An-Najah National University
Faculty of Graduate Studies

Developing Trip Generation Models Using Adaptive Neuro-Fuzzy Inference System: Salfit City as a Case Study

By

Mohammad Kamal Mohammad Irshaid

Supervisor

Prof. Sameer Abu-Eisheh

This Thesis is Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Roads and Transportation Engineering, Faculty of Graduate Studies, An-Najah National University, Nablus, Palestine.

2019

Developing Trip Generation Models Using Adaptive Neuro-Fuzzy Inference System: Salfit City as a Case Study

By

Mohammad Kamal Mohammad Irshaid

This thesis was defended successfully on 23/12/2019, and approved by:

Defense Committee Members:	**Signature**
- Prof. Sameer Abu-Eisheh/ Supervisor	……………
- Dr. Yahya Sarraj/ External Examiner	……………
- Dr. Khaled Al-Sahili/ Internal Examiner	……………

Dedication

This thesis is dedicated with special thanks to all whom provided me with their support and encouragement to achieve this work.

Acknowledgment

Special thanks to Prof. Sameer Abu-Eisheh for the continuous support accomplishing this thesis; for his motivation, and encouragement.

I would like also to express my gratitude and deep appreciation to Eng. Ahmad Amer, who provided the required dataset needed to achieve this work.

الاقرار

أنا الموقع أدناه مقدم الرسالة التي تحمل العنوان:

Developing Trip Generation Models Using Adaptive Neuro-Fuzzy Inference System: Salfit City as a Case Study

تطوير نماذج تولد الرحلات باستخدام نظام الاستدلال الضبابي المتكيف: مدينة سلفيت كحالة دراسية

أقر بأن ما اشـتملت عليه هذه الرسـالة إنما هي نتاج جهدي الخاص، باسـتثناء ما تمت الإشـارة إليه حيثما ورد، وإن هذه الرســـالة ككل، أو أي جزء منها لم يقدم من قبل لنيل أية درجة عملية أو بحث علمي لدى أي مؤسسة تعليمية أو بحثية أخرى.

Declaration

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted elsewhere for any other degree or qualification.

Student's name:	**اسم الطالب:**
Signature:	**التوقيع:**
Date:	**التاريخ:**

Table of Contents

Dedication III

Acknowledgment IV

Declaration V

Table of Contents VI

List of Tables X

List of Figures XII

List of Abbreviations XIV

Abstract XV

Chapter One: Introduction 2

1.1 Background 2

1.1.1 General 2

1.1.2 Trip Generation 3

1.1.3 Modeling of Trip Generation 4

1.2 Problem Definition 6

1.3 Objectives 8

1.4 Study Area 10

1.5 Thesis Structure 10

Chapter Two: Theoretical Background 13

2.1 Introduction 13

2.2 Artificial Neural Network (ANN) 13

2.2.1 Artificial Neuron Concept 14

2.2.2 ANN Architecture 16

2.2.3 Learning Process 17
2.3 Fuzzy Inference System (FIS) 17
2.3.1 Fuzzy Set and Fuzzy Logic 18
2.3.2 Fuzzy Inference Process 20
2.3.3 Types of Fuzzy Inference System 24
2.4 Adaptive Neuro-Fuzzy Inference System (ANFIS) 25
2.4.1 ANFIS Architecture 26
2.4.2 ANFIS Learning Algorithm 29
2.5 Multiple Linear Regression Analysis (MLR) 30
Chapter Three: Literature Review 34
3.1 Introduction 34
3.2 Application of ANN in Transportation and Trip Generation 35
3.3 Application of FIS in Transportation and Trip Generation 36
3.4 Application of ANFIS in Transportation 38
3.5 Application of ANFIS in Trip Generation 39
3.6 Trip Generation Models in Palestine 41
3.7 Summary 42
Chapter Four: Methodology 45
4.1 Introduction 45
4.2 Data Collection 45
4.3 Developing Trip Generation Models Using Regression Approach . 47
4.3.1 Linear Regression Model-Building Procedures 48
4.3.2 Statistical Tests 49
4.4 Developing Trip Generation Models Using ANFIS Approach 57
4.4.1 Loading Input-Output Training Dataset 58

4.4.2 Generating Initial Fuzzy Inference System (FIS) 58

4.4.3 Optimizing the FIS Using Training Algorithms 60

4.5 Models Comparison and Validation 61

Chapter Five: Data Collection 64

5.1 Introduction 64

5.2 Study Area and Zoning 64

5.3 Selecting Sample Size 67

5.4 Collecting Required Data 68

Chapter Six: Models Development and Comparison 73

6.1 Introduction 73

6.2 General Trip Generation Model (ALLTRIP) 73

6.2.1 Developing ALLTRIP Model Using MLR Approach 74

6.2.2 Developing ALLTRIP Model Using ANFIS Approach 77

6.2.3 Models Comparison and Validation 82

6.3 Home-Based Work Trip Generation Model (HBW) 85

6.3.1 Developing HBW Model Using Multiple Linear Regression ... 86

6.3.2 Developing HBW Model Using ANFIS 88

6.3.3 Models Comparison and Validation 92

6.4 Home-Based Education Trip Generation Model (HBE) 95

6.4.1 Developing HBE Model Using Multiple Linear Regression 95

6.4.2 Developing HBE Model Using ANFIS 96

6.4.3 Models Comparison and Validation 99

6.5 Home-Based Other Trips Generation Model (HBO) 102

6.5.1 Developing HBO Model Using Multiple Linear Regression .. 102

6.5.2 Developing HBO Model Using ANFIS Approach 104

6.5.3 Models Comparison and Validation 105

6.6 Summary 110

Chapter Seven: Summary and Conclusions 116

7.1 Summary 116

7.2 Conclusions 118

7.3 Recommendations 122

References 125

Appendices 132

Appendix A 133

Appendix B 134

الملخص .... ب

List of Tables

Table 4.1: ANOVA Table for Testing Model Overall Significance........... 54

Table 5.1: Traffic Analysis Zones Along with The Sample Size 67

Table 5.2: The Selected Explanatory (Independent) Variables 69

Table 5.3: Descriptive Statistics for the Selected Explanatory Variables .. 70

Table 5.4: The Selected Dependent Variables... 71

Table 5.5: Descriptive Statistics for the Selected Dependent Variables..... 71

Table 6.1: Regression Results for the ALLTRIP Model 74

Table 6.2: ANOVA Table for the ALLTRIP Model 77

Table 6.3: RMSE & MAE for ANFIS Configurations - ALLTRIP 78

Table 6.4: ANFIS Optimum Configuration Summary for ALLTRIP 79

Table 6.5: ALLTRIP Performance Comparison - MLR vs. ANFIS........... 83

Table 6.6: Actual and Predicted ALLTRIP Descriptive Statistics 85

Table 6.7: Regression Results for HBW Trip Generation Model............... 86

Table 6.8: ANOVA Table for HBW Trip Generation Model..................... 88

Table 6.9: RMSE & MAE for ANFIS Configurations - HBW 89

Table 6.10: ANFIS Optimum Configuration Summary for HBW 90

Table 6.11: HBW Performance Comparison - MLR vs. ANFIS................ 93

Table 6.12: Actual and Predicted HBW Descriptive Statistics 94

Table 6.13: Regression Results for HBE Trip Generation Model.............. 95

Table 6.14: ANOVA Table for HBE Trip Generation Model.................... 96

Table 6.15: RMSE & MAE for ANFIS Configurations - HBE.................. 97

Table 6.16: ANFIS Optimum Configuration Summary for HBE............... 98

Table 6.17: HBE Performance Comparison - MLR vs. ANFIS 100

Table 6.18: Actual and Predicted HBE Descriptive Statistics 101

Table 6.19: Regression Results for HBO Trips Generation Model 102

Table 6.20: ANOVA Table for HBO Trips Generation Model 104

Table 6.21: RMSE & MAE for ANFIS Configurations - HBO 104

Table 6.22: ANFIS Optimum Configuration Summary for HBO 105

Table 6.23: HBO Performance Comparison Among MLR and ANFIS ... 108

Table 6.24: Actual and Predicted HBO Descriptive Statistics 109

Table 6.25: Estimated Models Summary Considering MLR & ANFIS ... 111

List of Figures

Figure 2.1: Mathematical Modeling of Artificial Neuron........................... 15

Figure 2.2: Common Types of Neuron Activation Functions 15

Figure 2.3: Input Layer, Output Layer, and Hidden Layer of an ANN...... 16

Figure 2.4: Crisp Set verses Fuzzy Set Concept 19

Figure 2.5: Triangle, Trapezoidal, and Gaussian Membership Functions.. 20

Figure 2.6: Fuzzy Inference System 21

Figure 2.7: Fuzzy Inference Process Diagram 23

Figure 2.8: Takagi–Sugeno Fuzzy Model 27

Figure 2.9: The Equivalent ANFIS Architecture for Sugeno FIS 28

Figure 4.1: The Four Main Steps of the Study 46

Figure 5.1: Study Area Boundaries and Traffic Analysis Zones 66

Figure 6.1: The Optimized FIS for Estimating ALLTRIP 80

Figure 6.2: The Equivalent ANFIS Architecture for ALLTRIP Model 80

Figure 6.3: Optimized Gaussian MFs for EMP Input Variable 81

Figure 6.4: Optimized Gaussian MFs for EDU Input Variable 82

Figure 6.5: Optimized Gaussian MFs for DRIVE Input Variable 83

Figure 6.6: The Optimized FIS for Estimating HBW 90

Figure 6.7: The Equivalent ANFIS Architecture for HBW Model 91

Figure 6.8: Optimized Gaussian MFs for EMP Input Variable 91

Figure 6.9: Optimized Gaussian MFs for AGE_B Input Variable 92

Figure 6.10: Optimized Gaussian MFs for CAR Input Variable 92

Figure 6.11: The Optimized FIS for HBE Using Gaussian MFs 98

Figure 6.12: The Optimized FIS for HBE Using Trapezoidal MFs 98

Figure 6.13: The Equivalent ANFIS Architecture for HBE Model............ 99

Figure 6.14: Optimized Gaussian MFs for EDU Input Variable................ 99

Figure 6.15: Optimized Trapezoidal MFs for EDU Input Variable 100

Figure 6.16: The Optimized FIS for Estimating HBO.............................. 106

Figure 6.17: The Equivalent ANFIS Architecture for HBO Model 106

Figure 6.18: Optimized Gaussian MFs for SIZE Input Variable.............. 107

Figure 6.19: Optimized Gaussian MFs for DRIVE Input Variable.......... 107

Figure 6.20: Optimized Gaussian MFs for HHTYP Input Variable......... 107

Figure 6.21: Comparison Among Resulted R-Squared for Each Model .. 113

Figure 6.22: Comparison Among Resulted RMSE for Each Model 113

Figure 6.23: Comparison Among Resulted MAE for Each Model 114

Figure 6.24: Comparison Among Mean of Actual and Predicted Trips ... 114

List of Abbreviations

AI	Artificial Intelligence
ALLTRIP	Home-Based General Trips (All Trips)
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
FIS	Fuzzy Inference System
FL	Fuzzy Logic
HBE	Home-Based Education Trips
HBO	Home-Based Other Trips
HBW	Home-Based Work Trips
HH	Household or Housing Unit
MAE	Mean Absolute Error
MF	Membership Function
MLR	Multiple Linear Regression
PCBS	Palestinian Central Bureau of Statistics
RMSE	Root Mean Squared Error
SPSS	Statistical Package for Social Sciences
TAZ	Traffic Analysis Zones
UTMS	Urban Transportation Modeling System
VIF	Variance Inflation Factor

Developing Trip Generation Models Using Adaptive Neuro-Fuzzy Inference System: Salfit City as a Case Study

By
Mohammad Kamal Mohammad Irshaid

Supervisor
Prof. Sameer Abu-Eisheh

Abstract

In Palestine, few studies that are concerned with the development of trip generation models have been conducted. The lack of specialized studies for this purpose may be related to several challenges that encounter the Palestinian situation, such as the restricted financial support and the lack of reliable data, which makes it difficult to perform such studies. These limited studies were developed using mainly the Multiple Linear Regression (MLR) approach, which sometimes would not result in appropriate models when dealing with interrelated and complex relationships among several socioeconomic variables. Therefore, this study was devoted to investigating the feasibility of using a relatively new method for data analysis called the Adaptive Neuro-Fuzzy Inference System (ANFIS), as an alternative for the traditional MLR, and explore its application within the Palestinian context for the development of the home-based trip generation models.

Through this study, four types of trip generation models were developed for the Palestinian city of Salfit; the ALLTRIP model for estimating the total number of daily home-based trips generated, and the other three models for estimating the number of trips generated based on trip purpose, which are the Home-Based Work (HBW), the Home-Based Education (HBE), and the

Home-Based Other (HBO) trips generation models. These models were estimated and validated using a sample of 309 households, that was thoroughly collected for Salfit City in 2017. Each of these models was developed using the two competing approaches; MLR and ANFIS. The better performing and more suitable approach was then determined based on several evaluation criteria, such as the higher value of R-Squared, the lower RMSE, and the much closer outputs to the actual values.

In this study, the ANFIS was able to outperform and develop more accurate models than the MLR when dealing with the ALLTRIP and the HBO, which were considered to be more complex than others, as they include wider data range, and constitute more percentage of daily trips generated. Whereas for the HBW and the HBE, both modeling approaches were performed nearly at the same level, the R-Squared values were large enough to capture most of the variations, and the differences between the performance measures were very small which could be neglected. On the other hand, there was a little advantage for the MLR in the validation process. For these two cases, the use of the MLR was considered to be sufficient.

The robust comparison through this study reveals that the ANFIS represents a promising technique, that could be a good competitor for MLR approach, especially, when dealing with interrelated and complex relationships among several socioeconomic variables. The ANFIS was found to be a useful tool for modeling home-based trip generation for Salfit City, and its further applications in transportation planning studies were recommended.

Chapter One

Introduction

Chapter One
Introduction

1.1 Background

1.1.1 General

The urban transportation system is an essential component of the urban area, which reflects its economic health and quality of life, and has furthermore a considerable effect on land accessibility, land use patterns, and mobility of people and goods.

Urban transportation systems are usually designed to accommodate the transportation activities of the urban population. Hence, and in order to provide a proper system that meets community needs, it is necessary for transportation planners to predict the current and future demand for travel, which can be achieved by deeply understanding the relationships among land use, travel behavior, and socioeconomic and demographic characteristics of the urban area.

Travel demand modeling plays a major role in planning efficient transportation systems, as it provides useful information regarding traveler preferences, and forecasts within a rational framework the current and future demand for travel. However, despite the fact that the recently evolved advanced modeling techniques, such as tour- and activity-based models, enable more realistic representations of travel behavior, their implementation in Palestine may not be feasible, mainly due to the lack of sufficient financial support, reliable data, professional expertise, and the required technical resources.

Therefore, the conventional trip-based modeling techniques are still the prevailing methods that can be used in Palestine for modeling travel behavior, mainly the four-step travel demand forecasting process, or what is known by 'Urban Transportation Modeling System (UTMS)'. This process usually transfers urban activities into a number of trips, and attempts to quantify the relationship between urban activities and travel, through modeling trip generation, trip distribution, mode choice, and traffic assignment.

The trip in this context is often defined as a one-way single journey made by an individual between two points by a specified or combined mode of travel and for a defined purpose (Ben-Edigbe & Rahman, 2010). The output of this four-step process is usually an origin-destination trip matrix, for each mode of travel, that could be converted into traffic volumes over the network links, and as a result, the current and future transportation needs and problems could be predicted, and the mitigation measures could be identified and implemented accordingly.

1.1.2 Trip Generation

Trip generation is the first step in the traditional four-step transportation planning process. It estimates the number of daily trips generated by a household, or a zone, for various activities (such as work, education, shopping, and other), by developing relationships between trip ends and socioeconomic or activity characteristics of the land use. Trip ends generated by a household can be classified as being either production or attraction, with separate prediction models for each class (Meyer & Miller, 2001).

Trip generation models basically deal with two levels of analysis; aggregated at the zonal level (such as average trips per zone), and disaggregated at the household or individual level (such as total trips per household). The disaggregated level often provides more reliable and accurate results. Although the individual is the trip maker, the number of trips per household are usually estimated and preferred, because 1) the home is the basis where most trips start and end, 2) the income and the vehicle ownership are usually shared by all members of the household, and also because 3) the family constitutes the ‘cell society’ where all basic needs are usually met (Dodeen, 2014).

There are several household characteristics that affect, and can be used for, the prediction of household trips production, mainly the household monthly income, vehicle ownership, residential density, number of persons who are receiving education and/or working, driving license holders, household type, and others.

1.1.3 Modeling of Trip Generation

The estimation procedure for trip generation usually employs mathematical models that associate each trip purpose with one or more of the above household characteristics. However, two different approaches could be considered while developing trip generation models; the traditional statistical analysis approach, or the more recently evolved computational intelligence approach.

Statistical analysis is the most widely used technique for this purpose. It has solid and accepted mathematical foundations that can provide insights on the mechanisms creating the data (Karlaftis & Vlahogianni, 2011). Although many techniques have been suggested for this approach, such as linear regression, category analysis, and count data techniques, linear regression is the best established and most popular method.

The linear regression models capture the correlation patterns and study the relationship between variables that are considered as the determinants of behavior (explanatory variables), such as the household characteristics, and variables that are considered to estimate the number of trips as indicators of travel behavior (Přibyl & Goulias, 2003). However, statistical techniques frequently fail to develop an appropriate predictive model when dealing with complex and highly nonlinear problems.

Computational intelligence (machine learning-based approach) has been recently employed, especially when the models developed by statistical techniques fail to accurately simulate a problem. However, as transportation problems often contain complex and nonlinear relations among several variables that are describing their behavior, using computational intelligence methods are appealing. The application of Artificial Intelligence (AI) methods, which combine elements of self-learning, adaptation, and self-organization, enables the effective elaboration of modeling such problems (Pamuła, 2016).

Some of the commonly used elements of AI are the fuzzy logic and the artificial neural networks, in addition to their popular combinations, the neuro-fuzzy methods (i.e. Adaptive Neuro-Fuzzy Inference System - ANFIS), which have emerged as superior means in dealing with complex and highly nonlinear phenomena.

With the superiority of the computational intelligence among other methods in modeling complex and nonlinear phenomena, it is being motivated to explore its application within the Palestinian context for modeling trip generation, as an attempt to seek for more modeling accuracy. This study is devoted to develop trip generation models for one of the Palestinian cities using the two approaches; linear regression as a conventional statistical-based approach, and ANFIS as a computational intelligence-based approach, and conducting comparative analysis among their modeling performance, by testing their accuracy and prediction capabilities.

1.2 Problem Definition

In Palestine, there is little documented experience concerning transportation planning in general, and the development of trip generation models at specific. The lack of specialized studies for this purpose may be due to the economic, social, and political challenges that encounter the Palestinian situation, which makes it difficult to perform such studies. However, the state-of-practice approach, which is commonly used for this purpose (modeling trip generation), is the Multiple Linear Regression (MLR) technique. Despite the fact that this approach could be easily constructed, estimated, and interpreted, it has several limitations and drawbacks (Meyer & Miller, 2001):

1. It does not recognize the nonnegativity and the integer nature of household trips, which are treated as continuous variables that can be negative.

2. The ‘best fit’ equations between the dependent and explanatory variables may yield unrealistic results.

3. The correlation among explanatory variables may create estimation problems.

4. The assumption that the explanatory variables have linear and additive impacts may be wrong.

5. Usually fails to develop an appropriate and accurate model when dealing with complex and highly nonlinear problem, especially when the data range is being considerably large.

6. It requires a large amount of data for estimating the model, which can be costly and economically not feasible.

Count data models and category analysis techniques were emerged to overcome some of these problems. However, the empirical benefits of count data models have not been well established, and their application experience for trip generation models seems to be limited (Lim & Srinivasan, 2011). On the other hand, the category analysis avoids the assumption of linear, additive relationship between trip generation and its explanatory variables, but in contrast it has no statistical justification, and only key specific variables can be considered.

With the complicated nature of human-based activities, and as trip generation models usually deal with, and are influenced by, many interrelated relationships among several socioeconomic factors; the conventional statistical methods which have linear structure, particularly linear regression, seem to be not suitable for all cases, inaccurate, and inadequate for modeling and predicting the nonlinear and complex behavior of the urban travel.

There is a need to explore other modeling techniques, rather than the above mentioned, that would produce more accurate results using the same dataset at hand. The relevance of this study comes from the fact that it is pioneering in considering AI in transportation in general, and in trip generation at specific, in Palestine. This study is the first in its type that considers the comparison between trip generation models developed based on statistical analysis and those considering AI.

1.3 Objectives

This study aims to introduce the concept of a relatively new method for data analysis called the Adaptive Neuro-Fuzzy Inference System (ANFIS), and explore its application in the Palestinian context for the development of trip generation models.

The specific objectives of this study can be summarized as:

1. Investigate the feasibility of using an ANFIS approach for modeling home-based trip production for one of the Palestinian urban areas (Salfit City was taken as a study area).
2. Assess the performance of ANFIS as relative to other traditional and commonly used modeling techniques, specifically the multiple linear regression, considering the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R-Squared) statistical measures.
3. Examine the validity and study the predictive capability of ANFIS as compared with regression methods, and check the accuracy of these models in reflecting the actual data.
4. Evaluate the effectiveness, in terms of accuracy (i.e., minimum error), of using different options for building the ANFIS models. Several configurations will be considered to achieve the optimum ANFIS model structure.

Four types of trip generation models were developed in this study, considering both multiple linear regression and ANFIS approaches, which are:

1. **ALLTRIP Model:** for estimating the total number of daily general trips (all trips) generated by a household.
2. **HBW Model:** for estimating the number of daily work trips generated by a household, Home-Based Work trips.
3. **HBE Model:** for estimating the number of daily education trips generated by a household, Home-Based Education trips.

4. **HBO Model:** for estimating the number of daily other trips generated by a household, i.e., daily trips not for work or education purposes, Home-Based Other trips.

1.4 Study Area

Salfit City was selected as a case for this study. It is a small Palestinian city, located in the north-west part of the West Bank, which has a total population of 10,911 persons, distributed over a total number of 2,527 households, and a total area of 23 km^2, as was reported by the latest population census conducted by the Palestinian Central Bureau of Statistics (PCBS) in 2017.

Salfit City was selected due to the availability of thoroughly collected data through the recent study conducted by Amer (2017), which was concerned with the development of household trip generation models for the city. The study also coincided with the period of the latest Palestinian census.

This study relies on the data collected by Amer (2017), between the period of early October and late November 2016 for Salfit City. Amer collected 309 samples of households, which were distributed over 6 traffic analysis zones. Each collected sample includes data regarding household socioeconomic characteristics and the number of daily trips produced.

1.5 Thesis Structure

This thesis is composed of seven main chapters. Chapter Two shortly illustrates the theory behind the ANFIS, and the multiple linear regression approach of analysis. Chapter Three reviews the related literature on the applications of ANFIS in transportation and trip generation, and reviews also the application of trip generation in Palestine.

Chapter Four describes the methodology of this study, while Chapter Five reviews the procedures followed in collecting the required data. Chapter Six illustrates the development and validation process and the associated results of the four trip generation models, using both approaches; multiple linear regression and ANFIS. Finally, Chapter Seven presents the summary, conclusions, and recommendations of this study.

Chapter Two

Theoretical Background

Chapter Two
Theoretical Background

2.1 Introduction

This study aims to develop trip generation models using the Adaptive Neuro-Fuzzy Inference System (ANFIS) and compare its results with the traditional Multiple Linear Regression (MLR) approach, in an attempt to seek for more accurate modeling techniques. Hence, it is necessary first to explain the theoretical background behind both approaches, which is what intended to be achieved through this chapter.

ANFIS is an advanced modeling technique that has the capability of dealing with nonlinear and highly complex systems. ANFIS is an artificial intelligence-based approach, which was first proposed by Jang (1993), who integrated the best features of the Artificial Neural Network (ANN) and the Fuzzy Inference System (FIS) into a single framework for providing more enhanced prediction capabilities. The following sections briefly illustrate the concepts and the theories behind ANN, FIS, and ANFIS, while the last section discusses the multiple linear regression approach of analysis.

2.2 Artificial Neural Network (ANN)

(ANN) is a machine learning-based system developed for information processing, which is inspired by and tries to simulate the biological neural systems, such as human brain, in their way of working, learning, and operating techniques (Profillidis & Botzoris, 2018). ANN is not based on specific rules, but rather it is developed through trial and error procedure across successive calculations. Such a system "learns" to perform tasks by considering examples (dataset), generally without being programmed with any task-specific rules.

The advantage of ANN over the traditional method is that it does not need to know about the physical relationship for converting an input to output. The ANN can adapt itself to self-organize its structure, when the sample input-output training is presented.

ANN has been widely adopted and become a strong computational tool. In transportation research, ANN has been mainly used as data analytical tool for many reasons, including the capability of 1) dealing with large amounts of data, 2) recognizing patterns of operation and performance, 3) discovering linear and nonlinear relationships, 4) providing accurate and reliable predictions, 5) modeling flexibility and adaptability, and 6) finally, learning and generalization ability (Profillidis & Botzoris, 2018), (Karlaftis & Vlahogianni, 2011).

2.2.1 Artificial Neuron Concept

The basic structure of the ANN consists of an artificial neuron. The idea of the artificial neuron was first presented by McCulloch and Pitt in 1943 (Profillidis & Botzoris, 2018). This concept is illustrated in Figure 2.1, and can be represented by the following mathematical equation:

$$u_{(k)} = \sum_{j=1}^{n} w_{kj} x_j \text{ and } y_{(k)} = \varphi\left(u_{(k)}\right) + b_{(k)} \tag{2.1}$$

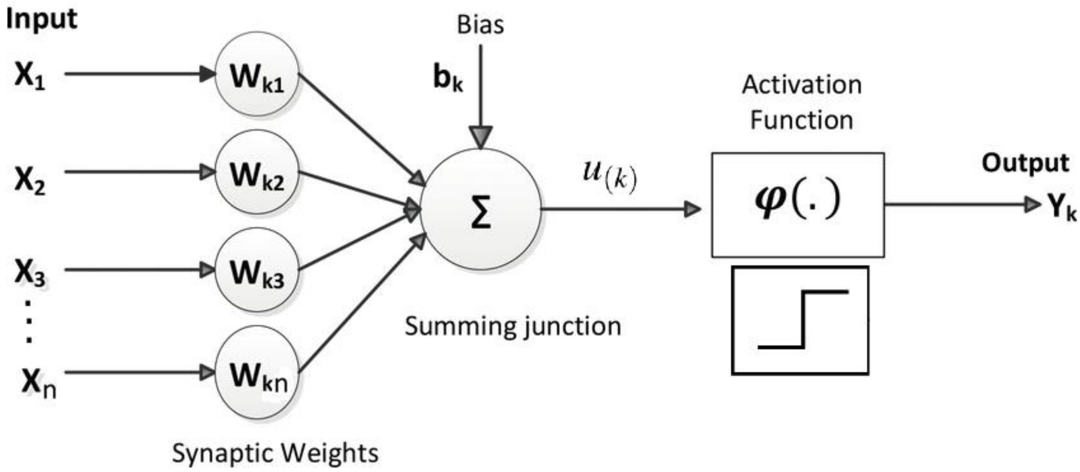

Figure 2.1: Mathematical Modeling of Artificial Neuron
Source: (Ahire, 2018)

Where simply, a number of inputs $x_{(j)}$ (maybe from external environment or other neurons), is each multiplied by a prespecified connection weight $w_{(kj)}$, and then the resulted summation product $u_{(k)}$ is compared with the neuron's threshold value for activation (activation function). If $u_{(k)}$ exceeds the threshold value, the neuron will be activated (or 'fired') and the output $y_{(k)}$ will result. Hence, $y_{(k)}$ depends on the activation function $\varphi(u_{(k)})$ and the bias $b_{(k)}$. However, several types of activation functions could be used for modeling neurons such as those illustrated in Figure 2.2, including linear, sigmoid, and gaussian functions.

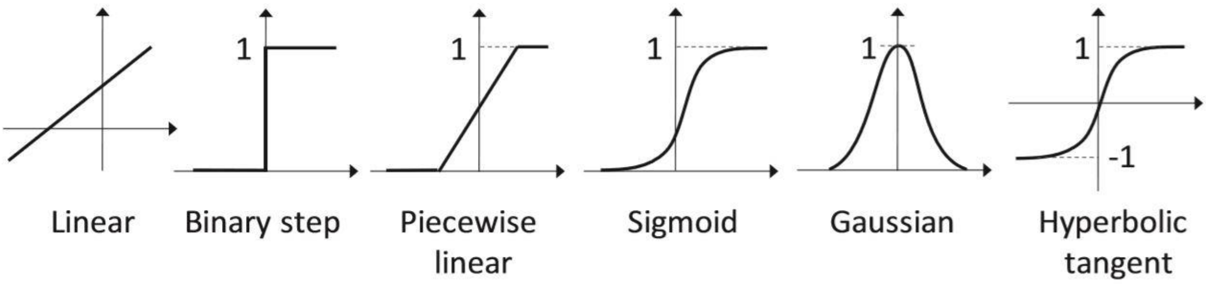

Figure 2.2: Common Types of Neuron Activation Functions
Source: (Profillidis & Botzoris, 2018)

2.2.2 ANN Architecture

The structure of ANN consisted of interconnected neurons arranged in a systematic manner to form a layered pattern. Its architecture usually consists of three different layers: input layer, an output layer, and one or more hidden layers, as illustrated in Figure 2.3. Each layer is composed of several processing neurons, each has specific function, which could be adaptive or fixed. The adaptive neuron can change its parameters over time, or during the learning process (i.e., learning epoch), whereas the fixed neuron has static function with no parameters. The architecture of an ANN describes the way the dependent variable of the output layer is associated with independent variables of the input layer.

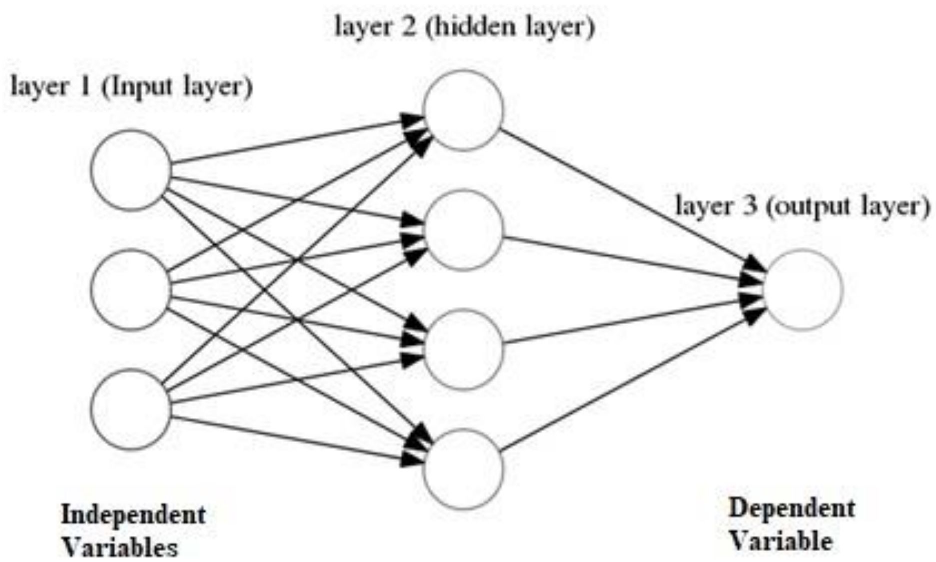

Figure 2.3: Input Layer, Output Layer, and Hidden Layer of an ANN
Source: (Profillidis & Botzoris, 2018)

Based on the pattern of connections between neurons, ANN architecture can be divided into two types, feedforward and feedback (recurrent) ANN. In the feedforward type, data are allowed to move in one direction only, so that the output of each layer will not give any effect to the previous layer. While in the feedback type, there are additional feedbacks on the previous layer. The data that are allowed to propagate forward and feedback can be an input to the neurons before.

2.2.3 Learning Process

The learning algorithm of the ANN plays a major role in the process of modifying the parameters and the values in the network to adapt its environment. The use of learning algorithms allows ANN to assemble itself for giving consistent responses to the input into the network. During the learning process, the parameters in the network will be modified. The level of learning (learning epochs) will expire when the resulting output becomes consistent with the desired output.

Several types of learning algorithms (learning rules) could be considered, however, the vast majority of applications of ANNs use the backpropagation learning algorithm, which is the most popular learning rule. The error computed at any step can be sent from the output layer backward to the hidden layer and next to the input layer. Many applications of ANNs in transportation use this backpropagation algorithm (Profillidis & Botzoris, 2018).

2.3 Fuzzy Inference System (FIS)

(FIS) is an effective technique that uses the Fuzzy Logic (FL) for modeling complex systems. The theory of FL was first initiated by Lotfi Zadeh in 1965, who is considered the father of FL (Jang & Gulley, 2018). This theory has been proven to be an effective means for dealing with objectives that are linguistically specified, such as low, medium, or high household income; child, young, or old family members; and so on.

The FIS simulates the human way of thinking and conclusion-making based on the linguistic variables, which are represented by fuzzy sets that linguistic expressions are associated with (Stojčić, 2018). However, trip generation models usually involve human decisions, which consist of vagueness and uncertainty. In addition, trip maker, in general, expresses various attributes of trip in the form of linguistic terms, which are rough and not accurate (Simha, 2017).

2.3.1 Fuzzy Set and Fuzzy Logic

Basically, fuzzy means not clear enough, not well-known. The fuzzy number is a generalization of a regular real number, a quantity whose value is imprecise, rather than exact as is the case with usual ordinary numbers. Fuzzy logic is an approach for computing, which is based on degrees of truth (represented by the degree of membership in a fuzzy set) rather than the usual true or false (0 or 1). Fuzzy logic starts with the concept of a fuzzy set.

A fuzzy set is a set without a crisp or clearly defined boundary. In the classic theory of sets, an element x even belongs or not to the set A, the membership (or not) of x within the set A is described by what's called the Membership Function (MF). $MF_{(x)}$ equals 1, if and only if x is a member of (or belong to) the set *A*, otherwise it will equal zero. Thus, in a classic set, the membership function of an element takes crisp values, either 0 or 1.

In contrast, the $MF_{(x)}$ of the fuzzy set can take any value from 0 to 1, the greater the value of $MF_{(x)}$, the greater the possibility that an element *x* belongs to the fuzzy set *A*. Thus, the degree of an object belongs to a fuzzy set is denoted by a membership value between 0 and 1. A fuzzy set is a collection of elements that might belong to the set to a certain degree, which varies from 1 (full belongingness) to 0 (full non-belongingness), through all intermediate values between 0 and 1. The value of the MF in a fuzzy set indicates the intensity of belongingness (Profillidis & Botzoris, 2018). Figure 2.4 illustrates both the classical set and the fuzzy set concepts, along with the MFs.

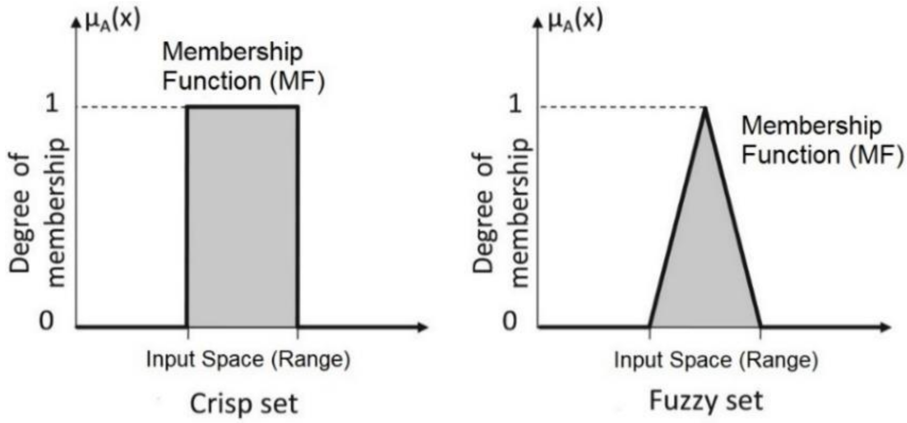

Figure 2.4: Crisp Set verses Fuzzy Set Concept
Source: (Profillidis & Botzoris, 2018)

A membership function is a curve that defines how each point in the input range is assigned to a membership value (or degree of membership) between 0 and 1. However, for each input variable (e.g., household income), different MFs could be identified, each will reflect a specific fuzzy set or a qualifying linguistic set (e.g., MF for low household income, and other MF for high household income). MFs can have several shapes, including triangle, trapezoidal, and gaussian as illustrated in Figure 2.5. The shape of the membership functions depends on specific parameters, and by changing these parameters the shape of the membership function will be changed consequently.

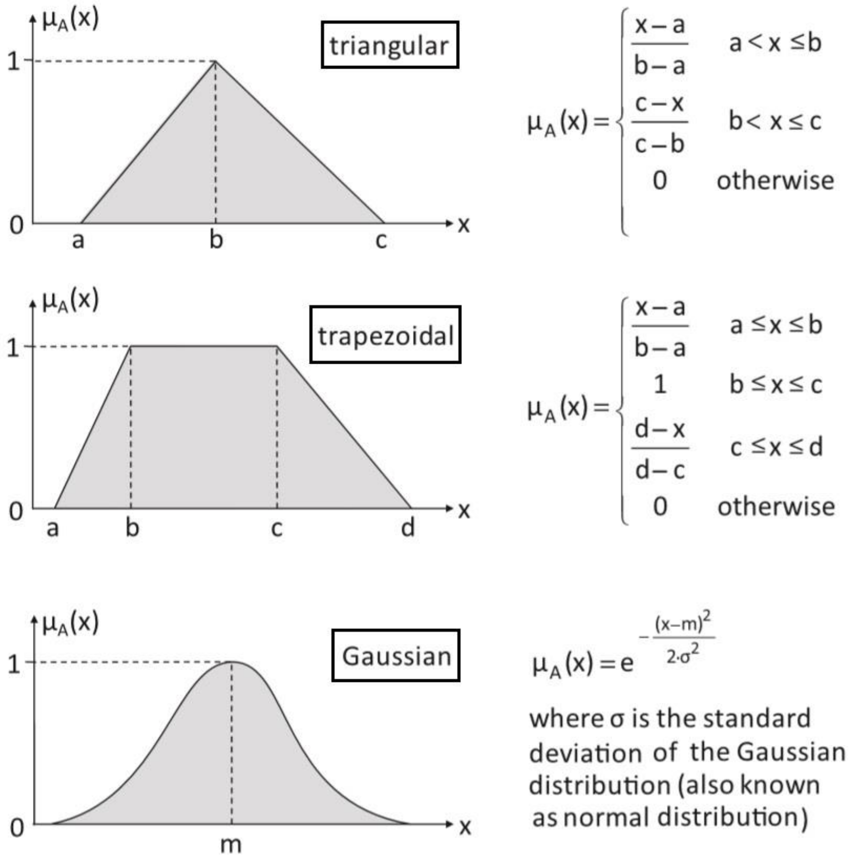

Figure 2.5: Triangle, Trapezoidal, and Gaussian Membership Functions
Source: (Profillidis & Botzoris, 2018)

2.3.2 Fuzzy Inference Process

In the fuzzy way of thinking there may be many truths between 1 (completely true) and 0 (completely false). However, the point of fuzzy logic is to map an input space to an output space, and the primary reasoning mechanism (the inference) for doing this is a list of If-Then statements called rules (Jang et.

al., 1997). These If-Then rules are used to formulate the conditional statements that comprise fuzzy logic, utilizing logical operators, such as the classical operators AND (minimum), OR (maximum), and NOT (additive complement). A single fuzzy If-Then rule has the form: If [(x is A) AND/OR (y is B)], then [(z is C)], the If-part of the rule is called the premise, while the Then-part is called the consequent.

The process of fuzzy inference, which is illustrated in Figure 2.6, involves membership functions, logical operators, and If-Then rules. The basic structure of a FIS consists of three conceptual components: a rule base, which contains a selection of fuzzy rules; a database which defines the MFs used in the fuzzy rules; and a reasoning mechanism, which performs the inference procedure upon the rules to derive an output (Přibyl & Goulias, 2003).

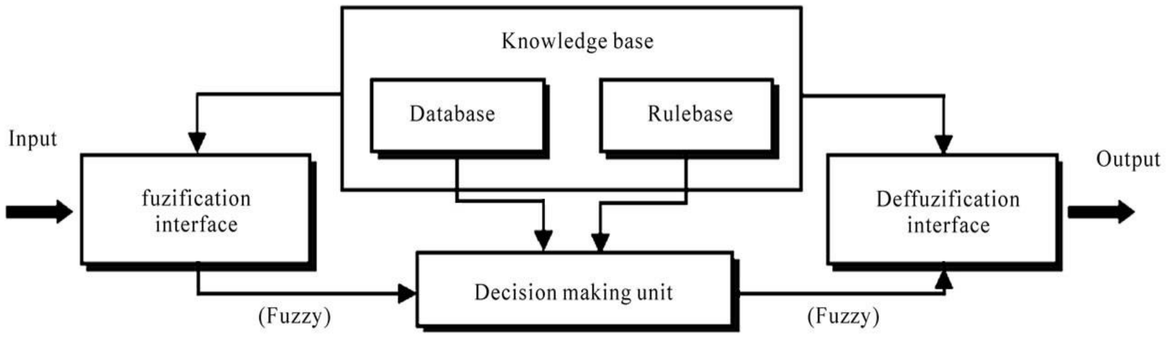

Figure 2.6: Fuzzy Inference System
Source: (Jang, 1993)

In the FIS, the crisp inputs are converted into fuzzy inputs (with values from 0 to 1) by using fuzzification interface based on the MFs. After fuzzification, the rule bases are developed considering fuzzy logical operators along with the If-Then statements. The rule bases and the database are mutually referred to as the knowledge base, which will be applied for each fuzzified input variables to decide whether an element belongs to the fuzzy set or not. Defuzzification is applied to transform the resulted fuzzy value to a crisp real-life value, which is the output.

FIS uses fixed membership functions that are chosen randomly and rules structure that is essentially predetermined by the user's explanation of the variable’s characteristics in the model (Jang & Gulley, 2018). However, the fuzzy inference process comprises of five sequential steps, illustrated in Figure 2.7, as per the following:

First Step - Fuzzification: Each input variable, which is always a crisp numerical value, will be fuzzified overall identified MFs for that variable. Fuzzifying includes the determination for each input the degrees of belongingness (membership value), to the appropriate fuzzy sets (ranging from 0 to 1 for each input).

Second Step - Apply Fuzzy Operator: The fuzzy logical operators along with the If-Then rules will be applied for each membership value, to obtain one number (output) that represents the result of the rules. This number is then applied to the output membership function.

Third Step - Apply Implication Method: The consequent is reshaped using a function associated with the premise (a single number). The input for the implication process is a single number given by the premise, and the output is a fuzzy set. The implication is implemented for each rule.

Forth Step - Aggregate All Outputs: Since decisions are based on testing all the rules in a FIS, the rule outputs must be combined. Aggregation is the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set. The input of the aggregation process is the list of reshaped output functions returned by the implication process for each rule.

Fifth Step – Defuzzification: The aggregate of a fuzzy set encompasses a range of output values, and must be de-fuzzified to obtain a single output numerical value. The defuzzification methods include: centroid, bisector, middle of maximum and so on. The most popular defuzzification method is the centroid calculation, which returns the center of area under the curve.

A fuzzy inference diagram that displays information flows through all parts of the fuzzy inference process from fuzzification through defuzzification is shown in Figure 2.7 for two inputs and one output variables.

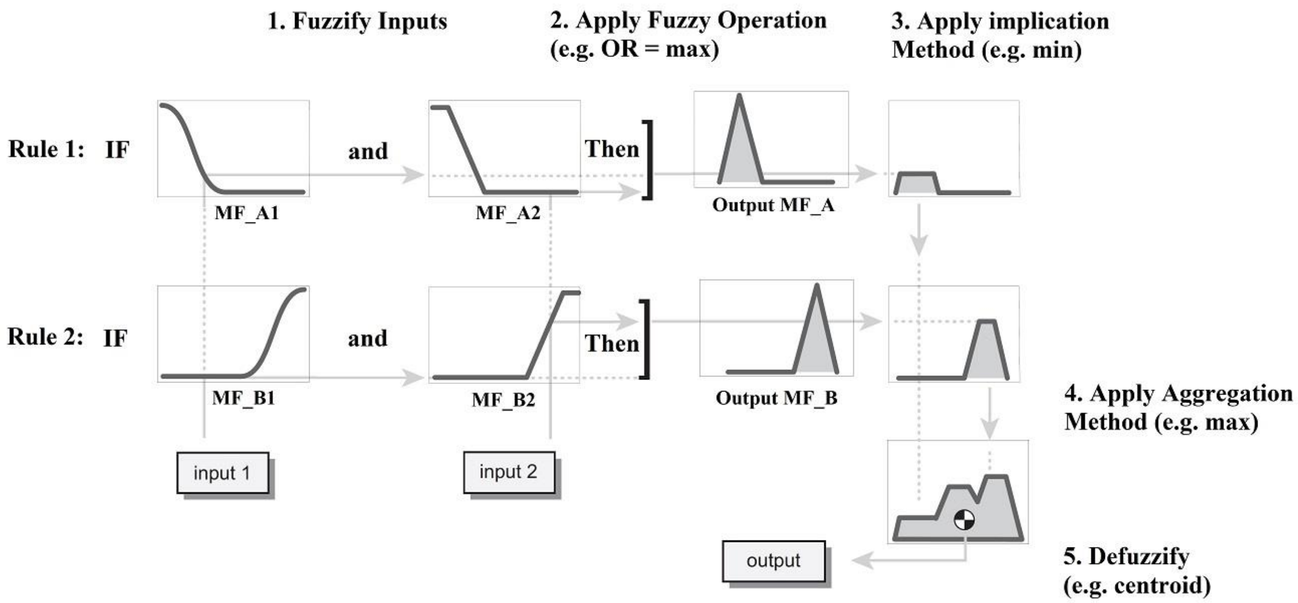

Figure 2.7: Fuzzy Inference Process Diagram
Source: (Jang & Gulley, 2018)

2.3.3 Types of Fuzzy Inference System

There are several types of FIS, mostly used are Mamdani type, and Takagi–Sugeno type. Mamdani fuzzy inference system was among the first systems built using fuzzy set theory, which was proposed in 1975, and expects the output membership functions to be fuzzy sets. After the aggregation process, there is a fuzzy set for each output variable that needs defuzzification (Profillidis & Botzoris, 2018).

Takagi-Sugeno fuzzy inference system, was introduced in 1985, is similar to the Mamdani method in many aspects, the principal difference is that the Takagi-Sugeno output membership functions are either linear or constant. A typical rule in a Sugeno fuzzy model has the form: $[\text{If Input 1 is x and Input 2 is y, then Output } f = ax + by + c]$. The final output of the system is the weighted average of all rule outputs. Sugeno systems always use product implication and sum aggregation (Jang, 1993).

Because it is more compact and computationally efficient representation than Mamdani system, Takagi-Sugeno system lends itself to the use of adaptive techniques for constructing fuzzy models. These adaptive techniques can be used to customize the membership functions so that the fuzzy system best models the data. However, FIS based on Takagi–Sugeno model was found to be widely used in the application of ANFIS method (Profillidis & Botzoris, 2018).

2.4 Adaptive Neuro-Fuzzy Inference System (ANFIS)

(ANFIS) is an adaptive neural network that is functionally based on the model of Takagi–Sugeno fuzzy inference system. It was first introduced, as mentioned earlier, by Jang (1993), who combined both learning capabilities of ANN and reasoning capabilities of FIS into one single framework. ANFIS can construct an input-output mapping based on both human knowledge (in the form of fuzzy if-then rules) and stipulated input-output data pairs (Jang, 1993).

The ANN can generate input-output models from sets of training data without being interested in the mechanism and the understanding of what happens between inputs and outputs (Profillidis & Botzoris, 2018). It appears as a black box, that does not have explicit knowledge representation. Therefore, many researchers face problems with explaining the meaning of its structure and the results obtained. The fuzzy logic, on the other hand, does not incorporate any learning mechanism, instead, it relies on the experience of people who already understand and are familiar with the mechanism of the system, and predefine the parameters of the MFs and the fuzzy rules (Profillidis & Botzoris, 2018).

In some modeling situations, it cannot be known what the membership functions, or the fuzzy rules should look like by simply looking at the data. In such a case, the ANFIS constructs a FIS from given input-output dataset, using the learning capability of ANN for extraction of the optimum fuzzy If-Then rules and MFs parameters. The optimization of these parameters is undertaken in such a way that the error between the estimated (target) and actual output is minimized (Tan et al., 2018). When using such an approach, the MFs parameters and the fuzzy rules are derived from training data instead of predefined.

Neuro-adaptive learning techniques provide a method for the fuzzy modeling procedure to learn information about a data set. The fuzzy rules and the membership functions are generated from training samples, and can be adjusted during the learning process (epoch) using a specific learning algorithm, which makes the assessment process closer to the real situation and adaptable to dynamic changes (Jang et al., 1997).

ANFIS method is flexible, can learn independently, and adapt itself to its environment. It has the ability to deal with uncertain human behavior, and easily incorporate both linguistic and numeric knowledge for problem-solving. In this context, neuro-fuzzy system arises as superior method for dealing with the urban travel behavior problems (Andrade et al., 2006).

2.4.1 ANFIS Architecture

ANFIS is structured basically by a five-layered feed-forward network and a specified learning algorithm, which is used to adjust the system. However, for simplicity, it assumed that a typical first-order Takagi–Sugeno fuzzy model, as shown in Figure 2.8, has two inputs x and y, two *If-Then* fuzzy rules, one output f, and each input variable has two associated MFs (Sugeno,1985). In this model, a basic rule set with two fuzzy *If-Then* rules can be expressed as per the following:

Rule 1: $\text{if } x = A_1 \text{ and } y = B_1, \text{ then } f_1 = p_1x + q_1y + r_1$

Rule 2: $\text{if } x = A_2 \text{ and } y = B_2, \text{ then } f_2 = p_2x + q_2y + r_2$

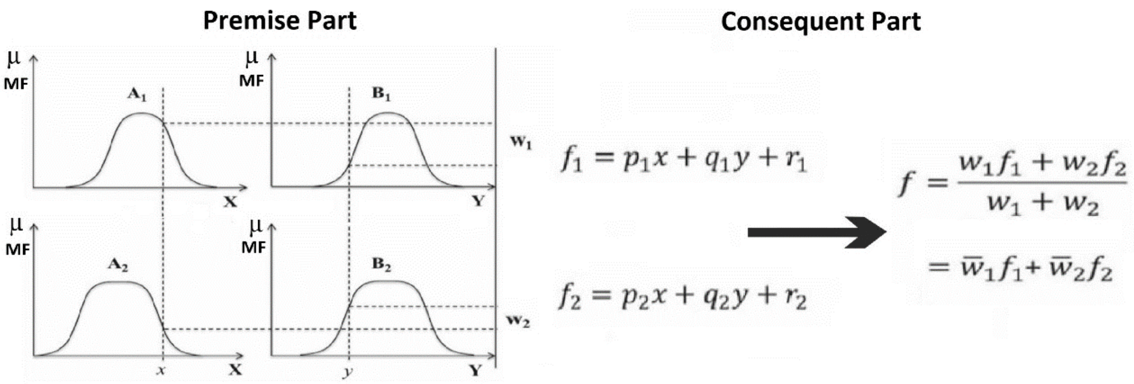

Figure 2.8: Takagi–Sugeno Fuzzy Model with Two Inputs and Two Rules
Source: (Jang, 1993)

Where A_1, A_2, B_1 and B_2 are the MFs of the inputs (*x*, *y*), respectively (part of the premises), and (p_i, q_i, r_i) for (i = 1, 2) are the linear parameters of the output function in part-Then (consequent part). However, the equivalent typical ANFIS architecture is shown in Figure 2.9, which consists of five layers that perform different functions. The first and fourth layers contain adaptive nodes (neurons) represented by squares, while the other layers have fixed nodes represented by circles. The following represents a brief description of each layer (Jang, 1993):

Layer 1 - Fuzzification Layer: All the nodes in this layer are adaptive, indicating that the parameters of the MFs can be modified during training epochs. The outputs of this layer are given by: $O_i^1 = \mu_{Ai}(x)$, where *x* is crisp input to node *i*, A_i is the linguistic label (MF name or number), and μ_{Ai} is the membership function of fuzzy set A_i, which can be linear or nonlinear. The parameters of the MF in this layer are referred to as premise parameters.

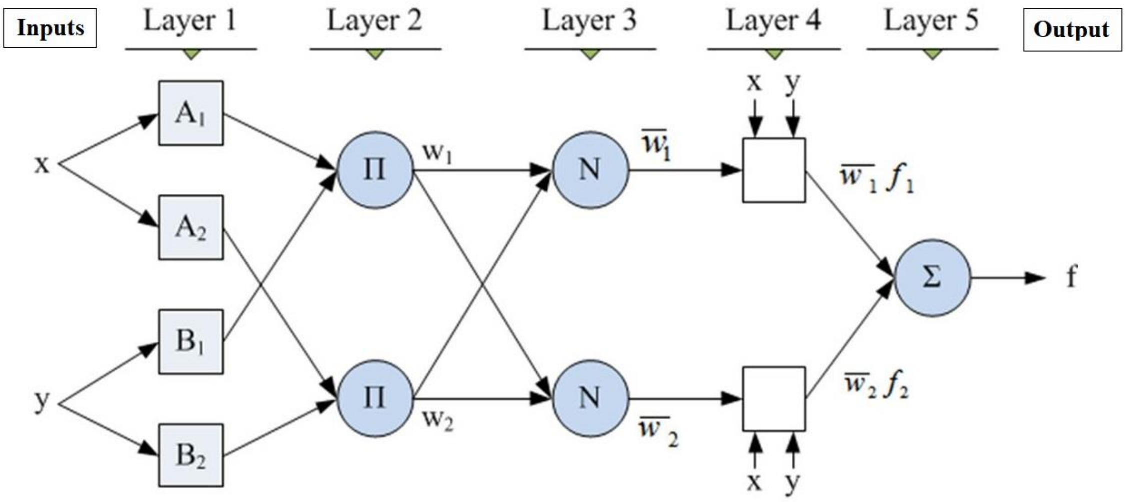

Figure 2.9: The Equivalent ANFIS Architecture for Takagi–Sugeno FIS
Source: (Jang, 1993)

Layer 2 - Rule Layer: This layer has circle fixed nodes labeled Π, indicating that they perform as a simple multiplier. The output is the product of all inputs, each node output represents the firing strength of each rule, and can be represented as:

$$W_i = \mu_{Ai}(x) \times \mu_{Bi}(x) \quad i = 1, 2 \tag{2.2}$$

Layer 3 - Normalization Layer: The nodes in this layer are also circle nodes labeled N. The i-th node is the ratio of the i-th rule's firing strength to the sum of all rule's firing strengths. The outputs of this layer, which is called normalized firing strengths, are given by:

$$\overline{W_i} = W_i/(W_1 + W_2) \quad i = 1, 2 \tag{2.3}$$

Layer 4 - Defuzzification Layer: Every node i in this layer is adaptive, the parameters in this layer are considered as consequent parameters. The outputs of this layer can be represented as:

$$O_i^4 = \overline{W_i} f_i = \overline{W_i}(p_i x + q_i y + r_i) \tag{2.4}$$

Layer 5 - Summation Layer: The node in the last layer is a fixed node labeled Σ, which computes the overall output as the summation of all incoming signals. The overall output is given as:

$$\text{Overall Output} = f = O_i^5 = \sum_i \overline{W_i} f_i = \frac{\sum_i W_i f_i}{\sum_i W_i} \qquad (2.5)$$

It is good to mention here that in the ANFIS system, each input variable may be clustered into several class values in layer 1 to build up fuzzy rules. Each fuzzy rule would be constructed using two or more membership functions in layer 2. Several methods have been proposed to classify the input data and to establish the rule-based relationship between the input and output variables, among which the most common being the grid partition and the subtractive fuzzy clustering. However, when there are few input variables, the grid partition is considered to be a suitable method for data classification (Srisaeng et al., 2015).

2.4.2 ANFIS Learning Algorithm

Training is a key part of the ANFIS model development process. The major task of the training process is to optimize the fuzzy rules and the associated parameters of the input and output MFs, which could be achieved by using a specific learning algorithm. The learning algorithm provides a measure of how well the fuzzy inference system is modeling the input-output data for a given set of parameters. It seeks to minimize some measure of error, for instance, the root means square error, between the observed and predicted data (Srisaeng et al., 2015).

Generally, two types of learning algorithm could be used for ANFIS, that are: the **backpropagation learning algorithm**, gradient descent method for all parameters (a steepest descent method), or the **hybrid learning algorithm**, which combines the backpropagation for the parameters associated with the input membership functions (premise parameters in layer 1), and least squares estimation for the parameters associated with the output membership functions (consequent parameters in layer 4). However, for more details, Jang (1993), Jang et al. (1997), and Rutkowska (2002) provide full insight into the working mechanism of these algorithms.

2.5 Multiple Linear Regression Analysis (MLR)

Basically, regression analysis is a conventional statistical technique that is commonly used for capturing the correlations and studying the patterns of relationships among two or more variables. For regression analysis, two types of variables can be identified:

1. **Dependent Variables:** which are the indicators of travel behavior (e.g. number of trips produced by household), also known as a response or output variables. These variables can be affected by other variables, and usually denoted by (Y).

2. **Explanatory (Independent) Variables:** which are the determinants of travel behavior (e.g. households socio-economic characteristics), also known as predictor or input variables. These variables are not affected by other variables, and usually denoted by (X).

The linear regression technique attempts to model the relationship between one dependent variable (Y) and one or more explanatory variables (Xs) by fitting a linear equation to observed data. However, if the linear equation has more than one explanatory variable, the regression is called 'multiple', and can be written as per the following:

$$\mathrm{Y} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{2.6}$$

where:

Y	=	Dependent Variable
X_n	=	Independent (explanatory) Variable
α	=	Constant
β_n	=	Coefficient of Independent Variable

The selection of the most relevant explanatory variables to be included in the model is usually achieved by following the stepwise regression procedure, which is based on the prespecified criterion that reflects the overall fitness of the model, and usually takes the form of a sequence of t-tests or F-tests. In each step, a variable is considered for addition to or subtraction from the set of explanatory variables. Stepwise regression has two common approaches, the backward elimination approach or the forward selection approach (Washington et al., 2010).

The backward elimination approach starts with a regression model that contains all explanatory variables, and sequentially removing one variable at each step. The variable removed is the one that contributes least to the overall fitness of the model. The removal process is repeated until removal of any variable results in a significant change in the overall fitness of the model.

However, the t-test is usually employed for this purpose, the explanatory variable with the least t-value is selected for removal.

The forward selection approach begins with a simple regression model that contains only the constant term, and sequentially grows at each step, by adding the variable with the largest contribution in the overall significance of the model. This process is repeated until there is no more variable that results in a significant increase in the overall fitness of the model.

In this research, the backward elimination approach was employed for building the required models using the Statistical Package for Social Sciences (SPSS) software. However, the unknown parameters associated with the regression model (i.e., α and β_n) can be estimated using the commonly used Ordinary Least-Squares (OLS) technique, which estimates the parameters based on minimizing the sum of squares of the differences between the actual and predicted values of the dependent variable.

Chapter Three

Literature Review

Chapter Three
Literature Review

3.1 Introduction

The Adaptive Neuro-Fuzzy Inference System (ANFIS) is an artificial intelligence-based approach, that was first introduced by Jang in 1993, as described in Chapter Two, which is an advanced modeling technique that has the capability of dealing with nonlinear and highly complex systems.

ANFIS incorporates the best features of the Artificial Neural Network (ANN) (learning capabilities) with the Fuzzy Inference System (FIS) (reasoning capabilities) into one single framework. It consists of an adaptive neural network that is functionally based on the model of the Takagi–Sugeno fuzzy inference system. Since the development of ANFIS, it had been successfully applied in many research areas, including transportation, as a prediction, knowledge discovery, decision-making, and evaluation tool.

Several studies in different scientific disciplines have proved the effectiveness and superiority of ANFIS in modeling complex nonlinear systems over the conventional techniques, such as the studies of Wahyudi et al. (2019) in construction management, Elhami et al. (2016) in the agricultural field, and Shafi et al. (2016) in the intelligent systems.

This chapter reviews some of the successful studies that considered the applications of ANN, FIS, and ANFIS in transportation in general and modeling trip generation at specific. It also reviews the latest empirical studies conducted in Palestine for the development of trip generation models.

3.2 Application of ANN in Transportation and Trip Generation

In transportation research, for example, Karlaftis & Vlahogianni (2011) discussed the differences and similarities, and compared between the applications of statistical methods and neural networks as a data analytics tool. They reviewed many relevant literatures that compare the performance of the two approaches in six distinct categories of transportation research: traffic operations, infrastructure management, maintenance and rehabilitation, planning, environment and transportation, and safety and human behavior. They found that when modeling complex datasets with possible nonlinearities or missing data, neural networks are often regarded as more flexible compared to statistical models.

Pamuła (2016) claimed that the features and computing capabilities of neural networks favor their wide application for solving transport problems, by presenting several examples that illustrate the effectiveness of this approach. The review shows that feedforward multilayer neural networks are the most often utilized configurations in transportation research. However, no systematic approach is reported on the optimization of the ANN configurations to achieve a set level of performance in solving modeling tasks.

At the Palestinian level, ElAstal (2014) successfully employed the ANN for modeling road accidents black spots, by analyzing the traffic accidents locations in Gaza Strip between 2000 to 2005. Several factors were considered, including the traffic volume, the surface type, the design speed, the number of lanes, and others. The R-squared value for the developed ANN model was 55%, which, based on the author's claims, could represent the real situation.

In trip generation, Tillema et al. (2004) explored the performance of artificial neural networks and commonly used regression models in dealing with trip generation, based on a set of 20 synthetic households' classes, that was generated using the Dutch national travel diary data. The Root Mean Square Error (RMSE) was used for performance comparison. However, they concluded that neural networks can be successfully used for modeling trip generation and usually able to outperform traditional regression methods. Moreover, they claimed that there is no one best neural network configuration for all proportions of available data, and the performance depends also on factors like the activation function, learning method, and stop criteria for learning.

Goel & Sinha (2008) also successfully demonstrated the application of artificial neural networks for modeling trip generation in Meerut City-India, where the error generated in training phase was quite low. The application demonstrates that the relationship between socioeconomic variables and transport variables is nonlinear, which could be taken care by ANN.

3.3 Application of FIS in Transportation and Trip Generation

In transportation research, for example, Teodorovic (1999) provided a comprehensive review of the results achieved by using fuzzy logic to model complex traffic and transportation problems. The Author reviewed the successful application of the fuzzy logic in modeling many fields of transportation, including trip generation, trip distribution, modal split, route choice, air transportation, network control, level of service, and many others. The author concluded that fuzzy logic could be used successfully to model situations in which people make decisions in complex environment where it is hard to develop a mathematical model. The author concluded also that fuzzy set theory and fuzzy logic present a promising mathematical approach to model complex transportation problems that are characterized by subjectivity, ambiguity, uncertainty, and imprecision.

In trip generation, Rassafi et al. (2012) developed a fuzzy expert system (i.e. FIS) for predicting the rate of trips generated in the city of Mashhad-Iran. They showed that the multiple linear regression models have several limitations including: the dependency on the exact prediction of independent variables in future, and that this approach has many assumptions, which raise challenging questions of its application. However, they concluded that FIS is able to make suitable predictions using uncertain and inexact data, and can be a good competitor for multiple linear regression method, especially, when there is no exact data for independent variables.

Pulugurta et al. (2012) employed the advantages of the fuzzy logic to model trip generation rates, based on household interview conducted in Port Blair-India. They illustrated the limitation of traditional linear regression models, where these models do not take into account subjectivity, imprecision, ambiguity, and vagueness of human minds. However, they observed that the results obtained from the fuzzy logic model gave better prediction accuracy in comparison to the traditional regression model, and they concluded that the fuzzy logic models were better able to capture and incorporate the human knowledge and reasoning into trip generation modeling.

Simha (2017) developed a fuzzy rule-based trip generation model for urban and suburban areas of Guntur City-India, using households socioeconomic and trip characteristics. Simha proved that the developed FIS can outperform the multiple linear regression approach, by comparing the number of trips predicted by both. Simha also found that the fuzzy logic approach was much more compatible to estimate the trip generation rates.

3.4 Application of ANFIS in Transportation

In transportation planning, the ANFIS had been applied mainly in mode choice modeling and traffic assignment. For example, Andrade et al. (2006) developed a hybrid model that combines the multinomial logit model with a neuro-fuzzy inference system. The model is applied for estimating traveler behavior in the context of transport mode choice, to investigate shopping traveler preferences regarding the modes of bus, subway, and automobile. The model was evaluated by comparing its results with the results of a multinomial logit model. The model demonstrated good performance by estimating a large number of right choices during the validation process. The results confirmed that the proposed model can describe uncertainties regarding traveler decisions on the time of transport mode choice.

Tortum et al. (2009) developed an ANFIS for modeling mode choice of intercity freight transport. They found that the traditional mathematical models are not only becoming almost intractable, but also data-intensive, difficult to calibrate and update, and not transferable. Moreover, they found that the ANFIS approach is highly adaptive and efficient in investigating non-linear relationships among different variables. It was tested on the freight transport market in Turkey, Germany, France, and Austria by using information on the freight flows and their attributes. They concluded that the ANFIS models are more successful in the representation of the non-linear behavior of mode choice of intercity freight transport as compared to the classical models.

Moreover, Stojčić (2018) reviewed the application of the ANFIS in the field of road traffic and transportation from the year 1993 till 2018. The author reviewed many areas including: vehicle routing, traffic control, vehicle steering and control, safety, modeling, traffic congestion, and others. From the literature reviewed by the author, it was concluded that ANFIS has been widely used due to its ability of modeling non-linear systems as well as its ability of adaptability (learning from examples). Moreover, the author concluded that ANFIS represents a promising modeling method, that can show better performance compared with the traditional methods.

3.5 Application of ANFIS in Trip Generation

In trip generation modeling, the application of ANFIS is not widely considered, and found to be limited to specific few studies, such as those presented hereafter.

Přibyl & Goulias (2003) developed an ANFIS for modeling trip generation of individuals (number of trips per person per day), using different options and settings, based on collected data from South Perth, Australia. They compared their results with two traditional analytical methods - linear regression and negative binomial models. They found that ANFIS is a potentially better data analytic method, which needs to be explored more in-depth, and compared to more sophisticated regression techniques that are already in use in transportation.

Ahmadpour et al. (2009) successfully developed a neuro-fuzzy inference system for modeling travel demand, specifically for full-time worker trip production, in Adelaide Metropolitan Area, Australia, based on the household/person characteristics. They concluded that the main advantage of the neuro-fuzzy technique is that both human knowledge in the form of linguistic terms and input-output data can be utilized in modeling, and also the model can be highly nonlinear from mathematical aspect.

Mahdavi & Mamdoohi (2018) developed an ANFIS for modeling trip generation based on land use data and socio-economic characteristics of 113 traffic analysis zone of Qazvin City, Iran. A comparison was performed with linear regression approach, and the result showed that the ANFIS has superiority over the linear regression techniques, where the R^2 for the ANFIS was 0.998, while for the linear regression it was 0.582. They indicated that the ANFIS model performs more accurately (higher R^2 and less RMSE) than the linear regression approach.

3.6 Trip Generation Models in Palestine

In Palestine, there are little documented experiences concerning transportation planning in general and the development of trip generation models at specific. Few specific studies were performed for this purpose, using the conventional linear regression statistical approach, as presented hereafter.

Moussa (2013) developed trip generation models for Gaza City, by using conventional and multiple cross-classification methods for trip production, and multiple linear regression technique for trip attraction. The author found that vehicle ownership, household size, income level, and a total number of licensed drivers are the primary factors that affect trip production in Gaza City. Moreover, the author found that the performance of these models will be improved by increasing the sample size.

Dodeen (2014) and Abu-Eisheh et. al. (2017) used the multiple linear regression technique for developing trip generation models for Jericho City. The authors found that the trip production rates are mainly affected by the household socioeconomic characteristics, including the number of employed persons, the number of persons who are receiving education, and the household monthly income. The authors developed three types of models, including general model for all trips generated by household, trip generation models based on trip purpose, and based on trip making period.

As an extension for Dodeen's works, Amer (2017) examined the potential for spatial transferability of the trip generation models estimated by Dodeen

(2014) from Jericho City to Salfit City. This transferability was investigated, and the results were compared with those resulting from the models already generated first for Salfit. Two approaches for testing transferability were used; “Native Transfer” and “Updating Constant”. The author concluded that if the existing variables in relevant model have similarities in socioeconomic characteristics between two cities, transfer effectiveness will improve. Moreover, the author concluded that the transferability of general trip generation models between cities is generally feasible, and could save cost, time, and effort.

There are more other studies that considered the development of trip generation rates for specific land uses in Palestine, such as the trip generation study for the West Bank, which was done by Al-Sahili et al. (2017), where they conducted a research to estimate trip generation rates for major land uses in the West Bank, including residential, office, commercial, school, hospital, and hotel. However, based on conducted traffic counting surveys for the selected sample, trip generation rates and equations were estimated for the selected land uses.

3.7 Summary

The review of previous literature reveals that the ANFIS represents a promising technique that can be successfully used for modeling complex and nonlinear systems. For trip generation modeling, it has been shown that ANFIS has better performance compared with traditional regression techniques.

Based on that, it is being motivated to investigate the feasibility of using the ANFIS for modeling trip generation in Palestine, which suffers from the limited number of studies that concerned with this purpose, by conducting a comparative analysis with the used traditional linear regression approach for this purpose.

It is to be stated that the ANFIS has not been implemented yet for modeling transportation or trip generation in Palestine. Moreover, there are no found literature or studies in Palestine that undertake comparative analysis among the performance of different trip generation modeling approaches.

Chapter Four

Methodology

Chapter Four
Methodology

4.1 Introduction

This study aimed to develop four main types of home-based trip generation models for Salfit City, as mentioned in Chapter One. Again, the first model was for estimating the total number of daily general trips produced by a household, denoted as ALLTRIP. The second and the third models were for estimating the number of daily home-based trips generated for two main purposes, work trips (HBW), and education trips (HBE). The fourth model was intended to estimate the number of other daily trips generated by a household (HBO), which are trips not for work or education purposes.

Each type of these models was developed using two main competing approaches; the Multiple Linear Regression (MLR) approach, and the Adaptive Neuro-Fuzzy Inference System (ANFIS), followed by conducting a comparative analysis among their performance accuracy. The procedures to achieve this purpose were divided into four main steps, which are summarized in Figure 4.1, and the subsequent sections. These steps, however, were repeated for each type of these trip generation models.

4.2 Data Collection

This study relies on the dataset collected by Amer (2017), who considered the development of trip generation models for Salfit City, as mentioned in Chapter One. Amer collected a total number of 309 samples of households,

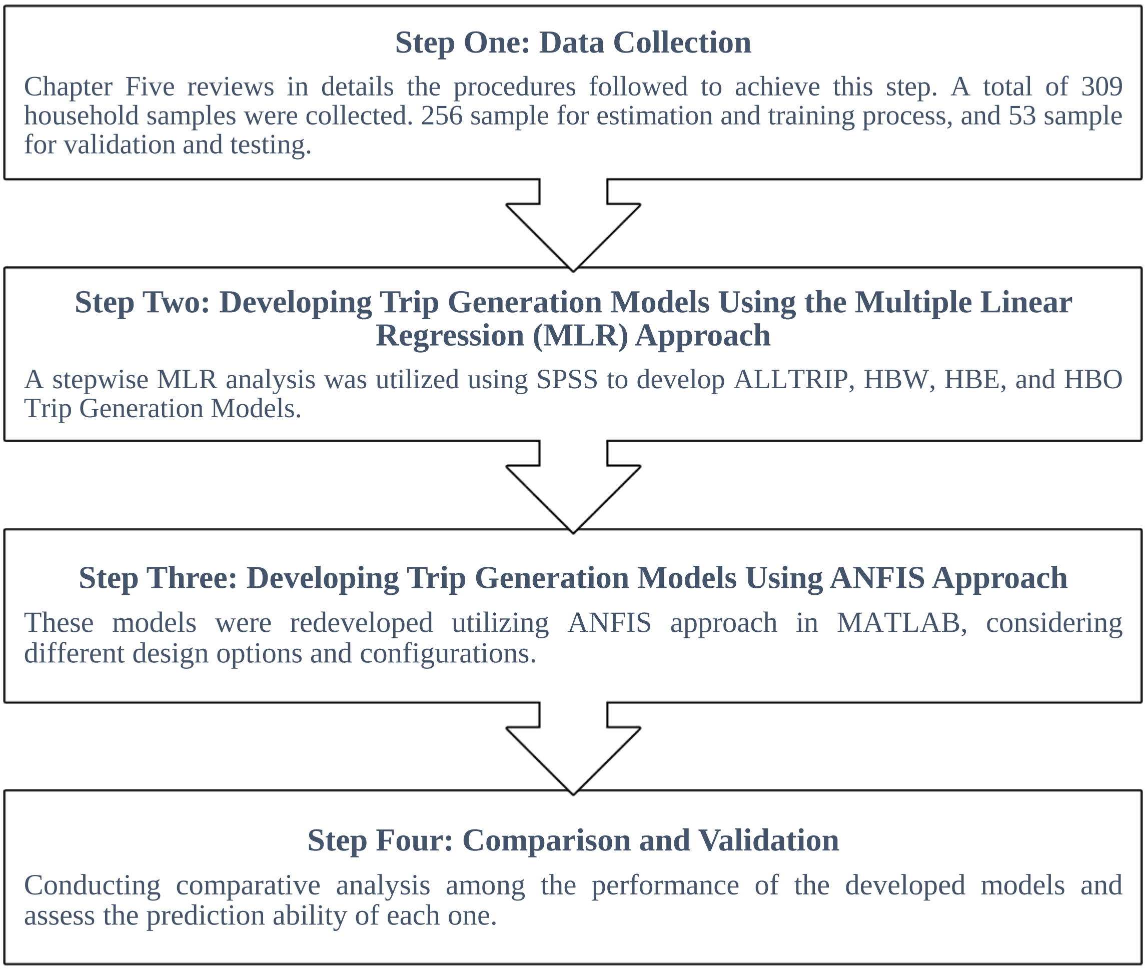

Figure 4.1: The Four Main Steps of the Study

which were distributed over 6 traffic analysis zones. Each collected sample included data regarding household socioeconomic characteristics and the associated number of daily trips produced.

These data were checked for sufficiency, adequacy, and consistency. The minimum recommended sample size was estimated to be 253 households, which is satisfied by the collected samples. However, the 256-households sample, which constitutes nearly 83% of the dataset, was used by Amer for estimation and calibration process. The same sample was also used in this study for the development and training of the desired models, and the remaining 53 samples were used for the validation process, to obtain fair and meaningful comparison among the performance of these models.

Chapter Five reviews in detail, and evaluates, the procedures followed by Amer in collecting these samples, summarizes the nature of the collected samples, and describes the selected household characteristics (explanatory variables) that have been correlated with the number of produced trips (dependent variables).

4.3 Developing Trip Generation Models Using Regression Approach

In this step, the development of the desired trip generation models was achieved by utilizing the MLR analysis using the Statistical Package for Social Sciences (SPSS). This step was done first by Amer (2017), however, it was repeated here in this study to ensure the proper selection of the relevant explanatory variables (household characteristics), that effectively describe each type of the four models, taking into consideration the correlations among these variables.

The backward elimination stepwise regression approach, based on the ordinary least-squares estimation technique, was considered for the estimation and calibration process (i.e., estimating the associated coefficients with the most relevant explanatory variables). Under this approach, the fitness of the regression analysis was evaluated based on different statistical tests that would quantify the significance of each explanatory variable that was assumed to be relevant. Accordingly, explanatory variables with coefficients that are least significant are excluded, as described in Chapter Two.

A total of 256 household samples were used for the development process of the four models (83% of the dataset), by following the linear regression model-building procedures described hereafter. However, several statistical tests were performed in this step to assess the goodness of the models, such as R-squared, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), F-test, and t-test, as will be described in the following sections.

4.3.1 Linear Regression Model-Building Procedures

While developing the multiple linear regression equations, the following steps have to be considered sequentially:

1. Detecting Nonlinearities by examining the relationships between the dependent variable and each of explanatory variables. However, if nonlinearities are detected, the relationship must be linearized by transforming the dependent or the explanatory variable, or both.

2. Developing the correlation matrix, which involves all explanatory variables, and detecting the potential sources of multicollinearity between each pair of these variables. If multicollinearity among two variables is detected (correlation value much closer to one), then eliminating one of them is necessary.

3. Selecting the appropriate and most relevant explanatory variables to be used in the model, which can be done by considering the stepwise regression procedures as described earlier. However, the unknown

parameters associated with the regression model (i.e. α and β_n in Chapter Two, Equation 2.6) can be estimated using several methods. These include the maximum-likelihood, and the commonly used ordinary least-squares estimation technique (considered for this study), which estimate the parameters based on minimizing the sum of squares of the differences between the actual and predicted values of the dependent variable.

4. Performing several statistical tests to ensure the selection of the "best fit" model among several competing models. These tests and measures are described in the following section. However, to ensure the validity of the developed model, logical aspects in terms of value, sign, and effect, must be reasonable and as expected, which should be taken into consideration in addition to the statistical tests.

5. Model verification, which is conducted to check the model ability in predicting future behavior, and can be done by comparing the predicted output of the model with the actual value for data set that are not used in the model estimation.

4.3.2 Statistical Tests

To ensure the goodness of the developed model, several relevant statistical tests were conducted in this step, which are 1) the Pearson's correlation and the Variance Inflation Factor (VIF) to check for multicollinearity, 2) the t-test to assess the significance of the individual regression coefficients (regression parameters), 3) the F-test for testing the overall significance of the model, and 4) the coefficient of determination (R-Squared), the Root Mean Squared Error (RMSE), and the Mean Absolute Error (MAE) to assess the developed model goodness of fit. These tests and measures are described in detail hereafter.

1) Pearson's Correlation & VIF: Testing for Multicollinearity

Multicollinearity among explanatory variables is a problem often encountered with observational data. The problem of multicollinearity arises when two or more variables included in the regression model have linear relationships. Multicollinearity can have serious effects on the estimates of the regression coefficients and on the general applicability of the estimated model (Montgomery & Runger, 2014). The effects of multicollinearity may be easily demonstrated and checked using Pearson's correlation and Variance Inflation Factor (VIF).

Pearson's Correlation Coefficient (r) is a quantitative measure of the strength of the linear relationship between two variables A and B, which could be defined as:

$$r = \frac{\sum_i (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_i (A_i - \bar{A})^2}\sqrt{\sum_i (B_i - \bar{B})^2}} \tag{4.1}$$

where:

r	=	Pearson's Correlation Coefficient
A & B	=	Explanatory Variables
$\bar{A}$ & $\bar{B}$	=	Mean of A & B
i	=	Number of Observations

It is necessary to develop the correlation matrix among each pair of explanatory variables, using the above equation, before start building regression models. The correlation parameter lies within the interval [–1, 1]. If the two variables are perfectly linearly, then r_{AB} = 1, and if no linear relationship between them, the coefficient r_{AB} = 0 (Montgomery & Runger, 2014).

As a rule of thumb, the strength of linear relationship could be considered small when Pearson's correlation absolute value is between (0.1 to 0.3), it could be medium when the value (0.3 to 0.5), and large when it is between (0.5 to 1.0) (Hunt & Broadstock, 2010).

The **Variance Inflation Factor (VIF)** quantifies how much the variance of a regression coefficient is inflated due to multicollinearity in the model. It provides an indicator that measures the effect of collinearity and how much the variance of an estimated regression coefficient is increased. To calculate VIF for the explanatory variable (X_j), the following formula could be used:

$$\mathrm{VIF}_j = \frac{1}{1 - \mathrm{R}_j^2} \tag{4.2}$$

Where R^2_j is the R-Squared value obtained by regression explanatory variable X_j on all other explanatory variables in the model. Any VIF that exceeds 1 indicates some level of multicollinearity in the data. The larger the VIF, the more severe the multicollinearity. However, some authors have suggested that if any variance inflation factor exceeds 10, multicollinearity is a problem (Montgomery & Runger, 2014).

2) T-Test: Testing Individual Coefficients

The t-test is used to test the significance of individual regression coefficients (βs). Such tests would be useful in determining the potential value of each of the explanatory variables in the regression model. The null hypothesis (H_0) for testing the significance of each individual coefficient (β_j) is H_0: $\beta_j = 0$, where the alternative hypothesis is H_1: $\beta_j \neq 0$. The t-statistic is used to test this hypothesis. If the calculated $|t_0| > t_{\alpha/2,n-p}$, the null hypothesis (H_0) should be rejected. To calculate the t-value, the following formula is used:

$$\mathrm{t}_0 = \frac{\hat{B}_j}{\mathrm{SE}\ (\hat{B}_j)} \tag{4.3}$$

where:

$\hat{B}_j$	=	Regression Coefficient
$\mathrm{SE}\ (\hat{B}_j)$	=	Standard Error of Regression Coefficient

Usually α-value is considered to be 0.05, However, as a rule of thumb, if the calculated t-statistic is greater than 2 in absolute value (i.e. $|t_0| > 2$), it is concluded that the estimate is statistically different from zero at 95% level of significance.

3) F-Test: Testing Overall Significance of the Model

The F-statistic is used to determine whether a linear relationship exists between the dependent variable (Y) and a subset of the explanatory variable (X_1, X_2, …, X_n). It's used to test whether the regression coefficients are jointly equal to zero or not. In other words, the F-test is used to test the overall significance of the regression model.

The null hypothesis (H_0) for testing the overall significance of the model is that the regression coefficients for the explanatory variables are all equal to zero (i.e. $\beta_1 = \beta_2 = \ldots = \beta_n = 0$). The alternative hypothesis (H_1) is that at least one of these coefficients is not equal to zero (i.e. $\beta_j \neq 0$ for at least one explanatory variable j).

Usually, a 95% level of significance for the F-value is accepted. The F-statistics is used to test the hypothesis that all regression coefficients are jointly equal to zero or not. When the values of the coefficients are zero, this indicates that all the explanatory variables have no impact on the dependent variable. To calculate the F-value, the following formula is used:

$$F_0 = \frac{SS_R \,/\, k}{SS_E \,/\, (n - p)} = \frac{MS_R}{MS_E} \tag{4.4}$$

where:

SS_R	=	Regression Sum of Squares
SS_E	=	Error Sum of Squares
SS_T	=	Total Sum of Squares ($SS_R + SS_E$)
MS_R	=	Mean Square of Regression
MS_E	=	Mean Square of Error
n	=	Number of Observations
p	=	Number of Parameters (Coefficients)
k	=	Regression Degree of Freedom

The null hypothesis (H_0) should be rejected if the above-computed F_0 value is greater than $f_{(\alpha,k,n-p)}$. This procedure is usually summarized in an analysis of variance (ANOVA) table, such as the following Table 4.1.

Table 4.1: ANOVA Table for Testing Model Overall Significance

Source	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Regression	SS_R	k	MS_R	MS_R / MS_E
Error	SS_E	$n - p$	MS_E	
Total	SS_T	$n - 1$		

4) R-Squared, RMSE, & MAE: Goodness of Fit Measures

The **R-squared**, also known as the coefficient of determination (R^2), measures the goodness of fit of the regression model. It measures the proportion of the total variation in the dependent variable that can be explained by the explanatory variables included in the model. The value of R-squared lies between 0 and 1. A value of R-Squared close to 1 indicates that the model has a good fit, whereas a value closer to 0 indicates that the model has a poor fit. However, there is no standard on how high R^2 value is "good" enough, it usually depends on the application or the phenomena. The R-squared can be calculated as per the following equation:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{4.5}$$

where:

SS_T	=	Total Sum of Squares	=	$\Sigma(Y_i - \bar{Y})^2$
SS_R	=	Regression Sum of Squares	=	$\Sigma(\hat{Y}_i - \bar{Y})^2$
SS_E	=	Error Sum of Squares	=	$\Sigma(Y_i - \hat{Y}_i)^2$
Y_i	=	Actual Value		
$\bar{Y}$	=	Average Value of Y_i		
$\hat{Y}_i$	=	Estimated Value by the Model		

The above equation is valid when the regression model includes the 'constant' term (i.e., the model intercept "α" is not zero). However, when there is no constant in the regression, that is regression line forced through the origin, the definition of SS_T as the sum of squared deviations from the mean is inappropriate, and in order to obtain the correct R-squared value, the average ($\bar{Y}$) in the above equation must be zero, as described in details by Eisenhauer (2003).

The **Root Mean Squared Error (RMSE)** is a measure of accuracy, frequently used to assess model performance, and to compare forecasting errors of different models for a particular dataset. RMSE measures the differences between the predicted values by a model and the actual observed values.

RMSE is always non-negative and ranges from zero to infinity. A value of zero (almost never achieved in practice) would indicate a perfect fit of the model to the data. In general, the lower the value of RMSE, the better the performance of the model is expected. The RMSE can be calculated using the following equation:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2} \tag{4.6}$$

where:

RMSE	=	Root Mean Squared Error
e_i	=	Model Error $(Y_i - \hat{Y}_i)$
Y_i	=	Actual Value
$\hat{Y}_i$	=	Estimated Value by the Model
n	=	Number of Observations

Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This indicates that the RMSE is useful when large errors are particularly undesirable.

The **Mean Absolute Error (MAE)** is another useful measure widely used in model evaluation, and usually assess model performance in reflecting the actual data. MAE measures also the differences between predicted and actual values. It can be calculated using the following equation:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|e_i| = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \qquad (4.7)$$

where:

MAE	=	Mean Absolute Error
e_i	=	Model Error $(Y_i - \hat{Y}_i)$
Y_i	=	Actual Value
$\hat{Y}_i$	=	Estimated Value by the Model
n	=	Number of Observations

The MAE and the RMSE can be used together to diagnose the variation in the errors in a set of forecasts. The RMSE will always be larger or equal to the MAE; the greater the difference between them, the greater the variance in the individual errors in the sample. If the RMSE equals MAE, then all the errors are of the same magnitude. Both the MAE and RMSE can range from zero to infinity, with preferencing of lower values.

In this study, RMSE, MAE, and R-squared were employed as indicators of the model performance. They were used for the assessment, evaluation, and comparison purposes among the developed models. In general, the model with the lowest RMSE and MAE, and highest R-Squared, is the best selected one among the competing models.

4.4 Developing Trip Generation Models Using ANFIS Approach

In this step, the four intended types of trip generation models were re-developed using the ANFIS approach. The MATLAB software package was utilized for this purpose. The same dataset and the same explanatory variables, which were used previously for the MLR approach, were also re-used here in this step. The input explanatory variables must be the same as in the case of the regression model, in order to obtain meaningful, fair, and reasonable comparisons among the performance of both approaches.

Referring to Chapter Two, the ANFIS is an adaptive neural network that is functionally based on the model of Takagi–Sugeno fuzzy inference system. The fuzzy rules and the membership functions are generated from training samples, and can be adjusted during the learning process using specific learning algorithms.

The procedures for building the ANFIS consist of three main steps: 1) loading training input-output dataset, 2) generating the initial Fuzzy Inference System (FIS) and the equivalent ANFIS architecture, and then 3) utilizing the training algorithms to optimize this FIS. These steps are described in detail hereafter.

4.4.1 Loading Input-Output Training Dataset

The ANFIS training dataset is the same set that was used for the estimation and calibration process in the previous approach, which consists of 256-households sample. The input explanatory variables and the output dependent variable are also the same as used previously in the linear regression approach. However, these training samples should be loaded in the form of numerical matrix in the MATLAB environment.

4.4.2 Generating Initial Fuzzy Inference System (FIS)

The Neuro-Fuzzy Designer in MATLAB was utilized to achieve this purpose. After loading the training dataset, the initial Takagi-Sugeno fuzzy inference system could be generated using the Grid Partitioning function that is embedded in the above tool, which splits the range of each input variable into equal intervals based on the selected number of the input Membership Functions (MFs), and creates particular decision rules. The grid partitioning function generates inputs MFs by uniformly partitioning the input variables ranges, and creates one rule for each input MF combination. The consequent of each rule corresponds to a different output linear MF.

The grid partitioning function allows the user to choose the desired number and type of the input MFs that associated with each input variable. Unfortunately, there are no simple ways to determine in advance what should be the number and the type of these MFs. It usually depends on trial and error procedure. A default value of three MFs for each input variable are usually identified in MATLAB, which seems to be reasonable, as it reflects effectively the linguistic terms of low, medium, and large for example.

In this study, three MFs for each input variable were considered. Using one MF may create estimation errors, where using more than three may increase the computational cost of ANFIS based model. However, by referring to the range of the input variables in Chapter Five of this thesis, three MFs seem to be sufficient, except when the input range consists of two values (e.g., 0 and 1), where selecting two MFs will be adequate. Moreover, three popular types of MFs were used separately in this study, which are: triangle, trapezoidal, and gaussian MFs, as shown in Chapter Two, Figure 2.5.

The grid partitioning function will create one fuzzy rule with one linear output MF for each input membership function combination. For example, if there are three input variables with three MFs for each variable, 27 fuzzy rules ($3\ \mathrm{MFs} \times 3\ \mathrm{MFs} \times 3\ \mathrm{MFs} = 27$) along with 27 output linear MFs will be created automatically. The function will also create an equivalent network to the generated FIS such as the one shown in Chapter Two, Figure 2.9.

After selecting the number and the type of the MFs associated with each input variable, the initial values of promise parameters are set in such a way that the MFs are equally spaced along the operating range of each input variable. These parameters, such as (σ and m) for the gaussian MF in Chapter Two, Figure 2.5, will be optimized next using special learning algorithms as being described in the following step.

4.4.3 Optimizing the FIS Using Training Algorithms

Training is a key part of the ANFIS model development process. The major task of the training process, sometimes known as a learning process, is to optimize the fuzzy rules and the associated parameters of the input and output MFs, by minimizing the output measure of error or maximizing the performance index. Two main training algorithms could be considered, which are provided by the MATLAB, as per the following:

1. **Backpropagation learning algorithm**, based on the gradient descent minimizing of the mean square error between the perfected and actual value, for all inputs and outputs associated MFs parameters.
2. **Hybrid learning algorithm**, combines the backpropagation for the parameters associated with the input MFs, and least-squares estimation for the parameters associated with the output MFs.

At this step, it could not be known which algorithm will perform more precisely among the other, hence, both were considered separately in this study. However, Jang (1993), Jang et al. (1997), and Rutkowska (2002) provide full insight into the working mechanism of these algorithms, which is out the scope of this thesis.

The training algorithm will use the training dataset to optimize the FIS, by training the FIS several times, or several cycles (known as training epochs), until reaching the minimum possible measure of error. The MATLAB is programmed to use RMSE as a measure of error between actual and predicted values. However, as the training process continues, the measure of error will keep decreasing, until reaching either a prespecified measure of error, or a selected number of training epochs, which is satisfied first.

The minimum possible error (tolerance) is usually preferred to be zero, which could not be achieved in practice, and could not be known what it should be by simply looking at the data. However, there are no simple ways to select the optimal number of training epochs, which usually depends on the size of the training dataset, as the smaller the training dataset is, the smaller the number of learning epochs is required. In this study, the error tolerance was set to be zero for all cases, and a different number of training epochs were used separately, which are 1, 5, 10, 50, 100, 200, 500, and 1000 epochs. These numbers were selected based on personal discretion, in an attempt to reach the lowest possible value of error.

In summary, in order to obtain the optimum model configuration with the minimum possible RMSE and MAE, several options were considered for the development of ANFIS based model, which are:

1. The number of inputs MFs: three for each input variable.
2. The type of inputs MFs: gaussian, trapezoidal, and triangle.
3. The type of learning algorithm: backpropagation and hybrid algorithms.
4. The number of training epochs: 1, 5, 10, 50, 100, 200, 500, and 1000 epochs.

4.5 Models Comparison and Validation

This study aims simply to explore which is the more accurate for modeling trip generation, the MLR approach or the ANFIS approach. To achieve this, the two approaches were used to develop four categories of trip generation models (i.e., ALLTRIP, HBW, HBE, and HBO). For each model category, a comparative analysis among the performance of each model was conducted.

Three statistical evaluation criteria were considered to assess the performance and the accuracy of each model, namely the coefficient of determination (R-Squared), the Root Mean Square Error (RMSE), and the Mean Absolute Error (MAE). These measures can be calculated using Equation 4.5, 4.6, and 4.7, respectively. The approach that develops the trip generation model with higher R-Squared, and lower RMSE and MAE, could be considered the more accurate one.

Testing and validation processes are also necessary to assess the generalization ability of the developed model and its capability in predicting future behavior. This was achieved by comparing the predicted number of trips with the actual values for a sample of 53 households, that were not used in model estimation or training process. These samples constitute nearly 17% of the whole dataset. This testing dataset was utilized to verify the accuracy and effectiveness of the developed model. The estimated outputs based on the MLR and the ANFIS were compared with the actual values. The more accurate the model, the more the average, the median, and the summation of the predicted values are closer to the actual values.

Chapter Five

Data Collection

Chapter Five
Data Collection

5.1 Introduction

This study tries to simulate and model the relationships among the socioeconomic characteristics of the household (basic study unit), and the associated number of daily trips produced. The required data for this purpose were collected from a previously conducted study by Amer (2017) for Salfit City, who considered the development of trip generation models for the city as mentioned earlier.

Basically, and before collecting the required data, Amer defined the boundaries of the study area and divided it into Traffic Analysis Zones (TAZs). Then, Amer determined the needed information and the required data to be collected, estimated the required sample size, and designed a proper household questionnaire that fit with the study purpose. These procedures are described hereafter and reviewed in the following sections.

5.2 Study Area and Zoning

Amer took the advantage and benefited from the maps and master plans issued by the Municipality of Salfit, and the Palestinian Central Bureau of Statistics (PCBS), for identifying and delineating the study area boundaries. These maps take into consideration the developed and expanding areas within the city, along with the areas that have a development chance in the near future. The study was limited to the urban populated area based on these maps, which has an estimated area of 4.5 km^2.

This area was divided by Amer into six internal traffic analysis zones and five external zones by a process called zoning. This usually includes dividing study area into smaller homogeneous zones based on a set of common criteria such as the type of existing and proposed land uses (i.e., residential, commercial or industrial use), residential densities, population, and size of each zone, the physical and historical boundaries, and the most important transportation system and roads network.

The same traffic analysis zones considered by Amer were also considered here in this study, which are shown in Figure 5.1 along with the study area boundaries. Moreover, the type of land use associated with each traffic analysis zone is shown in Table 5.1.

Zoning is an essential step that has to be done prior to the application of the four-step process for urban transportation planning. The TAZ forms the basis of these steps especially for trip distribution and traffic assignment. In the trip generation step, separate models are sometimes developed for each zone. However, in this study, one model (for each trip category) was considered for the whole city. The zones were used only to ensure the random distribution of the samples over the study area, and to facilitate the data collection procedure. It was assumed that the households will produce trips in the same pattern, regardless of their location. This assumption was supported by the fact that the study area has a small population (10,911 persons) with the same culture and traditions, and that more than 80% of the study area is residential in nature, as shown in Table 5.1.

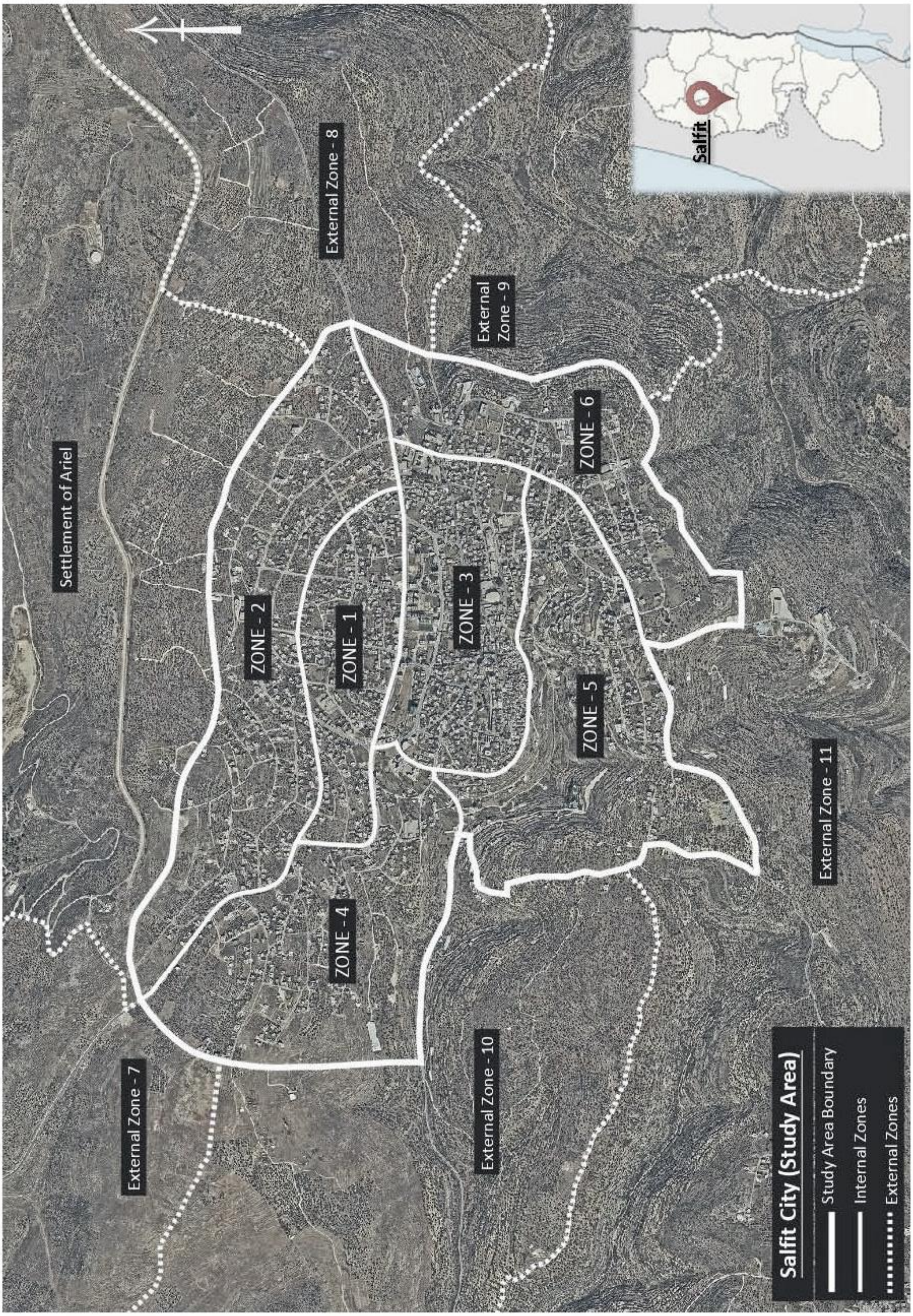

Figure 5.1: Study Area Boundaries and Traffic Analysis Zones
Source: (Amer, 2017)

Table 5.1: Traffic Analysis Zones Along with The Collected Sample Size

TAZ No.	Land Use	Area Approx. (km^2)	HH No.	% HH	Collected Sample
1	Residential	0.40	506	20%	51
2	Residential	1.00	404	16%	42
3	Commercial	0.60	581	23%	59
4	Residential	0.85	303	12%	30
5	Residential	1.00	430	17%	43
6	Residential & Industrial	0.65	303	12%	31
Total		**4.50**	**2527**	**100%**	**256**

TAZ: Traffic Analysis Zone
HH: Household or Housing Unit
Source: (Amer, 2017)

5.3 Selecting Sample Size

It is necessary to select a representative sample that reflects the performance of the population under study. Given the study's purpose, the sample size should not be too large, thus too expensive data collection process, and should not be too small, which may imply results with a high degree of variability (Ort´uzar & Willumsen, 2011). However, there are many practices that could be followed to determine the size of a representative sample. Amer considered the procedures described by the U.S. Bureau of Public Roads (BPR), which states that 10% of the dwelling units should be examined (minimum sample size), when the study area population is less than 50,000 persons.

The City of Salfit had a total population of 10,911 persons and 2,527 households in 2017 as mentioned earlier. Hence, a minimum sample size of 253 respondents should be used. However, household characteristics were collected by Amer from 309 respondents in the form of personal interviews of randomly selected households for each analysis zone. These samples were drawn from the household population between the period of early October and late November 2016 for Salfit City. The collected 309-households sample was divided into two main parts as per the following:

- A total of 256-households sample (randomly selected) was considered by Amer for estimating the models (more than the minimum), the sample size for each zone was selected based on its population and the number of households inside each zone (10% of household per zone), as shown in Table 5.1. The same data set was used here for the model's estimation and calibration (the training dataset).
- The balanced 53-households sample was randomly collected by Amer for model's verification, which represents nearly 17% of the sample size. The same set was also used here for the model's verification and validation (the testing dataset).

5.4 Collecting Required Data

For collecting the required data, Amer designed a proper household questionnaire that included two parts. The first one included the socioeconomic and demographic characteristics of the household, such as its size, type (independent or apartment), monthly income, resident age, number of persons who are employed or receiving education, vehicle ownership, and others. The second part included data regarding the associated number of daily trips produced.

For collecting and distributing the household questionnaire, Amer conducted personal face-to-face interviews with different households from different

TAZ areas. The travel data (number of trips produced) were gathered for representing typical working day in Palestine. This survey approach was used to ensure the highest response rates and data accuracy as well.

Given the purpose of this study, Table 5.2 and Table 5.3 summarize the selected explanatory variables (household characteristics) from Amer study, that were considered for estimating the intended models, followed by an illustration of the descriptive statistics associated with each of these variables, such as the mean, the median and the range, providing full insight into the nature of the study area composition.

Table 5.2: The Selected Explanatory (Independent) Variables

Variable	Description
SIZE	Number of **Persons** in the household (household size)
M	Number of **Males** in the household
F	Number of **Females** in the household
EMP	Number of persons who are **Employed** in the household
EDU	Number of persons who are **Receiving Education** in the household
AGE_A	Number of persons who are **under 16** years in the household
AGE_B	Number of persons who are **between 17 and 30** years in the household
AGE_C	Number of persons who are **between 31 and 50** years in the household
AGE_D	Number of persons who are **between 51 and 64** years in the household
AGE_E	Number of persons who are **above 65** years in the household
DRIVE	Number of **Licensed Drivers** in the household
CAR	Number of **Cars** owned by a household
BICY	Number of **Bicycles** owned by a household
MCYC	Number of **Motorcycles** owned by a household
INC	Monthly household **Income** (Thousand New Israeli Shekel)
HHTYP	**House Type**: 1 if Independent Residence, 0 if Apartment

Table 5.3: Descriptive Statistics for the Selected Explanatory Variables

Variable	Average	Standard Deviation	Median	Max.	Min.	Range	Sample Size
SIZE	3.99	1.749	4	9	1	8	309
M	2.01	1.181	2	6	0	6	
F	1.98	1.096	2	6	1	5	
EMP	1.38	0.899	1	4	0	4	
EDU	1.63	1.529	1	7	0	7	
AGE_A	1.24	1.373	1	6	0	6	
AGE_B	1.14	0.977	1	4	0	4	
AGE_C	0.92	0.805	1	2	0	2	
AGE_D	0.51	0.728	0	3	0	3	
AGE_E	0.19	0.485	0	2	0	2	
DRIVE	1.18	0.964	1	6	0	6	
CAR	0.54	0.605	0	3	0	3	
BICY	0.05	0.237	0	2	0	2	
MCYC	0.02	0.126	0	1	0	1	
INC	4.740	2.993	4.000	30.000	0.500	29.500	
HHTYP	**83.8%** Independent House / **16.2%** Apartment						

- **Gender Distribution: 50.4%** Males / **49.6%** Females
- **Age Distribution: Under 16:** (31.0%) **/ 17 – 30:** (28.4%) **/ 31 – 50:** (22.9%) **/ 51 – 64:** (12.7%) **/ Above 65:** (4.9%)
- **Vehicle Ownership Distribution: Cars:** (89.7%) **/ Bicycles:** (7.6%) **/ Motorcycles:** (2.7%)

Table 5.4 and Table 5.5 summarize the selected dependent variables that were desired for modeling from Amer study, and the associated descriptive statistics for each home-based trip category, that were derived from the collected dataset.

Based on Table 5.5, it could be noted that the ALLTRIP model will estimates all daily trips produced. At least one trip per day will be generated as minimum possible value. However, the HBW, HBE and HBO models constitute nearly 25%, 25%, and 50%, respectively, of the total daily trips, with zero trips per day as a minimum possible value of each category.

Table 5.4: The Selected Dependent Variables

Variable	Description
ALLTRIP	Number of daily **Total** (General) trips produced by household
HBW	Number of daily **Work** trips produced by household
HBE	Number of daily **Educational** trips produced by household
HBO	Number of daily **Other** trips produced by household

Table 5.5: Descriptive Statistics for the Selected Dependent Variables

Variable	Average	Standard Deviation	Median	Max.	Min.	Range	%	Sample Size
ALLTRIP	6.42	3.050	6	18	**1**	**17**	**100.0%**	309
HBW	1.56	1.044	2	5	**0**	**5**	**24.4%**	
HBE	1.59	1.546	1	7	**0**	**7**	**24.8%**	
HBO	3.26	1.774	3	11	**0**	**11**	**50.8%**	

Note: $\mathrm{ALLTRIP} = \mathrm{HBW} + \mathrm{HBE} + \mathrm{HBO}$

Chapter Six

Models Development and Comparison

Chapter Six
Models Development and Comparison

6.1 Introduction

In this chapter the four intended types of trip generation models for Salfit City were developed and validated using the two main approaches (i.e., MLR and ANFIS), followed by a comparison between their performance. Each of the following sections considers one type of these models, which are the ALLTRIP, the HBW, the HBE, and the HBO trips generation models.

For the MLR approach, different statistical tests were performed to ensure the selection of the most relevant explanatory variables, and the best estimation of the intended model. Furthermore, several configurations for the ANFIS were considered, in order to obtain the optimum possible structure with lowest measure of error.

6.2 General Trip Generation Model (ALLTRIP)

The ALLTRIP model intended to estimate the number of total daily trips produced by a household during typical working day. The estimated number of trips will include all possible trips (100% of the trips) that are generated regardless of their types or purposes. The predicted number of trips should lie between one trip per day as a minimum possible value, and up to 18 trips per day as a maximum value, based on the collected data as illustrated in Chapter Five, Table 5.5. However, the following sub-sections illustrate the development process for this model considering both the MLR and ANFIS.

6.2.1 Developing ALLTRIP Model Using Multiple Linear Regression

The stepwise MLR analysis was utilized for this purpose. The following equation suggests the best explanatory variables, among others, that would explain the number of total daily general trips produced by a household:

$$\mathrm{ALLTRIP} = 2.388 + 1.046\ \mathrm{EMP} + 1.237\ \mathrm{EDU} + 0.424\ \mathrm{DRIVE} \quad (6.1)$$

where:

ALLTRIP	=	Number of daily general trips produced by household
EMP	=	Number of household persons who are employed
EDU	=	Number of household persons who are receiving education
DRIVE	=	Number of licensed drivers in the household

The above equation represents the best estimation for the general trip generation model. However, the estimated model by Amer (2017), was based only on the EMP and EDU variables. In this study, an additional variable was used (i.e., DRIVE), which improved the R-Squared value, from 0.643 as in Amer's model to 0.659. The regression results for this model are summarized in Table 6.1.

Table 6.1: Regression Results for the General Trip Generation Model (ALLTRIP)

Model Summary					
Variables	**Coefficient**	**Standard Error**	**t-value**	**Sig.**	**VIF**
Intercept	2.388	0.223	10.721	0.000	
EMP	1.046	0.132	7.945	0.000	1.287
EDU	1.237	0.073	16.926	0.000	1.029
DRIVE	0.424	0.123	3.452	0.001	1.262
R-Squared: 0.659		**Adjusted R-Squared: 0.654**		**RMSE: 1.7112**	
MAE: 1.3882		**F-value: 161.982**		**Sample Size: 256**	

Interpretation of Regression Coefficients

Referring to Equation (6.1), the number of daily general trips produced by a household tends to increase with the increase in the household persons who are employed, receiving education, or having a driving license. There is direct proportional and logical relationship exists between the dependent and explanatory variables, as indicated by the positive sign associated with each coefficient in Equation (6.1).

Testing Coefficients: T-Test Individual

Referring to Table 6.1, the t-values for each coefficient included in the model are larger than two, which indicates that the selected explanatory variables are statistically significant (different from zero) at a 95% level of significance. More precisely, the t-value associated with the intercept, EMP, EDU, and DRIVE are 10.721, 7.945, 16.926, and 3.452, respectively, indicating a 99.99% level of significance. Hence, the null hypotheses that these explanatory variables have no effect on ALLTRIP are rejected, and the alternative hypotheses that each of the EMP, EDU, and DRIVE variables are positively correlated to ALLTRIP are accepted.

Testing for Multicollinearity: Pearson's Correlation and VIF

Referring to Appendix A, Table A-I: Pearson's Correlation Matrix, which was considered while developing the regression models, the correlation among EMP & EDU, EMP & DRIVE, and EDU & DRIVE are 0.166 (small), 0.455 (medium), and 0.088 (small), respectively. No large correlation exists among any of these variables, as the correlation values are less than 0.500, which are not severe and could be accepted.

The Variance Inflation Factors (VIF) for each explanatory variable included in the model are shown in Table 6.1. All of these values are less than 10, which indicates, in addition to the correlation values, that there are no multicollinearity problems in the estimated model.

Testing Goodness of Fit: R-Squared

The R-Squared and the adjusted R-Squared for the estimated model are 0.659, and 0.654, respectively. The R-Squared value implies that the explanatory variables (EMP, EDU, and DRIVE) explained nearly 65% of the variation in the dependent variable (ALLTRIP). Such value could be considered reasonable and shows a good explanation of data variability.

It is good to mention that the value of the R-Squared could be raised to 0.913, by simply not including intercept in the model. In other words, performing the regression through the origin will highly improve the R-Squared value. However, this approach was not considered here for three reasons.

First, the improvement in the R-Squared does not necessary imply an improvement in the model performance, the values of the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) will be raised from 1.7112 and 1.3882 to 2.0604 and 1.6707, respectively, which is not preferred.

Second, it could not be declared that the household will produce no trips per day if the selected variables (i.e., EMP, EDU, and DRIVE) were set to be zero. Moreover, and based on the collected dataset, a minimum of one possible trip could be generated by a household during the day, regardless of its characteristics.

Third, the existence of the intercept is significant at a 99.99% level of significance, as indicated by the associated t-value in Table 6.1, which is 10.721.

Testing Overall Significance: F-Test

Referring to Table 6.1, the F-value for the estimated model is 161.982. Considering this high value, the null hypothesis that the explanatory variables EMP, EDU, and DRIVE have no impact on the total number of daily trips produced by a household (all equal zero) is rejected statistically at the 99.99% level of significance. Consequently, the alternative hypothesis that these variables jointly affect ALLTRIP is accepted.

The F-value for the estimated model can be obtained from Table 6.2, which presents the analysis of variance (ANOVA) of the ALLTRIP model, summarizes the procedure for calculating this value.

Table 6.2: ANOVA Table for General Trip Generation Model (ALLTRIP)

Source	Sum of Squares	df	Mean Square	F-Value	Sig.
Regression	1445.506	3	481.835	161.982	0.000
Residual	749.604	252	2.975		
Total	2195.109	255			

6.2.2 Developing ALLTRIP Model Using ANFIS Approach

For the development of the general trip generation model (ALLTRIP) using the ANFIS approach, the same dataset (i.e., 256-households sample) and the same explanatory variables (i.e., EMP, EDU, and DRIVE), which were considered by the MLR approach, were also used here in this approach, to obtain fair and reasonable comparison of their performance.

In order to obtain the optimum configuration of the ANFIS, that is intended to estimate the number of total daily trips produced by a household, with lowest possible RMSE and MAE, different design options and combinations were evaluated, as shown in Table 6.3, which are:

1. Number and type of inputs Membership Functions (MFs): Three types were considered: gaussian, trapezoidal, and triangle. For each input, three MFs were selected.
2. Type of learning algorithm: backpropagation, and hybrid algorithms.
3. Number of training epochs: 1, 5, 10, 50, 100, 200, 500, and 1000 epochs were considered.

Table 6.3: Resulted RMSE and MAE for Different ANFIS Configurations – ALLTRIP

Learning Algorithm	Index	Training Epochs							
		1	5	10	50	100	200	500	1000
Gaussian Membership Functions (3 MFs for each input)									
Back-propagation	RMSE	6.9612	6.8865	6.7931	6.0037	4.7763	2.4853	1.6729	1.6653
	MAE	6.3188	6.2504	6.1646	5.4286	4.2249	1.9868	1.3269	1.3230
Hybrid	RMSE	1.5114	1.5108	1.5101	1.5068	1.5028	1.4970	1.4895	**1.4880**
	MAE	1.1504	1.1486	1.1467	1.1393	1.1340	1.1288	1.1220	**1.1203**
Trapezoidal Membership Functions (3 MFs for each input)									
Back-propagation	RMSE	6.9590	6.8757	6.7718	5.9453	4.8184	2.9659	1.6110	1.5769
	MAE	6.3171	6.2419	6.1476	5.3760	4.2410	2.2799	1.2937	1.2599
Hybrid	RMSE	1.5453	1.5451	1.5448	1.5434	1.5400	1.5288	1.5288	1.5288
	MAE	1.2063	1.2061	1.2059	1.2017	1.1943	1.1825	1.1825	1.1825
Triangle Membership Functions (3 MFs for each input)									
Back-propagation	RMSE	6.9604	6.8825	6.7854	6.0190	5.0845	3.3590	1.6817	1.5533
	MAE	6.3181	6.2467	6.1573	5.4380	4.5152	2.7353	1.3532	1.2389
Hybrid	RMSE	1.5289	1.5196	1.5178	1.5152	1.5130	1.5130	1.5130	1.5130
	MAE	1.1896	1.1751	1.1689	1.1649	1.1616	1.1616	1.1616	1.1616

Referring to Table 6.3, the optimum structure of the ANFIS was achieved first at three gaussian membership functions for each input variable, hybrid learning algorithm, and 1000 training epochs.

An initial Takagi–Sugeno fuzzy inference system was first developed, using the same input variables that were used in the regression approach, which are the EMP, EDU, and DRIVE.

Three gaussian membership functions for each input variable were then identified. Consequently, 27 fuzzy rules (i.e., $3 \times 3 \times 3 = 27$) along with 27 output linear MFs were created. The associated parameters with the inputs and outputs MFs were optimized at the end using the hybrid learning algorithm for 1000 training epochs.

Table 6.4 summarizes the results associated with the optimized FIS for estimating the ALLTRIP. The values of R-Squared, RMSE, and MAE were 74.18%, 1.488, and 1.1203, respectively. However, the optimized Takagi–Sugeno FIS is presented in Figure 6.1, which sometimes denoted by ANFIS Rule Viewer. The Equivalent ANFIS architecture for the developed FIS is shown in Figure 6.2.

Table 6.4: ANFIS Optimum Configuration Summary for ALLTRIP

R-Square: 74.18%		**RMSE: 1.4880**
MAE: 1.1203		**Sample Size: 256**
ANFIS Sum of Squares:	1628.324	
Residual Sum of Squares:	566.786	
Total Sum of Squares:	2195.109	

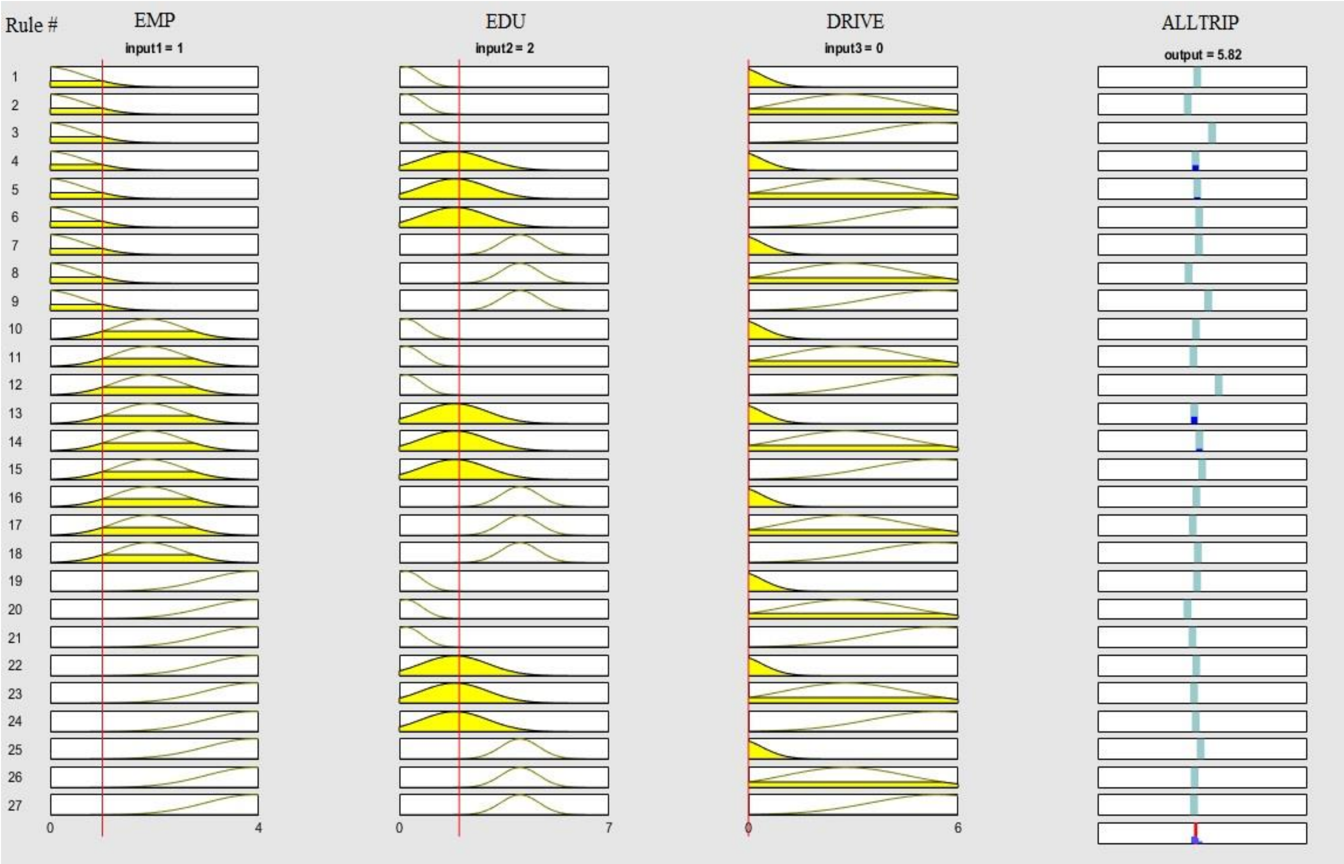

Figure 6.1: The Optimized FIS for Estimating ALLTRIP (ANFIS Rule Viewer)

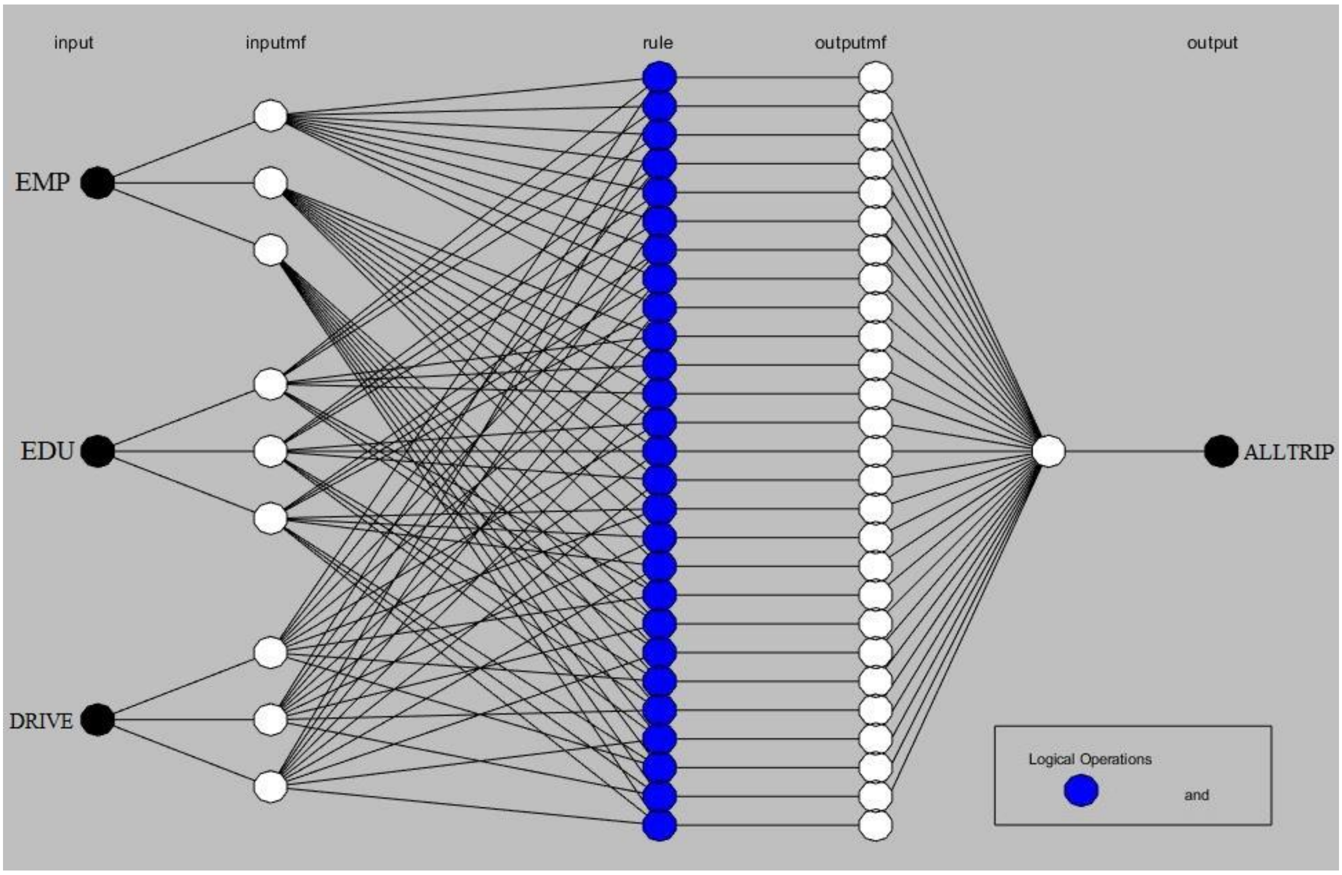

Figure 6.2: The Equivalent ANFIS Architecture for ALLTRIP Model

Three gaussian MFs were associated with each input variable. Figure 6.3, 6.4, and 6.5 illustrate these MFs for EMP, EDU, and DRIVE input variables, respectively. In each figure, the initial and the optimized parameters of these MFs (before and after training) were illustrated. The gaussian MF can be determined using two parameters (i.e., σ, and m). For example, the initial parameters of the second MF associated with the EDU variable (EDUmf2) were [1.486, 3.5], and after training, they were optimized to [1.093, 1.844].

Figure 6.3: Optimized Gaussian MFs for EMP Input Variable

Figure 6.4: Optimized Gaussian MFs for EDU Input Variable

6.2.3 Models Comparison and Validation

Performance Comparison

Table 6.5 illustrates a comparison of the performance of these two approaches based on the R-Squared, RMSE, and MAE measures. It could be concluded that the ANFIS can be used for modeling ALLTRIP more accurate with better performance compared with the MLR approach.

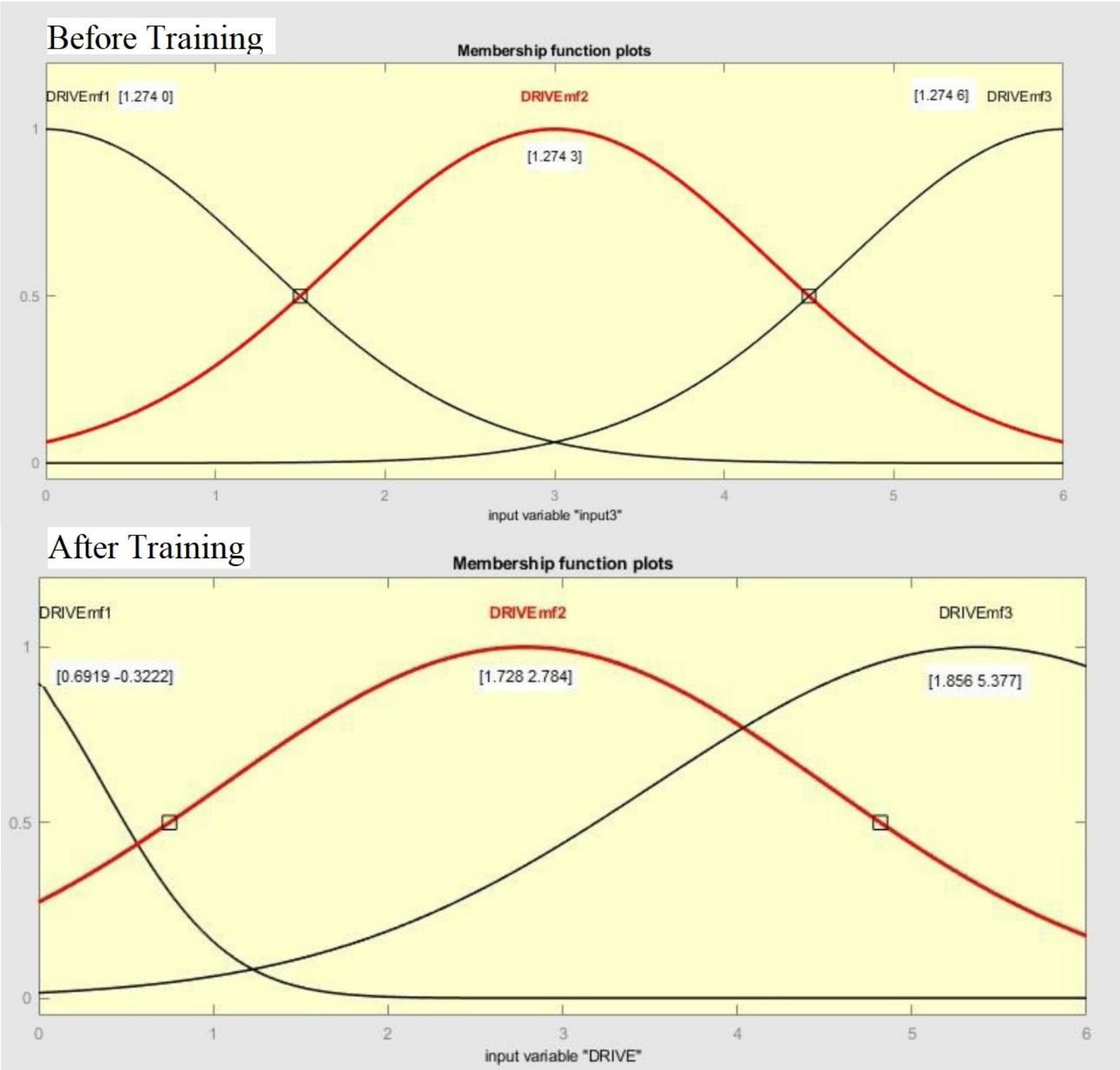

Figure 6.5: Optimized Gaussian MFs for DRIVE Input Variable

Table 6.5: ALLTRIP Performance Comparison Among MLR and ANFIS

	MLR	ANFIS	Difference
RMSE	1.7112	**1.488**	-0.2232
MAE	1.3882	**1.1203**	-0.2679
R-Squared	65.85%	**74.18%**	+8.33%
Residual Sum of Squares	749.604	**566.786**	-182.812

Referring to Table 6.5, the R-Squared value for ALLTRIP was raised from 65.85% by using MLR to 74.18% by using ANFIS, and the explanatory power for the variation in the ALLTRIP was improved by 8.33% accordingly. Furthermore, the residual sum of squares, the RMSE, and the MAE were all reduced form 749.604, 1.7112, and 1.3882, to 566.786, 1.488, and 1.1203, respectively. However, and based on these results, the ANFIS can be used for developing more accurate related model with better performance than the MLR approach.

Models Validation

For validation purpose, the additional 53-households testing sample was considered. A comparison among the actual and predicted ALLTRIP for this testing sample, using both MLR and ANFIS, is presented in Appendix B, Table B-I.

For both modeling approaches, the predicted results and the actual values are within acceptable conformity, with better results associated with ANFIS. This testing sample was not included in the model estimation, calibration, or training processes. They were used here as a testing device to assess the prediction capability of the estimated models.

Table 6.6 summarizes a comparison of several statistical measures between the actual and predicted values of this sample. The means of the predicted values using MLR and ANFIS (μ_{MLR} & μ_{ANFIS}) were compared with the mean of the actual values (μ_{Actual}) using the t-test. The results indicated that the null hypothesis H_0: $\mu_{\mathrm{Actual}} = \mu_{\mathrm{ANFIS}}$ and $\mu_{\mathrm{Actual}} = \mu_{\mathrm{MLR}}$ could not be statistically rejected, and the two means are not significantly different from

μ_{Actual} at 90% level of significance (i.e., $t_{statistic} < t_{critical\,(\alpha=10\%)}$ for MLR and ANFIS).

The ANFIS developed more accurate model, as the predicted mean and standard deviation using ANFIS were closer to the actual values than with the MLR. Furthermore, the difference between the summation of the actual ALLTRIP and of the related predicted values using ANFIS (9.34 trip) were smaller as compared with the MLR (19.65 trips). In addition to the other measures, which indicate that the ANFIS can provide more accurate and closer predictions to actual values than the MLR.

Table 6.6: Actual and Predicted ALLTRIP Descriptive Statistics

Validation (Statistics) / Testing Dataset					
	Actual		**MLR**		**ANFIS**
Mean	**6.81**		**6.44**		**6.64**
St. Deviation	3.563		2.639		3.026
Median	6		6.14		6.47
Sum	361		341.35		351.66
Difference	0		**19.65**		**9.34**
Count	53		53		53
t-test	t-critical =1.66	>	t-statistic = 0.608	>	t-statistic = 0.265
	α = 0.1	<	P-value = 0.5448	<	P-value = 0.7917

6.3 Home-Based Work Trip Generation Model (HBW)

This model intended to estimate the number of daily work trips generated by a household during a typical working day, the home-based work trip generation model (HBW). Based on the collected data, the home-based work trips constitute nearly 25% of all daily generated trips, which range from zero trips to a maximum of 5 trips per day. The same steps followed for developing the ALLTRIP model will be repeated in this section.

6.3.1 Developing HBW Model Using Multiple Linear Regression

Using stepwise MLR analysis, Equation 6.2 represents the best estimated HBW trips generation model. The regression results are summarized in Table 6.7.

$$\mathrm{HBW} = 0.875\ \mathrm{EMP} + 0.153\ \mathrm{AGE_B} + 0.280\ \mathrm{CAR} \qquad (6.2)$$

where:

HBW	=	Number of daily work trips produced by household
EMP	=	Number of household persons who are employed
AGE_B	=	Number of young household persons between (17 & 30 years)
CAR	=	Number of cars owned by a household

The only difference between this model and the estimated one by Amer (2017), is that this model has no constant. Removing the constant increases the value of R-Squared from 0.695 as in Amer's model, to 0.904 as in this model, which is logical as will be described later.

Table 6.7: Regression Results for Home-Based Work Trip Generation Model

Model Summary					
Variables	**Coefficient**	**Standard Error**	**t-value**	**Sig.**	**VIF**
Intercept	0.000	N/A	N/A	N/A	
EMP	0.875	0.041	21.605	0.000	1.312
AGE_B	0.153	0.040	3.813	0.000	1.228
CAR	0.280	0.062	4.534	0.000	1.086
R-Square: 0.904		**Adjusted R-Square: 0.899**		**RMSE: 0.5932**	
MAE: 0.4163		**F-value: 790.181**		**Sample Size: 256**	

Interpretation: Referring to Equation 6.2, the number of daily work trips produced by a household tends to increase with the increase in household EMP, AGE_B, and CAR. There is a direct proportional-logical relationship exists between the dependent and explanatory variables, as indicated by the positive sign associated with each coefficient in Equation 6.2 as expected.

Testing Coefficients: Referring to Table 6.7, the coefficients of EMP, AGE_B, and CAR are significant at a 99.99% level of significance, as indicate by the associated t-value.

Testing for Multicollinearity: Referring to Appendix A, Table A-I: Pearson's Correlation Matrix, the correlation among EMP & AGE_B, EMP & CAR, and AGE_B & CAR are 0.431 (medium), 0.281 (small), and 0.127 (small), respectively, which is accepted as all are less than 0.500.

The Variance Inflation Factors (VIF) for each explanatory variable included in the model are shown in Table 6.7. All of these values are less than 10, which indicates, in addition to the correlation values, that there are no multicollinearity problems in the estimated model.

Testing Goodness of Fit: The R-squared value for the estimated model is 0.904, which indicates that the variables (EMP, AGE_B, and CAR) can explain nearly 90% of the variation in the dependent variable (HBW).

The intercept (constant) was excluded from this model (i.e., considered to be zero) for three reasons. First, an improvement in the R-square value from 0.695 to 0.904 will be gained by simply removing the intercept. Second, it is logical that the household could produce zero work trip during the day. Third, removing the intercept will improve the coefficients t-value for EMP, AGE_B, and CAR, from 18.194, 2.983, and 3.952, to 21.605, 3.813, and 4.534, respectively, where the intercept, which was equal to 0.149, was having the smallest t-value among the other variables, which is 2.045.

Testing Overall Significance of Model: Referring to Table 6.7, the F-value for the estimated model is 790.181. Considering this high value, the alternative hypothesis that the EMP, AGE_B, and CAR are jointly affecting the HBW is accepted at the 99.99% level of significance. The calculation of the F-value for the estimated model can be obtained from Table 6.8.

Table 6.8: ANOVA Table for Home-Based Work Trip Generation Model

Source	Sum of Squares	df	Mean Square	F-value	Sig.
Regression	843.930	3	281.310	790.181	0.000
Residual	90.070	253	0.356		
Total	934.000	256			

6.3.2 Developing HBW Model Using ANFIS

For the development of the HBW trips generation model using ANFIS, the same dataset and the same explanatory variables, as shown in Table 6.7, were used here for the training process. However, Table 6.9 illustrates a comparison among several design options for obtaining the optimum configuration of ANFIS for HBW.

Referring to Table 6.9, the optimum structure of the ANFIS was achieved first at three gaussian MFs for each input variable (i.e., EMP, AGE_B, and

CAR), hybrid learning algorithm, and only 1 training epochs. The same error was obtained also at three triangle MFs for each input variable, hybrid learning algorithm, but after 5 training epochs. However, the first configuration was considered here.

Table 6.9: Resulted RMSE and MAE for Different ANFIS Configurations - HBW

Learning Algorithm	Index	Training Epochs							
		1	5	10	50	100	200	500	1000
Gaussian Membership Functions (3 for each input)									
Back-propagation	RMSE	1.8965	1.8422	1.7747	1.2322	0.7376	0.5735	0.5642	0.5534
	MAE	1.5741	1.5270	1.4678	0.9628	0.5472	0.4044	0.3938	0.3860
Hybrid	RMSE	0.5465	0.5465	0.5465	0.5465	0.5465	0.5465	0.5465	0.5465
	MAE	0.3664	0.3664	0.3664	0.3664	0.3664	0.3664	0.3664	0.3664
Trapezoidal Membership Functions (3 for each input)									
Back-propagation	RMSE	1.8950	1.8351	1.7610	1.2075	0.7428	0.5702	0.5598	0.5558
	MAE	1.5730	1.5212	1.4564	0.9404	0.5417	0.4037	0.3964	0.3872
Hybrid	RMSE	0.5484	0.5483	0.5481	0.5475	0.5471	0.5471	0.5471	0.5471
	MAE	0.3727	0.3728	0.3729	0.3720	0.3709	0.3709	0.3709	0.3709
Triangle Membership Functions (3 for each input)									
Back-propagation	RMSE	1.8954	1.8370	1.7649	1.2377	0.7916	0.5781	0.5530	0.5483
	MAE	1.5730	1.5212	1.4565	0.9581	0.5871	0.4103	0.3845	0.3772
Hybrid	RMSE	0.5466	0.5465	0.5465	0.5465	0.5465	0.5465	0.5465	0.5465
	MAE	0.3668	0.3664	0.3664	0.3664	0.3664	0.3664	0.3664	0.3664

After the development of the initial FIS, and identifying the MFs, 27 fuzzy rules along with 27 output linear MFs were created consequently, to be optimized during the learning process.

Table 6.10 summarizes the results associated with the optimized FIS. The values of R-Squared, RMSE, and MAE were 92.74%, 0.5465, and 0.3664, respectively. However, the optimized Takagi–Sugeno FIS for HBW is presented in Figure 6.6, and the equivalent ANFIS architecture for the developed FIS is shown in Figure 6.7.

Table 6.10: ANFIS Optimum Configuration Summary for HBW

R-Square: 92.74%		**RMSE: 0.5465**
MAE: 0.3664		**Sample Size: 256**
Model Sum of Squares:	866.152	
Residual Sum of Squares:	76.456	
Total Sum of Squares:	934	

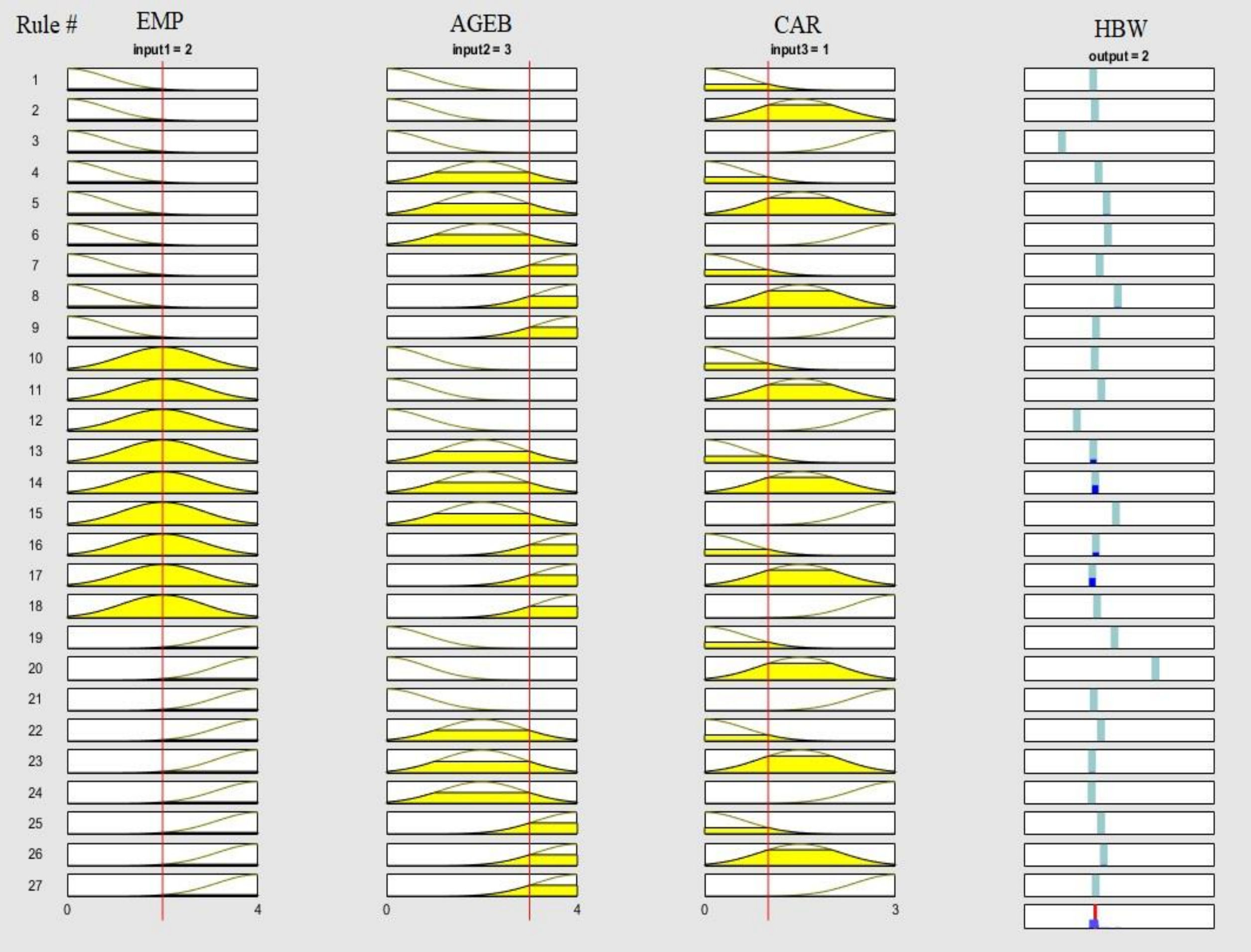

Figure 6.6: The Optimized FIS for Estimating HBW (ANFIS Rule Viewer)

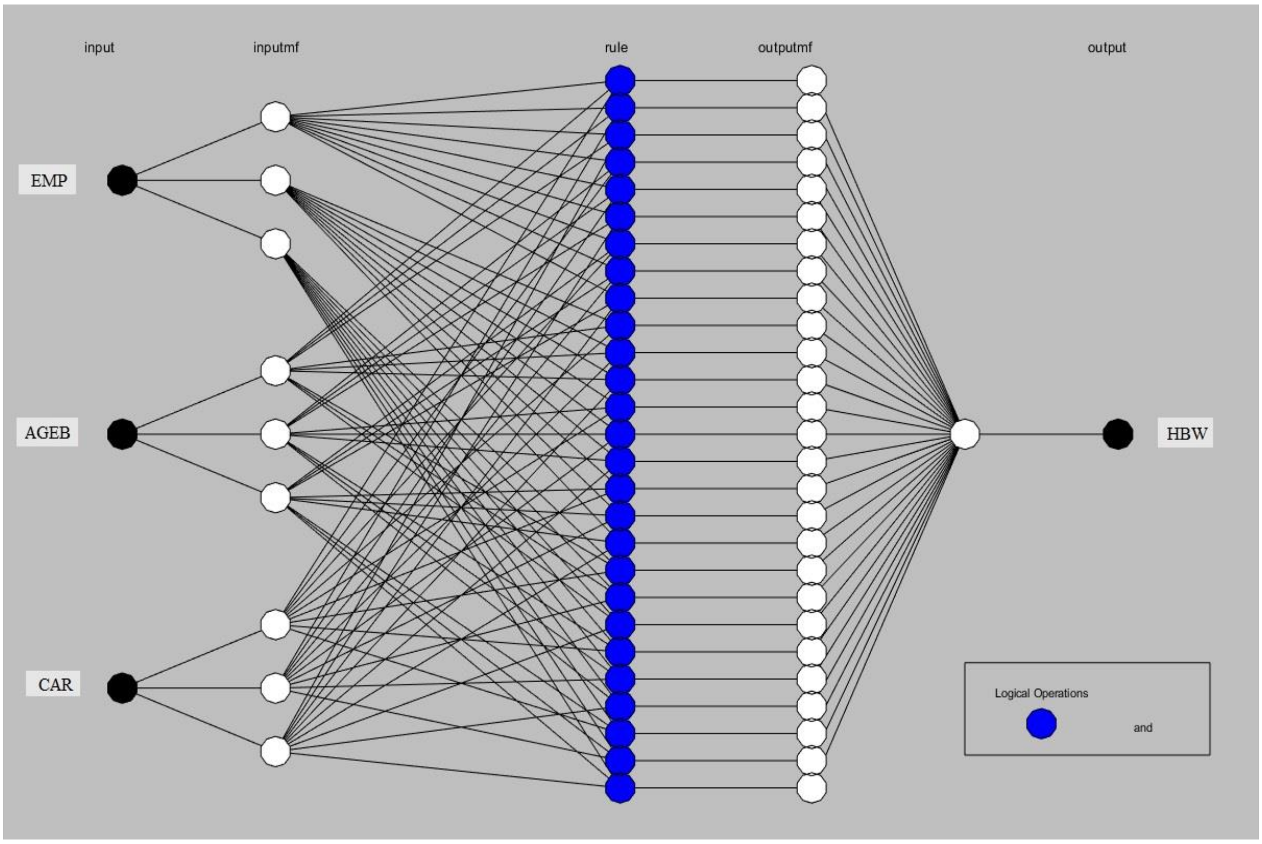

Figure 6.7: The Equivalent ANFIS Architecture for HBW Model

Three gaussian MFs were associated with each input variable. Figure 6.8, 6.9, and 6.10 illustrate these MFs for EMP, AGE_B, and CAR input variables, with their optimized parameters [σ, m], respectively.

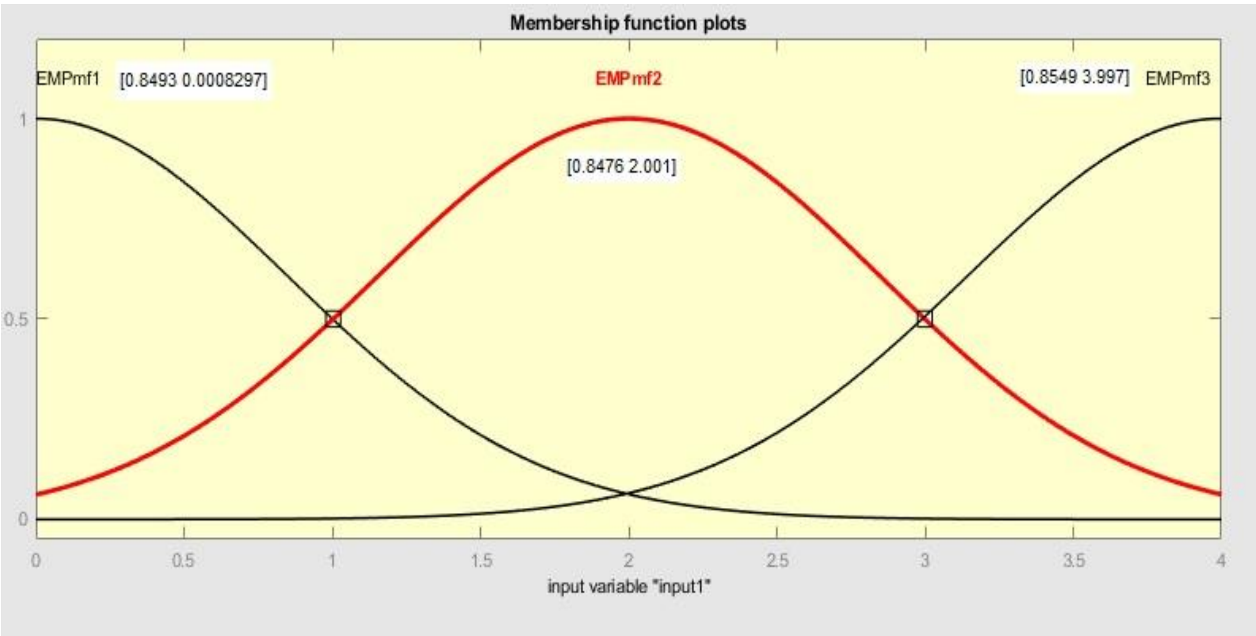

Figure 6.8: Optimized Gaussian MFs for EMP Input Variable

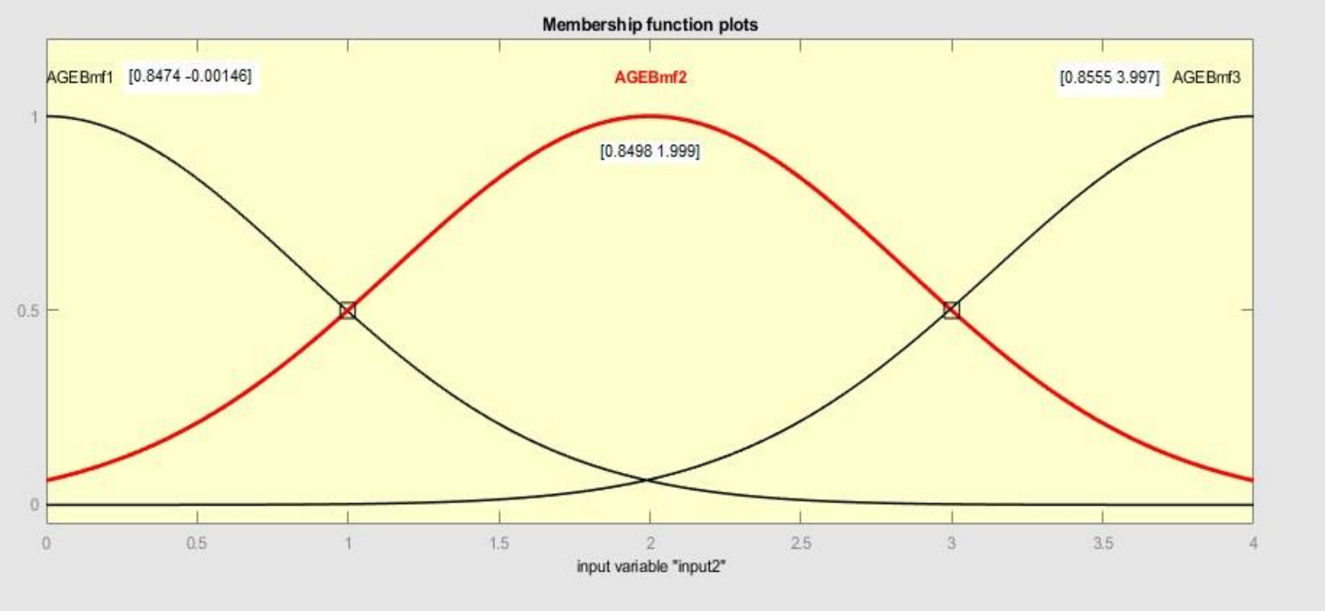

Figure 6.9: Optimized Gaussian MFs for AGE_B Input Variable

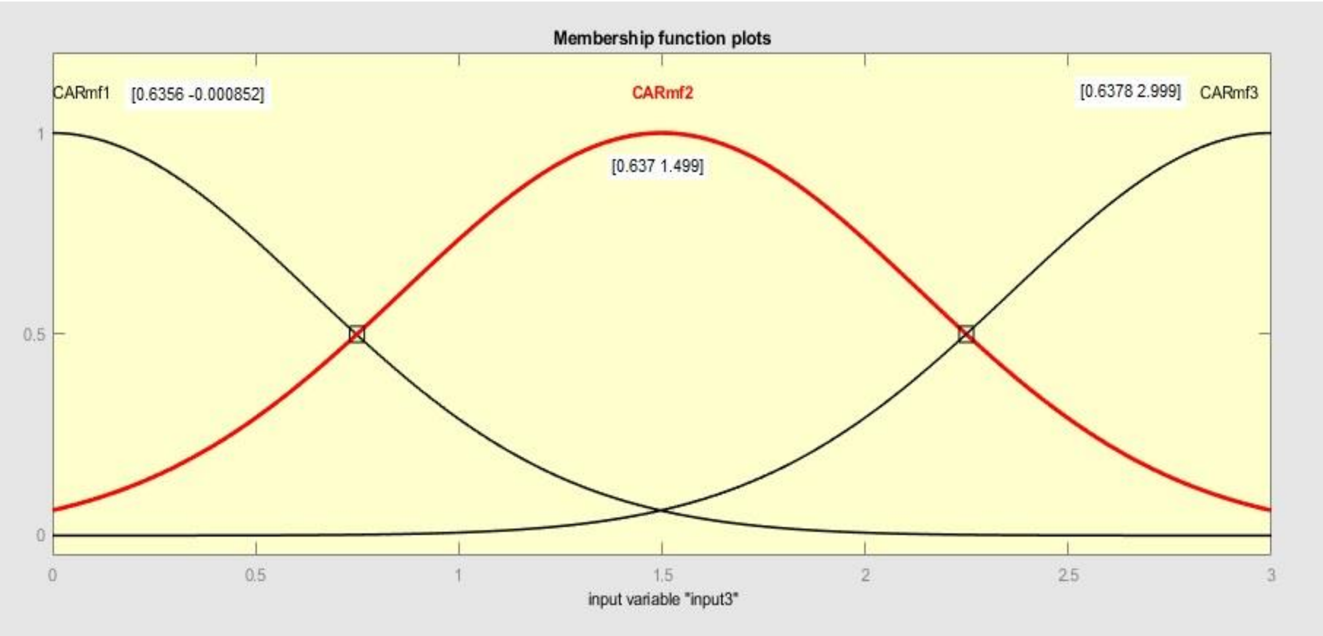

Figure 6.10: Optimized Gaussian MFs for CAR Input Variable

6.3.3 Models Comparison and Validation

Performance Comparison

Table 6.11 illustrates a comparison among the performance of these two approaches based on the R-Squared, RMSE, and MAE measures. It could be noticed that both approaches are performing nearly the same, with little improvement achieved using the ANFIS.

Table 6.11: HBW Performance Comparison Among MLR and ANFIS

	MLR	ANFIS	Difference
RMSE	0.5932	**0.5465**	-0.0467
MAE	0.4163	**0.3664**	-0.0499
R-Squared	90.36%	**92.74%**	+2.38%
Residual Sum of Squares	90.070	**76.456**	-13.614

The R-Squared value for HBW was raised from 90.36% by using MLR to 92.74% by using ANFIS, a small improvement was detected by only 2.38%. Furthermore, the residual sum of squares, the RMSE, and the MAE were all reduced form 90.070, 0.5932, and 0.4163, to 76.456, 0.5465, and 0.3664, respectively. The differences among these measures are small and could be neglected.

Models Validation

A comparison among the actual and predicted HBW for the testing sample, using both MLR and ANFIS, is presented in Appendix B, Table B-II. For both modeling approaches, the predicted results and the actual values are within acceptable conformity, with better results associated with MLR.

Table 6.12, summarizes several statistical measures for the actual and predicted values. The means of the predicted values using MLR and ANFIS (μ_{MLR} & μ_{ANFIS}) were compared with the mean of the actual values (μ_{Actual}) using the t-test. The results indicated that the null hypothesis H_0: $\mu_{\mathrm{Actual}} = \mu_{\mathrm{ANFIS}}$ and $\mu_{\mathrm{Actual}} = \mu_{\mathrm{MLR}}$ could not be statistically rejected, and hence, the μ_{Actual} is not significantly different from both μ_{MLR} and μ_{ANFIS} at 90% level of significance (i.e., $t_{\text{statistic}} < t_{\text{critical}\ (\alpha = 10\%)}$).

The prediction capabilities of the MLR is better than the ANFIS, as μ_{MLR} was much closer to μ_{Actual} than the μ_{ANFIS}, and the difference between the summation of the actual HBW and of the related predicted values using MLR (-0.72 trip) was much smaller as compared with the ANFIS (-4.66 trip), in addition to the other measures.

Table 6.12: Actual and Predicted HBW Descriptive Statistics

Validation (Statistics) / Testing Dataset					
	Actual		**MLR**		**ANFIS**
Mean	**1.45**		**1.47**		**1.54**
St. Deviation	0.932		0.764		0.739
Median	1		1.31		1.58
Sum	77		77.72		81.66
Difference	0		**-0.72**		**-4.66**
Count	53		53		53
t-test	t-critical =1.66	>	t-statistic = 0.121	<	t-statistic = 0.551
	α = 0.1	<	P-value = 0.9041	>	P-value = 0.5829

Based on these results, the ANFIS provides neglected improvement in the model performance, which is not enough to say that the ANFIS can outperform or even be better than the MLR. Furthermore, the validation assessment indicates that the MLR can provide more precise and closer predictions to actual values than the ANFIS. It seems that when dealing with systems that have small data range, such as in this case, the MLR will be enough for developing efficient models, that fit the data in a very good way.

6.4 Home-Based Education Trip Generation Model (HBE)

This model intended to estimate the number of daily education trips generated by a household during typical working day, i.e., home-based education (HBE) trips generation model. Based on the collected data, the HBE constitute nearly 25% of total daily trips, which ranges from zero up to a maximum of 7 trips per day.

6.4.1 Developing HBE Model Using Multiple Linear Regression

Equation 6.3 represents the best estimated HBE trips generation model. The regression results are summarized in Table 6.13. However, this model is exactly the same as the one developed previously by Amer (2017). It was found that the number of persons who are receiving education in the household (EDU) is the most relevant factor.

$$\mathrm{HBE} = 0.984\ \mathrm{EDU} \tag{6.3}$$

where:

HBE = Number of daily education trips produced by household

EDU = Number of household persons who are receiving education

Table 6.13: Regression Results for Home-Based Education Trip Generation Model

Model Summary					
Variables	**Coefficient**	**Standard Error**	**t-Value**	**Sig.**	**VIF**
Intercept	0.000	N/A	N/A	N/A	
EDU	0.984	0.011	85.556	0.000	1.000
R-Square: 0.966		**Adjusted R-Square: 0.962**		**RMSE: 0.4035**	
MAE: 0.1318		**F-value: 7319.795**		**Sample Size: 256**	

Referring to Equation (6.13), the HBE trips is highly and positively correlated with the household EDU, as expected and indicated by the positive coefficient sign. The coefficient of EDU is significant at a 99.99% level of significance, as indicated by the associated t-value. The R-square value is 0.966, which indicates that the EDU can explain nearly 97% of the variation in the dependent variable (HBE). Such value shows a very good explanation of data variability.

The F-value for the estimated model is 7319.795. Hence, the hypothesis that the EDU has no impact on the HBE is rejected at the 99.99% level of significance. The calculation of the F-value for the estimated model can be obtained from the following Table 6.14.

Table 6.14: ANOVA Table for Home-Based Education Trip Generation Model

Source	Sum of Squares	df	Mean Square	F-value	Sig.
Regression	1196.324	1	1196.324	7319.795	0.000
Residual	41.676	255	0.163		
Total	1238	256			

6.4.2 Developing HBE Model Using ANFIS

Table 6.15 illustrates a comparison among several design options for obtaining the optimum configuration of ANFIS for the estimation of the HBE trips, which was achieved first at both three gaussian and three trapezoidal MFs for the EDU input variable, hybrid learning algorithm, and 200 training epochs.

Table 6.15: Resulted RMSE and MAE for Different ANFIS Configurations - HBE

Learning Algorithm	Index	Training Epochs							
		1	5	10	50	100	200	500	1000
Gaussian Membership Functions (3 for EDU)									
Back-propagation	RMSE	2.1666	2.0372	1.8764	0.7140	0.4068	0.4031	0.4028	0.4027
	MAE	1.5713	1.4816	1.3694	0.4868	0.1784	0.1423	0.1408	0.1401
Hybrid	RMSE	0.4023	0.4023	0.4023	0.4022	0.4021	0.4020	0.4020	0.4020
	MAE	0.1442	0.1442	0.1442	0.1443	0.1443	0.1430	0.1430	0.1430
Trapezoidal Membership Functions (3 for EDU)									
Back-propagation	RMSE	2.1641	2.0249	1.8536	0.7533	0.4101	0.4028	0.4027	0.4027
	MAE	1.5699	1.4746	1.3555	0.4758	0.1874	0.1495	0.1489	0.1488
Hybrid	RMSE	0.4023	0.4023	0.4023	0.4022	0.4021	0.4020	0.4020	0.4020
	MAE	0.1443	0.1443	0.1443	0.1444	0.1443	0.1430	0.1430	0.1430
Triangle Membership Functions (3 for EDU)									
Back-propagation	RMSE	2.1668	2.0379	1.8780	0.7341	0.4046	0.4027	0.4025	0.4025
	MAE	1.5713	1.4813	1.3690	0.5045	0.1691	0.1405	0.1409	0.1409
Hybrid	RMSE	0.4022	0.4021	0.4021	0.4021	0.4021	0.4021	0.4021	0.4021
	MAE	0.1449	0.1443	0.1443	0.1443	0.1443	0.1443	0.1443	0.1443

The same results can be obtained by using either gaussian or trapezoidal MFs. Only one input variable was considered in this approach (i.e., EDU), with three assigned MFs. Hence, three fuzzy rules along with three output linear MFs were created consequently.

Table 6.16 summarizes the results associated with the optimized FIS. The values of R-Squared, RMSE, and MAE were 96.66%, 0.4020, and 0.1430, respectively. The optimized FISs for estimating HBE using the gaussian and the trapezoidal MFs are presented in Figure 6.11 and 6.12, respectively. The equivalent ANFIS architecture for the developed FIS is shown in Figure 6.13.

Table 6.16: ANFIS Optimum Configuration Summary for HBE

R-Square: 96.66%	**RMSE: 0.4020**
MAE: 0.1430	**Sample Size: 256**
Model Sum of Squares:	1196.623
Residual Sum of Squares:	41.377
Total Sum of Squares:	1238

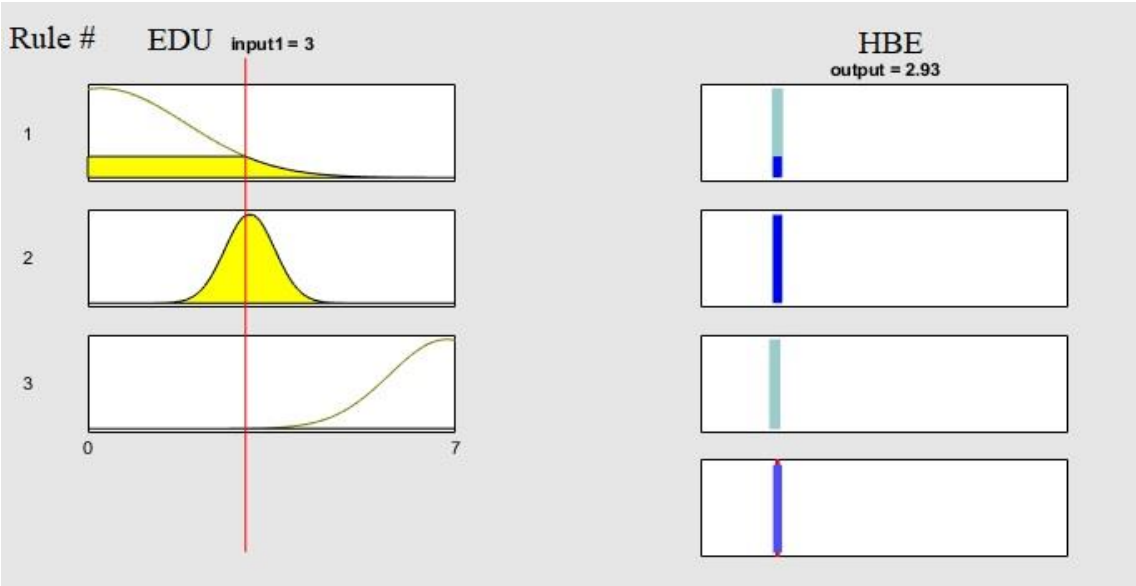

Figure 6.11: The Optimized FIS for Estimating HBE Using Gaussian MFs

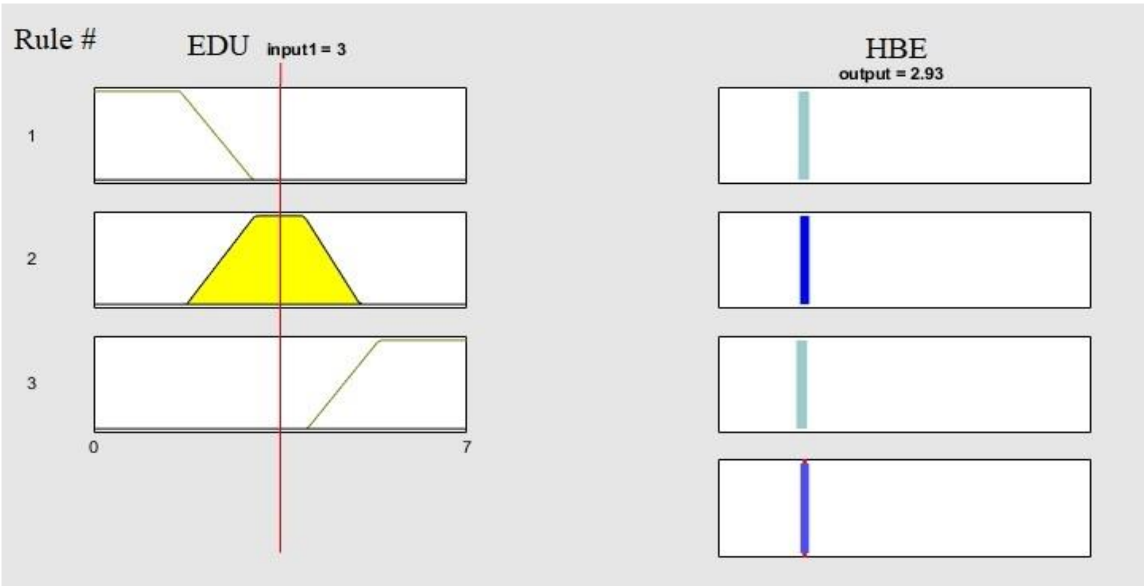

Figure 6.12: The Optimized FIS for Estimating HBE Using Trapezoidal MFs

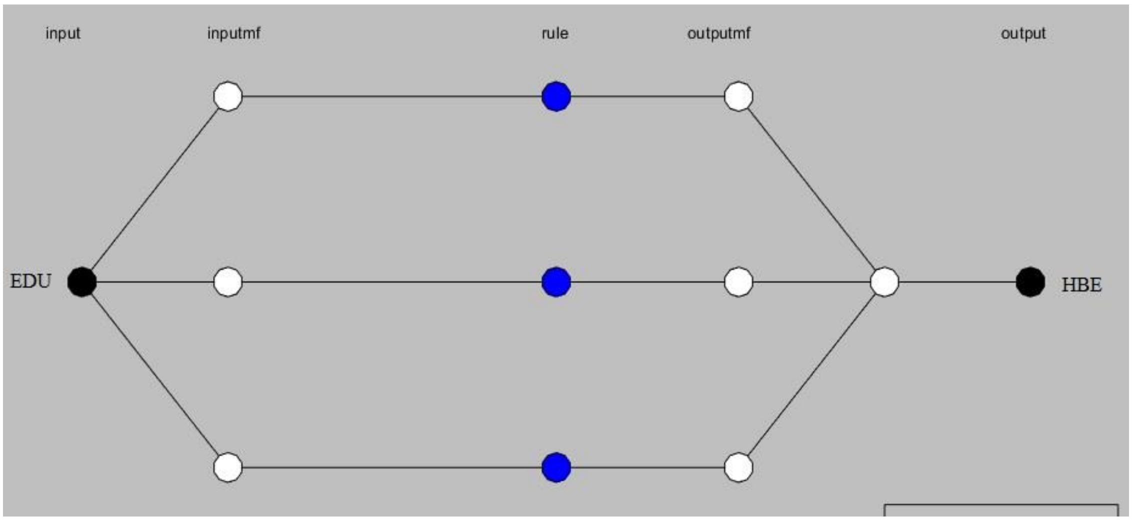

Figure 6.13: The Equivalent ANFIS Architecture for HBE Model

In this approach, the same results were obtained by assigning either gaussian or trapezoidal MFs to the EDU input variable. Figure 6.15 and 6.16 illustrate the optimized gaussian MFs with its parameters [σ, m], and the optimized trapezoidal MFs with its parameters [a, b, c, and d], respectively.

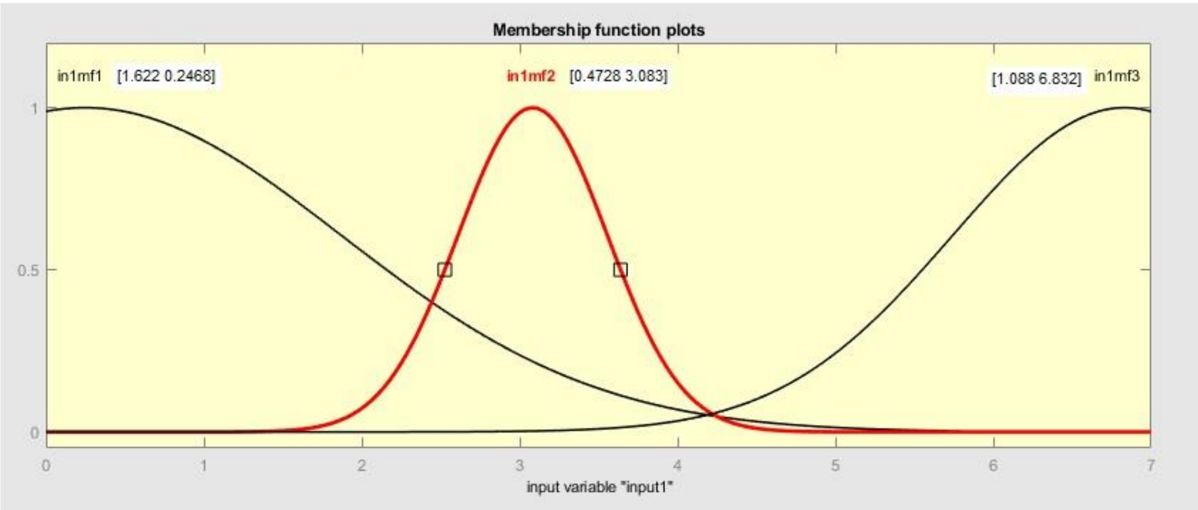

Figure 6.14: Optimized Gaussian MFs for EDU Input Variable

6.4.3 Models Comparison and Validation

Performance Comparison

Table 6.17 illustrates a comparison among the performance of the two approaches. The differences between the performance measures are very small and could be neglected, which indicates that both approaches are performing at the same level of accuracy, with no advantage among each other.

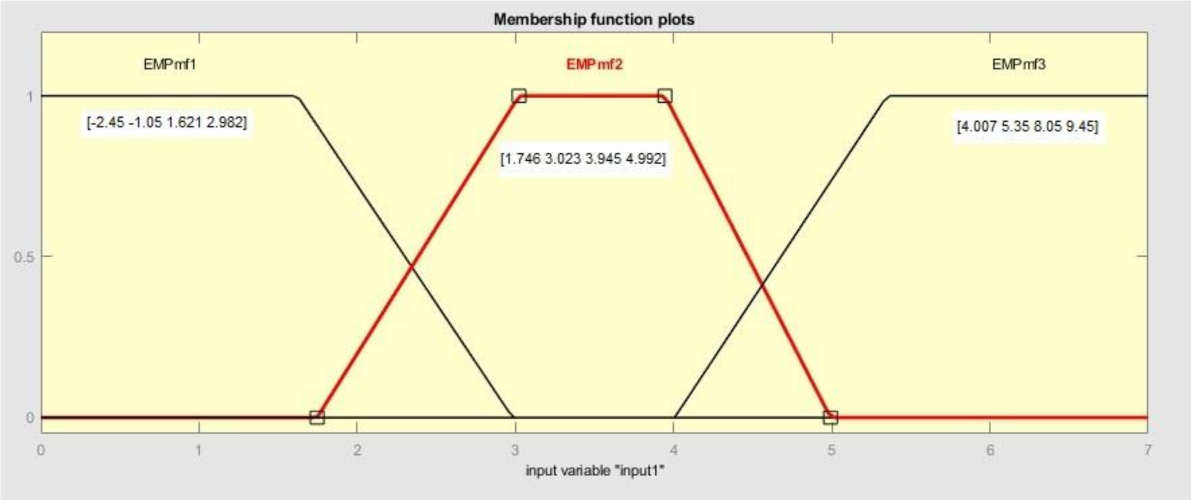

Figure 6.15: Optimized Trapezoidal MFs for EDU Input Variable

Table 6.17: HBE Performance Comparison Among MLR and ANFIS

	MLR	ANFIS	Difference
RMSE	0.4035	**0.4020**	-0.0015
MAE	**0.1318**	0.1430	+0.0112
R-Squared	96.63%	**96.66%**	+0.03%
Residual Sum of Squares	41.676	**41.377**	-0.300

Models Validation

A comparison among the actual and predicted HBE for the testing sample is presented in Appendix B, Table B-III. For both modeling approaches, the predicted results and the actual values are within acceptable and very well conformity, with better results associated with the MLR approach.

Table 6.18, summarizes several statistical measures for the actual and predicted values. The means of the predicted values using MLR and ANFIS (μ_{MLR} & μ_{ANFIS}) were compared with the mean of the actual values (μ_{Actual}) using the t-test. The results indicated that the null hypothesis H_0:

$\mu_{Actual} = \mu_{ANFIS}$ or $\mu_{Actual} = \mu_{MLR}$ could not be statistically rejected, and hence, the μ_{Actual} is not significantly different from both μ_{MLR} and μ_{ANFIS} at a 90% level of significance (i.e., $t_{statistic} < t_{critical\ (\alpha=10\%)}$ for MLR and ANFIS).

The prediction capabilities of the MLR can be considered better than the ANFIS, as μ_{MLR} was closer to μ_{Actual} than the μ_{ANFIS}, and the difference between the summation of the actual HBE and of the related predicted values using MLR (-3.22 trip) was smaller as compared with the ANFIS (-7.21 trip). In addition to the other measures, which indicate that the MLR can provide more precise and closer predictions to actual values than the ANFIS.

Table 6.18: Actual and Predicted HBE Descriptive Statistics

Validation (Statistics) / Testing Dataset					
	Actual		**MLR**		**ANFIS**
Mean	**1.58**		**1.65**		**1.72**
St. Deviation	1.69		1.59		1.64
Median	1		0.95		1.00
Sum	84		87.22		91.21
Difference	0		**-3.22**		**-7.21**
Count	53		53		53
t-test	t-critical =1.66	>	t-statistic = 0.22	<	t-statistic = 0.433
	α = 0.1	<	P-value = 0.8266	>	P-value = 0.6661

As the HBE model falls within a small data range from 0 to 7, and depends only on one highly correlated explanatory variable (EDU), the system could be considered simple without complexities. In this case, the MLR will be able to fit the data in an efficient way with high accuracy, and it will be enough for developing such model.

6.5 Home-Based Other Trips Generation Model (HBO)

Based on the collected data, the HBO trips constitute nearly 50% of total daily trips generated by a household, which range from zero up to a maximum of 11 trips per day.

6.5.1 Developing HBO Model Using Multiple Linear Regression

Equation 6.4 represents the best estimated HBO trips generation model (HBO). The regression results are summarized in Table 6.19.

$$\mathrm{HBO} = 0.415\ \mathrm{SIZE} + 0.289\ \mathrm{DRIVE} + 1.218\ \mathrm{HHTYP} \qquad (6.4)$$

where:

HBO = Number of daily other trips produced by household

SIZE = Number of persons in the household (household size)

DRIVE = Number of licensed drivers in the household

HHTYP = House Type: 1 if independent residence, 0 if apartment

This type of model was not considered by Amer (2017). Instead, Amer divided the other trips into three main categories: shopping, social, and recreational trips; and developed a separate model for each one. In this study, these three categories were combined into a single type (i.e., HBO), and were estimated using one single model.

Table 6.19: Regression Results for Home-Based Other Trips Generation Model

Model Summary					
Variables	**Coefficient**	**Standard Error**	**t-value**	**Sig.**	**VIF**
Intercept	0.000	#N/A	#N/A	#N/A	
SIZE	0.415	0.050	8.307	0.000	4.847
DRIVE	0.289	0.103	2.809	0.005	2.559
HHTYP	1.218	0.221	5.497	0.000	4.362
R-Square: 0.807		**Adjusted R-Square: 0.801**		**RMSE: 1.5778**	
MAE: 1.2825		**F-value: 351.584**		**Sample Size: 256**	

Referring to Equation (6.4), the the daily HBO trips is positively correlated with the household SIZE, DRIVE, and HHTYP, as indicated by the positive sign associated with each coefficient. Referring to Table 6.19, the coefficients of SIZE, DRIVE, and HHTYP are significant at 99.50% level of significance, as indicated by the associated t-value.

Referring to Appendix A, Table A-I: Pearson's Correlation Matrix, the correlation among SIZE & DRIVE, SIZE & HHTYP, and DRIVE & HHTYP are 0.247 (small), 0.097 (small), and 0.082 (small), respectively, which all are less than 0.500. The VIF for each explanatory variable included in the model are shown in Table 6.19. All of these values are less than 10, which indicates that there are no multicollinearity problems in the estimated model.

The R-square value for the estimated model is 0.807, which indicates that the SIZE, DRIVE, and HHTYP can explain nearly 80% of the variation in the dependent variable (HBO). Such value shows a good explanation of data variability. However, in addition to the high R-Squared value that was obtained, removing the constant from the model seems to be reasonable, as the minimum possible value for HBO trips is zero, and it is logical to obtain this value if the above explanatory variables were set to be zero.

The F-value for the estimated model is 351.584. Considering this high value, the hypothesis that the SIZE, DRIVE, and HHTYP have no impact on the HBO is rejected at the 99.99% level of significance. The calculation of the F-value for the estimated model can be obtained from Table 6.20.

Table 6.20: ANOVA Table for Home-Based Other Trips Generation Model (HBO)

Source	Sum of Squares	df	Mean Square	F-Value	Sig.
Regression	2656.737	3	885.579	351.584	0.000
Residual	637.263	253	2.519		
Total	3294	256			

6.5.2 Developing HBO Model Using ANFIS Approach

Table 6.21 illustrates a comparison among several design options for obtaining the optimum configuration of ANFIS for HBO model.

Table 6.21: Resulted RMSE and MAE for Different ANFIS Configurations - HBO

Learning Algorithm	Index	Training Epochs							
		1	5	10	50	100	200	500	1000
Gaussian Membership Functions (3-MFs for SIZE & DRIVE, 2-MFs for HHTYP)									
Back-propagation	**RMSE**	3.5521	3.4133	3.2423	2.0813	1.6374	1.4973	1.4818	1.4710
	MAE	3.1228	2.9890	2.8203	1.7105	1.3250	1.1957	1.1761	1.1561
Hybrid	**RMSE**	1.4506	1.4502	1.4495	1.4467	1.4466	1.4444	**1.4419**	1.4419
	MAE	1.1332	1.1332	1.1331	1.1331	1.1330	1.1333	**1.1252**	1.1252
Trapezoidal Membership Functions (3-MFs for SIZE & DRIVE, 2-MFs for HHTYP)									
Back-propagation	**RMSE**	3.5466	3.3873	3.1949	2.0800	1.6178	1.4921	1.4791	1.4756
	MAE	3.1181	2.9657	2.7751	1.7179	1.3262	1.1884	1.1707	1.1667
Hybrid	**RMSE**	1.4628	1.4629	1.4629	1.4635	1.4637	1.4631	1.4631	1.4631
	MAE	1.1440	1.1440	1.1440	1.1452	1.1454	1.1439	1.1439	1.1439
Triangle Membership Functions (3-MFs for SIZE & DRIVE, 2-MFs for HHTYP)									
Back-propagation	**RMSE**	3.5499	3.4029	3.2238	2.0959	1.6597	1.4980	1.4757	1.4701
	MAE	3.1206	2.9778	2.7995	1.7247	1.3393	1.1934	1.1690	1.1565
Hybrid	**RMSE**	1.4537	1.4472	1.4461	1.4455	1.4455	1.4455	1.4455	1.4455
	MAE	1.1407	1.1344	1.1319	1.1289	1.1289	1.1289	1.1289	1.1289

Referring to Table 6.21, the optimum structure was achieved first at three gaussian MFs for the SIZE and DRIVE, two gaussian MFs for the HHTYP, hybrid learning algorithm, and 500 training epochs. Two gaussian membership functions were identified for the HHTYP, as this input variable could be, based on the type of the household, either 1 if independent residency or 0 if apartment. Hence, two MFs seem to be sufficient here.

After the development of the initial FIS, 18 fuzzy rules (i.e., $3 \times 3 \times 2 = 18$) along with 18 output linear MFs were created consequently. Table 6.22 summarizes the results associated with the optimized FIS. The values of R-Squared, RMSE, and MAE were 83.94%, 1.4419, and 1.1252, respectively. The optimized FIS is presented in Figure 6.16, and the equivalent ANFIS architecture for the developed FIS is shown in Figure 6.17. Figure 6.18, 6.19, and 6.20 illustrate the optimized gaussian MFs for the SIZE, DRIVE, and HHTYP input variables, respectively.

Table 6.22: ANFIS Optimum Configuration Summary for HBO

R-Square: 83.94%	**RMSE: 1.4419**
MAE: 1.1252	**Sample Size: 256**
Model Sum of Squares:	2764.854
Residual Sum of Squares:	529.146
Total Sum of Squares:	3294

6.5.3 Models Comparison and Validation

Performance Comparison

Table 6.23 illustrates a comparison among the performance of the HBO models developed by both the MLR and the ANFIS approaches. This comparison indicates that the ANFIS has the ability to model the HBO trips more accurate with better performance as compared with the MLR approach.

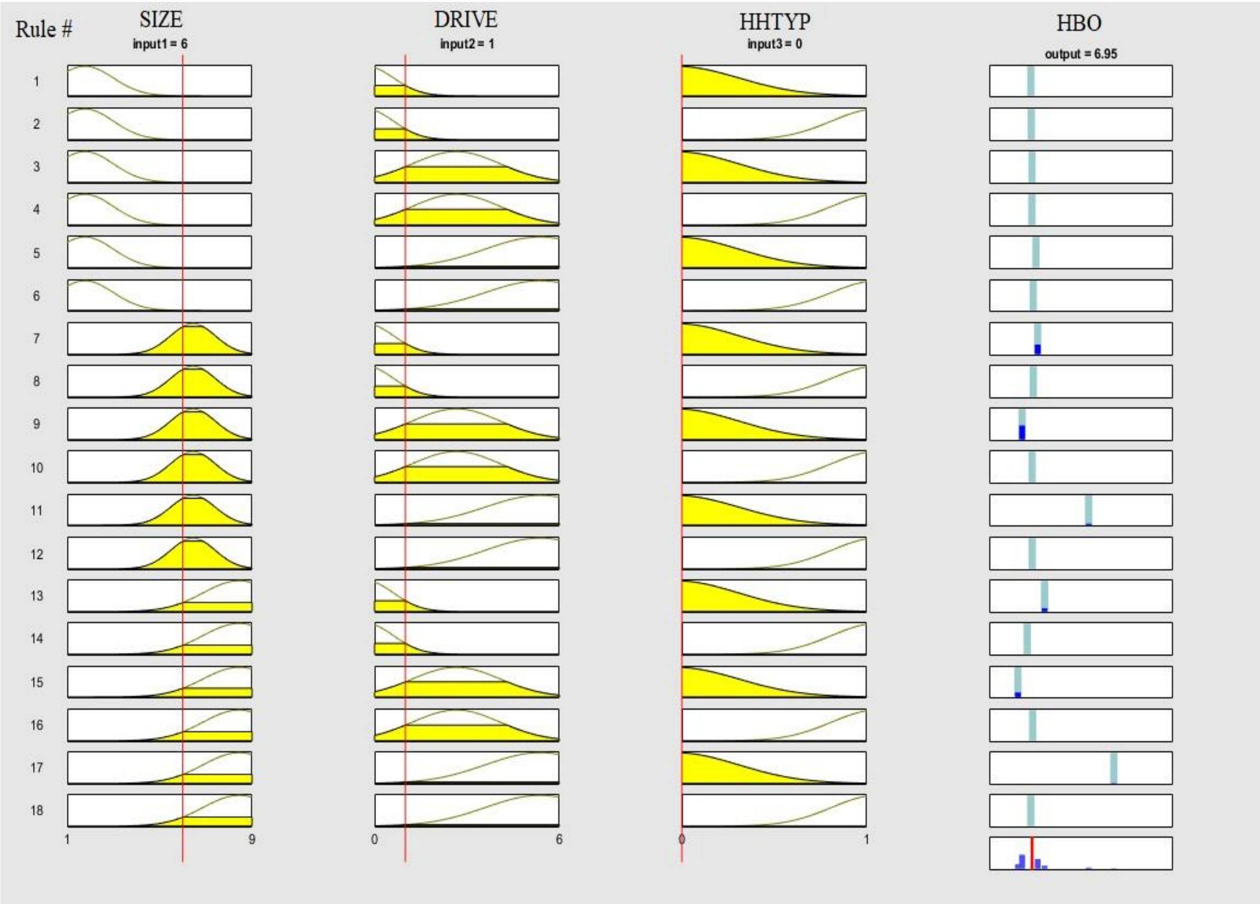

Figure 6.16: The Optimized FIS for Estimating HBO

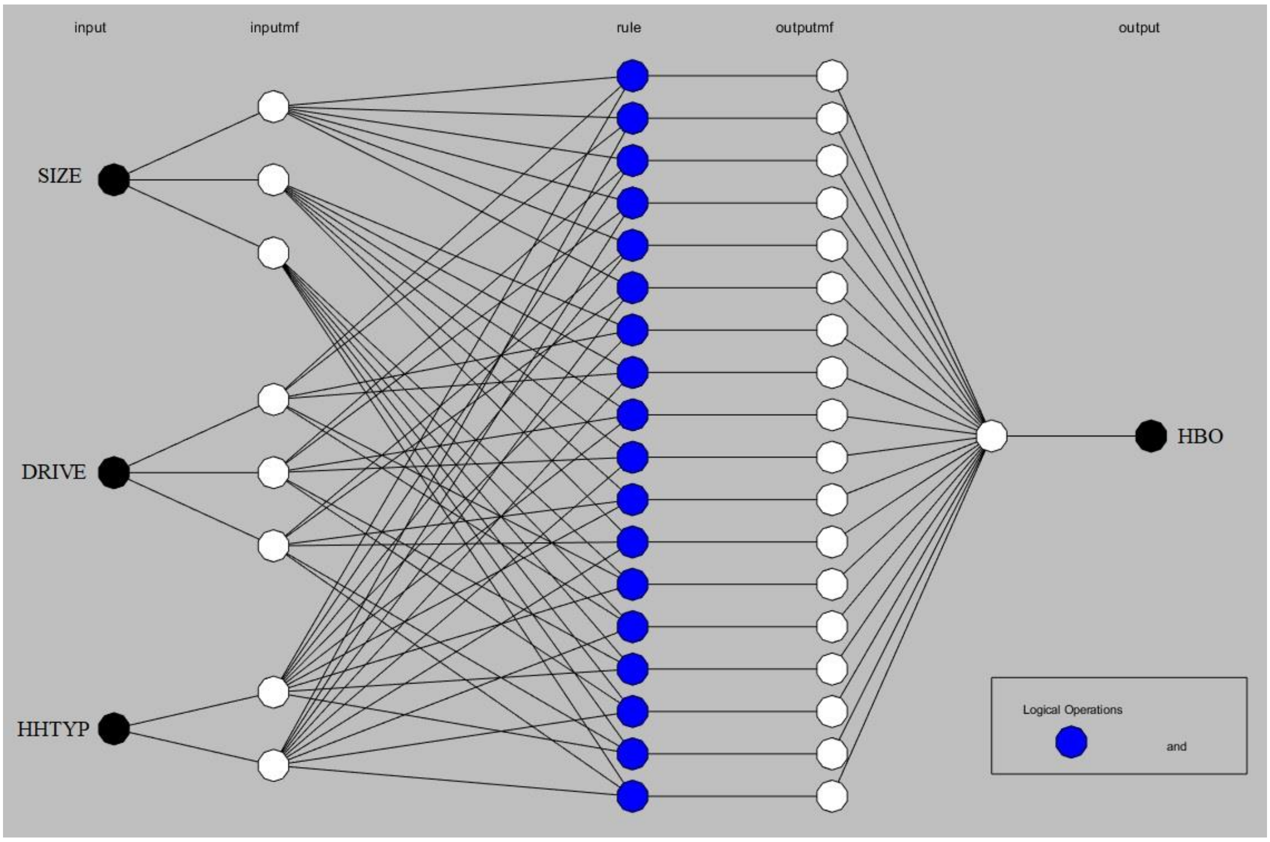

Figure 6.17: The Equivalent ANFIS Architecture for HBO Model

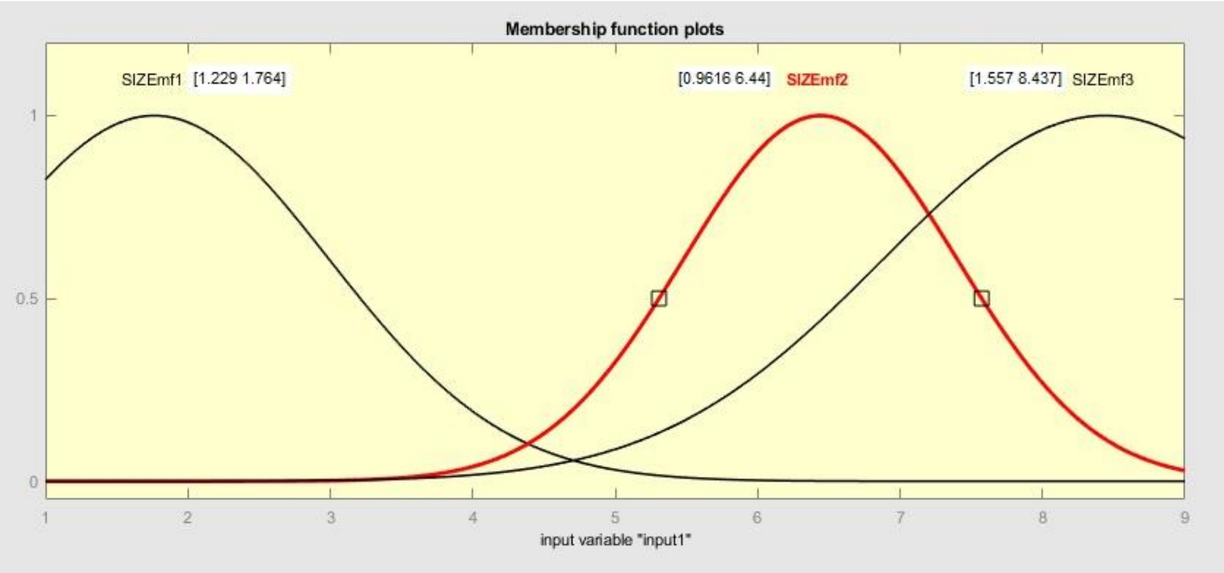

Figure 6.18: Optimized Gaussian MFs for SIZE Input Variable

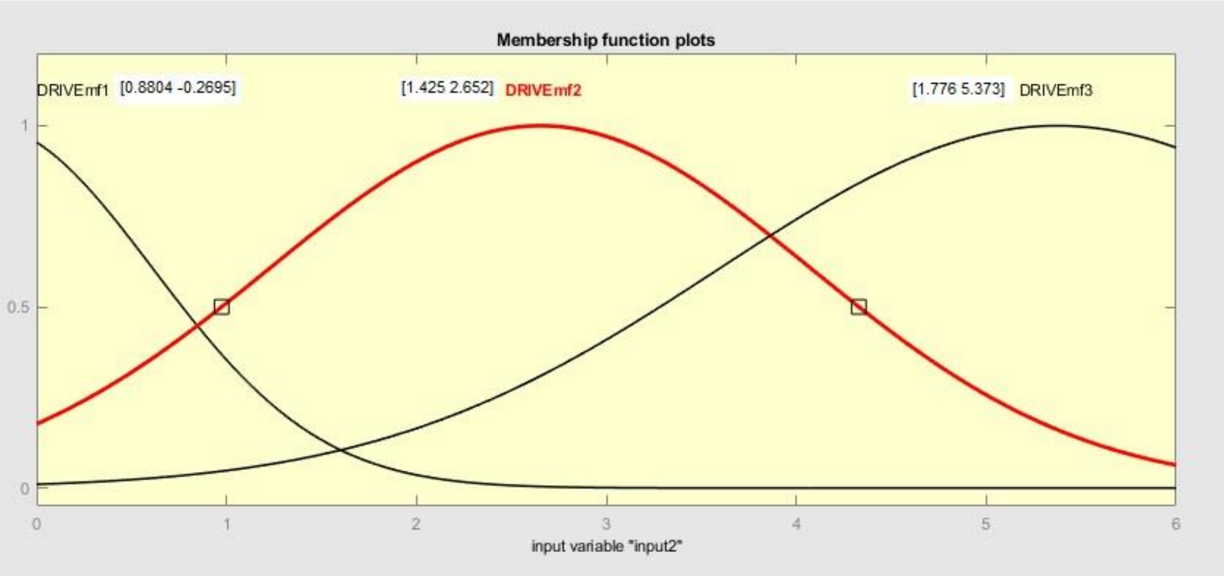

Figure 6.19: Optimized Gaussian MFs for DRIVE Input Variable

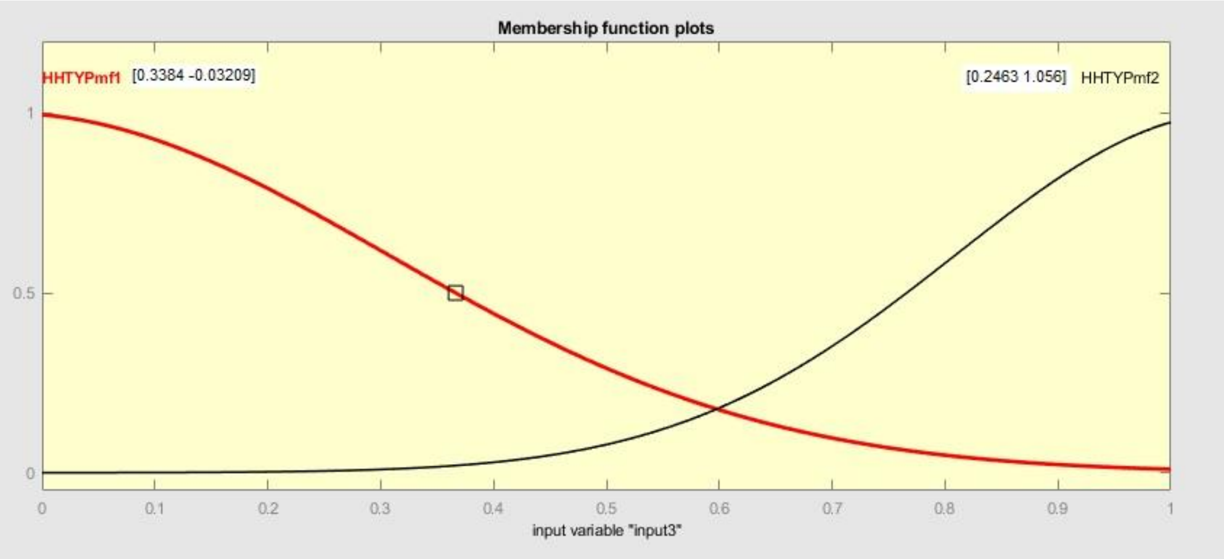

Figure 6.20: Optimized Gaussian MFs for HHTYP Input Variable

Referring to Table 6.23, the R-Squared value for HBO was raised from 80.65% by using MLR, to 83.94% by using ANFIS, which indicates an improvement by 3.29%. Furthermore, the residual sum of squares, RMSE, and MAE were all reduced form 637.263, 1.5778 and 1.2825, to 529.146, 1.4419 and 1.1252, respectively.

Table 6.23: HBO Performance Comparison Among MLR and ANFIS

	MLR	ANFIS	Difference
RMSE	1.5778	**1.4419**	-0.1359
MAE	1.2825	**1.1252**	-0.1572
R-Squared	80.65%	**83.94%**	+3.29%
Residual Sum of Squares	637.263	**529.146**	-108.117

It is noticed that the ANFIS is able to outperform the MLR approach in modeling the HBO trips as expected, especially with the complexity that associated with the HBO, as it has a data range of 0 to 11 trips per day and three input explanatory variables.

Models Validation

A comparison among the actual and predicted HBO for the testing sample, is presented in Appendix B, Table B-IV. For both modeling approaches, the predicted results and the actual values are within acceptable conformity, with better results associated with ANFIS.

Table 6.24 summarizes a comparison of several statistical measures among the actual and predicted values of this sample. The means of the predicted values using MLR and ANFIS (μ_{MLR} & μ_{ANFIS}) were compared with the mean of the actual values (μ_{Actual}) using the t-test.

The results indicated that the μ_{Actual} is not significantly different form μ_{ANFIS} at a 90% level of significance, while it is different from μ_{MLR} at the same level of significance. Hence, the hypothesis H_0: $\mu_{\text{Actual}} = \mu_{\text{ANFIS}}$ could be accepted, where the hypothesis H_0: $\mu_{\text{Actual}} = \mu_{\text{MLR}}$ is rejected (i.e., $t_{\text{statistic}} > t_{\text{critical}\ (\alpha = 10\%)}$). The ANFIS provides closer predicted values to the actual ones than the MLR approach.

The predicted mean using ANFIS was much closer to the actual mean than with the MLR. Furthermore, the difference between the summation of the actual HBO and of the related predicted values using ANFIS (23.01 trip), were much smaller as compared with the MLR (52.82 trips). In addition to the other measures, which indicate that the ANFIS can provide more precise and closer predictions to actual values than the MLR.

Table 6.24: Actual and Predicted HBO Descriptive Statistics

<table>
<tr><th colspan="6">Validation (Statistics) / Testing Dataset</th></tr>
<tr><th></th><th colspan="2">Actual</th><th colspan="2">MLR</th><th>ANFIS</th></tr>
<tr><td>Mean</td><td colspan="2">3.77</td><td colspan="2">2.78</td><td>3.34</td></tr>
<tr><td>St. Deviation</td><td colspan="2">2.006</td><td colspan="2">0.974</td><td>1.093</td></tr>
<tr><td>Median</td><td colspan="2">4</td><td colspan="2">2.75</td><td>3.15</td></tr>
<tr><td>Sum</td><td colspan="2">200</td><td colspan="2">147.18</td><td>176.99</td></tr>
<tr><td>Difference</td><td colspan="2">0</td><td colspan="2">52.82</td><td>23.01</td></tr>
<tr><td>Count</td><td colspan="2">53</td><td colspan="2">53</td><td>53</td></tr>
<tr><td rowspan="2">t-test</td><td>t-critical =1.66</td><td><</td><td>t-statistic = 3.232</td><td>></td><td>t-statistic = 1.37</td></tr>
<tr><td>α = 0.1</td><td>></td><td>P-value = 0.0016</td><td><</td><td>P-value = 0.1735</td></tr>
</table>

6.6 Summary

Through this chapter, four key trip generation models were developed using the data collected by Amer in 2017 for Salfit City. These models are ALLTRIP, HBW, HBE, and HBO as illustrated in Table 6.25. Each of these models was developed using two main approaches, the MLR and the ANFIS.

A comparison among the performance of the optimum models developed by each of these approaches was then conducted, in an attempt to seek the more accurate modeling technique.

A summary of the estimated models along with their performance and validation comparison is illustrated in Table 6.25. Figures 6.21, 6.22, and 6.23 illustrate a comparison among the resulted R-Squared, RMSE, and MAE, respectively, for each developed model considering both approaches. Where Figure 6.24 illustrates for each model a comparison among the mean of the actual and predicted trips using both approaches.

For the ALLTRIP and the HBO models, the ANFIS was able to outperform the MLR approach, and was more suitable for modeling these trips. In contrast, for the HBW and the HBE models, both modeling approaches were performing nearly at the same level, with better prediction results associated with the MLR approach, and the MLR approach was enough for modeling such trips.

Table 6.25: Summary of the Estimated Models Considering MLR & ANFIS

Variables		Models Estimation or Training			Models Performance Index					Models Validation & Testing			
Dependent (Output)	**Independent** (Input Variable)	**MLR**	**ANFIS**	**Sample Size**	**Index**	**MLR**		**ANFIS**	**Difference**	**Index**	**Actual**	**MLR**	**ANFIS**
Number of Daily Total Trips Produced by Household													
ALLTRIP (1-18)	**EMP EDU DRIVE**	ALLTRIP = +2.388 +1.046 EMP +1.237 EDU +0.424 DRIVE	3-Gaussian MFs for Each Input Variable 27-Fuzzy Rules Hybrid LA 1000-Training Epochs	256	**RMSE**	**1.7112**	>	**1.488**	-0.2232	**Average**	**6.81**	**6.44**	**6.64**
										St. Dev.	3.563	2.639	3.026
					MAE	**1.3882**	>	**1.1203**	-0.2679	**Median**	6	6.14	6.47
										Sum	361	341.35	351.66
					R-Squared	**65.85%**	<	**74.18%**	8.33%	**Difference**	0	**19.65**	**9.34**
										Count	53	53	53
Number of Daily Work Trips Produced by Household													
HBW (0-5)	**EMP AGEB CAR**	HBW = +0.875 EMP +0.153 AGEB +0.280 CAR	3-Gaussian MFs for Each Input Variable 27-Fuzzy Rules Hybrid LA 1-Training Epochs	256	**RMSE**	**0.5932**	>	**0.5465**	-0.0467	**Average**	**1.45**	**1.47**	**1.54**
										St. Dev.	0.932	0.764	0.739
					MAE	**0.4163**	>	**0.3664**	-0.0499	**Median**	1	1.31	1.58
										Sum	77	77.72	81.66
					R-Squared	**90.36%**	<	**92.74%**	2.38%	**Difference**	0	**-0.72**	**-4.66**
										Count	53	53	53
Number of Daily Educational Trips Produced by Household													
HBE (0-7)	**EDU**	HBE = +0.984 EDU	3-Gaussian MFs or 3-Trapezoidal MFs 3-Fuzzy Rules Hybrid LA 200-Training Epochs	256	**RMSE**	**0.4035**	>	**0.4020**	-0.0015	**Average**	**1.58**	**1.65**	**1.72**
										St. Dev.	1.69	1.59	1.64
					MAE	**0.1318**	<	**0.1430**	0.0112	**Median**	1	0.95	1
										Sum	84	87.22	91.21
					R-Squared	**96.63%**	<	**96.66%**	0.03%	**Difference**	0	**-3.22**	**-7.21**
										Count	53	53	53

Table 6.25: Summary of the Estimated Models Considering MLR & ANFIS (Continued)

Variables		**Models Estimation or Training**			**Models Performance Index**					**Models Validation & Testing**			
Dependent (Output)	**Independent** (Input Variable)	**MLR**	**ANFIS**	**Sample Size**	**Index**	**MLR**		**ANFIS**	**Difference**	**Index**	**Actual**	**MLR**	**ANFIS**
Number of Daily Other Trips Produced by Household													
HBO (0-11)	**SIZE DRIVE HHTYP**	HBO = +0.415 SIZE + 0.289 DRIVE + 1.218 HHTYP	3-Gaussian MFs for SIZE & DRIVE 2-Gaussian MFs for HHTYP 18-Fuzzy Rules Hybrid LA 500-Training Epochs	256	**RMSE**	**1.5778**	>	**1.4419**	-0.1359	**Average**	**3.77**	**2.78**	**3.34**
										St. Dev.	2.006	0.974	1.093
					MAE	**1.2825**	>	**1.1252**	-0.1572	**Median**	4	2.75	3.15
										Sum	200	147.18	176.99
					R-Squared	**80.65%**	<	**83.94%**	3.29%	**Difference**	0	**52.82**	**23.01**
										Count	53	53	53

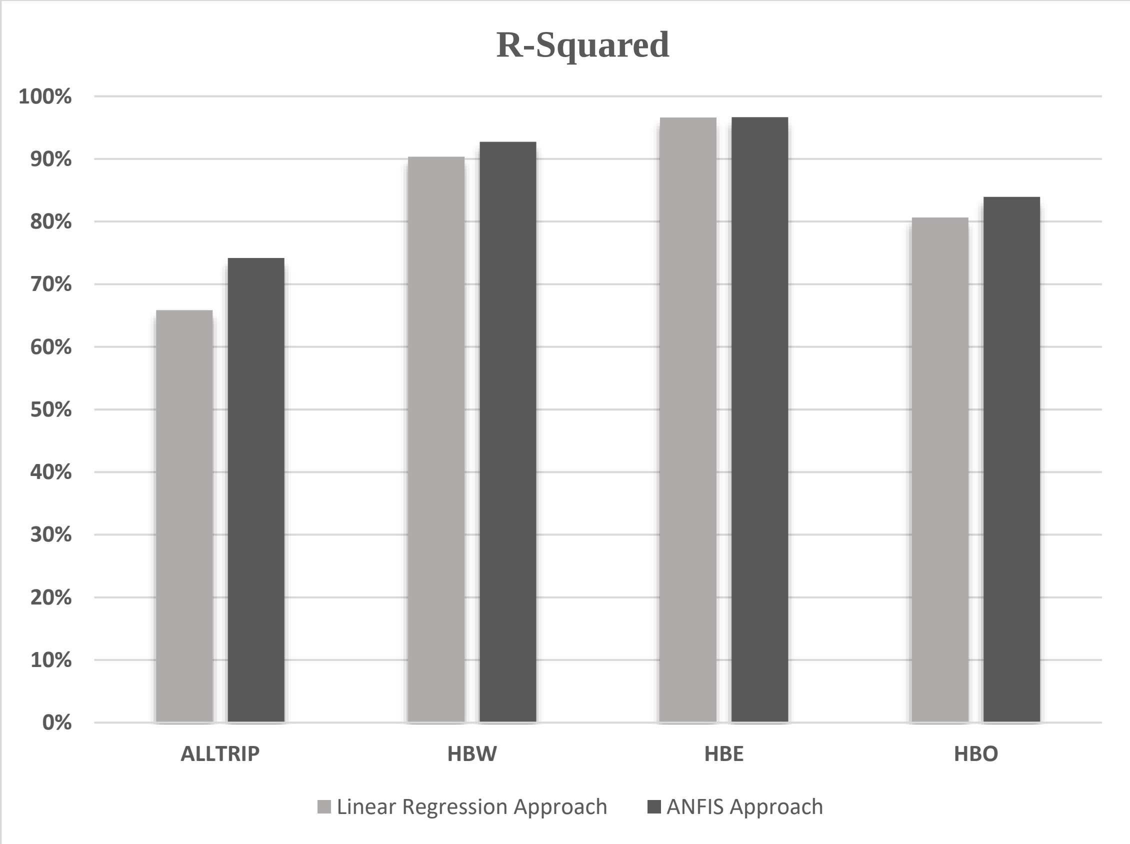

Figure 6.21: Comparison Among the Resulted R-Squared for Each Model

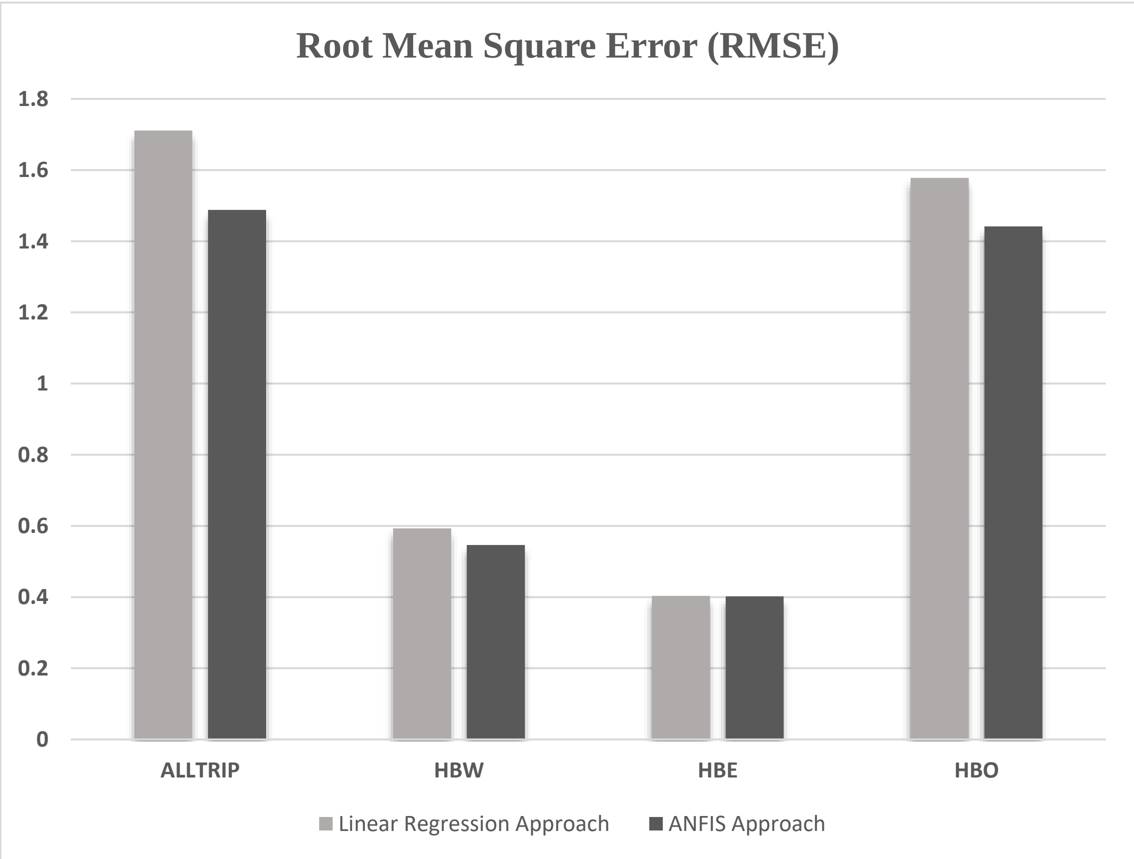

Figure 6.22: Comparison Among the Resulted RMSE for Each Model

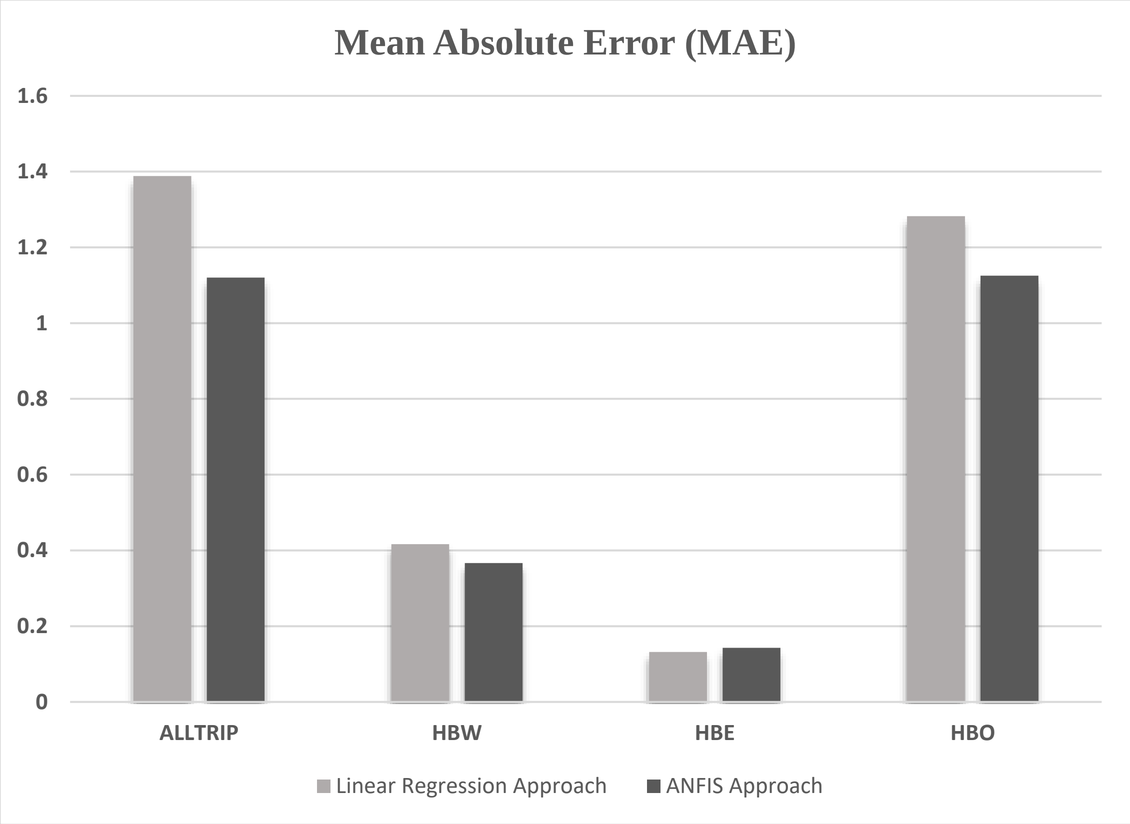

Figure 6.23: Comparison Among the Resulted MAE for Each Model

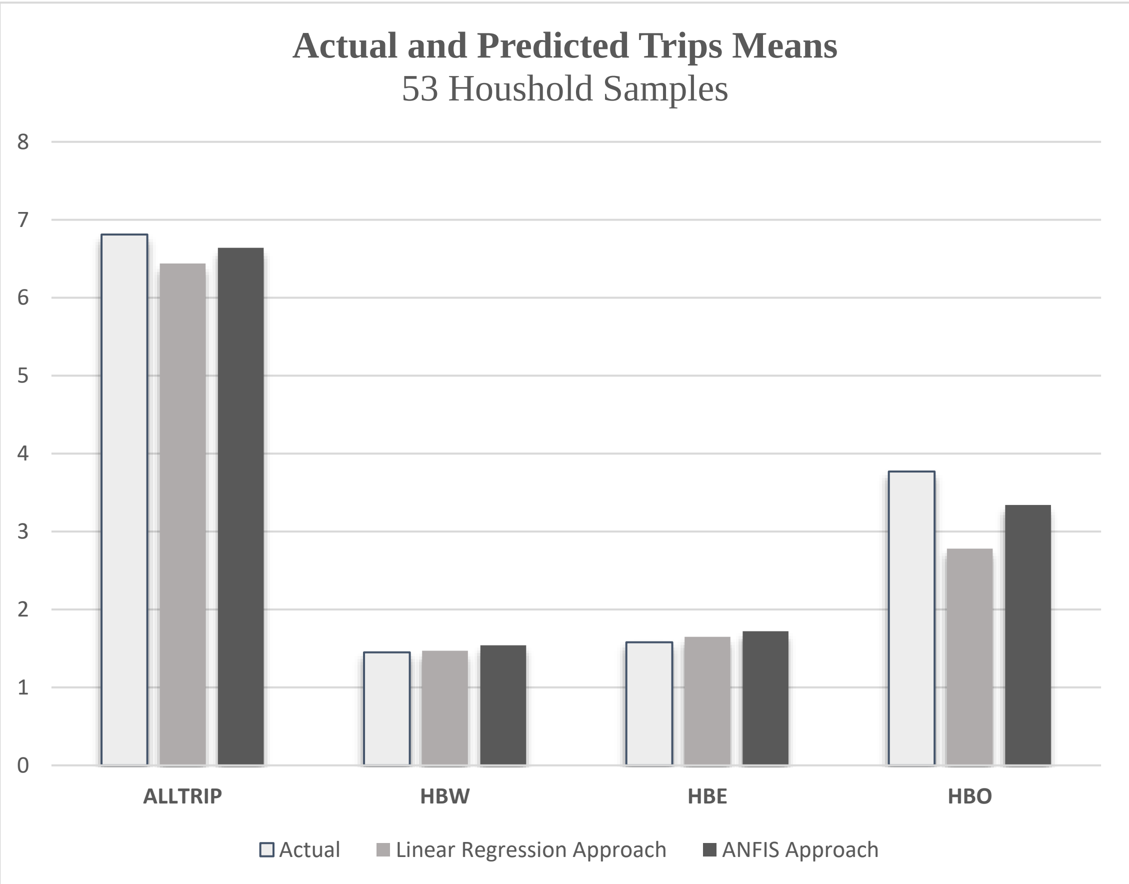

Figure 6.24: Comparison Among the Mean of Actual and Predicted Trips

Chapter Seven

Summary and Conclusions

Chapter Seven
Summary and Conclusions

7.1 Summary

In Palestine, there is little documented experience concerning transportation planning in general, and the development of trip generation models at specific. The lack of specialized studies for this purpose may be due to the economic, social, and political challenges that encounter the Palestinian situation, especially the restricted financial support and the lack of reliable data, which makes it difficult to perform such studies.

Limited few studies were conducted for this purpose using mainly the conventional Multiple Linear Regression (MLR) approach. This approach sometimes would not be able to develop an appropriate trip generation models, especially when dealing with many complex interrelated relationships among several socioeconomic factors. Hence, and for the Palestinian case, which suffers from the limited data resources, it is necessary to explore other modeling techniques, rather than the MLR approach, that would produce more accurate and reliable prediction models.

This study was devoted to investigating the feasibility of using a relatively new method for data analysis called the Adaptive Neuro-Fuzzy Inference System (ANFIS), as an alternative for the traditional MLR approach, and explore its application within the Palestinian context for the development of the home-based trip generation models, which was achieved by conducting a comparative analysis among the performance of several trip generation models that were developed using both approaches.

Salfit City was selected as a case for this study, due to the availability of thoroughly collected data through the recently conducted study, which was concerned with the development of home-based trip generation models for the city. For each modeling approach, a sample of 256 households was used for the estimation and training process (minimum sample size), and an additional sample of 53 households was considered for validation and testing purposes.

Through this study, four types of trip generation models were developed for Salfit City, which are the ALLTRIP model for estimating the number total daily trips generated by a household, the HBW and HBE models for estimating the number of daily home-based trip generated for work and education purposes, respectively, and the HBO model for estimating the number of daily home-based other trips generated not for work and education purposes.

Each of these models was developed using the two main competing approaches; the MLR and the ANFIS. For the MLR approach, different statistical tests were performed to ensure the selection of the most relevant explanatory variables, and the best estimation of the intended model. Furthermore, several configurations for the ANFIS were considered, in an attempt to obtain the optimum possible structure with the lowest RMSE.

A comparison among the performance of the optimum models developed by each of these approaches was then conducted. The more suitable and accurate approach was then determined based on several evaluation criteria, such as the higher value of R-Squared, the lower RMSE and MAE, and the much closer outputs to the actual values.

7.2 Conclusions

Through this study, it was found that ANFIS will be able to produce more accurate models than the MLR approach, especially when the relationships become more complex among the socioeconomic variables that are describing the trips generation. In such a case, the power of the ANFIS is expected to emerge as a better approach that will develop more reliable models with better performance. For each type of the four intended models, the following is a brief discussion that demonstrates these findings:

1. In the development of the **ALLTRIP** model, the ANFIS was able to outperform the MLR approach. This model can be considered a complex system as it constitutes all the trips generated by a household during the day (i.e., 100% of trips) with data ranges from 1 to 18 trips. The R-Squared was improved by 8.33%, from 65.85% to 74.18%. The ANFIS was able to capture the variation among the ALLTRIP more accurately than the MLR approach. Moreover, the validation assessment indicates that the ANFIS can provide more precise and closer predictions to actual values than the MLR.

2. In the **HBW** model, both modeling approaches, the ANFIS and the MLR, were performing nearly at the same level. The HBW model was considered a simple system as it constitutes nearly 25% of total daily trips, with small data ranges from 0 to 5 trips. In this case, the improvement in the R-Squared, or the reduction in the errors, were not that large to say that the ANFIS can outperform the MLR. The R-Squared value associated with the MLR is 90.36%, which is good enough to capture most of the variation among the HBW trips. Moreover, the validation assessment indicates that the MLR can provide more precise and closer predictions to actual values than the ANFIS. In such a case, the MLR approach is enough.

3. In the **HBE** case, both ANFIS and MLR were also performing at the same level. This model was considered the simplest among the other models, as it uses only one explanatory variable to explain its behavior with high correlation value, and constitutes nearly 25% of all trips generated, with small data ranges from 0 to 7 trips. The R-Squared value associated with the MLR approach was 96.63%, which is large enough to capture most of the variation among the HBE trips. The differences between the performance measures were very small and could be neglected. Furthermore, the validation assessment indicates that the MLR can provide more precise and closer predictions to actual values than the ANFIS. In such a case, the MLR approach is enough.

4. In the **HBO** model, the ANFIS was able to outperform the MLR approach, especially with the associated complexity. The HBO trips

constitute nearly 50% of total daily trips, with data ranges from 0 to 11 trips per day. By using the ANFIS, the R-Squared value was raised from 80.65% to 83.94% with an improvement of 3.29%. Moreover, the validation assessment indicates that the ANFIS can provide more precise and closer predictions to actual values than the MLR.

Based on the above discussion, it could be concluded that when the developed model by the MLR approach has large R-Squared value, which is good enough to capture most of the variation among the system, the use of ANFIS as an alternative approach will not produce significant improvements in the model. Efficient models could be developed using the MLR approach, when dealing with a simple system, with small data range at hand. However, the claims by the reviewed literature, that the ANFIS can be successfully used for modeling complex and nonlinear systems, and usually outperform the MLR approach in modeling trip generation, had been successfully confirmed by this study for ALLTRIP and HBO. The utilized performance measures confirmed higher forecasting accuracy of ANFIS in comparison to MLR for more complex systems.

It is to be stated here that, and through this study, a particular emphasis was given to the effect of different design options in optimizing and building the desired ANFIS models. The development of ANFIS involves 48 possible configurations, which were considered for each type of the four intended models. The configuration with the lowest RMSE and MAE was selected to represent the optimum ANFIS structure. It was noticed that the use of gaussian MFs along with hybrid learning algorithm would usually produce lowest possible RMSE. However, as there are no simple ways to determine in advance what should be the number and type of MFs, the type of the training algorithm, or the optimal number of training epochs, the following are the main findings in this regard:

1. the membership functions of the gaussian type always reached the optimum structure first with minimum possible error at the lowest number of training epochs, comparing with trapezoidal and triangle types.
2. For each input variable, three MFs seems to be sufficient, as using one MF may create estimation errors, where using more than three may increase the computational cost and time, and the number of rules.
3. The hybrid learning algorithm will reach the minimum possible measure of error quickly after several training epochs, which is considered more suitable for modeling trip generation in this context, comparing with the backpropagation algorithm, which slowly approaches the optimum results after too many training epochs.
4. There are no simple ways to select the optimal number of training epochs, it usually depends on the size of the training dataset. However, it was found that as the smaller the training dataset is, the smaller the number of learning epochs is required.

In summary, and through this study, the robust comparison and validation process reveal that the ANFIS represents a promising modeling technique, that can be a good competitor for MLR approach, especially, when dealing with interrelated and complex relationships among several socioeconomic variables. It was found that the ANFIS can be used successfully for modeling ALLTRIP and HBO trips generation for Salfit City, and usually able to outperform traditional MLR approach. The ANFIS is a potentially better data analytic method for complex systems, which needs to be explored more in-depth and compared to more sophisticated regression techniques that are already in use in transportation.

7.3 Recommendations

The following recommendations can be drawn from the results of this thesis:

1. The Palestinian municipalities, and other related agencies, are encouraged to use advanced modeling approaches, including the use of ANFIS approach, in transportation planning, which could provide more accurate results when dealing with complex systems. This approach can be used side by side with the MLR approach, in transportation planning in general, and the development of trip generation models at specific. The Palestinian municipalities, including Salfit Municipality, are also encouraged to integrate the developed models by this thesis with more comprehensive urban transportation planning process in their cities.

2. The developed models by this study rely on the data collected for Salfit City in the year of 2017. Hence, it is recommended to collect new data that represent the current situation of the city, and use these data to check and ensure the validity of the developed models.

3. Further studies are recommended to be conducted that concern with the applications of the ANFIS approach in modeling trip generation, as this approach is considerably new, which needs to be explored more in-depth and compared to more sophisticated regression techniques that are already in use in transportation.

4. In this research, the Takagi-Sugeno type fuzzy inference system was considered due to its simplicity, as the output MFs are being either linear or constant. However, it is recommended to explore the benefits of using the Mamdani type fuzzy inference system instead, which expects the output MFs to be fuzzy sets that need defuzzification after the aggregation process.

5. It is also recommended to investigate the feasibility of normalizing the inputs and outputs dataset by scaling between -1 and 1, when considering the application ANFIS approach. By this it is expected to eliminate their dimensions and ensure that all variables receive the same treatment during the training of the model. In addition, the normalized input and output could accelerate the convergence process during the model training.

6. A recent study was conducted which considered the development of transportation mode choice models for Palestinian universities students, using the multinomial logistic regression approach. It is being recommended to explore the potential of using the ANFIS approach for developing such models, as this approach is more able to capture and

interpret the uncertainties associated with traveler decisions, and better deals with subjective information about the attributes of trip.

7. The Artificial Neural Network (ANN) provides another promising technique for modeling trip generation. It allows the user to identify all desired inputs and outputs variables in one step. For example, all explanatory variables that were collected by Amer (2017) could be inserted into the network as inputs variables, and the desired outputs such as HBW, HBE and ALLTRIP could be all identified together. The network will then adapt itself to self-organize its structure, when the 256-households sample is presented. It is recommended to investigate the feasibility of using the ANN approach by conducting a comparative analysis among the performance of different trip generation models developed using this approach and both of the ANFIS and MLR approaches.

References

Abu-Eisheh, S., Ghanim, M., & Dodeen, A. (2017). **Modeling Travel Behavior Based on Household Characteristics**. MATEC Web of Conferences, Volume 120, Article Number 07005, ASCMCES-17, PP:1-5.

Ahire, J. (2018). **The Artificial Neural Networks Handbook: Part 4**. Retrieved from Medium Corporation (Accessed: 2019, Mar 15): https://medium.com/@jayeshbahire/the-artificial-neural-networks-handbook-part-4-d2087d1f583e

Ahmadpour, M., Yue, W., & Mohammadzaheri, M. (2009). **Neuro-fuzzy Modelling of Workers Trip Production**. Retrieved from (Accessed: 2019, Mar 20): https://www.semanticscholar.org/paper/Neuro-fuzzy-modelling-of-workers-trip-production-AhmadpourYue/67a562de25044a1bf3157a166b2488ff478f268f

Al-Sahili, K., Abu-Eisheh, S., & Kobari, F. (2017). **Estimation of New Development Trip Impacts Through Trip Generation Rates for Major Land Uses in Palestine**. Presented at the Seventh Jordanian International Civil Engineering Conference, Engineers Association, Amman.

Amer, A. (2017). **Transferability of Trip Generation Models for Palestinian Cities: Jericho and Salfit**. Master thesis, An-Najah National University, Nablus, Palestine.

Andrade, K., Uchida, K., & Kagaya, S. (2006). **Development of Transport Mode Choice Model by Using Adaptive Neuro-Fuzzy Inference System**. Transportation Research Record: Journal of the Transportation Research Board, No. 1977, Transportation Research Board of the National Academies. Washington, D.C., PP: 8-16.

Ben-Edigbe, J., & Rahman, R. (2010). **Multivariate School Travel Demand Regression Based on Trip Attraction**. International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, Vol. 4, No. 6, PP:1156-1160.

Chang, J. S., Jung, D., Kim, J., & Kang, T. (2014). **Comparative Analysis of Trip Generation Models: Results Using Home-Based Work Trips in the Seoul Metropolitan Area**. Transportation Letters, The International Journal of Transportation Research, Vol. 6, No. 2, PP:78-88.

Dodeen, A. (2014). **Developing Trip Generation Models Utilizing Linear Regression Analysis: Jericho City as a Case Study**. Master thesis, An-Najah National University, Nablus, Palestine.

Eisenhauer, J. (2003). **Regression Through the Origin**. Teaching Statistics. Volume 25, Number 3, PP: 76-80.

ElAstal, I. (2014). **Modeling Road Accident Black Spots in Gaza Strip Using Artificial Neural Network (ANN) 2000-2013**. Master thesis, The Islamic University, Gaza, Palestine.

Elhami, B., Akram, A., Khanali, M., & Mousavi-Avval, H. (2016). **Application of ANFIS and Linear Regression Models to Analyze the Energy and Economics of Lentil and Chickpea Production in Iran**. Energy Equipment and Systems, Vol. 4, No. 2, PP: 255-270.

Goel, S., & Sinha, K. (2008). **Trip Generation Modeling Using Artificial Neural Network**. INDIACom–Computing for Nation Development, Bharati Vidyapeeth’s Institute of Computer Applications and Management, New Delhi.

Hunt, L., & Broadstock, D. (2010). **Modeling Car Trip Generations for UK Residential Developments Using Data from TRICS**. Transportation Planning and Technology, 33 (8), 671-678.

Jang, J., Sun, C., & Mizutain, E. (1997). **Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence**. Prentice Hall, Inc.

Jang, R. (1993). **ANFIS: Adaptive Network-Based Fuzzy Inference Systems**. IEEE Transactions on Systems, Man, and Cybernetics, Vol 23, No. 3, PP: 665–685.

Jang, R., & Gulley, N. (2018). **Fuzzy Logic Toolbox™ User's Guide**, **MATLAB R2018b**. The MathWorks, Inc.

Karlaftis, M., & Vlahogianni, E. (2011). **Statistical Methods Versus Neural Networks in Transportation Research: Differences, Similarities and Some Insights**. Transportation Research Part C, Vol.19, Elsevier Ltd, PP: 387–399.

Lim, K., & Srinivasan, S. (2011). C**omparative Analysis of Alternate Econometric Structures for Trip Generation Models**. Transportation Research Record: Journal of the Transportation Research Board, No. 2254, TRB, Washington, D.C., PP: 68–78.

Mahdavi, R., & Mamdoohi, R. (2018). **Comparison of Linear Regression Model and ANFIS Model for Prediction of Arbitrary Trip Attraction: Case Study Qazvin City**. Presented at the 17th International Conference of Traffic and Transportation Engineering, Tehran, Iran.

Meyer, M. D., & Miller, E. J. (2001). **Urban Transportation Planning: A Decision Oriented Approach**. Second Edition. New York: McGraw-Hill.

Montgomery, D., & Runger, G. (2014). **Applied Statistics and Probability for Engineers**. Sixth Edition, John Wiley & Sons, Inc.

Moussa, H. (2013). **Development of a Trip Generation Model for Gaza City**. Master thesis. The Islamic University, Gaza, Palestine.

Ort´uzar, J. d., & Willumsen, L. G. (2011). **Modelling Transport**. Fourth Edition, John Wiley and Sons, Ltd., Publication.

Palestinian Central Bureau of Statistics (PCBS): **Population, Housing and Establishments Census 2017, Salfit Governorate**.

Pamuła, T. (2016). **Neural Networks in Transportation Research – Recent Applications**. Transport Problems, Volume 11, Issue 2.

Přibyl, O., & Goulias, K. (2003). **Application of Adaptive Neuro-Fuzzy Inference System to Analysis of Travel Behavior**. Transportation Research Record: Journal of the Transportation Research Board, No.1854, TRB, National Research Council. Washington, D.C., PP: 180-188.

Profillidis, V., & Botzoris, G. (2018). **Modeling of Transport Demand: Analyzing, Calculating, and Forecasting Transport Demand**. Elsevier Inc., Science Publishers.

Pulugurta, S., Errampalli, M., & Ravinder, K. (2012). **Prediction of Future Trips Using Fuzzy Logic Based Trip Generation Model**. CSIR-Central Road Research Institute. New Delhi.

Rassafi, A., Rezaei, R., & Hajizamani, M. (2012). P**redicting Urban Trip Generation Using a Fuzzy Expert System**. Iranian Journal of Fuzzy Systems, Vol. 9, No. 3, PP: 127-146.

Rutkowska, D. (2002). **Neuro-Fuzzy Architectures and Hybrid Learning**. Springer-Verlag Berlin Heidelberg GmbH.

Shafi, J., Angelov, P., & Umair, M. (2016). **Prediction of the Attention Area in Ambient Intelligence Tasks**. Innovative Issues in Intelligent Systems, Studies in Computational Intelligence 623. Springer International Publishing Switzerland. PP: 33-56.

Simha, G. H. (2017). **Modelling Trip Generation for Urban & Suburban Areas: A Comparative Case Study**. Imperial Journal of Interdisciplinary Research (IJIR), Volume-3, Issue-5, PP: 446-461.

Srisaeng, P., Baxter, S., & Wild, G. (2015). **An Adaptive Neuro-Fuzzy Inference System for Forecasting Australia's Domestic Low Cost Carrier Passenger Demand**. AVIATION, Vol. 19, No. 3, Taylor & Francis Group, PP: 150–163.

Stojčić, M. (2018). **Application of The ANFIS Model in Road Traffic and Transportation: A Literature Review From 1993 To 2018**. Operational Research in Engineering Sciences: Theory and Applications, Volume-1, Issue-1, PP: 40-61.

Sugeno, M. (1985). **Industrial Applications of Fuzzy Control**. Elsevier Science Publishing Company.

Tan, Y., Shuai, C., Jiao, L., & Shen, L. (2018). **Adaptive Neuro-Fuzzy Inference System Approach for Urban Sustainability Assessment: A China Case Study**. Sustainable Development. Volume 26, Issue 6. ERP Environment . John Wiley & Sons, Ltd. PP: 1-16.

Teodorovic, D. (1999). **Fuzzy Logic Systems for Transportation Engineering: The State of the Art**. Transportation Research Part A: Policy and Practice, Vol. 33, PP: 337-364.

Tillema, F., Zuilekom, K., & Maarseveen, M. F. (2004). **Trip Generation: Comparison of Neural Networks and Regression Models**. Urban Transport X, WIT Transactions on The Built Environment, Volume 75, PP: 121-130.

Tortum, A., Yayla, N., & Gökdag, M. (2009). **The Modeling of Mode Choices of Intercity Freight Transportation with The Artificial Neural Networks and Adaptive Neuro-Fuzzy Inference System**. Expert Systems with Applications. Vol.36, Elsevier Ltd, PP: 6199–6217.

Wahyudi, U., Albert, C., Hafiz, Z., Ran, G., & Dwifitra, J. (2019). **Making Decision Toward Overseas Construction Projects: An Application Based on Adaptive-Neuro Fuzzy System**. Engineering, Construction and Architectural Management. PP:1-19.

Washington, S., Karlaftis, M., & Mannering, F. (2010). **Statistical and Econometric Methods for Transportation Data Analysis**. Second Edition. CRC Press, Chapman & Hall Bool.

Appendices

Appendix A

Table A-I: Pearson's Correlation Matrix

	SIZE	M	F	EMP	EDU	AGE$_A$	AGE$_B$	AGE$_C$	AGE$_D$	AGE$_E$	DRIVE	CAR	BICY	MCYC	INC	HHTYP
SIZE	1	0.802	0.742	0.500	0.838	0.763	0.370	0.501	0.066	-0.182	0.247	0.146	0.173	0.113	0.235	0.097
M	0.802	1	0.195	0.503	0.642	0.597	0.348	0.408	0.043	-0.179	0.352	0.208	0.170	0.116	0.268	0.077
F	0.742	0.195	1	0.256	0.656	0.582	0.217	0.365	0.060	-0.098	0.010	0.007	0.093	0.055	0.085	0.072
EMP	0.500	0.503	0.256	1	0.166	0.182	0.431	0.256	0.161	-0.239	0.455	0.281	0.074	0.214	0.430	0.095
EDU	0.838	0.642	0.656	0.166	1	0.865	0.119	0.564	-0.188	-0.266	0.088	0.047	0.167	-0.020	0.128	0.028
AGE$_A$	0.763	0.597	0.582	0.182	0.865	1	-0.144	0.579	-0.267	-0.263	-0.031	0.009	0.176	-0.021	0.080	0.012
AGE$_B$	0.370	0.348	0.217	0.431	0.119	-0.144	1	-0.201	0.163	-0.191	0.345	0.127	0.028	0.175	0.138	0.048
AGE$_C$	0.501	0.408	0.365	0.256	0.564	0.579	-0.201	1	-0.478	-0.293	0.032	0.000	0.107	-0.015	0.087	0.033
AGE$_D$	0.066	0.043	0.060	0.161	-0.188	-0.267	0.163	-0.478	1	-0.009	0.212	0.174	-0.009	0.126	0.181	0.034
AGE$_E$	-0.182	-0.179	-0.098	-0.239	-0.266	-0.263	-0.191	-0.293	-0.009	1	-0.092	0.005	-0.078	-0.058	-0.093	0.131
DRIVE	0.247	0.352	0.010	0.455	0.088	-0.031	0.345	0.032	0.212	-0.092	1	0.597	0.061	0.174	0.419	0.082
CAR	0.146	0.208	0.007	0.281	0.047	0.009	0.127	0.000	0.174	0.005	0.597	1	0.068	0.108	0.448	0.077
BICY	0.173	0.170	0.093	0.074	0.167	0.176	0.028	0.107	-0.009	-0.078	0.061	0.068	1	0.088	-0.009	0.024
MCYC	0.113	0.116	0.055	0.214	-0.020	-0.021	0.175	-0.015	0.126	-0.058	0.174	0.108	0.088	1	0.067	0.053
INC	0.235	0.268	0.085	0.430	0.128	0.080	0.138	0.087	0.181	-0.093	0.419	0.448	-0.009	0.067	1	0.110
HHTYP	0.097	0.077	0.072	0.095	0.028	0.012	0.048	0.033	0.034	0.131	0.082	0.077	0.024	0.053	0.110	1

Appendix B

Table B-I: Actual vs. Predicted ALLTRIP Using MLR & ANFIS

SAMPLE #	Actual	MLR		ANFIS	
	ALLTRIP	Predicted	Residuals	Predicted	Residuals
257	8	6.57	1.44	7.00	1.00
258	8	7.57	0.43	7.20	0.80
259	4	2.81	1.19	4.36	-0.36
260	6	3.86	2.14	3.73	2.27
261	3	7.99	-4.99	7.55	-4.55
262	9	6.57	2.44	7.00	2.00
263	6	4.67	1.33	4.80	1.20
264	3	5.33	-2.33	5.25	-2.25
265	5	3.86	1.14	3.73	1.27
266	10	9.04	0.96	9.25	0.75
267	6	7.57	-1.57	7.20	-1.20
268	7	7.80	-0.80	7.33	-0.33
269	4	5.10	-1.10	4.25	-0.25
270	10	7.57	2.43	7.20	2.80
271	8	5.52	2.48	6.00	2.00
272	7	6.14	0.86	5.78	1.22
273	6	7.38	-1.38	7.46	-1.46
274	5	5.10	-0.10	4.25	0.75
275	4	5.52	-1.52	6.00	-2.00
276	4	3.43	0.57	3.15	0.85
277	2	2.39	-0.39	2.29	-0.29
278	12	9.23	2.77	10.02	1.98
279	11	10.28	0.72	9.50	1.50
280	12	11.51	0.49	15.00	-3.00
281	5	3.43	1.57	3.15	1.85
282	8	7.99	0.01	7.55	0.45
283	8	10.28	-2.28	9.50	-1.50
284	18	11.51	6.49	15.00	3.00
285	6	5.33	0.67	5.25	0.75
286	10	10.28	-0.28	9.50	0.50
287	11	6.95	4.05	8.00	3.00
288	9	9.27	-0.27	9.00	0.00
289	6	5.52	0.48	6.00	0.00
290	5	3.86	1.14	3.73	1.27
291	12	10.86	1.14	11.29	0.71
292	8	7.99	0.01	7.55	0.45
293	1	2.39	-1.39	2.29	-1.29

294	9	7.80	1.20	7.33	1.67
295	3	3.86	-0.86	3.73	-0.73
296	10	9.04	0.96	9.25	0.75
297	13	10.47	2.53	12.98	0.02
298	10	5.72	4.28	7.00	3.00
299	4	4.67	-0.67	4.80	-0.80
300	3	3.86	-0.86	3.73	-0.73
301	12	11.32	0.68	12.00	0.00
302	1	3.43	-2.43	3.15	-2.15
303	6	5.53	0.47	6.00	0.00
304	4	3.43	0.57	3.15	0.85
305	1	2.39	-1.39	2.29	-1.29
306	7	6.33	0.67	6.47	0.53
307	3	2.81	0.19	4.36	-1.36
308	5	5.91	-0.91	5.82	-0.82
309	3	6.33	-3.33	6.47	-3.47
Sum	**361**	**341.35**	**19.65**	**351.66**	**9.34**

Table B-II: Actual vs. Predicted HBW Using MLR & ANFIS

SAMPLE #	Actual	MLR		ANFIS	
	HBW	Predicted	Residuals	Predicted	Residuals
257	2	2.03	-0.03	1.82	0.18
258	1	1.46	-0.46	1.58	-0.58
259	0	0.00	0.00	0.00	0.00
260	3	1.16	1.85	1.62	1.38
261	1	1.31	-0.31	1.44	-0.44
262	2	1.90	0.10	1.92	0.08
263	1	1.18	-0.18	1.35	-0.35
264	2	1.75	0.25	1.91	0.09
265	1	1.31	-0.31	1.44	-0.44
266	2	2.18	-0.18	2.36	-0.36
267	2	1.31	0.69	1.44	0.56
268	2	2.18	-0.18	2.36	-0.36
269	1	1.03	-0.03	1.03	-0.03
270	1	1.31	-0.31	1.44	-0.44
271	1	1.16	-0.16	1.62	-0.62
272	2	2.18	-0.18	2.36	-0.36
273	2	2.03	-0.03	1.82	0.18
274	0	1.03	-1.03	1.03	-1.03
275	1	1.03	-0.03	1.03	-0.03
276	1	1.03	-0.03	1.03	-0.03
277	0	0.00	0.00	0.00	0.00

278	1	1.46	-0.46	1.58	-0.58
279	2	2.31	-0.31	2.00	0.00
280	2	2.03	-0.03	1.82	0.18
281	1	1.18	-0.18	1.35	-0.35
282	1	1.31	-0.31	1.44	-0.44
283	2	1.75	0.25	1.91	0.09
284	3	2.03	0.97	1.82	1.18
285	2	2.18	-0.18	2.36	-0.36
286	2	2.03	-0.03	1.82	0.18
287	2	2.21	-0.21	2.40	-0.40
288	4	3.49	0.51	3.50	0.50
289	3	1.46	1.54	1.58	1.42
290	3	1.31	1.69	1.44	1.56
291	1	0.88	0.13	0.95	0.05
292	1	1.16	-0.16	1.62	-0.62
293	0	0.00	0.00	0.00	0.00
294	2	2.18	-0.18	2.36	-0.36
295	1	1.03	-0.03	1.03	-0.03
296	2	2.03	-0.03	1.82	0.18
297	1	1.16	-0.16	1.62	-0.62
298	2	2.06	-0.06	2.25	-0.25
299	1	1.18	-0.18	1.35	-0.35
300	2	1.03	0.97	1.03	0.97
301	2	3.52	-1.52	3.33	-1.33
302	0	0.88	-0.88	0.95	-0.95
303	3	2.93	0.07	2.75	0.25
304	1	1.18	-0.18	1.35	-0.35
305	0	0.00	0.00	0.00	0.00
306	1	1.18	-0.18	1.35	-0.35
307	0	0.28	-0.28	0.33	-0.33
308	1	0.88	0.13	0.95	0.05
309	0	0.88	-0.88	0.95	-0.95
Sum	**77**	**77.72**	**-0.72**	**81.66**	**-4.66**

Table B-III: Actual vs. Predicted HBE Using MLR & ANFIS

SAMPLE #	Dependent	Linear Regression		ANFIS	
	HBE	Predicted	Residuals	Predicted	Residuals
257	2	0.95	1.05	1.00	1.00
258	3	2.84	0.16	2.93	0.07
259	0	0.00	0.00	0.04	-0.04
260	0	0.00	0.00	0.04	-0.04
261	0	2.84	-2.84	2.93	-2.93

262	1	0.95	0.05	1.00	0.00
263	1	0.95	0.05	1.00	0.00
264	0	0.00	0.00	0.04	-0.04
265	0	0.00	0.00	0.04	-0.04
266	3	2.84	0.16	2.93	0.07
267	0	2.84	-2.84	2.93	-2.93
268	2	1.90	0.10	1.94	0.06
269	1	0.95	0.05	1.00	0.00
270	3	2.84	0.16	2.93	0.07
271	1	0.95	0.05	1.00	0.00
272	1	0.95	0.05	1.00	0.00
273	2	1.90	0.10	1.94	0.06
274	1	0.95	0.05	1.00	0.00
275	1	0.95	0.05	1.00	0.00
276	0	0.00	0.00	0.04	-0.04
277	0	0.00	0.00	0.04	-0.04
278	4	3.79	0.21	4.00	0.00
279	4	3.79	0.21	4.00	0.00
280	5	4.74	0.26	4.88	0.12
281	0	0.00	0.00	0.04	-0.04
282	3	2.84	0.16	2.93	0.07
283	4	3.79	0.21	4.00	0.00
284	4	4.74	-0.74	4.88	-0.88
285	0	0.00	0.00	0.04	-0.04
286	2	3.79	-1.79	4.00	-2.00
287	2	1.90	0.10	1.94	0.06
288	2	1.90	0.10	1.94	0.06
289	1	0.95	0.05	1.00	0.00
290	0	0.00	0.00	0.04	-0.04
291	7	5.69	1.31	5.96	1.04
292	3	2.84	0.16	2.93	0.07
293	0	0.00	0.00	0.04	-0.04
294	2	1.90	0.10	1.94	0.06
295	0	0.00	0.00	0.04	-0.04
296	3	2.84	0.16	2.93	0.07
297	5	4.74	0.26	4.88	0.12
298	1	0.95	0.05	1.00	0.00
299	0	0.95	-0.95	1.00	-1.00
300	0	0.00	0.00	0.04	-0.04
301	4	3.79	0.21	4.00	0.00
302	0	0.00	0.00	0.04	-0.04
303	0	0.00	0.00	0.04	-0.04
304	0	0.00	0.00	0.04	-0.04
305	0	0.00	0.00	0.04	-0.04
306	2	1.90	0.10	1.94	0.06

307	0	0.00	0.00	0.04	-0.04
308	2	1.90	0.10	1.94	0.06
309	2	1.90	0.10	1.94	0.06
Sum	**84**	**87.22**	**-3.22**	**91.21**	**-7.21**

Table B-IV: Actual vs. Predicted HBO Using MLR & ANFIS

SAMPLE #	**Actual**	**Linear Regression**		**ANFIS**	
	HBO	**Predicted**	**Residuals**	**Predicted**	**Residuals**
257	4	2.24	1.76	3.01	0.99
258	4	2.36	1.64	3.58	0.42
259	4	1.12	2.88	2.48	1.52
260	3	1.53	1.47	1.85	1.15
261	2	1.82	0.18	1.87	0.13
262	6	3.46	2.54	3.78	2.22
263	4	2.46	1.54	2.93	1.07
264	1	1.41	-0.41	1.53	-0.53
265	4	1.12	2.88	2.48	1.52
266	5	3.87	1.13	3.52	1.48
267	4	3.58	0.42	3.15	0.85
268	3	3.87	-0.87	3.52	-0.52
269	2	2.75	-0.75	3.15	-1.15
270	6	2.36	3.64	3.58	2.42
271	6	3.04	2.96	3.92	2.08
272	4	1.95	2.05	3.94	0.06
273	2	2.36	-0.36	3.58	-1.58
274	4	2.75	1.25	3.15	0.85
275	2	2.63	-0.63	3.05	-1.05
276	3	2.88	0.12	3.31	-0.31
277	2	2.05	-0.05	2.10	-0.10
278	7	3.07	3.93	4.68	2.32
279	5	3.07	1.93	4.68	0.32
280	5	3.48	1.52	4.99	0.01
281	4	0.83	3.17	2.79	1.21
282	4	3.87	0.13	3.52	0.48
283	2	3.07	-1.07	4.68	-2.68
284	11	4.29	6.71	5.14	5.86
285	4	3.04	0.96	3.92	0.08
286	6	4.29	1.71	5.14	0.86
287	7	3.71	3.29	4.64	2.36
288	3	4.16	-1.16	3.59	-0.59
289	2	2.63	-0.63	3.05	-1.05
290	2	2.34	-0.34	2.40	-0.40

291	4	3.32	0.68	4.00	0.00
292	4	3.87	0.13	3.52	0.48
293	1	2.46	-1.46	2.93	-1.93
294	5	2.24	2.76	3.01	1.99
295	2	2.34	-0.34	2.40	-0.40
296	5	3.87	1.13	3.52	1.48
297	7	5.12	1.88	6.60	0.40
298	7	3.29	3.71	3.92	3.08
299	3	0.83	2.17	2.79	0.21
300	1	2.34	-1.34	2.40	-1.40
301	6	4.70	1.30	5.88	0.12
302	1	1.63	-0.63	1.23	-0.23
303	3	2.46	0.54	2.93	0.07
304	3	2.05	0.95	2.10	0.90
305	1	1.63	-0.63	1.23	-0.23
306	4	3.17	0.83	3.05	0.95
307	3	2.34	0.66	2.40	0.60
308	2	2.88	-0.88	3.31	-1.31
309	1	3.17	-2.17	3.05	-2.05
Sum	**200**	**147.18**	**52.82**	**176.99**	**23.01**

جامعة النجاح الوطنية
كلية الدراسات العليا

تطوير نماذج تولد الرحلات بإستخدام نظام الإستدلال الضبابي المتكيف: مدينة سلفيت كحالة دراسية

إعداد
محمد كمال محمد إرشيد

إشراف
أ.د. سمير أبو عيشة

قدمت هذه الأطروحة استكمالا لمتطلبات الحصول على درجة الماجستير في هندسة الطرق والمواصلات بكلية الدراسات العليا في جامعة النجاح الوطنية في نابلس، فلسطين.

2019م

تطوير نماذج تولد الرحلات بإستخدام نظام الإستدلال الضبابي المتكيف: مدينة سلفيت كحالة دراسية

إعداد

محمد كمال محمد إرشيد

إشراف

أ.د. سمير أبو عيشة

الملخص

تعتبر نمذجة توليد الرحلات الخطوة الأولى والأساسية لعملية تخطيط المواصلات في المناطق الحضرية، حيث تهدف هذه النماذج إلى تقدير عدد الرحلات اليومية التي تنتجها الأسرة لعدة أنشطة مختلفة. في فلسطين، يعزى السبب وراء محدودية الدراسات التي تعنى بتطوير مثل هذه نماذج إلى مجموعة من التحديات الإقتصادية والإجتماعية والسياسية التي تواجه الواقع الفلسطيني، ولا سيما نقص الدعم المالي ومحدودية البيانات الموثوقة، والتي تجعل من الصعب إجراء مثل هذه الدراسات. وعلى الرغم من ذلك، تم إجراء عدد قليل من الدراسات المتخصصة لهذا الغرض، حيث تم تطويرها بإستخدام طريقة تحليل الانحدار الخطي المتعدد (MLR). هذه الطريقة تكون غير قادرة في بعض الأحيان على تطوير نماذج ملائِمة وموثوقة عند التعامل مع مجموعة من العلاقات المتداخلة والمعقدة بين العديد من المتغيرات الإجتماعية والإقتصادية.

تم تكريس هذا البحث لدراسة جدوى إستخدام طريقة جديدة نسبيًا لتحليل البيانات تسمى نظام الإستدلال الضبابي المتكيف (ANFIS)، وذلك كبديل لطريقة (MLR) المذكورة سابقا، بالإضافة إلى إستكشاف تطبيقاته ضمن السياق الفلسطيني، من خلال مقارنة أداء نماذج الرحلات المتولدة من المنازل التي تم تطويرها باستخدام هذه الطريقة، وتلك التي تم تطويرها بإستخدام (MLR).

خلال هذه الدراسة، تم تطوير أربعة أنواع من نماذج توليد الرحلات لمدينة سلفيت، نموذج (ALLTRIP) لتقدير العدد الكلي للرحلات اليومية الناتجة عن الأسرة، بغض النظر عن وقت وهدف الرحلة، ونماذج (HBW) و(HBE) لتقدير عدد الرحلات المنزلية اليومية المتولدة لأغراض

العمل والتعليم، تباعا، ونموذج (HBO) لتقدير عدد الرحلات اليومية الأخرى المتولدة لأغراض غير العمل والتعليم. وقد تم تطوير كل نموذج من هذه النماذج بإســـتخدام الطريقتين المتنافســـتين الرئيســـيتين وهما: (MLR) و (ANFIS). ثم تم اختيار النهج الأكثر ملاءمة اســـتنادًا إلى العديد من معايير التقييم، مثل القيمة الأعلى لـ (R-Squared)، والأدنى لكل من (RMSE) و(MAE)، والأقرب إلى القيم الفعلية.

في هذه الدراســـة، كانت طريقة (ANFIS) أكثر ملاءمة، وقادرة على تطوير نماذج أكثر دقة، مقارنة بطريقة (MLR) التقليدية، وذلك خلال نمذجة كلا من (ALLTRIP) و(HBO)، والتي اعتبرت أعقد نسبيا من غيرها، حيث شملت على مدى بيانات أوسع وعدد رحلات أكثر. أما بالنسبة لنماذج (HBW) و(HBW)، والتي تعتبر أقل تعقيدا، فإن كلا النهجين أنتجا نفس المســـتوى من الأداء تقريبًا، حيث كانت قيم (R-Squared) كبيرة بما فيه الكفاية لإلتقاط وتفسير معظم التباينات، كما وكانت الإختلافات بين مقاييس الأداء لهذه النماذج صـــغيرة حيث يمكن إهمالها، وبالتالي، أعتبرت طريقة (MLR) كافية لنمذجة مثل هذه الرحلات.

خلصت الدراسة إلى أن طريقة (ANFIS) تمثل تقنية نمذجة واعدة، حيث تشكل منافسًا جيدًا لنهج (MLR) التقليدي، خاصـة عند التعامل مع العديد من العلاقات المتداخلة بين العديد من المتغيرات الإجتماعية والإقتصـادية. كما وشـكلت طريقة (ANFIS) أداة مفيدة لتطوير نماذج توليد الرحلات المنزلية لمدينة ســـلفيت، حيث توصـــي هذه الدراســـة بإســـتخدام هذا النهج في تطبيقات تخطيط المواصلات ذات الصلة.