributes approximately 0.3 partial  $R^2$  to the March, April, June, and December models. The coefficient is negative in the models for March, April, and December. This suggests that the closer a count station is to an urban area, the more traffic it may experience during these months.

The *SRtlP* variable appears in seven models and contributes high partial  $R^2$  values to the January, July, August, and October models. For the January model, the coefficient of this variable is negative, while the coefficient is positive in the other three models. This indicates that retail-related employment in South Florida tends to increase traffic during January. In contrast, decreased traffic occurs during July, August, and October.

The *SSHPP* variable is selected by four models: January, February, September, and October. The coefficients for January and February are negative, while those for September and October are positive. This indicates that the seasonal households in South Florida were active during the first two months of a year, but not in September and October.

The variables *PPA18\_64* (age group 18-64) and *PPA22\_64* (age group 22-64) also appear frequently, with relatively high partial  $R^2$  in the January, February, June, and July models. They are correlated with an increase in traffic during winter time and contribute to a decrease in traffic during summer time.

Table 5.16 lists all of the variables included in the models above, along with their partial  $R^2$  values and the months for which they are significant. Tables 5.17 and 5.18 list

all the variables that have a partial  $R^2$  larger than 0.05 by name and partial  $R^2$  value, respectively.

Table 5.15 Regression Models for the Single-Peak Group for Rural Areas

Month	Seasonal Factor Equation (Single Peak Group)	$R^2$	Adj. $R^2$	RMSE
JAN	$JAN\_SF=0.85572+0.00039208 \times Indexdist2-1.16530 \times ST23+0.48158 \times PPA18\_64-0.00268 \times SSHP-0.00623 \times SRtlP+0.01111 \times SHotlP-0.41836 \times CMseumP-0.00392 \times TranP-0.00269 \times WholP+0.00153 \times EdP+0.00159 \times ManuP$	0.9468	0.9189	0.0244
FEB	$FEB\_SF=0.76702+0.45899 \times PPA18\_64-0.00092932 \times NSHP-0.00263 \times SSHP-0.00698 \times SRtlP-0.00240 \times WholP$	0.8480	0.8198	0.0367
MAR	$MAR\_SF=0.81821-0.00204 \times TF-0.00000214 \times Distl+0.35360 \times PPA22\_64-0.00120 \times NSHP-0.00446 \times SRtlP-0.19837 \times NMseumP-0.06868 \times SMseumP$	0.8299	0.7823	0.0310
APR	$APR\_SF=1.00503-0.00000112 \times Distl-4.77704E-7 \times Interdist+0.00077162 \times NRtlP+0.41417 \times CMseumP-0.00166 \times RestP-0.00130 \times ManuP$	0.7035	0.6351	0.0253
MAY	$MAY\_SF=0.98524+9.210628E-7 \times Distl-6.25641E-7 \times Interdist+0.00479 \times TranP-0.00124 \times ManuP$	0.5529	0.4890	0.0295
JUN	$JUN\_SF=1.22743+0.00281 \times TF+0.00000168 \times Distl-1.41281 \times ST21+0.92899 \times ST23-0.51406 \times PPA22\_64$	0.6805	0.6213	0.0406
JUL	$JUL\_SF=1.12175+0.00212 \times TF-0.00039189 \times Indexdist2+8.1128E-7 \times Interdist-0.39404 \times PPA18\_64+0.00589 \times SRtlP+0.00147 \times ManuP$	0.7872	0.7381	0.0371
AUG	$AUG\_SF=1.07595+0.00304 \times TF+7.357061E-7 \times Interdist-0.23121 \times PPA18\_64+0.00600 \times SRtlP+0.05256 \times FishP$	0.6967	0.6405	0.0345
SEP	$SEP\_SF=1.00854+0.47485 \times ST21+0.00895 \times CSHP+0.00121 \times SSHP+0.00541 \times SRtlP+0.00826 \times NHotlP+0.03512 \times FishP+0.00336 \times TranP$	0.8530	0.8119	0.0225
OCT	$OCT\_SF=1.02075+0.00080427 \times SSHP+0.00326 \times SRtlP$	0.4093	0.3699	0.0300
NOV	$NOV\_SF=1.06294+0.05240 \times MA+0.00051468 \times Indexdist2-1.20256 \times ST23+0.00101 \times ServP$	0.7330	0.6948	0.0271
DEC	$DEC\_SF=0.98965-0.00000303 \times Distl+0.22647 \times PPA18\_64+0.01098 \times NHotlP+0.01003 \times RcServP$	0.6560	0.6069	0.0383

Table 5.16 Model Variables for Rural SP Group Sorted by Month and Partial R<sup>2</sup> Value

Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month
<i>SRtlP</i>	0.3413	JAN	<i>SMseumP</i>	0.0326	MAR	<i>SRtlP</i>	0.2331	AUG
<i>PPA18_64</i>	0.2069	JAN	<i>Dist1</i>	0.2743	APR	<i>Interdist</i>	0.1622	AUG
<i>SSHP</i>	0.1173	JAN	<i>Interdist</i>	0.1279	APR	<i>TF</i>	0.1278	AUG
<i>WholP</i>	0.0907	JAN	<i>ManuP</i>	0.0769	APR	<i>FishP</i>	0.0889	AUG
<i>Indexdist2</i>	0.0409	JAN	<i>RestP</i>	0.0992	APR	<i>PPA18_64</i>	0.0847	AUG
<i>TranP</i>	0.0313	JAN	<i>CMseumP</i>	0.0733	APR	<i>SSHP</i>	0.3317	SEP
<i>ST23</i>	0.0297	JAN	<i>NRtlP</i>	0.0520	APR	<i>CSHP</i>	0.2332	SEP
<i>EdP</i>	0.0275	JAN	<i>TranP</i>	0.1901	MAY	<i>SRtlP</i>	0.0830	SEP
<i>ManuP</i>	0.0181	JAN	<i>Interdist</i>	0.1924	MAY	<i>NHotlP</i>	0.0809	SEP
<i>SHotlP</i>	0.0221	JAN	<i>ManuP</i>	0.0909	MAY	<i>FishP</i>	0.0571	SEP
<i>CMseumP</i>	0.0208	JAN	<i>Dist1</i>	0.0795	MAY	<i>ST21</i>	0.0362	SEP
<i>SSHP</i>	0.4765	FEB	<i>Dist1</i>	0.2853	JUN	<i>TranP</i>	0.0310	SEP
<i>WholP</i>	0.1594	FEB	<i>PPA22_64</i>	0.1752	JUN	<i>SSHP</i>	0.2954	OCT
<i>PPA18_64</i>	0.1022	FEB	<i>ST21</i>	0.0886	JUN	<i>SRtlP</i>	0.1139	OCT
<i>SRtlP</i>	0.0624	FEB	<i>TF</i>	0.0769	JUN	<i>Indexdist2</i>	0.2770	NOV
<i>NSHP</i>	0.0474	FEB	<i>ST23</i>	0.0544	JUN	<i>ServP</i>	0.1990	NOV
<i>Dist1</i>	0.3867	MAR	<i>ManuP</i>	0.2628	JUL	<i>ST23</i>	0.1645	NOV
<i>NMseumP</i>	0.1638	MAR	<i>SRtlP</i>	0.1595	JUL	<i>MA</i>	0.0924	NOV
<i>SRtlP</i>	0.0781	MAR	<i>PPA18_64</i>	0.1402	JUL	<i>Dist1</i>	0.3387	DEC
<i>NSHP</i>	0.0625	MAR	<i>Indexdist2</i>	0.0935	JUL	<i>NHotlP</i>	0.1597	DEC
<i>PPA22_64</i>	0.0717	MAR	<i>Interdist</i>	0.0800	JUL	<i>RcServP</i>	0.0879	DEC
<i>TF</i>	0.0345	MAR	<i>TF</i>	0.0511	JUL	<i>PPA18_64</i>	0.0698	DEC

Table 5.17 Model Variables for Rural SP Group Sorted by Name and Partial R<sup>2</sup> Value

Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month
<i>CMseumP</i>	0.0733	APR	<i>NMseumP</i>	0.1638	MAR	<i>SSHP</i>	0.3317	SEP
<i>CSHP</i>	0.2332	SEP	<i>NRtlP</i>	0.0520	APR	<i>SSHP</i>	0.2954	OCT
<i>Dist1</i>	0.3867	MAR	<i>NSHP</i>	0.0625	MAR	<i>SSHP</i>	0.1173	JAN
<i>Dist1</i>	0.3387	DEC	<i>PPA18_64</i>	0.2069	JAN	<i>ST21</i>	0.0886	JUN
<i>Dist1</i>	0.2853	JUN	<i>PPA18_64</i>	0.1402	JUL	<i>ST23</i>	0.1645	NOV
<i>Dist1</i>	0.2743	APR	<i>PPA18_64</i>	0.1022	FEB	<i>ST23</i>	0.0544	JUN
<i>Dist1</i>	0.0795	MAY	<i>PPA18_64</i>	0.0847	AUG	<i>TF</i>	0.1278	AUG
<i>FishP</i>	0.0889	AUG	<i>PPA18_64</i>	0.0698	DEC	<i>TF</i>	0.0769	JUN
<i>FishP</i>	0.0571	SEP	<i>PPA22_64</i>	0.1752	JUN	<i>TF</i>	0.0511	JUL
<i>Indexdist2</i>	0.2770	NOV	<i>PPA22_64</i>	0.0717	MAR	<i>TranP</i>	0.1901	MAY
<i>Indexdist2</i>	0.0935	JUL	<i>RcServP</i>	0.0879	DEC	<i>WholP</i>	0.1594	FEB
<i>Interdist</i>	0.1924	MAY	<i>RestP</i>	0.0992	APR	<i>WholP</i>	0.0907	JAN
<i>Interdist</i>	0.1622	AUG	<i>ServP</i>	0.1990	NOV			
<i>Interdist</i>	0.1279	APR	<i>SRtlP</i>	0.3413	JAN			
<i>Interdist</i>	0.0800	JUL	<i>SRtlP</i>	0.2331	AUG			
<i>MA</i>	0.0924	NOV	<i>SRtlP</i>	0.1595	JUL			
<i>ManuP</i>	0.2628	JUL	<i>SRtlP</i>	0.1139	OCT			
<i>ManuP</i>	0.0909	MAY	<i>SRtlP</i>	0.0830	SEP			
<i>ManuP</i>	0.0769	APR	<i>SRtlP</i>	0.0781	MAR			
<i>NHotlP</i>	0.1597	DEC	<i>SRtlP</i>	0.0624	FEB			
<i>NHotlP</i>	0.0809	SEP	<i>SSHP</i>	0.4765	FEB			

Table 5.18 Model Variables for Rural SP Group Sorted by Partial R<sup>2</sup> Value

Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month
<i>SSHP</i>	0.4765	FEB	<i>NMseumP</i>	0.1638	MAR	<i>RcServP</i>	0.0879	DEC
<i>Dist1</i>	0.3867	MAR	<i>Interdist</i>	0.1622	AUG	<i>PPA18 64</i>	0.0847	AUG
<i>SRtlP</i>	0.3413	JAN	<i>NHotlP</i>	0.1597	DEC	<i>SRtlP</i>	0.0830	SEP
<i>Dist1</i>	0.3387	DEC	<i>SRtlP</i>	0.1595	JUL	<i>NHotlP</i>	0.0809	SEP
<i>SSHP</i>	0.3317	SEP	<i>WholP</i>	0.1594	FEB	<i>Interdist</i>	0.0800	JUL
<i>SSHP</i>	0.2954	OCT	<i>PPA18 64</i>	0.1402	JUL	<i>Dist1</i>	0.0795	MAY
<i>Dist1</i>	0.2853	JUN	<i>Interdist</i>	0.1279	APR	<i>SRtlP</i>	0.0781	MAR
<i>Indexdist2</i>	0.2770	NOV	<i>TF</i>	0.1278	AUG	<i>ManuP</i>	0.0769	APR
<i>Dist1</i>	0.2743	APR	<i>SSHP</i>	0.1173	JAN	<i>TF</i>	0.0769	JUN
<i>ManuP</i>	0.2628	JUL	<i>SRtlP</i>	0.1139	OCT	<i>CMseumP</i>	0.0733	APR
<i>CSHP</i>	0.2332	SEP	<i>PPA18 64</i>	0.1022	FEB	<i>PPA22 64</i>	0.0717	MAR
<i>SRtlP</i>	0.2331	AUG	<i>RestP</i>	0.0992	APR	<i>PPA18 64</i>	0.0698	DEC
<i>PPA18 64</i>	0.2069	JAN	<i>Indexdist2</i>	0.0935	JUL	<i>NSHP</i>	0.0625	MAR
<i>ServP</i>	0.1990	NOV	<i>MA</i>	0.0924	NOV	<i>SRtlP</i>	0.0624	FEB
<i>Interdist</i>	0.1924	MAY	<i>ManuP</i>	0.0909	MAY	<i>FishP</i>	0.0571	SEP
<i>TranP</i>	0.1901	MAY	<i>WholP</i>	0.0907	JAN	<i>ST23</i>	0.0544	JUN
<i>PPA22 64</i>	0.1752	JUN	<i>FishP</i>	0.0889	AUG	<i>NRtlP</i>	0.0520	APR
<i>ST23</i>	0.1645	NOV	<i>ST21</i>	0.0886	JUN	<i>TF</i>	0.0511	JUL

### 5.2.3 Double-Peak Models

There are 83 TTMSs in the DP group. The regression models for the 12 MSFs are shown in Table 5.19. For the DP group models, the most significant variables are *Dist1*, *SRtlP*, and *SSHP*. The *SRtlP* variable contributes the largest partial R<sup>2</sup> value to the January, February, August, and September models. The coefficients in the January and February models are negative. Those for August and September are positive. This indicates that retail-related employment in South Florida tends to generate more traffic in January and February, and does the opposite in August and September.

The variable *SSHP* is present in the March model with a negative coefficient and in the May, and October models with a positive coefficient. This indicates that the traffic generated by seasonal households in South Florida is more noticeable in March, whereas these households have less of an impact on traffic in May and October.



The variable *Dist1* appears in eight models: January, February, March, May, June, July, November, and December. The coefficients in the June and July models are positive. They are negative for the other six models. This suggests that, for a count station near an urban area, traffic tends to decrease in June and July, but increase in the other six months.

Table 5.20 lists the variables included in the models above, along with their partial  $R^2$  value and the months for which they are significant. Tables 5.21 and 5.22 list all the variables that have a partial  $R^2$  larger than 0.05 by name and by partial  $R^2$  value, respectively.

Table 5.19 Regression Models for Rural DP Group

Month	Seasonal Factor Equation (Double Peak Group)	$R^2$	Adj. $R^2$	RMSE
JAN	$JAN\_SF = 1.14688 - 0.00000164 \times Dist1 - 0.00075402 \times Beachdist - 0.00487 \times Rt\_High + 0.00233 \times NSHP - 0.00530 \times SRtlP + 0.01445 \times CHotlP - 0.00052938 \times ServP$	0.6095	0.5730	0.0464
FEB	$FEB\_SF = 1.12776 - 0.00000111 \times Dist1 - 0.00310 \times RETIRE - 1.28081 \times ST22 + 0.00275 \times NSHP - 0.00655 \times SRtlP$	0.6298	0.6057	0.0453
MAR	$MAR\_SF = 1.01308 - 0.02003 \times PA - 8.86801E-7 \times Dist1 - 0.00050155 \times Indexdist2 - 0.00376 \times Rt\_High - 0.00265 \times SSHP - 0.00688 \times SHotlP$	0.5847	0.5519	0.0359
APR	$APR\_SF = 0.98908 - 0.01506 \times PA - 0.00304 \times Rt\_High + 0.00043424 \times EdP$	0.2581	0.2299	0.0269
MAY	$MAY\_SF = 0.96930 + 3.997491E-7 \times Dist1 + 0.00213 \times SSHP + 0.00910 \times CHotlP$	0.2830	0.2558	0.0306
JUN	$JUN\_SF = 0.88950 - 0.04795 \times CO + 0.00000154 \times Dist1 + 0.00057774 \times Beachdist + 0.00528 \times RETIRE - 0.00404 \times Rt\_Low - 0.00248 \times NSHP$	0.5278	0.4906	0.0464
JUL	$JUL\_SF = 0.78991 + 0.00000188 \times Dist1 + 0.00108 \times Beachdist + 0.00750 \times RETIRE - 0.00487 \times Rt\_Low + 0.71992 \times PPA18\_21 - 0.00377 \times NSHP - 0.01512 \times CHotlP + 0.00220 \times AgriP + 0.00646 \times RcServP + 0.00251 \times MineP$	0.7122	0.6722	0.0478
AUG	$AUG\_SF = 0.96453 + 8.381719E-7 \times Dist1 + 0.24575 \times PPA65up + 0.00816 \times SRtlP + 0.00456 \times RcServP + 0.00259 \times MineP$	0.5686	0.5406	0.0460
SEP	$SEP\_SF = 0.99057 + 0.00220 \times TF + 0.00394 \times Rt\_High + 0.00511 \times SRtlP + 0.00343 \times RcServP$	0.4717	0.4446	0.0405

Month	Seasonal Factor Equation (Double Peak Group)	R <sup>2</sup>	Adj. R <sup>2</sup>	RMSE
OCT	$OCT\_SF = 1.01007 + 0.01853 \times PA + 0.00234 \times SSHP + 0.00755 \times SHotIP$	0.2690	0.2412	0.0356
NOV	$NOV\_SF = 1.05748 - 0.00000149 \times Dist1 - 0.00066973 \times Beachdist + 0.54413 \times ST23$	0.3476	0.3229	0.0432
DEC	$DEC\_SF = 1.07997 - 0.00000149 \times Dist1 - 0.00071958 \times Beachdist + 0.00294 \times NSHP$	0.3813	0.3578	0.0517

Table 5.20 Model Variables for Rural DP Group Sorted by Month and Partial R<sup>2</sup> value

Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month
<i>SRtlP</i>	0.3306	JAN	<i>Rt_High</i>	0.1602	APR	<i>SRtlP</i>	0.4058	AUG
<i>Dist1</i>	0.0994	JAN	<i>PA</i>	0.0598	APR	<i>MineP</i>	0.0423	AUG
<i>Beachdist</i>	0.0499	JAN	<i>EdP</i>	0.0381	APR	<i>PPA65up</i>	0.0393	AUG
<i>Rt_High</i>	0.0512	JAN	<i>SSHP</i>	0.1688	MAY	<i>Dist1</i>	0.0476	AUG
<i>NSHP</i>	0.0258	JAN	<i>Dist1</i>	0.0664	MAY	<i>RcServP</i>	0.0337	AUG
<i>ServP</i>	0.0240	JAN	<i>CHotIP</i>	0.0479	MAY	<i>SRtlP</i>	0.3540	SEP
<i>CHotIP</i>	0.0286	JAN	<i>Dist1</i>	0.3032	JUN	<i>Rt_High</i>	0.0451	SEP
<i>SRtlP</i>	0.3832	FEB	<i>RETIRE</i>	0.0952	JUN	<i>TF</i>	0.0429	SEP
<i>Dist1</i>	0.0857	FEB	<i>NSHP</i>	0.0309	JUN	<i>RcServP</i>	0.0296	SEP
<i>RETIRE</i>	0.0690	FEB	<i>CO</i>	0.0349	JUN	<i>SSHP</i>	0.1435	OCT
<i>NSHP</i>	0.0520	FEB	<i>Beachdist</i>	0.0315	JUN	<i>SHotIP</i>	0.0763	OCT
<i>ST22</i>	0.0399	FEB	<i>Rt_Low</i>	0.0322	JUN	<i>PA</i>	0.0492	OCT
<i>SSHP</i>	0.3411	MAR	<i>Dist1</i>	0.2465	JUL	<i>Dist1</i>	0.1976	NOV
<i>Dist1</i>	0.1216	MAR	<i>NSHP</i>	0.0691	JUL	<i>Beachdist</i>	0.1137	NOV
<i>Rt_High</i>	0.0409	MAR	<i>Beachdist</i>	0.0860	JUL	<i>ST23</i>	0.0364	NOV
<i>PA</i>	0.0337	MAR	<i>AgriP</i>	0.0478	JUL	<i>Dist1</i>	0.2414	DEC
<i>SHotIP</i>	0.0249	MAR	<i>RcServP</i>	0.0358	JUL	<i>Beachdist</i>	0.0656	DEC
<i>Indexdist2</i>	0.0225	MAR	<i>MineP</i>	0.0291	JUL	<i>NSHP</i>	0.0742	DEC

Table 5.21 Model Variables for Rural DP Group Sorted by Name and Partial R<sup>2</sup> Value

Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month
<i>Beachdist</i>	0.1137	NOV	<i>Dist1</i>	0.0664	MAY	<i>SRtlP</i>	0.3540	SEP
<i>Beachdist</i>	0.0860	JUL	<i>PA</i>	0.0598	APR	<i>SRtlP</i>	0.3306	JAN
<i>Beachdist</i>	0.0656	DEC	<i>NSHP</i>	0.0742	DEC	<i>SSHP</i>	0.3411	MAR
<i>Dist1</i>	0.3032	JUN	<i>NSHP</i>	0.0691	JUL	<i>SSHP</i>	0.1688	MAY
<i>Dist1</i>	0.2465	JUL	<i>NSHP</i>	0.0520	FEB	<i>SSHP</i>	0.1435	OCT
<i>Dist1</i>	0.2414	DEC	<i>RETIRE</i>	0.0952	JUN	<i>Rt_High</i>	0.1602	APR
<i>Dist1</i>	0.1976	NOV	<i>RETIRE</i>	0.0690	FEB	<i>Rt_High</i>	0.0512	JAN
<i>Dist1</i>	0.1216	MAR	<i>SHotIP</i>	0.0763	OCT			
<i>Dist1</i>	0.0994	JAN	<i>SRtlP</i>	0.4058	AUG			
<i>Dist1</i>	0.0857	FEB	<i>SRtlP</i>	0.3832	FEB			

Table 5.22 Model Variables for Rural DP Group Sorted by Partial R<sup>2</sup> Value

Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month	Variable	Partial R <sup>2</sup>	Month
<i>SRtlP</i>	0.4058	AUG	<i>SSHP</i>	0.1688	MAY	<i>SHotlP</i>	0.0763	OCT
<i>SRtlP</i>	0.3832	FEB	<i>Rt_High</i>	0.1602	APR	<i>NSHP</i>	0.0742	DEC
<i>SRtlP</i>	0.3540	SEP	<i>SSHP</i>	0.1435	OCT	<i>NSHP</i>	0.0691	JUL
<i>SSHP</i>	0.3411	MAR	<i>Distl</i>	0.1216	MAR	<i>RETIRE</i>	0.0690	FEB
<i>SRtlP</i>	0.3306	JAN	<i>Beachdist</i>	0.1137	NOV	<i>Distl</i>	0.0664	MAY
<i>Distl</i>	0.3032	JUN	<i>Distl</i>	0.0994	JAN	<i>Beachdist</i>	0.0656	DEC
<i>Distl</i>	0.2465	JUL	<i>RETIRE</i>	0.0952	JUN	<i>PA</i>	0.0598	APR
<i>Distl</i>	0.2414	DEC	<i>Beachdist</i>	0.0860	JUL	<i>NSHP</i>	0.0520	FEB
<i>Distl</i>	0.1976	NOV	<i>Distl</i>	0.0857	FEB	<i>Rt_High</i>	0.0512	JAN

### 5.3 Summary

This chapter described the regression analyses for identifying variables that potentially influence monthly seasonal factors. Since the seasonal effects of the same variables may vary depending on latitude, TTMSs in the urban areas were modeled by sub-regions, i.e., North Florida, Central Florida and South Florida. However, because the land use for the rural areas is more difficult to model, a strategy was developed to separate rural TTMSs from commuting and non-commuting groups based on their daily traffic patterns. A commuting traffic pattern is characterized by double peaks (DP) during a day: one in the morning and the other in the afternoon, while non-commuting traffic has a single peak (SP) during the day. Regression analysis was separately performed for TTMSs in the SP and DP groups to model relationships between SFs and land use explanatory variables.

## **CHAPTER VI**

### **SEASONAL FACTOR ASSIGNMENT**

The models described in Chapter V indicate that there is a relationship between the MSFs and land use variables. Even though the models cannot be used to directly predict the MSFs, they provide likely connections between the seasonal factors and the various variables being modeled. These variables may be used to develop a metric to determine which TTMS(s) may be used for the assignment of seasonal factors to a coverage count site. This metric is based on the similarity between land use and other characteristics. This chapter describes a preliminary investigation to explore a simple and practical assignment method. Results from the application of this method are also described. Section 6.1 explains the methodology used for assignment. Sections 6.2 and 6.3 present assignment results in urban and rural areas, respectively. Evaluation of the assignment results is discussed in Section 6.4.

#### **6.1 Methodology for Measuring Similarity between Two Count Sites**

As mentioned in Chapter I, the current practice in Florida for assigning seasonal factors to coverage counts is to create seasonal factor groups and use the group averages as the seasonal factors. These seasonal groups are then assigned to coverage count sites. Finally, their averaged seasonal factors are applied to convert ADTs to AADTs.

Although the regression model described in the previous chapter is not strong enough to predict SFs based on the independent variables, the causal relation between SFs and influential variables implies that MSFs are similar when a PTMS shares similar

characteristics with a TTMS. The goal of this assignment is to identify one or more best-matched TTMSs for any given short-term count site based on their similarity scores.

The similarity score,  $S$ , is calculated based on a set of variables that are identified in the regression analyses. Measurement of the similarity involves computing the differences between the values of each of the variables for the two count stations. Recall that there are 12 regression equations in each model set, and that these variables may appear repeatedly in different equations and are associated with different partial  $R^2$  values. Hence, the partial  $R$ -squares of a variable from the monthly models is summed and used as weight to be applied to normalized differences. The sum of these weighted differences yields a score that measures weighted normalized differences for two count stations,  $i$  and  $j$ , expressed as follows:

$$S_{ij} = \sum_{k=1}^p \frac{|V_{ki} - V_{kj}| \times SPR_k}{\max\{V_k\}} \quad (6-1)$$

where

$S_{ij}$  = similarity score defined for count stations  $i$  and  $j$  ( $i \neq j$ ),

$p$  = number of influential variables,

$V_{ki}$  = value of the  $k$ th variable in the 12-month models for count station  $i$ ,

$V_{kj}$  = value of the  $k$ th variable in the 12-month models for count station  $j$ ,

$SPR_k$  = sum of partial  $R^2$  for the  $k$ th variable in the appeared months, and

$\max(V_k)$  = maximum value for the variable  $V_k$  among all TTMSs.

Using Equation (6-1), a similarity score can be computed for any pair of count stations. If multiple TTMSs are matched to a given count site, they may be ranked based on their similarity scores; the first best-match is the one with the lowest value of

similarity score, and so on. In this way, SFs may be assigned from a TTMS or TTMSs with the smaller value of similarity score.

In the following sections, the above discussed method is applied to TTMS assignments for both urban and rural areas.

## 6.2 Urban TTMS Assignment within Model Regions

The urban area TTMSs were modeled for three different regions: North, Central, and South Florida. There is one set of models for each of the three regions, and the assignment was conducted for TTMSs within each region. In other words, only TTMSs within the same region were considered as candidates for any given count site. For any TTMS within a region that was tested as a short-count site, all of the TTMSs in the same region were treated as potential candidates. Their similarity scores were then computed. The full and reduced sets of influential variables identified by regression analysis were employed to calculate the similarity scores.

### 6.2.1 Urban TTMS Assignment Based on Full Variable Set

The variables used to calculate the similarity scores for TTMSs are summarized in Table 6.1. These variables are hypothesized to be significant for each of the model regions based on the regression analyses described in Chapter V. The definition of the variables can be found in Section 4.2.

Table 6.1 Variable Sets Used for Assignment for Model Regions

Model Region	Variables
North	<i>EdP, FishP, FR, HotlP, LEG, MA, MseumP, OffP, RcServP, RestP, RETIRE, SHP, ST1, ST22, ST23, Rt Low, SU, TranP, and WholP</i>
Central	<i>AgriP, CO, DISN, EdP, MA, MInc, MineP, MseumP, OffP, PA, RETIRE, RtlP, ServP, SHP, ST22, ST23, Rt High, Rt Low, and TranP</i>
South	<i>AgriP, HotlP, MA, ManuP, MseumP, RcServP, RETIRE, SHP, ST1, ST23, and SU</i>

The first five best matching TTMSs are given. Table 6.2 shows several examples of the assignment result for TTMSs in North Florida. The column “Test Sites” displays the identification number of a TTMS. This TTMS is treated as a short-term count site. The next five columns list the first five best matching sites. The corresponding similarity score for each test site is provided in parentheses.

Table 6.2 Sample Assignment Results for TTMSs in North Florida with Full Variable Set

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
260323	550208 (0.63)	550207 (0.65)	550226 (0.73)	550300 (0.8)	550206 (0.86)
290286	729923 (0.42)	290320 (0.49)	480156 (0.51)	720161 (0.52)	509940 (0.54)
550206	550300 (0.19)	550208 (0.25)	550212 (0.31)	559908 (0.33)	550207 (0.37)

The complete assignment results for each TTMS in urban areas are summarized in Appendix B. Tables B.1 to B.3 are for North Florida, Central Florida, and South Florida, respectively.

### 6.2.2 Urban TTMS Assignment Based on a Reduced Variable Set

One concern regarding the above discussed assignment method is that it involves too many variables. Some of these have relatively low  $R^2$  values. The desirability is to have as few variables as possible. Since variables with low  $R^2$  values have small weighting factors, it may be possible to exclude them entirely from the assignment process. A reduced variable set is tested in this section. For each of the three regions, the partial  $R^2$  values for each of the variables from the 12 regression models are summed. For instance, if a variable appears four times in the models for one region, the sum will contain four partial  $R^2$  values. If this sum of partial  $R^2$  is less than a certain value, the variable is not used for assignment purposes. For North and Central Florida, the variables are reduced from 19 to 9, with a cutoff value of 0.2 for the sum of partial  $R^2$

values. For South Florida, the size of the variable set is reduced to 6 from 11, using a cutoff value of 0.15. Tables B.4, B.5, and B.6 show the assignment results for the three regions, respectively. The reduced variables used to calculate the similarity scores for sites in three model regions are summarized in Table 6.3.

**Table 6.3 Reduced Variable Sets Used for Assignment for Three Model Regions**

District	Variables
North	<i>FR, HotIP, MseumP, RETIRE, SHP, ST22, ST23, Rt Low, and SU</i>
Central	<i>AgriP, CO, MA, MInc, PA, RETIRE, SHP, ST23, and Rt Low</i>
South	<i>HotIP, MA, ManuP, MseumP, RETIRE, and SHP</i>

Compared to the assignment results obtained based on the full set of variables, 40 of 57 sites remained unchanged for North Florida. For Central Florida, 25 of 38 sites remained unchanged. For South Florida, 46 of 56 sites remained unchanged.

### **6.3 Rural TTMS Assignment within Hourly Traffic Pattern Groups**

Since the TTMSs are separated into two hourly traffic pattern groups instead of model region, the assignment for MSFs in the rural areas is conducted within hourly pattern group. Using the same method as the urban areas, every TTMS in the rural areas was first treated as a PTMS, and a similarity score between the test site and all other TTMSs within the same group were computed with both a full and reduced set of influential variables identified by regression analysis. The first five sites with the lowest similarity score were selected as the candidates for the evaluation of the assignment results.

#### **6.3.1 Rural TTMS Assignment Based on Full Variable Set**

The complete variable set that has been identified in the regression analyses and used to calculate similarity scores for the TTMSs in the rural area is summarized in Table



6.4. For the SP group, the complete variable set includes 25 variables, while the DP group is composed of 23 variables. The definition of the variables can be found in Section 4.2. The assignment results for the SP group and DP group are shown in Table B.7 and Table B.8, respectively.

Table 6.4 Variable Sets Used for Assignment for Model Groups

Model Region	Variables
SP Group	<i>WholP, TranP, TF, SSHP, SRtlP, SMseumP, RestP, RcServP, PPA65up, PPA6_21, PPA5_10, PPA22_64, PPA18_64, PPA14_17, PPA11_13, NSHP, NRtlP, NMseumP, NHotlP, ManuP, Interdist, Indexdist2, FishP, EdP, Dist1</i>
DP Group	<i>TF, SSHP, SRtlP, SHotlP, ServP, Rt_Low, Rt_High, RETIRE, RcServP, PPA65up, PPA22_64, PPA14_17, PPA11_13, NSHP, MineP, Indexdist2, FR, EdP, Dist1, CO, CHotlP, Beachdist, AgriP</i>

### 6.3.2 Rural TTMS Assignment Based on a Reduced Variable Set

For the same reason as the urban areas, due to the concern that too many influential variables were involved, a reduced variable set consisting of nine variables with the highest partial  $R^2$  values instead of the original 25 was tested for the SP group, and a reduced set of eight variables was tested for the DP group. The variables that have been included in the reduced variable set are summarized in Table 6.5. The assignment results for the SP group and DP group with the reduced variables set are shown in Tables B.9 and B.10, respectively.

Table 6.5 Variable Sets Used for Assignment for Model Groups

Model Region	Variables
SP Group	<i>TranP, TF, SSHP, SRtlP, PPA18_64, ManuP, Interdist, Indexdist2, Dist1</i>
DP Group	<i>TF, SSHP, SRtlP, Rt_High, RETIRE, NSHP, FR, Dist1, and Beachdist</i>

## 6.4 Summary

In this chapter, a metric was developed based on the assumption that if the variables identified in the regression analysis are the underlying causes of seasonal traffic variations, they may be used to directly link one count station to a TTMS based on the

similarity between the variables' values. The TTMS assignments were conducted within each model region/group for both urban and rural areas, and the best matching sites for each TTMS are displayed in Appendix B. A full set of influential variables identified by the regression models was first employed for the assignment procedure. In order to reduce the complexity of the method, a reduced set of influential variables, selected based on the summation of partial  $R^2$  in regression models, was also tested. The assignment results obtained based on both the variable sets are comparable. For a large portion of the TTMSs, the best matching site remains unchanged.

## CHAPTER VII

### EVALUATION OF THE ASSIGNMENT RESULTS

In this chapter, the accuracy of the assignment method developed in the previous chapter is first evaluated by estimating the 12 MSFs for all the TTMSs. The MSFs of the TTMS are assumed to be a short-count site, and its best matches are compared. Verification of whether the similarity scores are good indicators of matches is carried out by a further discussion on the relationship between the site difference in terms of influential variables and MSF patterns between each pair of TTMS. Finally, the TTMSs locations are also evaluated based on substitutability of the MSFs. The sites with the most distinctive MSFs patterns and the sites with the most common MSFs patterns are identified.

#### 7.1 Evaluation of the Assignment Results

For validation purposes, the assignment was conducted based on TTMSs. Since the MSFs for these TTMSs already known, it is possible to verify whether the similarity scores are good indicators of matches. The effectiveness of the proposed similarity score can be determined by comparing the differences between the MSFs of an assumed short-count and that of matched TTMSs, thus to evaluate whether the MSFs are similar when the related sites share the similar values of influential variables. The percentage differences between a pair of matched sites are computed as follows:

$$e_{ij} = \frac{1}{12} \sum_{m=1}^{12} \left| \frac{MSF_{mi} - MSF_{mj}}{MSF_{mi}} \right| \quad (7-1)$$

where,

$e_{ij}$  = measure of difference between the monthly seasonal factors of count site  $i$  and  $j$  being compared,

$MSF_{mi}$  = monthly seasonal factor for count site  $i$  for month  $m$ , and

$MSF_{mj}$  = monthly seasonal factor for count site  $j$  for month  $m$  (could be the mean of matched sites).

The average differences between MSFs estimated from matching TTMSs and true MSFs for all TTMSs are calculated with both complete and reduced influence variable sets. As an example, the seasonal factors for TTMS 899921 and its first five best matches are listed in Table 7.1, which also lists the percentage differences between the MSFs of each matched pair of sites. The last two rows of the table display the average MSFs of the first two matched sites and the corresponding percentage differences. Coincidentally, the assignment results of this site computed based on full and reduced variable sets are exactly the same. The MSFs for all the TTMS sites are plotted in Figure 7.1. Note that the first and fourth best matches contain seasonal factors closest to those of site 899921.

Table 7.1 Seasonal Factors for Sample Site 899921 and First Five Best Matched Sites

	Site	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Test Site	899921	0.94	0.87	0.88	0.94	1.01	1.07	1.13	1.10	1.12	1.03	0.99	0.99
1 <sup>st</sup>	030094	0.95	0.84	0.87	0.95	1.07	1.13	1.15	1.11	1.15	1.01	0.95	0.92
		1%	-3%	-1%	1%	6%	6%	2%	1%	3%	-2%	-4%	-7%
2 <sup>nd</sup>	860214	1.01	0.96	0.94	0.97	1.02	1.00	1.04	1.00	1.25	0.98	0.96	0.95
		7%	10%	7%	3%	1%	-7%	-8%	-9%	12%	-5%	-3%	-4%
3 <sup>rd</sup>	130333	1.03	0.95	0.94	0.99	1.03	1.07	1.12	1.03	1.03	0.95	0.95	0.94
		10%	9%	7%	5%	2%	0%	-1%	-6%	-8%	-8%	-4%	-5%
4 <sup>th</sup>	030191	0.96	0.87	0.85	0.96	1.05	1.12	1.09	1.10	1.13	1.04	0.94	1.00
		2%	0%	-3%	2%	4%	5%	4%	0%	1%	1%	5%	1%
5 <sup>th</sup>	860176	0.97	0.91	0.93	1.00	1.05	1.05	1.07	1.04	1.07	1.01	0.99	0.94
		3%	5%	6%	6%	4%	-2%	-5%	-5%	-4%	-2%	0%	-5%
(1 <sup>st</sup> and 2 <sup>nd</sup> )/2		0.98	0.90	0.91	0.96	1.05	1.07	1.10	1.06	1.20	1.00	0.96	0.94
		4%	3%	3%	2%	3%	0%	-3%	-4%	7%	-3%	-4%	-6%

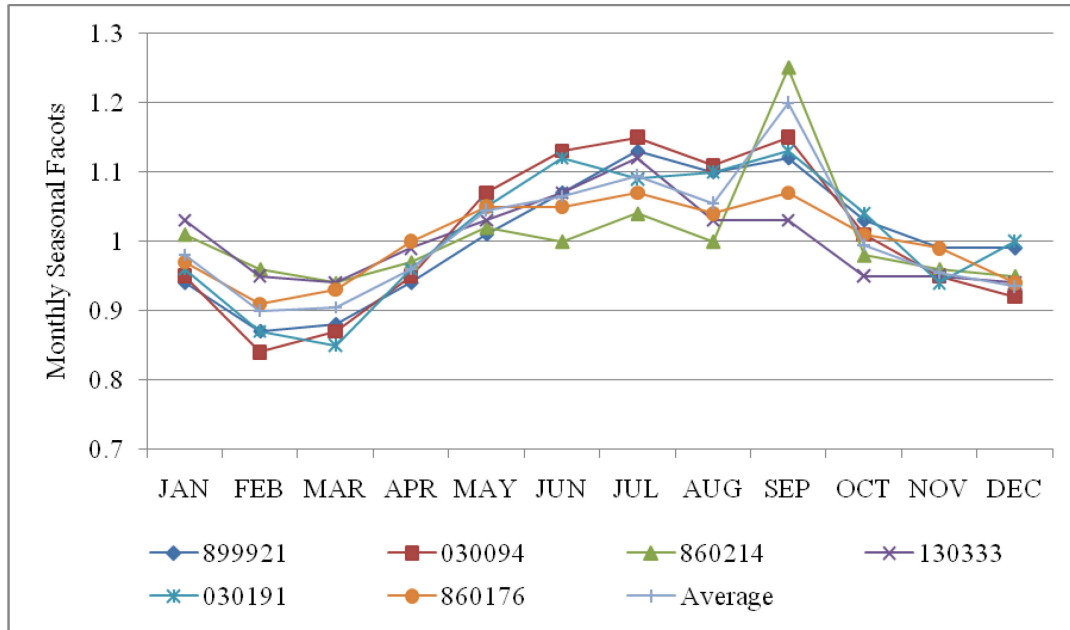


Figure 7.1 MSFs for Test Site 899921 and First Five Closest Matching Sites

Figure 7.2 plots the percentage differences in the MSFs of test site 899921 and its matching sites. The averages of the 12-month percentage differences for each of the matching sites are 3.1%, 6.3%, 5.4%, 2.3%, 4.0%, and 3.6% for the first, second, third, fourth, fifth, and average, respectively. It can be seen that the fourth site is most similar to the test site, and the first site and the average are also rather close.

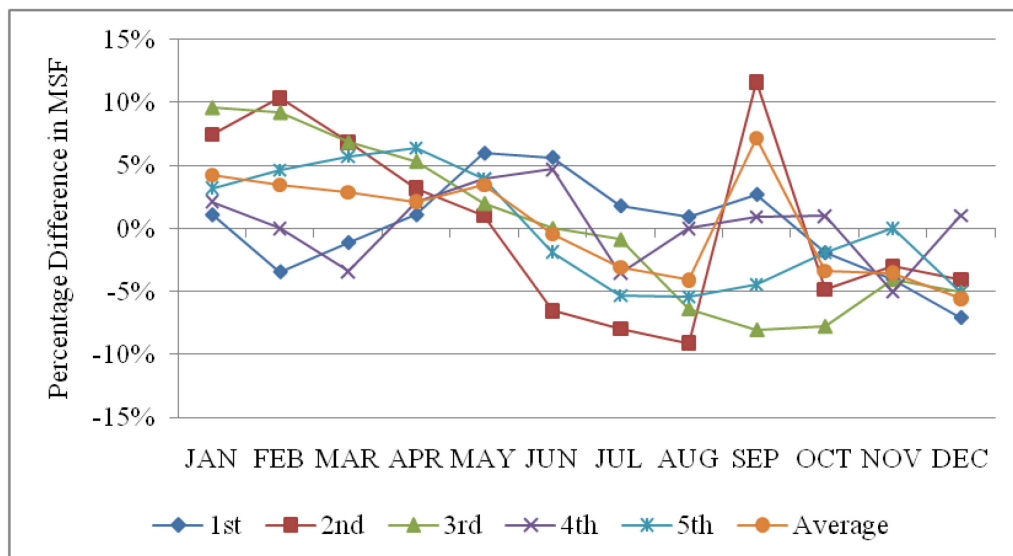


Figure 7.2 Percentage Difference in the MSFs for Site 899921 and the Best Matches

Tables 7.2 and 7.3 summarize the average percentage difference of 12 MSFs between the matching sites and sites of interest for all TTMSs in urban and rural areas, respectively.

Table 7.2 Average Percentage Errors of the Assignment Results for Urban Area.

Best Match	Complete Variable Set			Reduced Variable Set		
	North FL	Central FL	South FL	North FL	Central FL	South FL
1 <sup>st</sup>	4.8%	3.4%	3.7%	4.6%	3.5%	3.9%
2 <sup>nd</sup>	4.8%	3.6%	4.2%	5.3%	3.6%	4.2%
3 <sup>rd</sup>	4.9%	3.7%	4.0%	4.7%	3.8%	4.0%
4 <sup>th</sup>	4.8%	4.0%	3.9%	5.1%	4.3%	4.1%
5 <sup>th</sup>	4.4%	3.7%	3.8%	5.2%	4.0%	3.9%
(1 <sup>st</sup> and 2 <sup>nd</sup> )/2	4.3%	3.1%	3.5%	4.4%	3.1%	3.6%

Table 7.3 Average Errors of the Assignment Results for Rural Area.

Best Match	Complete Variable Set		Reduced Variable Set	
	SP Model	DP Model	SP Model	DP Model
1 <sup>st</sup>	4.4%	4.1%	4.1%	4.6%
2 <sup>nd</sup>	5.7%	4.1%	5.4%	4.1%
3 <sup>rd</sup>	5.9%	4.5%	5.4%	5.0%
4 <sup>th</sup>	5.4%	4.3%	5.6%	4.7%
5 <sup>th</sup>	5.5%	4.9%	6.1%	4.9%
(1 <sup>st</sup> and 2 <sup>nd</sup> )/2	4.4%	3.6%	4.2%	3.7%

Most of the models showed that the differences of MSFs are small when the sites contain the close values of influential variables, which suggests that the identified influential variables can be used for SF assignment with the proposed method. However, after carefully checking the seasonal factor patterns of all TTMSs and their matching sites, it was found that the first two matches often have a seasonal factor pattern that is similar to that of the site of interest. Furthermore, their MSF values are also a close match to those of the site of interest. It is thus better to estimate MSFs for a count site by averaging the corresponding MSFs of the first two (or multiple) best matches. An advantage of using average values is reliance on more TTMSs and avoiding the occasional exceptions. Moreover, the results show that the reduced variable set also

produced assignment results with relatively good accuracy, which was only slightly worse than those of the complete variable set.

#### *7.1.1 Evaluation of the Assignment Results for Urban Areas*

For the urban areas, there were a total of 151 count stations. Figure 7.3 plots the  $e_{ij}$  value for each TTMS for which the MSFs are estimated. It can be seen that the distributions of variances based on the full and reduced variable sets are similar. This indicates that the reduced variable set can be used to replace the full variable set. Table 7.4 summarizes the percentage distribution and accumulated percentage distribution for each interval. In Table 7.4, the first column defines the interval for the  $e_{ij}$  values; the second and third columns respectively show the percentage distribution and the accumulated percentage distribution of the  $e_{ij}$  values resulting from the full variable set. Similarly, the fourth and fifth columns respectively show the percentage distribution and the accumulated percentage distribution of the  $e_{ij}$  values obtained based on the reduced variable set. Figure 7.4 shows the percentage distribution of the  $e_{ij}$ , and Figure 7.5 illustrates the accumulated percentage distribution of the  $e_{ij}$  values. Around 65% of the sites showed that the variances are lower than 0.05 (5%), and 97% of the sites have a variance that is lower than 0.10 (10%). The majority of the  $e_{ij}$  values (about 70%) fall into the range of 0.02-0.06.

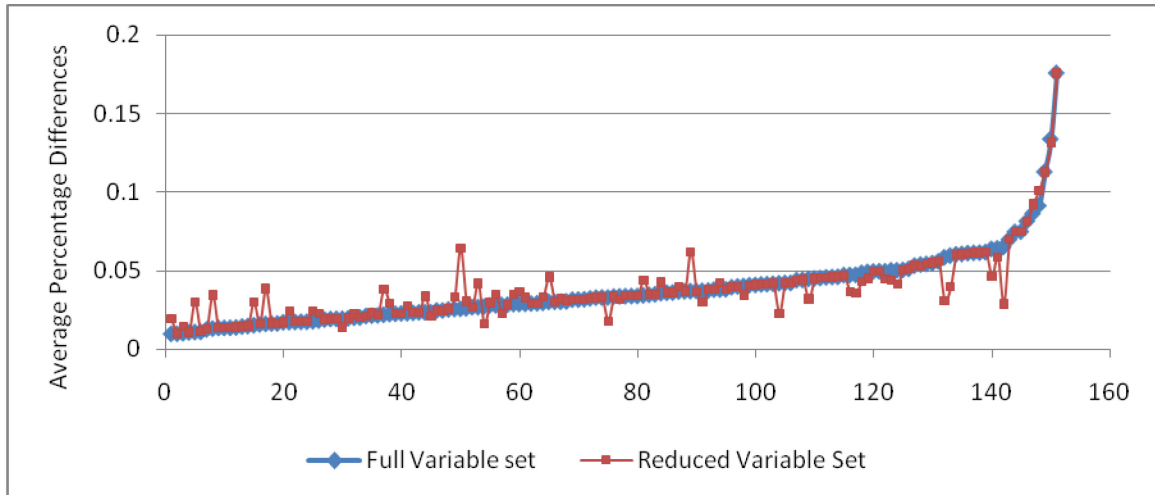


Figure 7.3 12-Month Average Percentage Differences in Urban Areas

Table 7.4 Distribution for Percentage Differences in Urban Areas

$e_{ij}$	Models with Full Variable Set		Models with Reduced Variable Set	
	% Distribution	% Acc_Distribution	% Distribution	% Acc_Distribution
0.00-0.01	0.66	0.66	0.66	0.66
0.01-0.02	19.87	20.53	19.87	20.53
0.02-0.03	21.85	42.38	21.85	42.38
0.03-0.04	21.19	63.58	21.19	63.58
0.04-0.05	17.22	80.79	17.22	80.79
0.05-0.06	7.28	88.08	7.28	88.08
0.06-0.07	6.62	94.70	6.62	94.70
0.07-0.08	1.32	96.03	1.32	96.03
0.08-0.09	1.32	97.35	1.32	97.35
0.09-0.10	0.66	98.01	0.66	98.01
0.10-0.11	0.00	98.01	0.00	98.01
0.11-0.12	0.66	98.68	0.66	98.68
0.12-0.13	0.00	98.68	0.00	98.68
0.13-0.14	0.66	99.34	0.66	99.34
0.14-0.15	0.00	99.34	0.00	99.34
0.15-	0.66	100.00	0.66	100.00



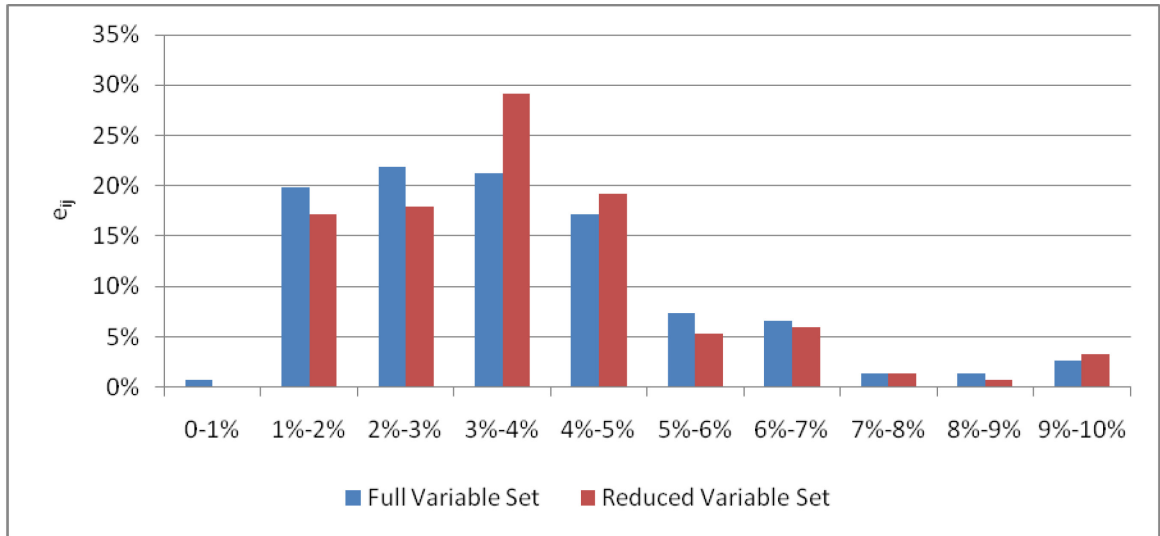


Figure 7.4 Distributions of Variances for TTMSs in Urban Areas

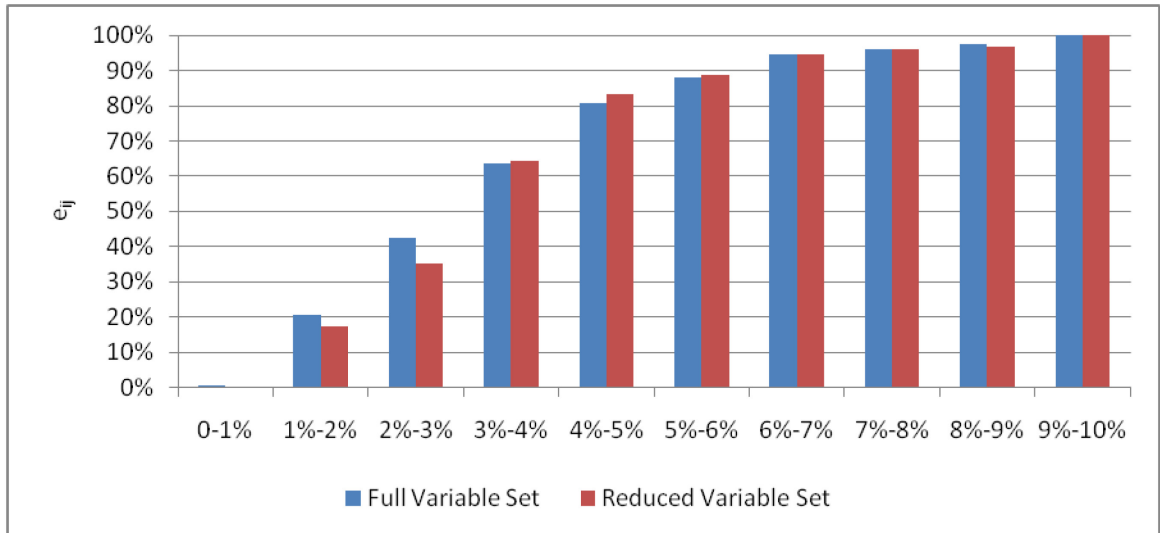


Figure 7.5 Accumulated Percentage Distributions of Variances Urban Area TTMSs

### 7.1.2 Evaluation of the Assignment Results for Rural Area

For rural areas, there are a total of 116 count stations. Figure 7.6 plots the  $e_{ij}$  values for each TTMS. Table 7.5 provides the percentage distribution and accumulated percentage distribution for each interval. Figure 7.7 illustrates the percentage distribution of the  $e_{ij}$  values for rural areas, and Figure 7.8 describes the accumulated percentage distribution of the  $e_{ij}$  values for rural areas. It can be seen that, similar to the results of

the urban areas, the accumulated percentages for results obtained based on both full and reduced variable sets are close. However, for the results based on the full variable set, a large percentage (over 30%) of the variances are in the range of 0.03-0.04, while the results based on the reduced variable set fall mainly in the intervals of 0.02-0.03 (15.5%) and 0.04-0.05 (20%). The overall assignment results for the rural areas are worse compared to the results for the urban areas. Around 60% of the sites have an  $e_{ij}$  value that is lower than 0.05 (5%).

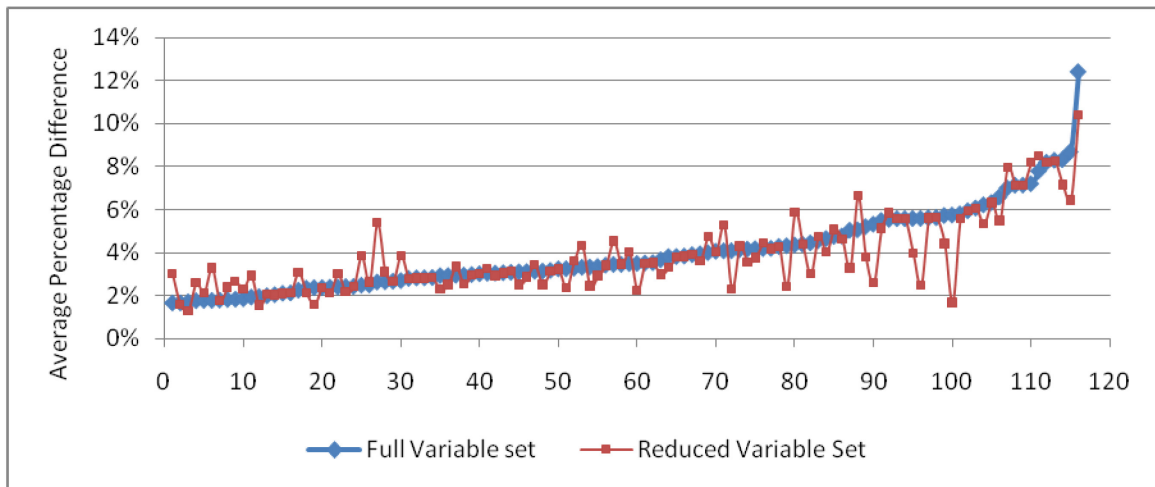


Figure 7.6 12-Month Average Variances Plot for Rural Areas

Table 7.5 Distribution of Percentage Differences for TTMSs in Rural Areas

$e_{ij}$	Models with Full Variable Set		Models with Reduced Variable Set	
	% Distribution	% Acc. Distribution	% Distribution	% Acc. Distribution
0.00-0.01	0.00	0.00	0.00	0.00
0.01-0.02	11.21	11.21	5.17	5.17
0.02-0.03	22.41	33.62	32.76	37.93
0.03-0.04	25.86	59.48	26.72	64.66
0.04-0.05	14.66	74.14	12.07	76.72
0.05-0.06	13.79	87.93	12.07	88.79
0.06-0.07	3.45	91.38	3.45	92.24
0.07-0.08	4.31	95.69	3.45	95.69
0.08-0.09	3.45	99.14	3.45	99.14
0.09-0.10	0.00	99.14	0.00	99.14
0.10-0.11	0.00	99.14	0.86	100.00
0.11-0.12	0.00	99.14	0.00	100.00
0.12-0.13	0.86	100.00	0.00	100.00

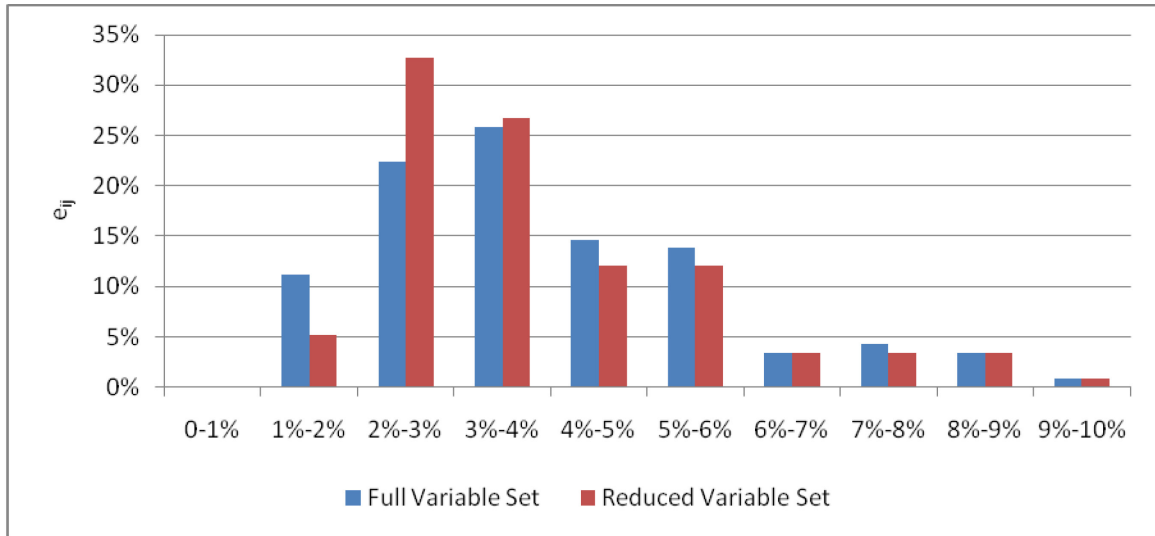


Figure 7.7 Distributions of Variances for TTMSs in Rural Areas

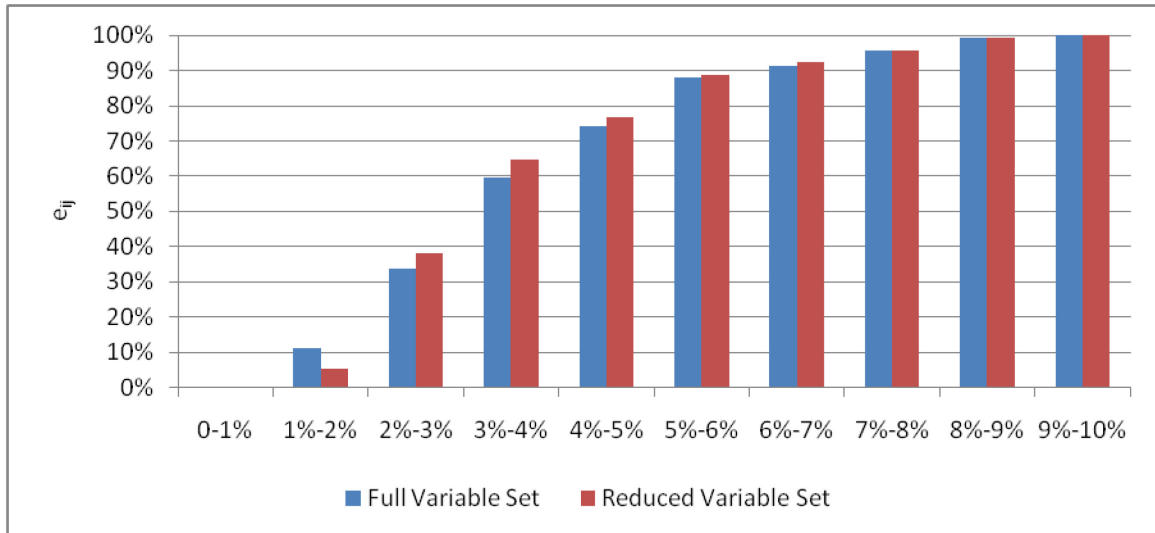


Figure 7.8 Accumulated Percentage Distributions of Variances Rural Area TTMSs

## 7.2 Evaluation of the Assignment Methods

The MSFs from the first best and few matching sites show good assignment results. Thus, it can be considered that when land use and social economic characteristics of the surrounding areas of the sites are similar, the MSFs patterns are also close. However, whether the similarity score between site pairs is correlated with the MSFs, the difference between site pairs cannot be answered by solely testing the first few matches.

In order to further understand the relationship, a similarity score is calculated between each site pair within the same model region or group by using Equation (6-1). An average value  $\bar{S}_i$  is then computed for each site based on Equation (7-2). This average value represents the average difference between site  $i$  and all other sites in terms of the similarity score.

$$\bar{S}_i = \frac{1}{q-1} \sum_{j=1}^{q-1} S_{ij} \quad (7-2)$$

where,

$\bar{S}_i$  = average similarity score for count stations  $i$ ,

$S_{ij}$  = similarity score defined for count stations  $i$  and  $j$  ( $i \neq j$ ), refer to Equation (6-1),

$q$  = number of sites for a certain model region or group.

Similarly, the average MSFs' percentage difference was also computed for each pair of sites within the same model region by Equation (7-1). An average percentage difference  $\bar{e}_i$  is then computed for each site by Equation (7-3).

$$\bar{e}_i = \frac{1}{q-1} \sum_{j=1}^{q-1} e_{ij} \quad (7-3)$$

where,

$\bar{e}_i$  = average percentage differences for count station  $i$ ,

$e_{ij}$  = measure of difference between the monthly seasonal factors of count site  $i$  and  $j$  being compared ( $i \neq j$ ), refer to Equation (7-1),

$q$  = number of sites for a certain model region or group.

The trends of the two average values above are compared for each model region. To adjust the values measured in different scales to comparable range, the maximum  $\bar{S}_i$  and  $\bar{e}_i$  for each sub-model are used for normalization. The average similarity score and MSFs percentage difference for each site for North Florida, Central Florida and South Florida models are plotted in Figures 7.9 to 7.11, respectively. The blue line represents the  $\bar{S}_i$  for each site, and the red line represents the corresponding  $\bar{e}_i$  value. As can be seen in the figures for the urban models, the trends for these two average values are relatively consistent. The peak value of MSFs difference always occurs at the sites where the variation of land use characteristics is obvious. This also proved that the main causes for the seasonal traffic variation can be explained by the influential variables identified by the regression models for urban areas. Distinctive MSF patterns tend to occur at locations with special land use or socioeconomic characteristics, such as an area with high density of seasonal households.

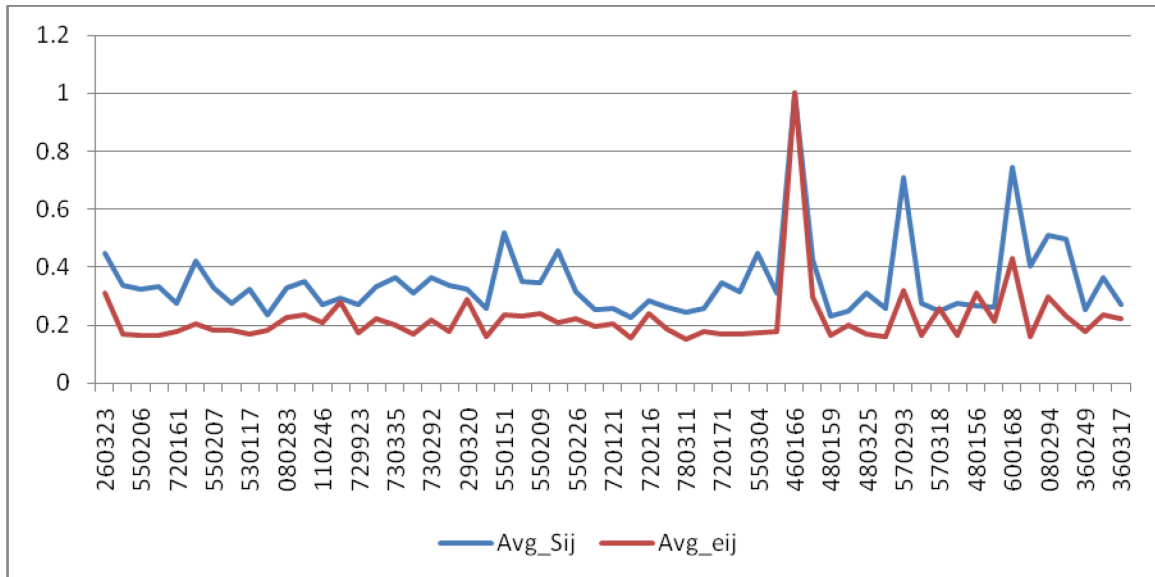


Figure 7.9  $\bar{S}_i$  and  $\bar{e}_i$  for Sites in North Florida

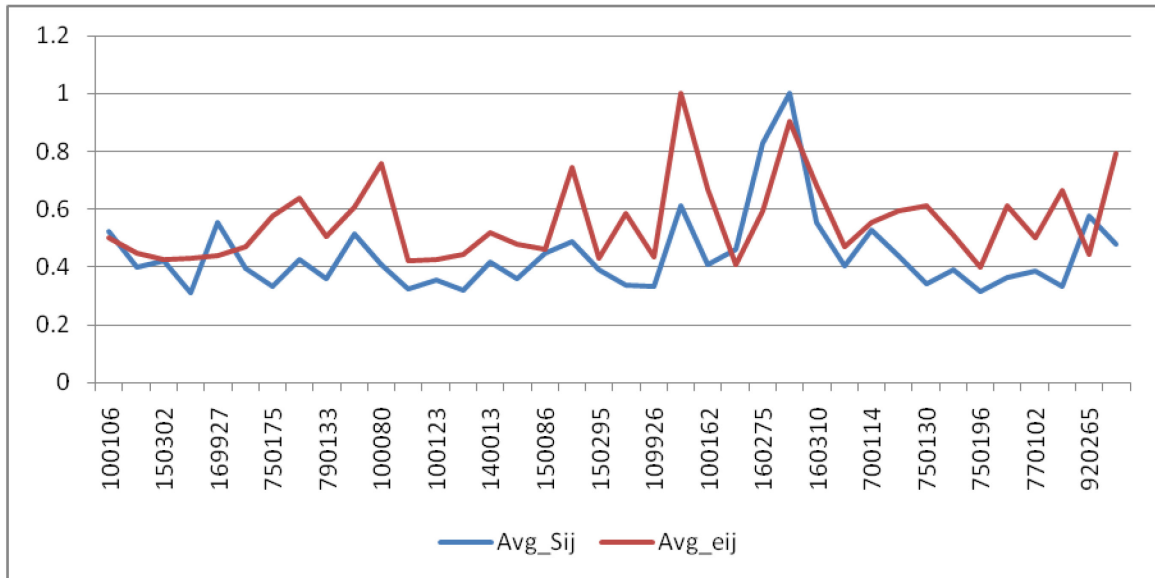


Figure 7.10  $\bar{S}_i$  and  $\bar{e}_i$  for Sites in Central Florida

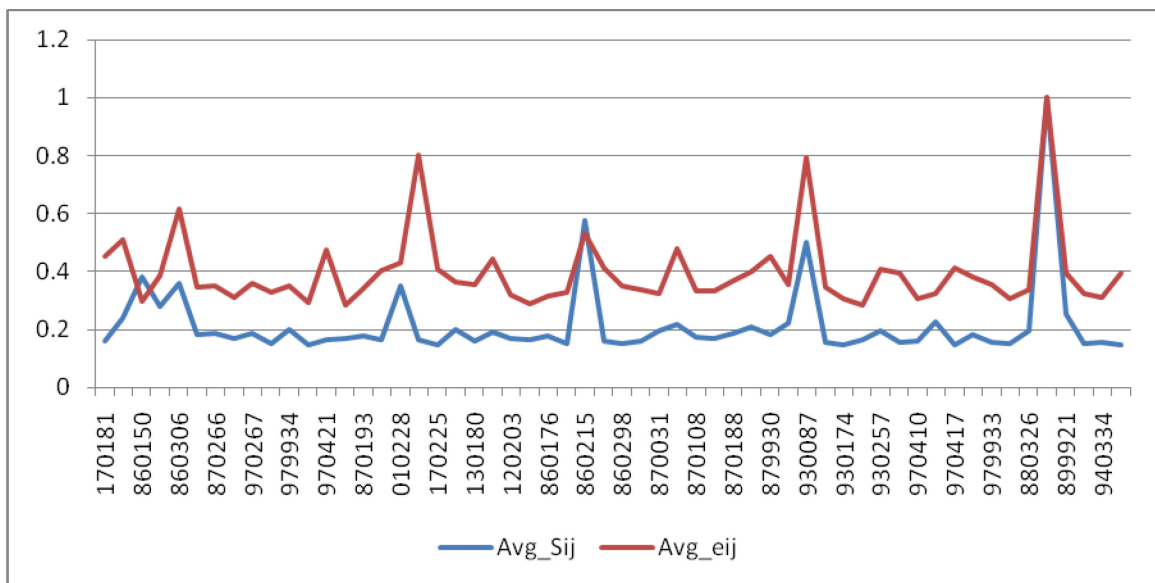


Figure 7.11  $\bar{S}_i$  and  $\bar{e}_i$  for Sites in South Florida

Figures 7.12 and 7.13 show the comparison between  $\bar{S}_i$  and  $\bar{e}_i$  for the rural models. The double-peak model includes the largest number of TTMSs, and the trends are not consistent for several sites. This indicates that the current influential variables do not fully reflect the causes for traffic seasonality of double-peak TTMSs. Additional variables need to be examined for potential inclusion to better explain the seasonal

variation for commuting traffic-dominated roadway sections in rural areas. However, the chart for the single-peak group model shows a strong correlation between  $\bar{S}_i$  and  $\bar{e}_i$ .

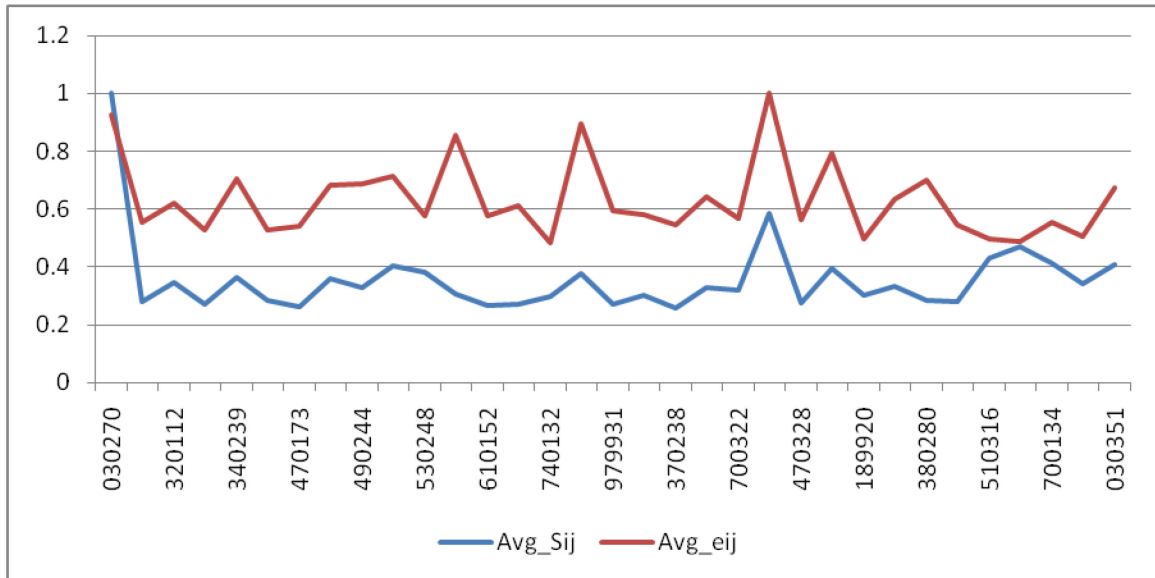


Figure 7.13  $\bar{S}_i$  and  $\bar{e}_i$  for Sites in Single Peak Model

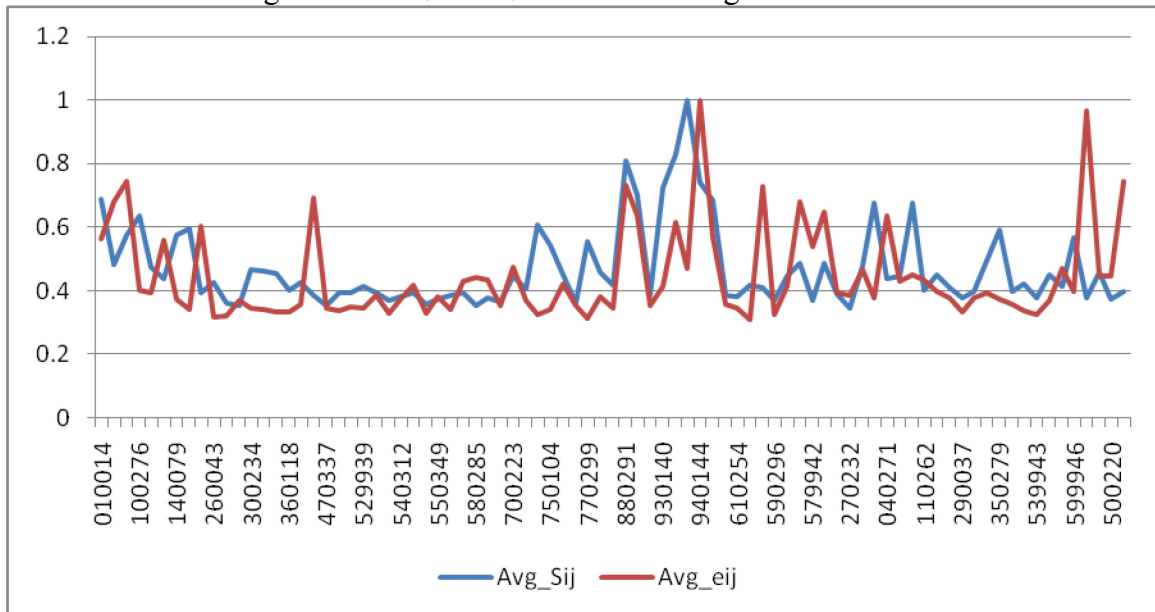


Figure 7.12  $\bar{S}_i$  and  $\bar{e}_i$  for Sites in Double Peak Model

### 7.3 Evaluation of the Current TTMSs

As aforementioned, due to the high operating and maintenance costs, there are only about 300 TTMSs statewide. The MSF patterns for these existing TTMSs were examined to determine which sites carried the most special seasonality information. The redundant TTMSs can also be identified if many in the relative spatial proximity are found to be similar to nearby TTMSs in both their land use and roadway characteristics, as well as in their seasonal factors. The distribution for current TTMSs was also investigated to determine whether they provide adequate coverage by representing a wide range of commonly encountered combinations of land uses and roadway types. Areas are identified according to the need for additional TTMSs because of their unique land use and roadway functions.

#### 7.3.1 Evaluation of the MSFs Pattern for Current TTMSs

Recall that  $e_{ij}$  is the measure of difference between the monthly seasonal factors of count sites  $i$  and  $j$ , which can be considered an indicator for whether two sites share similar MSFs. The distribution of  $e_{ij}$  for each site is plotted by a sub-model. As shown in Figures 7.14 to 7.18, each column represents one count station, and different colors depict the intervals of the  $e_{ij}$  value between the given site and all other sites within the same model. The dark blue color in the bottom represents the percentage of sites with an  $e_{ij}$  value lower than 0.05, the red color represents the percentage of sites with an  $e_{ij}$  value between 0.05-0.10, and the green, purple, and the light blue colors represent the intervals for 0.10-0.15, 0.15-0.20, and larger than 0.20, respectively. As shown in Figures 7.14 to



7.16, a large percentage of the sites in the urban areas share similar MSFs. The average percentage MSF difference is lower than 0.10 for most of the site pairs in Central Florida.

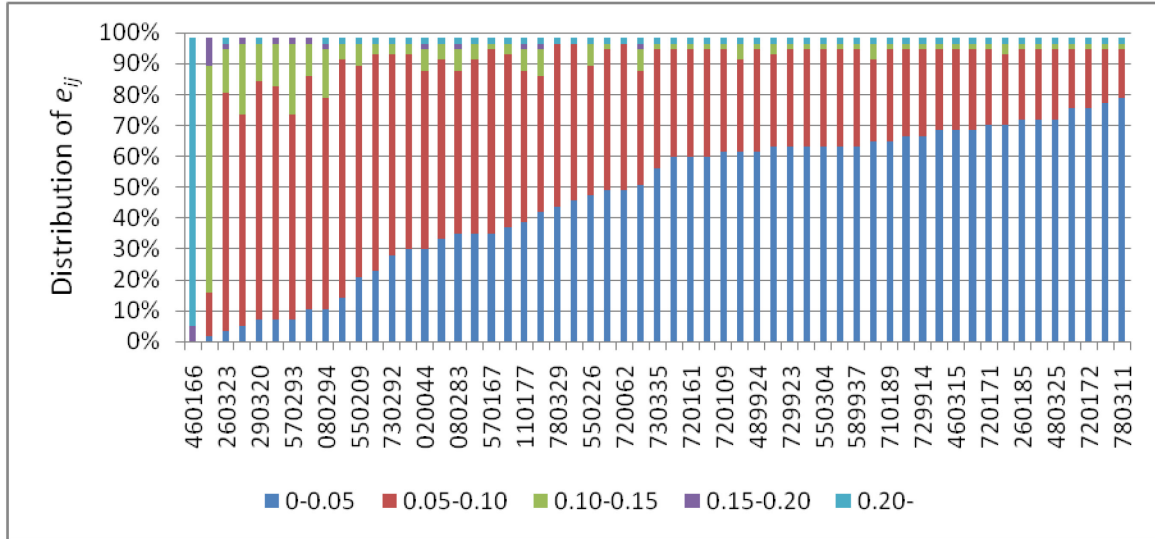


Figure 7.14 Distribution of  $e_{ij}$  for Sites in North Florida

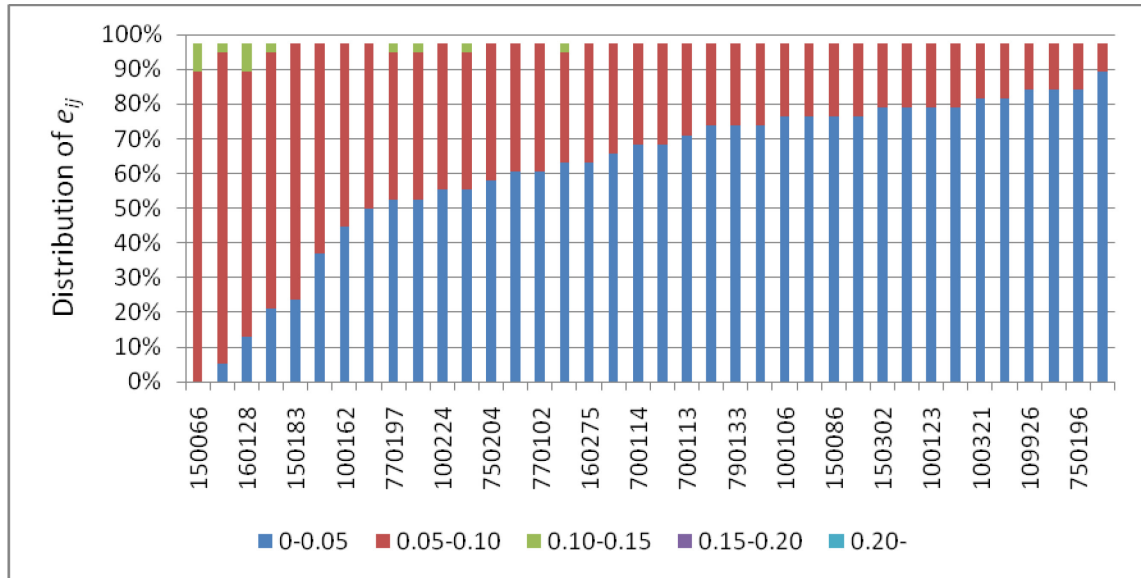


Figure 7.15 Distribution of  $e_{ij}$  for Sites in Central Florida

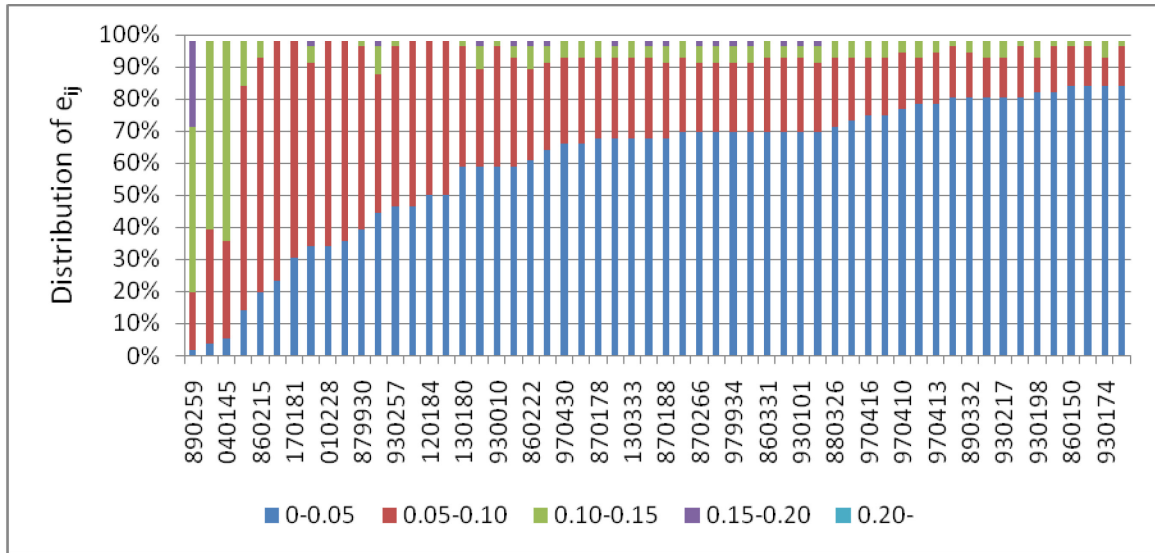


Figure 7.16 Distribution of  $e_{ij}$  for Sites in South Florida

However, the variation of MSFs is more apparent for rural areas. As shown in Figures 7.17 and 7.18, in several sites, the  $e_{ij}$  value is larger than 0.05 in all other sites. This also means that these sites have distinctive MSF patterns and carry more information in terms of MSFs than other sites.

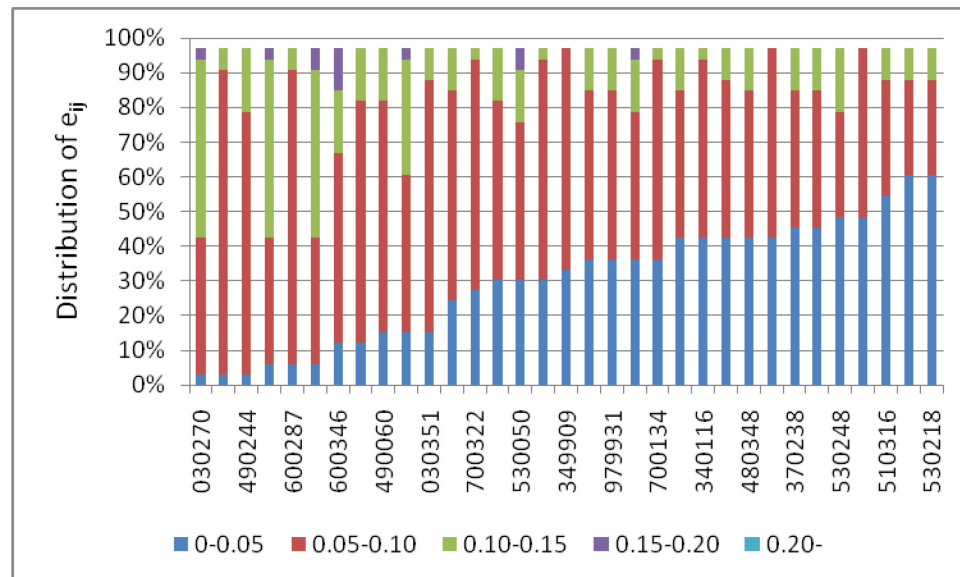


Figure 7.17 Distribution of  $e_{ij}$  for Sites in Single-Peak Group

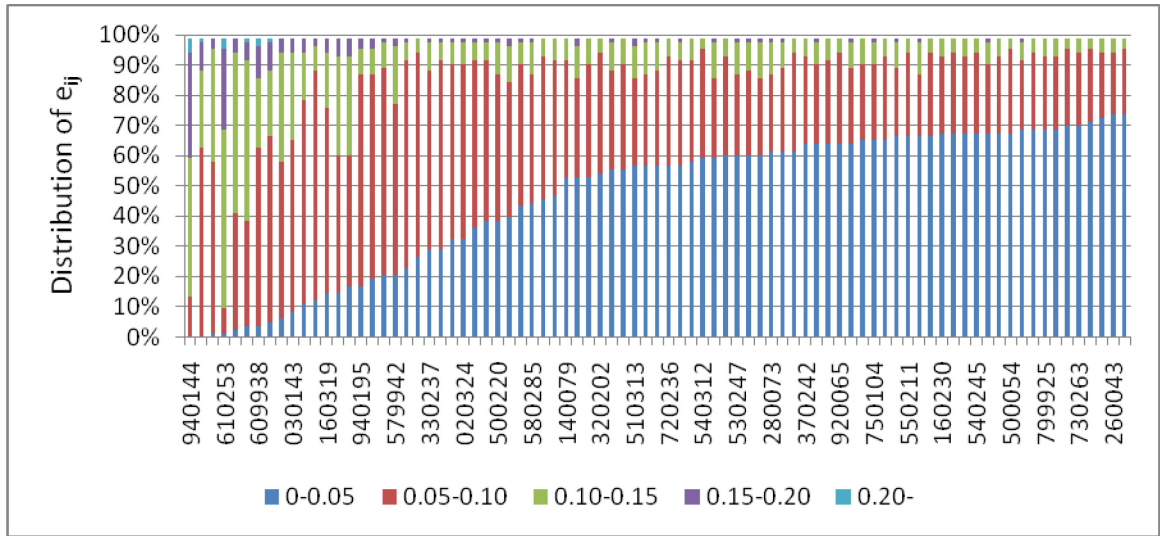


Figure 7.18 Distribution of  $e_{ij}$  for Sites in Double-Peak Group

Figure 7.19 shows the  $\bar{e}_i$  value for all of the TTMSs. TTMSs in different models are distinguished by different symbols, and the size of the symbol indicates the magnitude of the value. The sites that have distinctive MSF patterns are emphasized with a larger size of symbols.

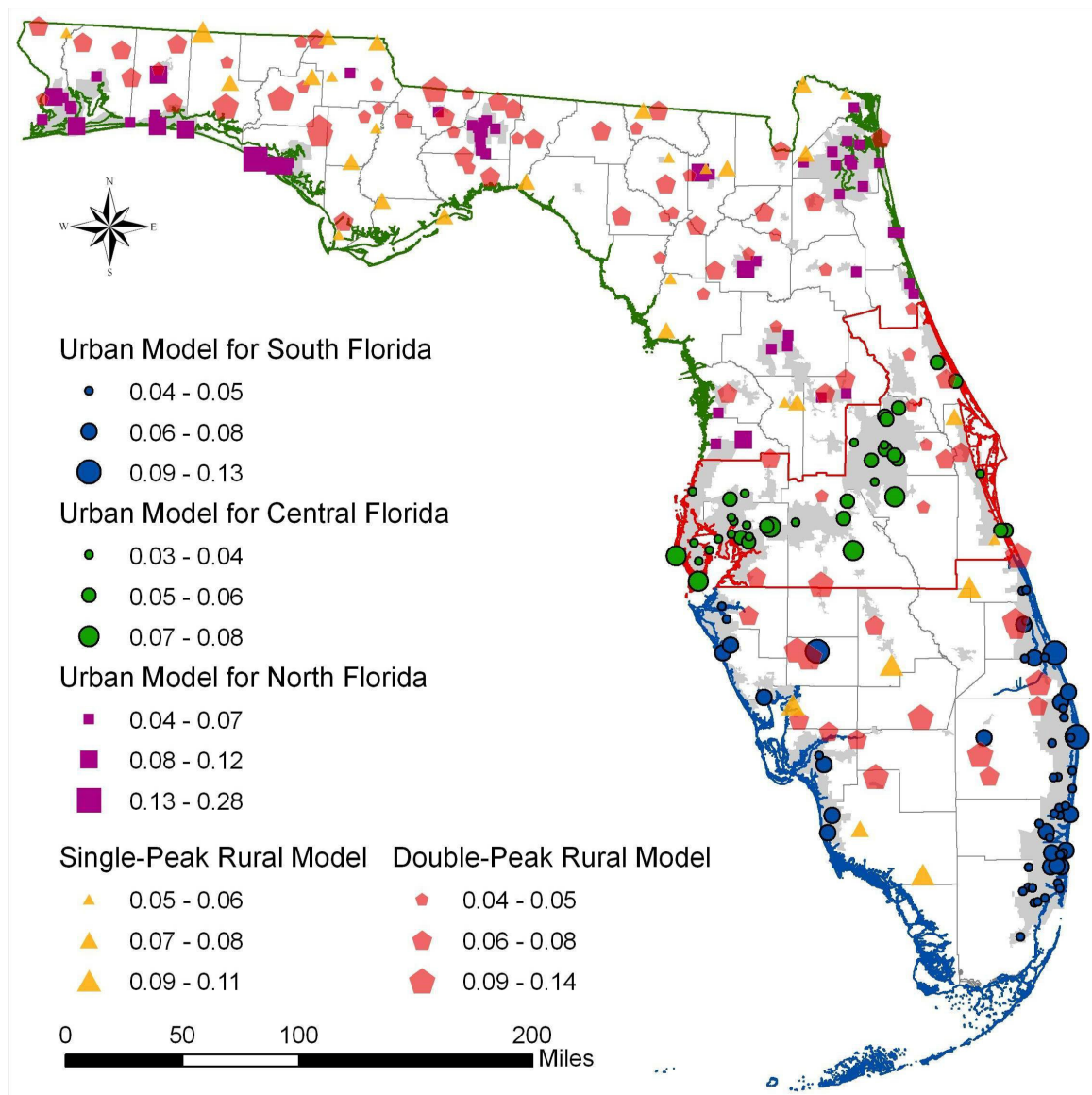


Figure 7.19  $\bar{e}_i$  Value for all TTMSs

### 7.3.2 Evaluation of the Coverage for Current TTMSs in Urban Area

Due to the limited availability of existing TTMSs, there are large areas without any TTMS that is remotely close. Based on the buffer area/service area defined in Chapter III, Figure 7.20 shows the census block groups that are neither completely nor partially covered by the buffer area of any TTMS. Under such conditions, the existing

TTMSs may not adequately reflect the entire range of variation in demographics, socioeconomics, land use, and roadway characteristics of all locations.

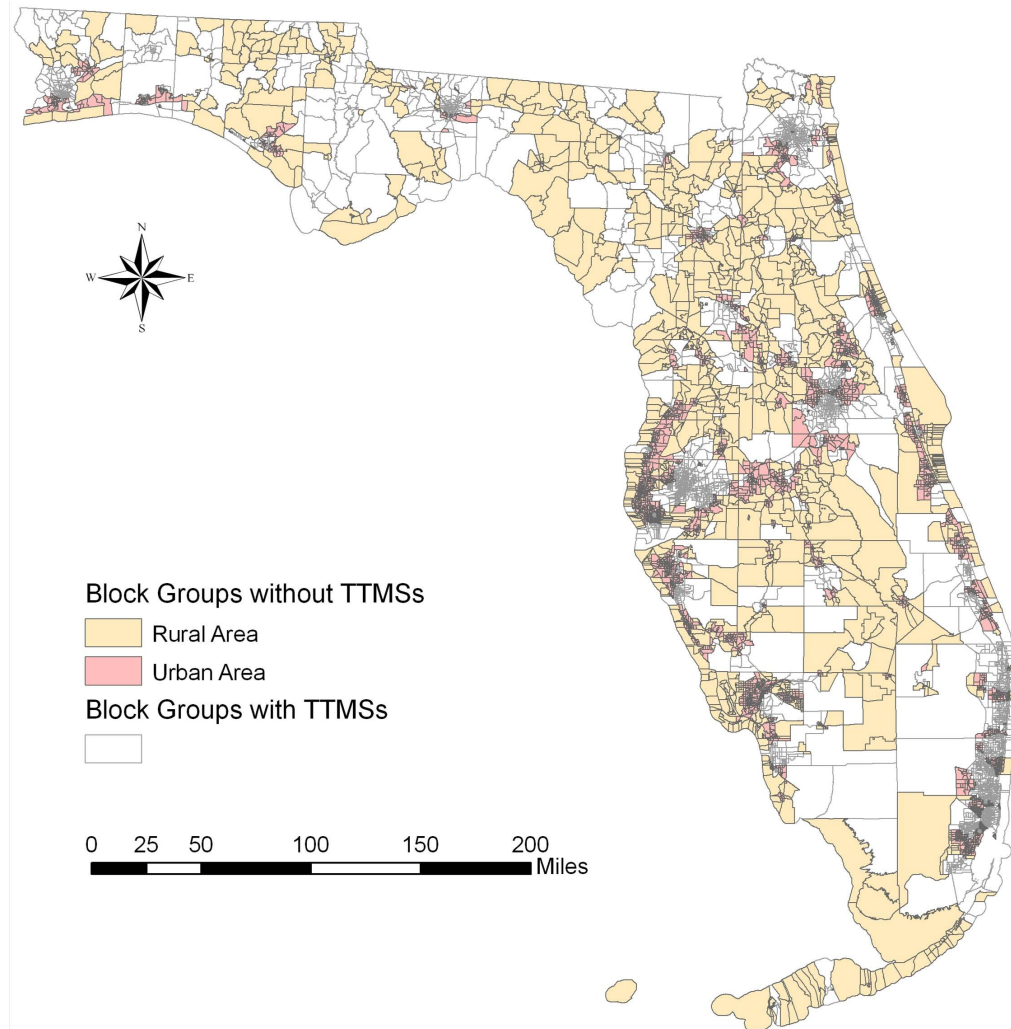


Figure 7.20 Census Block Groups Not Covered by any TTMS Buffer

The reduced set of variables for the urban areas was examined in order to determine the areas that contain influential variables that exceed the range of that for current TTMSs. Nine quantified variables were selected based on the summation of their partial  $R^2$  for each model, as well as the frequency of the variable that appears in different models. The variables are *SHP*, *RETIRE*, *MseumP*, *AgriP*, *HotlP*, *ManuP*, *Rt\_Low*, *ST22*, and *ST23*.

The demographics and socioeconomic variables above were computed for each census block groups in the urban areas, such that the range of each variable is compared with that of the TTMSs buffer area. In the census block groups that are not overlapped with the TTMSs buffer area, there are more than 900 that have at least one variable exceeding the range defined by the existing TTMSs. In other words, the characteristics of such areas cannot be statistically represented by the current TTMSs. Since such block groups are scattered, Figures 7.21 to 7.24 show the location of the census block groups that have special land use and demographic characteristics by FDOT districts. These areas can be considered a higher priority when new TTMSs are constructed.

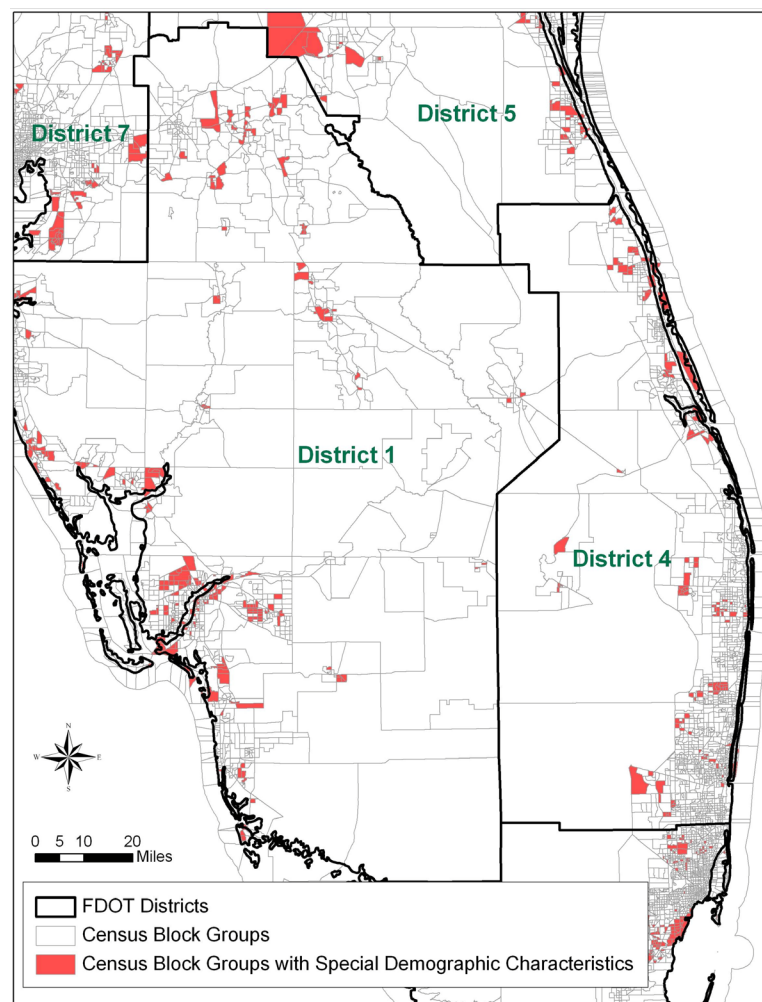


Figure 7.21 Census Block Groups with Special Demographic Characteristics in D1 & D4



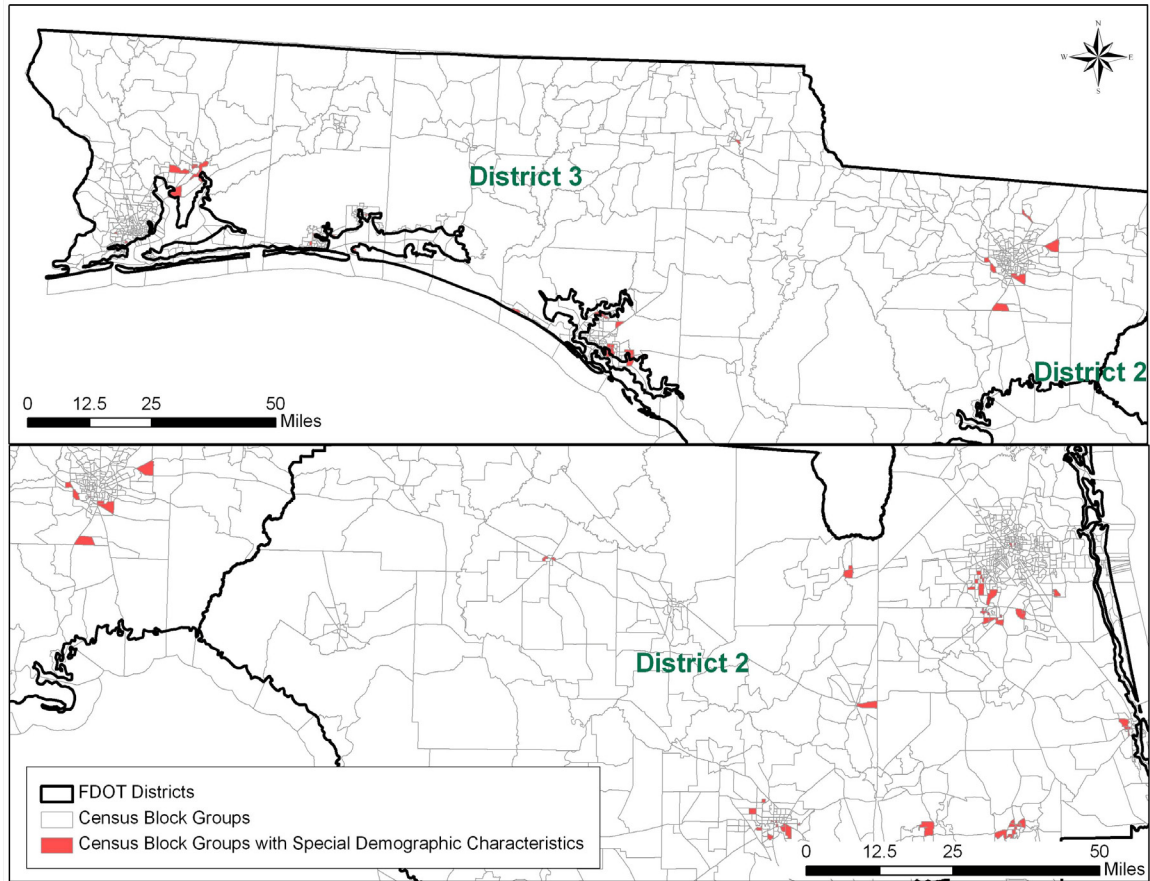


Figure 7.22 Census Block Groups with Special Demographic Characteristics in D2 & D3

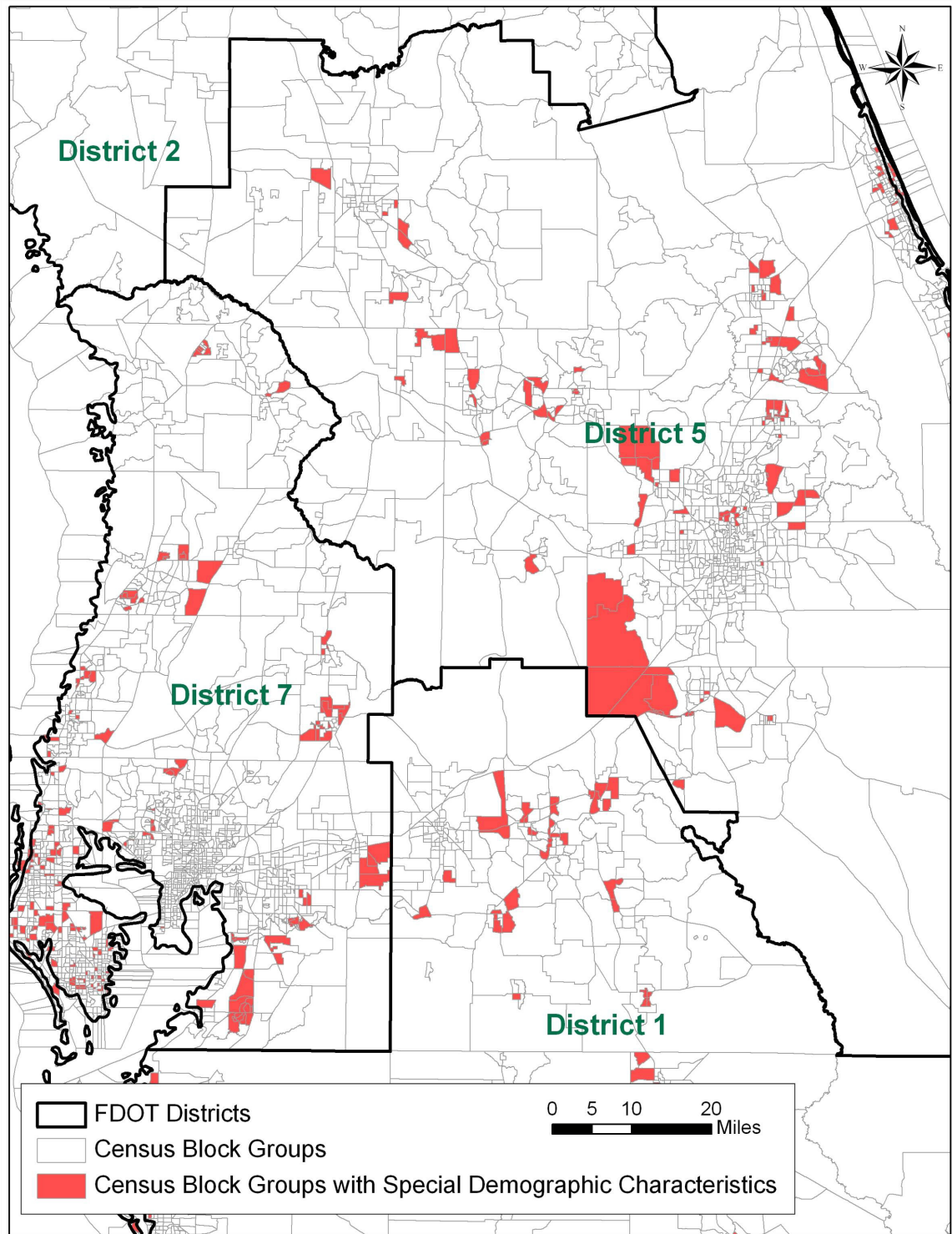


Figure 7.23 Census Block Groups with Special Demographic Characteristics in D5 & D7



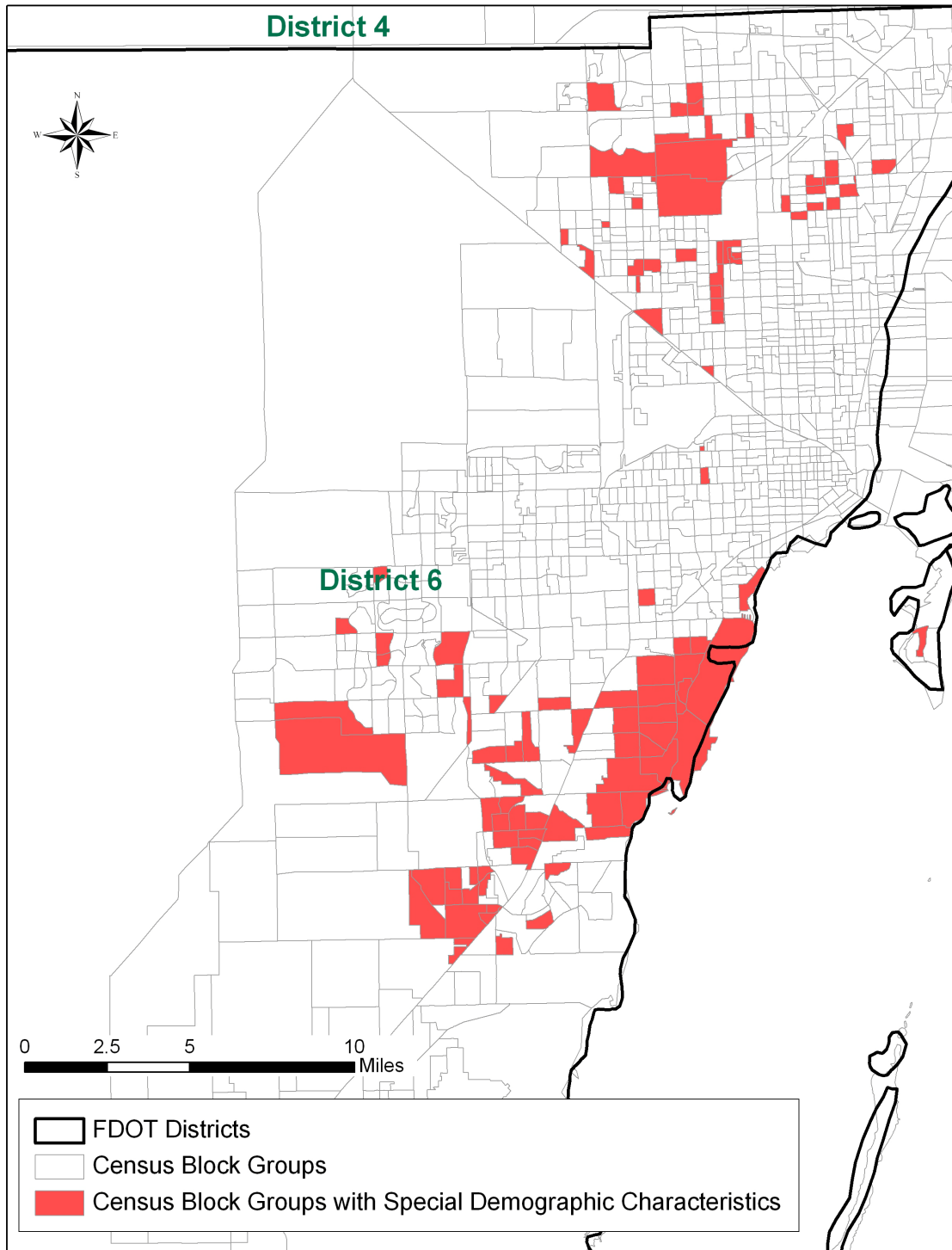


Figure 7.24 Census Block Groups with Special Demographic Characteristics in D6

## 7.4 Summary

In this chapter, the accuracy of the assignment results from Chapter VI was evaluated based on the average differences between 12 MSFs estimated from matching TTMSs and true MSFs. Most of the models showed that the differences among the MSFs are small when the sites contain similar values of influential variables. The averages of the corresponding MSFs of the first two (or more) best matches are recommended as the estimates. The assignment results obtained based on both the full and reduced variable sets are similar.

The similarity of influential variables was examined for all the TTMS pairs within same sub-model group, as well as the similarity of MSFs. Consistent patterns can be observed for the urban areas. This indicates that the factors underlying the seasonal traffic patterns were well explained by influential variables identified for the urban areas. However, the current variables cannot fully reflect the cause of traffic seasonality for the rural areas.

Finally, the TTMS locations were also evaluated based on the substitutability of the MSFs. The sites with the most distinctive MSF patterns and the sites with the most common MSFs patterns were identified.

## **CHAPTER VIII**

### **SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS**

Seasonal factors are a complex subject. While there have been many studies that applied various methods to determine seasonal groups, determining the underlying causes of season variations in traffic and developing models to predict seasonal factors has proven to be a significant challenge. This dissertation research aims to improve the accuracy of the existing seasonal factor estimation methods using a more objective and data-driven method.

#### **8.1 Summary and Conclusions**

An extensive literature search and review was first performed to investigate and assess the state-of-the-art techniques and theories regarding TTMS grouping and PTMS assignment to a seasonal group. The main review tasks included: the current practice in Florida of estimating the SF at a given short-count station; and modeling techniques relevant to the factoring process, including grouping and assignment procedures.

There are 300 TTMSs that are continuously in service throughout the state, 24.2% of them are missing data from the year 2000. Among the 69 sites that have missing data, imputation technique was adopted to replace the missing MSFs for 54 sites with substitute values. The advantage of the imputation method is that it is easily understood and implemented. The disadvantage of this method is that it is time-consuming because manual adjustments have to be made for each site. This process, however, may be made more efficient through automation.

Regression models were developed to identify influential variables for the seasonal factors of TTMSs. It was found that influential variables for the seasonal factors of the TTMSs are different in urban and rural areas, and in different climate zones. To account for differences in climate, the urban areas are divided into three regions: North, Central, and South Florida.

The approach for modeling urban area TTMSs cannot be applied to rural TTMSs as there were insufficient number of TTMSs in the rural areas. An alternative approach was developed to separately model those TTMSs for which the hourly traffic pattern on a typical weekday shows a single peak (i.e., when recreational travel dominates) and those for which the weekday hourly traffic pattern has a double peak (i.e., when commute travel dominates). The models were improved as a result, particularly for the single-peak TTMSs, which were more difficult to model due to the low land use intensity and through traffic. Hence, this improvement may be attributed to the intrinsic connection between daily traffic patterns and land use variables. These, in turn, are connected to the seasonal traffic variations. The more noticeable improvement in the models for the single-peak group may be because recreational roadways (single-peak pattern) have more seasonal variations, while the traffic on commuting roadways (double-peak pattern) varies less seasonally.

The influential variables selected by regression analysis were used in developing a method to identify a TTMS that is similar to a given count station in terms of its land uses or functions. A similarity score was developed to measure the similarity between two count sites. This score was based on the identified influential variables, which were weighted by their partial  $R^2$  values in the models. This approach showed promising

results: the average of standard errors among estimated seasonal factors was about 5% overall.

The relationship between the influential variables and the MSFs were examined for all TTMSs. Better correlations were observed for urban area models than that for rural areas. Because of the strength of explanation for influential variables for urban models, the ranges of the important variables were computed to define the coverage of application for assignment method for urban areas. The areas with influential variables fall outside the ranges were located and considered as the areas with higher priority for new TTMSs.

The regression variables used in this research were carefully selected to ensure that they capture rich information that may affect MSFs while being readily available from either census or transportation planning data. Although land use and demographics may evolve over time, this evolution is usually slow, and frequent model updates may not be needed. Therefore, model updates at intervals of five to ten years may be adequate. The models, however, need to be developed for specific regions if applied outside Florida.

For North Florida, the most influential variables were found to be the proximity of a TTMS to an urban freeway; employment related to hotels, camps, museums, art galleries, and gardens; retired households; seasonal households; population ages 11-13 and 14-17; retired households with low income; residential university; etc. For Central Florida, important variables were agriculture workers; proximity of a TTMS to an urban collector, minor arterial, or principal arterial; median household income; retired households; seasonal households; population ages 14-17; low-income retired households;

proximity to a large residential university; etc. For South Florida, hotels, museums, manufacturing-related employment; retired and seasonal households; and proximity to principal arterials, were significant variables.

For the rural areas, some variables were found to be significant regardless of whether the TTMSs' hourly traffic patterns show a single or double peak. These included the distance to and population of a nearby urban area, seasonal households in South Florida, and retail employees in South Florida. For TTMSs with a single peak, manufacturing employment, truck factor, population ages 18-64, and proximity to a freeway interchange were also found to be important. For the double-peak TTMSs, the distance from a TTMS to the closest public beach was found to be important. Several variables describe the spatial proximity to nearby urban areas and roadway function classes. The significance of these variables suggests that the basis for current practices considering the function class and roadway use may be valid in rural areas.

The assignment method developed in this research offers at least three advantages. First, no additional TTMSs are required to validate the assignment results. This makes this approach more practical and less expensive when compared to, for example, a fuzzy decision tree. Second, a count site may be linked to multiple TTMSs. This provides the analyst with alternative TTMSs in case there is a sufficient basis to reject the best matching TTMS based on the selected variables. Third, this method can be tested with the same TTMSs that are used in the regression analysis. Although this is not to say that there is no need for independent testing using an entirely different set of data, this method allows the development of some understanding of how well the method works. Finally, this method has the potential to eliminate the need to conduct seasonal factor grouping.

The similarity of MSFs patterns were plotted for all existing TTMSs by model region/group, so the roadway section with more obvious seasonal traffic can be easily observed. The urban areas with unique land use and socioeconomic characteristics were also located on the maps as the potential locations for new TTMSs to improve the accuracy for AADT estimation.

## **8.2 Recommendations for Future Research**

The following research efforts are recommended to further develop the results from this research for implementation.

- 1) Study additional categories of variables for rural areas to better explain the variation for traffic seasonality.
- 2) Investigate the effect of inclusion or exclusion of different variables. This is especially true for some of the variables that are correlated with others, even though they are not included in the same equations. For instance, if the exclusion of a variable does not change the assignment results, this variable may be left out to simplify the problem. Conversely, if a variable improves assignment results, it may be included even if its partial  $R^2$  from regression analysis is low. The effectiveness of the weighting scheme and alternative weighting schemes may also be examined to improve the assignment results.
- 3) Explore the feasibility of estimating the seasonal factor for a given PTMS for each month based on those of one or more TTMSs that share similarities on a monthly basis, as opposed to matching a PTMS to a TTMS and borrowing all of the monthly seasonal factors from that TTMS.

In addition, there are many opportunities to improve the efficiency of the entire process. Tasks that may be made highly automated include TTMS data imputation, investigation of abnormal data, variable compilation, regression analysis, and seasonal factor assignment.



## REFERENCES

- Aunet, B. (2000). "Wisconsin's Approach to Variation in Traffic Data." *North American Travel Monitoring Exhibition and Conference CD*, Wisconsin Department of Transportation, Madison, Wisconsin, August 2000.
- Banfield, J.D. and A.E. Raftery (1993). "Model-Based Gaussian and Non-Gaussian Clustering," *Biometrics*, Vol. 49, No. 3, 1993, pp. 803-821.
- Bellamy, P.H. (1978). *Seasonal Variations in Traffic Flows*, Supplementary Report 437, prepared for the Department of the Environment and the Department of Transport, Prepared by Traffic Engineering Department, Transport and Road Research Laboratory, Berkshire, Great Britain, 1978.
- Bureau of Public Roads (1965). *Guide for Traffic Volume Counting Manual*, 2<sup>nd</sup> Edition, U.S. Department of Commerce,
- Capparuccini, D.M., Faghri, A., Polus, A., and Suarez, R.E. (2008). "Fluctuation and Seasonality of Hourly Traffic and Accuracy of Design Hourly Volume Estimates." *Transportation Research Record: Journal of the Transportation Research Board*, No. 2049, Transportation Research Board of the National Academies, Washington, D.C., pp. 63-70.
- Carpenter, G.A. and S. Grossberg (1988). "The Art of Adaptive Pattern Recognition by a Self-Organizing Neural Network," *Computer*, No. 3, Vol. 21, 1988, pp. 77-88.
- Davis, G.A. and Y. Guan (1996). "Bayesian Assignment of Coverage Count Locations to Factor Groups and Estimation of Mean Daily Traffic," *Transportation Research Record 1542*, Transportation Research Board, National Research Council, Washington, D.C., 1996, pp. 30-37.
- Davis, G.A. (1997). *Estimation Theory Approach to Monitoring and Updating Average Daily Traffic*, Final Report, Minnesota Department of Transportation, St. Paul, Minnesota, January 1997.
- Dundar, M.M. and D. Landgrebe (2002). "A Model Based Mixture Supervised Classification Approach in Hyperspectral Data Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 40, No. 12, December 2002, pp. 2692-2699.
- Erhunmwunsee, P.O. (1991) "Estimating Average Annual Daily Traffic Flow from Short Period Counts," *ITE Journal*, November 1991, pp. 23-30.
- Faghri, A., M. Glaubitz, and J. Parameswaran (1986). "Development of Integrated Traffic Monitoring System for Delaware," *Transportation Research Record 1536*, Transportation Research Board, National Research Council, Washington, D.C., 1986, pp. 40-44.

Faghri, A. and J. Hua (1995). "Roadway Seasonal Classification Using Neural Networks," *Journal of Computing in Civil Engineering*, No. 3, Vol. 9, July 1995, pp. 209-215.

FDOT (2002). *Project Traffic Forecasting Handbook*, Draft, Florida Department of Transportation, Tallahassee, Florida, FL.

Flaherty, J. (1993). "Cluster Analysis of Arizona Automatic Traffic Record Data," *Transportation Research Record 1410*, Transportation Research Board, National Research Council, Washington, D.C., 1993, pp. 93-99.

Fraley, C. (1996). *Algorithms for Model-Based Gaussian Hierarchical Clustering*, Technical Report, No. 311, Department of Statistics, University of Washington, Seattle, Washington, October 1996.

Fraley, C. and A.E. Raftery (1998). "How Many Clusters? Which Clustering Method? Answers Via Model Based Cluster Analysis," *The Computer Journal*, Vol. 41, No. 8, 1998, pp. 578-588.

Fraley, C. and A.E. Raftery (2002). *MCLUST: Software for Model-Based Clustering, Density Estimation and Discriminant Analysis*, Technical Report No. 415, Department of Statistics, University of Washington, Seattle, Washington, October 2002.

HCM (2000). *Highway Capacity Manual 2000*, Transportation Research Board, National Research Council, Washington, D.C.

Jang, J., C. Sun, and E. Mizutani (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice-Hall, Inc., Upper Saddle River, New Jersey, 1997.

Janikow, C.Z. (1998). "Fuzzy Decision Trees: Issues and Methods," *IEEE Transactions on Systems, Man, and Cybernetics*, Part B, Vol. 28, No. 1, February 1998, pp. 1-14.

Kamvar, S.D., D. Klein, and C.D. Manning (2002). "Interpreting and Extending Classical Agglomerative Clustering Algorithms Using A Model-Based Approach," *Proceedings of the 19th International Conference on Machine Learning*, 2002, pp. 283-290.

Lam, W. and J. Xu (2000). "Estimation of AADT from Short Period Counts in Hong Kong—A Comparison between Neural Network Method and Regression Analysis," *Journal of Advanced Transportation*, No. 2, Vol. 34, 2000, pp. 249-268.

Li, M.T., F. Zhao, Y. Wu, and A. Gan (2003). "Evaluation of Agglomerative Hierarchical Clustering Methods," Presented at the 82nd Transportation Research Board Annual Meeting, National Research Council, Washington, D.C., 2003.

Li, M.-T., F. Zhao, and L.-F. Chow (2006). *Assignment of Seasonal Factor Categories to Urban Coverage Count Stations Using a Fuzzy Decision Tree*, *ASCE Journal of Transportation Engineering*, Vol. 132, No. 8, 2006, pp. 654-662.

Lingras, P. (1995). "Classifying Highways: Hierarchical Grouping versus Kohonen Neural Networks", *Journal of Transportation Engineering*, No. 4, Vol. 121, 1995, pp. 364-368.

Lingras, P., S.C. Sharma, P. Osborne, and I. Kalyar (2000). "Traffic Volume Time-Series Analysis According to the Type of Road Use," *Journal of Computer-Aided Civil and Infrastructure Engineering*, No. 5, Vol. 15, 2000, pp. 365-373.

Lingras, P. (2001). "Statistical and Genetic Algorithms Classification of Highways," *Journal of Transportation Engineering*, No. 3, Vol. 127, 2001, pp. 237-243.

McDonald, S.M. (1999). *Prototype Demonstration of a Geographic Information System Application for the Seasonal Analysis of Traffic Data, Development of Seasonal Factors and Seasonal Adjustment of Roadways*, Final Report, Prepared for Federal Highway Administration, August 1999.

Ritchie, S.G. (1986). "A Statistical Approach to Statewide Traffic Counting", *Transportation Research Record 1090*, Transportation Research Board, National Research Council, Washington D.C., 1986, pp. 14-21.

Sakawa, M. (2002). *Genetic Algorithms and Fuzzy Multiobjective Optimization*, Kluwer Academic Publishers, Norwell, Massachusetts, 2002.

*SAS-OnlineDOC*, Version 8, SAS Institute Inc., Cary, North Carolina, 1999.

Seaver, W.L., A. chatterjee, and M.L. Seaver (2000), "Estimation of Traffic Volume on Rural Local Roads," *Transportation Research Record 1719*, Transportation Research Board, National Research Council, Washington D.C., 2000, pp 121-128.

Sharma, S.C. and A. Werner (1981). "Improved Method of Grouping Provincewide Permanent Traffic Counters," *Transportation Research Record 815*, Transportation Research Board, National Research Council, Washington, D.C., 1981, pp. 12-18.

Sharma, S.C. (1983). *Improved Classification of Canadian Primary Highways According to Type of Road Use*, *Canadian Journal of Civil Engineering*, No. 3, Vol. 10, pp. 497-509.

Sharma, S.C., P.J. Lingras, M.U. Hassan, and N.A.S. Murthy (1986). *Road Classification According to Driver Population*, *Transportation Research Record 1090*, Transportation Research Board, National Research Council, Washington, D.C., pp. 61-69.

Sharma, S.C. and R.R. Allipuram (1993). "Duration and Frequency of Seasonal Traffic Counts," *Journal of Transportation Engineering*, No. 3, Vol. 119, 1993, pp. 344-359.

Sharma, S.C. and Y. Leng (1994). "Seasonal Traffic Counts for a Precise Estimation of AADT," *ITE Journal*, September 1994, pp. 21-28.

Sharma, S.C., B.M. Gulati, and S.N. Rizak (1996). "Statewide Traffic Volume Studies and Precision of AADT Estimates," *Journal of Transportation Engineering*, No. 6, Vol. 122, 1996, pp. 430-439.

Sharma, S.C., P. Lingras, F. Xu, and G.X. Liu (1999). "Neural networks as Alternative to Traditional Factor Approach of Annual Average Daily Traffic Estimation from Traffic Counts", *Transportation Research Record 1660*, Transportation Research Board, National Research Council, Washington D.C., 1999, pp. 24-31.

Sharma, S.C., P. Lingras, G.X. Liu, and F. Xu (2000). "Estimation of Annual Average Daily Traffic on Low-Volume Roads—Factor Approach Versus Neural Networks," *Transportation Research Record 1719*, Transportation Research Board, National Research Council, Washington D.C., 2000, pp. 103-111.

Sharma, S.C., P. Lingras, F. Xu, and P. Kilburn (2001). "Application of Neural Networks to Estimate AADT on Low-Volume Roads", *Journal of Transportation Engineering*, ASCE, Vol. 127, No. 5, 2001, pp. 426-432.

Smith, B.L. and M.J. Demtsky (1997) "Traffic Flow Forecasting: Comparison of Modeling Approaches," *Journal of Transportation Engineering*, No.4, Vol. 123, 1997, pp. 261-266.

Stamatiadis, N. and D.L. Allen (1997) "Seasonal Factors Using Vehicle Classification Data," *Transportation Research Record 1593*, Transportation Research Board, National Research Council, Washington D.C., 1997, pp. 23-28.

Tantrum, J., A. Murua, and W. Stuetzle (2002), "Hierarchical Model-Based Clustering of Large Datasets through Fraction and Refractionation," *Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining*, 2002, pp. 183-190.

USDOT (2001). *Traffic Monitoring Guide*, Office of Highway Policy Information, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C.

Zhao, F. and S. Chung (2001) *Contributing Factors of Annual Average Daily Traffic in a Florida County*, Transportation Research Record 1769, Transportation Research Board, National Research Council, Washington D.C., 2001, pp. 113-122.

Zhao, F., M.-T. Li, and L.-F. Chow (2004). *Alternatives for Estimating Seasonal Factors on Rural and Urban Roads In Florida*. Final Report for Project BD015-03, Research Office, Florida Department of Transportation, Tallahassee, Florida.

Zhao, F. and Park, N. (2004). "Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic." *Transportation Research Record: Journal of the Transportation Research Board*, No. 1879, Transportation Research Board of the National Academies, Washington, D.C., pp. 99–107.

Zhao, F., Yang, S.S., and Lu, C.X. (2008). *Alternatives for Estimating Seasonal Factors on Rural and Urban Roads in Florida (Phase II)*. Final Report for Project BD015-17, Research Office, Florida Department of Transportation, Tallahassee, FL.

## APPENDIX A

In this appendix, the TTMSs with one or more missing MSFs are listed for urban and rural areas, respectively.

Table A.1 List of TTMSs with Missing Data in Urban Areas

Index	COSITE	Description
1	140013	US 41, 0.4 MI. NORTH OF DALE MABRY HIGHWAY
2	150086	US 92 1 MI EAST OF SAN MARTIN BLVD.
3	460305	US-98/SR-30, APPROX. 250' WEST OF HATHAWAY BRIDGE
4	480159	US 29,0.8 MI N OF US-90-A , WIM#16
5	509940	SR-267, 1 MI. NORTH OF I-10, QUINCY
6	530117	US 90, WEST OF RUSS STREET, MARIANNA
7	559908	US319, 0.3 MI E OF SR 61, TALLAHASSEE, WIM#8
8	589937	SR-87, 180 FEET NORTH OF BASS LN., MILTON
9	729923	I-95, 0.75MI S OF DUNN AVE, JACKSONVILLE, WIM#23
10	799929	US1, 0.25MI N OF RIO GRANDE RD, EDGEWATER, WIM#29
11	870187	SR-836,0.8 MI E OF NW 107TH AVE UNDERPASS,DADE CO.
12	930099	SR-7/US-441 ONE MI. N OF SR-806,(REF 0694) TTMS
13	930174	I-95, S.E. CORNER OF CONGRESS AVE. O.P.,W PALM BCH
14	930257	SR 715, .7 MILES SOUTH OF HOOKER HIGHWAY (TTMS)
15	970267	SR 821, APPROX. 0.5 MI. SOUTH OF NW 25TH ST.
16	970403	TPK, 0.2 MI N OF PEMBROKE RD (TTMS)
17	970410	TPK, 1500 FT N OF SR834/SAMPLE RD
18	970413	TPK, 2627 FT N OF SR806/ATLANTIC AVE TTMS
19	979913	FL TURNPIKE AT BECKER RD OP, SOUTH OF FT PIERCE
20	979934	HOMESTEAD EXTN., SOUTH OF I-75 INTERCHANGE
21	109922	I-275 TAMPA, 0.25MI N OF FLETCHER AVE., WIM#22
22	140199	US-19,1.4 MI. N. OF SR-54,NEW PORT RICHEY,PASCO CO
23	360317	I-75, SB SHOULDER, 0.35 MILES N OF WILLIAMS RD.
24	550207	MERIDIAN RD., NORTH OF BRADFORD RD., TALLAHASSEE
25	710189	US-17,0.6 MI SOUTH OF CR-220,CLAY CO.-- UC 6/94
26	729914	I-295, 3.0 MI N OF I-10
27	799906	I-4, 0.4 MI E ENTERPRISE RD OP -- REPL TTMS 0179
28	920303	I-4/SR-400, APPROX. 0.4 MI. SW OF ORANGE CTY. LINE
29	720157	I-295,3.0 MI N OF I-10,WIM#14 -- UC 9/94
30	100341	SR674-COLLEGE AV, 285 FT W CYPRESS V BLVD-HILLS#53
31	100338	SR583 (56TH ST), 1216 FT S OF SLIGH AVE - HILLS#03
32	100342	SR45/US41, 574 FT N OF TRENTON ST - HILLS#58
33	100339	SR60 (CC CSWY), 1996 FT W ROCKY PT DR - HILLS#18
34	860255	SR 834/SAMPLE RD. 0.14 MI.W OF NW 14TH AVE. TTMS
35	860256	SR 818/GRIFFIN RD, 112' WEST OF SW 70TH AVE. TTMS
36	550201	US-319(CAPITAL CIRCLE), 0.3 MI. EAST OF SR-61

Table A.2 List of TTMSs with Missing Data in Rural Areas

Index	COSITE	Description
1	010014	US 41, 1.4 MI N OF OIL WELL ROAD (R-117,1000,9917)
2	130146	SR64, 1 MI W OF CR675, E OF DESOTO SPDWY @ PTMS 18
3	140079	US 98/301, 0.5 MI SOUTH OF US 301 & 98 JCT.
4	290269	I 10, 0.45 MI EAST OF US41, LAKE CITY
5	300234	SR 349, 0.1 MILES NORTH OF FOREST HILLS
6	479944	SR-69, 2.5 MILES S. OF CITY LINE, SELMAN
7	540245	SR 59, 1150' NORTH OF US 27
8	550211	SR-20, BTWN COES LANDING RD & WILLIAMS LANDING RD
9	550349	SR-61/US-319, MP-15.033, 300' N. OF CHEROKEE ROAD
10	700223	SR-407,0.7 MI. SOUTHWEST OF I-95,BREVARD CO.
11	740047	US 1, 7.0 MI N OF HILLIARD AT STATE LINE
12	750104	SR-50,0.19 MI. W. OF SR-520 NEAR BITHLO (TTMS)
13	799925	US92,0.25MI E OF CLARK'S BAY RD,E OF DELAND,WIM#25
14	890289	SR 76/KANNER HWY, 3 MILES WEST OF CR 711 - TTMS
15	920065	SR-500, 2.0 MI. W OF SR-15 (IN HOLOPAW) (TTMS-C)
16	939935	US-27/SR-25, 1.9 MI. N OF TALISMAN SUGARMILL RD.
17	560301	SR-12,1.7 MILES SOUTH OF GADSEN COUNTY LINE
18	010350	I-75, AIRPORT RD OVERPASS, PUNTA GORDA MP-13.480
19	040271	SR 72, 600' WEST OF CR661
20	090229	SR 66, 430' EAST OF SPARTA ROAD
21	120273	SR 31, 202' NORTH OF FOXHILL ROAD
22	299936	I-10, 50 FT. WEST OF CR-250 OVERPASS, LAKE CITY
23	480348	SR-95/US-29, MP-15.984, 450' N. OF CHURCH ROAD
24	580251	US 90, 0.9 MILES WEST OF OKALOOSA COUNTY
25	599946	SR-363, 1.1 MILES S. OF US-98, ST. MARKS
26	700134	SR-9/I-95,3.34 MI. S. OF SR-514
27	030351	COLLIER CO. I-75, GOLDEN GATE W OF EVERGLADES BLVD
28	609938	US-331/SR-83, APPROX. 3.2 MILES NORTH OF FREEPORT
29	019917	US41, 4.8 MI N OF LEE CO (NEAR R 14, 1000 & 117)
30	079918	SR 25/80, US 27 1.6 MI EAST OF SR 80 R-160
31	549901	I10 JEFFERSON CO, APPROX 1.0 MI E OF SR257, WIM#1
32	609928	I-10/SR-8, APPROX. 1.3 MI. WEST OF BOY SCOUT ROAD
33	269904	I-75/SR-93, 3 MILES NORTH OF MARION COUNTY LINE

## APPENDIX B

Table B.1 Assignment Results for TTMSs in North Florida with Full Variable Set

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
260323	550208 (0.63)	550207 (0.65)	550226 (0.73)	550300 (0.8)	550206 (0.86)
290286	729923 (0.42)	290320 (0.49)	480156 (0.51)	720161 (0.52)	509940 (0.54)
550206	550300 (0.19)	550208 (0.25)	550212 (0.31)	559908 (0.33)	550207 (0.37)
559908	550212 (0.32)	550206 (0.33)	260185 (0.35)	550300 (0.39)	550226 (0.54)
720161	729923 (0.24)	720172 (0.33)	290320 (0.38)	720171 (0.38)	480159 (0.43)
780329	480282 (0.8)	720172 (0.88)	570250 (0.91)	480159 (0.94)	110246 (0.95)
550207	550226 (0.15)	550300 (0.31)	550208 (0.35)	550206 (0.37)	550209 (0.41)
460315	360317 (0.25)	509940 (0.26)	530117 (0.33)	720172 (0.34)	360264 (0.40)
530117	460315 (0.33)	360317 (0.39)	110246 (0.42)	509940 (0.43)	480282 (0.51)
460308	720121 (0.19)	570318 (0.2)	720062 (0.21)	480159 (0.21)	740182 (0.22)
080283	360317 (0.28)	730335 (0.31)	460315 (0.43)	730292 (0.50)	760105 (0.52)
360264	460315 (0.40)	730335 (0.46)	360317 (0.48)	020044 (0.55)	080283 (0.58)
110246	480282 (0.28)	780311 (0.33)	360317 (0.33)	570250 (0.35)	720172 (0.36)
580261	570318 (0.30)	729923 (0.38)	460308 (0.44)	780311 (0.48)	720109 (0.49)
729923	720161 (0.24)	780311 (0.30)	570318 (0.32)	729914 (0.35)	720172 (0.37)
550212	550206 (0.31)	559908 (0.32)	260185 (0.35)	550300 (0.38)	550226 (0.46)
730335	080283 (0.31)	360264 (0.46)	020044 (0.47)	360317 (0.47)	730292 (0.50)
260185	550212 (0.35)	559908 (0.35)	550206 (0.42)	720109 (0.46)	780311 (0.47)
730292	360317 (0.46)	730335 (0.50)	080283 (0.50)	489924 (0.57)	760105 (0.61)
760105	360317 (0.47)	730335 (0.52)	080283 (0.52)	460315 (0.61)	730292 (0.61)
290320	720161 (0.38)	729923 (0.42)	290286 (0.49)	720171 (0.55)	570318 (0.56)
509940	780311 (0.23)	460315 (0.26)	720172 (0.29)	360317 (0.31)	480159 (0.33)
550151	550304 (0.78)	550300 (1.01)	550207 (1.02)	550226 (1.08)	550208 (1.10)
550208	550300 (0.21)	550206 (0.25)	550207 (0.35)	550226 (0.38)	550212 (0.52)
550209	550226 (0.33)	550207 (0.41)	550300 (0.48)	260185 (0.60)	480282 (0.60)
550213	550206 (0.65)	550300 (0.73)	550208 (0.75)	559908 (0.83)	550212 (0.83)
550226	550207 (0.15)	550300 (0.33)	550209 (0.33)	550208 (0.38)	550206 (0.41)
720062	720121 (0.17)	460308 (0.21)	570167 (0.24)	480159 (0.33)	720172 (0.35)
720121	570167 (0.12)	720062 (0.17)	460308 (0.19)	480159 (0.28)	480325 (0.29)
720172	480159 (0.19)	780311 (0.21)	480282 (0.24)	570250 (0.25)	460308 (0.27)
720216	489924 (0.22)	460308 (0.34)	480159 (0.36)	729914 (0.36)	720121 (0.39)
740182	360249 (0.15)	460308 (0.22)	720109 (0.24)	570318 (0.3)	589937 (0.33)
780311	720172 (0.21)	480159 (0.22)	509940 (0.23)	360249 (0.24)	720109 (0.29)
720109	589937 (0.17)	460308 (0.22)	740182 (0.24)	360249 (0.25)	480159 (0.26)
720171	720161 (0.38)	489924 (0.45)	729914 (0.46)	720216 (0.5)	290320 (0.55)
550300	550206 (0.19)	550208 (0.21)	550207 (0.31)	550226 (0.33)	550212 (0.38)
550304	720171 (0.69)	550300 (0.77)	550151 (0.78)	550208 (0.79)	550207 (0.82)
729914	489924 (0.27)	729923 (0.35)	720216 (0.36)	780311 (0.45)	720171 (0.46)
460166	570293 (2.16)	460305 (2.19)	780329 (2.47)	760105 (2.56)	600168 (2.62)
460305	720161 (0.75)	290320 (0.80)	720171 (0.86)	580261 (0.88)	729923 (0.95)
480159	720172 (0.19)	360249 (0.21)	460308 (0.21)	780311 (0.22)	720109 (0.26)
480282	720172 (0.24)	570250 (0.25)	110246 (0.28)	480159 (0.29)	360249 (0.30)
480325	570167 (0.25)	720121 (0.29)	589937 (0.31)	720109 (0.36)	720062 (0.42)



Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
570250	720172 (0.25)	480282 (0.25)	360249 (0.26)	740182 (0.33)	110246 (0.35)
570293	290320 (1.65)	720161 (1.66)	550151 (1.77)	720171 (1.79)	460305 (1.79)
589937	720109 (0.17)	570167 (0.29)	720121 (0.30)	460308 (0.31)	480325 (0.31)
570318	460308 (0.20)	480159 (0.28)	720109 (0.29)	360249 (0.29)	780311 (0.29)
489924	720216 (0.22)	729914 (0.27)	720172 (0.3)	480159 (0.39)	780311 (0.42)
480156	460308 (0.26)	480159 (0.28)	570318 (0.31)	360249 (0.31)	720172 (0.32)
570167	720121 (0.12)	460308 (0.23)	720062 (0.24)	480325 (0.25)	589937 (0.29)
600168	760105 (1.73)	460308 (2.00)	720062 (2.00)	720172 (2.02)	480159 (2.02)
710189	720121 (0.73)	720062 (0.74)	570167 (0.79)	480156 (0.80)	480159 (0.85)
080294	110177 (0.50)	760105 (0.76)	360264 (0.77)	730335 (0.78)	020044 (0.81)
110177	080294 (0.50)	020044 (0.67)	360264 (0.71)	730335 (0.71)	760105 (0.72)
360249	740182 (0.15)	480159 (0.21)	780311 (0.24)	720109 (0.25)	570250 (0.26)
020044	730335 (0.47)	360317 (0.49)	360264 (0.55)	080283 (0.57)	730292 (0.62)
360317	460315 (0.25)	080283 (0.28)	509940 (0.31)	110246 (0.33)	530117 (0.39)

Table B.2 Assignment Results for TTMSs in Central Florida with Full Variable Set

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5 <sup>th</sup>
100106	109926 (0.64)	100224 (0.66)	100194 (0.67)	109922 (0.69)	750196 (0.69)
100321	770102 (0.18)	100162 (0.2)	750154 (0.23)	750038 (0.32)	150295 (0.33)
150302	109922 (0.49)	100194 (0.51)	100224 (0.53)	750130 (0.53)	109926 (0.54)
109922	100194 (0.09)	109926 (0.13)	750196 (0.13)	100224 (0.14)	770343 (0.15)
169927	140199 (0.45)	150295 (0.58)	700113 (0.62)	750038 (0.72)	100162 (0.72)
750038	700113 (0.17)	770102 (0.23)	150295 (0.28)	100321 (0.32)	770197 (0.35)
750175	100110 (0.21)	109922 (0.23)	750196 (0.24)	100194 (0.26)	109926 (0.28)
770197	700113 (0.33)	750038 (0.35)	150295 (0.36)	140013 (0.40)	770102 (0.41)
790133	100110 (0.35)	750196 (0.39)	750175 (0.40)	109922 (0.40)	100123 (0.43)
799929	790133 (0.55)	700284 (0.70)	750175 (0.72)	100110 (0.79)	160275 (0.79)
100080	750175 (0.48)	750196 (0.51)	100194 (0.51)	109922 (0.51)	100110 (0.52)
100110	750196 (0.17)	750175 (0.21)	100123 (0.21)	100194 (0.22)	109922 (0.24)
100123	750196 (0.21)	100110 (0.21)	100194 (0.25)	109922 (0.28)	109926 (0.32)
100194	109922 (0.09)	109926 (0.11)	100224 (0.14)	750196 (0.15)	770343 (0.19)
140013	770102 (0.29)	100321 (0.33)	150295 (0.35)	750038 (0.36)	770197 (0.40)
140190	750130 (0.11)	100224 (0.20)	109922 (0.22)	770343 (0.22)	109926 (0.23)
150086	750154 (0.28)	100162 (0.32)	770102 (0.44)	100321 (0.47)	150295 (0.52)
150183	790133 (0.50)	100080 (0.54)	160310 (0.56)	750196 (0.69)	100110 (0.69)
150295	700113 (0.23)	750038 (0.28)	100321 (0.33)	750154 (0.34)	140013 (0.35)
100224	109926 (0.04)	100194 (0.14)	109922 (0.14)	750130 (0.16)	770343 (0.17)
109926	100224 (0.04)	100194 (0.11)	109922 (0.13)	770343 (0.17)	750130 (0.19)
150066	160310 (0.83)	150302 (0.92)	100110 (1.00)	750175 (1.03)	700284 (1.07)
100162	100321 (0.200)	750154 (0.24)	770102 (0.25)	150086 (0.32)	750038 (0.36)
140199	150295 (0.37)	700113 (0.39)	169927 (0.45)	750038 (0.49)	100321 (0.56)
160275	799929 (0.79)	160310 (1.20)	790133 (1.21)	169927 (1.36)	700284 (1.43)
160128	100106 (1.30)	150183 (1.64)	790133 (1.72)	100080 (1.72)	100123 (1.73)
160310	150183 (0.56)	100080 (0.70)	790133 (0.81)	150066 (0.83)	750175 (0.86)
700113	750038 (0.17)	150295 (0.23)	770102 (0.31)	100321 (0.33)	770197 (0.33)
700114	750154 (0.65)	150295 (0.66)	150086 (0.66)	750038 (0.70)	100162 (0.70)
700284	790133 (0.46)	750130 (0.51)	140190 (0.52)	770343 (0.54)	100080 (0.55)
750130	140190 (0.11)	100224 (0.16)	770343 (0.16)	109922 (0.18)	109926 (0.19)
750154	770102 (0.20)	100321 (0.23)	100162 (0.24)	150086 (0.28)	150295 (0.34)
750196	109922 (0.13)	100194 (0.15)	100110 (0.17)	770343 (0.21)	100123 (0.21)
750204	109922 (0.24)	100194 (0.24)	750196 (0.24)	109926 (0.25)	100224 (0.27)
770102	100321 (0.18)	750154 (0.20)	750038 (0.23)	100162 (0.25)	140013 (0.29)
770343	109922 (0.15)	750130 (0.16)	109926 (0.17)	100224 (0.17)	100194 (0.19)
920265	150086 (0.56)	750154 (0.63)	100321 (0.69)	770102 (0.71)	100162 (0.71)
979932	750175 (0.55)	100110 (0.55)	750196 (0.59)	109922 (0.63)	790133 (0.64)

Table B.3 Assignment Results for TTMSs in South Florida with Full Variable Set

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
170181	130180 (0.09)	979913 (0.12)	880314 (0.20)	890332 (0.21)	170225 (0.22)
030094	899921 (0.16)	860214 (0.22)	130333 (0.29)	030191 (0.36)	860176 (0.45)
860150	010228 (0.60)	030094 (0.75)	860306 (0.76)	899921 (0.86)	860214 (0.89)
860214	899921 (0.21)	030094 (0.22)	130333 (0.44)	030191 (0.52)	010228 (0.57)
860306	030094 (0.61)	860214 (0.64)	899921 (0.75)	860150 (0.76)	930010 (0.80)
870178	870266 (0.12)	870188 (0.12)	870193 (0.13)	930257 (0.14)	860222 (0.16)
870266	870188 (0.07)	970267 (0.07)	979934 (0.11)	930257 (0.11)	870178 (0.12)
930099	930101 (0.12)	940260 (0.2)	970403 (0.21)	860186 (0.21)	120203 (0.21)
970267	870266 (0.07)	979934 (0.08)	870188 (0.11)	970430 (0.12)	930257 (0.14)
970416	930174 (0.08)	970417 (0.13)	890332 (0.15)	970410 (0.17)	880314 (0.19)
979934	970267 (0.08)	870266 (0.11)	870188 (0.13)	970430 (0.16)	930257 (0.18)
890332	970417 (0.10)	880314 (0.10)	930174 (0.11)	979913 (0.14)	970416 (0.15)
970421	860163 (0.11)	880314 (0.15)	890332 (0.21)	930198 (0.22)	970416 (0.23)
930198	860331 (0.17)	930217 (0.20)	120184 (0.21)	970421 (0.22)	880314 (0.23)
870193	870188 (0.12)	870178 (0.13)	870266 (0.15)	930257 (0.16)	970267 (0.17)
120184	860176 (0.11)	860331 (0.15)	930217 (0.17)	930198 (0.21)	030191 (0.22)
010228	030094 (0.53)	860214 (0.57)	899921 (0.57)	860150 (0.60)	130333 (0.76)
040145	930217 (0.15)	860331 (0.15)	120184 (0.22)	970417 (0.22)	930198 (0.23)
170225	979913 (0.13)	940260 (0.15)	979933 (0.16)	970410 (0.18)	860186 (0.18)
130333	030191 (0.13)	860176 (0.22)	899921 (0.27)	030094 (0.29)	120184 (0.29)
130180	170181 (0.09)	979913 (0.16)	170225 (0.20)	880314 (0.25)	890332 (0.27)
030191	130333 (0.13)	860176 (0.16)	120184 (0.22)	930217 (0.25)	860331 (0.32)
120203	879930 (0.10)	860186 (0.13)	970403 (0.14)	870187 (0.16)	940260 (0.16)
860163	970421 (0.11)	970410 (0.16)	880314 (0.18)	860331 (0.21)	890332 (0.21)
860176	120184 (0.11)	030191 (0.16)	130333 (0.22)	930217 (0.23)	860331 (0.25)
860186	970403 (0.03)	940260 (0.08)	940334 (0.09)	860298 (0.11)	930101 (0.11)
860215	930087 (0.37)	010228 (0.95)	860214 (1.46)	030094 (1.46)	899921 (1.51)
860222	870187 (0.08)	979933 (0.09)	860298 (0.10)	860186 (0.13)	970403 (0.15)
860298	979933 (0.09)	860222 (0.10)	860186 (0.11)	970403 (0.13)	930101 (0.14)
860331	930217 (0.10)	040145 (0.15)	120184 (0.15)	930198 (0.17)	970416 (0.20)
870031	870108 (0.25)	930174 (0.35)	970416 (0.35)	970417 (0.38)	979913 (0.39)
870096	870187 (0.26)	970267 (0.28)	979934 (0.30)	860222 (0.31)	979933 (0.31)
870108	940260 (0.23)	940334 (0.23)	170225 (0.23)	970416 (0.24)	870031 (0.25)
870187	860222 (0.08)	979933 (0.11)	860298 (0.14)	860186 (0.14)	120203 (0.16)
870188	870266 (0.07)	970267 (0.11)	930257 (0.12)	870193 (0.12)	870178 (0.12)
870258	860186 (0.31)	120203 (0.31)	970403 (0.31)	940334 (0.33)	879930 (0.34)
879930	120203 (0.10)	860186 (0.16)	970403 (0.17)	940260 (0.19)	979933 (0.21)
930010	860176 (0.30)	120184 (0.33)	930198 (0.41)	930217 (0.41)	860331 (0.41)
930087	860215 (0.37)	010228 (0.86)	860150 (1.19)	030094 (1.19)	899921 (1.22)
930101	940260 (0.10)	970403 (0.10)	860186 (0.11)	930099 (0.12)	860298 (0.14)
930174	970416 (0.08)	970417 (0.11)	890332 (0.11)	880314 (0.15)	970410 (0.16)
930217	860331 (0.10)	040145 (0.15)	120184 (0.17)	930198 (0.2)	970417 (0.21)
930257	870266 (0.11)	870188 (0.12)	970267 (0.14)	870178 (0.14)	870193 (0.16)
970403	860186 (0.03)	940260 (0.07)	940334 (0.09)	930101 (0.10)	860298 (0.13)
970410	930174 (0.16)	860163 (0.16)	979913 (0.17)	970416 (0.17)	170225 (0.18)

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
970413	930198 (0.31)	120184 (0.36)	860176 (0.40)	930217 (0.41)	130333 (0.41)
970417	890332 (0.10)	930174 (0.11)	970416 (0.13)	880314 (0.14)	979913 (0.17)
970430	970267 (0.12)	870266 (0.14)	979934 (0.16)	870188 (0.17)	870187 (0.17)
979933	860298 (0.09)	860222 (0.09)	870187 (0.11)	860186 (0.12)	970403 (0.14)
880314	890332 (0.10)	970417 (0.14)	970421 (0.15)	930174 (0.15)	979913 (0.17)
880326	870193 (0.25)	870178 (0.26)	870188 (0.26)	860222 (0.28)	860298 (0.28)
890259	860215 (1.75)	930087 (2.04)	010228 (2.62)	860150 (2.89)	860306 (2.90)
899921	030094 (0.16)	860214 (0.21)	130333 (0.27)	030191 (0.38)	860176 (0.48)
940260	970403 (0.07)	860186 (0.08)	930101 (0.10)	940334 (0.12)	860298 (0.14)
940334	970403 (0.09)	860186 (0.09)	940260 (0.12)	979933 (0.15)	860298 (0.15)
979913	170181 (0.12)	170225 (0.13)	890332 (0.14)	130180 (0.16)	930174 (0.16)

Table B.4 Assignment Results for TTMSs in North Florida with Reduced Variable Set

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
260323	550208 (0.41)	550207 (0.55)	550300 (0.59)	710189 (0.59)	550226 (0.63)
290286	729923 (0.22)	720161 (0.31)	509940 (0.33)	290320 (0.35)	720172 (0.42)
550206	550300 (0.17)	550207 (0.23)	550208 (0.24)	550226 (0.26)	550212 (0.27)
559908	260185 (0.25)	550212 (0.26)	550206 (0.28)	550300 (0.35)	550226 (0.38)
720161	729923 (0.21)	290320 (0.30)	720172 (0.31)	290286 (0.31)	720171 (0.36)
780329	720172 (0.76)	480282 (0.77)	480159 (0.8)	480156 (0.82)	570250 (0.84)
550207	550226 (0.11)	550300 (0.17)	550208 (0.20)	550206 (0.23)	550209 (0.30)
460315	530117 (0.19)	509940 (0.23)	360317 (0.23)	110246 (0.3)	570250 (0.31)
530117	460315 (0.19)	360317 (0.25)	509940 (0.30)	110246 (0.33)	080283 (0.38)
460308	740182 (0.11)	480156 (0.12)	360249 (0.16)	720121 (0.17)	570318 (0.18)
080283	360317 (0.22)	730335 (0.27)	460315 (0.37)	530117 (0.38)	110246 (0.41)
360264	460315 (0.38)	730335 (0.43)	360317 (0.47)	020044 (0.51)	530117 (0.52)
110246	360317 (0.21)	780311 (0.22)	720172 (0.23)	509940 (0.23)	480282 (0.25)
580261	570318 (0.27)	729923 (0.32)	480156 (0.36)	740182 (0.37)	360249 (0.38)
729923	720161 (0.21)	290286 (0.22)	480156 (0.26)	780311 (0.27)	570318 (0.29)
550212	559908 (0.26)	260185 (0.27)	550206 (0.27)	550226 (0.34)	550207 (0.36)
730335	080283 (0.27)	020044 (0.42)	360264 (0.43)	360317 (0.43)	730292 (0.48)
260185	559908 (0.25)	550212 (0.27)	550206 (0.33)	550226 (0.35)	550209 (0.35)
730292	360317 (0.44)	080283 (0.46)	730335 (0.48)	489924 (0.54)	760105 (0.57)
760105	360317 (0.42)	080283 (0.45)	730335 (0.49)	110246 (0.53)	509940 (0.55)
290320	720161 (0.30)	729923 (0.33)	290286 (0.35)	720171 (0.47)	570318 (0.49)
509940	780311 (0.20)	110246 (0.23)	460315 (0.23)	360249 (0.25)	480156 (0.26)
550151	550304 (0.71)	550207 (0.82)	550226 (0.88)	550300 (0.90)	550208 (0.97)
550208	550300 (0.19)	550207 (0.20)	550226 (0.24)	550206 (0.24)	550209 (0.40)
550209	550226 (0.21)	550300 (0.28)	550207 (0.30)	260185 (0.35)	550208 (0.40)
550213	550206 (0.55)	550300 (0.63)	550208 (0.64)	550207 (0.71)	550212 (0.72)
550226	550207 (0.11)	550209 (0.21)	550300 (0.21)	550208 (0.24)	550206 (0.26)
720062	720121 (0.13)	460308 (0.18)	570167 (0.2)	740182 (0.25)	480159 (0.30)
720121	570167 (0.09)	720062 (0.13)	460308 (0.17)	480325 (0.18)	589937 (0.24)
720172	570250 (0.09)	480282 (0.13)	480159 (0.17)	480156 (0.17)	360249 (0.17)
720216	489924 (0.20)	460308 (0.31)	729914 (0.34)	480159 (0.34)	360249 (0.34)
740182	460308 (0.11)	720109 (0.12)	360249 (0.13)	480156 (0.14)	570318 (0.17)
780311	360249 (0.13)	480156 (0.17)	720172 (0.19)	480282 (0.2)	509940 (0.20)
720109	740182 (0.12)	360249 (0.13)	589937 (0.14)	460308 (0.18)	480156 (0.19)
720171	720161 (0.36)	729914 (0.42)	489924 (0.44)	290320 (0.47)	720216 (0.48)
550300	550206 (0.17)	550207 (0.17)	550208 (0.19)	550226 (0.21)	550209 (0.28)
550304	720171 (0.58)	550207 (0.69)	550300 (0.7)	550208 (0.71)	550151 (0.71)
729914	489924 (0.24)	729923 (0.32)	720216 (0.34)	780311 (0.40)	480156 (0.41)
460166	570293 (2.01)	460305 (2.04)	760105 (2.41)	780329 (2.43)	290320 (2.47)
460305	290320 (0.70)	720161 (0.72)	720171 (0.82)	580261 (0.85)	729923 (0.89)
480159	360249 (0.11)	480156 (0.13)	720172 (0.17)	480282 (0.18)	570250 (0.19)
480282	720172 (0.13)	480159 (0.18)	570250 (0.18)	780311 (0.20)	110246 (0.25)
480325	570167 (0.15)	720121 (0.18)	589937 (0.21)	720109 (0.27)	360249 (0.29)
570250	720172 (0.09)	480282 (0.18)	480159 (0.19)	360249 (0.19)	480156 (0.21)
570293	290320 (1.57)	550151 (1.61)	720161 (1.64)	460305 (1.75)	720171 (1.77)

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
589937	720109 (0.14)	740182 (0.20)	480325 (0.21)	360249 (0.21)	720121 (0.24)
570318	480156 (0.15)	740182 (0.17)	360249 (0.18)	460308 (0.18)	480159 (0.25)
489924	720216 (0.20)	729914 (0.24)	720172 (0.28)	480156 (0.32)	570250 (0.35)
480156	360249 (0.09)	460308 (0.12)	480159 (0.13)	740182 (0.14)	570318 (0.15)
570167	720121 (0.09)	480325 (0.15)	460308 (0.19)	720062 (0.20)	480156 (0.23)
600168	760105 (1.69)	480282 (1.92)	570250 (1.94)	460308 (1.96)	480159 (1.96)
710189	720121 (0.51)	720062 (0.53)	480325 (0.56)	570167 (0.58)	260323 (0.59)
080294	110177 (0.41)	760105 (0.72)	360264 (0.73)	730335 (0.74)	020044 (0.74)
110177	080294 (0.41)	020044 (0.62)	360264 (0.64)	760105 (0.65)	730335 (0.67)
360249	480156 (0.09)	480159 (0.11)	780311 (0.13)	740182 (0.13)	720109 (0.13)
020044	730335 (0.42)	360317 (0.45)	080283 (0.50)	360264 (0.51)	760105 (0.56)
360317	110246 (0.21)	080283 (0.22)	460315 (0.23)	530117 (0.25)	509940 (0.27)

Table B.5 Assignment Results for TTMSs in Central Florida with Reduced Variable Set

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
100106	100080 (0.55)	150183 (0.57)	750204 (0.58)	750196 (0.58)	109926 (0.59)
100321	100162 (0.04)	770102 (0.08)	750154 (0.11)	750038 (0.19)	140013 (0.23)
150302	750204 (0.42)	109922 (0.42)	100194 (0.42)	100224 (0.44)	109926 (0.44)
109922	100194 (0.03)	750204 (0.04)	109926 (0.06)	750196 (0.06)	100224 (0.08)
169927	140199 (0.32)	150295 (0.43)	700113 (0.45)	750038 (0.53)	100162 (0.60)
750038	700113 (0.10)	770102 (0.12)	150295 (0.13)	100162 (0.15)	750154 (0.17)
750175	100110 (0.07)	750196 (0.08)	109922 (0.11)	100123 (0.11)	100194 (0.12)
770197	140013 (0.18)	750038 (0.20)	700113 (0.22)	150295 (0.22)	750154 (0.23)
790133	100110 (0.19)	700284 (0.24)	100123 (0.24)	750175 (0.25)	750196 (0.25)
799929	790133 (0.42)	700284 (0.53)	979932 (0.57)	100123 (0.60)	100110 (0.60)
100080	150183 (0.22)	100123 (0.28)	750196 (0.31)	100110 (0.32)	100194 (0.34)
100110	750175 (0.07)	750196 (0.08)	100123 (0.09)	750204 (0.12)	109922 (0.13)
100123	750196 (0.07)	100110 (0.09)	109922 (0.11)	750175 (0.11)	100194 (0.12)
100194	109922 (0.03)	109926 (0.06)	750196 (0.06)	750204 (0.07)	100224 (0.08)
140013	770102 (0.15)	770197 (0.18)	750154 (0.19)	100162 (0.19)	750038 (0.19)
140190	750130 (0.06)	750204 (0.12)	100224 (0.12)	770343 (0.13)	109926 (0.14)
150086	750154 (0.16)	770102 (0.20)	140013 (0.22)	100162 (0.22)	920265 (0.24)
150183	100080 (0.22)	790133 (0.36)	160310 (0.36)	700284 (0.47)	100110 (0.48)
150295	700113 (0.12)	750038 (0.13)	750154 (0.22)	770197 (0.22)	100162 (0.23)
100224	109926 (0.03)	770343 (0.07)	109922 (0.08)	100194 (0.08)	750204 (0.09)
109926	100224 (0.03)	100194 (0.06)	109922 (0.06)	770343 (0.07)	750204 (0.07)
150066	150302 (0.64)	160310 (0.67)	700284 (0.82)	750175 (0.83)	100110 (0.84)
100162	100321 (0.04)	770102 (0.05)	750154 (0.09)	750038 (0.15)	140013 (0.19)
140199	150295 (0.26)	700113 (0.31)	169927 (0.32)	750038 (0.39)	770197 (0.46)
160275	799929 (0.65)	160310 (1.03)	790133 (1.06)	979932 (1.14)	169927 (1.17)
160128	100106 (1.08)	150183 (1.29)	100080 (1.41)	790133 (1.43)	100123 (1.49)
160310	150183 (0.36)	100080 (0.51)	150066 (0.67)	790133 (0.68)	979932 (0.74)
700113	750038 (0.10)	150295 (0.12)	100162 (0.19)	770102 (0.21)	770197 (0.22)
700114	150295 (0.50)	750038 (0.52)	140013 (0.54)	750154 (0.55)	770197 (0.55)
700284	790133 (0.24)	100110 (0.34)	770343 (0.34)	140190 (0.35)	750130 (0.36)
750130	140190 (0.06)	100224 (0.09)	770343 (0.10)	750204 (0.10)	109926 (0.11)
750154	770102 (0.07)	100162 (0.09)	100321 (0.11)	150086 (0.16)	750038 (0.17)
750196	100194 (0.06)	109922 (0.06)	100123 (0.07)	750175 (0.08)	100110 (0.08)
750204	109922 (0.04)	100194 (0.07)	109926 (0.07)	100224 (0.09)	750196 (0.09)
770102	100162 (0.05)	750154 (0.07)	100321 (0.08)	750038 (0.12)	140013 (0.15)
770343	109926 (0.07)	100224 (0.07)	750130 (0.10)	750204 (0.10)	109922 (0.11)
920265	150086 (0.24)	750154 (0.31)	100162 (0.35)	770102 (0.36)	100321 (0.37)
979932	100110 (0.23)	750175 (0.24)	790133 (0.25)	100123 (0.26)	750196 (0.27)

Table B.6 Assignment Results for TTMSs in South Florida with Reduced Variable Set

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
170181	130180 (0.06)	979913 (0.09)	880314 (0.18)	890332 (0.19)	170225 (0.19)
030094	899921 (0.10)	860214 (0.14)	130333 (0.28)	030191 (0.33)	860176 (0.44)
860150	010228 (0.54)	030094 (0.73)	860306 (0.75)	899921 (0.78)	860214 (0.80)
860214	030094 (0.14)	899921 (0.17)	130333 (0.35)	010228 (0.45)	030191 (0.45)
860306	860214 (0.56)	030094 (0.60)	899921 (0.69)	860150 (0.75)	930010 (0.78)
870178	870266 (0.08)	870188 (0.09)	930257 (0.10)	870193 (0.12)	970267 (0.13)
870266	930257 (0.03)	870188 (0.04)	970267 (0.05)	979934 (0.08)	870178 (0.08)
930099	930101 (0.09)	940260 (0.16)	120203 (0.17)	970403 (0.17)	860186 (0.18)
970267	870266 (0.05)	979934 (0.06)	930257 (0.06)	970430 (0.09)	870188 (0.09)
970416	930174 (0.07)	970417 (0.11)	870108 (0.12)	890332 (0.13)	970410 (0.16)
979934	970267 (0.06)	870266 (0.08)	930257 (0.09)	870188 (0.11)	970430 (0.13)
890332	970417 (0.06)	880314 (0.08)	930174 (0.09)	979913 (0.11)	970416 (0.13)
970421	860163 (0.09)	880314 (0.13)	930198 (0.19)	970416 (0.19)	970410 (0.20)
930198	860331 (0.14)	930217 (0.15)	120184 (0.18)	040145 (0.18)	970421 (0.19)
870193	870188 (0.10)	870266 (0.11)	970430 (0.11)	870178 (0.12)	930257 (0.12)
120184	860176 (0.09)	930217 (0.13)	860331 (0.14)	930198 (0.18)	040145 (0.19)
010228	860214 (0.45)	030094 (0.48)	899921 (0.50)	860150 (0.54)	130333 (0.71)
040145	930217 (0.11)	860331 (0.13)	970417 (0.18)	930198 (0.18)	120184 (0.19)
170225	870108 (0.11)	940260 (0.11)	979913 (0.12)	979933 (0.15)	970403 (0.15)
130333	030191 (0.10)	860176 (0.21)	899921 (0.22)	120184 (0.27)	030094 (0.28)
130180	170181 (0.06)	979913 (0.12)	170225 (0.17)	870108 (0.22)	880314 (0.23)
030191	130333 (0.10)	860176 (0.14)	120184 (0.19)	930217 (0.25)	860331 (0.30)
120203	879930 (0.06)	860186 (0.10)	970403 (0.10)	940260 (0.14)	870187 (0.14)
860163	970421 (0.09)	970410 (0.16)	880314 (0.17)	860331 (0.19)	890332 (0.20)
860176	120184 (0.09)	030191 (0.14)	930217 (0.21)	130333 (0.21)	860331 (0.23)
860186	970403 (0.02)	940260 (0.06)	940334 (0.07)	860298 (0.10)	860222 (0.10)
860215	930087 (0.31)	010228 (0.95)	860214 (1.33)	030094 (1.41)	899921 (1.43)
860222	870187 (0.07)	979933 (0.08)	860298 (0.08)	860186 (0.10)	970403 (0.11)
860298	860222 (0.08)	979933 (0.08)	860186 (0.10)	970403 (0.11)	930101 (0.11)
860331	930217 (0.08)	040145 (0.13)	930198 (0.14)	120184 (0.14)	970416 (0.19)
870031	870108 (0.22)	930174 (0.22)	970416 (0.22)	970417 (0.23)	979913 (0.25)
870096	870187 (0.22)	860222 (0.26)	970267 (0.26)	979933 (0.28)	979934 (0.29)
870108	940334 (0.11)	170225 (0.11)	940260 (0.11)	970416 (0.12)	970403 (0.15)
870187	860222 (0.07)	979933 (0.11)	970430 (0.11)	860186 (0.13)	860298 (0.14)
870188	870266 (0.04)	930257 (0.05)	970267 (0.09)	870178 (0.09)	870193 (0.10)
870258	120203 (0.26)	940334 (0.28)	860186 (0.28)	970403 (0.29)	879930 (0.31)
879930	120203 (0.06)	860186 (0.14)	970403 (0.15)	870187 (0.17)	940260 (0.17)
930010	860176 (0.29)	120184 (0.31)	930217 (0.37)	930198 (0.38)	860331 (0.38)
930087	860215 (0.31)	010228 (0.80)	899921 (1.15)	860150 (1.18)	030094 (1.18)
930101	940260 (0.08)	930099 (0.09)	970403 (0.09)	860186 (0.1)	860298 (0.11)
930174	970416 (0.07)	970417 (0.09)	890332 (0.09)	880314 (0.12)	979913 (0.14)
930217	860331 (0.08)	040145 (0.11)	120184 (0.13)	930198 (0.15)	970417 (0.21)
930257	870266 (0.03)	870188 (0.05)	970267 (0.06)	979934 (0.09)	870178 (0.10)
970403	860186 (0.02)	940260 (0.05)	940334 (0.06)	930101 (0.09)	120203 (0.10)
970410	930174 (0.15)	979913 (0.15)	860163 (0.16)	970416 (0.16)	170225 (0.17)



Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
970413	930198 (0.23)	120184 (0.25)	930217 (0.29)	860176 (0.30)	130333 (0.32)
970417	890332 (0.06)	930174 (0.09)	880314 (0.1)	970416 (0.11)	979913 (0.16)
970430	970267 (0.09)	870266 (0.09)	870193 (0.11)	870188 (0.11)	860222 (0.11)
979933	860222 (0.08)	860298 (0.08)	860186 (0.10)	870187 (0.11)	970403 (0.11)
880314	890332 (0.08)	970417 (0.10)	930174 (0.12)	970421 (0.13)	979913 (0.15)
880326	870188 (0.18)	870193 (0.18)	870178 (0.18)	870266 (0.20)	860222 (0.20)
890259	860215 (1.74)	930087 (1.98)	010228 (2.61)	860150 (2.82)	860306 (2.84)
899921	030094 (0.10)	860214 (0.17)	130333 (0.22)	030191 (0.31)	860176 (0.43)
940260	970403 (0.05)	860186 (0.06)	940334 (0.08)	930101 (0.08)	870108 (0.11)
940334	970403 (0.06)	860186 (0.07)	940260 (0.08)	870108 (0.11)	860298 (0.11)
979913	170181 (0.09)	890332 (0.11)	170225 (0.12)	130180 (0.12)	930174 (0.14)

Table B.7 Assignment Results for Rrural TTMSs in Single Peak Group with Full Variable set

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
030270	030351 (2.84)	530218 (2.86)	010350 (2.87)	700134 (2.89)	709919 (2.96)
290269	299936 (0.52)	370238 (0.75)	610152 (0.77)	380280 (0.83)	480348 (0.85)
320112	370238 (0.94)	610152 (1.02)	740132 (1.22)	340116 (1.24)	290269 (1.25)
340116	349909 (0.14)	470328 (0.64)	740047 (0.66)	480348 (0.74)	470173 (0.86)
340239	349909 (0.93)	340116 (1.06)	480348 (1.12)	470173 (1.19)	470328 (1.22)
349909	340116 (0.14)	480348 (0.67)	470328 (0.68)	470173 (0.75)	740047 (0.78)
470173	480348 (0.43)	600346 (0.48)	610152 (0.60)	490244 (0.68)	470328 (0.73)
490060	480348 (1.21)	490244 (1.32)	979931 (1.35)	470328 (1.35)	470173 (1.39)
490244	480348 (0.65)	470173 (0.68)	600346 (0.83)	470328 (0.94)	510316 (0.97)
530050	530248 (0.91)	470328 (1.15)	370238 (1.17)	480348 (1.23)	470173 (1.30)
530248	530050 (0.91)	380280 (1.04)	290269 (1.04)	470173 (1.09)	470328 (1.10)
600346	470173 (0.48)	480348 (0.52)	610152 (0.82)	490244 (0.83)	470328 (0.84)
610152	470173 (0.60)	290269 (0.77)	600346 (0.82)	370238 (0.82)	480348 (0.85)
740047	340116 (0.66)	349909 (0.78)	470173 (0.82)	480348 (0.83)	470328 (0.91)
740132	610152 (0.85)	290269 (0.93)	299936 (1.09)	189920 (1.12)	709919 (1.17)
880139	709919 (0.88)	189920 (0.98)	720235 (0.98)	290269 (1.05)	299936 (1.07)
979931	189920 (1.02)	700322 (1.23)	290269 (1.25)	490060 (1.35)	370238 (1.36)
720235	709919 (0.83)	740047 (0.98)	880139 (0.98)	189920 (1.00)	290269 (1.02)
370238	290269 (0.75)	610152 (0.82)	340116 (0.87)	470328 (0.93)	320112 (0.94)
600287	290269 (0.90)	610152 (0.93)	189920 (1.02)	299936 (1.10)	470173 (1.13)
700322	189920 (0.70)	709919 (0.71)	010350 (0.73)	700134 (0.88)	740047 (1.14)
090327	880139 (1.35)	380280 (1.73)	720235 (1.76)	189920 (1.82)	030351 (1.83)
470328	480348 (0.55)	340116 (0.64)	349909 (0.68)	470173 (0.73)	600346 (0.84)
010350	700134 (0.49)	700322 (0.73)	189920 (0.92)	709919 (0.93)	880139 (1.20)
189920	700322 (0.70)	709919 (0.87)	010350 (0.92)	290269 (0.95)	880139 (0.98)
299936	290269 (0.52)	380280 (1.02)	880139 (1.07)	740132 (1.09)	600287 (1.10)
380280	290269 (0.83)	480348 (0.95)	470328 (0.98)	340116 (0.98)	470173 (0.99)
480348	470173 (0.43)	600346 (0.52)	470328 (0.55)	490244 (0.65)	349909 (0.67)
510316	600346 (0.90)	490244 (0.97)	470173 (1.04)	480348 (1.10)	349909 (1.15)
530218	470173 (1.11)	610152 (1.14)	600346 (1.24)	480348 (1.24)	740047 (1.29)
700134	010350 (0.49)	709919 (0.64)	700322 (0.88)	189920 (1.03)	720235 (1.31)
709919	700134 (0.64)	700322 (0.71)	720235 (0.83)	189920 (0.87)	880139 (0.88)
030351	880139 (1.13)	700322 (1.22)	720235 (1.27)	010350 (1.33)	709919 (1.39)

Table B.8 Assignment Results for Rural TTMSs in Double Peak Group with Full Variable Set

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
010014	940195 (0.44)	090229 (0.46)	940144 (0.54)	020324 (0.56)	040271 (0.60)
040068	920065 (0.32)	040271 (0.35)	070039 (0.41)	020324 (0.43)	130146 (0.43)
050272	710233 (0.52)	770299 (0.53)	030143 (0.56)	120273 (0.56)	260043 (0.59)
100276	750104 (0.34)	930140 (0.41)	030143 (0.47)	890289 (0.47)	770299 (0.55)
110136	110262 (0.32)	020324 (0.33)	040271 (0.38)	160319 (0.44)	270232 (0.45)
130146	750336 (0.19)	799925 (0.25)	070039 (0.26)	030143 (0.29)	920065 (0.32)
140079	030143 (0.42)	070039 (0.45)	750104 (0.48)	160319 (0.50)	920065 (0.55)
160230	260043 (0.54)	110262 (0.56)	710233 (0.59)	160319 (0.62)	110136 (0.64)
160319	280018 (0.25)	920065 (0.27)	070039 (0.30)	020324 (0.32)	550349 (0.33)
260043	710233 (0.35)	270232 (0.35)	110262 (0.36)	280073 (0.37)	540245 (0.39)
280018	550211 (0.18)	760240 (0.18)	550349 (0.19)	340278 (0.21)	500220 (0.21)
280073	270232 (0.19)	540245 (0.20)	290037 (0.23)	560301 (0.23)	280018 (0.24)
300234	320277 (0.15)	290297 (0.16)	370241 (0.18)	330237 (0.21)	370242 (0.25)
320277	370241 (0.15)	300234 (0.15)	290297 (0.18)	330237 (0.21)	290037 (0.22)
330149	260231 (0.19)	340278 (0.19)	370242 (0.22)	280018 (0.29)	530247 (0.29)
360118	550349 (0.20)	260231 (0.21)	280018 (0.22)	550211 (0.22)	340278 (0.23)
370242	290297 (0.14)	370241 (0.16)	290037 (0.21)	330149 (0.22)	300234 (0.25)
460192	579942 (0.17)	610254 (0.18)	760240 (0.20)	590296 (0.20)	530247 (0.22)
470337	580285 (0.16)	529939 (0.17)	560301 (0.17)	480243 (0.18)	600051 (0.19)
479944	560301 (0.15)	480243 (0.18)	470337 (0.19)	539943 (0.2)	529939 (0.20)
480048	700223 (0.20)	270232 (0.23)	609938 (0.34)	280073 (0.34)	790170 (0.34)
529939	560301 (0.13)	470337 (0.17)	479944 (0.20)	550210 (0.23)	500281 (0.25)
530247	540312 (0.19)	610254 (0.20)	500220 (0.21)	550349 (0.22)	460192 (0.22)
540245	539943 (0.18)	470337 (0.19)	540312 (0.20)	280073 (0.20)	479944 (0.21)
540312	500220 (0.14)	530247 (0.19)	550349 (0.20)	540245 (0.20)	500054 (0.22)
550210	500281 (0.15)	290037 (0.18)	539943 (0.20)	590252 (0.20)	479944 (0.21)
550211	550349 (0.09)	590296 (0.17)	500220 (0.17)	280018 (0.18)	760240 (0.22)
550349	550211 (0.09)	500220 (0.18)	280018 (0.19)	540312 (0.2)	360118 (0.2)
570122	600051 (0.12)	570219 (0.14)	609938 (0.16)	580285 (0.17)	610253 (0.17)
570219	609938 (0.07)	570122 (0.14)	600051 (0.15)	610253 (0.19)	580285 (0.21)
580285	600051 (0.15)	470337 (0.16)	580330 (0.16)	570122 (0.17)	610253 (0.17)
590252	580285 (0.18)	550210 (0.20)	500281 (0.22)	600051 (0.23)	570122 (0.23)
600051	570122 (0.12)	610253 (0.13)	580285 (0.15)	570219 (0.15)	609938 (0.18)
700223	480048 (0.20)	799925 (0.30)	110262 (0.37)	270232 (0.37)	750336 (0.38)
710233	120273 (0.25)	270232 (0.25)	130146 (0.33)	110262 (0.34)	260043 (0.35)
730263	020324 (0.57)	460192 (0.62)	760240 (0.62)	510313 (0.62)	280018 (0.64)
750104	030143 (0.23)	770299 (0.34)	100276 (0.34)	130146 (0.35)	750336 (0.36)
750336	130146 (0.19)	799925 (0.20)	920065 (0.24)	070039 (0.31)	160319 (0.35)
760240	590296 (0.17)	280018 (0.18)	610254 (0.19)	460192 (0.20)	579942 (0.21)
770299	750104 (0.34)	710233 (0.43)	030143 (0.48)	120273 (0.52)	750336 (0.52)
790170	480048 (0.34)	270232 (0.38)	510313 (0.40)	700223 (0.41)	570219 (0.41)
799925	920065 (0.18)	750336 (0.20)	130146 (0.25)	700223 (0.30)	480048 (0.35)
880291	940144 (0.59)	010014 (0.70)	940195 (0.71)	040271 (0.76)	090229 (0.78)
890289	940195 (0.41)	100276 (0.47)	070039 (0.56)	750104 (0.57)	930140 (0.58)

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
920065	799925 (0.18)	750336 (0.24)	160319 (0.27)	760240 (0.31)	280018 (0.32)
930140	100276 (0.41)	930268 (0.48)	750104 (0.49)	030143 (0.57)	890289 (0.58)
930268	939935 (0.36)	930140 (0.48)	030143 (0.65)	100276 (0.66)	750104 (0.67)
939935	930268 (0.36)	930140 (0.78)	750104 (0.92)	030143 (0.92)	100276 (0.94)
940144	940195 (0.32)	010014 (0.54)	880291 (0.59)	090229 (0.61)	040271 (0.66)
940195	940144 (0.32)	890289 (0.41)	010014 (0.44)	070039 (0.53)	020324 (0.63)
480243	470337 (0.18)	479944 (0.18)	600051 (0.19)	580285 (0.2)	610253 (0.22)
610254	579942 (0.17)	460192 (0.18)	760240 (0.19)	590296 (0.19)	530247 (0.20)
340278	260231 (0.16)	330149 (0.19)	280018 (0.21)	360118 (0.23)	370241 (0.25)
500281	550210 (0.15)	539943 (0.18)	560301 (0.20)	580285 (0.22)	290037 (0.22)
590296	760240 (0.17)	550211 (0.17)	579942 (0.17)	610254 (0.19)	460192 (0.20)
510313	570219 (0.25)	609938 (0.28)	470337 (0.30)	600051 (0.30)	570122 (0.31)
469907	460192 (0.32)	579942 (0.33)	590296 (0.34)	610254 (0.35)	580251 (0.36)
579942	610254 (0.17)	460192 (0.17)	590296 (0.17)	500220 (0.20)	760240 (0.21)
030143	750104 (0.23)	070039 (0.28)	130146 (0.29)	710233 (0.36)	750336 (0.37)
560301	529939 (0.13)	479944 (0.15)	470337 (0.17)	500281 (0.20)	580285 (0.22)
270232	280073 (0.19)	480048 (0.23)	580285 (0.24)	710233 (0.25)	590252 (0.27)
330237	320202 (0.16)	320277 (0.21)	300234 (0.21)	370241 (0.22)	290037 (0.26)
720236	480048 (0.60)	790170 (0.62)	510313 (0.67)	110136 (0.70)	040271 (0.74)
040271	110262 (0.29)	040068 (0.35)	480048 (0.35)	110136 (0.38)	270232 (0.39)
070039	130146 (0.26)	030143 (0.28)	160319 (0.30)	750336 (0.31)	920065 (0.35)
090229	010014 (0.46)	110136 (0.52)	040271 (0.56)	940144 (0.61)	020324 (0.64)
110262	270232 (0.29)	040271 (0.29)	110136 (0.32)	710233 (0.34)	480048 (0.35)
120273	710233 (0.25)	130146 (0.35)	110262 (0.37)	270232 (0.37)	480048 (0.42)
260231	340278 (0.16)	330149 (0.19)	360118 (0.21)	550349 (0.22)	550211 (0.24)
290037	370241 (0.13)	290297 (0.13)	550210 (0.18)	539943 (0.19)	370242 (0.21)
290297	370241 (0.06)	290037 (0.13)	370242 (0.14)	300234 (0.16)	320277 (0.18)
320202	330237 (0.16)	320277 (0.24)	370241 (0.26)	300234 (0.26)	290037 (0.30)
350279	320202 (0.35)	370242 (0.39)	320277 (0.41)	370241 (0.42)	330237 (0.43)
370241	290297 (0.06)	290037 (0.13)	320277 (0.15)	370242 (0.16)	300234 (0.18)
500054	500220 (0.15)	540312 (0.22)	530247 (0.26)	610254 (0.26)	360118 (0.28)
539943	540245 (0.18)	500281 (0.18)	580285 (0.19)	290037 (0.19)	479944 (0.20)
580251	600051 (0.26)	480243 (0.29)	570122 (0.33)	479944 (0.35)	580285 (0.35)
580330	580285 (0.16)	600051 (0.23)	610253 (0.24)	590252 (0.24)	570122 (0.26)
599946	580251 (0.48)	600051 (0.49)	470337 (0.50)	480243 (0.50)	540245 (0.51)
610253	600051 (0.13)	609938 (0.15)	570122 (0.17)	580285 (0.17)	570219 (0.19)
020324	160319 (0.32)	110136 (0.33)	280018 (0.36)	920065 (0.38)	110262 (0.40)
500220	540312 (0.14)	500054 (0.15)	550211 (0.17)	550349 (0.18)	579942 (0.20)
609938	570219 (0.07)	610253 (0.15)	570122 (0.16)	600051 (0.18)	580285 (0.23)

Table B.9 Assignment Results for Rural TTMSs in Single Peak Group with Reduced Variable Set

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
030270	030351 (2.15)	700134 (2.18)	530218 (2.19)	010350 (2.23)	340239 (2.24)
290269	370238 (0.16)	740132 (0.21)	299936 (0.27)	610152 (0.28)	600287 (0.41)
320112	370238 (0.34)	610152 (0.36)	290269 (0.42)	530248 (0.44)	530050 (0.54)
340116	349909 (0.13)	470328 (0.26)	380280 (0.32)	740047 (0.35)	480348 (0.36)
340239	349909 (0.42)	340116 (0.54)	490244 (0.56)	480348 (0.60)	470328 (0.60)
349909	340116 (0.13)	480348 (0.29)	470328 (0.31)	470173 (0.31)	490244 (0.40)
470173	480348 (0.22)	600346 (0.28)	470328 (0.30)	349909 (0.31)	610152 (0.34)
490060	470328 (0.31)	490244 (0.34)	480348 (0.37)	340116 (0.55)	349909 (0.56)
490244	480348 (0.23)	470328 (0.26)	490060 (0.34)	470173 (0.38)	349909 (0.40)
530050	530248 (0.23)	320112 (0.54)	470173 (0.58)	480348 (0.63)	470328 (0.65)
530248	530050 (0.23)	320112 (0.44)	610152 (0.60)	470173 (0.61)	470328 (0.62)
600346	470173 (0.28)	480348 (0.31)	470328 (0.37)	490244 (0.40)	740047 (0.46)
610152	370238 (0.12)	290269 (0.28)	600287 (0.30)	470173 (0.34)	320112 (0.36)
740047	340116 (0.35)	380280 (0.45)	470328 (0.45)	600346 (0.46)	470173 (0.46)
740132	290269 (0.21)	979931 (0.34)	370238 (0.35)	299936 (0.35)	189920 (0.37)
880139	720235 (0.44)	189920 (0.52)	740132 (0.56)	709919 (0.62)	979931 (0.63)
979931	740132 (0.34)	720235 (0.35)	189920 (0.37)	370238 (0.38)	470173 (0.43)
720235	189920 (0.25)	709919 (0.26)	979931 (0.35)	700322 (0.36)	740132 (0.41)
370238	610152 (0.12)	290269 (0.16)	320112 (0.34)	740132 (0.35)	600287 (0.36)
600287	610152 (0.30)	370238 (0.36)	290269 (0.41)	380280 (0.41)	470173 (0.57)
700322	189920 (0.23)	709919 (0.26)	720235 (0.36)	010350 (0.41)	979931 (0.43)
090327	880139 (0.92)	340116 (1.19)	380280 (1.21)	189920 (1.23)	720235 (1.28)
470328	480348 (0.12)	490244 (0.26)	340116 (0.26)	470173 (0.30)	490060 (0.31)
010350	700134 (0.25)	700322 (0.41)	709919 (0.57)	189920 (0.58)	979931 (0.60)
189920	700322 (0.23)	720235 (0.25)	740132 (0.37)	979931 (0.37)	709919 (0.38)
299936	290269 (0.27)	740132 (0.35)	370238 (0.41)	380280 (0.5)	610152 (0.52)
380280	340116 (0.32)	470328 (0.35)	370238 (0.37)	610152 (0.38)	480348 (0.40)
480348	470328 (0.12)	470173 (0.22)	490244 (0.23)	349909 (0.29)	600346 (0.31)
510316	490244 (0.51)	340239 (0.61)	349909 (0.62)	600346 (0.69)	470328 (0.73)
530218	610152 (0.84)	470173 (0.89)	600346 (0.91)	370238 (0.92)	340239 (0.94)
700134	010350 (0.25)	709919 (0.46)	700322 (0.49)	720235 (0.65)	189920 (0.66)
709919	720235 (0.26)	700322 (0.26)	189920 (0.38)	700134 (0.46)	740132 (0.52)
030351	720235 (0.57)	189920 (0.66)	979931 (0.67)	880139 (0.70)	010350 (0.77)

Table B.10 Assignment Results for Rural TTMSs in Double Peak Group with Reduced Variable Set

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
010014	090229 (0.48)	940195 (0.6)	940144 (0.66)	040068 (0.73)	020324 (0.83)
040068	920065 (0.40)	040271 (0.42)	799925 (0.55)	160319 (0.56)	020324 (0.60)
050272	770299 (0.55)	120273 (0.68)	030143 (0.76)	160230 (0.76)	710233 (0.77)
100276	750104 (0.41)	890289 (0.44)	930140 (0.56)	770299 (0.61)	030143 (0.64)
110136	720236 (0.28)	110262 (0.35)	020324 (0.41)	040271 (0.45)	160230 (0.45)
130146	750336 (0.15)	799925 (0.27)	070039 (0.29)	160319 (0.35)	920065 (0.38)
140079	030143 (0.37)	750104 (0.54)	070039 (0.55)	160319 (0.59)	750336 (0.62)
160230	260043 (0.27)	710233 (0.3)	110262 (0.31)	160319 (0.40)	700223 (0.41)
160319	920065 (0.21)	750336 (0.23)	799925 (0.29)	130146 (0.35)	260043 (0.36)
260043	710233 (0.25)	160230 (0.27)	110262 (0.28)	700223 (0.29)	270232 (0.35)
280018	760240 (0.17)	340278 (0.18)	550211 (0.19)	550349 (0.20)	360118 (0.20)
280073	540245 (0.20)	290037 (0.21)	270232 (0.21)	550210 (0.25)	479944 (0.25)
300234	290297 (0.17)	320277 (0.25)	370241 (0.28)	330237 (0.30)	350279 (0.32)
320277	370241 (0.07)	290037 (0.18)	320202 (0.19)	350279 (0.19)	550210 (0.22)
330149	290297 (0.29)	260231 (0.30)	460192 (0.32)	340278 (0.32)	550349 (0.33)
360118	550349 (0.18)	550211 (0.19)	340278 (0.20)	280018 (0.20)	260231 (0.24)
370242	370241 (0.18)	540245 (0.2)	290037 (0.21)	350279 (0.21)	529939 (0.22)
460192	469907 (0.19)	579942 (0.23)	530247 (0.25)	760240 (0.31)	610254 (0.31)
470337	560301 (0.07)	529939 (0.08)	480243 (0.14)	479944 (0.16)	599946 (0.18)
479944	560301 (0.10)	600051 (0.14)	539943 (0.15)	470337 (0.16)	480243 (0.17)
480048	700223 (0.16)	270232 (0.23)	790170 (0.25)	110262 (0.37)	720236 (0.37)
529939	470337 (0.08)	560301 (0.10)	480243 (0.14)	540245 (0.17)	479944 (0.19)
530247	579942 (0.15)	540312 (0.17)	610254 (0.2)	500220 (0.22)	500054 (0.23)
540245	529939 (0.17)	470337 (0.19)	290037 (0.2)	370242 (0.20)	280073 (0.20)
540312	530247 (0.17)	550349 (0.17)	500220 (0.18)	260231 (0.22)	500054 (0.23)
550210	290037 (0.10)	539943 (0.13)	500281 (0.16)	580251 (0.18)	480243 (0.18)
550211	550349 (0.08)	500220 (0.19)	280018 (0.19)	360118 (0.19)	760240 (0.20)
550349	550211 (0.08)	260231 (0.17)	540312 (0.17)	360118 (0.18)	500220 (0.19)
570122	570219 (0.08)	600051 (0.11)	480243 (0.12)	580285 (0.12)	539943 (0.16)
570219	570122 (0.08)	600051 (0.12)	480243 (0.13)	580251 (0.17)	580285 (0.18)
580285	500281 (0.11)	570122 (0.12)	539943 (0.13)	480243 (0.14)	600051 (0.18)
590252	580251 (0.17)	580285 (0.18)	550210 (0.18)	290037 (0.20)	539943 (0.21)
600051	480243 (0.09)	570122 (0.11)	570219 (0.12)	479944 (0.14)	560301 (0.15)
700223	480048 (0.16)	110262 (0.25)	799925 (0.28)	790170 (0.29)	260043 (0.29)
710233	260043 (0.25)	160230 (0.30)	120273 (0.30)	270232 (0.36)	700223 (0.36)
730263	720236 (0.47)	020324 (0.49)	760240 (0.59)	280018 (0.59)	110136 (0.66)
750104	030143 (0.26)	770299 (0.30)	100276 (0.41)	750336 (0.53)	130146 (0.54)
750336	130146 (0.15)	799925 (0.22)	160319 (0.23)	920065 (0.27)	070039 (0.29)
760240	280018 (0.17)	550211 (0.2)	469907 (0.25)	590296 (0.25)	579942 (0.26)
770299	750104 (0.30)	710233 (0.42)	030143 (0.49)	050272 (0.55)	160230 (0.56)
790170	270232 (0.19)	480048 (0.25)	700223 (0.29)	280073 (0.36)	570219 (0.37)
799925	920065 (0.19)	750336 (0.22)	130146 (0.27)	700223 (0.28)	160319 (0.29)
880291	940144 (0.79)	010014 (0.89)	720236 (0.94)	090229 (0.96)	940195 (0.98)
890289	100276 (0.44)	940195 (0.6)	750104 (0.72)	930140 (0.76)	070039 (0.78)

Test Sites	Best Matching Sites				
	1st	2nd	3rd	4th	5th
920065	799925 (0.19)	160319 (0.21)	750336 (0.27)	280018 (0.37)	130146 (0.38)
930140	100276 (0.56)	930268 (0.58)	750104 (0.6)	890289 (0.76)	030143 (0.77)
930268	930140 (0.58)	939935 (0.64)	100276 (1.02)	750104 (1.04)	030143 (1.11)
939935	930268 (0.64)	930140 (1.13)	100276 (1.50)	750104 (1.52)	030143 (1.59)
940144	940195 (0.36)	010014 (0.66)	090229 (0.69)	880291 (0.79)	040271 (0.85)
940195	940144 (0.36)	010014 (0.60)	890289 (0.6)	070039 (0.66)	030143 (0.85)
480243	600051 (0.09)	560301 (0.11)	570122 (0.12)	570219 (0.13)	529939 (0.14)
610254	579942 (0.16)	500054 (0.16)	590296 (0.19)	469907 (0.20)	530247 (0.20)
340278	280018 (0.18)	360118 (0.20)	260231 (0.22)	550211 (0.23)	500220 (0.26)
500281	539943 (0.09)	580285 (0.11)	550210 (0.16)	290037 (0.18)	570122 (0.22)
590296	579942 (0.18)	610254 (0.19)	550211 (0.20)	469907 (0.21)	500220 (0.21)
510313	470337 (0.24)	599946 (0.26)	529939 (0.30)	600051 (0.30)	610253 (0.30)
469907	460192 (0.19)	579942 (0.19)	610254 (0.2)	590296 (0.21)	500220 (0.22)
579942	530247 (0.15)	610254 (0.16)	590296 (0.18)	469907 (0.19)	500220 (0.21)
030143	750104 (0.26)	140079 (0.37)	130146 (0.41)	070039 (0.41)	750336 (0.45)
560301	470337 (0.07)	479944 (0.10)	529939 (0.10)	480243 (0.11)	539943 (0.15)
270232	790170 (0.19)	280073 (0.21)	480048 (0.23)	290037 (0.30)	700223 (0.31)
330237	300234 (0.3)	290297 (0.35)	320277 (0.37)	320202 (0.39)	370241 (0.40)
720236	110136 (0.28)	480048 (0.37)	700223 (0.44)	790170 (0.46)	730263 (0.47)
040271	110262 (0.41)	040068 (0.42)	700223 (0.43)	110136 (0.45)	480048 (0.45)
070039	750336 (0.29)	130146 (0.29)	160319 (0.37)	030143 (0.41)	920065 (0.48)
090229	010014 (0.48)	110136 (0.68)	940144 (0.69)	040271 (0.72)	940195 (0.93)
110262	700223 (0.25)	260043 (0.28)	160230 (0.31)	110136 (0.35)	480048 (0.37)
120273	710233 (0.30)	130146 (0.42)	260043 (0.45)	750336 (0.45)	160230 (0.46)
260231	550349 (0.17)	550211 (0.22)	340278 (0.22)	540312 (0.22)	360118 (0.24)
290037	550210 (0.10)	370241 (0.12)	539943 (0.17)	500281 (0.18)	320277 (0.18)
290297	300234 (0.17)	370241 (0.18)	320277 (0.23)	290037 (0.26)	370242 (0.26)
320202	320277 (0.19)	370241 (0.21)	500281 (0.23)	350279 (0.25)	290037 (0.25)
350279	320277 (0.19)	370241 (0.21)	370242 (0.21)	290037 (0.24)	500281 (0.24)
370241	320277 (0.07)	290037 (0.12)	370242 (0.18)	290297 (0.18)	550210 (0.20)
500054	500220 (0.09)	610254 (0.16)	540312 (0.23)	530247 (0.23)	550349 (0.25)
539943	500281 (0.09)	550210 (0.13)	580285 (0.13)	560301 (0.15)	480243 (0.15)
580251	560301 (0.16)	600051 (0.16)	570219 (0.17)	590252 (0.17)	599946 (0.17)
580330	580285 (0.24)	500281 (0.25)	590252 (0.27)	539943 (0.27)	570122 (0.29)
599946	580251 (0.17)	479944 (0.18)	470337 (0.18)	540245 (0.2)	560301 (0.21)
610253	609938 (0.17)	600051 (0.17)	570122 (0.19)	480243 (0.21)	570219 (0.25)
020324	110136 (0.41)	160319 (0.49)	730263 (0.49)	110262 (0.52)	920065 (0.54)
500220	500054 (0.09)	540312 (0.18)	550349 (0.19)	550211 (0.19)	590296 (0.21)
609938	610253 (0.17)	600051 (0.24)	570219 (0.24)	570122 (0.29)	480243 (0.32)

VITA  
SHANSHAN YANG

EDUCATION

- 1997 – 2001    B.S., Civil Engineering  
                    Tsinghua University  
                    Beijing, China
- 2001 – 2004    M.S., Transportation Engineering  
                    Tsinghua University  
                    Beijing, China
- 2005 – 2012    Doctoral Candidate in Civil Engineering  
                    Florida International University  
                    Miami, Florida

EMPLOYMENT

- 2005– 2012    Graduate Teaching/Research Assistant  
                    Lehman Center for Transportation Research  
                    Department of Civil and Environmental Engineering  
                    Florida International University, Miami, Florida

AFFILIATIONS

- Student member, Women in Transportation, 2011 – Present
- Student member, Chi Epsilon Civil Engineering Honor Society, 2006 – Present
- Student member, Institute of Transportation Engineers, 2005 – Present

PUBLICATIONS AND PRESENTATIONS

- C. Lu, S. Yang, and F. Zhao. “Seasonal Factor Assignment Based on the Similarity of Hourly Traffic Patterns and Influential Variables,” Proceedings of the 91st Annual Meeting of the Transportation Research Board, National Research Council, Washington, D.C., January 2012.
- S. Yang, C. Lu, F. Zhao, R.L. Reel, and Doug O’Hara. “Estimation for Seasonal Factors of Similarity-Based Traffic for Urban Roads in Florida,” Transportation Research Record 2121, Journal of the Transportation Research Board, National Research Council, Washington, D.C., 2009, pp. 74-80.
- F. Zhao, K. Xing, S. Yang, C. Lu, and S. Chung, “Hurricane Evacuation Planning for Disadvantaged Population,” Final Report FTA-FL-04-7104-2010.04, Prepared for the Federal Transit Administration, December 2009.



- F. Zhao, S. Yang, S. Chung, and N. Cartya, "Transportation Impact Analysis - 107th Avenue Expansion," City of Sweetwater, Florida, Prepared for the Florida Department of Transportation, District 6, Miami, FL, January 2009.
- C. Lu, F. Zhao, and S. Yang. "Exploitation of Hourly Traffic Patterns for Modeling Seasonal Factors," Proceedings of the 10th International Conference on Application of Advanced Technologies in Transportation, ASCE, Athens, Greece, May 27-31, 2008.
- F. Zhao, S. Yang, and C. Lu, "Alternatives for Estimating Seasonal Factors on Rural and Urban Roads in Florida, Phase 2," Final Report, Prepared for the Florida Department of Transportation, Tallahassee, FL, February 2008.
- H. Lu, S. Yang, and P. Jiao, "Transfer Analysis of Bicycle to Public Transit." Proceedings of the "Velo-city 2003" International Conference, Paris, France, 2003.
- S. Yang, H. Lu, and P. Jiao, "Study on Models and Algorithms of Tourism Demand and Environmental Capacity." Proceedings of 5th Conference of the Eastern Asia Society for Transportation Studies, Fukuoka, Japan, 2003.

#### HONORS AND AWARDS

- Winner of ITE Gold Coast Chapter Book Scholarship Essay Competition, Spring 2011.
- Certificate of Appreciation from Miami-Dade Public School System Community Traffic Safety Committee, for volunteer work in Safe Route to School Project, 2009, 2010.