

Predicting Bus Travel Times in Washington, DC Using Artificial Neural Networks (ANNs)

Stephen Arhin, PE, PTOE, PMP, CRA, F. ITE

Babin Manandhar, EIT

Hamdiat Baba Adam

Adam Gatiba, EIT



MINETA TRANSPORTATION INSTITUTE

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the [Mineta Consortium for Transportation Mobility \(MCTM\)](#) funded by the U.S. Department of Transportation and the [California State University Transportation Consortium \(CSUTC\)](#) funded by the State of California through Senate Bill 1. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce Development

To ensure the efficient movement of people and products, we must prepare a new cohort of transportation professionals who are ready to lead a more diverse, inclusive, and equitable transportation industry. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and accuracy of the information presented herein. This document is disseminated in the interest of information exchange. MTI's research is funded, partially or entirely, by grants from the U.S. Department of Transportation, the U.S. Department of Homeland Security, the California Department of Transportation, and the California State University Office of the Chancellor, whom assume no liability for the contents or use thereof. This report does not constitute a standard specification, design standard, or regulation.

Report 21-05

Predicting Bus Travel Times in Washington, DC Using Artificial Neural Networks (ANNs)

Dr. Stephen Arhin, PE, PTOE, PMP, CRA, F. ITE
Babin Manandhar, EIT
Hamdiat Baba-Adam
Adam Gatiba, EIT

April 2021

A publication of the
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-02

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 21-05	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Predicting Bus Travel Times in Washington, DC Using Artificial Neural Networks (ANNs)		5. Report Date April 2021	
		6. Performing Organization Code	
7. Authors Stephen Arhin - https://orcid.org/0000-0002-1306-4017 Babin Manandhar - https://orcid.org/0000-0003-3623-8347 Hamdiat Baba Adam - https://orcid.org/0000-0003-0974-2812 Adam Gatiba - https://orcid.org/0000-0003-3131-0960		8. Performing Organization Report CA-MTI-1943	
		9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219	
11. Contract or Grant No. 69A3551747127			
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology University Transportation Centers Program 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period Covered Final Report	
		14. Sponsoring Agency Code	
15. Supplemental Notes DOI: 10.31979/mti.2021.1943			
16. Abstract Washington, DC is ranked second among cities in terms of highest public transit commuters in the United States, with approximately 9% of the working population using the Washington Metropolitan Area Transit Authority (WMATA) Metrobuses to commute. Deducing accurate travel times of these metrobuses is an important task for transit authorities to provide reliable service to its patrons. This study, using Artificial Neural Networks (ANN), developed prediction models for transit buses to assist decision-makers to improve service quality and patronage. For this study, we used six months of Automatic Vehicle Location (AVL) and Automatic Passenger Counting (APC) data for six Washington Metropolitan Area Transit Authority (WMATA) bus routes operating in Washington, DC. We developed regression models and Artificial Neural Network (ANN) models for predicting travel times of buses for different peak periods (AM, Mid-Day and PM). Our analysis included variables such as number of served bus stops, length of route between bus stops, average number of passengers in the bus, average dwell time of buses, and number of intersections between bus stops. We obtained ANN models for travel times by using approximation technique incorporating two separate algorithms: Quasi-Newton and Levenberg-Marquardt. The training strategy for neural network models involved feed forward and errorback processes that minimized the generated errors. We also evaluated the models with a Comparison of the Normalized Squared Errors (NSE). From the results, we observed that the travel times of buses and the dwell times at bus stops generally increased over time of the day. We gathered travel time equations for buses for the AM, Mid-Day and PM Peaks. The lowest NSE for the AM, Mid-Day and PM Peak periods corresponded to training processes using Quasi-Newton algorithm, which had 3, 2 and 5 perceptron layers, respectively. These prediction models could be adapted by transit agencies to provide the patrons with accurate travel time information at bus stops or online.			
17. Key Words Travel Time, Artificial Neural Network, Quasi-Newton Optimization Algorithm, Order Selection, Bus Transit		18. Distribution Statement No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 114	22. Price

Copyright © 2021

by **Mineta Transportation Institute**

All rights reserved.

DOI: 10.31979/mti.2021.1943

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219

Tel: (408) 924-7560

Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/1943

ACKNOWLEDGMENTS

The authors extend their appreciation to the U.S. Department of Transportation's University Transportation Centers Program and the staff of the Mineta Transportation Institute staff, including Executive Director Karen Philbrick, PhD; Deputy Executive Director Hilary Nixon, PhD, for funding and contributing to this study. The authors also thank Washington Metropolitan Area Transit Authority for providing Metrobus data that was used for the study. Any opinions, findings, and conclusions or recommendations in this material are those of the authors and do not necessarily reflect the views of the U.S. Department of Transportation.

CONTENTS

List of Figures	vi
List of Tables	x
Executive Summary	1
I. Introduction.....	4
II. Literature Review	6
2.1 Public Transportation in the United States	6
2.2 The Bus Transit System in Washington, DC	6
2.3 Advantages of the Public Transport System	7
2.4 Issues Affecting the Bus Transit System	8
2.5 Strategies to Improve the Bus Transit Service.....	10
2.6 Bus Arrival Time Prediction.....	11
2.7 Significant Variables for Predicting Bus Travel Times.....	13
2.8 Bus Transit Times Prediction Models.....	14
2.9 Artificial Neural Networks (ANNs) and Their Applications	16
2.10 Summary of Literature Review	19
III. Research Methodology.....	20
3.1 Description of the Study Area	20
3.2 Data Collection	21
3.3 Data Analysis	27
IV. Results.....	37
4.1 Summary Statistics	37
4.2 Regression Analysis.....	42
4.3 Model Testing	45
4.4 Neural Network.....	47
4.5 Comparison of Errors and Error Statistics.....	60
V. Discussion.....	69
VI. Conclusions and Recommendations	72
Endnotes	73
Bibliography.....	79
About the Authors	84
Appendices.....	85
Appendix A: SPSS Output.....	86
Appendix B: Multilayer Perceptron/Scaling and Unscaling Layers.....	89
Appendix C: Neural Network Model Architecture	92
Appendix D: Neural Designer Model Training Plots.....	97
Appendix E: Neural Network Error Tables	105
Appendix F: Neural Network Travel Time Equations	108

LIST OF FIGURES

Figure 1: WMATA Metrobus with APC and AVL Schematics	7
Figure 2: Illustration of ANN	17
Figure 3: Sigmoid Activation Function Curve	18
Figure 4: Hyperbolic Tangent Function Curve	18
Figure 5: Map of Washington, DC	20
Figure 6: Map of Operation Paths of Selected Routes.....	22
Figure 7: Sample Datasheet Obtained from WMATA (Route 32)	24
Figure 8: Sample Representation of a Bus Service Route	25
Figure 9: Snapshot of SPSS Interface and Sample Output.....	27
Figure 10: Snapshot of Neural Designer Interface and Sample Output	28
Figure 11: Example Normal Probability Plot.....	30
Figure 12: Example Scatter Plot Showing Homoscedasticity	31
Figure 13: ANN Architecture	33
Figure 14: Total Number of Data Points per Peak Period used for the Analysis	37
Figure 15: Distribution of Data Points per Peak Period for the Different Numbers of Served Bus Stops	38
Figure 16: Average Dwell Time During Different Peak Periods.....	39
Figure 17: Average Number of Passengers Onboard During Different Peak Periods.....	40
Figure 18: Graph Representing Travel Time of Buses vs. Number of Intersections in between Served Bus Stops During the AM Peak Period.....	40
Figure 19: Graph Representing Travel Time of Buses vs. Number of Intersections in between Served Bus Stops During the Mid-Day Peak Period	41
Figure 20: Graph Representing Travel Time of Buses vs. Number of Intersections in between Served Bus Stops During the PM Peak Period	41
Figure 21: Normality Probability Curve Obtained for AM Peak Period.....	46
Figure 22: Homoscedasticity Regression Plot for AM Peak Period	47
Figure 23: Typical Neural Network Architecture for a Network with Two Perceptron Layers	48
Figure 24: Quasi-Newton Method Error History Plot for AM Peak Period (Two Perceptron Layers).....	49
Figure 25: Training and Selection Errors Obtained Before and After Order Selection (Quasi-Newton Method with Two Perceptron Layers)	61
Figure 26: Training and Selection Errors Obtained Before and After Order Selection (Quasi-Newton Method with Three Perceptron Layers)	62
Figure 27: Training and Selection Errors Obtained Before and After Order Selection (Quasi-Newton Method with Five Perceptron Layers)	63
Figure 28: Training and Selection Errors Obtained Before and After Order Selection (Levenberg-Marquardt Method).....	64

Figure 29: Normalized Squared Errors Obtained for AM Peak Period Neural Network Analyses	65
Figure 30: Normalized Squared Errors Obtained for Mid-Day Peak Period Neural Network Analyses	65
Figure 31: Normalized Squared Errors Obtained for PM Peak Period Neural Network Analyses	66
Figure 32: Mean Absolute and Mean Percentage Errors Obtained after Neural Network Analyses	67
Figure 33: Normality Probability Curve Obtained for Mid-Day Peak Period (top) and PM Peak Period (bottom)	86
Figure 34: Homoscedasticity Regression Plot for Mid-Day Peak Period.....	87
Figure 35: Homoscedasticity Regression Plot for PM Peak Period.....	88
Figure 36 Quasi-Newton Method Error History Plot for Mid-Day Peak Period (Two Perceptron Layers).....	92
Figure 37: Quasi-Newton Method Error History Plot for PM Peak Period (Two Perceptron Layers).....	92
Figure 38: Quasi-Newton Method Error History Plot after Order Selection for AM Peak Period (Two Perceptron Layers).....	93
Figure 39: Quasi-Newton Method Error History Plot after Order Selection for Mid-Day Peak Period (Two Perceptron Layers)	93
Figure 40: Quasi-Newton Method Error History Plot after Order Selection for PM Peak Period (Two Perceptron Layers)	93
Figure 41: Quasi-Newton Method Error History Plot for AM Peak Period (Three Perceptron Layers).....	94
Figure 42: Quasi-Newton Method Error History Plot for Mid-Day Peak Period (Three Perceptron Layers).....	94
Figure 43: Quasi-Newton Method Error History Plot for PM Peak Period (Three Perceptron Layers).....	94
Figure 44: Quasi-Newton Method Error History Plot after Order Selection for AM Peak Period (Three Perceptron Layers).....	95
Figure 45: Quasi-Newton Method Error History Plot after Order Selection for Mid-Day Peak Period (Three Perceptron Layers)	95
Figure 46: Quasi-Newton Method Error History Plot after Order Selection for PM Peak Period (Three Perceptron Layers)	96
Figure 47: Quasi-Newton Method Error History Plot for AM Peak Period (Five Perceptron Layers).....	96
Figure 48: Quasi-Newton Method Error History Plot for Mid-Day Peak Period (Five Perceptron Layers).....	97
Figure 49: Quasi-Newton Method Error History Plot for PM Peak Period (Five Perceptron Layers).....	97

Figure 50: Quasi-Newton Method Error History Plot after Order Selection for AM Peak Period (Five Perceptron Layers)	98
Figure 51: Quasi-Newton Method Error History Plot after Order Selection for Mid-Day Peak Period (Five Perceptron Layers).....	98
Figure 52: Quasi-Newton Method Error History Plot after Order Selection for PM Peak Period (Five Perceptron Layers)	98
Figure 53: Levenberg-Marquardt Method Error History Plot for AM Peak Period (Three Perceptron Layers).....	99
Figure 54: Levenberg-Marquardt Method Error History Plot for Mid-Day Peak Period (Three Perceptron Layers)	99
Figure 55: Levenberg-Marquardt Method Error History Plot for PM Peak Period (Three Perceptron Layers)	99
Figure 56: Levenberg-Marquardt Method Error History Plot after Order Selection for AM Peak Period (Three Perceptron Layers)	100
Figure 57: Levenberg-Marquardt Method Error History Plot after Order Selection for Mid-Day Peak Period (Three Perceptron Layers)	100
Figure 58: Levenberg-Marquardt Method Error History Plot after Order Selection for PM Peak Period (Three Perceptron Layers)	100
Figure 59: Neural Network Architecture for AM Peak Model with Two Perceptron Layers (Quasi-Newton Analysis).....	101
Figure 60: Neural Network Architecture for AM Peak Model with Three Perceptron Layers (Quasi-Newton Analysis).....	101
Figure 61: Neural Network Architecture for AM Peak Model with Five Perceptron Layers (Quasi-Newton Analysis).....	101
Figure 62: Neural Network Architecture for AM Peak Model with Three Perceptron Layers (Levenberg-Marquardt Analysis).....	102
Figure 63: Neural Network Architecture for Mid-Day Peak Model with Two Perceptron Layers (Quasi-Newton Analysis).....	102
Figure 64: Neural Network Architecture for Mid-Day Peak Model with Three Perceptron Layers (Quasi-Newton Analysis).....	102
Figure 65: Neural Network Architecture for Mid-Day Peak Model with Five Perceptron Layers (Quasi-Newton Analysis).....	103
Figure 66: Neural Network Architecture for Mid-Day Peak Model with Three Perceptron Layers (Levenberg-Marquardt Analysis).....	103
Figure 67: Neural Network Architecture for PM Peak Model with Two Perceptron Layers (Quasi-Newton Analysis).....	103
Figure 68: Neural Network Architecture for PM Peak Model with Three Perceptron Layers (Quasi-Newton Analysis).....	104

Figure 69: Neural Network Architecture for PM Peak Model with Five
Perceptron Layers (Levenberg-Marquardt Analysis)..... 104

Figure 70: Neural Network Architecture for PM Peak Model with Three
Perceptron Layers (Quasi-Newton Analysis)..... 104

LIST OF TABLES

Table 1: Summary of Regression Analysis Results.....	2
Table 2: Summary of ANN Model Analysis Results	3
Table 3: Sample Peak Period ANN Data Matrix Model for a Bus Route	32
Table 4: Average Travel Time by Number of Served Bus Stops	38
Table 5: Descriptive Statistics.....	39
Table 6: Inputs-Target Correlation for AM Peak Period.....	43
Table 7: Results of the Regression Analyses for AM Peak Period	43
Table 8: Results of the Regression Analyses for Mid-Day Peak Period.....	44
Table 9: Results of the Regression Analyses for PM Peak Period.....	45
Table 10: SPSS Coefficient Output for AM Peak Period	46
Table 11: Summary of Results.....	47
Table 12: Quasi-Newton Method Results for Two Perceptron Layers.....	49
Table 13: Quasi-Newton Method Results for Two Perceptron Layers.....	50
Table 14: Quasi-Newton Method Results of All Peak Periods for Two Perceptron Layers (After Order Selection).....	50
Table 15: Normalized Squared Errors for Training, Selection and Testing Instances (Two-Layer Quasi-Newton Method).....	51
Table 16: Error Statistics for Quasi-Newton Method for All Peak Periods (Two Perceptron Layers).....	51
Table 17: Neural Network Analysis Output Equations	51
Table 18: Quasi-Newton Method Results for Three Perceptron Layers.....	52
Table 19: Quasi-Newton Method Results for Three Perceptron Layers (All Peak Periods).....	53
Table 20: Quasi-Newton Method Results of All Peak Periods for Three Perceptron Layers (After Order Selection).....	53
Table 21: Errors for Training, Selection and Testing Instances (Three-Layer Quasi-Newton Method).....	54
Table 22: Error Statistics for Quasi-Newton Method for All Peak Periods (Three-Perceptron Layers)	54
Table 23: Quasi-Newton Method Results for Five Perceptron Layers	55
Table 24: Quasi-Newton Method Results for Five Perceptron Layers (All Peak Periods).....	56
Table 25: Quasi-Newton Method Results of All Peak Periods for Five Perceptron Layers (After Order Selection).....	56
Table 26: Errors for Training, Selection and Testing Instances (Five-Layer Quasi-Newton Method).....	57
Table 27: Error Statistics for Quasi-Newton Method for All Peak Periods (Five Perceptron Layers).....	57

Table 28: Levenberg-Marquardt Method Results for Three Perceptron Layers (All Peak Periods).....	58
Table 29: Levenberg-Marquardt Method Results (All Peak Periods).....	59
Table 30: Levenberg-Marquardt Method Results of All Peak Periods (After Order Selection).....	59
Table 31: Errors for Training, Selection and Testing Instances (Three-Layer Levenberg-Marquardt)	60
Table 32: Error Statistics for Levenberg-Marquardt Method for All Peak Periods (Three Perceptron Layers).....	60
Table 33: SPSS Coefficient Output for Mid-Day Peak Period.....	87
Table 34: SPSS Coefficient Output for PM Peak Period.....	87
Table 35: Perceptron Layers for Quasi-Newton Analysis (Two Layers).....	89
Table 36: Perceptron Layers for Quasi-Newton Analysis (Three Layers).....	89
Table 37: Perceptron Layers for Quasi-Newton Analysis (Five Layers).....	89
Table 38: Perceptron Layers for Levenberg-Marquardt Analysis (Three Layers).....	89
Table 39: Scaling and Unscaling Values for Quasi-Newton Analysis (Two Perceptron Layers).....	89
Table 40: Scaling and Unscaling Values for Quasi-Newton Analysis (Three Perceptron Layers).....	90
Table 41: Scaling and Unscaling Layers for Quasi-Newton Analysis (Five Perceptron Layers).....	91
Table 42: Scaling and Unscaling Layers for Levenberg-Marquardt Analysis (Three Perceptron Layers).....	91
Table 43: Errors for Training, Selection, and Testing Instances (Two Layers, Quasi-Newton Method).....	105
Table 44: Error Statistics for Quasi-Newton Method for All Peak Periods (Two Perceptron Layers).....	105
Table 45: Errors for Training, Selection, and Testing Instances (Three Layers, Quasi-Newton Method).....	105
Table 46: Error Statistics for Quasi-Newton Method for All Peak Periods (Three Perceptron Layers).....	106
Table 47: Errors for Training, Selection and Testing Instances (Five Layers, Quasi-Newton Method).....	106
Table 48: Error Statistics for Quasi-Newton Method for All Peak Periods (Five Perceptron Layers).....	106
Table 49: Errors for Training, Selection, and Testing Instances (Three Layers, Levenberg-Marquardt)	107
Table 50: Error Statistics for Levenberg-Marquardt Method for All Peak Periods (Three Perceptron Layers).....	107

Executive Summary

The accurate prediction of travel time has become necessary, and research in that direction has made huge leaps and bounds over the years to enable public transit agencies to provide patrons with efficient transit services and to help them better plan their commutes. The Washington, DC area is ranked second in terms of states in the US with the highest percentage of public transit commuters. Over the years, many interventions have been introduced to improve the overall public transit system by upgrading to Intelligent Transport Systems, which can enable the provision of better transit travel times prediction, fare payment systems, passenger boarding structures, monitoring processes, etc.

This report presents the findings of a study conducted to predict bus transit travel times in Washington, DC using Artificial Neural Network (ANN). The ANN models were developed based on AVL (Automatic Vehicle Location) and APC (Automatic Passenger Counters) data with the aim of helping Washington Metropolitan Area Transit Authority (WMATA) to improve bus stop arrival times and consequently improve transit travel times, which would potentially lead to rider satisfaction and increased ridership. In this project, six months' worth (January–June 2019) of AVL and APC data for bus routes operating in the DC area were used. Bus routes on arterial and collector roadways were selected for this study. Other criteria for route selection considered included routes with high patronage bus stops, routes with longer headways, and routes with stops in close proximity to metro rail stations. Based on the selection criteria, six bus routes were selected for the study: 70, 32, 52, 42, D4, and S1.

For this research, information extracted from the data for each route included departure and arrival times, length of route, passenger loading and unloading, dwell time, travel time, and bus location. Field observations were conducted to collect the number of intersections between the origin and destination and the speed limit of each bus route. To reduce the effects of the variability in traffic characteristics, the extracted data were partitioned according to AM Peak (7:00 AM–9:30 AM), Mid-Day Peak (10:00 AM–2:30 PM), and PM Peak (4:00 PM–6:30 PM) periods, and separate ANN models were developed for each period. For each route, a minimum sample size of 500 origin-to-destination trips per peak period was extracted, and a total minimum sample of 1,500 was used in the ANN model. For the ANN model, travel time was designated as the dependent variable while the independent variables included number of served bus stops, route length between bus stops, average number of passengers in the bus, average dwell time at a bus stop, and number of intersections between served bus stops. Regression and ANN models were developed to predict travel times for the selected bus routes.

The following are the summary of the results from the regression analyses:

- Travel times generally increased from morning through the evening peak periods for all the bus routes selected.

- In the AM Peak period, the independent variable, “number of bus stops served,” appeared to have the greatest effect on the total bus travel time, followed by the “number of intersections between two served stops” independent variable. Average dwell time had the least effect on bus travel time. The independent variable “number of passengers” had an inverse relationship with travel time.
- For the Mid-Day Peak period, distance between served bus stops variable appeared to have the greatest effect on bus travel time, followed by the number of intersections between two served stops. The average dwell time variable had the least effect on bus travel time.
- The number of intersections between two served stops variable appeared to have the greatest effect on the travel time model for the PM Peak period. This was followed by the number of served bus stops, with the average dwell time having the least effect on bus travel time.

A summary of the models with the accompanying coefficient of determination (R^2) values are presented in Table 1.

Table 1: Summary of Regression Analysis Results

#	Peak Period	Model	R^2
1	AM	$TT_A = 40.36 + 62.50SBS + 0.11L - 1.37P + 2.17DT + 7.41T$	0.76
2	Mid-Day	$TT_M = 41.07 + 31.94SBS + 0.08L + 0.12P - 1.51DT - 6.97T$	0.85
3	PM	$TT_P = -6.063 + 55.67SBS + 0.11L + 1.88P + 0.89DT + 15.24T$	0.68

Artificial Neural Network (ANN) models were developed to predict buses’ travel times during different peaks using Neural Designer software based on Quasi-Newton and Levenberg-Marquardt optimization algorithms. Optimization algorithms determine how parameters are adjusted in the neural network training process that would yield the lowest error.

A summary of the results from the ANN analyses follows:

- The Quasi-Newton optimization algorithm is the default optimization method in Neural Designer which was used for training the datasets for the three peak periods for all six bus routes. The algorithm helps determine the local maxima and minima of functions (greatest and least amount of errors) and is a faster alternative to finding zeros of a function via Newton’s method. Three sets of perceptron layers (two layers, three layers, and five layers) were used to determine the optimum number of hidden layers for travel time prediction. Perceptron layers can be visualized as algorithms that enable a neural network to learn. Increasing the number of layers can aid to solve complex problems by outperforming single layer networks.
- The Levenberg-Marquardt optimization algorithm was also used to train the model for all three peaks of the six bus routes using an analysis with three perceptron layers. The algorithm can be used to find the local minimum of a function while fitting a set of observations with a non-linear model.

- From the results (Table 2), the two-layer perceptron network using the Quasi-Newton optimization algorithm produced the lowest Mean Absolute Error (MAE) for the Mid-Day Peak period. The lowest MAE for the PM Peak period was obtained in the Quasi-Newton five-layer perceptron network. The Levenberg-Marquardt model had the highest MAE values for all the peak periods.
- After both training and testing errors were analyzed, the Quasi-Newton algorithm two-layer perceptron model had the lowest Normalized Squared Errors (NSE) consistently for most of the peak periods. The three-layer perceptron Quasi-Newton model produced the second-lowest testing errors. The Levenberg-Marquardt Optimization model resulted in the highest testing NSE for all peak periods, indicating that it was not suitable for the datasets.

Table 2 shows a summary of the analysis for each peak period, highlighting the lowest obtained measures of effectiveness.

Table 2: Summary of ANN Model Analysis Results

Peak Period	Lowest Normalized Squared Error	Lowest Mean Percentage Error (%)
AM Peak Period	0.145 ¹	4.591 ²
Mid-Day Peak Period	0.053 ²	1.510 ²
PM Peak Period	0.254 ³	4.797 ³

¹ Quasi-Newton Analysis (three perceptron layers)

² Quasi-Newton Analysis (two perceptron layers)

³ Quasi-Newton Analysis (five perceptron layers)

The results of the analyses indicate that ANN models can effectively predict the travel times of buses on selected routes with minimal percentage errors when combined with traditional multiple regression analyses. The ANN models could be incorporated into existing predictive models used by WMATA to provide patrons with travel time information at bus stops and online. These models could be adopted by transit agencies in other jurisdictions with similar characteristics to the Washington, DC area. For future work, these models could be calibrated using real-time data for arterial and collector bus routes. Also, similar models could be developed for bus routes that serve residential or local roads.

I. Introduction

Washington, DC has the second-highest percentage of public transit commuters in the United States, with New York City listed as number one. Of the working population that commute to and from Washington, DC, approximately 9% use Metrobus,¹ which is a bus service overseen by the Washington Metropolitan Area Transit Authority (WMATA). WMATA has of approximately 1,600 buses that operate in all parts of the District of Columbia, some parts of Maryland and Virginia with over 269 bus routes. The buses serve 11,129 stops, including 2,554 stops with bus shelters.² The accurate prediction of travel time is necessary to enable public transit agencies to provide patrons with efficient transit service and for them to effectively plan their commute or travel in the region. Transit agencies are continuously evaluating best practices available to improve the reliability of their services. The use of technology, particularly in bus transit, has been critical for this purpose. This includes the use of Automatic Vehicle Location (AVL) technology installed in public transit buses that track of buses in real time. AVL technology employs Global Positioning System (GPS) installed onboard transit buses to track vehicle location and display it on a geographical map of the area. The anticipated arrival times and the travel times of transit buses are estimated by using the AVL data, and this information is collected by transit agencies and shared with patrons at 30-second intervals via Long-Term Evolution (LTE), Third Generation (3G), or Fourth Generation (4G) telecommunications technology.

Automatic Passenger Counters (APCs) installed on buses count the number of passengers getting on and off at each bus stop, which helps in the computation of the total number of patrons onboard. Despite these efforts, there are various factors—including traffic congestion, roadway conditions, the frequency of bus stops, and/or unforeseeable circumstances—that affect the public transit agencies' ability to provide accurate arrival and travel times to patrons.

Arhin et al. (2013) evaluated the reliability of transit buses in Washington, DC. The study found that transit bus services had an overall on-time performance of approximately 75%. The mean deviation between the scheduled arrival times and the actual arrival times ranged between 1.99 and 5.03 minutes. These deviations in arrival times cumulatively affected the accuracy of the predicted travel times between the origin and destination points.³ This may result in a decrease in patrons' perception of the bus services' reliability. It is therefore important that travel time prediction models are developed to provide more accurate information to patrons based on pertinent factors that affect travel time.

Previous studies have used conventional linear regression methods to develop models that predict travel times. Jeong and Rilett (2004) proposed a set of multilinear regression models to estimate the travel times of buses from an origin bus stop to a target bus stop. However, in terms of accuracy, these models were found to be outperformed by machine learning models such as Artificial Neural Networks (ANNs) which provide a much more effective alternative to the conventional models.⁴

ANN are mathematical models that are inspired by the biological neural networks in the human brain. The effectiveness of ANN is based on its ability to approximate both linear and nonlinear functions to a required degree of accuracy using a learning algorithm, and to build “piece-wise” approximations of the functions.

Hence, this research is aimed at developing ANN models to predict the travel times of transit buses in Washington, DC using AVL and APC data. The models will enable public transit agencies to provide more accurate information to patrons to improve reliability and consequently increase bus ridership.

II. Literature Review

2.1 Public Transportation in the United States

Public transportation became popular in the 1800s and has since gone through many technological advancements to become what it is presently.⁵ In the United States, the public transportation system includes bus, rail, ferry, and airline services which are publicly financed and accessible to the general populace. They are, however, more established and accessible in central urban areas where there is higher public demand compared to suburban and rural areas. Depending on the type of service rendered, public transit services are classified as either national, regional, or commuter/local.⁶

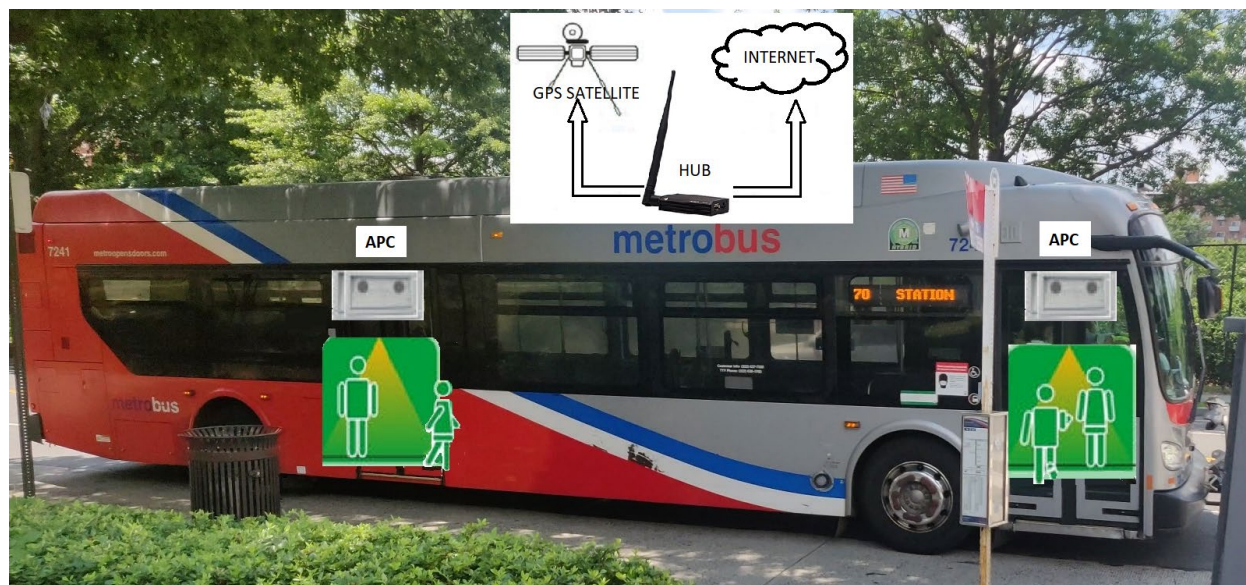
Conventional or commuter buses, bus rapid transit, and intercity buses are the three common bus services in the United States. The hours, service frequency, and routes of these bus services are determined by the needs of the city or community with the goal of providing maximum convenience while catering to the needs of the majority.⁷ Despite increases in auto vehicle ownership, public transportation ridership has steadily increased to an average national rate of 21% over the years. This makes the public transportation sector a viable and important aspect of the American economy.⁸

2.2 The Bus Transit System in Washington, DC

The metro system in Washington, DC, comprises the regional bus and rail public transit which operates within DC, Maryland, and Virginia (DMV). The system is run by the Washington Metropolitan Area Transit Authority (WMATA) with contributions from the various cities and counties towards its operational costs.⁹

The region's Metrobus service is the fifth-largest bus system in the United States. It has over 1,450 buses and services approximately 350 routes across the DMV area. Its services are supplemented by the bus systems of the various local jurisdictions.¹⁰ Figure 1 shows a WMATA bus with schematics of APC and AVL systems. The APC counts the number of passengers getting on and off the bus while the AVL tracks the location of the bus in real time. The data is shared via internet to the WMATA database.

Figure 1: WMATA Metrobus with APC and AVL Schematics



Public transportation is regarded as one of the safest and most convenient ways of moving around the DC area, especially due to traffic congestion, several parking restrictions, and inadequate and/or expensive parking facilities. The DC Circulator and the Metrobus are the two main options for bus commuters and provide services to varied locations within DC and the larger DMV area. The Metrobus, for instance, provides 24-hour daily service seven days a week and connects to the rail system and other local bus systems.¹¹

2.3 Advantages of the Public Transport System

Despite increased ownership of auto vehicles, studies show that public transport ridership has increased by 21% since 1997. This indicates higher patronage, especially when compared to the 19% increase in population over that same period. Public transportation benefits have been far-reaching on both individual and community levels. Notable among these benefits is increased economic development as areas farther out and hitherto undeveloped become easily accessible. According to the American Public Transportation Association (APTA), a return of \$4 is generated for every \$1 invested, and for every \$10 million of capital injected into public transportation, there was a \$30 million increase in business sales. In the United States, public transportation agencies directly employ over 400,000 people and create and support another 50,000 jobs for every \$1 billion invested.¹²

Further, studies show that using public transportation can reduce a person's risk of being in a motor vehicle accident by 90%.¹³ This could be attributed to the level and regularity of public vehicle maintenance as well as the training and enforced driving habits of public transit drivers.¹⁴

Public transport also helps individuals to save money on vehicle purchases, maintenance, fuel, and parking. According to APTA, in the United States, approximately 16 cents of every dollar earned goes into transportation—the second-largest expenditure after housing. Thus, households that

patronized public transportation and had one fewer automobile could save about \$10,000 annually.¹⁵

Moreover, public transportation provides increased mobility and ease in going about one's daily business without having to rely on the help of others. It enables riders to save time on commutes as well as engaging in other activities such as reading, working, or studying that would not have been possible if they were driving. Studies also show that people who use public transportation get more physical activity compared to those who don't, since most of them walk to and from their transit stops. This leads to an overall healthier lifestyle.¹⁶

Additionally, using public transportation helps to reduce air pollution from carbon emissions. Studies show that communities with increased public transit usage reduce carbon emissions by 37 million metric tons yearly.¹⁷ Comparatively, buses emit 20% less carbon monoxide, 90% less hydrocarbons, and 25% less nitrogen oxides per passenger mile when compared to a vehicle with a single occupant.¹⁸ Also, public transportation leads to fuel savings for the country as a whole. According to APTA, public transportation saves the United States approximately 4.2 billion gallons of gasoline every year.¹⁹ This could be attributed to the significantly decreased amount of fuel per passenger mile due to the transportation efficiency of public modes.²⁰

2.4 Issues Affecting the Bus Transit System

Although public transportation ridership has increased over the years, there are still some challenges faced by the system that make it unpopular among the people. Notable among these include inadequate accessibility, overcrowding on buses, delayed bus arrival times, delayed or longer commute times, and the overall cost to the rider per commute.²¹

Studies show that about 45% of Americans have no access to public transportation despite the progress made in public transportation over the years.²² This can be generally attributed to rapid development in housing infrastructure farther away from town or city centers without corresponding development in the public transit system. Thus, housing communities were built farther away from downtowns with no planning for public transport within those areas. This creates complex transit systems and leads to automobile dependence.²³

In areas where public transport is accessible, there tends to be increased patronage, especially around peak times. This usually leads to overcrowding on buses where riders are subjected to standing, sometimes throughout their travel time. A study conducted by Li and Hensher (2013) on crowding in public transport in some developed countries indicated that despite the various measures put in place by transport authorities to identify and control crowding on public transportation, overcrowding still occurs on many public transit systems. Survey results indicated that apart from the discomfort associated with standing for a long time, riders mentioned that it also leads to stress, physical exhaustion, less privacy, health concerns due to rider proximity, and overall dissatisfaction. These experiences were very likely to influence a rider's choice to use public transportation or rely on other alternate modes. The study only took into consideration the density

of people and the physical space available without considering riders' social and emotional perceptions of crowding.²⁴

Over the years, technology has been applied in the public transportation sector to improve its effectiveness and efficiency as well as rider satisfaction. One such introduction is the bus arrival prediction times provided using the AVL technology mounted on buses to provide up-to-date real-time bus location data. A study conducted by Arhin et al. (2013) to evaluate the reliability of transit buses in Washington, DC showed an overall performance of 75% with an average deviation of two to five minutes from the predicted arrival times. These results were consistent with WMATA's window of two-minutes-early and seven-minutes-late bus arrival times. However, the statistics from the study showed that about 82% of the bus arrival times failed to meet the general transit industry threshold.²⁵ However, for increased ridership, retention, and overall customer satisfaction, bus service reliability is important.

Delayed commute times, which are directly related to inconsistencies in the buses' actual arrival and predicted arrival times, represent one major reason that makes public transportation unattractive for riders. This is due to a combination of factors which include lower speed limits, traffic congestion, delay from traffic signals, and delays from a large number of passenger stops along the bus route. The average bus speed through downtown DC is less than 12 mph.²⁶ A study conducted by Estrada et al. analyzed strategies for bus travel time control along corridors with signalized intersections. One significant conclusion from their study is that the reduction in the speed of buses below the recommended segment speed resulted in a significantly higher round trip travel time.²⁷

Traffic congestion in the United States caused an average loss of 97 hours a year at an average cost of \$1,348 per driver, totaling \$87 billion dollars for the year 2018. This amount is projected to increase to about \$186 billion by 2030. Washington, DC was identified as the third most congested city with drivers having to spend about 102 hours in traffic per year. Apparently, congestion is more prevalent in cities and urban areas with sprawling settlements situated more towards the outskirts, whereas the hub of business is typically within the town center.²⁸ The efforts of city officials to designate bus-only lanes has not been totally successful, as buses are still affected in traffic congestion zones like other auto vehicles, especially during peak periods. The impact of signalized intersections, although relatively small, has been found to add up, especially on longer routes with more signalized intersections. A study conducted by Albright and Figliozzi (2012) determined that an average of 8 to 26 seconds was added to bus travel time for each signalized intersection on the bus route.²⁹ In another study conducted to gain insight into time variability in public transport travel, Mazloumi et al. (2010) found that travel time variation increased by up to 22% for each additional signalized intersection per kilometer.³⁰ Figliozzi and Feng (2012) found out that for each stop sign, an average of 12 to 16 seconds was added to bus travel time, and turning left and right at intersections added an average 5 to 38 seconds to bus travel time.³¹ Mazloumi et al. (2010) found that the number of bus stops along a route had an influence on bus travel time variability especially in the AM Peak and that the length of the route along with the number of

bus stops affected the bus arrival and departure times and thereby increased the total travel time of the bus trip.³²

Generally, public transportation provides cheaper or more competitive pricing when it comes to mobility. This is usually the case, especially in urban areas where riders are travelling shorter distances. However, in the case where a rider would require two or more transfers, which in most cases encompass several public transport options, it becomes costly both in time and money. A study by Owen and Levinson (2015) found that bus transit ridership was higher among low-income households compared to higher income households. In addition, where riders had to wait for up to 30 minutes to access the adjoining ride to continue their trip, they preferred to use a different mode of transportation, particularly auto vehicles.³³ A study on the impact of fare integration on travel behavior and transit ridership by Sharaby and Shiftan (2012) showed that the introduction of a single fare system and free transfers across buses along routes led to an increase of 25% in bus transit ticket sales, 7.7% in passenger trips, and 18.6% in bus boarding.³⁴ This implies that the reduction in fares for single ticket purchasing and free transfers for riders who had to use more than one bus to reach their destination increased rider patronage of the bus transit system.

2.5 Strategies to Improve the Bus Transit Service

There have been several interventions within cities and urban areas, especially where public transportation is often used, to improve service to patrons. These have been implemented to address some of the dissatisfactions experienced by riders to make the public transportation system more efficient to current users and appealing enough to attract more users. Designated bus lanes, improved traffic management and AVL technology, improved facilities and amenities, and faster boarding at bus stops are a few of the strategies that have been employed to improve the efficiency of bus transit systems.

Research has shown that designated bus lanes have the tendency to reduce bus delay times by about 10%. A study conducted in New York showed that bus delay times was reduced from 34% to 42% and ridership increased by 10% after the introduction of designated or dedicated bus lanes. In other areas within the United States, two-lane busways were also employed on highways and freeways to enable rapid bus transit services.³⁵

Traffic signal priority for buses has also been considered in some jurisdictions as a strategy to reduce bus service delays and make it attractive for ridership and a preferred mode of transport over auto vehicles. Over the years, two general types of systems have been employed. One is a pre-installed electronic communications device that enables the bus driver to advance the traffic signal cycle to green to let them pass through the intersection if maintaining the bus schedule is needed. The second is the use of Automatic Vehicle Location (AVL) technology coupled with an advanced radio communications system that allows a computerized system to determine the bus location in relation to its schedule; the system is able to control traffic signals, giving priority to buses as and when needed.³⁶ Estrada et al. (2016) proposed a dynamic bus control model that utilized real-time

bus tracking data (AVL) at stops to control bus speeds and potentially also induce signalized intersections to extend the green phase for significantly delayed buses. The effects of this simulated model, which took into consideration passenger travel time, operating costs, and bus travel time variability, reduced the total system cost for both user and agency by 15–40% and reduced bus travel time variability by 53–78%.³⁷

Improved facilities and amenities at subways, bus stops, and aboard buses give riders the full experience from their origin to their destination. The overall satisfaction with service provided is very important for ridership retention and the attraction of more riders. A study of the Curitiba's bus transit system in Brazil showed that improved station platforms provided easier same-level boarding access, shielded riders from weather conditions, and provided a means for passengers to pay their fare before boarding. This helped to reduce the bus dwell time by 15 to 19 seconds. Additional improvements in the transit system included single fare with unlimited transfers, exclusive bus lanes, signal priority for buses, bi-articulated buses with large capacity doors and express bus transit services. These interventions led to a reduction of 27 million auto trips per year, saving about 27 million liters of fuel annually. On the whole, Curitiba uses about 30% less fuel per capita, and public transportation accounts for 55% of all private trips in the city.³⁸

According to an Organisation Gestion Marketing report, results obtained from eight European cities after the implementation of fare integration platforms reported a significant increase in public travel behavior after the first one to two years.³⁹ These results were consistent with those from a study conducted by Sharaby and Shiftan (2012) that investigated the impact of fare integration on transit ridership and travel behaviors in Haifa, Israel. In that case, the change of the complex pre-boarding fare system into a single ticket system and the provision of free transfers between the five zones reduced fares for many passengers and led to a 25% increase in single ticket sales, a 7.7% increase in passenger trips and an 18.6% increase in bus boarding.⁴⁰

2.6 Bus Arrival Time Prediction

Several studies have been conducted over the years on the deviations that exist between the predicted arrival times and the actual arrival times of buses at designated stops along a route. Different types of prediction models have been used in these analyses that include Date model, Time Series Model, Regression model, Kalman filtering model, Artificial Neural Network (ANN) and Support Vector Machines (SVM).⁴¹ The following paragraphs provide an overview of the application of different prediction models in various bus travel time prediction studies.

One of the earliest models used to predict bus travel time was conducted by D'Angelo et al. (1999). In that study, a non-linear time series model was used to predict short-term travel time on a freeway network. To predict corridor travel times and recurrent congestion locations along a road corridor, a multiple-variable prediction model and a calibrated single-variable model were compared. The single variable was based on minimum speed prediction permitted, whereas the multivariable model was based on speed, occupancy, and volume data. Despite the low accuracy

and unsatisfactory real-time performance, the results from the calibrated single-variable prediction model were better than those of the multiple-variable prediction model with reasonable errors for short-term travel time predictions.⁴²

Patnaik et al. (2004) used data from Automatic Passenger Counting (APC) systems installed on buses to predict bus travel time. The buses operated along a 30-mile urban bus route (inbound or outbound trips) length over a period of six months. Right and left turn movements at all signalized intersections as well as all bus stops along these routes were recorded to be used in the prediction model. Multivariable regression models were used to estimate bus arrival times under various conditions. These models made use of data obtained from the APC; service direction (inbound or outbound), number of passengers boarding and/or alighting, distance between stops, dwell time, peak periods and other variables. Results from the study showed that variations in the predicted and actual bus arrival times were less pronounced in the AM Peak than in the PM Peak and that the actual and estimated travel times were reasonably close to each other with a deviation of about 2.45 minutes.⁴³ However, the regression models are based on the assumption that the variables are independent of each other, which, in terms of traffic data, would be impractical.

Shalaby and Farhan (2002) conducted a study to develop a bus travel time prediction model for dynamic operations control and passenger information systems. In the research, dynamic data from AVL and APC systems were used to develop various prediction models and tested using VISSIM microsimulation under different scenarios. The results revealed that the Kalman Filter algorithm model outperformed all the other models developed in terms of accuracy and also demonstrated the ability to dynamically update itself to reflect new data and changing scenarios along the transit route.⁴⁴ A study by Vanajakshi et al. (2009) was conducted to predict bus travel times under heterogeneous traffic conditions using global positioning system data and an algorithm based on the Kalman Filtering technique. The results from the study determined that the predicted and actual bus arrival times were reasonably close, with an average percentage error varying from 0.22 to 13.75%.⁴⁵

A study by Yin et al. (2017) relied on the use of ANN and Support Vector Machine (SVM) models to predict bus arrival times at stops with multiple bus services. Their study had three main input variables: the weighted travel time of preceding buses along the chosen route, the travel speed of the selected bus, and the weighted travel time of preceding buses on different routes intersecting with the chosen route. The results showed that both the ANN and SVM models have high accuracy in predicting bus arrival times, but the ANN model performed significantly better with a mean absolute percentage error of less than 10%. These models also performed better than earlier regression models used.⁴⁶ The outcome of the study corroborated that of an earlier one by Kumar et al. (2014) comparing model-based approaches and machine learning methods for bus arrival time predictions. In the latter study, the ANN model and Kalman filter algorithm model methods were compared using the same field data. The results showed that the ANN model performed better than the Kalman Filter model when used with a large database. The researchers also

concluded that in the case without any database, the Kalman Filter model would be a better choice.⁴⁷

2.7 Significant Variables for Predicting Bus Travel Times

There have been various studies on predicting bus travel times which have used various models and data to develop prediction models with better accuracies. The significant variables that have been prevalent across the various studies include time of the day, distance between the origin and destination and/or stops along the route, bus dwell time, number of passengers boarding and/or alighting, number of signalized and unsignalized intersections between the origin and destination points, etc.

Wei Fan et al. (2015) conducted a study predicting the travel time of buses using only GPS data. The study compared Historical Average (HA), Kalman Filtering (KF), and Artificial Neural Network (ANN) models developed based on the arrival and departure time data at bus stops collected by the GPS. The basis of the study was to predict bus travel time only with the GPS data in the absence of traffic stream and geometric data such as flow, speed, weather, distance, etc. Their results determined that the ANN model performed better in overall prediction accuracy and robustness and that—though these models gave a better a mean absolute percentage error compared to other existing bus time travel prediction models—the addition of explanatory variables from traffic-stream data in the predictive models played a huge role in the accurate prediction of bus travel times. It was concluded that the time of the day significantly predicts travel time.⁴⁸ Since collecting real time traffic-stream data can be expensive, this approach works well in situations of limited resources (e.g., budget).

Within the context of Intelligent Transport Systems (ITS) and Advanced Public Transportation Systems (APTS), the ideal situation for transit agencies is to collect, process, and disseminate real-time information to the managing agency and riders. Some studies have been conducted using data from Automatic Vehicle Location (AVL) and Automatic Passenger Counting (APC) systems installed on buses to generate real-time data—vehicle arrival/departure, passenger boarding/alighting, en-route traffic conditions, speed, distance, etc.—which in most cases results in better predictive models. Shalaby and Farhan (2002), Patnaik et al. (2004), Chen et al. (2004) and Kumar et al. (2014) are some studies that relied on AVL and/or APC collected data to produce bus travel time prediction models.

Shalaby and Farhan (2002) used data from AVL and APC systems to develop bus travel time prediction models based on the Kalman Filter algorithm.⁴⁹ Patnaik et al. (2004) developed regression models to estimate bus arrival times using data collected from APC systems installed on buses.⁵⁰ Based on APC data, Chen et al. (2004) developed a dynamic model to predict bus arrival times by combining ANN and Kalman Filter algorithms.⁵¹ Kumar et al. (2014) compared the performance of an ANN model and the Kalman Filter algorithmic model to predict bus travel time using APC data.⁵² For both studies, the AVL and/or APC data included variables such as

direction of bus, speed of the vehicle, distance between stops and cumulative distance of trip, dwell time, records for any time the bus stopped or opened its doors, number of passengers alighting and/or boarding, date, day and time of service, etc.

Ranjitkar et al. (2019) conducted a study that introduced ten independent variables (factors) in seven different models to compare their accuracies in predicting bus travel time. The developed models used included multivariate linear regression, ANN, decision tree, and gene expression programming models, among others. The study relied on AVL data and passenger flow data that were combined into one dataset. The study identified significant variables for the models that included inter-stop distance, route completion/distance, dwell time, number of passengers boarding and/or alighting, date and day of the week, time of the day, route number, etc. This is consistent with the initial studies that utilized AVL and/or APC data in their predictive models. The results of the research showed that the ANN time series was the best-performing model with respect to computational effort and accuracy. It was also determined that the inter-stop distance (i.e., the distance between two stops) was the most significant variable or factor across all regression models, while, on the other hand, the dwell time and number of passengers alighting and/or boarding did not appear to have an impact on bus travel time.⁵³

Inter-stop distance and total travel distance, speed, dwell time, number of passengers and time of the day are some of the common variables seen among the discussed bus travel time prediction studies which utilized AVL and/or APC data. It is noted that in most cases inter-stop distance, cumulative travel distance and the bus travel speed were determined to be significant to bus travel time prediction accuracy.

2.8 Bus Transit Times Prediction Models

Real-time bus travel time, though ideal and highly sought after (especially by patrons, since it would enable them to better plan their commute), cannot be easily assessed, and even when made available, it comes with some level of inaccuracy. Over the years, a multiplicity of mathematical models have been developed to predict bus travel times with reasonable accuracy, and the most widely used can be classified into four main categories. These include historical average, regression and Machine Learning models.

Historical Average Models are based on historical bus travel time data that have been collated from previous trips. They assume that the current traffic conditions are the same or do not change, and that based on them, current and future travel times can be extrapolated. Due to very few constraints and the assumption of static traffic conditions, historical average models have simple algorithms and require relatively small computational time. However, unless the traffic conditions remain stable over a period of time with no other unforeseen changes, the performance of these type of models is weak.⁵⁴

Jeong and Rilett (2004) used AVL data to develop models to predict bus arrival time using Houston, Texas as a case study. The study developed and compared the performance of three types

of models: a historical data-based model, regression models and ANN models using the same dataset. The prediction of bus arrival time was based on dwell time at stops and traffic congestion in terms of average delay at intersections as well as bus arrival time delays. The historical data-based model involved using a simple statistical model to calculate the travel time between stops on the bus route by finding the difference between the average arrival time and average departure time. The average departure time was obtained by adding the arrival time and dwell time at a stop. From these deductions, a recursive formula was used to predict the bus arrival times at the remaining stops on the route. The limitation of this model was that it did not consider dwell times at signalized intersections or traffic congestion zones. Five linear multiple regression models were tested using the distance between stops, bus schedule adherence and dwell time as independent variables with bus arrival time at each stop as the dependent variable. The results from the study indicated that, in terms of prediction accuracy, the ANN models outperformed the historical data model and the regression models.⁵⁵

Farhan et al. (2002) used AVL and APC data to develop a bus travel time model which could potentially provide real-time information to passengers and transit management. The study also used historical average, regression analysis, and neural network techniques to develop the models and test their accuracies in predicting bus travel time. The historical average model was developed using the AVL data; the travel time was calculated by finding the difference between the arrival times of two consecutive stops. The average time between stops was computed for both the peak and off-peak periods. The researchers concluded that, though the historical average model could suffice for informing passengers of bus arrival times, it was not appropriate for evaluating and/or implementing control measures, since it did not include independent variables that were likely to affect bus travel time.⁵⁶

Regression models make use of mathematical functions to explain the relationship between a dependent variable and a set of independent variables. Since these models measure the concurrent impact on the dependent variable due to different factors that are not related to each other, better bus time arrival models were produced.⁵⁷ A study conducted by Patnaik et al. (2004) developed a set of multiple linear regression models to predict bus arrival times using number of stops, dwell times, boarding and alighting passengers, and the weather as potential explanatory variables. Results from the study indicated that the model accuracy was higher in predicting bus arrival times at downstream stops. The advantage of the multiple linear regression model is the ability to establish the relationship between the dependent variable and each of the independent variables, thereby indicating which ones are more significant to bus arrival prediction times. Thus, from their study, Patnaik et al. determined that the weather was not as significant as the dwell time at stops or signalized intersection delays.⁵⁸ Another regression model was developed in a study by Ramakrishna et al. (2006) to predict bus arrival times which used GPS data. From the results, the researchers determined that intersection delay and bus stop dwell time as input variables did not impact predicted bus arrival times.⁵⁹

Finally, machine learning models include Artificial Neural Network (ANN) models and can find nonlinear relationships between a dependent variable and its independent variables. These models are suitable for complex data and can eliminate noise to produce an effective prediction model without indicating the specific traffic processes. ANNs are constructed with multiple layers of processing units known as artificial neurons which are interconnected by weights and contain activation functions. The input–output relationship is mapped automatically by using the weights in a learning process. Due to their ability to solve complex non-linear relationships, they are quite popular among bus arrival time prediction studies.⁶⁰ The following sections presents a detailed discussion of ANN and some of its applications in bus travel time prediction.

2.9 Artificial Neural Networks (ANNs) and Their Applications

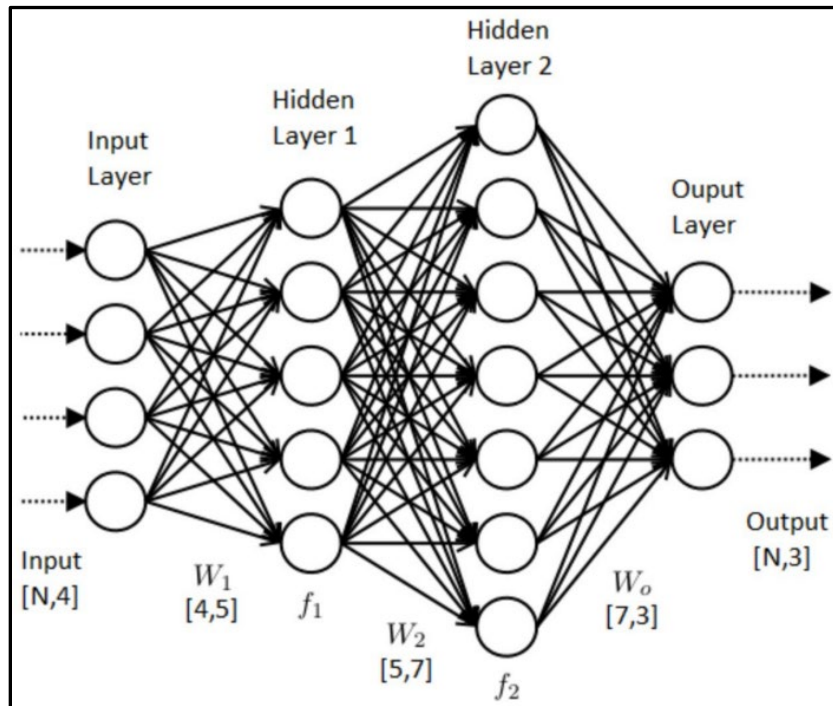
ANNs are mathematical models that are inspired by and work similarly to biological neural networks in the human brain. ANNs are used in engineering to perform complex tasks such as pattern recognition, forecasting, data compression, and classification. ANN can approximate both linear and nonlinear functions to a good degree of accuracy. It does this through a learning algorithm where it can be trained to learn from examples and identify underlying functional relationships among datasets and build “piece-wise” approximations of the functions.⁶¹

Classification or forecasting using ANN involves training and learning procedures. Large amounts of historical data (a set of input data with known outputs) are presented to the network to improve the training and gain maximum benefit from the network. The learning process involves the construction of a network of inputs and outputs with weights assigned to each mapping; the weights are adjusted at each iteration. The learning rule used determines how the weights and bias levels of a network are updated. Thus, the learning rule enables a neural network to learn from the existing conditions and improve its performance. There are several learning rules used in training neural networks. Notable among the rules are the hebbian, perceptron (error-correction), delta, correlation, and outstar learning rules.

Multilayer Perceptron (MLP) is the most commonly known ANN rule. Perceptron layers serve as algorithms that enable the supervised learning through which classification can be achieved. MLP basically consists of three layers: input layer, hidden layer, and output layer. It operates on a feed forward network in which information flows from the input layer through the hidden layer to the output layer to produce results. These layers have nodes (neurons) which are interconnected and assigned weights based on the input. The weighted sum is computed with the help of the incoming connections. The outputs for specific inputs are obtained by adjusting the weights to minimize the errors between the produced output and the desired output by error-back propagation.⁶² In error-back propagation, the error in the output is put back into the network and passed from one layer to the other by adjusting the weights of connections. This is the most frequently used technique in ANN transportation and other engineering applications.⁶³

Figure 2 illustrates the interconnections between the various layers in MLP ANN.

Figure 2: Illustration of ANN⁶⁴



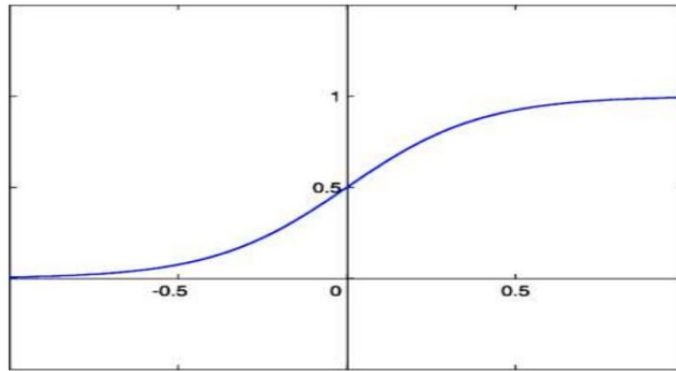
In ANN, Activation functions are used to create nonlinearity into the network. They are also known as transfer functions and operate by calculating the weighted sum of their inputs and adding a bias, then deciding whether a neuron should be activated or not. According to a study by Amita et al. to predict bus travel time using Artificial Neural Network, the back-propagation neural network training algorithm requires four processes: Weight initialization, Forward feed, Back propagation of errors and updating the weights and biases. In back-propagation neural network, the learning transfer function needs to be distinguishable and hence the activation functions are used to provide nonlinearity (remove any direct relationships) within the network.⁶⁵

The three most common types of activation functions used in an ANN are the sigmoid, hyperbolic tangent, and rectified linear unit functions.⁶⁶ The sigmoid function is a logistic function and is expressed as

$$y_i = \frac{1}{(1 + \exp(-v_j))}$$

where v_j is the weighted sum of all synaptic inputs plus the bias of the j^{th} neuron and y_j is the output of the neuron. Figure 3 is a graphical presentation of the sigmoid function.

Figure 3: Sigmoid Activation Function Curve⁶⁷

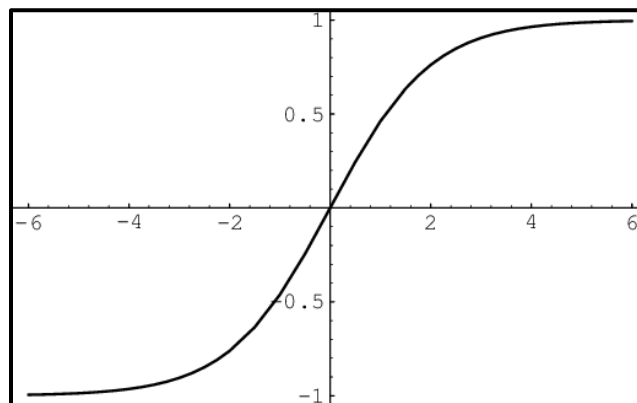


The hyperbolic tangent function is similar to the sigmoid function but ranges from -1 to 1. The function is expressed as

$$y_i = \frac{\exp(v_j) - \exp(-v_j)}{\exp(v_j) + \exp(-v_j)}$$

Figure 4 is a graphical presentation of the hyperbolic tangent function.

Figure 4: Hyperbolic Tangent Function Curve⁶⁸



This research utilized a hyperbolic tangent function, as this type of activation function is effective in centering the data by bringing the mean close to 0 and makes learning for consecutive layers much easier.⁶⁹

Because of ANN's ability of learning patterns and predicting output from previous data, it has been incorporated in many studies to predict bus arrival times. Jeong and Rilett (2004) developed an ANN model and two non-ANN models (historical databased model and regression models) and compared their performance using AVL data to predict bus arrival times in Houston, Texas. The prediction of bus arrival time was based on dwell time at stops and traffic congestion. The ANN architecture used in this research was based on the MLP rule, with input, hidden, and output layers. The ANN model had the lowest Mean Absolute Percentage Error (MAPE) when

compared to the historical data-based and the regression models. On the average, the ANN models resulted in a 54.24% and 48.61% improvement of the MAPE in the downtown and northern areas, respectively. The results from the study indicated that in terms of prediction accuracy, the ANN models outperformed the historical data model and regression models.⁷⁰

Chien et al. (2002) conducted a study in which they developed models to accurately predict bus arrival times in an urban road network using ANN. In their study, two ANN predictive models were developed and trained by link-based and stop-based data, respectively. Afterwards, the ANN models were integrated with an adaptive algorithm to enable them to dynamically adjust prediction errors. They were then tested via simulations developed in the CORSIM software program and then the results were compared. From the results, the enhanced ANN models outperformed the ANN models without the adaptive features, yielding consistently lower RMSEs under different scenarios. The enhanced stop-based ANN was noted to perform better at multiple intersections between stops, while the enhanced link-based ANN performed better for stops with fewer intersections between them. Chien et al. suggested the development of a hybrid ANN model which would integrate both the link-based and stop-based models to further improve the accuracy of bus arrival time predictions.⁷¹

2.10 Summary of Literature Review

The public transport system has gone through several improvement phases over the years with the goal of serving the greater interest of the public while remaining profitable. As part of various transit agencies' efforts to improve transit services, AVL and APC systems have been introduced as part of the developments in Intelligent Transport Systems. This allows for the collection of real-time data at bus stops and on buses that can be relayed to both riders and transit management. From the literature, several bus arrival prediction models have been developed using various techniques that can be broadly categorized into four main types: historical average models, regression models, Kalman Filter models, and ANN models. The significant variables for predicting bus travel time include distance, speed of bus, dwell time, number of passengers boarding and/or alighting, and number of stops, among others. Despite the studies and models developed, none has been developed specifically to predict bus travel time using ANN and considering the unique nature of traffic patterns in an urban area like DC. As a result, this research aims to develop a model to predict bus travel time in DC using ANN.

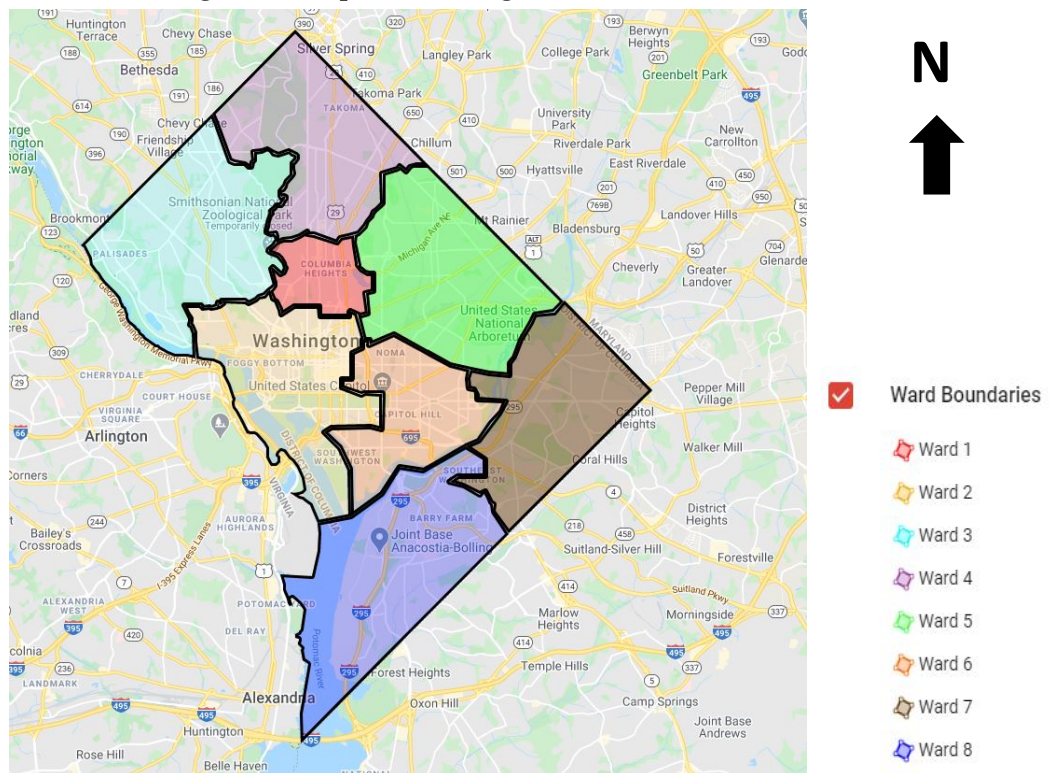
III. Research Methodology

3.1 Description of the Study Area

This research is based on data obtained describing WMATA buses that operate in Washington, DC. The city is divided into four (unequal) quadrants: Northwest (NW), Northeast (NE), Southwest (SW), and Southeast (SE); these are further divided into eight wards overlapping the quadrant boundaries. As of July 2018, the population of Washington, DC was approximately 702,455 with an annual growth rate of approximately 1.4%.⁷² The city is highly urbanized and is ranked as the sixth most congested city in the United States, with each driver spending an average of 63 hours per year in traffic.⁷³

The Washington Metropolitan Area Transit Authority (WMATA) is the agency that oversees the operations of Metrobus service in the jurisdiction. WMATA has a bus fleet of approximately 1,600 buses that operate on 325 routes in Washington, DC, in portions of Maryland, and in Northern Virginia, covering a total land area of about 1,500 square miles. Metrobuses operate 24 hours a day, seven days a week, and they make more than 400,000 trips each weekday. Of the total number of bus stops, 2,556 (22.2%) have shelters, while the remainder do not. Figure 5 shows a map of Washington, DC divided into wards.

Figure 5: Map of Washington, DC⁷⁴



Created with DC Ward boundary data overlay in Google Maps

3.2 Data Collection

The following steps were followed to collect the data required for the analysis.

Selection of Bus Routes

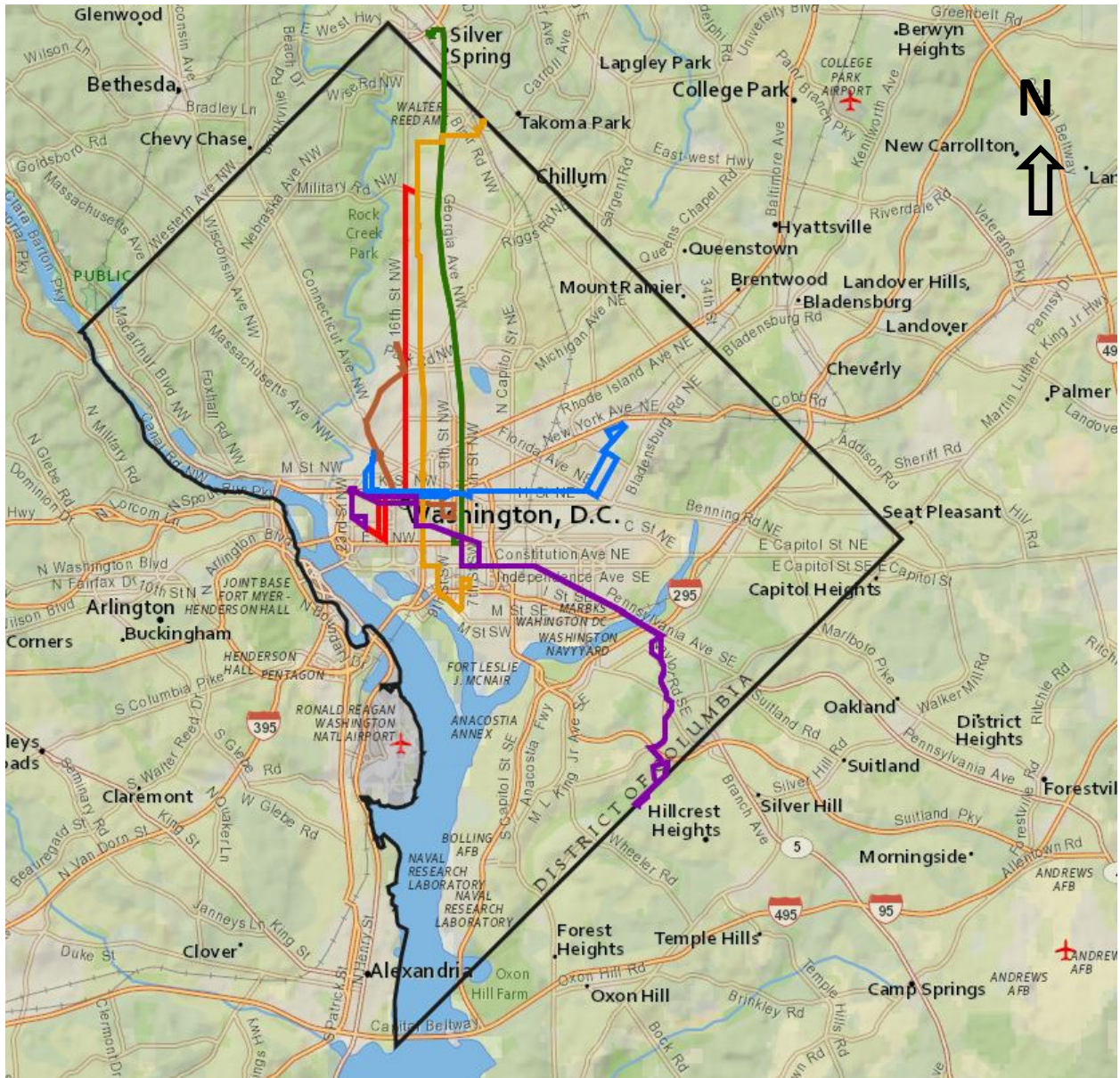
Six months (January–June 2019) of AVL and APC data for six bus routes operating in DC was obtained from the WMATA. The bus routes operating along two functional roadway types were considered: arterials and collectors. In general, bus routes with certain characteristics were considered to avoid erratic bus schedule and patronage patterns. For instance, officials from WMATA were consulted for the provision of bus route data that had relatively higher patronage. Bus stops on routes with longer headways were also considered for analysis since routes can have several bus stops that amass larger groups of patrons boarding or alighting buses. Finally, bus stops in a proximity to metro rail stations grant rail patrons access to bus services and vice versa were also considered. Such bus stops usually have a high number of bus patrons either boarding or alighting the buses.

Based on the criteria, the following bus routes were selected for the study:

1. Route 70 (Georgia Avenue Line)
2. Route 32 (Pennsylvania Avenue Line)
3. Route 52 (14th Street Line)
4. Route 42 (Mount Pleasant Line)
5. Route D4 (Ivy City–Franklin Square Line)
6. Route S1 (16th Street–Potomac Park Line)

Figure 6 shows the operation paths of the selected routes.

Figure 6: Map of Operation Paths of Selected Routes



Route Characteristics

The bus routes selected for this study have the following characteristics. Classifications of the primary operational roadway of transit buses are listed according to the District Department of Transportation's (DDOT) 2016 Street Function Classification System.

Buses on Route 70 operate primarily on Georgia Avenue, NW with the main direction of travel being Northbound/Southbound. Georgia Avenue, NW is classified as a Principal. Route 70 buses run a round trip from 9th Street and Constitution Avenue, NW to Silver Spring Station in Maryland. The length of the route is approximately 7.5 miles, and the buses serve 51 bus stops in the northbound (NB) direction of travel and 55 bus stops in the southbound (SB) travel direction. There are 108 intersections between the first and the last bus stops on this route. Route 70 operates

24 hours a day, seven days a week with buses arriving at a stop every 12 minutes between 6:00 AM–1:00 PM during weekdays (Monday–Friday) and every 15 minutes between 6:00 AM–11:00 PM during weekends (Saturday and Sunday).⁷⁵

Buses on Route 32 primarily operate on Pennsylvania Avenue, SE, Naylor Road, SE, Alabama Avenue, SE, and Southern Avenue, SE with the main of travel being Northeast-bound (NEB) and Southwest-bound (SWB) directions (the routes are round trips). Pennsylvania Avenue, NW is classified as a Principal Arterial, while Naylor Road, Alabama Avenue, and Southern Avenue, SE are classified as Minor Arterials. Route 32 transit buses run from Virginia Avenue and E Street, NW to Southern Avenue Bus Station in the Southeast DC round trip. The length of the route is approximately 8.75 miles, and the buses serve 47 bus stops along both the NEB and SWB directions of travel. There are 106 intersections between the first and the last bus stops. Route 32 operates from 4:00 AM–1:00 AM during weekdays (Monday–Friday) and from 5:00 AM–1:00 AM during weekends (Saturday and Sunday).⁷⁶

Transit buses on Route 52 operate primarily on 14th Street, NW with the main direction of travel being Northbound/Southbound. Fourteenth Street, NW is classified as a Principal Arterial from 14th Street and Independence Avenue, NW to 14th Street and U Street, NW and as a Minor Arterial from 14th Street and U Street, NW to Aspen Street, NW. Route 52 buses run from Tacoma Station Bus Stop to D Street and 7th Street, NW and back. The length of the route is approximately 7.95 miles, and the buses serve 63 bus stops along the NB direction of travel and 60 bus stops along the SB travel direction. There are 121 intersections between the first and the last bus stops. Route 52 operates from 4:00 AM–1:00 AM during weekdays (Monday–Friday) and from 5:00 AM–1:00 AM during weekends (Saturday and Sunday).⁷⁷

Route 42 transit buses operate primarily on Connecticut Avenue, NW and Columbia Road, NW with the main direction of travel being Northbound/Southbound. Connecticut Avenue, NW is classified as a Principal Arterial, while Columbia Road, NW is classified as a Minor Arterial. Route 42 buses run from 9th Street and F Street, NW to Lamont Street and Mt. Pleasant Street, NW (round trip). The length of the route is approximately 3.5 miles, and the buses serve 25 bus stops along the Northbound and Southbound directions of travel. There are 50 intersections between the first and last bus stops. Route 42 operates 24 hours a day, seven days a week.⁷⁸

Transit buses on Route D4 operate primarily on K Street, NW and NE with the main direction of travel being Eastbound/Westbound. K Street, NW is classified as a Principal Arterial, while K Street, NE is classified as a Minor Arterial according to the District Department of Transportation's (DDOT) 2016 Street Functional Classification System. Route D4 buses run a round trip from Okie Street and 16th Street, NE to Massachusetts Avenue and 20th Street, NW and back. The length of the route is approximately 4.7 miles; the buses serve 40 bus stops along the Eastbound direction of travel and 37 stops along the Westbound travel direction. There are 82 intersections between the first and the last bus stops. Route 42 operates from 4:00 AM–1:00 AM, seven days a week.⁷⁹

Route S1 transit buses operate primarily on 16th Street, NW with the main direction of travel being Northbound/Southbound. According to the District Department of Transportation's (DDOT) 2016 Street Functional Classification System, 16th Street, NW is classified as a Principal Arterial. Route S1 buses run from Virginia Avenue and E Street, NW to Colorado Avenue and 16th Street, NW and back. The length of the route is approximately 4.75 miles, and the buses service 37 bus stops along the NB direction of travel and 47 stops along the Southbound travel direction. There are 91 intersections on the route within the study limits. Buses traveling SB on Route S1 operate from 5:00 AM–10:00 AM, seven days a week, while buses traveling NB on Route S1 operate from 3:00 PM–8:00 PM, also on a daily basis.⁸⁰

Data Extraction

AVL and APC data were obtained from WMATA officials for the selected bus routes: bus data describing a six-month duration from January 2019 through June 2019 were obtained for the study for each of the six routes. Excel sheets containing the data from the first week of every month were filtered to obtain only the necessary information required for the analysis. A sample of the Excel data for a bus route obtained from the WMATA database is shown in Figure 7.

Figure 7: Sample Datasheet Obtained from WMATA (Route 32)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	
	BUSTOOL S_VERSION	ROUTE ID	STOP_ SEQUENCE	Stop Desc	BUS ID	EVENT TIME	ODOMETER DISTANCE	ROUTE STATUS	DWELL TIME	DELTA TIME	STOP SPEED	STOP_F DOOR_ ENTRY	STOP_ BACK_ DOOR_ ENTRY	STOP_F DOOR_ EXIT	STOP_ BACK_ DOOR_ EXIT	LATITUD E	LONGITU DE	HEAD ING	NAV_ STATE	EVENT _TYPE	Event_De scription	TRIP_ID	TRIP_ TYPE	schedule d_time	TRIP_ST ART_TI ME	STOP ID	TA_GEO ID	
1																												
2	S1000056	3201	0	NULL	2114	48:03.0	54100	0	0	603	0	0	0	0	0	38.8296	-76.9905	49	1	9	NULL	9.41E+08	2	NULL	48:00.0	NULL	BT-SH	
3	S1000056	3201	-1	NULL	2114	48:03.0	54100	4	0	2339	0	0	0	0	0	38.8296	-76.9905	49	0	6	NULL	9.41E+08	2	NULL	48:00.0	NULL		
4	S1000056	3201	-1	NULL	2114	48:03.0	54100	4	0	603	0	0	0	0	0	38.8296	-76.9905	49	0	720	NULL	9.41E+08	2	NULL	48:00.0	NULL		
5	S1000056	3201	0	NULL	2114	00:01.0	54205	0	0	1321	0	0	0	0	0	38.8297	-76.9904	46	1	104	NULL	9.41E+08	2	NULL	48:00.0	NULL	BT-SH	
6	S1000056	3201	0	NULL	2114	00:25.0	54205	0	0	1345	0	0	0	0	0	38.8297	-76.9904	46	1	104	NULL	9.41E+08	2	NULL	48:00.0	NULL	BT-SH	
7	S1000056	3201	0	NULL	2114	00:51.0	54205	0	0	1371	0	0	0	0	0	38.8297	-76.9904	46	1	104	NULL	9.41E+08	2	NULL	48:00.0	NULL	BT-SH	
8	S1000056	3201	0	NULL	2114	01:16.0	54205	0	0	1396	0	0	0	0	0	38.8297	-76.9904	46	1	104	NULL	9.41E+08	2	NULL	48:00.0	NULL	BT-SH	
9	S1000056	3201	0	NULL	2114	01:40.0	54205	0	0	1420	0	0	0	0	0	38.8297	-76.9904	46	1	104	NULL	9.41E+08	2	NULL	48:00.0	NULL	BT-SH	
10	S1000056	3201	0	NULL	2114	06:01.0	54205	0	229	1910	0	0	0	0	0	38.8297	-76.9904	46	1	5	NULL	9.41E+08	2	09:50.0	48:00.0	NULL	BT-SH	
11	S1000056	3201	0	NULL	2114	06:02.0	54205	0	0	1681	0	0	0	0	0	38.8297	-76.9904	46	1	101	NULL	9.41E+08	2	NULL	48:00.0	NULL	BT-SH	
12	S1000056	3201	0	NULL	2114	15:00.0	54205	0	0	2219	0	0	0	0	0	38.8297	-76.9904	46	1	104	NULL	9.41E+08	2	NULL	48:00.0	NULL	BT-SH	
13	S1000056	3201	0	NULL	2114	30:00.0	54205	0	0	3120	0	0	0	0	0	38.8297	-76.9904	46	1	104	NULL	9.41E+08	2	NULL	48:00.0	NULL	BT-SH	
14	S1000056	3201	0	NULL	2114	45:01.0	54205	0	0	4021	0	0	0	0	0	38.8297	-76.9904	46	1	104	NULL	9.41E+08	2	NULL	48:00.0	NULL	BT-SH	
15	S1000056	3201	0	NULL	2114	00:01.0	54205	0	0	4921	0	0	0	0	0	38.8297	-76.9904	46	1	104	NULL	9.41E+08	2	NULL	48:00.0	NULL	BT-SH	
16	S1000056	3201	0	NULL	2114	02:00.0	54205	0	0	5039	0	0	0	0	0	38.8297	-76.9904	46	0	15	NULL	9.41E+08	2	NULL	02:00.0	NULL	1003537	
17	S1000056	3201	0	NULL	2114	02:00.0	54205	0	0	5039	0	0	0	0	0	38.8297	-76.9904	46	0	720	NULL	9.41E+08	2	NULL	02:00.0	NULL	1003537	
18	S1000056	3201	21	NULL	2114	02:00.0	54205	0	0	5039	0	0	0	0	0	38.8297	-76.9904	46	1	16	NULL	9.41E+08	2	02:00.0	48:00.0	NULL	1000593	
19	S1000056	3201	35	NULL	2114	02:00.0	54205	0	0	5039	0	0	0	0	0	38.8297	-76.9904	46	1	16	NULL	9.41E+08	2	02:00.0	48:00.0	NULL	1000948	
20	S1000056	3201	2	NULL	2114	02:00.0	54205	0	0	5039	0	0	0	0	0	38.8297	-76.9904	46	1	16	NULL	9.41E+08	2	02:00.0	48:00.0	NULL	3002667	
21	S1000056	3201	20	NULL	2114	02:00.0	54205	0	0	5039	0	0	0	0	0	38.8297	-76.9904	46	1	16	NULL	9.41E+08	2	02:00.0	48:00.0	NULL	1000576	
22	S1000056	3201	13	NULL	2114	02:00.0	54205	0	0	5039	0	0	0	0	0	38.8297	-76.9904	46	1	16	NULL	9.41E+08	2	02:00.0	48:00.0	NULL	1003927	
23	S1000056	3201	42	NULL	2114	02:00.0	54205	0	0	5039	0	0	0	0	0	38.8297	-76.9904	46	1	16	NULL	9.41E+08	2	02:00.0	48:00.0	NULL	1001179	
24	S1000056	3201	25	NULL	2114	02:00.0	54205	0	0	5039	0	0	0	0	0	38.8297	-76.9904	46	1	16	NULL	9.41E+08	2	02:00.0	48:00.0	NULL	1000702	
25	S1000056	3201	48	NULL	2114	02:00.0	54205	0	0	5039	0	0	0	0	0	38.8297	-76.9904	46	1	16	NULL	9.41E+08	2	02:00.0	48:00.0	NULL	1003537	
26	S1000056	3201	0	NULL	2114	02:02.0	54205	3	0	5039	0	0	0	0	0	38.8297	-76.9904	46	0	11	NULL	9.41E+08	2	NULL	02:00.0	NULL	1003537	
27	S1000056	3201	0	NULL	2114	22:35.0	149563	0	0	95	101	0	0	0	0	38.8685	-76.9709	162	1	9	NULL	9.41E+08	2	NULL	22:33.0	NULL	1000437	
28	S1000056	3201	-1	NULL	2114	22:35.0	149563	4	0	1508	101	0	0	0	0	38.8685	-76.9709	162	0	6	NULL	9.41E+08	2	NULL	22:33.0	NULL		

From Figure 7, it can be seen that the APC and AVL systems (for Route 32 buses) recorded information including odometer reading, geolocation, and dwell time, amongst others. The datasheets were filtered to display only the information required for the neural and regression analyses. Based on the significance of their impact on the travel time of transit buses (from the previous literature), the following independent variables for a bus trip were extracted for each week and for the selected routes from the data:

1. Departure and Arrival Times (Event Time)
2. Length of Routes

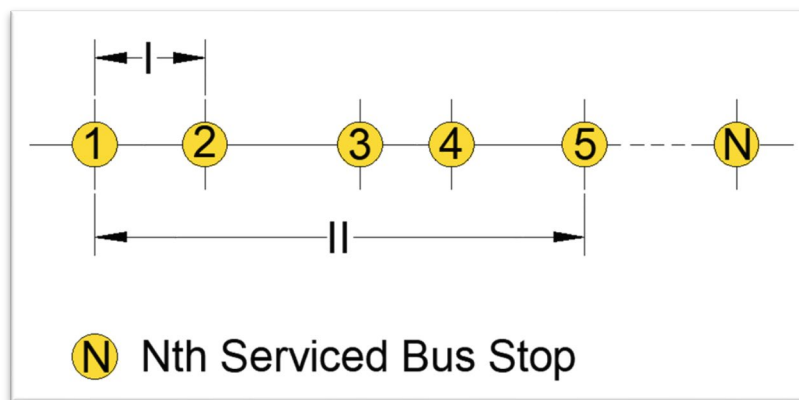
3. Passengers Loading/Unloading
4. Dwell Time
5. Travel Time
6. Location of the Bus

Field data observations were also conducted to obtain the number of intersections on the selected bus routes and the associated roadway speed limits. The information obtained was used in the Neural Network training. The descriptions of the variables included in the ANN Model are as follows.

1. *Number of Served Bus Stops (X_1) Between Origin and Destination*

A bus can “serve” a stop if there are passengers entering or exiting at that particular stop. Hence, for the purpose of this research, X_1 denotes the number of bus stops between any two “origin” and “destination” points along a bus route. Figure 8 represents a sample bus route. The yellow circles represent bus stops along the route. From Figure 8, if a bus serves Stop 1 and Stop 2 (situation I), $X_1 = 0$. Similarly, X_1 for a data point used in the analysis is 3 (Stops 2, 3, and 4) if the bus starts serving from Stop 1 and reaches Stop 5 (situation II), serving all bus stops in between.

Figure 8: Sample Representation of a Bus Service Route



2. *Length of Route between Bus Stops (X_2)*

The data provided by WMATA included the odometer readings of all the buses along a route. Hence, X_2 was obtained by taking the difference of the odometer readings between any two served bus stops. For example, from Figure 8, X_2 for situation II can be represented as:

$$X_2 = Od5 - Od1$$

where

Od5 = Odometer reading of a WMATA bus at stop 5 (in feet)

Od1 = Odometer reading of a WMATA bus at stop 1 (in feet).

3. *Average Number of Passengers in the Bus (X_3)*

Automatic Passenger Counting (APC) systems installed on all the WMATA buses count the number of passengers getting in and out of the bus from the front and rear doors. At any point, the number of passengers will be the sum of passengers already on the bus (excluding the origin bus stop) and the number of patrons boarding the bus minus the number of patrons alighting the bus at any stop. The average number (X_3) can be found by taking the mean of the number of passengers on the bus between any two particular bus stops served. Average number of onboard passengers was computed as:

$$\text{Average number of Passengers} = \frac{\Sigma P}{N}$$

where

N = number of served bus stops between the origin and the destination (inclusive) along a route

P = Number of passengers onboard at the n^{th} bus stop.

4. *Average Dwell Time at Served Bus Stops (X_4)*

The dwell time is the period during which the front and the rear doors of a bus at a bus stop remained open to serve patrons. Average dwell time (X_4) for the analysis is the mean of dwell times of the served bus stops between and origin and destination point. Thus, dwell time was computed as

$$\text{Average Dwell Time} = \frac{\Sigma DT}{N}$$

where

N = number of served bus stops between the origin and the destination (inclusive) along a route

DT = dwell time of the bus at the n^{th} bus stop.

5. *Number of Intersections between Bus Stops (X_5)*

Field data observations were conducted to determine the number of intersections (unsignalized and signalized) between any two served bus stops along a route (X_5).

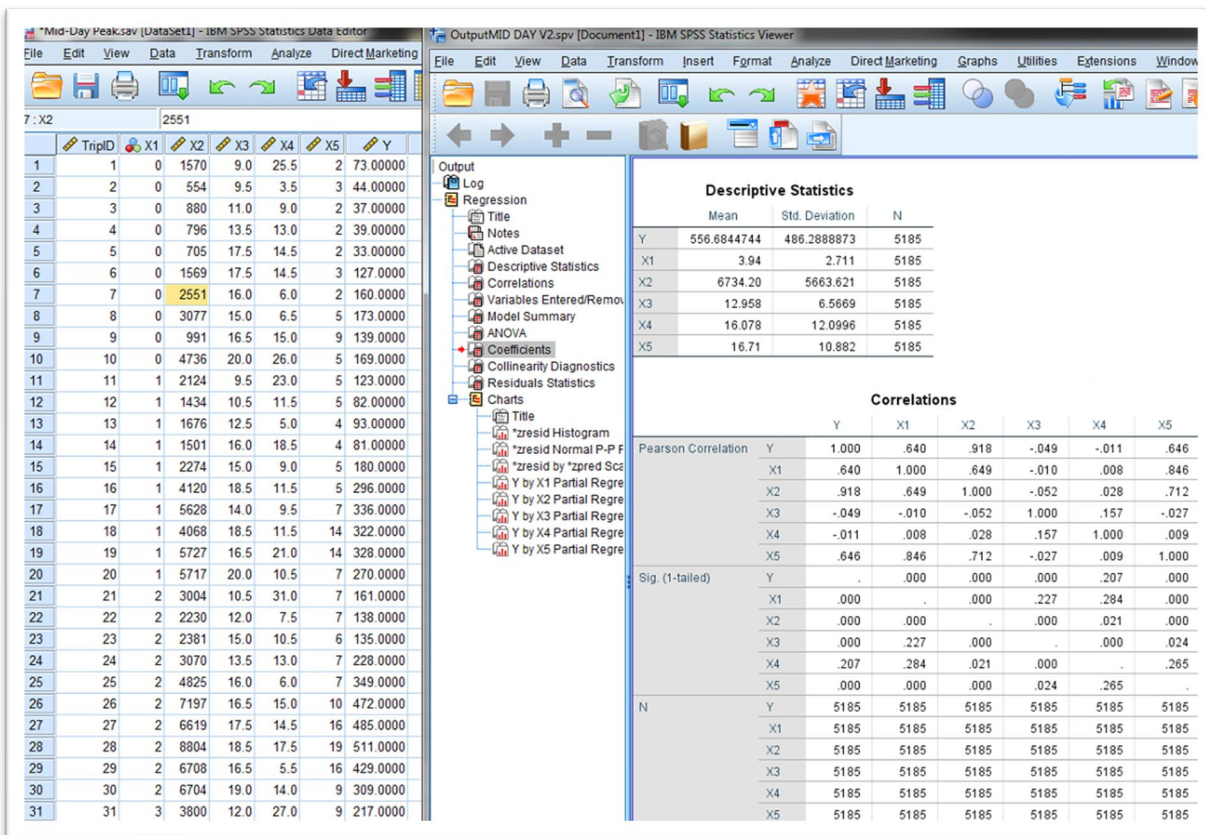
Due to the variability of traffic characteristics during the day, separate ANN models were developed for AM Peak (7:00 AM–9:30 AM), PM Peak (4:00 PM–6:30 PM) and Mid-Day Peak (10:00 AM–2:30 PM) periods. Thus, the extracted data were partitioned into these peak periods. Since a sample matrix contained data for a bus traveling in both directions of a route, analysis by bus direction was excluded.

3.3 Data Analysis

Two software programs were used for the data analysis. IBM's *SPSS Statistics 25* software (SPSS) was used for the regression analysis of the peak models, while *Neural Designer* was used for the neural network analysis.

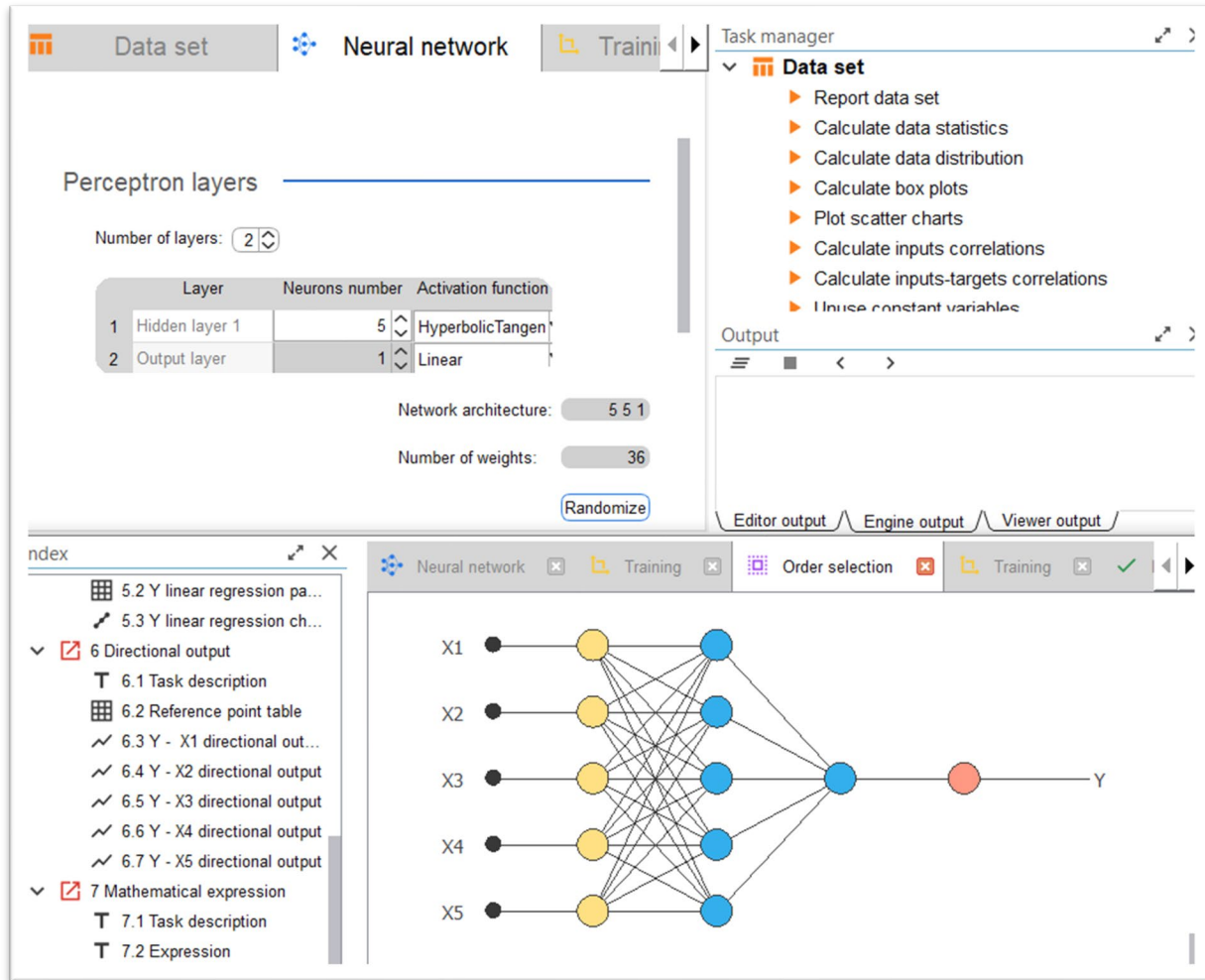
SPSS is a powerful statistical software that allow users to better understand the data under consideration based on the extraction of statistical insights. Figure 9 shows a sample analysis of a model in SPSS software.

Figure 9: Snapshot of SPSS Interface and Sample Output



Neural Designer is a software that incorporates data science and machine learning techniques and which helps to build, train, and deploy neural network models. The software provides access to advanced techniques for data preparation, machine learning, and model deployment. The high operability of *Neural Designer* allows it to be integrated in numerous projects from different sectors for approximation and classification problems. A snapshot of the software interface along with the sample output is presented in Figure 10.

Figure 10: Snapshot of Neural Designer Interface and Sample Output



Determination of Sample Size

ANNs are highly data-dependent, and the amount of data required depends on the complexity of the model to be developed in addition to the number of input variables. Even if no specific minimum amount of data is required, it should be noted that the model will predict with better accuracy on unseen data if it is trained with large amounts of data. For this study, a minimum sample size of 500 origin-to-destination trips of multiple transit buses on a route were extracted for each peak period from the six-month AVL/APC data obtained from WMATA. Thus, a minimum sample of 1,500 origin-to-destination trips was extracted and exported into a Comma Separated Values (CSV) file for further analysis for the three peak periods for each bus route.

Descriptive Statistics

Descriptive statistics including the mean, median and standard deviation were computed for the bus travel times as well as the other predictor or independent variables. The averages of predictors such as dwell times and number of passengers per peak periods were also obtained.

Regression Analysis

Multiple regression analyses were conducted to examine the relationships between the travel times of buses and variables such as number of served bus stops, length of routes, average number of passengers, average dwell time, and number of intersections between bus stops. The regression models were developed for AM, Mid-Day, and PM Peak periods, indexed as $i = 1, 2,$ and $3,$ respectively. The multiple regression models for travel times can be represented as:

$$TT = \beta_{0i} + \beta_{1i}SBS + \beta_{2i}L + \beta_{3i}P + \beta_{4i}DT + \beta_{5i}IT + \varepsilon_i$$

where

TT = Travel Time

SBS = Number of Served Bus Stops

L = Average Number of Passengers

P = Average Dwell Time

DT = Length of the Route between Served Bus Stops

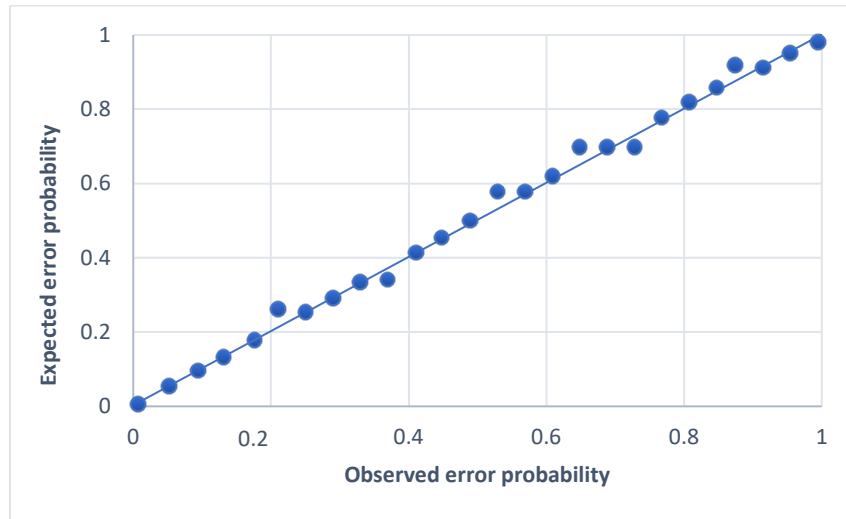
IT = Number of Intersections between Served Bus Stops.

The independent variables $SBS, L, P, DT,$ and IT affect the dependent variable TT . The term β_{0i} is the intercept while β_{ki} are the regression coefficients for the predictors. Further, ε_i is the error residual (distributed error).

The variables used for the neural network were tested to ensure that they met the assumptions of multiple linear regression including normality of errors, homoscedasticity and with no multicollinearity.

When considering normality of errors, the errors (residual terms) for a multiple regression model should be approximately normally distributed. Hence, the relationship between the explanatory variables and the response variables should be explained by the residuals. A histogram representing the plots of observed cumulative probabilities and expected cumulative probabilities can be used to test the normal probability plot. In the case of normally distributed errors, the plotted points will form a straight diagonal line (line of best fit) as shown in Figure 11.

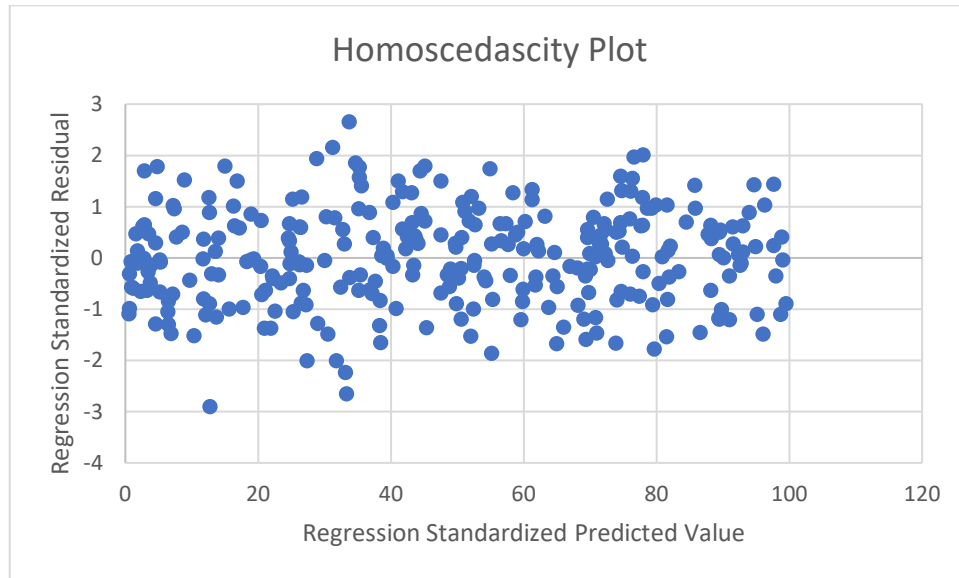
Figure 11: Example Normal Probability Plot



Multicollinearity occurs when two or more independent variables in a multiple regression model are highly correlated with each other. The presence of multicollinearity affects the prediction, as it becomes harder to infer which independent variable affects the outcome the most (variance explained in the dependent variable). Multicollinearity can be tested by inspecting the Variance Inflation Factor (VIF) and the tolerance level using the multicollinearity diagnosis (VIF is the ratio of the variance of the model with multiple variables to the variance of the model with one variable). A high VIF (value greater than 10) and a low tolerance level (value less than 0.1) indicate that two or more independent variables are multicollinear. Another indication of multicollinearity is the correlation between independent variables. Correlation values greater than 0.5 (positive or negative) indicate the presence of multicollinearity.

The multiple regression model also needs to exhibit homoscedasticity, which exists when the variances along the line of best fit remain similar at any point along the line. Violations of this assumption (heteroscedasticity) contribute to the inaccuracy of the significance tests of the regression coefficients and estimations of confidence intervals. Homoscedasticity can be tested by plotting the regression standardized predicted value against the regression standardized residual. The assumption is met if there is an even distribution about the zero line, as shown in Figure 12.

Figure 12: Example Scatter Plot Showing Homoscedasticity



Model Evaluation for Regression Analysis

The multiple regression models were evaluated using the p -values of the F test, R^2 , and adjusted R^2 values. These evaluative parameters are typically used to assess the performance of the models.

The F -test evaluates the null hypothesis that, for the population from which a sample was drawn, all regression coefficients are equal to zero, against the alternative hypothesis that at least one regression coefficient is not. Thus, the F -test determines whether the proposed relationship between the travel time and the set of predictors is statistically significant. The F -test is computed by taking the ratio:

$$F \text{ statistic} = \frac{MSM}{MSE}$$

where MSM is the mean of squares for model and MSE is the mean of squares for the error. The statistical significance of the F -statistic is then determined using the p -value. The significance level for this study was set at 5%.

Another method of evaluation for a regression model is the coefficient of determination, R^2 , which is a measure of the goodness-of-fit of a model. It is defined as the percentage of the variance of the dependent variable that can be explained by the model. R^2 is expressed mathematically as:

$$R^2 = \frac{SST - SSE}{SST}$$

where

SST = Sum of Squares Total (sum of the squares of the difference of the dependent variable and its mean)

SSE = Sum of Squares of Error (sum of the squares of the difference of the predicted dependent variable from actual values of the data).

Generally, R^2 increases as predictors are added to the model. However, this increase does not always result in the actual improvement of the model, as this could also be an indication of overfitting of the model. An adjusted R^2 is also used to assess the model to prevent overfitting.

Like R^2 , $R^2_{adjusted}$ is a measure of the percentage of total variance in the dependent variable that is explained by the model. Unlike R^2 , $R^2_{adjusted}$ takes into account the model's degrees of freedom, paying a penalty when too many predictor variables are added; $R^2_{adjusted}$ will decrease as independent variables are added, if the increase in model fit is not enough to make up for the loss of degrees of freedom. It is expressed as:

$$R^2_{adjusted} = 1 - \frac{MSE}{MST}$$

where MST = Mean of Squares Total

MSE = Mean of Squares for Error.

ANN Model Development

The purpose of developing an ANN model in this research is to determine the travel time of the bus on a route using the Approximation technique in the Neural Designer software. In the approximation technique, the neural network learns from the input–target examples provided by the user. It can be regarded as the process of finding the best fit from the dataset and representing it as a function. It should be noted that the objective of approximation is to produce a neural network which performs well in generalization and makes good predictions for unseen data (good fit) rather than capturing specific details in the dataset (overfitting).

For the neural network analysis, the dependent variable for the model is the travel time (Y). The independent variables that affect the travel time were filtered from the WMATA data to obtain a matrix for each peak period. The general form of the matrix containing the independent variables and the dependent variable (travel time) is represented in Table 3.

Table 3: Sample Peak Period ANN Data Matrix Model for a Bus Route

Trip ID	No. of SERVED Bus Stops, X_1	Length of Route, X_2	Average Number of Passengers, X_3	Average Dwell Time, X_4	No. of Intersections, X_5	Travel Time, Y
1	A	D	G	J	M	P
2	B	E	H	K	N	Q
3..	C	F	I	L	O	R
...	-	-	-	-	-	-
500	S	T	U	V	W	X

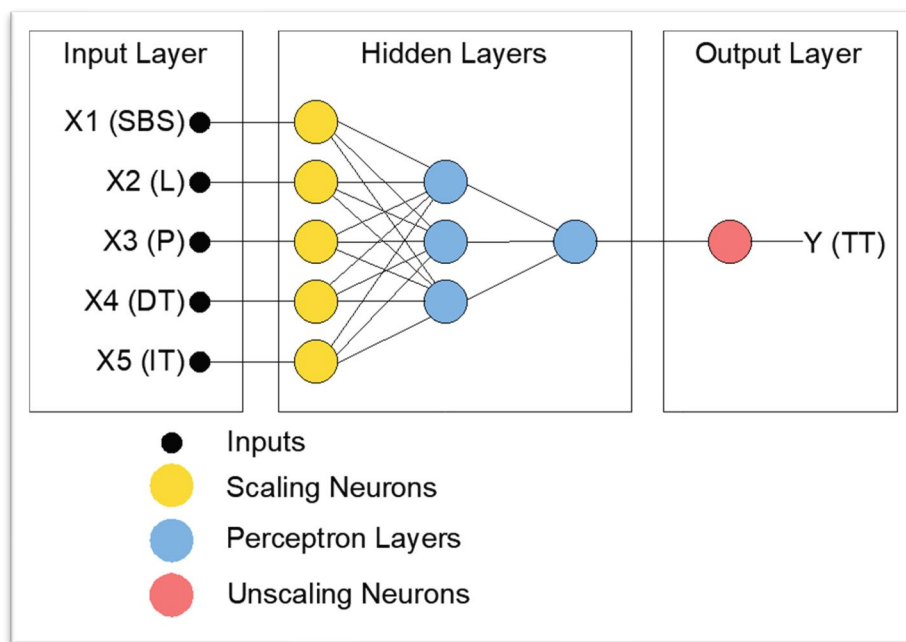
The data matrix was exported as a CSV file and analyzed in the Neural Designer software, which is a data analytic tool that incorporates neural networks to recognize patterns and make predictions from the data. The software was used to split the data into a training set (75%) and a testing set (25%). The training dataset was used to train and develop the model, while the testing dataset was used to validate the model.

Training was conducted through an iterative process of feed forward and error-back propagations until the gradient normalization goal or the stopping criterion of 1,000 epochs (iterations) was met. The parameters that were adjusted in the software to analyze the data set are as follows.

Furthermore, the models were trained using Multilayer Perceptron (MLP). Perceptron layers (also known as dense layers) are important layers that enable the neural network to learn. Numerical values are inputs (X_1, \dots, X_n) for the perceptron neurons in a network to produce a numerical output y . The output is also affected by the combination of bias (b) and the sum of individual weights of independent variables (w_1, \dots, w_n).

The MLP used for this research consisted of three layers: input layer, hidden layer, and output layer. A typical ANN architecture is presented in Figure 13.

Figure 13: ANN Architecture



Data Standardization was also required since the inputs in the data sets did not have the same ranges.

For standardization, a scaling layer was applied to make the values of all the independent variable comparable. For the purpose of this research, an Automatic Scaling Layer that determined the optimal data distribution was applied to obtain the data sets for all bus routes. The scaled outputs were unscaled back to the original units using the unscaling layer on the perceptron layers. The

minimum and maximum unscaling method was used for variables that had previously been scaled to fit within the range of -1 to +1.

Training Strategy

The purpose of approximation is to find the target based on the inputs with the least possible error. The function in neural network that is used to evaluate the generated solution is known as the *loss function* (and the value calculated by the function is referred to as loss). The training strategy refers to the procedure used to carry out the learning process which is applied to neural network to obtain the minimum possible loss. Minimizing the error can be done by finding a set of parameters that fit the neural network to the data set. The two concepts that compose the training strategy are the optimization algorithm and the loss index.

The goal of the learning process of a neural network is to search for a set of parameters that minimizes the loss index. At the lowest value of the loss index, the gradient is zero. The optimization algorithm enables the capability of varying the parameters to obtain the ideal value for each training iteration or epoch, which gradually decreases the loss. The optimization algorithm stops during training after specific conditions or criteria have been met. The common stopping criteria that terminate the learning process are loss reaching the minimum desired value, the process reaching maximum number of epochs or maximum computing time and increment of the selection subset error while training.

Quasi-Newton algorithm and Levenberg-Marquardt algorithm are the different types of optimization algorithms in Neural Designer that were used for the neural network analysis.

The Quasi-Newton method is an optimization algorithm that is based on Newton's method that finds the stationary point of a function (gradient = 0). While Newton's method is used for approximating quadratic functions by using the first- and second-order derivatives to find the stationary point, the successive gradient vectors are analyzed to avoid excessive computational cost.⁸¹

The Quasi-Newton algorithm yields a function with low loss and high accuracy. The algorithm is the default optimization method in Neural Designer and is also recommended for training medium-sized data sets (10–1,000 variables, 1,000–1,000,000 instances). Hence, this optimization algorithm was used for training the datasets of all peak periods of the six bus routes.

The Levenberg-Marquardt training method does not compute the Hessian matrix but achieves second-order training speed. It is applicable when dealing with loss index represented by sum of squares (Mean Squared Error, sum of squared error, etc.). The method can result in a faster and more stable convergence of the model and is used more often to train small to medium-sized data sets.

As mentioned earlier, the learning goal for neural networks involves searching for a set of parameters for which the loss index is at a minimum. The gradient of the learning curve is zero when the condition of minimum loss index is achieved.

The loss index evaluates the performance of a neural network by assessing its parameters. It is a sum of a regularization term and an error term(s) which can be represented mathematically as:

$$LI = RT + ET$$

where

LI = Loss Index

RT = Regularization Term

ET = Error Term.

A regularization term is introduced to the neural network to control the complexity of the neural network model, thereby preventing overfitting. The error term is an important part of the loss expression that measures how well the neural network fits the dataset. There are three types of error in approximation problems based on the different subsets of data: training, selection, and testing errors are measured for the training, selection, and testing instances, respectively.

Model Selection, Testing and Evaluation

Model selection in the Neural Designer software program refers to finding the optimal network architecture with the best generalization properties. Thus, the purpose of model selection is to minimize the error of the selection data set (improving final selection error) for the neural network. Order selection was performed to achieve the best selection model that generated an adequate fit to the data provided. The incremental order selection process was used to obtain the optimal order, training and selection errors to evaluate how well the models were trained. Following the order selection, a second training was done for each peak period of all six bus routes to improve the accuracy of the training model.

The final step in the neural network design was to calculate and document the errors in Neural Designer by testing and evaluating the models. The software was also used to obtain mathematical expressions for the approximation of travel times for different peak periods.

After training the network for the required number of epochs, the models are tested using the set-aside test dataset. For the purpose of this research, Normalized Squared Error (NSE) was used to evaluate the models. NSE is the default error term used when solving approximation problems. The NSE predicts the data and yields a value between 0 (perfect prediction) to 1 (predicting on the basis of the mean). It can be represented by using the following:

$$NSE = \frac{\sum(O - T)^2}{NC}$$

where

NSE = Normalized Squared Error

O = Outputs

T = Targets

NC = Normalized Coefficient.

The Mean Percentage Error (MPE) was also calculated for all the models for each peak period to determine the accuracy. MPE represents the error made in the predicted values as compared to the observed values. It can be represented as:

$$MPE = \frac{1}{n} \frac{\sum |A - P|}{A} * 100$$

where

A = Observed values

P = Predicted values

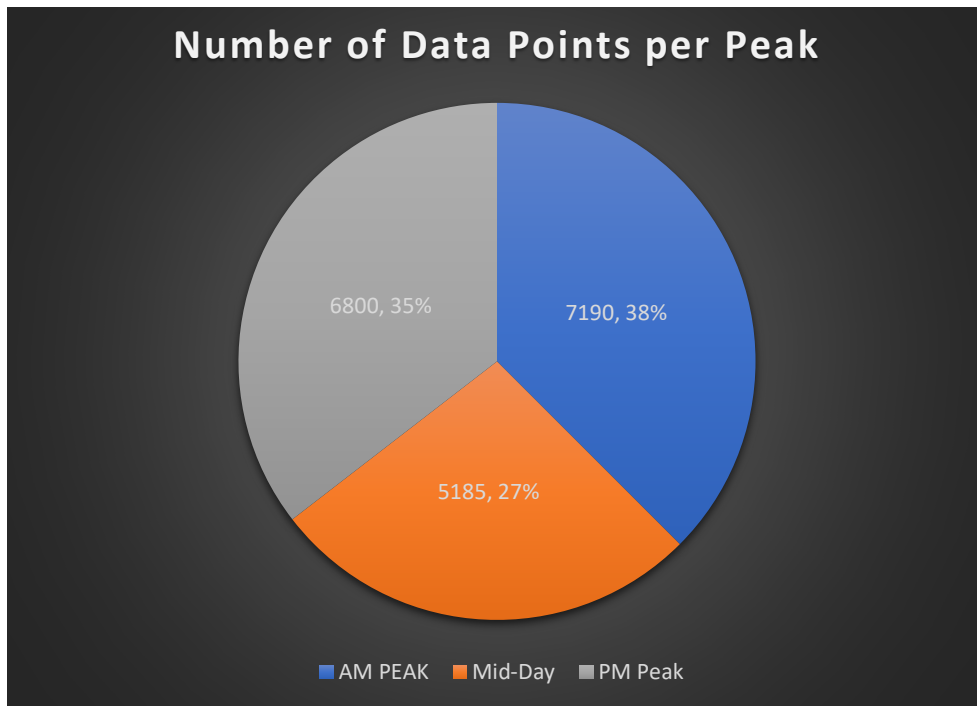
n = Number of observations.

IV. Results

4.1 Summary Statistics

This section presents an overview of the data obtained for all the bus routes during the AM, PM, and Mid-Day Peak hours. The distribution of data points that were used to develop the models is presented in Figure 14.

Figure 14: Total Number of Data Points per Peak Period used for the Analysis



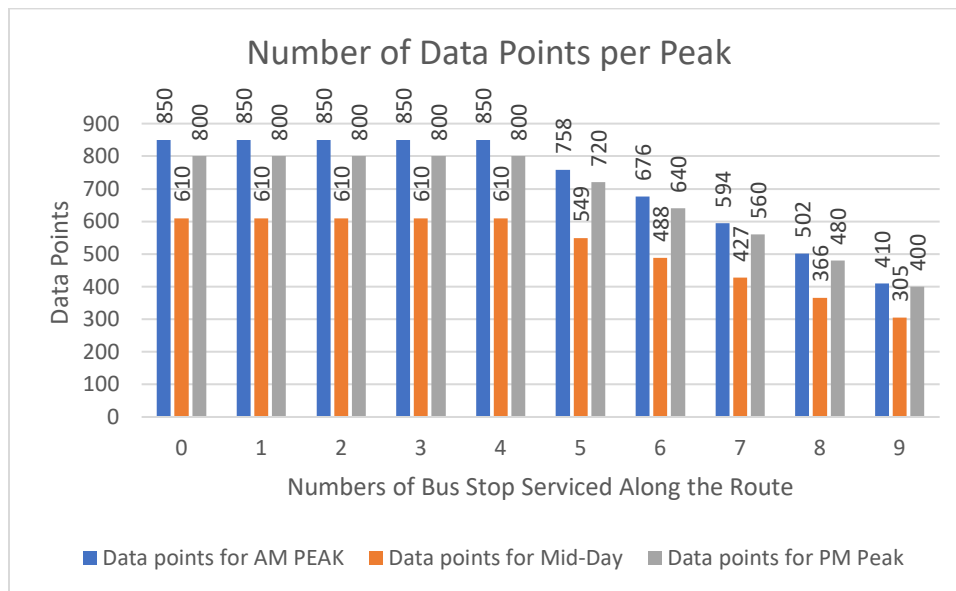
From Figure 14, a total of 7,190, 5,185, and 6,800 data points were used to develop the neural network models respectively for the AM, Mid-Day, and PM Peak periods. Table 4 presents a summary of the average travel time of the buses based on the number of bus stops served during each peak period.

Table 4: Average Travel Time by Number of Served Bus Stops

Number of Served Bus Stops	Average Travel Time (in seconds)		
	AM Peak Period	Mid-Day Peak Period	PM Peak Period
0	87.74	104.64	101.54
1	189.21	221.32	218.95
2	287.27	334.10	333.84
3	389.17	447.58	451.05
4	489.92	562.17	571.99
5	590.48	676.73	686.69
6	690.10	791.78	802.48
7	791.66	907.42	919.30
8	885.84	1023.47	1037.80
9	988.84	1140.49	1160.04

It can be observed from Table 4 that the travel times generally increased over time of the day for all the number of bus stops served (0–9) along the route. The number of data points used for each number (0–9) of bus stops served in the neural network is presented in Figure 15.

Figure 15: Distribution of Data Points per Peak Period for the Different Numbers of Served Bus Stops



From Figure 15, buses serving 0–5 stops had the highest number of data points. Buses serving 9 bus stops along the route contributed to the lowest number of data points used for the matrices analyzed in Neural Designer. Overall, AM Peak period had the most data, followed by the PM and the Mid-Day Peak periods. Hence, the matrices for the AM, PM and Mid-Day Peak periods respectively contained 7,190, 6,800 and 5,185 data points. The descriptive statistics of data sets are presented in Table 5.

Table 5: Descriptive Statistics

	Mean			Standard Deviation		
	AM Peak	Mid-Day Peak	PM Peak	AM Peak	Mid-Day Peak	PM Peak
SBS (X_1)	3.92	3.94	3.94	2.70	2.71	2.71
L (X_2)	6718.90	6734.20	6016.89	5817.77	5663.62	5186.32
P (X_3)	15.58	12.96	16.92	9.38	6.57	9.71
DT (X_4)	13.81	16.08	17.54	9.92	12.10	17.71
IT (X_5)	15.72	16.71	15.67	10.26	10.88	10.39
TT (Y)	481.64	556.68	562.78	324.42	486.29	401.39

Table 5 presents the mean and standard deviation of all the variables (dependent and independent) that were used in all three peak periods. Figures 16 and 17 represent the average dwell time at bus stops and the average number of passengers in a bus at different peak periods throughout the day, respectively.

Figure 16: Average Dwell Time During Different Peak Periods

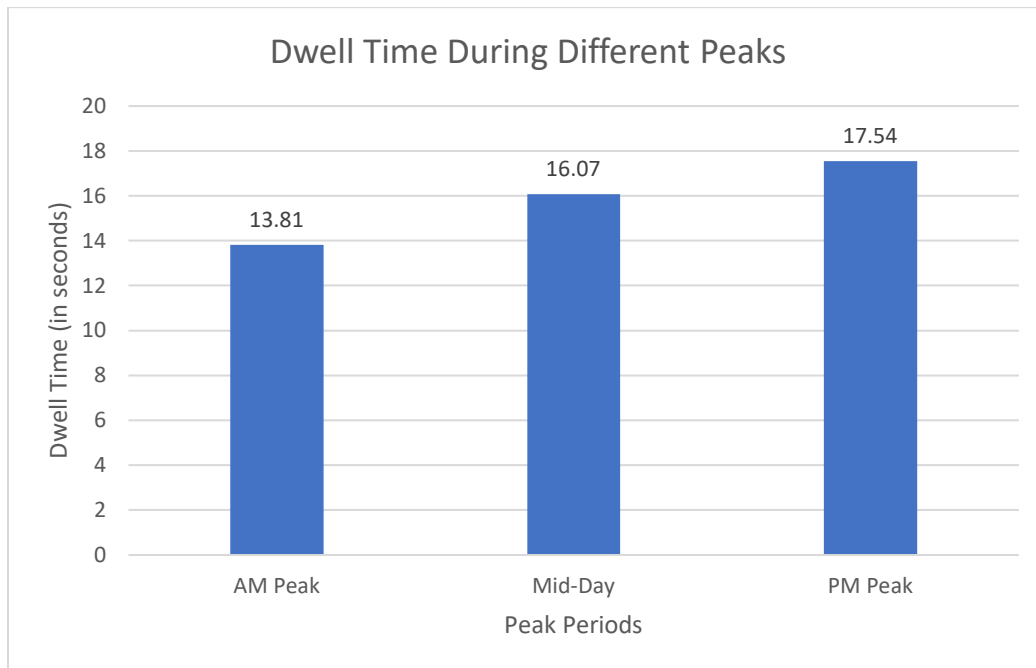
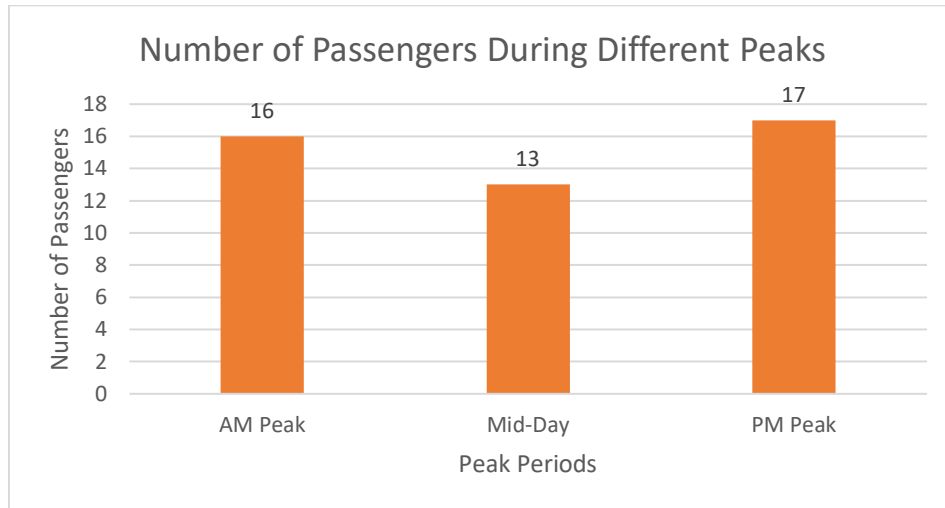


Figure 17: Average Number of Passengers Onboard During Different Peak Periods



It can be observed from Figure 16 that the dwell times at bus stops gradually increase over time. From Figure 17, buses operating during Mid-Day Peak period had the lowest average number of passengers (approximately 13) while PM Peak period had the highest average number of passengers (approximately 17). Figures 18 through 20 present the graphs of the travel times of the buses with respect to the number of intersections the buses pass through along a route for the AM, Mid-Day, and PM models, respectively.

Figure 18: Graph Representing Travel Time of Buses vs. Number of Intersections in between Served Bus Stops During the AM Peak Period

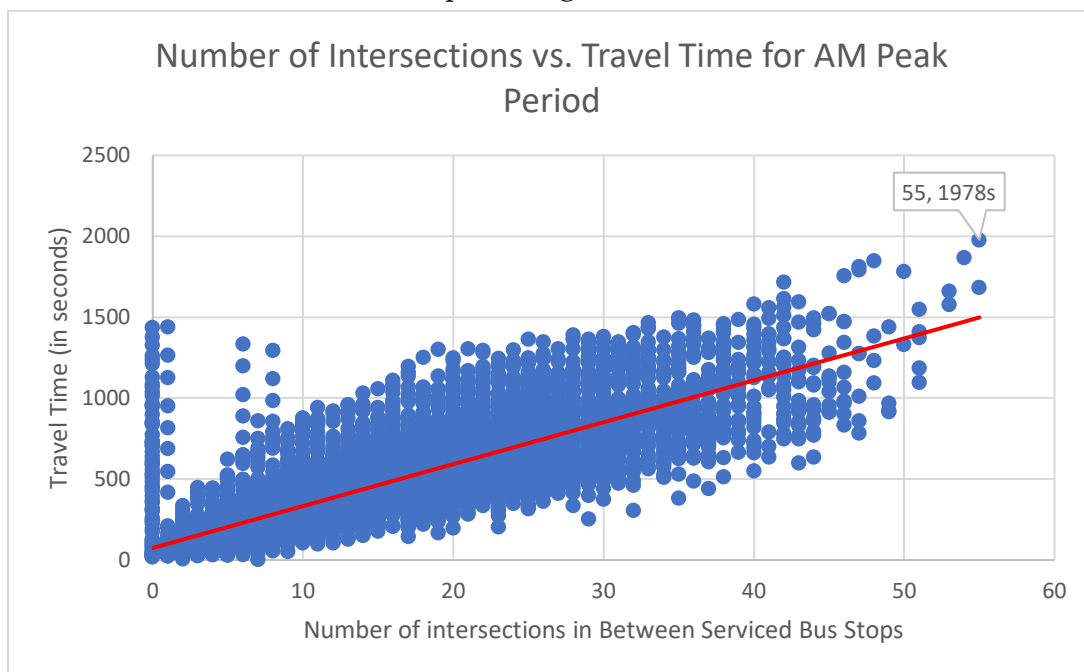


Figure 19: Graph Representing Travel Time of Buses vs. Number of Intersections in between Served Bus Stops During the Mid-Day Peak Period

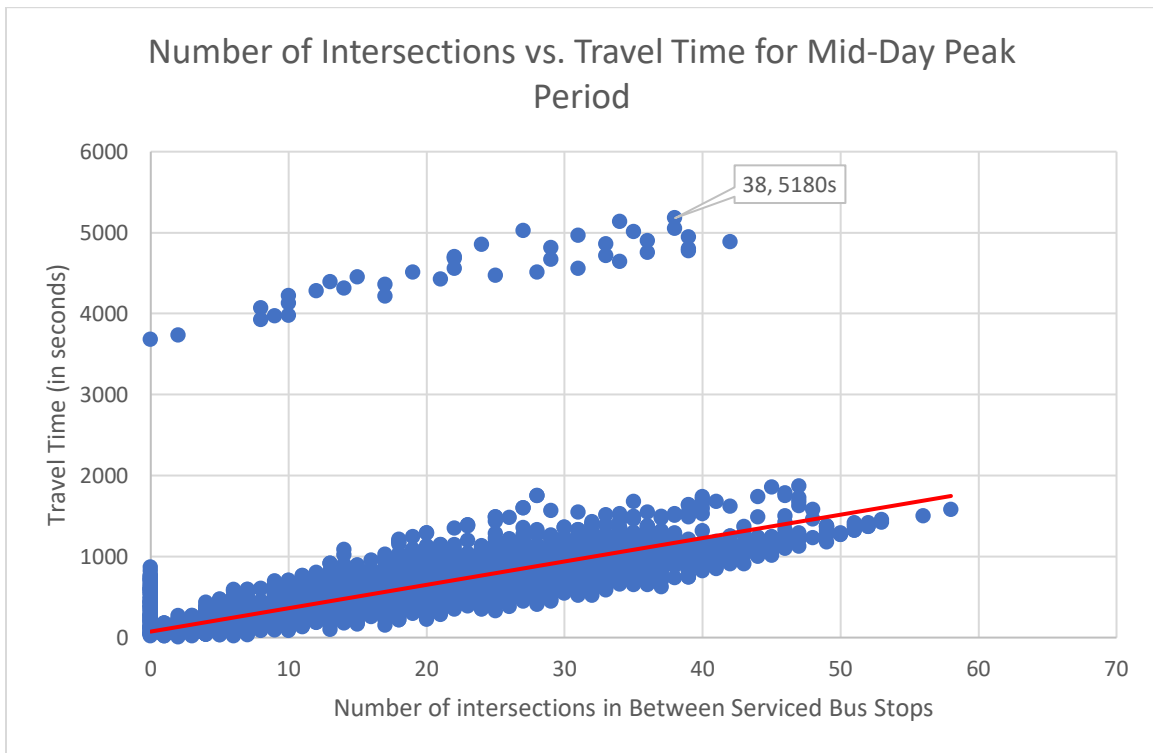
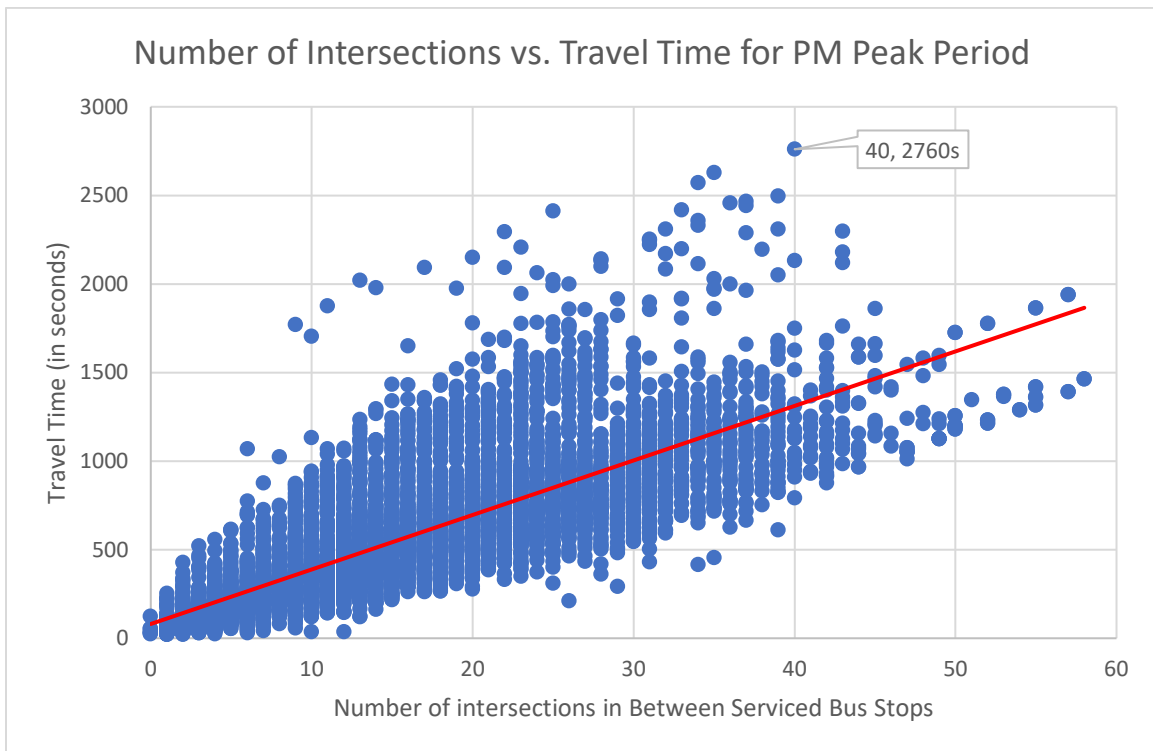


Figure 20: Graph Representing Travel Time of Buses vs. Number of Intersections in between Served Bus Stops During the PM Peak Period



From Figure 18, it can be observed that during the AM Peak period, the longest travel time was approximately 33 minutes (1,978 seconds) for the bus that passed through 55 intersections while serving a route. The longest travel time during the Mid-Day Peak period was approximately 86 minutes (5,180 seconds), as can be seen in Figure 19. The bus traveling 86 minutes passed through 38 intersections on the route. The PM Peak period had the longest bus travel time of 46 minutes (2,760 seconds) as shown in Figure 20 for the bus passing through 40 intersections along the service route.

4.2 Regression Analysis

This section presents results of the regression analyses developed to predict the travel time of buses during different peak periods. Three models were developed: one for the AM Peak (7:00 AM–9:30 AM), one for the Mid-Day Peak (10:00 AM–2:30 PM), and one for the PM Peak (4:00 PM–6:30 PM). The models were multiple linear regressions taking the form:

$$TT_i = \beta_{oi} + \beta_{1i}SBS + \beta_{2i}L + \beta_{3i}P + \beta_{4i}DT + \beta_{5i}IT + \epsilon_i$$

The level of significance for testing the models in SPSS was set to 5%. The following evaluative criteria were used to measure the performance of the models:

1. Statistical significance (using the p-values of the F-statistics with 5% significance level)
2. Goodness-of-fit (using the R² and the adjusted R² values)
3. Statistical significance for the models' predictors (p-values of t-statistics).

While the t-test for each of the predictor variable tests the model against the null hypothesis (true regression coefficient = 0), the F-test compares the fit of the regression model with the fit of a null model with an intercept (β_{oi}) but no predictor variables (where all regression coefficients are set to 0).

The regression analysis used for the study presented outputs that generated five different models per peak period. The model number (1–5) represented the numbers of predictors each model used to generate the outputs. Hence, Model 1 had only the first predictor as a constant value (X_1) while Model 5 had all five predictors (X_1, X_2, X_3, X_4, X_5). Since the prediction of travel time is dependent on all the input variables, the model with all the predictors incorporated is used for each peak period.

Correlation between Travel Time and Independent Variables

The dependencies between a single input independent variable and the target variable are represented by correlations. The range of correlation ranges from -1 to 1 where a value close to -1 signifies a strong negative correlation and a value close to 1 signifies a strong positive relation. A value close to 0 signifies a weak correlation or no correlation. Table 6 presents the correlations between the independent and dependent (target) variables for the AM, Mid-Day, and PM Peak periods.

Table 6: Inputs-Target Correlations

Inputs	Correlation for AM Peak	Correlation for Mid-Day Peak	Correlation for PM Peak
SBS (X_1)	0.834	0.639	0.792
L (X_2)	0.693	0.920	0.641
P (X_3)	-0.076	-0.041	-0.022
DT (X_4)	0.027	-0.011	0.014
IT (X_5)	0.819	0.646	0.796

It can be observed from Table 6 that during the AM Peak period, X_1 (number of bus stops served) had the highest correlation with travel time followed by X_5 (number of intersections between two served bus stops). The lowest correlation was observed for average dwell time, while number of people in the bus had a negative correlation with buses' travel time. During the Mid-Day Peak period, X_2 (length of the route between served bus stops) had the highest correlation with travel time followed by X_5 (number of intersections between two served bus stops). The lowest correlation was observed for average dwell time. The average dwell time and the average number of passengers on the bus are negatively correlated with the travel time during the Mid-Day Peak period. During the PM Peak period, X_5 (number of intersections between two served bus stops) had the highest correlation with travel time followed by X_1 (number of served bus stops). The lowest correlation was observed for average dwell time, while the average number of passengers in the bus had a negative correlation with travel time of the buses.

Travel Time Prediction Models for AM Peak Period

The results of the regression analyses conducted using SPSS for the AM Peak period are presented in Table 7.

Table 7: Results of the Regression Analyses for AM Peak Period

MODEL SUMMARY			
R	R-Squared	Adjusted R-Squared	Standard Error
0.873	0.762	0.762	158.240
ANOVA SUMMARY			
Model	df	F	Sig
Regression	5	4606.017	0.000
COEFFICIENTS			
Variable	Unstandardized B	t	Sig.
Constant	40.364	7.691	0.000
SBS (X_1)	62.504	47.880	0.000
L (X_2)	0.0110	21.112	0.000
P (X_3)	-1.366	-6.469	0.000
DT (X_4)	2.177	11.208	0.000
IT (X_5)	7.406	17.859	0.000

The results showed that the multiple regression model for the AM Peak period is statistically significant, at a 5% level of significance. The F -value was 4,606.02 (degrees of freedom = 5), which was also significant (p -value < 0.05). The effects of all the independent predictor variables were determined to be statistically significant, at the 5% significance level (p -values of less than 0.05). The best-fitting model for the AM Peak period, with an R^2 value of 0.76, was determined to be $TT_A = 40.36 + 62.50SBS + 0.11L - 1.37P + 2.17DT + 7.41T$, where TT_A is the travel time of the bus during AM Peak period.

Travel Time Prediction Models for Mid-Day Peak Period

The results of the regression analyses for the Mid-Day Peak period are presented in Table 8.

Table 8: Results of the Regression Analyses for Mid-Day Peak Period

MODEL SUMMARY			
R	R-Squared	Adjusted R-Squared	Standard Error
0.924	0.854	0.854	186.120
ANOVA SUMMARY			
Model	df	F	Sig.
Regression	5	6041.668	0.000
COEFFICIENTS			
Variable	Unstandardized B	t	Sig.
Constant	41.069	5.439	0.000
SBS (X_1)	31.942	17.739	0.000
L (X_2)	0.079	119.604	0.000
P (X_3)	0.115	0.288	0.000
DT (X_4)	-1.508	-6.966	0.000
IT (X_5)	-6.973	-14.346	0.000

The results showed that the multiple regression model for the Mid-Day Peak period is statistically significant, at a 95% confidence interval. The F -value was 6,041.67 (degrees of freedom = 5), which was also significant (p -value < 0.05). The effects of all the independent predictor variables were determined to be statistically significant at a 5% level of significance (p -values of < 0.05). The best-fitting model for the Mid-Day Peak period, with an R^2 value of 0.85, was determined to be $TT_M = 41.07 + 31.94SBS + 0.08L + 0.115P - 1.51DT - 6.97T$, where TT_M is the travel time of the bus during Mid-Day Peak period.

Travel Time Prediction Models for PM Peak Period

Table 9 presents the results of the regression analyses for the PM Peak period.

Table 9: Results of the Regression Analyses for PM Peak Period

MODEL SUMMARY			
R	R-Squared	Adjusted R-Squared	Standard Error
0.836	0.699	0.688	220.458
ANOVA SUMMARY			
Model	df	F	Sig
Regression	5	3148.810	0.000
COEFFICIENTS			
Variable	Unstandardized B	t	Sig.
Constant	-6.063	-0.808	0.419
SBS (X ₁)	55.669	30.141	0.000
L (X ₂)	0.011	14.231	0.000
P (X ₃)	1.883	6.585	0.000
DT (X ₄)	0.888	5.810	0.000
IT (X ₅)	15.242	29.538	0.000

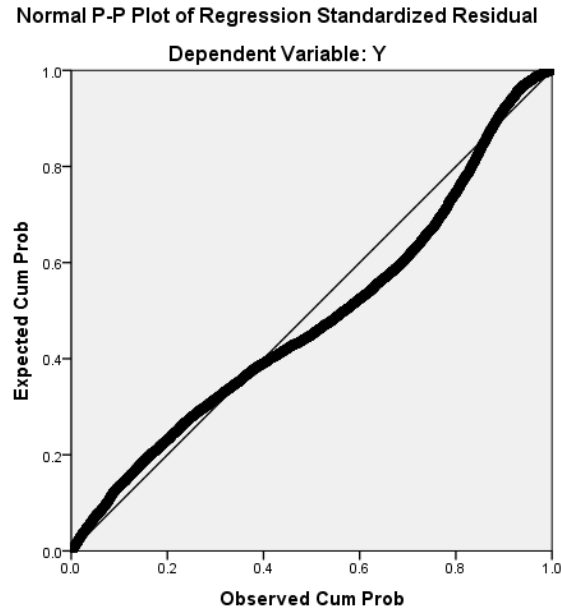
The results showed that the multiple regression model for the PM Peak period is statistically significant at a 95% confidence interval. The F -value was 3,148.81 (degrees of freedom = 5), which was also statistically significant (p -value < 0.05). The effects of all the independent predictor variables were determined to be statistically significant at a significance level of 5% (p -values of less than 0.05). However, the constant term was not statistically significant (p -value = 0.419). The best-fitting model for the PM Peak period, with an R^2 value of 0.68, was determined to be $TT_P = -6.063 + 55.67SBS + 0.11L + 1.88P + 0.89DT + 15.24T$, where TT_P is the travel time of the bus during PM Peak period.

4.3 Model Testing

To test the models, normality of errors, multicollinearity and homoscedasticity of all peak periods were evaluated.

The normal probability plot was used to test the assumption of normality of errors. The observed cumulative probabilities of the standardized residuals were plotted against the expected cumulative probabilities of the standardized residuals. By visually inspecting the plots for all three models (presented in Appendix A), it can be observed that the curves closely follow the diagonals of the plots. This is an indication that the errors are normally distributed. Figure 21 shows the normal plot for the AM peak period, while those for the Mid-Day and PM Peak periods are presented in Appendix A.

Figure 21: Normality Probability Curve Obtained for AM Peak Period



Following normality of errors, the test for multicollinearity showed that the VIF for all the variables in all three models was less than the maximum value of 10. Thus, there was no multicollinearity between the independent variables. Table 10 shows the summary of the VIF values for the AM Peak models.

Table 10: SPSS Coefficient Output for AM Peak Period

Model	Coefficients ^a								
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
(Constant)	40.364	5.248		7.691	.000	30.076	50.652		
SBS (X ₁)	62.504	1.305	.521	47.880	.000	59.945	65.063	.280	3.576
L (X ₂)	.011	.000	.189	21.112	.000	.010	.012	.412	2.424
P (X ₃)	-1.366	.211	-.040	-6.469	.000	-1.780	-.952	.888	1.126
DT (X ₄)	2.177	.194	.067	11.208	.000	1.796	2.557	.938	1.066
IT (X ₅)	7.406	.415	.234	17.859	.000	6.593	8.219	.192	5.199

The values used to determine multicollinearity based on tolerance and VIF for the Mid-Day and PM Peak models are presented in Appendix A.

Finally, visual inspection of the residual plots of the three models (presented in Appendix A) shows a relatively even distribution about the zero line, confirming that the variances of the residuals of the dependent variables are the same for all values of the independent variables. Figure 22 shows the residual plot for the AM Peak period, while those for the Mid-Day and PM Peak periods are presented in Appendix A.

Figure 22: Homoscedasticity Regression Plot for AM Peak Period

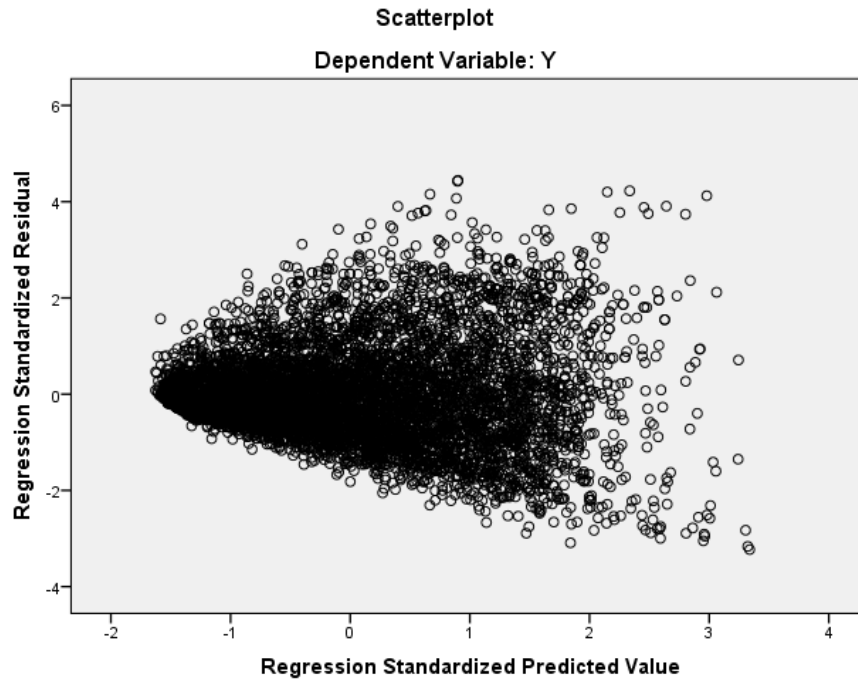


Table 11 presents a summary of the models developed for different peak periods using the Multiple Linear Regression approach.

Table 11: Summary of Results

#	Peak Period	Model	R ²
1	AM	$TT_A = 40.36 + 62.50SBS + 0.11L - 1.37P + 2.17DT + 7.41T$	0.76
2	Mid-Day	$TT_M = 41.07 + 31.94SBS + 0.08L + 0.12P - 1.51DT - 6.97T$	0.85
3	PM	$TT_P = -6.063 + 55.67SBS + 0.11L + 1.88P + 0.89DT + 15.24T$	0.68

4.4 Neural Network

This section presents results from the neural networks which were developed to predict the travel time of buses during different peak periods using the Neural Designer software. Matrices for all three peaks were analyzed individually using the Quasi-Newton and the Levenberg-Marquardt algorithms separately.

Quasi-Newton Optimization Algorithm

The Quasi-Newton optimization algorithm was used for the first set of neural network analyses for all three peak period matrices. The algorithm is the default optimization method in Neural Designer and is also recommended for training medium-sized data sets (10–1,000 variables, 1,000–1,000,000 instances). Hence, this optimization algorithm was used to train the data sets of all the peak periods for the six bus routes. Each matrix was subjected to a different number of perceptron layers for the analysis. Three sets of perceptron layers (two layers, three layers, and five layers) were

used to determine the ideal model with the optimum number of hidden layers for travel time prediction.

The inputs used for all Quasi-Newton analyses were scaled using the automatic scaling method. The size of the scaling layer was 5 (number of inputs). The scaled outputs were unscaled back to the original units using the unscaling layers on the perceptron layers for all peak periods. The unscaling method for the output layer was the minimum and maximum method. The size of the unscaling layer is 1 (number of outputs). The unscaling minimum, maximum, mean, and standard deviation for the models were the same across the different number of perceptron layers (two, three, and five layers).

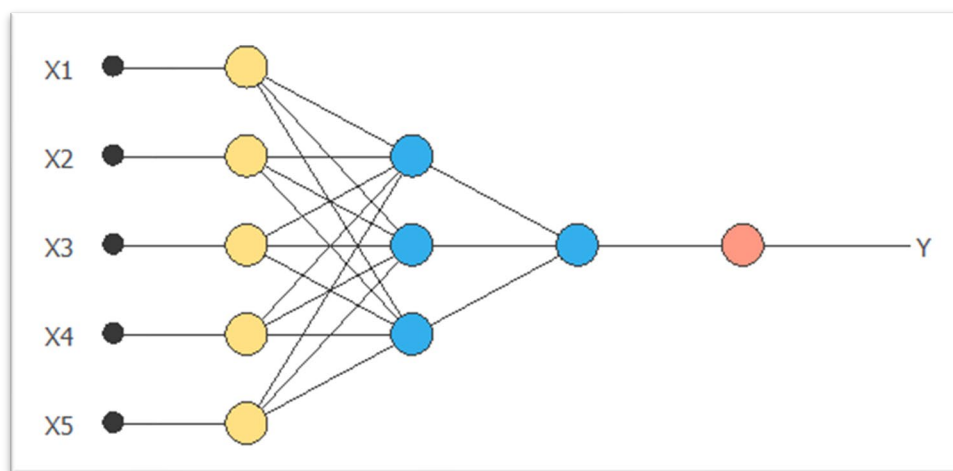
The values used for scaling and unscaling for all three peak periods are presented in Appendix B.

Neural networks were developed for all three peak periods using two, three and five perceptron layers (separately) and analyzed using the Quasi-Newton optimization algorithm.

For the models that had two perceptron layers, the first perceptron layer of all peak periods used all independent input variables for this analysis, while the second layer only used one of the five inputs. The sizes of all layers and the corresponding activation functions for all three peak periods incorporating data from six routes are presented in Appendix B. The activation functions of the first layer were set to a hyperbolic tangent, while those of second layer were set as linear.

Figure 23 presents the typical neural architecture for networks with two perceptron layers (all peak periods).

Figure 23: Typical Neural Network Architecture for a Network with Two Perceptron Layers



The yellow, blue, and red circles represent the scaling neurons, the perceptron neurons, and the unscaling neurons, respectively. Additional perceptron layers are represented by the increment of the hidden layers, which apply different weights to the input variables of the neural network to minimize the errors of the output. The number of perceptron neurons in a layer can change after

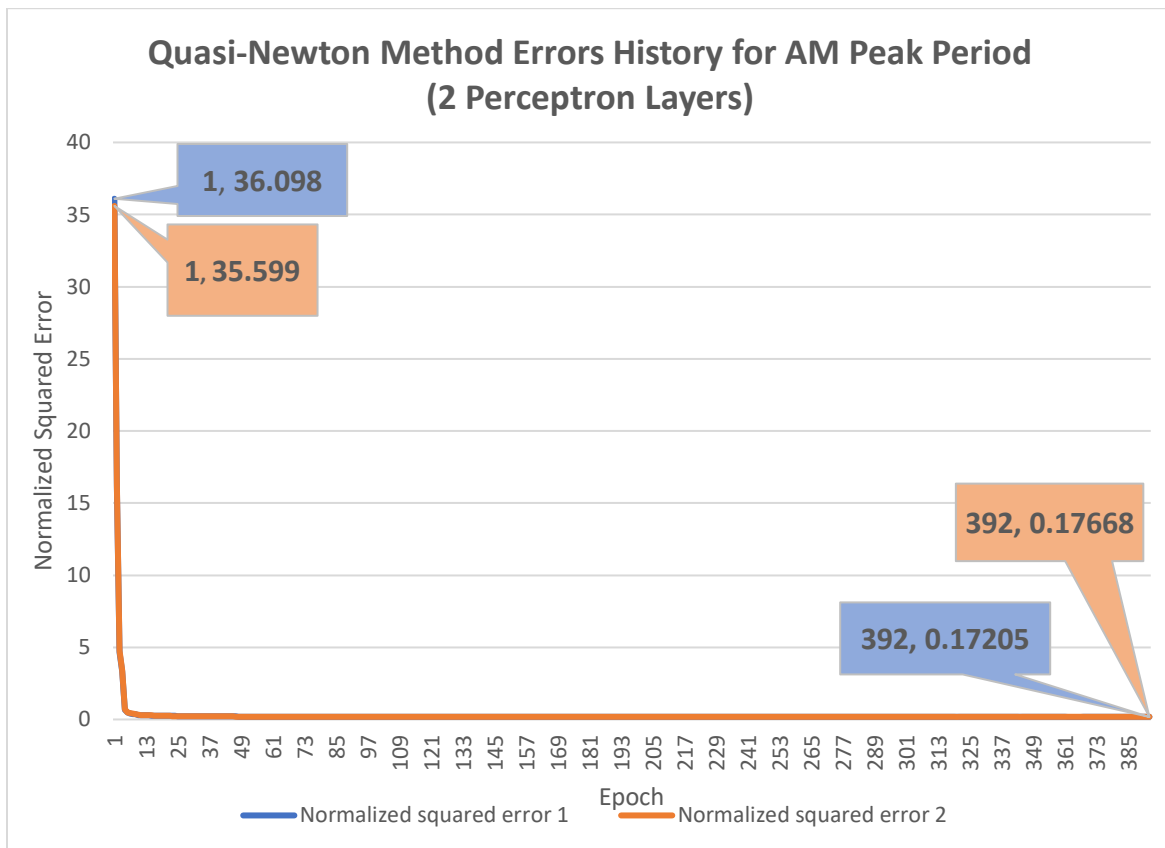
order selection. The neural network architecture for all the analytical methods used in the research (before and after the order selection) is presented in Appendix C.

The results of the Quasi-Newton optimization algorithm (two perceptron layers) are presented in Table 12. Plots showing the training and selection errors in each iteration for the AM Peak period is shown in Figure 24. The training plots for the Mid-Day and PM Peak periods can be found in Appendix D.

Table 12: Quasi-Newton Method Results for Two Perceptron Layers

	AM Peak Value	Mid-Day Peak Value	PM Peak Value
Initial Training Error	36.100	150.010	33.860
Final Training Error	0.172	0.0537	0.248
Initial Selection Error	35.600	118.750	35.550
Final Selection Error	0.177	0.042	0.265
Epochs Number	392	403	164
Elapsed Time (min:sec)	00:00	00:01	00:00
Stopping Criterion	Gradient norm goal	Gradient norm goal	Gradient norm goal

Figure 24: Quasi-Newton Method Error History Plot for AM Peak Period (Two Perceptron Layers)



The blue line in Figure 24 represents the training errors, and the orange line (overlapping the blue line) represent the selection errors. It can be observed from Table 12 and Figure 24 that during

the AM Peak period, for the initial training error value of 36.098, the final value after 392 epochs decreased to 0.172. The initial value of the selection error in the AM Peak period was 35.599, and the final value after 392 epochs decreased to 0.176. The error histories for the Mid-Day and PM Peak periods are presented in Appendix D. For the Mid-Day Peak period, the initial training error of 150.005 decreased to 0.0537 after 403 epochs. Similarly, after 403 epochs, the initial value of selection error for the Mid-Day Peak period decreased from 118.745 to 0.0418. For the PM Peak period, the initial training error decreased from 33.859 to 0.248 after 164 epochs, and the initial selection error decreased from 35.551 to 0.246. From the table, it appears that the Mid-Day Peak period had the highest initial training and selection errors. The Mid-Day Peak errors had the highest percentage reduction after training when compared to other peak periods.

Order selection was performed to achieve the best selection model that generated an adequate fit for all the peak periods. The incremental order selection process was used to obtain the optimal order, training, and selection errors. Table 13 shows the order selection results by the incremental order algorithm for the Quasi-Newton method with two perceptron layers for all three peak periods.

Table 13: Quasi-Newton Method Results for Two Perceptron Layers

	AM Peak Value	Mid-Day Peak Value	PM Peak Value
Optimal Order	10	4	6
Optimal Training Error	0.139	0.517	0.218
Optimal Selection Error	0.149	0.041	0.229
Iterations Number	10	10	10
Elapsed Time (min:sec)	01:38	00:48	00:47

Following the order selection process, another training was conducted to achieve better neural network performance. The optimal order represents the complexity of the hidden layers in the neural network. Table 14 shows the results of the second training for the model for all peak periods. Plots showing the training and selection errors in each iteration after order selection for the AM, Mid-Day, and PM Peak periods are presented in Appendix D.

Table 14: Quasi-Newton Method Results of All Peak Periods for Two Perceptron Layers (after Order Selection)

	AM Peak Value	Mid-Day Peak Value	PM Peak Value
Initial Training Error	7.670	60.580	48.390
Final Training Error	0.138	0.052	0.218
Initial Selection Error	8.250	46.510	51.460
Final Selection Error	0.145	0.041	0.230
Epochs Number	615	336	308
Elapsed Time (min:sec)	00:03	00:01	00:01
Stopping Criterion	Gradient norm goal	Gradient norm goal	Gradient norm goal

It can be observed from Table 14 that for the AM Peak period, the initial training error value of 7.67 decreased to 0.138 after 615 epochs. The initial value of the selection error for the AM Peak

period was 8.25, and the final value after 615 epochs decreased to 0.145. For the Mid-Day Peak period, the initial training error of 60.58 decreased to 0.052 after 336 epochs. Similarly, after 336 epochs, the initial value of selection error for the Mid-Day Peak period decreased from 46.51 to 0.041. For the PM Peak period, the initial training error decreased from 48.39 to 0.218 after 308 epochs, and the initial selection error decreased from 51.46 to 0.229. It can be observed that the initial selection and training errors for all the peak periods, except the PM Peak period, were higher before the order selection process was carried out. Order selection reduced the final values of training and selection errors across all peak periods.

The normalized squared errors were measured for all the instances and were also used to evaluate all models. Table 15 presents the normalized squared errors for the training, selection, and testing instances that were obtained from the model containing two perceptron layers in Neural Designer. Other errors, like mean squared errors and sum of squared errors, are provided in Appendix E.

Table 15: Normalized Squared Errors for Training, Selection and Testing Instances
(Two-Layer Quasi-Newton Method)

	AM Peak	Mid-Day Peak	PM Peak
Training Error	0.138	0.052	0.218
Selection Error	0.145	0.041	0.229
Testing Error	0.147	0.053	0.257

It can be observed from Table 15 that the lowest testing error was obtained for the Mid-Day Peak model. The error statistics highlighting errors between the neural network and the testing instances in the data set for all peak periods are presented in Table 16. The minimum, maximum, and standard deviation errors are provided in Appendix E.

Table 16: Error Statistics for Quasi-Newton Method for All Peak periods
(Two Perceptron Layers)

	AM Peak Error	Mid-Day Peak Error	PM Peak Error
Absolute Error	90.727	78.129	134.590
Percentage Error	4.591	1.510	4.914

The expressions take the inputs X_1 through X_5 to produce the output Y where the information is propagated in a feed forward fashion through the scaling layer, the perceptron layers, and the unscaling layer.

Table 17 presents the mathematical expressions that were obtained from the neural network analysis for all peak models undergoing Quasi-Newton and Levenberg-Marquardt analyses.

Table 17: Neural Network Analysis Output Equations

Travel Time	Equation
AM	$0.5 * (\text{scaled_Y}_{AM} + 1.0) * (1976) + 2$
Mid-Day	$0.5 * (\text{scaled_Y}_{\text{Mid-Day}} + 1.0) * (5174) + 6$
PM	$0.5 * (\text{scaled_Y}_{PM} + 1.0) * (2739) + 21$

Table 17 represents general equations for all three peak periods for the different analyses (with multiple perceptron layers). It should be noted that the scaled values of Y_{AM} , $Y_{Mid-Day}$, and Y_{PM} change with different peak periods and different analyses. The mathematical expressions for the scaled Y outputs that were obtained from the neural network analysis for all peak models undergoing Quasi-Newton analysis (two perceptron layers) can be represented as:

$$\text{scaled_}Y_{AM} = (-0.403933 + (y_{1_1} * -0.827554) + (y_{1_2} * 0.475845) + (y_{1_3} * 1.51826) + (y_{1_4} * 0.184144) + (y_{1_5} * -0.241722) + (y_{1_6} * -1.17179) + (y_{1_7} * 0.385873) + (y_{1_8} * -0.688366) + (y_{1_9} * 0.695668) + (y_{1_{10}} * 2.37657))$$

$$\text{scaled_}Y_{Mid-Day} = (-0.208024 + (y_{1_1} * -0.159253) + (y_{1_2} * -1.66349) + (y_{1_3} * 0.801582) + (y_{1_4} * 1.1423))$$

$$\text{scaled_}Y_{PM} = (-0.109188 + (y_{1_1} * 1.08879) + (y_{1_2} * -2.69854) + (y_{1_3} * -0.551765) + (y_{1_4} * -2.13805) + (y_{1_5} * -0.937324) + (y_{1_6} * 1.2105))$$

The values of y_{x_x} for the AM, Mid-Day, and PM Peak periods are presented in Appendix E.

Neural networks for all three peak periods having three perceptron layers were also modeled and analyzed using the Quasi-Newton optimization algorithm.

The first and second perceptron layers of all peak periods used all independent input variables, while the third layer only used one of the five inputs. The activation functions of the first two layers were set to hyperbolic tangent, while those of third layer of each peak period were set to linear. The sizes of all layers and the corresponding activation functions for all three peak periods incorporating data from six routes are presented in Appendix B.

The results of the Quasi-Newton optimization algorithm obtained by applying three perceptron layers are shown in Table 18. Plots showing the training and selection errors in each iteration for the AM, Mid-Day, and PM Peak periods are presented in Appendix D.

Table 18: Quasi-Newton Method Results for Three Perceptron Layers

	AM Peak Value	Mid-Day Peak Value	PM Peak Value
Initial Training Error	12.460	99.100	29.950
Final Training Error	0.161	0.053	0.233
Initial Selection Error	11.680	81.540	31.800
Final Selection Error	0.161	0.043	0.259
Epochs Number	1,000	578	1,000
Elapsed Time (min:sec)	00:03	00:02	00:03
Stopping Criterion	Maximum number of iterations	Gradient norm goal	Maximum number of iterations

It can be observed from Table 18 that during the AM Peak period, for the initial training error value of 12.46, the final value after 1,000 epochs decreased to 0.161. The initial value of the selection error in the AM Peak period was 11.68, which decreased to 0.161 after 1,000 epochs.

For the Mid-Day Peak period, the initial training error of 99.1 decreased to 0.05 after 578 epochs. Similarly, after 578 epochs, the initial value of selection error for the Mid-Day Peak period decreased from 81.54 to 0.043. For the PM Peak period, the initial training error decreased from 29.95 to 0.233 after 1,000 epochs, and the initial selection error decreased from 31.799 to 0.259. From Table 18, it appears that the Mid-Day Peak period had both the highest and the lowest initial/ final training and selection errors. Additionally, the Mid-Day Peak period converged earlier (578 epochs) compared to the AM and PM Peak periods (1,000 epochs), as the gradient normal goal was met in the training process.

Order selection was performed to achieve the best selection model that generated the adequate fit for all the peak periods. The incremental order selection process was used to obtain the optimal order, training, and selection errors. Table 19 shows the order selection results obtained by the incremental order algorithm for the Quasi-Newton method with three perceptron layers for all three peak periods.

Table 19: Quasi-Newton Method Results for Three Perceptron Layers (All Peak Periods)

	AM Peak Value	Mid-Day Peak Value	PM Peak Value
Optimal Order	9	1	10
Optimal Training Error	0.145	0.053	0.208
Optimal Selection Error	0.153	0.042	0.229
Iterations Number	10	10	10
Elapsed Time (min:sec)	01:23	00:53	04:08

Following the order selection process, another training was conducted to achieve better neural network performance for the three-layered model. Table 20 shows the results of the second training for the model for all the peak periods. A plot showing the training and selection errors in each iteration after the order selection process for the AM, Mid-Day, and PM Peak periods is presented in Appendix D.

Table 20: Quasi-Newton Method Results of All Peak Periods for Three Perceptron Layers (After Order Selection)

	AM Peak Value	Mid-Day Peak Value	PM Peak Value
Initial Training Error	16.170	93.280	18.000
Final Training Error	0.145	0.053	0.205
Initial Selection Error	16.080	72.120	18.320
Final Selection Error	0.154	0.042	0.229
Epochs Number	812	404	1,000
Elapsed Time (min:sec)	00:04	00:01	00:06
Stopping Criterion	Gradient norm goal	Gradient norm goal	Maximum number of iterations

It can be observed from Table 20 that for the AM Peak period, the initial training error value of 16.17 decreased to 0.145 after 821 epochs. The initial value of the selection error for the AM Peak

period was 16.08, and the final value after 821 epochs decreased to 0.154. From Table 20, it can also be observed that for the Mid-Day Peak period, the initial training error of 93.28 decreased to 0.053 after 404 epochs. Similarly, after 404 epochs, the initial value of selection error for the Mid-Day Peak period decreased from 72.12 to 0.042. For the PM Peak period, initial training error decreased from 18.00 to 0.205 after 1,000 epochs, and the initial selection error decreased from 18.32 to 0.229. It can be observed that the initial selection and training errors for all the peak periods, except the PM Peak period, were higher before the order selection process was carried out. Order selection reduced the final values of training and selection errors for all the three peak periods.

The normalized squared errors were obtained for all the instances in order to evaluate the model for each use. Table 21 presents the normalized squared errors for the training, selection, and testing instances that were obtained from the model containing three perceptron layers in Neural Designer. Other errors including mean squared errors, sum of squared errors, and others, are provided in Appendix E.

Table 21: Errors for Training, Selection and Testing Instances
(Three-Layer Quasi-Newton Method)

	AM Peak Error	Mid-Day Peak Error	PM Peak Error
Training Error	0.145	0.053	0.205
Selection Error	0.154	0.042	0.229
Testing Error	0.163	0.055	0.254

It can be observed from Table 21 that the lowest testing error was obtained for the Mid-Day Peak model. The error statistics highlighting errors between the neural network and the testing instances in the dataset for all peak periods are presented in Table 22. The minimum, maximum, and standard deviation errors are provided in Appendix E.

Table 22: Error Statistics for Quasi-Newton Method for All Peak Periods
(Three Perceptron Layers)

	AM Peak Error	Mid-Day Peak Error	PM Peak Error
Absolute Error	94.572	79.485	132.040
Percentage Error	4.786	1.536	4.821

The mathematical expressions for the scaled Y outputs that were obtained from the neural network analysis for all peak period models undergoing Quasi-Newton analysis (three perceptron layers) can be represented as follows:

$$\text{scaled_Y}_{AM} = (0.215174 + (y_{2_1} * -0.975811) + (y_{2_2} * 1.23331) + (y_{2_3} * 1.12438) + (y_{2_4} * -1.60024) + (y_{2_5} * -1.04541) + (y_{2_6} * 1.33817) + (y_{2_7} * 0.271241) + (y_{2_8} * -0.351073) + (y_{2_9} * 1.20335))$$

$$\text{scaled_Y}_{Mid-Day} = (0.103787 + (y_{2_1} * 1.6249))$$

$$\text{scaled_Y}_{\text{PM}} = (0.684445 + (y_{2_1} * 1.35074) + (y_{2_2} * -1.28872) + (y_{2_3} * 2.40162) + (y_{2_4} * 1.26585) + (y_{2_5} * -1.08219) + (y_{2_6} * 1.46912) + (y_{2_7} * -0.888777) + (y_{2_8} * -1.32063) + (y_{2_9} * -0.494677) + (y_{2_10} * 1.01097))$$

The values of $y_{x,x}$ for the AM, Mid-Day, and PM Peak periods are presented in Appendix F.

Finally, neural networks for all three peak periods having five perceptron layers were also modeled and analyzed using the Quasi-Newton optimization algorithm. The activation functions of the first four layers were set to hyperbolic tangent, while those of last layer of each peak period were set to linear. The sizes of all layers and the corresponding activation functions for all three peak periods incorporating data from six routes are presented in Appendix B.

The results of the Quasi-Newton optimization algorithm obtained by applying five perceptron layers are shown in Table 23. Plots showing the training and selection errors in each iteration for the AM, Mid-Day, and PM Peak periods are presented in Appendix D.

Table 23: Quasi-Newton Method Results for Five Perceptron Layers

	AM Peak Value	Mid-Day Peak Value	PM Peak Value
Initial Training Error	83.040	15.130	35.540
Final Training Error	0.150	0.097	0.235
Initial Selection Error	80.730	14.450	36.970
Final Selection Error	0.166	0.103	0.257
Epochs Number	1,000	634	927
Elapsed Time (min:sec)	00:12	00:03	00:05
Stopping Criterion	Maximum number of iterations	Gradient norm goal	Gradient norm goal

It can be observed from Table 23 that during the AM Peak period, for the initial training error value of 83.04, the final value after 1,000 epochs decreased to 0.150. The initial value of the selection error in the AM Peak period was 80.73 and decreased to 0.166 after 1,000 epochs. From Table 23, it can also be observed that for the Mid-Day Peak period, the initial training error of 15.13 decreased to 0.097 after 578 epochs. Similarly, after 634 epochs, the initial value of selection error for the Mid-Day Peak period decreased from 14.45 to 0.103. For the PM Peak period, the initial training error decreased from 61.08 to 0.218 after 1,000 epochs, and the initial selection error decreased from 65.75 to 0.238. From the table, it appears that the Mid-Day Peak period had the highest initial training and selection errors, which decreased to the lowest values for both training in comparison to other peak periods. Additionally, Mid-Day and PM Peak periods converged earlier (634 and 927 epochs) compared to the AM Peak period (1,000 epochs) as the gradient normal goals were met during the training process.

Order selection was performed to achieve the best selection model that generated the adequate fit for all the peak periods. The incremental order selection process was used to obtain the optimal order, training, and selection errors. Table 24 shows the order selection results obtained by using the incremental order algorithm for the Quasi-Newton method with five perceptron layers for all three peak periods.

Table 24: Quasi-Newton Method Results for Five Perceptron Layers (All Peak Periods)

	AM Peak Value	Mid-Day Peak Value	PM Peak Value
Optimal Order	1	1	2
Optimal Training Error	0.151	0.095	0.212
Optimal Selection Error	0.159	0.099	0.225
Iterations Number	10	10	10
Elapsed Time (min:sec)	06:28	02:01	02:44

Following the order selection process, another training was conducted to achieve better neural network performance for the five-layered model. Table 25 shows the results of the second training for the model for all peak periods. A plot showing the training and selection errors in each iteration after the order selection process for the AM, Mid-Day, and PM Peak periods is presented in Appendix D.

Table 25: Quasi-Newton Method Results of All Peak Periods for Five Perceptron Layers (After Order Selection)

	AM Peak Value	Mid-Day Peak Value	PM Peak Value
Initial Training Error	96.800	59.490	61.080
Final Training Error	0.150	0.094	0.235
Initial Selection Error	96.640	56.440	65.750
Final Selection Error	0.159	0.106	0.257
Epochs Number	1,000	1,000	927
Elapsed Time (min:sec)	00:05	00:03	00:05
Stopping Criterion	Maximum number of iterations	Maximum number of iterations	Gradient norm goal

It can be observed from Table 25 that for the AM peak period, the initial training error value of 96.80 decreased to 0.150 after 1,000 epochs. The initial value of the selection error for the AM Peak period was 96.64, and the final value (after 1,000 epochs) decreased to 0.159. From Table 25, it can also be observed that for the Mid-Day Peak period, the initial training error of 59.49 decreased to 0.094 after 1,000 epochs. Similarly, after 1,000 epochs, the initial value of selection error for the Mid-Day Peak period decreased from 56.44 to 0.106. For the PM Peak period, initial training error decreased from 61.08 to 0.235 after 1,000 epochs and the initial selection error decreased from 65.75 to 0.257. It can be observed that the initial selection and training errors for the AM and PM Peak periods were higher after the order selection process was conducted. Order selection reduced the final values of training and selection errors for the AM Peak period after introducing five perceptron layers. Moreover, the final training error for the PM Peak period was also reduced after order selection.

The normalized squared errors were measured for all the instances to evaluate the model for each use. Table 26 presents the normalized squared errors for the training, selection, and testing instances that were obtained from the model containing five perceptron layers in Neural Designer.

Other errors, including mean squared errors, sum of squared errors, and others, are provided in Appendix E.

Table 26: Errors for Training, Selection and Testing Instances
(Five-Layer Quasi-Newton Method)

	AM Peak Error	Mid-Day Peak Error	PM Peak Error
Training Error	0.150	0.094	0.212
Selection Error	0.159	0.106	0.225
Testing Error	0.169	0.119	0.258

It can be observed from Table 26 that the lowest testing error was obtained for the Mid-Day Peak period model. The error statistics highlighting errors between the neural network and the testing instances in the data set for all peak periods are presented in Table 27. The minimum, maximum, and standard deviation errors are presented in Appendix E.

Table 27: Error Statistics for Quasi-Newton Method for All Peak Periods
(Five Perceptron Layers)

	AM Peak Error	Mid-Day Peak Error	PM Peak Error
Absolute Error	97.242	79.237	131.381
Percentage Error	4.921	4.251	4.797

The mathematical expressions for the scaled Y outputs that were obtained from the neural network analysis for all peak models undergoing Quasi-Newton analysis (five perceptron layers) can be represented as:

$$\text{scaled_Y}_{AM} = (-0.154104 + (y_{4_1} * -1.67074))$$

$$\text{scaled_Y}_{Mid-Day} = (0.217368 + (y_{4_1} * 2.53726))$$

$$\text{scaled_Y}_{PM} = (0.870799 + (y_{4_1} * -0.468192) + (y_{4_2} * 2.5343))$$

The values of y_{x_x} for the AM, Mid-Day, and PM Peak periods are presented in Appendix F.

Levenberg-Marquardt Algorithm

The Levenberg-Marquardt (LM) algorithm was also used to train all three peak period matrices in Neural Designer. The algorithm utilizes the damped least-squares method where the loss functions take the form of sum of squared errors. The LM algorithm can be useful to train data sets that consist of a few thousand instances and a few hundred parameters. Although the data used for the project have five parameters, the data set is considerably large and can be suitable for LM algorithm training. The optimization algorithm was used for training the data sets of all peak periods of the six bus routes. Each matrix was subjected to three perceptron layers for the analysis. The errors obtained during training using the LM method have been discussed.

The first and second perceptron layers of all peak periods used all independent input variables, while the third layer only used one of the five inputs. The activation functions of the first two

layers were set to hyperbolic tangent, while those of third layer of each peak period were set to linear. The sizes of all layers and the corresponding activation functions for all three peak periods incorporating data from six routes analyzed using the LM algorithm are presented in Appendix B.

The inputs used in the LM algorithm were also scaled using the automatic scaling method. The size of the scaling layer is 5 (number of inputs). Unscaling for the output layer was done using the minimum and maximum method. The unscaling values of minimum, maximum, mean, and standard deviation for the LM algorithm model were the same as the unscaling values used for Quasi-Newton method. The values used for scaling and unscaling have been provided in Appendix B.

The results of the LM optimization algorithm are shown in Table 28. Plots showing the training and selection errors in each iteration for the AM, Mid-Day, and PM Peak periods are presented in Appendix D.

Table 28: Levenberg-Marquardt Method Results for Three Perceptron Layers
(All Peak Periods)

	AM Peak Value	Mid-Day Peak Value	PM Peak Value
Initial Training Error	0.564	0.284	1.725
Final Training Error	0.248	0.133	0.317
Initial Selection Error	0.568	0.268	1.735
Final Selection Error	0.253	0.136	0.342
Epochs Number	4	6	9
Elapsed Time (min:sec)	00:00	00:00	00:00
Stopping Criterion	Maximum parameters increment norm	Maximum parameters increment norm	Maximum parameters increment norm

It can be observed from Table 28 that during the AM Peak period, for the initial training error value of 0.564, the final value after four epochs decreased to 0.248. The initial value of the selection error in the AM Peak period was 0.568 which decreased to 0.253 after four epochs. From Table 28 it can also be observed that for the Mid-Day Peak period, the initial training error of 0.284 decreased to 0.133 after six epochs. Similarly, after six epochs, the initial value of selection error for the Mid-Day Peak period decreased from 0.268 to 0.136. For the PM Peak period, initial training error decreased from 1.725 to 0.317 after nine epochs, and the initial selection error decreased from 1.735 to 0. From the tables, it appears that the PM Peak period had the highest initial and final training and selection errors.

Order selection was performed to achieve the best selection model that generated an adequate fit for all the peak periods. The incremental order selection process was used to obtain the optimal order, training, and selection errors. Table 29 shows the order selection results from the incremental order algorithm for Levenberg-Marquardt Method for all three peak periods.

Table 29: Levenberg-Marquardt Method Results (All Peak Periods)

	AM Peak Value	Mid-Day Peak Value	PM Peak Value
Optimal Order	4	1	5
Optimal Training Error	0.207	0.122	0.253
Optimal Selection Error	0.211	0.123	0.275
Iterations Number	8	10	10
Elapsed Time (min:sec)	00:01	00:01	00:02

Following the order selection process, another training was conducted to achieve better neural network performance of the LM model. Table 30 shows the results of the second training for the model for all peak periods. Plots showing the training and selection errors in each iteration after order selection for the AM, Mid-Day, and PM Peak periods are presented in Appendix D.

Table 30: Levenberg-Marquardt Method Results of All Peak Periods (After Order Selection)

	AM Peak Value	Mid-Day Peak Value	PM Peak Value
Initial Training Error	0.359	0.175	0.291
Final Training Error	0.190	0.121	0.252
Initial Selection Error	0.345	0.165	0.315
Final Selection Error	0.201	0.122	0.279
Epochs Number	5	3	3
Elapsed Time (min:sec)	00:00	00:00	00:00
Stopping Criterion	Minimum parameters increment norm	Minimum parameters increment norm	Minimum parameters increment norm

It can be observed from Table 30 that for the AM Peak period, the initial training error value of 0.359 decreased to 0.190 after five epochs. The initial value of the selection error for the AM Peak period was 0.345, and the final value after five epochs decreased to 0.201. From Table 30 it can also be observed that for the Mid-Day Peak period, the initial training error of 0.175 decreased to 0.121 after three epochs. Similarly, after three epochs, the initial value of selection error for the Mid-Day Peak period decreased from 0.165 to 0.122. For the PM Peak period, initial training error decreased from 0.291 to 0.252 after three epochs, and the initial selection error decreased from 0.315 to 0.279. It can be observed that the order selection reduced the initial and final values of training and selection errors for all three peak periods.

The normalized squared errors were measured for all the instances to evaluate the model for each instance. Table 31 presents the normalized squared errors for the training, selection, and testing instances that were obtained from the model analyzed using the Levenberg-Marquardt algorithm in Neural Designer. Other errors, including mean squared errors, sum of squared errors, and others, are provided in Appendix E.

Table 31: Errors for Training, Selection and Testing Instances
(Three-Layer Levenberg-Marquardt)

	AM Peak Error	Mid-Day Peak Error	PM Peak Error
Training Error	0.190	0.121	0.252
Selection Error	0.201	0.122	0.279
Testing Error	0.207	0.137	0.275

It can be observed from Table 31 that the lowest testing error was obtained for the Mid-Day Peak model. The error statistics highlighting errors between the neural network and the testing instances in the data set for all peak periods are presented in Table 32. The minimum, maximum, and standard deviation errors are provided in Appendix E.

Table 32: Error Statistics for Levenberg-Marquardt Method for All Peak Periods
(Three Perceptron Layers)

	AM Peak Error	Mid-Day Peak Error	PM Peak Error
Absolute Error	109.063	83.776	143.402
Percentage Error	5.519	4.494	5.235

The mathematical expressions for the scaled Y outputs that were obtained from the neural network analysis for all peak models undergoing Levenberg-Marquardt analysis (five perceptron layers) can be represented as:

$$\text{scaled_Y}_{AM} = (-0.746019 + (y_{2_1} * 1.0558) + (y_{2_2} * 0.491962) + (y_{2_3} * 0.691286) + (y_{2_4} * -0.0592317))$$

$$\text{scaled_Y}_{Mid-Day} = (-0.28863 + (y_{1_1} * 1.33105))$$

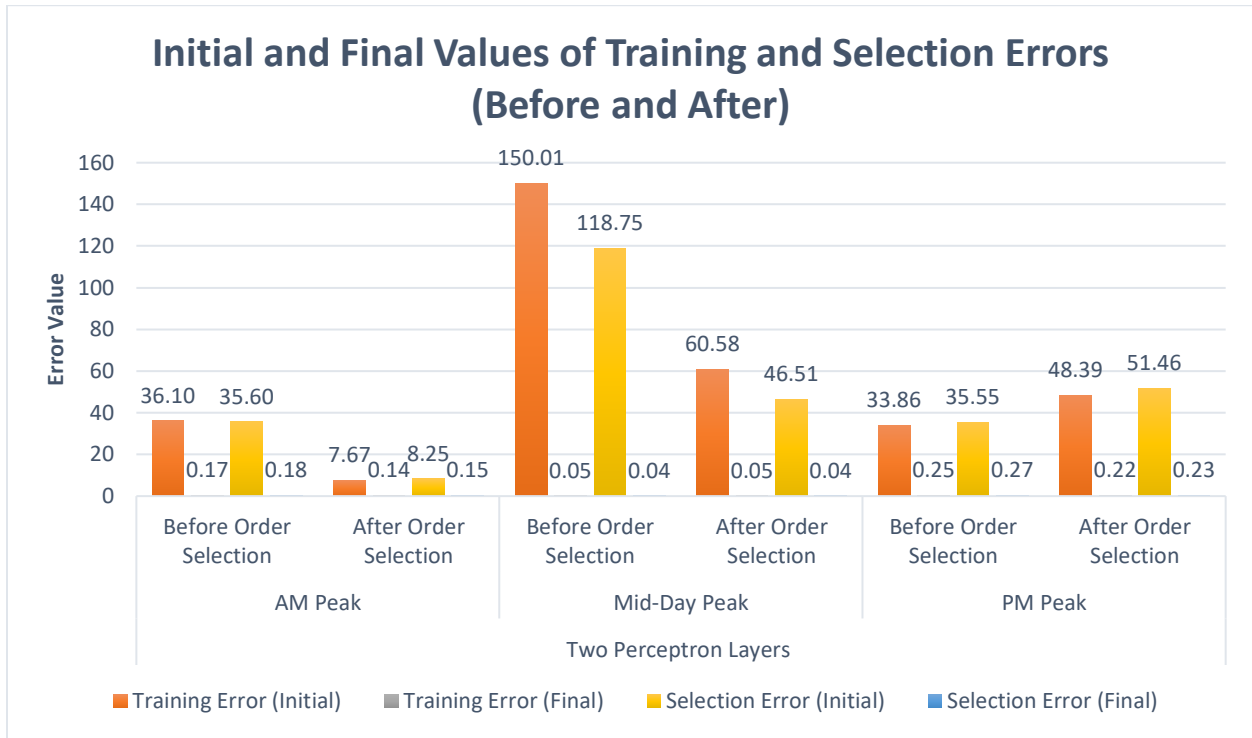
$$\text{scaled_Y}_{PM} = (-0.541352 + (y_{1_1} * 0.101253) + (y_{1_2} * 0.209238) + (y_{1_3} * 0.741562) + (y_{1_4} * 0.183389) + (y_{1_5} * -0.0125264))$$

The values of y_{x_x} for the AM, Mid-Day, and PM Peak periods are presented in Appendix F.

4.5 Comparison of Errors and Error Statistics

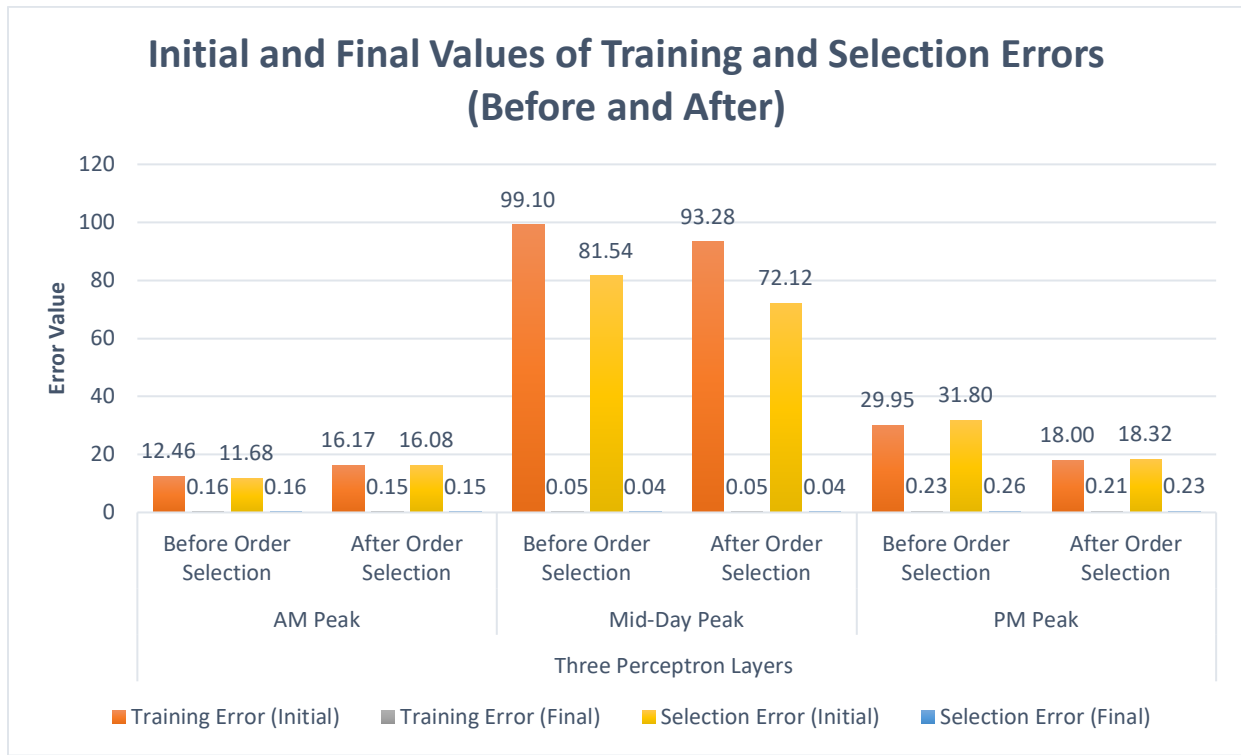
In training a model to predict data from the given inputs, the performance can be assessed by comparing the predicted values of the model with the true values of the output. Training error can be found when the model is applied to the data used for training, and selection error can be obtained by providing the model with unseen data (data not used in the training phase). Training and selection errors were calculated for all peak periods undergoing the Quasi-Newton and Levenberg-Marquardt optimization algorithms. Figures 25 through 28 present the comparison of the initial and final training and selection errors (before and after undergoing order selection) for the different optimization techniques used.

Figure 25: Training and Selection Errors Obtained Before and After Order Selection (Quasi-Newton Method with Two Perceptron Layers)



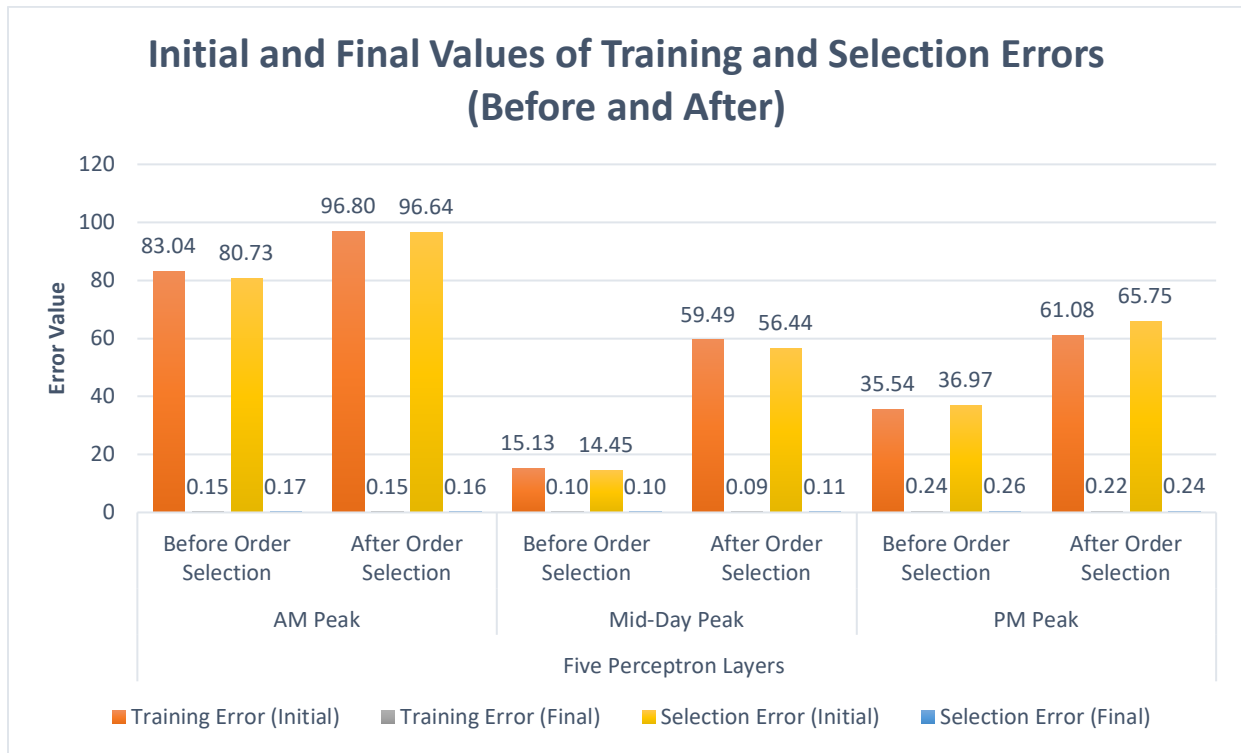
It can be observed from Figure 25 that order selection reduced the initial training and selection errors for the AM and Mid-Day Peak periods. However, after order selection and training, the final training and selection errors across all peak periods were reduced for the models with two perceptron layers undergoing the Quasi-Newton method. The highest final training and selection errors were observed for the PM Peak model while the Mid-Day Peak model had the lowest final training and selection errors. The selection error for the Mid-Day Peak period before order selection had the greatest change in error (99.97%), while the lowest change in error (98.2%) was obtained for the selection error of the AM Peak model after order selection.

Figure 26: Training and Selection Errors Obtained Before and After Order Selection (Quasi-Newton Method with Three Perceptron Layers)



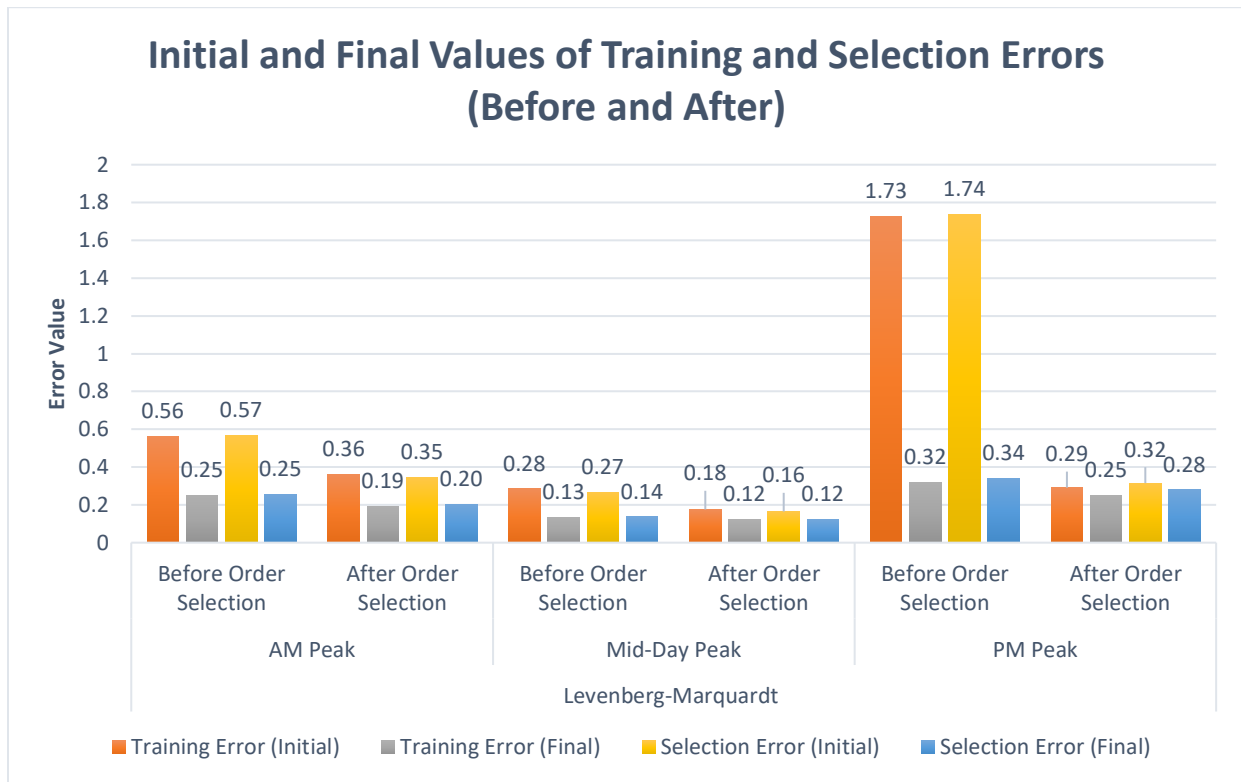
It can be observed from Figure 26 that order selection reduced the initial training and selection errors for the Mid-Day and PM Peak periods. However, after order selection and training, the final training and selection errors across all peak periods decreased for the models with three perceptron layers undergoing the Quasi-Newton method. The highest final training and selection errors were observed for the PM Peak period, while the Mid-Day Peak period had the lowest final training and selection errors. The selection error for the Mid-Day Peak period before order selection had the greatest change in error (99.95%), while the lowest change in error (98.62%) was obtained for the selection error of the AM Peak period before order selection.

Figure 27: Training and Selection Errors Obtained Before and After Order Selection (Quasi-Newton Method with Five Perceptron Layers)



It can be observed from Figure 27 that order selection reduced the initial training and selection errors for the AM and Mid-Day Peak periods. The final training and selection errors across all peak periods were greater than errors obtained before order selection for all the models with five perceptron layers undergoing the Quasi-Newton method. The highest final training and selection errors were observed in the AM Peak model, while the Mid-Day Peak period had the lowest final training and selection errors. The training error for the AM Peak period after order selection had the highest change in error (99.85%), while the lowest change in error (99.29%) was obtained for the selection error of the Mid-Day Peak period before order selection.

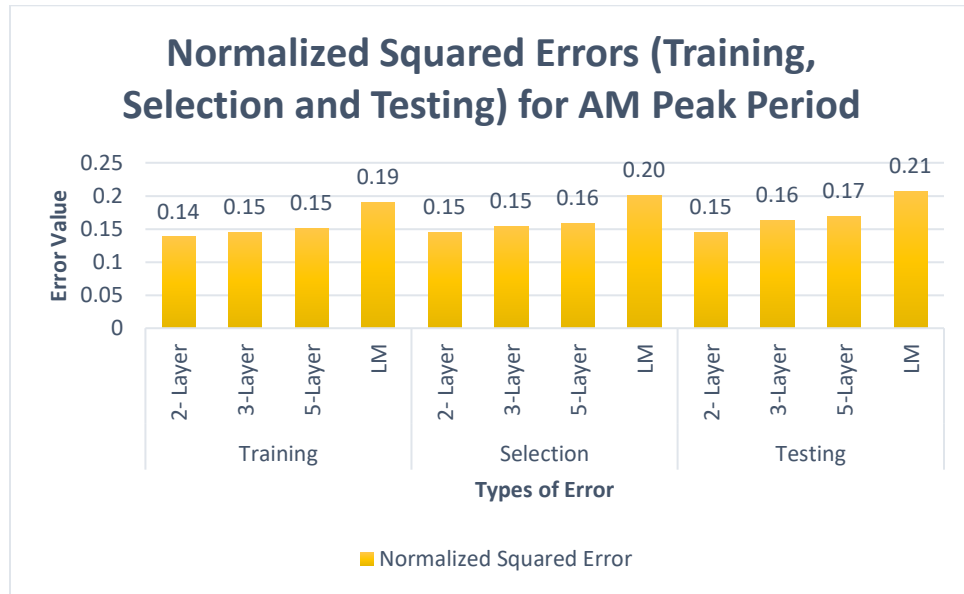
Figure 28: Training and Selection Errors Obtained Before and After Order Selection (Levenberg-Marquardt Method)



It can be observed from Figure 28 that order selection reduced the initial training and selection errors for all the peak periods. The final training and selection errors across all peak periods were reduced after order selection for all the models undergoing the Levenberg-Marquardt method. The highest final training and selection errors were observed for the PM Peak model, while the Mid-Day Peak model had the lowest final training and selection errors. The training error for the PM Peak model before order selection had the highest change in error (81.62%), whereas the lowest change in error (11.43%) was obtained for the selection error of the PM Peak model after order selection.

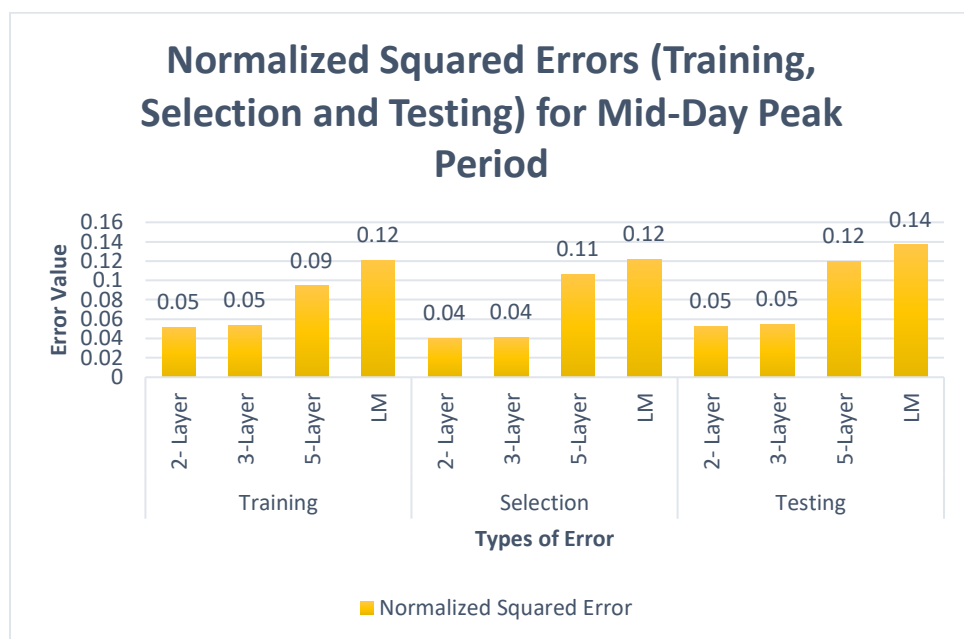
Normalized Squared Errors (NSEs) were obtained for all models to compare and assess the error metric associated with the neural network predictions. The NSE values obtained for the training, selection, and testing errors for all peak periods are presented in Figures 29 through 31.

Figure 29: Normalized Squared Errors Obtained for AM Peak Period Neural Network Analyses



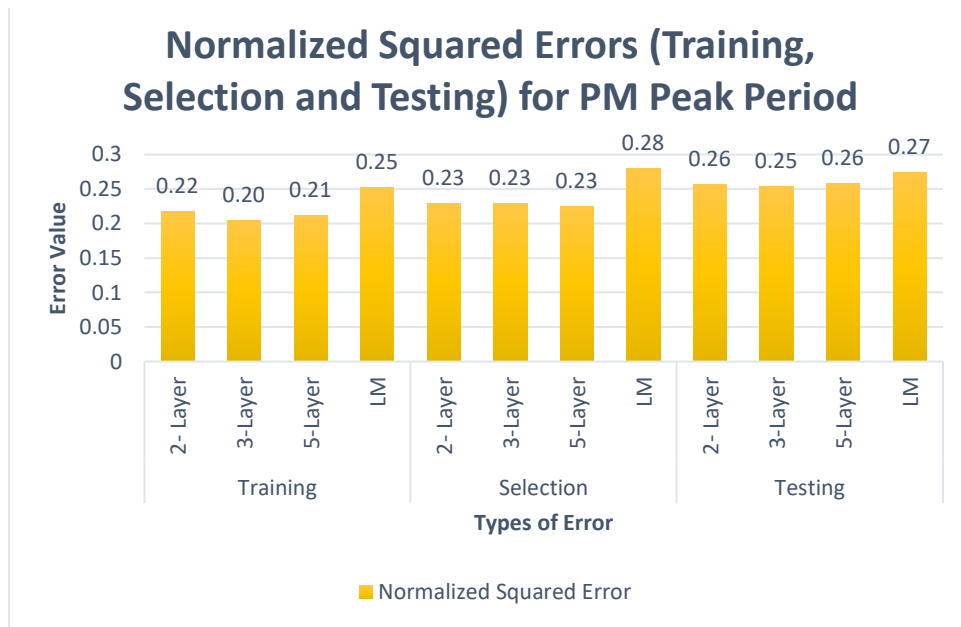
From Figure 29, it can be observed that the highest NSE for the training, selection, and testing errors of the AM Peak model were obtained using the Levenberg-Marquardt method. The models with two perceptron layers undergoing the Quasi-Newton method for the AM Peak period produced the lowest NSE for training, selection, and testing sets. Moreover, the lowest testing error (0.145) for the AM Peak period was obtained for the model with two perceptron layers undergoing the Quasi-Newton method.

Figure 30: Normalized Squared Errors Obtained for Mid-Day Peak Period Neural Network Analyses



The highest NSE values for the training, selection, and testing errors for the Mid-Day Peak period were obtained for the Levenberg-Marquardt method, as shown in Figure 30. The models with two perceptron layers undergoing the Quasi-Newton method for the Mid-Day Peak model produced the lowest NSE for training, selection, and testing sets. Moreover, the lowest testing error (0.053) for the Mid-Day Peak period was obtained for the model with two perceptron layers undergoing the Quasi-Newton method.

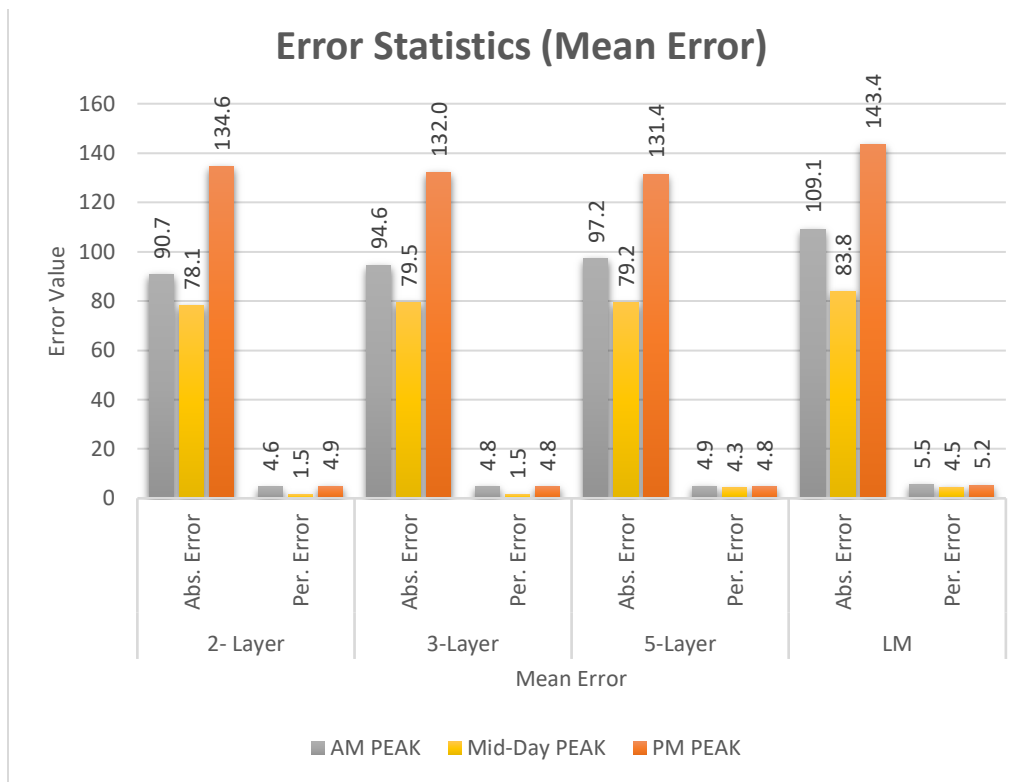
Figure 31: Normalized Squared Errors Obtained for PM Peak Period Neural Network Analyses



From Figure 31, it can be seen that the highest NSE for the training, selection, and testing errors of PM Peak period were obtained for the Levenberg-Marquardt method. The models with three perceptron layers undergoing the Quasi-Newton method for the PM Peak period produced the lowest NSE for training and testing sets. The lowest selection error was obtained for the model with five perceptron layers (Quasi-Newton method). The lowest testing error (0.254) for the PM Peak period was obtained for the model with three perceptron layers undergoing the Quasi-Newton method.

The Mean Absolute Errors (MAE) and Mean Percentage Errors (MPE) were also obtained for the models during the neural network analyses. MAE and MPE are (respectively) a measure of errors and the computed average of percentage errors between paired observations, and these measures can be used to determine a model's accuracy. Figure 32 represents the MAE and MPE for all three peak period models.

Figure 32: Mean Absolute and Mean Percentage Errors Obtained after Neural Network Analyses



It can be observed from Figure 32 that the lowest Mean Absolute Errors for the AM and Mid-Day Peak periods were obtained in the Quasi-Newton model with two layers, whereas the lowest MAE for the PM Peak period was obtained in the Quasi-Newton model with five layers. The Levenberg-Marquardt model had the highest MAE values for all the three peak periods.

Root Mean Squared Error or Root Mean Squared Deviation is a common way of measuring the quality of the fit of a model. That is, if the predicted values are close to the observed/actual values, the RMSE will be small; if there are significant differences between the observed and the predicted values, the RMSE will be large. Thus, an RMSE value of zero (0) would indicate a perfect fit with about 68% of the observed values within one RMSE if the data are normally distributed. However, Normalized Squared Error is preferable when comparing different models based on the same actual values. To compare model fits of different variables, a standardization is required to ensure some level of accuracy. Calculating the Normalized Mean Squared Errors of the variables provides a way of standardizing them to allow effective comparison.⁸²

The Mean Absolute Error parameter measures the difference between two continuous variables. The prediction error, that is, the difference between the observed and predicted value, is converted to positive to give the absolute error.⁸³ The mean (average sum) of all recorded absolute errors (MAE) gives an idea of the average error one can expect from the prediction model. However, the MAE does not provide the relative size of the error, especially when comparing several models.

The Mean Percentage Error (MPE) allows for the computation of the mean error in percentage form. This determines how large or small an error is and provides a better means of comparing the various models.⁸⁴

V. Discussion

This research aimed at developing neural network models for predicting the travel times of buses in Washington, DC using AVL and APC data. Previous literature showed that the travel times of the buses can be affected by time of the day (AM, Mid-Day and PM Peak periods), vehicle arrival and departure, passenger boarding and/or alighting, speed, distance, and en-route traffic conditions, among other factors.

Six months of data were obtained for six selected bus routes (70, 32, 52, 42, D4, and S1) from WMATA. The data obtained from WMATA were filtered, which enabled the extraction of variables including number of served bus stops within the study limits, length of routes between bus stops, average number of passengers in the bus, average dwell time, and number of intersections between served bus stops for three peak periods (AM, Mid-Day, and PM). The travel times associated with individual data points or factors following the extraction were compiled into matrices (per peak period) which were analyzed using multiple linear regression and neural network analyses.

The regression analysis conducted on the bus data generated an equation for each peak period. The independent variables “numbers of bus stops served,” “length of the route between served bus stops,” and “number of intersections between two served bus stops” had the highest correlation values for the AM, Mid-Day, and PM Peak periods, respectively. The F -statistics showed that the models were statistically significant at a 95% confidence interval (p -values < 0.05). To verify the validity of the models, normality of errors, multicollinearity, and homoscedasticity tests were conducted. The observed cumulative probabilities of the standardized residuals were plotted against the expected cumulative probabilities of the standardized residuals, which generated curves that followed the diagonals of the different peak plots. Hence, the errors were normally distributed. Normal distribution of errors ensures that there is no bias in the data. In other words, there are not many outliers that can be affecting the model’s estimation process. Multicollinearity was checked for all models by assessing the values of Variance Inflation Factors (VIFs) and tolerances to ensure that no two independent predictor variables used in this study had high correlations between themselves. Multicollinearity indicates the presence of redundant information which skews the results in a predictive model. All three models had VIFs lower than the maximum value of 10, which indicated that the independent variables were not multicollinear. A VIF value closer to 1 indicates that the predictors are not correlated with each other. The even distribution of the residual plots along the zero line for all three models also indicated homoscedasticity. Hence, the variance between the residuals of the dependent variable were similar across the models.

Neural Network Analyses were also conducted for different peak periods to develop predictive models for the travel time using Neural Designer software. The method used was approximation, which refers to prediction of a dependent value based on the combination of different independent variables. The Quasi-Newton and Levenberg-Marquardt algorithms were used separately for all

peak periods (AM, Mid-Day and PM Peak periods) to obtain separate bus travel time equations. The Quasi-Newton algorithm is the default optimization method in Neural Designer which is also recommended for training medium-sized data sets (10–1,000 variables). Using the Quasi-Newton algorithm, the study also aimed at exploring the complexity of the training network that would affect the predictive models. Hence, three Quasi-Newton algorithms were used for three peak periods where the number of perceptron layers varied (two, three, and five layers separately). In addition to the Quasi-Newton algorithm, the Levenberg-Marquardt algorithm was also used for all three models to compare the differences between two different approximation algorithms. All peak period models were subjected to three perceptron layers for the Levenberg-Marquardt analyses.

The neural network analyses required scaling of the inputs for training, after which the outputs were unscaled back to the original units. Equations to determine the travel time based on the values of the independent variables were provided for all the three peak periods using all the algorithms. A major part of the training process was the documentation of initial and final training, selection, and testing errors before and after order selection. In approximation training processes, the training errors gradually decrease over time, while the testing errors increase. It could be seen that generally, the initial selection and training errors for all the peak periods decreased after conducting the order selection process. Hence, it could be seen that order selection increased the accuracy of the training models whereby final values of training and selection errors across most of the peak periods were reduced.

The training error may not be a good validation of the predictive model, as lower training error results in overfitting. Overfitting leads to inaccuracy in predicting the correct output from unseen data while testing. Hence, testing error was analyzed for all the models to evaluate the accuracy of approximation. The minimums, maximums, means, and standard deviations of the error statistics were also obtained. The mean absolute error and the mean percentage error were compared to test the quality of the predictive models. Both measures of errors were obtained to observe the difference between the predicted output and the actual output values. The low MAE and MPE errors obtained for the different models in this study indicate better accuracy of the continuous variables.

Even though lower errors are desirable, low training error indicates to overfitting, which leads to inaccuracy in predicting the correct output from unseen data while testing. Hence, testing errors were for all the models were compared to evaluate the accuracy of approximation of the neural networks. In this study, the models that were trained using Quasi-Newton Algorithm with 2 perceptron layers generally yielded the lowest normalized squared errors (testing) for all peak periods which were followed by models that had 3 perceptron layers applied. In other words, Quasi-Newton optimization using 2 perceptron layers (hidden Multilayer Perceptron) demonstrated the best travel time prediction performance in this study. It could be observed that the two-layer models trained using Quasi-Newton algorithm had the lowest normalized squared errors (testing) followed by training models having three perceptron layers. Mid-Day Peak models

had the lowest overall errors for both training algorithms. The Levenberg-Marquardt algorithm resulted in the highest testing errors, which was an indication that it was not suitable for the data sets.

VI. Conclusions and Recommendations

This research developed neural network models which can be helpful for transit agencies to improve travel time prediction for their patrons. General improvement in the credibility of WMATA's online NextBus Arrival service can increase ridership. Moreover, the prediction models can serve as an excellent tool to build schedules for new bus routes in or around Washington, DC. It is recommended that neural network training including more factors (independent variables) be conducted to study the impact of these variables on bus travel time. Models can also be developed to analyze travel times by bus direction for a route. The research focused on specific peak periods of the day to ensure that traffic patterns differ by different times of the day. However, a single model with low errors can be developed in future research that can predict travel time on any given time of the day.

The results of the analyses indicate that ANN models can effectively predict the travel times of buses on selected routes with minimal percentage errors and can be used in combination with traditional regression analyses. The ANN models could be incorporated into several other predictive models used by WMATA to provide patrons with travel time information at bus stops and online. These models could be adopted by transit agencies in other jurisdictions with similar characteristics to those of the Washington, DC area. For future work, these models could be calibrated using real-time data for arterial and collector bus routes. Also, similar models could be developed for bus routes on residential/local roads.

Endnotes

- ¹ “Sex of Workers By Means of Transportation To Work,” accessed September 25, 2020, <https://data.census.gov/cedsci/table?q=B08&d=ACS%201-Year%20Estimates%20Detailed%20Tables&tid=ACSDT1Y2018.B08006&hidePreview=false>.
- ² “WMATA. Metro Facts, 2017,” July 2, 2018, <https://web.archive.org/web/20180702201723/https://www.wmata.com/about/upload/Metro-Facts-2017-FINAL.pdf>.
- ³ Stephen Arhin, Errol C. Noel, and Mineta National Transit Research Consortium, “Evaluation of Bus Transit Reliability in the District of Columbia.” (Mineta Transportation Institute, 2013).
- ⁴ R. Jeong and R. Rilett, “Bus Arrival Time Prediction Using Artificial Neural Network Model,” in Proceedings: The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No.04TH8749), 2004, 988–93, <https://doi.org/10.1109/ITSC.2004.1399041>.
- ⁵ “The History of Public Bus Transportation,” Blog (blog), August 14, 2017, <https://gogocharters.com/blog/history-of-public-bus-transportation/>.
- ⁶ Alan Black, *Urban Mass Transportation Planning* (New York: McGraw-Hill, 1995).
- ⁷ “Public Transportation in the United States,” in Wikipedia, February 6, 2020, https://en.wikipedia.org/w/index.php?title=Public_transportation_in_the_United_States&oldid=939382661.
- ⁸ APTAAdmin, “Public Transportation Facts,” American Public Transportation Association (blog), accessed March 11, 2020, <https://www.apta.com/news-publications/public-transportation-facts/>.
- ⁹ “Metro System,” accessed March 13, 2020, <https://www.commuterpage.com/ways-to-get-around/metro-system/>.
- ¹⁰ “Metro System.”
- ¹¹ “How Do I Get Around Washington, DC?,” Washington.org, March 24, 2016, <https://washington.org/dc-guide-to/how-do-i-get-around-washington-dc>.
- ¹² APTAAdmin, “Public Transportation Facts.”
- ¹³ APTAAdmin, “Public Transportation Facts.”

- ¹⁴ “9 Benefits of Public Transportation,” accessed March 27, 2020, <https://www.nationalexpresstransit.com/blog/9-benefits-of-public-transportation/>.
- ¹⁵ APTAAdmin, “Public Transportation Facts.”
- ¹⁶ “9 Benefits of Public Transportation.”
- ¹⁷ APTAAdmin, “Public Transportation Facts.”
- ¹⁸ “9 Benefits of Public Transportation.”
- ¹⁹ APTAAdmin, “Public Transportation Facts.”
- ²⁰ “9 Benefits of Public Transportation.”
- ²¹ “Characteristics of Bus Rapid Transport for Decision Making,” April 15, 2016, https://web.archive.org/web/20160415111548/http://www.nbrti.org/docs/pdf/Characteristics_BRT_Decision-Making.pdf.
- ²² APTAAdmin, “Public Transportation Facts.”
- ²³ “5 Transportation Challenges in Urban Areas,” accessed April 16, 2020, <https://www.nationalexpresstransit.com/blog/5-transportation-challenges-in-urban-areas/>.
- ²⁴ Zheng Li and David Hensher, “Crowding in Public Transport: A Review of Objective and Subjective Measures,” *Journal of Public Transportation* 16, no. 2 (June 2013): 107–34, <https://doi.org/10.5038/2375-0901.16.2.6>.
- ²⁵ “Evaluation of Bus Transit Reliability in the District of Columbia - Dot_26843_DS1.Pdf,” accessed April 17, 2020, https://rosap.ntl.bts.gov/pdfjs/web/viewer.html?file=https://rosap.ntl.bts.gov/view/dot/26843/dot_26843_DS1.pdf#page=1&zoom=auto,592,798.
- ²⁶ “DDOT, Automated Enforcement Final Report_Final_mm.Pdf,” accessed April 18, 2020, http://d92016.eos-intl.net/eLibSQL14_D92016_Documents/Automated%20Enforcement%20Final%20Report_Final_mm.pdf.
- ²⁷ Miquel Estrada et al., “Bus Control Strategies in Corridors with Signalized Intersections,” *Transportation Research Part C: Emerging Technologies* 71 (October 1, 2016): 500–520, <https://doi.org/10.1016/j.trc.2016.08.013>.
- ²⁸ Benjamin Schneider, “The High Cost of Global Traffic Jams,” *CityLab*, accessed April 17, 2020, <https://www.citylab.com/transportation/2018/02/traffics-mind-boggling-economic-toll/552488/>.

- ²⁹ Eric Albright and Miguel Figliozzi, “Factors Influencing Effectiveness of Transit Signal Priority and Late-Bus Recovery at Signalized-Intersection Level,” *World Transit Research*, January 1, 2012, <https://www.worldtransitresearch.info/research/4821>.
- ³⁰ Ehsan Mazloumi, Graham Currie, and Geoffrey Rose, “Using GPS Data to Gain Insight into Public Transport Travel Time Variability,” *Journal of Transportation Engineering* 136, no. 7 (July 1, 2010): 623–31, [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000126](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000126).
- ³¹ Miguel Figliozzi et al., “A Study of Headway Maintenance for Bus Routes: Causes and Effects of ‘Bus Bunching’ in Extensive and Congested Service Areas,” *Civil and Environmental Engineering Faculty Publications and Presentations*, July 1, 2012, <https://doi.org/10.15760/trec.107>.
- ³² Ehsan Mazloumi, Graham Currie, and Geoffrey Rose, “Using GPS Data to Gain Insight into Public Transport Travel Time Variability.”
- ³³ Andrew Owen and David M. Levinson, “Modeling the Commute Mode Share of Transit Using Continuous Accessibility to Jobs,” *Transportation Research Part A: Policy and Practice* 74 (April 1, 2015): 110–22, <https://doi.org/10.1016/j.tra.2015.02.002>.
- ³⁴ Nir Sharaby and Yoram Shiftan, “The Impact of Fare Integration on Travel Behavior and Transit Ridership,” *Transport Policy* 21 (May 1, 2012): 63–70, <https://doi.org/10.1016/j.tranpol.2012.01.015>.
- ³⁵ “ISSUES IN BUS RAPID TRANSIT,” n.d., 29.
- ³⁶ “ISSUES IN BUS RAPID TRANSIT.”
- ³⁷ Estrada et al., “Bus Control Strategies in Corridors with Signalized Intersections.”
- ³⁸ “ISSUES IN BUS RAPID TRANSIT.”
- ³⁹ Henry Britton Ogm et al., *Contents*, 2003.
- ⁴⁰ Sharaby and Shiftan, “The Impact of Fare Integration on Travel Behavior and Transit Ridership.”
- ⁴¹ Tingting Yin et al., “A Prediction Model of Bus Arrival Time at Stops with Multi-Routes,” *Transportation Research Procedia, World Conference on Transport Research, Shanghai, 10–15 July 2016*, 25 (January 1, 2017): 4623–36, <https://doi.org/10.1016/j.trpro.2017.05.381>.
- ⁴² Matthew P. D’Angelo, Haitham M. Al-Deek, and Morgan C. Wang, “Travel-Time Prediction for Freeway Corridors,” *Transportation Research Record* 1676, no. 1 (January 1, 1999): 184–91, <https://doi.org/10.3141/1676-23>.

- ⁴³ Jayakrishna Patnaik, Steven Chien, and Athanassios Bladikas, “Estimation of Bus Arrival Times Using APC Data,” *Journal of Public Transportation* 7, no. 1 (March 1, 2004), <https://doi.org/10.5038/2375-0901.7.1.1>.
- ⁴⁴ Amer Shalaby et al., *Bus Travel Time Prediction Model for Dynamic Operations Control and Passenger Information Systems*, 2002.
- ⁴⁵ L. Vanajakshi, S. C. Subramanian, and R. Sivanandan, “Travel Time Prediction under Heterogeneous Traffic Conditions Using Global Positioning System Data from Buses,” *IET Intelligent Transport Systems* 3, no. 1 (March 1, 2009): 1–9, <https://doi.org/10.1049/iet-its:20080013>.
- ⁴⁶ Yin et al., “A Prediction Model of Bus Arrival Time at Stops with Multi-Routes.”
- ⁴⁷ Vivek Kumar et al., “Comparison of Model Based and Machine Learning Approaches for Bus Arrival Time Prediction,” 2014, <https://trid.trb.org/view/1288350>.
- ⁴⁸ Wei Fan and Zegeye Gurmu, “Dynamic Travel Time Prediction Models for Buses Using Only GPS Data,” *International Journal of Transportation Science and Technology* 4, no. 4 (January 1, 2015): 353–66, [https://doi.org/10.1016/S2046-0430\(16\)30168-X](https://doi.org/10.1016/S2046-0430(16)30168-X).
- ⁴⁹ Shalaby et al., *Bus Travel Time Prediction Model for Dynamic Operations Control and Passenger Information Systems*.
- ⁵⁰ Jayakrishna Patnaik, Steven Chien, and Athanassios Bladikas, “Estimation of Bus Arrival Times Using APC Data.”
- ⁵¹ Mei Chen et al., “A Dynamic Bus-Arrival Time Prediction Model Based on APC Data,” *Computer-Aided Civil and Infrastructure Engineering* 19, no. 5 (2004): 364–76, <https://doi.org/10.1111/j.1467-8667.2004.00363.x>.
- ⁵² Vivek Kumar et al., “Comparison of Model Based and Machine Learning Approaches for Bus Arrival Time Prediction.”
- ⁵³ Prakash Ranjitkar et al., “Bus Arrival Time Modeling Based on Auckland Data:,” *Transportation Research Record*, April 16, 2019, <https://doi.org/10.1177/0361198119840620>.
- ⁵⁴ Fan and Gurmu, “Dynamic Travel Time Prediction Models for Buses Using Only GPS Data.”
- ⁵⁵ Jeong and Rilett, “Bus Arrival Time Prediction Using Artificial Neural Network Model.”
- ⁵⁶ Farhan Ali, Shalaby Amer, and Sayed Tarek, “Bus Travel Time Prediction Using AVL and APC,” *Applications of Advanced Technologies in Transportation* (2002), Proceedings, n.d., 616–23, [https://doi.org/10.1061/40632\(245\)78](https://doi.org/10.1061/40632(245)78).
- ⁵⁷ Fan and Gurmu, “Dynamic Travel Time Prediction Models for Buses Using Only GPS Data.”

- ⁵⁸ Jayakrishna Patnaik, Steven Chien, and Athanassios Bladikas, “Estimation of Bus Arrival Times Using APC Data.”
- ⁵⁹ Y. Ramakrishna et al., “Bus Travel Time Prediction Using GPS Data,” Proceedings Map India, 2006.
- ⁶⁰ Fan and Gurmu, “Dynamic Travel Time Prediction Models for Buses Using Only GPS Data.”
- ⁶¹ Goutam Dutta et al., “Artificial Neural Network Models for Forecasting Stock Price Index in the Bombay Stock Exchange,” *Journal of Emerging Market Finance* 5, no. 3 (December 1, 2006): 283–95, <https://doi.org/10.1177/097265270600500305>.
- ⁶² “Fundamentals of Artificial Neural Networks,” accessed April 20, 2020, https://neuron.eng.wayne.edu/tarek/MITbook/t_contents.html.
- ⁶³ Jeong and Rilett, “Bus Arrival Time Prediction Using Artificial Neural Network Model.”
- ⁶⁴ Jayesh Bapu Ahire, “The Artificial Neural Networks Handbook: Part 1,” Medium, September 1, 2020, <https://medium.com/coinmonks/the-artificial-neural-networks-handbook-part-1-f9ceb0e376b4>.
- ⁶⁵ Johar Amita, Jain Sukhvir Singh, and Garg Pradeep Kumar, “PREDICTION OF BUS TRAVEL TIME USING ARTIFICIAL NEURAL NETWORK,” *INTERNATIONAL JOURNAL FOR TRAFFIC AND TRANSPORT ENGINEERING* 5, no. 4 (December 2015): 410–24, [https://doi.org/10.7708/ijtte.2015.5\(4\).06](https://doi.org/10.7708/ijtte.2015.5(4).06).
- ⁶⁶ Sagar Sharma, “Activation Functions in Neural Networks,” Medium, February 14, 2019, <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>.
- ⁶⁷ “Artificial Neural Networks/Print Version,” Wikibooks, Open Books for an Open World, accessed April 20, 2020, https://en.wikibooks.org/wiki/Artificial_Neural_Networks/Print_Version.
- ⁶⁸ Mark Seidenberg, “Figure 2. Activation Curve for Hyperbolic Tangent Function Used In...,” ResearchGate, accessed September 28, 2020, https://www.researchgate.net/figure/Activation-curve-for-hyperbolic-tangent-function-used-in-simulations-y-tanh-x-2_fig2_2439950.
- ⁶⁹ Akanksha Rai, “Activation Functions in Neural Networks,” GeeksforGeeks (blog), January 29, 2018, <https://www.geeksforgeeks.org/activation-functions-neural-networks/>.
- ⁷⁰ Jeong and Rilett, “Bus Arrival Time Prediction Using Artificial Neural Network Model.”
- ⁷¹ Chien Steven I-Jy, Ding Yuqing, and Wei Chienhung, “Dynamic Bus Arrival Time Prediction with Artificial Neural Networks,” *Journal of Transportation Engineering* 128,

no. 5 (September 1, 2002): 429–38, [https://doi.org/10.1061/\(ASCE\)0733-947X\(2002\)128:5\(429\)](https://doi.org/10.1061/(ASCE)0733-947X(2002)128:5(429)).

- ⁷² “International Business | Dmped,” accessed July 13, 2020, <https://dmped.dc.gov/page/international-business>.
- ⁷³ “DC Ranks As One Of World’s Worst Traffic Cities: Report,” Washington DC, DC Patch, February 13, 2018, <https://patch.com/district-columbia/washingtondc/dc-ranks-one-worlds-worst-traffic-cities-report>.
- ⁷⁴ “DC Ward Map Overlay,” Google My Maps, accessed June 14, 2020, <https://www.google.com/maps/d/viewer?mid=1DpJu2DbYssvxH0XsK6f6iXV6BN4>.
- ⁷⁵ “WMATA_Metrobus Timetable for Route 70,” accessed June 18, 2020, https://www.wmata.com/schedules/timetables/upload/70_180624.pdf.
- ⁷⁶ “WMATA_Metrobus Timetable for Routes 32,34,36,” accessed June 18, 2020, https://www.wmata.com/schedules/timetables/upload/32,34,36_151219.pdf.
- ⁷⁷ “WMATA_Metrobus Timetable for Route 52-54,” accessed June 18, 2020, https://www.wmata.com/schedules/timetables/upload/52-54_151219.pdf.
- ⁷⁸ “WMATA_Metrobus Timetable for Route 42,43,” accessed June 18, 2020, https://www.wmata.com/schedules/timetables/upload/42,43_170924.pdf.
- ⁷⁹ “WMATA_Metrobus Timetable for Route D4,” accessed June 18, 2020, https://www.wmata.com/schedules/timetables/upload/D4_180624.pdf.
- ⁸⁰ “WMATA_Metrobus Timetable for Route S1,” accessed June 18, 2020, https://www.wmata.com/schedules/timetables/upload/S1_190623.pdf.
- ⁸¹ “5 Algorithms to Train a Neural Network,” accessed July 13, 2020, https://www.neuraldesigner.com/blog/5_algorithms_to_train_a_neural_network#Quasi-Newton.
- ⁸² S.A. Otto, “How to Normalize the RMSE [Blog Post],” accessed September 28, 2020, <https://www.marinedatascience.co/blog/2019/01/07/normalizing-the-rmse/>.
- ⁸³ E | M (Medium user @ewuramaminka), “Mean Absolute Error ~ MAE [Machine Learning(ML)],” Medium, February 22, 2018, <https://medium.com/@ewuramaminka/mean-absolute-error-mae-machine-learning-ml-b9b4afc63077>.
- ⁸⁴ Nate Watson, “Using Mean Absolute Error to Forecast Accuracy,” Contemporary Analysis (blog), January 23, 2012, <https://canworksmart.com/using-mean-absolute-error-forecast-accuracy/>.

Bibliography

- “5 Algorithms to Train a Neural Network.” Accessed July 13, 2020. https://www.neuraldesigner.com/blog/5_algorithms_to_train_a_neural_network#Quasi-Newton.
- “5 Transportation Challenges in Urban Areas.” Accessed April 16, 2020. <https://www.nationalexpresstransit.com/blog/5-transportation-challenges-in-urban-areas/>.
- “9 Benefits of Public Transportation.” Accessed March 27, 2020. <https://www.nationalexpresstransit.com/blog/9-benefits-of-public-transportation/>.
- Ahire, Jayesh Bapu. “The Artificial Neural Networks Handbook: Part 1.” Medium, September 1, 2020. <https://medium.com/coinmonks/the-artificial-neural-networks-handbook-part-1-f9ceb0e376b4>.
- Albright, Eric, and Miguel Figliozzi. “Factors Influencing Effectiveness of Transit Signal Priority and Late-Bus Recovery at Signalized-Intersection Level.” *World Transit Research*, January 1, 2012. <https://www.worldtransitresearch.info/research/4821>.
- Amita, Johar, Jain Sukhvir Singh, and Garg Pradeep Kumar. “PREDICTION OF BUS TRAVEL TIME USING ARTIFICIAL NEURAL NETWORK.” *INTERNATIONAL JOURNAL FOR TRAFFIC AND TRANSPORT ENGINEERING* 5, no. 4 (December 2015): 410–24. [https://doi.org/10.7708/ijtte.2015.5\(4\).06](https://doi.org/10.7708/ijtte.2015.5(4).06).
- APTAAAdmin. “Public Transportation Facts.” *American Public Transportation Association* (blog). Accessed March 11, 2020. <https://www.apta.com/news-publications/public-transportation-facts/>.
- Arhin, Stephen, Errol C. Noel, and Mineta National Transit Research Consortium. “Evaluation of Bus Transit Reliability in the District of Columbia.” Mineta Transportation Institute, 2013.
- “Artificial Neural Networks/Print Version - Wikibooks, Open Books for an Open World.” Accessed April 20, 2020. https://en.wikibooks.org/wiki/Artificial_Neural_Networks/Print_Version.
- Black, Alan. *Urban Mass Transportation Planning*. New York: McGraw-Hill, 1995.
- “Characteristics of Bus Rapid Transport for Decision Making,” April 15, 2016. https://web.archive.org/web/20160415111548/http://www.nbrti.org/docs/pdf/Characteristics_BRT_Decision-Making.pdf.
- Chen, Mei, Xiaobo Liu, Jingxin Xia, and Steven I. Chien. “A Dynamic Bus-Arrival Time Prediction Model Based on APC Data.” *Computer-Aided Civil and Infrastructure Engineering* 19, no. 5 (2004): 364–76. <https://doi.org/10.1111/j.1467-8667.2004.00363.x>.
- Chien Steven I-Jy, Ding Yuqing, and Wei Chienhung. “Dynamic Bus Arrival Time Prediction with Artificial Neural Networks.” *Journal of Transportation Engineering* 128, no. 5 (September 1, 2002): 429–38. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2002\)128:5\(429\)](https://doi.org/10.1061/(ASCE)0733-947X(2002)128:5(429)).

- D'Angelo, Matthew P., Haitham M. Al-Deek, and Morgan C. Wang. "Travel-Time Prediction for Freeway Corridors." *Transportation Research Record* 1676, no. 1 (January 1, 1999): 184–91. <https://doi.org/10.3141/1676-23>.
- Washington DC, DC Patch. "DC Ranks As One Of World's Worst Traffic Cities: Report," February 13, 2018. <https://patch.com/district-columbia/washingtondc/dc-ranks-one-worlds-worst-traffic-cities-report>.
- Google My Maps. "DC Ward Map Overlay." Accessed June 14, 2020. <https://www.google.com/maps/d/viewer?mid=1DpJu2DbYssvxH0XsK6f6iXV6BN4>.
- "DDOT, Automated Enforcement Final Report_Final_mm.Pdf." Accessed April 18, 2020. http://d92016.eos-intl.net/eLibSQL14_D92016_Documents/Automated%20Enforcement%20Final%20Report_Final_mm.pdf.
- Dutta, Goutam, Pankaj Jha, Arnab Kumar Laha, and Neeraj Mohan. "Artificial Neural Network Models for Forecasting Stock Price Index in the Bombay Stock Exchange." *Journal of Emerging Market Finance* 5, no. 3 (December 1, 2006): 283–95. <https://doi.org/10.1177/097265270600500305>.
- Estrada, Miquel, Josep Mensión, Josep M. Aymamí, and Laura Torres. "Bus Control Strategies in Corridors with Signalized Intersections." *Transportation Research Part C: Emerging Technologies* 71 (October 1, 2016): 500–520. <https://doi.org/10.1016/j.trc.2016.08.013>.
- "Evaluation of Bus Transit Reliability in the District of Columbia - Dot_26843_DS1.Pdf." Accessed April 17, 2020. https://rosap.ntl.bts.gov/pdfjs/web/viewer.html?file=https://rosap.ntl.bts.gov/view/dot/26843/dot_26843_DS1.pdf#page=1&zoom=auto,592,798.
- Fan, Wei, and Zegeye Gurmu. "Dynamic Travel Time Prediction Models for Buses Using Only GPS Data." *International Journal of Transportation Science and Technology* 4, no. 4 (January 1, 2015): 353–66. [https://doi.org/10.1016/S2046-0430\(16\)30168-X](https://doi.org/10.1016/S2046-0430(16)30168-X).
- Farhan Ali, Shalaby Amer, and Sayed Tarek. "Bus Travel Time Prediction Using AVL and APC." *Applications of Advanced Technologies in Transportation (2002)*, Proceedings, n.d., 616–23. [https://doi.org/10.1061/40632\(245\)78](https://doi.org/10.1061/40632(245)78).
- Figliozzi, Miguel, Wu-chi Feng, Gerardo Lafferriere, and Wei Feng. "A Study of Headway Maintenance for Bus Routes: Causes and Effects of 'Bus Bunching' in Extensive and Congested Service Areas." *Civil and Environmental Engineering Faculty Publications and Presentations*, July 1, 2012. <https://doi.org/10.15760/trec.107>.
- "Fundamentals of Artificial Neural Networks." Accessed April 20, 2020. https://neuron.eng.wayne.edu/tarek/MITbook/t_contents.html.
- Washington.org. "How Do I Get Around Washington, DC?," March 24, 2016. <https://washington.org/dc-guide-to/how-do-i-get-around-washington-dc>.
- "International Business | Dmped." Accessed July 13, 2020. <https://dmped.dc.gov/page/international-business>.
- "ISSUES IN BUS RAPID TRANSIT," n.d., 29.

- Jeong, R., and R. Rilett. “Bus Arrival Time Prediction Using Artificial Neural Network Model.” In *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No.04TH8749)*, 988–93, 2004. <https://doi.org/10.1109/ITSC.2004.1399041>.
- Kumar, Vivek, B. Anil Kumar, Lelitha Devi Vanajakshi, and Shankar C. Subramanian. “Comparison of Model Based and Machine Learning Approaches for Bus Arrival Time Prediction,” 2014. <https://trid.trb.org/view/1288350>.
- Li, Zheng, and David Hensher. “Crowding in Public Transport: A Review of Objective and Subjective Measures.” *Journal of Public Transportation* 16, no. 2 (June 2013): 107–34. <https://doi.org/10.5038/2375-0901.16.2.6>.
- M, E. |. “Mean Absolute Error ~ MAE [Machine Learning(ML)].” Medium, February 22, 2018. <https://medium.com/@ewuramaminka/mean-absolute-error-mae-machine-learning-ml-b9b4afc63077>.
- Mazloumi, Ehsan, Graham Currie, and Geoffrey Rose. “Using GPS Data to Gain Insight into Public Transport Travel Time Variability.” *Journal of Transportation Engineering* 136, no. 7 (July 1, 2010): 623–31. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000126](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000126).
- “Metro System.” Accessed March 13, 2020. <https://www.commuterpage.com/ways-to-get-around/metro-system/>.
- Ogm, Henry Britton, Pieter Hilferink Nea, Rosario Macario Tis Pt, Stefano Proietti Isis, Adriaan Roest, Crollius Nea, and Carlo Sessa Isis. *Contents*, 2003.
- Otto, S.A. “How to Normalize the RMSE [Blog Post].” Accessed September 28, 2020. <https://www.marinedatascience.co/blog/2019/01/07/normalizing-the-rmse/>.
- Owen, Andrew, and David M. Levinson. “Modeling the Commute Mode Share of Transit Using Continuous Accessibility to Jobs.” *Transportation Research Part A: Policy and Practice* 74 (April 1, 2015): 110–22. <https://doi.org/10.1016/j.tra.2015.02.002>.
- Patnaik, Jayakrishna, Steven Chien, and Athanassios Bladikas. “Estimation of Bus Arrival Times Using APC Data.” *Journal of Public Transportation* 7, no. 1 (March 1, 2004). <https://doi.org/10.5038/2375-0901.7.1.1>.
- “Public Transportation in the United States.” In *Wikipedia*, February 6, 2020. https://en.wikipedia.org/w/index.php?title=Public_transportation_in_the_United_States&oldid=939382661.
- Rai, Akanksha. “Activation Functions in Neural Networks.” *GeeksforGeeks* (blog), January 29, 2018. <https://www.geeksforgeeks.org/activation-functions-neural-networks/>.
- Ramakrishna, Y., P. Ramakrishna, V. Lakshmanan, and R. Sivanandan. “Bus Travel Time Prediction Using GPS Data.” *Proceedings Map India*, 2006.
- Ranjitkar, Prakash, Li-Sian Tey, Enakshi Chakravorty, and Kirsten L. Hurley. “Bus Arrival Time Modeling Based on Auckland Data.” *Transportation Research Record*, April 16, 2019. <https://doi.org/10.1177/0361198119840620>.
- Schneider, Benjamin. “The High Cost of Global Traffic Jams.” CityLab. Accessed April 17, 2020. <https://www.citylab.com/transportation/2018/02/traffics-mind-boggling-economic-toll/552488/>.

- Seidenberg, Mark. “Figure 2. Activation Curve for Hyperbolic Tangent Function Used In...” ResearchGate. Accessed September 28, 2020. https://www.researchgate.net/figure/Activation-curve-for-hyperbolic-tangent-function-used-in-simulations-y-tanh-x-2_fig2_2439950.
- “Sex of Workers By Means of Transportation To Work.” Accessed September 25, 2020. <https://data.census.gov/cedsci/table?q=B08&d=ACS%201-Year%20Estimates%20Detailed%20Tables&tid=ACSDT1Y2018.B08006&hidePreview=false>.
- Shalaby, Amer, Ph D, P. Eng, and Ali Farhan M. A. Sc. *Bus Travel Time Prediction Model for Dynamic Operations Control and Passenger Information Systems*, 2002.
- Sharaby, Nir, and Yoram Shiftan. “The Impact of Fare Integration on Travel Behavior and Transit Ridership.” *Transport Policy* 21 (May 1, 2012): 63–70. <https://doi.org/10.1016/j.tranpol.2012.01.015>.
- SHARMA, SAGAR. “Activation Functions in Neural Networks.” Medium, February 14, 2019. <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>.
- Blog. “The History of Public Bus Transportation,” August 14, 2017. <https://gogocharters.com/blog/history-of-public-bus-transportation/>.
- Vanajakshi, L., S. C. Subramanian, and R. Sivanandan. “Travel Time Prediction under Heterogeneous Traffic Conditions Using Global Positioning System Data from Buses.” *IET Intelligent Transport Systems* 3, no. 1 (March 1, 2009): 1–9. <https://doi.org/10.1049/iet-its:20080013>.
- Watson, Nate. “Using Mean Absolute Error to Forecast Accuracy.” *Contemporary Analysis* (blog), January 23, 2012. <https://canworksmart.com/using-mean-absolute-error-forecast-accuracy/>.
- “WMATA. Metro Facts, 2017,” July 2, 2018. <https://web.archive.org/web/20180702201723/https://www.wmata.com/about/upload/Metro-Facts-2017-FINAL.pdf>.
- “WMATA_Metrobus Timetable for Route 42,43.” Accessed June 18, 2020. https://www.wmata.com/schedules/timetables/upload/42,43_170924.pdf.
- “WMATA_Metrobus Timetable for Route 52-54.” Accessed June 18, 2020. https://www.wmata.com/schedules/timetables/upload/52-54_151219.pdf.
- “WMATA_Metrobus Timetable for Route 70.” Accessed June 18, 2020. https://www.wmata.com/schedules/timetables/upload/70_180624.pdf.
- “WMATA_Metrobus Timetable for Route D4.” Accessed June 18, 2020. https://www.wmata.com/schedules/timetables/upload/D4_180624.pdf.
- “WMATA_Metrobus Timetable for Route S1.” Accessed June 18, 2020. https://www.wmata.com/schedules/timetables/upload/S1_190623.pdf.
- “WMATA_Metrobus Timetable for Routes 32,34,36.” Accessed June 18, 2020. https://www.wmata.com/schedules/timetables/upload/32,34,36_151219.pdf.

Yin, Tingting, Gang Zhong, Jian Zhang, Shanglu He, and Bin Ran. “A Prediction Model of Bus Arrival Time at Stops with Multi-Routes.” *Transportation Research Procedia*, World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016, 25 (January 1, 2017): 4623–36. <https://doi.org/10.1016/j.trpro.2017.05.381>.

About the Authors

Stephen Arhin, PhD, P.E., PTOE, PMP, CRA, Fellow ITE

Dr. Stephen Arhin is an Associate Professor in the Civil and Environmental Engineering Department at Howard University. He has more than 25 years of experience in all facets of traffic and transportation engineering. He is a registered Professional Engineer (P.E.) in DC, DE, MD, and VA, and a registered Professional Traffic Operations Engineer (PTOE). He is also a Certified Research Administrator (CRA), a certified Project Management Professional (PMP) and a Fellow of the Institute of Transportation Engineers (ITE). He has extensive experience in working with state and local transportation agencies on a wide variety of safety, operations, and design projects in addition to private industry. Dr. Arhin has authored and co-authored several project reports, published articles in peer-reviewed journals and presented at conferences on such topics as pedestrian and bicycle safety, countdown pedestrian traffic signals, intelligent transportation technologies, pavement condition monitoring, crash data analysis, traffic volume trends, mitigation of reflective cracking in composite highway pavements, truck weight enforcement and red-light violation. He is a member of ITE, TRB and ASCE, and serves as a Program Evaluator for ABET.

Babin Manandhar, EIT

Babin Manandhar is an Engineer in Training (EIT) and a research assistant at the Howard University Transportation Research and Data Center. He has a Master of Engineering degree in Civil and Environmental Engineering from Howard University.

Adam Gatiba, EIT

Adam Gatiba is a research assistant at the Howard University Transportation Research Center. He has a Master of Science degree in Civil Engineering from Howard University. Adam also works at Volkert Inc. as a Traffic Engineer with focus on traffic safety and operations.

Hamdiat Baba Adam

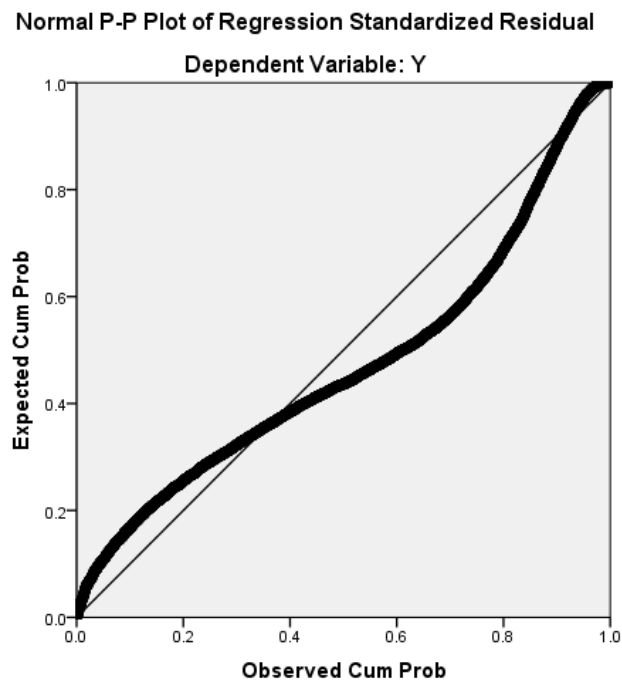
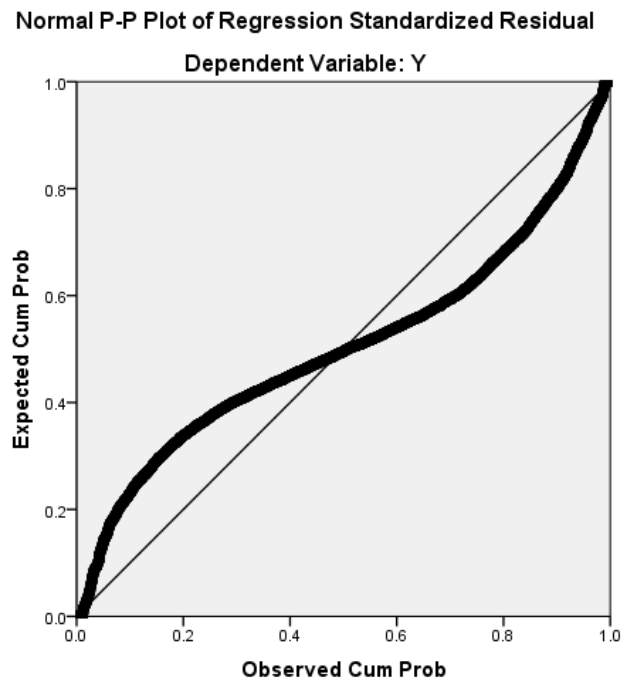
Hamdiat Baba Adam holds a Master of Science degree in Geospatial Information Sciences (GIS) from the University of Maryland. She has Geomatic Engineering experience in the Transportation and Energy Industry. Hamdiat is currently a Research Assistant at the Howard University Transportation Research Center.

Appendices to Project 1943

APPENDIX A: SPSS OUTPUT

Normality Curves (SPSS Output):

Figure 33: Normality Probability Curve Obtained for Mid-Day Peak Period (top) and PM Peak Period (bottom)



Multicollinearity Check (SPSS Output):

Table 33: SPSS Coefficient Output for Mid-Day Peak Period

Model	Coefficients ^a								
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
(Constant)	41.069	7.551		5.439	.000	26.267	55.872		
SBS (X ₁)	31.942	1.801	.178	17.739	.000	28.412	35.472	.280	3.567
L (X ₂)	.079	.001	.915	119.604	.000	.077	.080	.483	2.071
P (X ₃)	.115	.399	.002	.288	.000	-.668	.898	.971	1.030
DT (X ₄)	-1.508	.217	-.038	-6.966	.000	-1.933	-1.084	.974	1.027
IT (X ₅)	-6.973	.486	-.156	-14.346	.000	-7.926	-6.020	.239	4.186

Table 34: SPSS Coefficient Output for PM Peak Period

Model	Coefficients ^a								
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
(Constant)	-6.063	7.504		-.808	.419	-20.774	8.648		
SBS (X ₁)	55.669	1.847	.376	30.141	.000	52.048	59.289	.285	3.508
L (X ₂)	.011	.001	.136	14.231	.000	.009	.012	.488	2.051
P (X ₃)	1.883	.286	.046	6.585	.000	1.322	2.444	.927	1.079
DT (X ₄)	.888	.153	.039	5.810	.000	.588	1.187	.976	1.024
IT (X ₅)	15.242	.516	.394	29.538	.000	14.230	16.253	.249	4.020

Homoscedasticity (SPSS Output):

Figure 34: Homoscedasticity Regression Plot for Mid-Day Peak Period

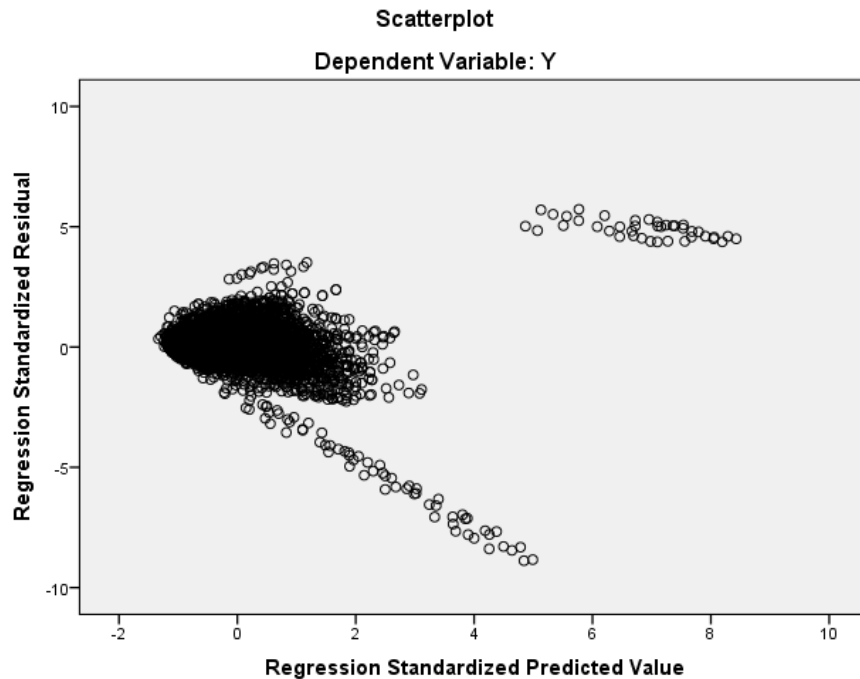
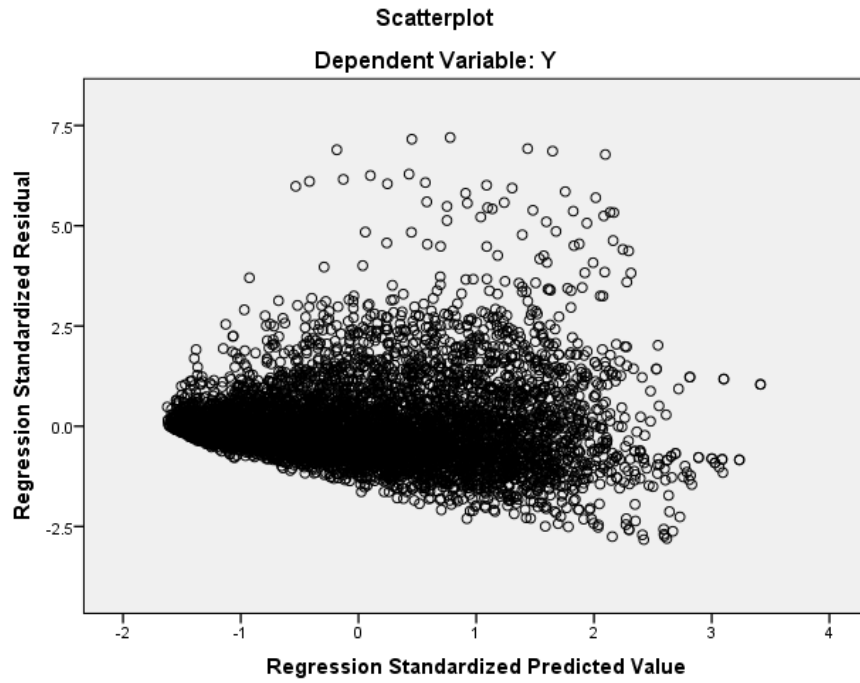


Figure 35: Homoscedasticity Regression Plot for PM Peak Period



APPENDIX B: MULTILAYER PERCEPTRON/SCALING AND UNSCALING LAYERS

Perceptron Layers (Neural Designer Output):

Table 35: Perceptron Layers for Quasi-Newton Analysis (Two Layers)

Peak Periods	Layer	Number of Inputs	Perceptron Number	Activation Function
All Peaks (AM, Mid-Day and PM)	1	5	3	Hyperbolic Tangent
	2	3	1	Linear

Table 36: Perceptron Layers for Quasi-Newton Analysis (Three Layers)

Peak Periods	Layer	Number of Inputs	Perceptron Number	Activation Function
All Peaks (AM, Mid-Day and PM)	1	5	3	Hyperbolic Tangent
	2	3	3	Hyperbolic Tangent
	3	3	1	Linear

Table 37: Perceptron Layers for Quasi-Newton Analysis (Five Layers)

Peak Periods	Layer	Number of Inputs	Perceptron Number	Activation Function
All Peaks (AM, Mid-Day and PM)	1	5	3	Hyperbolic Tangent
	2	3	3	Hyperbolic Tangent
	3	3	3	Hyperbolic Tangent
	4	3	3	Hyperbolic Tangent
	5	3	1	Linear

Table 38: Perceptron Layers for Levenberg-Marquardt Analysis (Three Layers)

Peak Periods	Layer	Number of Inputs	Perceptron Number	Activation Function
All Peaks (AM, Mid-Day and PM)	1	5	3	Hyperbolic Tangent
	2	3	3	Hyperbolic Tangent
	3	3	1	Linear

Table 39: Scaling and Unscaling Values for Quasi-Newton Analysis (Two Perceptron Layers)

Peak Periods		Scaling Layers			
		Minimum	Maximum	Mean	Deviation
AM Peak	X ₁ (SBS)	0	9	0	1
	X ₂ (L)	0	5.46e+04*	0	1
	X ₃ (P)	0	58	0	1
	X ₄ (DT)	1	76	0	1
	X ₅ (IT)	0	55	0	1
Mid-Day Peak	X ₁ (SBS)	0	9	0	1
	X ₂ (L)	0	5.46e+04	0	1
	X ₃ (P)	0.5	47.5	0	1
	X ₄ (DT)	1	90	0	1
	X ₅ (IT)	0	58	0	1
PM Peak	X ₁ (SBS)	0	9	0	1
	X ₂ (L)	0	4.14e+04	0	1
	X ₃ (P)	0.5	61.5	0	1

Scaling Layers					
Peak Periods		Minimum	Maximum	Mean	Deviation
	X ₄ (DT)	1	331	0	1
	X ₅ (IT)	0	58	0	1
Unscaling Layers					
		Minimum	Maximum	Mean	Standard Deviation
Y for AM Peak		2	1.98e+03	0	1
Y for Mid-Day Peak		6	5.18e+03	0	1
Y for PM Peak		21	2.76e+03	0	1

***5.46e+04 = 54,600**

Table 40: Scaling and Unscaling Values for Quasi-Newton Analysis (Three Perceptron Layers)

Scaling Layers					
Peak Periods		Minimum	Maximum	Mean	Deviation
AM Peak	X ₁ (SBS)	0	9	3.92	2.7
	X ₂ (L)	0	5.46e+04	6.72e+03	5.82e+03
	X ₃ (P)	0	58	15.6	9.38
	X ₄ (DT)	1	76	13.8	9.92
	X ₅ (IT)	0	55	15.7	10.3
Mid-Day Peak	X ₁ (SBS)	0	9	0	1
	X ₂ (L)	0	5.47e+04	0	1
	X ₃ (P)	0.5	47.5	0	1
	X ₄ (DT)	1	90	0	1
	X ₅ (IT)	0	58	0	1
PM Peak	X ₁ (SBS)	0	9	0	1
	X ₂ (L)	0	4.14e+04	0	1
	X ₃ (P)	0.5	61.5	0	1
	X ₄ (DT)	1	331	0	1
	X ₅ (IT)	0	58	0	1
Unscaling Layers					
		Minimum	Maximum	Mean	Standard Deviation
Y for AM Peak		2	1.98e+03	0	1
Y for Mid-Day Peak		6	5.18e+03	0	1
Y for PM Peak		21	2.76e+03	0	1

Table 41: Scaling and Unscaling Layers for Quasi-Newton Analysis (Five Perceptron Layers)

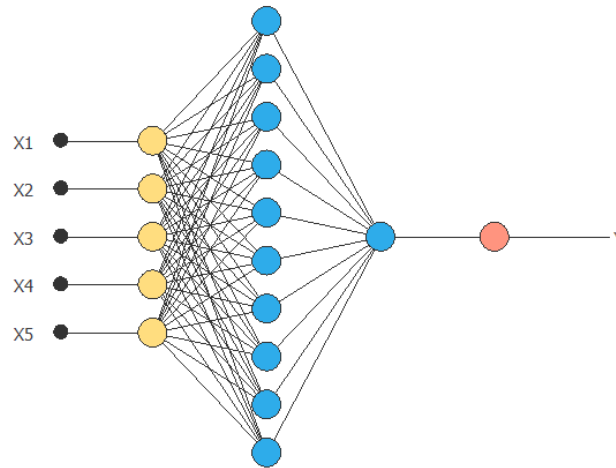
		Scaling Layers			
Peak Periods		Minimum	Maximum	Mean	Deviation
AM Peak	X ₁ (SBS)	0	9	0	1
	X ₂ (L)	0	5.46e+04	0	1
	X ₃ (P)	0	58	0	1
	X ₄ (DT)	1	76	0	1
	X ₅ (IT)	0	55	0	1
Mid-Day Peak	X ₁ (SBS)	0	9	3.91	2.7
	X ₂ (L)	0	2.69e+04	6.34e+03	4.26e+03
	X ₃ (P)	0.5	38.5	12.8	6.12
	X ₄ (DT)	1	58	15.5	11
	X ₅ (IT)	0	58	16.6	10.8
PM Peak	X ₁ (SBS)	0	9	0	1
	X ₂ (L)	0	4.14e+04	0	1
	X ₃ (P)	0.5	61.5	0	1
	X ₄ (DT)	1	331	0	1
	X ₅ (IT)	0	58	0	1
		Unscaling Layers			
		Minimum	Maximum	Mean	Standard Deviation
	Y for AM Peak	2	1.98e+03	0	1
	Y for Mid-Day Peak	6	5.18e+03	0	1
	Y for PM Peak	21	2.76e+03	0	1

Table 42: Scaling and Unscaling Layers for Levenberg-Marquardt Analysis (Three Perceptron Layers)

		Scaling Layers			
Peak Periods		Minimum	Maximum	Mean	Deviation
AM Peak	X ₁ (SBS)	0	9	0	1
	X ₂ (L)	0	5.46e+04	0	1
	X ₃ (P)	0	58	0	1
	X ₄ (DT)	1	76	0	1
	X ₅ (IT)	0	55	0	1
Mid-Day Peak	X ₁ (SBS)	0	9	3.91	2.7
	X ₂ (L)	0	2.69e+04	6.34e+03	4.26e+03
	X ₃ (P)	0.5	38.5	12.8	6.12
	X ₄ (DT)	1	58	15.5	11
	X ₅ (IT)	0	58	16.6	10.8
PM Peak	X ₁ (SBS)	0	9	0	1
	X ₂ (L)	0	4.14e+04	0	1
	X ₃ (P)	0.5	61.5	0	1
	X ₄ (DT)	1	331	0	1
	X ₅ (IT)	0	58	0	1
		Unscaling Layers			
		Minimum	Maximum	Mean	Standard Deviation
	Y for AM Peak	2	1.98e+03	0	1
	Y for Mid-Day Peak	6	5.18e+03	0	1
	Y for PM Peak	21	2.76e+03	0	1

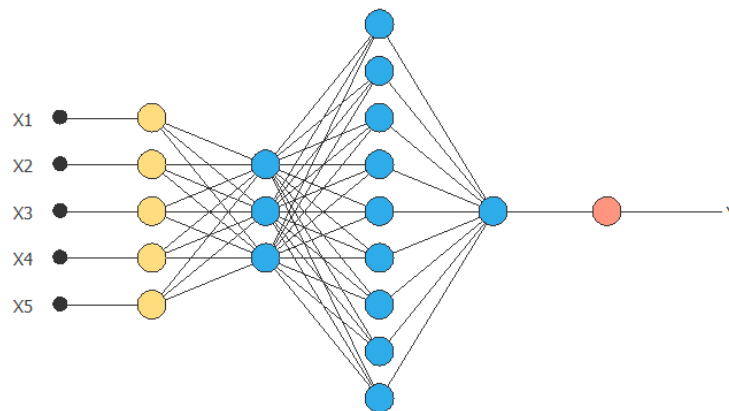
APPENDIX C: NEURAL NETWORK MODEL ARCHITECTURE

Figure 36: Neural Network Architecture for AM Peak Model with Two Perceptron Layers (Quasi-Newton Analysis)



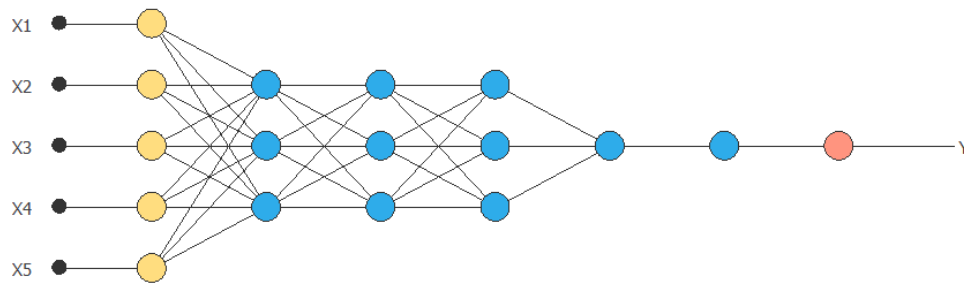
Complexity: Number of hidden neurons is ten.

Figure 37: Neural Network Architecture for AM Peak Model with Three Perceptron Layers (Quasi-Newton Analysis)



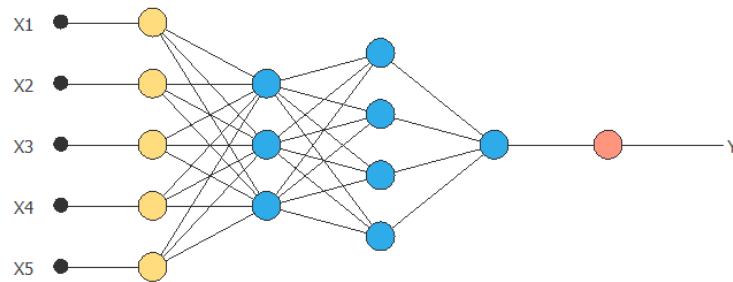
Complexity: Number of hidden neurons is 3:9.

Figure 38: Neural Network Architecture for AM Peak Model with Five Perceptron Layers (Quasi-Newton Analysis)



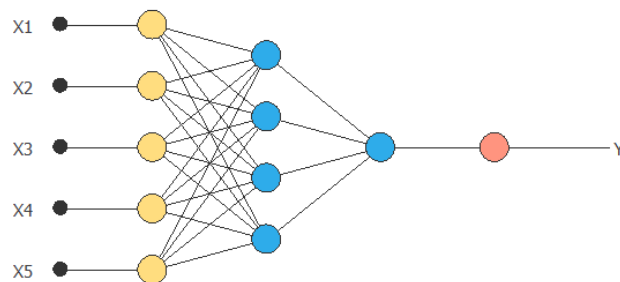
Complexity: Number of hidden neurons is 3:3:3:1.

Figure 39: Neural Network Architecture for AM Peak Model with Three Perceptron Layers (Levenberg-Marquardt Analysis)



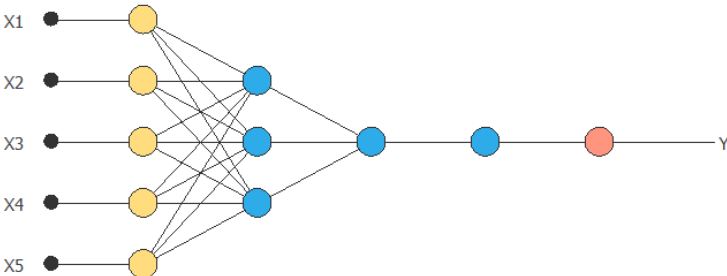
Complexity: Number of hidden neurons is 3:4.

Figure 40: Neural Network Architecture for Mid-Day Peak Model with Two Perceptron Layers (Quasi-Newton Analysis)



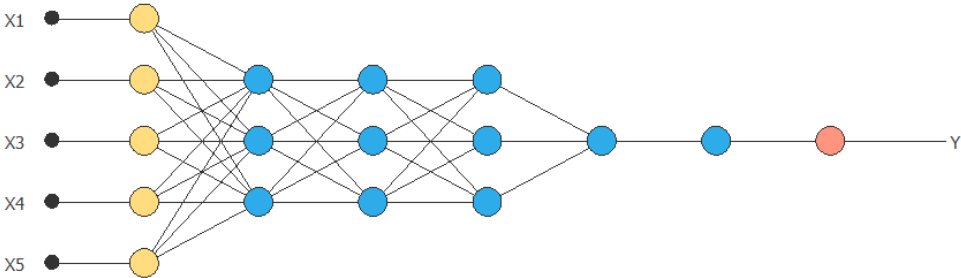
Complexity: Number of hidden neurons is four.

Figure 41: Neural Network Architecture for Mid-Day Peak Model with Three Perceptron Layers (Quasi-Newton Analysis)



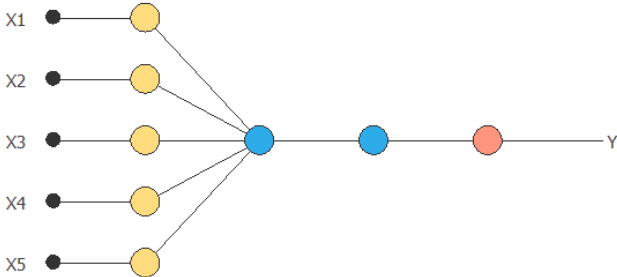
Complexity: Number of hidden neurons is 3:1.

Figure 42: Neural Network Architecture for Mid-Day Peak Model with Five Perceptron Layers (Quasi-Newton Analysis)



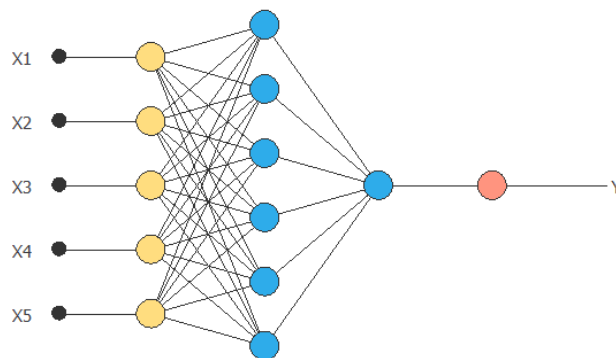
Complexity: Number of hidden neurons is 3:3:3:1.

Figure 43: Neural Network Architecture for Mid-Day Peak Model with Three Perceptron Layers (Levenberg-Marquardt Analysis)



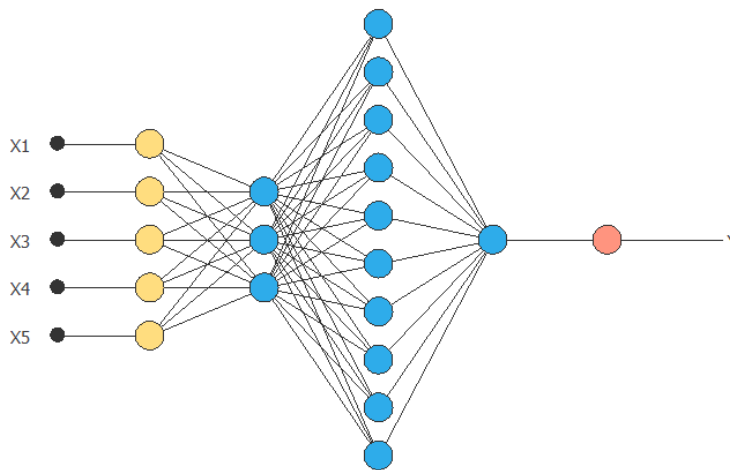
Complexity: Number of hidden neurons is one.

Figure 44: Neural Network Architecture for PM Peak Model with Two Perceptron Layers (Quasi-Newton Analysis)



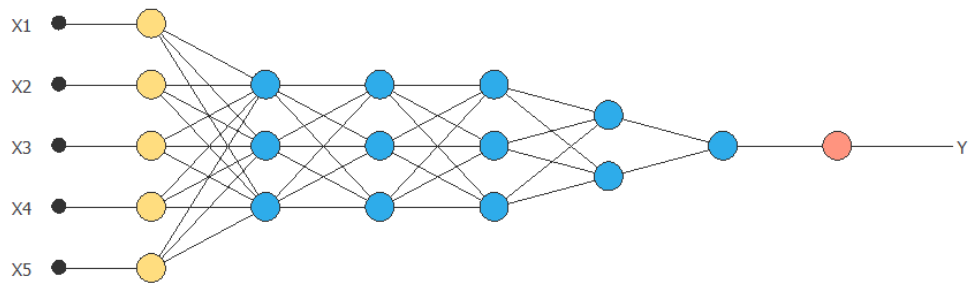
Complexity: Number of hidden neurons is six.

Figure 45: Neural Network Architecture for PM Peak Model with Three Perceptron Layers (Quasi-Newton Analysis)



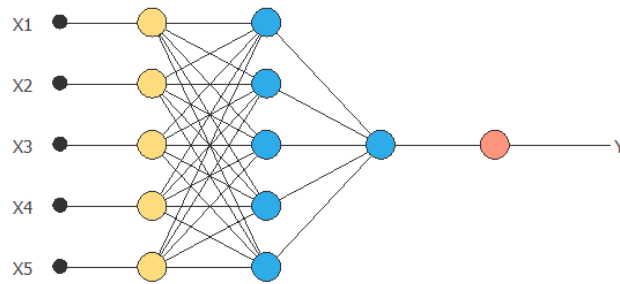
Complexity: Number of hidden neurons is 3:10.

Figure 46: Neural Network Architecture for PM Peak Model with Five Perceptron Layers (Levenberg-Marquardt Analysis)



Complexity: Number of hidden neurons is 3:3:3:2.

Figure 47: Neural Network Architecture for PM Peak Model with Three Perceptron Layers (Quasi-Newton Analysis)



Complexity: Number of hidden neurons is five.

APPENDIX D: NEURAL DESIGNER MODEL TRAINING PLOTS

Error Plots (Neural Designer Output):

Figure 48: Quasi-Newton Method Error History Plot for Mid-Day Peak Period
(Two Perceptron Layers)

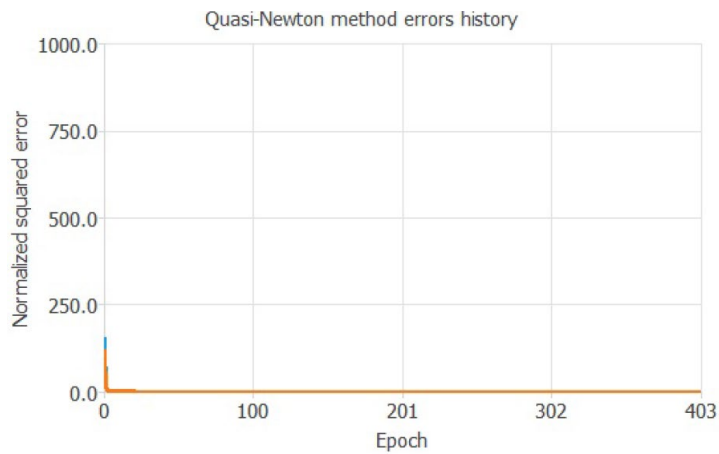


Figure 49: Quasi-Newton Method Error History Plot for PM Peak Period
(Two Perceptron Layers)

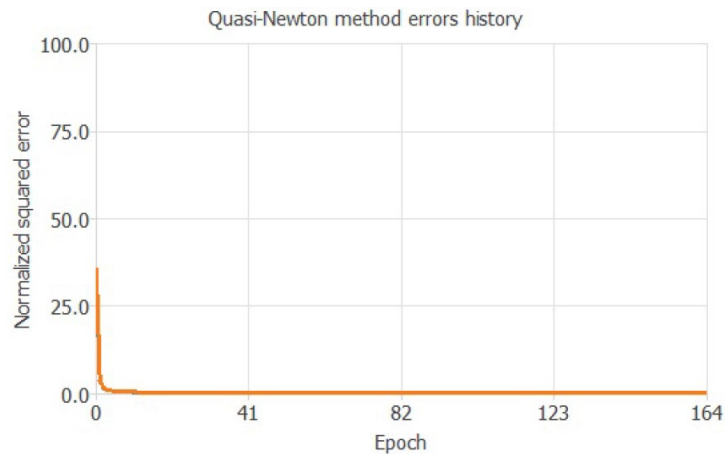


Figure 50: Quasi-Newton Method Error History Plot after Order Selection for AM Peak Period (Two Perceptron Layers)

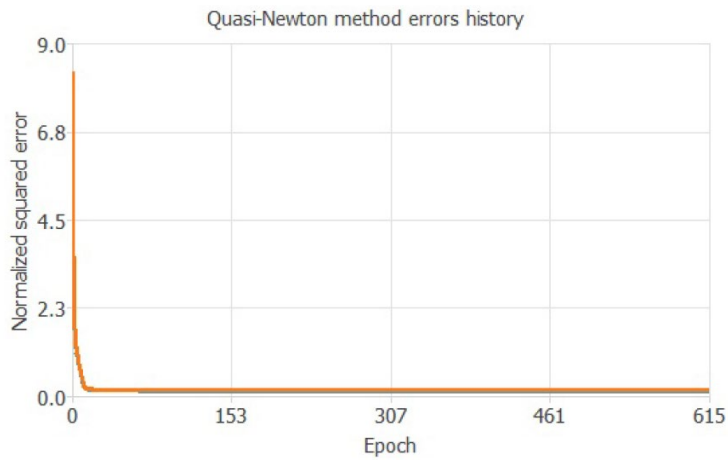


Figure 51: Quasi-Newton Method Error History Plot after Order Selection for Mid-Day Peak Period (Two Perceptron Layers)

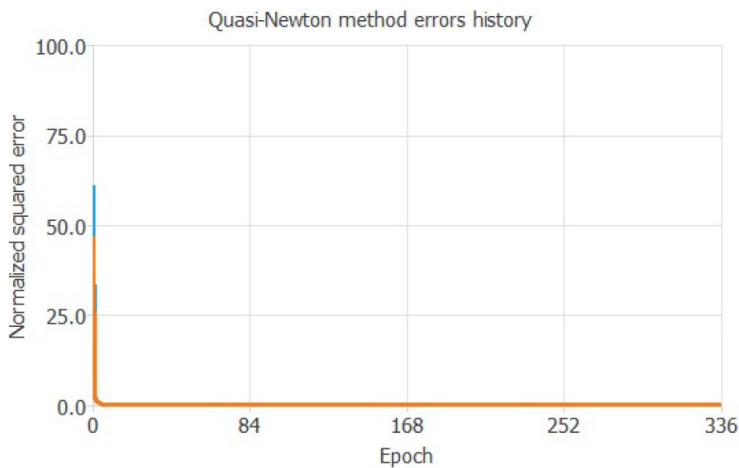


Figure 52: Quasi-Newton Method Error History Plot after Order Selection for PM Peak Period (Two Perceptron Layers)

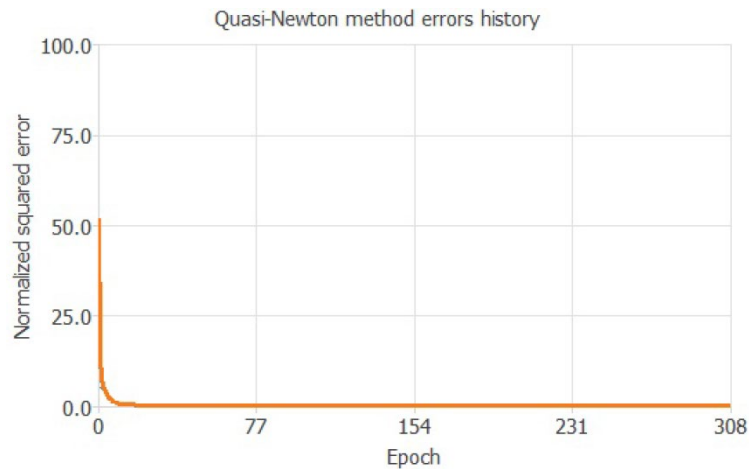


Figure 53: Quasi-Newton Method Error History Plot for AM Peak Period (Three Perceptron Layers)

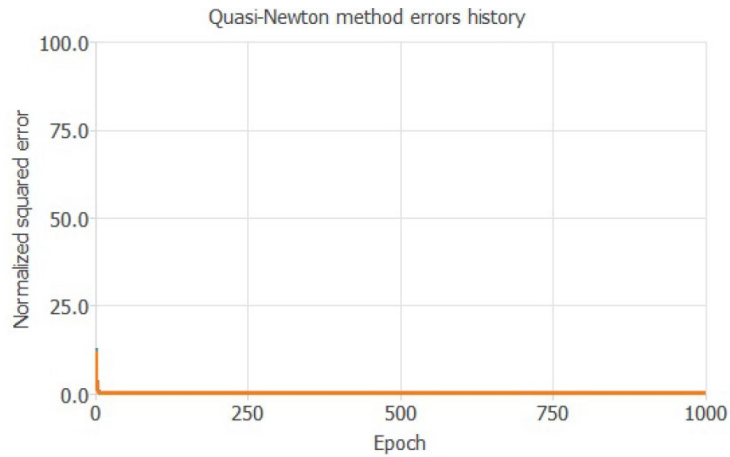


Figure 54: Quasi-Newton Method Error History Plot for Mid-Day Peak Period (Three Perceptron Layers)

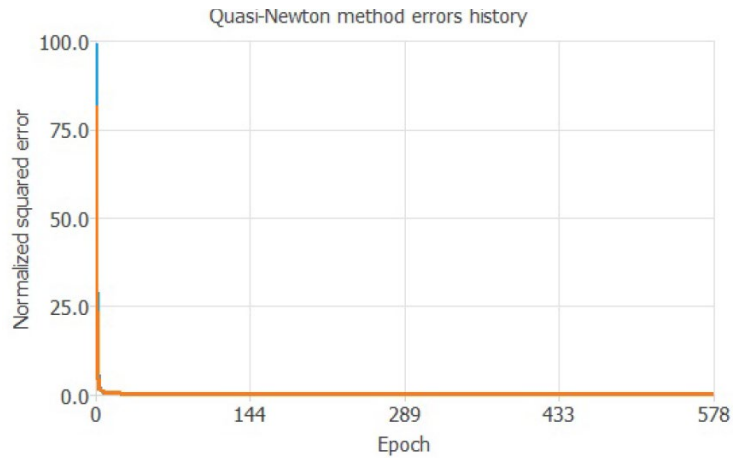


Figure 55: Quasi-Newton Method Error History Plot for PM Peak Period (Three Perceptron Layers)

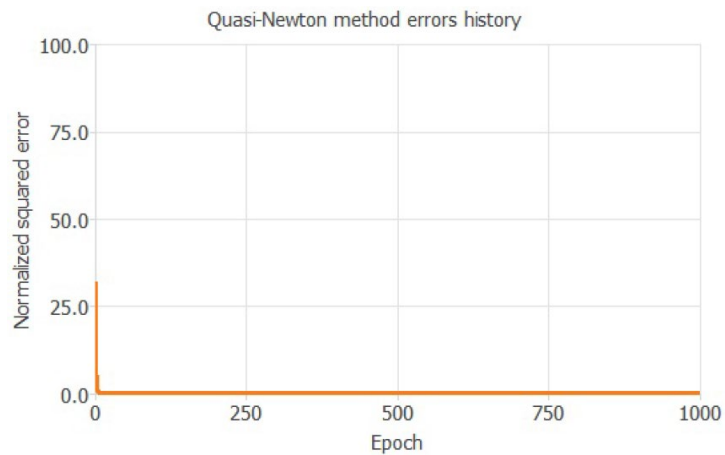


Figure 56: Quasi-Newton Method Error History Plot after Order Selection for AM Peak Period (Three Perceptron Layers)

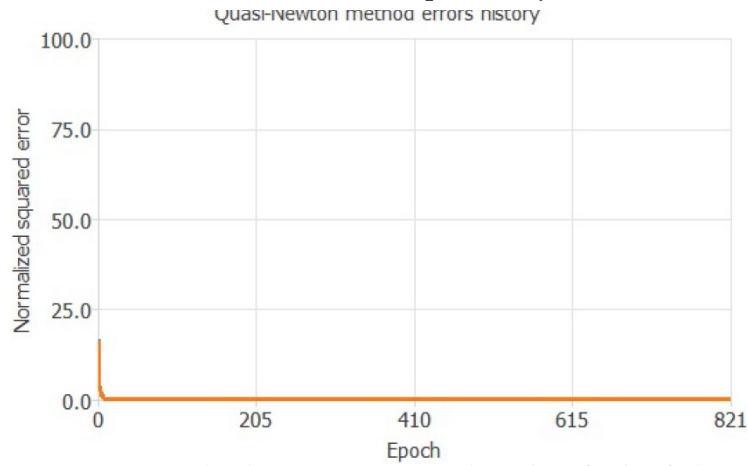


Figure 57: Quasi-Newton Method Error History Plot after Order Selection for Mid-Day Peak Period (Three Perceptron Layers)

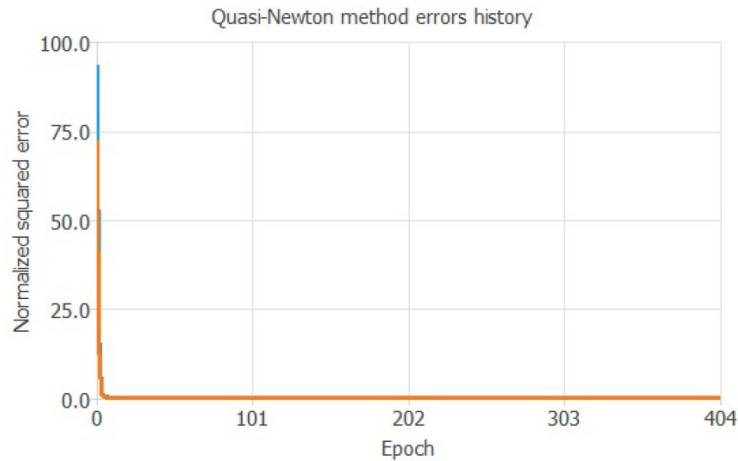


Figure 58: Quasi-Newton Method Error History Plot after Order Selection for PM Peak Period (Three Perceptron Layers)

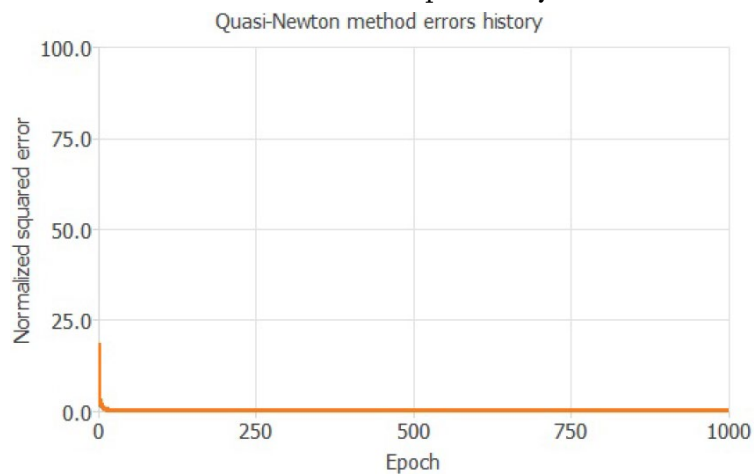


Figure 59: Quasi-Newton Method Error History Plot for AM Peak Period (Five Perceptron Layers)

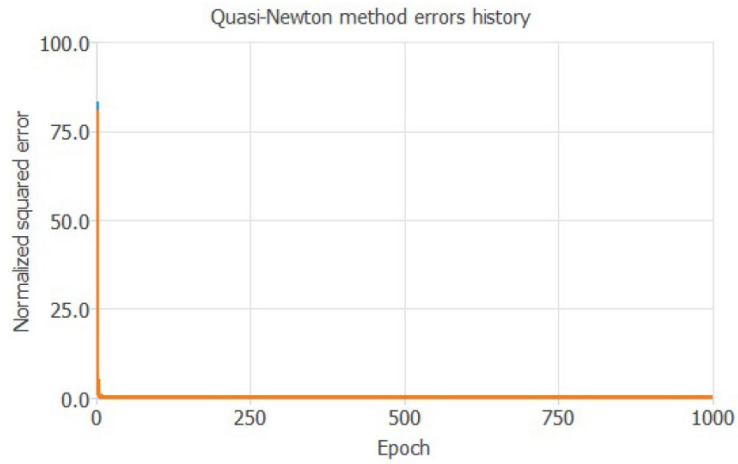


Figure 60: Quasi-Newton Method Error History Plot for Mid-Day Peak Period (Five Perceptron Layers)

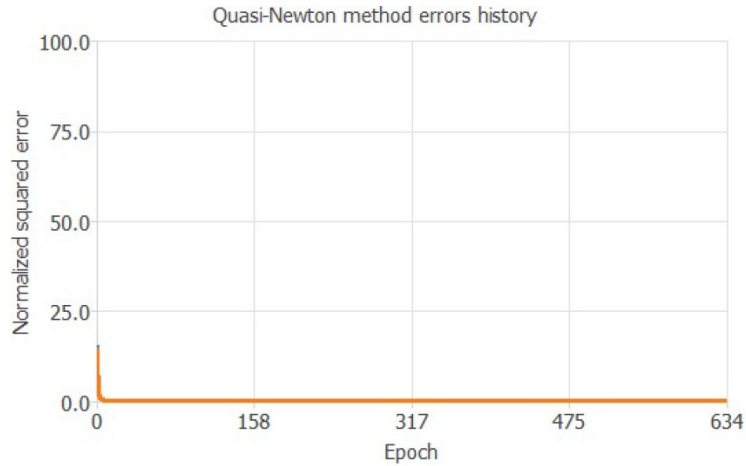


Figure 61: Quasi-Newton Method Error History Plot for PM Peak Period (Five Perceptron Layers)

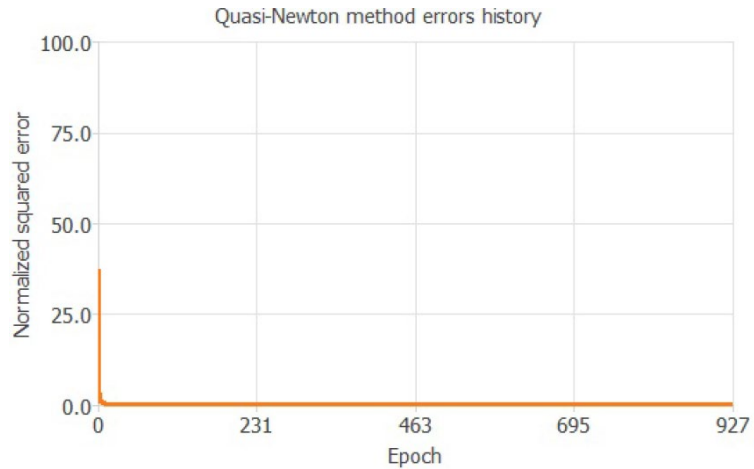


Figure 62: Quasi-Newton Method Error History Plot after Order Selection for AM Peak Period (Five Perceptron Layers)

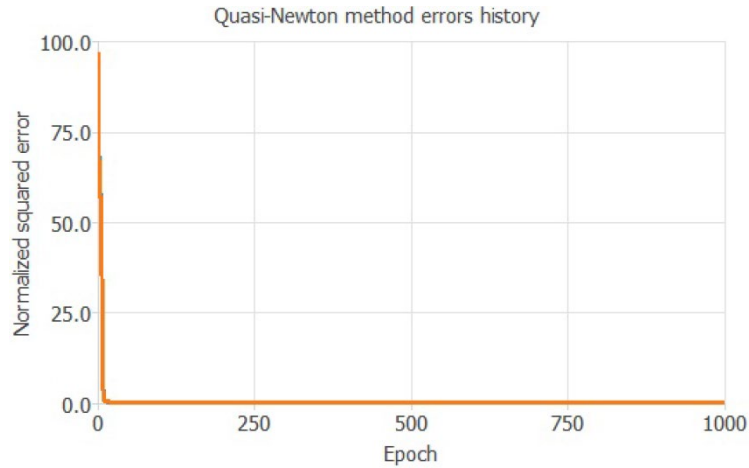


Figure 63: Quasi-Newton Method Error History Plot after Order Selection for Mid-Day Peak Period (Five Perceptron Layers)

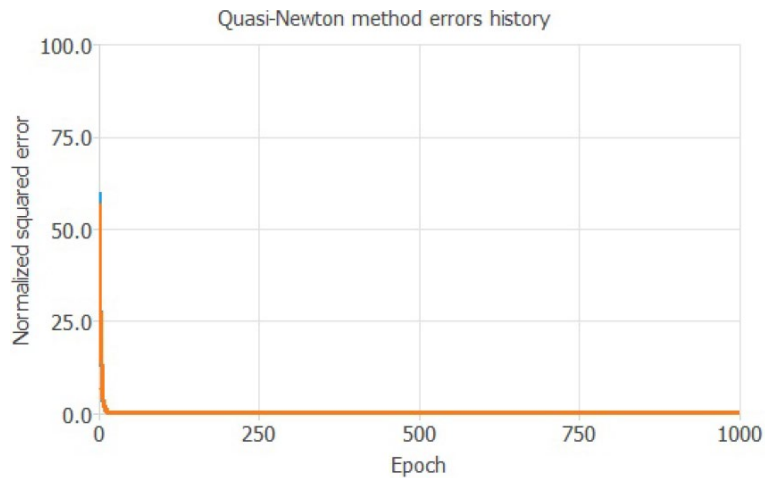


Figure 64: Quasi-Newton Method Error History Plot after Order Selection for PM Peak Period (Five Perceptron Layers)

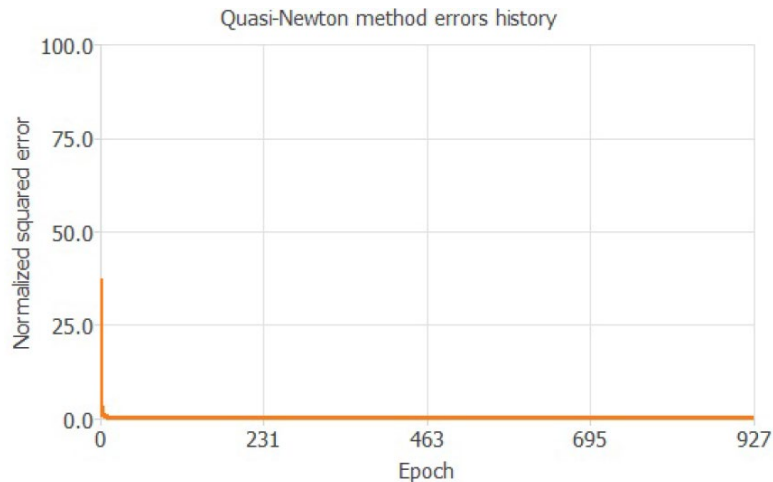


Figure 65: Levenberg-Marquardt Method Error History Plot for AM Peak Period (Three Perceptron Layers)

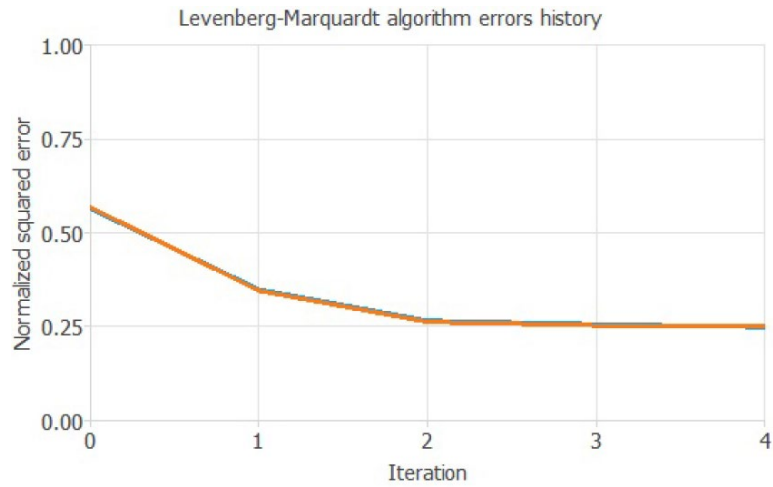


Figure 66: Levenberg-Marquardt Method Error History Plot for Mid-Day Peak Period (Three Perceptron Layers)

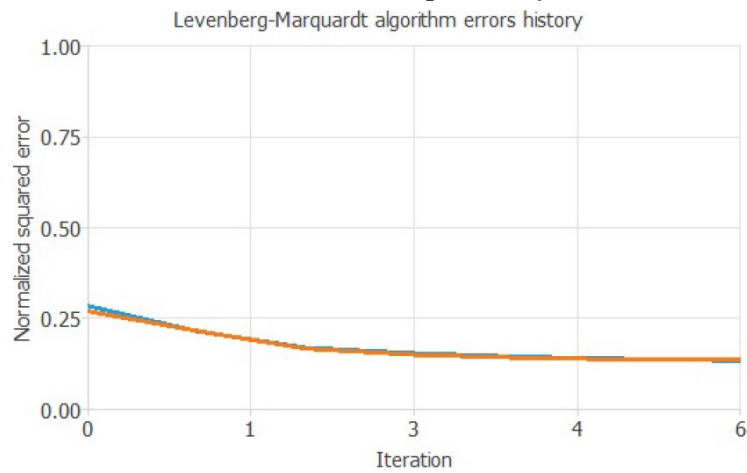


Figure 67: Levenberg-Marquardt Method Error History Plot for PM Peak Period (Three Perceptron Layers)

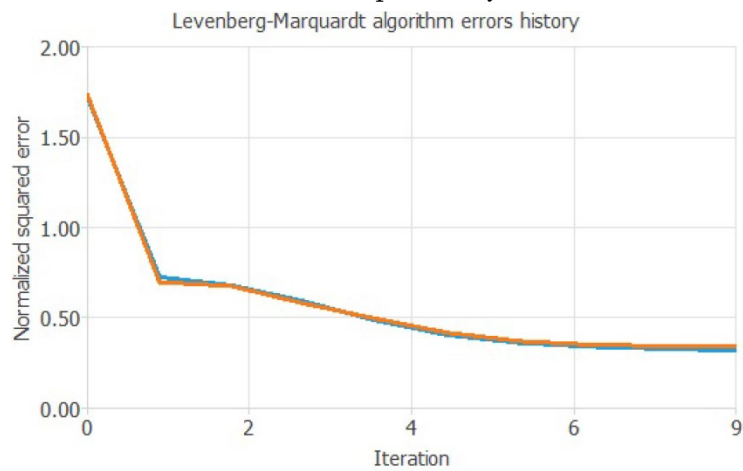


Figure 68: Levenberg-Marquardt Method Error History Plot after Order Selection for AM Peak Period (Three Perceptron Layers)

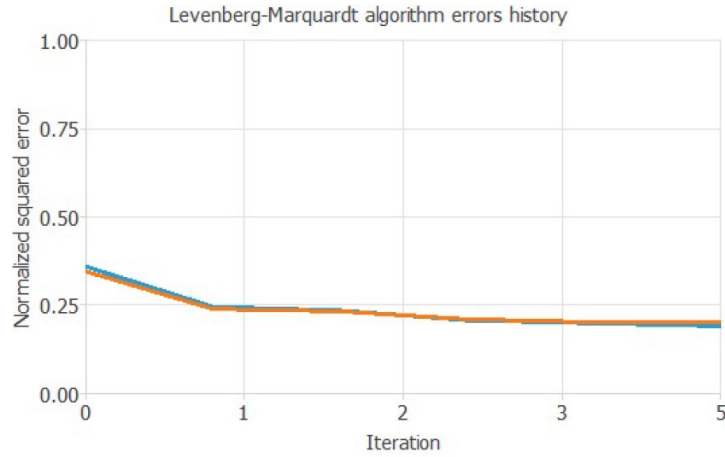


Figure 69: Levenberg-Marquardt Method Error History Plot after Order Selection for Mid-Day Peak Period (Three Perceptron Layers)

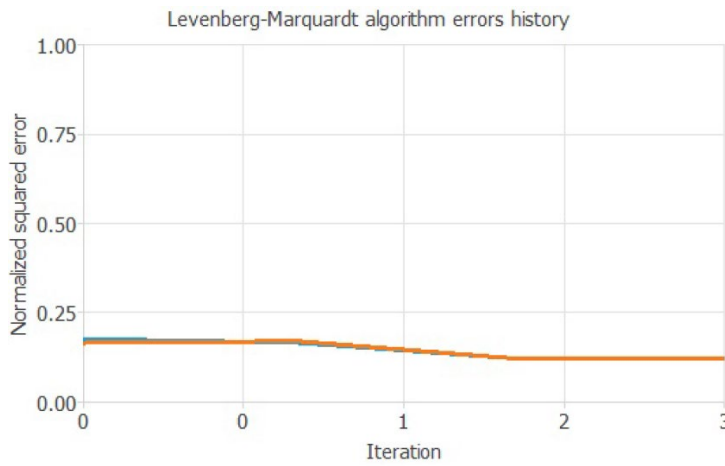
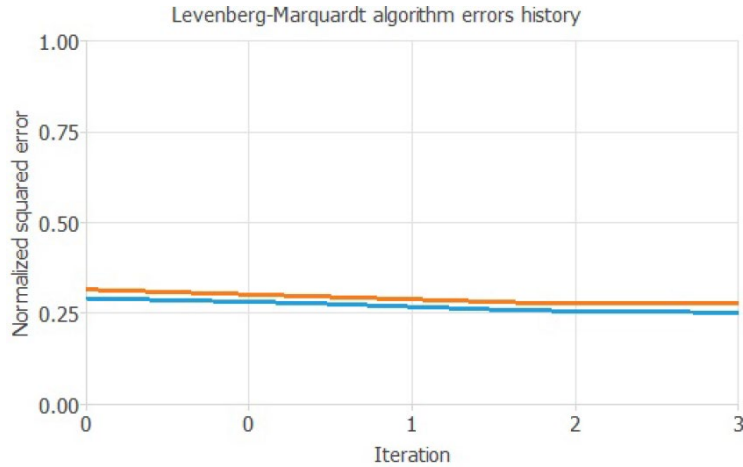


Figure 70: Levenberg-Marquardt Method Error History Plot after Order Selection for PM Peak Period (Three Perceptron Layers)



APPENDIX E: NEURAL NETWORK ERROR TABLES

Neural Network Training, Selection, and Testing Errors (Neural Designer Output):

Table 43: Errors for Training, Selection, and Testing Instances (Two Layers, Quasi-Newton Method)

Error Type	AM Peak Error			Mid-Day Peak Error			PM Peak Error		
	Training	Selection	Testing	Training	Selection	Testing	Training	Selection	Testing
Sum of Squared Error	6.38e+07	2.16e+07	2.13e+07	3.57e+07	1.21e+07	1.25e+07	1.45e+08	4.86e+07	5.63e+07
Mean Squared Error	14788.8	15088.7	14825.1	11491.2	11669.3	12057.4	35541.5	35808.7	41439.5
Root Mean Squared Error	121.609	122.836	121.758	107.197	108.025	109.806	188.525	189.232	203.567
Normalized Squared Error	0.138	0.145	0.147	0.052	0.041	0.053	0.218	0.229	0.257
Minkowski Error	4.69e+06	1.62e+06	1.58e+06	2.76e+06	921756	939850	8.10e+06	2.73e+06	2.93e+06

Table 44: Error Statistics for Quasi-Newton Method for All Peak Periods (Two Perceptron Layers)

Peak Periods	Error Type	Minimum	Maximum	Mean	Standard Deviation
AM Peak	Absolute Error	0.068	718.353	90.727	81.230
	Percentage Error	0.003	36.354	4.591	4.111
Mid-Day Peak	Absolute Error	0.030	604.830	78.129	77.194
	Percentage Error	0.0006	11.689	1.510	1.492
PM Peak	Absolute Error	0.009	1549.100	134.589	152.782
	Percentage Error	0.0003	56.557	4.914	5.578

Table 45: Errors for Training, Selection, and Testing Instances (Three Layers, Quasi-Newton Method)

Error Type	AM Peak Error			Mid-Day Peak Error			PM Peak Error		
	Training	Selection	Testing	Training	Selection	Testing	Training	Selection	Testing
Sum of Squared Error	6.61e+07	2.39e+07	2.78e+07	3.68e+07	1.24e+07	1.29e+07	1.36e+08	4.86e+07	5.56e+07
Mean Squared Error	15330.5	16651.2	16550.9	11826.7	11942.6	12498.1	33309.9	35710.4	40933.2
Root Mean Squared Error	123.816	129.039	128.650	108.751	109.282	111.795	182.510	188.972	202.320
Normalized Squared Error	0.145	0.154	0.163	0.053	0.041	0.055	0.205	0.229	0.254
Minkowski Error	4.85e+06	1.70e+06	1.70e+06	2.83e+06	933523	965816	7.68e+06	2.71e+06	2.88e+06

Table 46: Error Statistics for Quasi-Newton Method for All Peak Periods (Three Perceptron Layers)

Peak Periods	Error Type	Minimum	Maximum	Mean	Standard Deviation
AM Peak	Absolute Error	0.109	575.922	94.572	87.249
	Percentage Error	0.005	29.146	4.786	4.415
Mid-Day Peak	Absolute Error	0.069	609.131	79.485	78.651
	Percentage Error	0.001	11.773	1.536	1.520
PM Peak	Absolute Error	0.121	1463.190	132.041	153.348
	Percentage Error	0.004	53.421	4.821	5.598

Table 47: Errors for Training, Selection and Testing Instances (Five Layers, Quasi-Newton Method)

Error Type	AM Peak Error			Mid-Day Peak Error			PM Peak Error		
	Training	Selection	Testing	Training	Selection	Testing	Training	Selection	Testing
Sum of Squared Error	6.84e+0 7	2.46e+07	2.46e+0 7	3.22e+0 7	1.23e+07	1.26e+0 7	1.40e+08	4.77e+07	5.64e+07
Mean Squared Error	15867.8	17132.2	17184.4	10628.5	12314.6	12570.1	34437.9	35120.2	41491.9
Root Mean Squared Error	125.967	130.890	131.088	103.094	110.971	112.116	185.575	187.404	203.696
Normalized Squared Error	0.150	0.158	0.168	0.094	0.106	0.119	0.211	0.225	0.257
Minkowski Error	4.99e+0 6	1.74e+06	1.76e+0 6	2.54e+0 6	909141	934319	7.80e+06	2.68e+06	2.88e+06

Table 48: Error Statistics for Quasi-Newton Method for All Peak Periods (Five Perceptron Layers)

Peak Periods	Error Type	Minimum	Maximum	Mean	Standard Deviation
AM Peak	Absolute Error	0.042	621.841	97.242	87.939
	Percentage Error	0.002	31.469	4.921	4.450
Mid-Day Peak	Absolute Error	0.192	615.793	79.237	79.358
	Percentage Error	0.010	33.036	4.250	4.257
PM Peak	Absolute Error	0.013	1522.14	131.381	155.720
	Percentage Error	0.0004	55.572	4.796	5.685

Table 49: Errors for Training, Selection, and Testing Instances (Three Layers, Levenberg-Marquardt)

Error Type	AM Peak Error			Mid-Day Peak Error			PM Peak Error		
	Training	Selection	Testing	Training	Selection	Testing	Training	Selection	Testing
Sum of Squared Error	8.59e+07	3.22e+07	2.97e+07	4.12e+07	1.41e+07	1.44e+07	1.67e+08	5.92e+07	6.01e+07
Mean Squared Error	19919.9	22427	20705.2	13594.2	14119.6	14317.9	40982.3	43581.1	44201.7
Root Mean Squared Error	141.138	149.757	143.893	116.594	118.826	119.657	202.441	208.761	210.242
Normalized Squared Error	0.190	0.201	0.206	0.120	0.121	0.136	0.251	0.279	0.274
Minkowski Error	5.97e+06	2.17e+06	2.05e+06	3.01e+06	1.01e+06	1.02e+06	9.14e+06	3.20e+06	3.14e+06

Table 50: Error Statistics for Levenberg-Marquardt Method for All Peak Periods (Three Perceptron Layers)

Peak Periods	Error Type	Minimum	Maximum	Mean	Standard Deviation
AM Peak	Absolute Error	0.119	641.214	109.063	93.896
	Percentage Error	0.006	32.450	5.519	4.751
Mid-Day Peak	Absolute Error	0.091	613.172	83.776	85.478
	Percentage Error	0.004	32.8955	4.494	4.585
PM Peak	Absolute Error	0.040	1497.860	143.402	153.802
	Percentage Error	0.001	54.686	5.235	5.615

APPENDIX F: NEURAL NETWORK TRAVEL TIME EQUATIONS

Travel Time Equations Obtained from Neural Network Analyses (Neural Designer Output):

Equations for Quasi-Newton (Two Perceptron Layers) AM Model

$$scaled_X_1 = \frac{X_1 - 3.92}{2.7}$$

$$scaled_X_2 = \frac{X_2 - 6718.9}{5817.77}$$

$$scaled_X_3 = \frac{X_1 - 15.58}{9.37}$$

$$scaled_X_4 = \frac{X_4 - 13.81}{9.92}$$

$$scaled_X_5 = \frac{X_5 - 15.72}{10.26}$$

$$y_{1_1} = \tanh (0.27 + (scaled_X_1 * 0.16) + (scaled_X_2 * 1.38) + (scaled_X_3 * 0.77) + (scaled_X_4 * -0.1) + (scaled_X_5 * -1.59))$$

$$y_{1_2} = \tanh (0.76 + (scaled_X_1 * 0.28) + (scaled_X_2 * -0.32) + (scaled_X_3 * -0.43) + (scaled_X_4 * -0.08) + (scaled_X_5 * -0.56))$$

$$y_{1_3} = \tanh (-0.77 + (scaled_X_1 * -0.1) + (scaled_X_2 * -0.52) + (scaled_X_3 * 0.33) + (scaled_X_4 * 0.06) + (scaled_X_5 * 0.78))$$

$$y_{1_4} = \tanh (-0.73 + (scaled_X_1 * -1.33) + (scaled_X_2 * 0.17) + (scaled_X_3 * -0.60) + (scaled_X_4 * -0.13) + (scaled_X_5 * 1.49))$$

$$y_{1_5} = \tanh (-0.70 + (scaled_X_1 * -0.80) + (scaled_X_2 * 1.77) + (scaled_X_3 * -0.15) + (scaled_X_4 * 0.25) + (scaled_X_5 * -0.59))$$

$$y_{1_6} = \tanh (-0.62 + (scaled_X_1 * 0.15) + (scaled_X_2 * -0.73) + (scaled_X_3 * 0.38) + (scaled_X_4 * 0.04) + (scaled_X_5 * 0.59))$$

$$y_{1_7} = \tanh (-0.26 + (scaled_X_1 * 0.05) + (scaled_X_2 * 1.26) + (scaled_X_3 * -1.06) + (scaled_X_4 * 0.02) + (scaled_X_5 * -0.64))$$

$$y_{1_8} = \tanh (-0.51 + (scaled_X_1 * -0.23) + (scaled_X_2 * -1.86) + (scaled_X_3 * 0.81) + (scaled_X_4 * 0.08) + (scaled_X_5 * 1.92))$$

$$y_{1_9} = \tanh (-0.01 + (scaled_X_1 * -0.06) + (scaled_X_2 * -0.70) + (scaled_X_3 * 0.69) + (scaled_X_4 * -0.02) + (scaled_X_5 * 0.22))$$

$$y_{1_10} = \tanh(-0.03 + (\text{scaled_}X_1 * 0.13) + (\text{scaled_}X_2 * 0.19) + (\text{scaled_}X_3 * 0.04) + (\text{scaled_}X_4 * 0.02) + (\text{scaled_}X_5 * -0.13))$$

$$\text{Scaled_}Y_{AM} = (-0.4 + (y_{1_1} * -0.83) + (y_{1_2} * 0.48) + (y_{1_3} * 1.52) + (y_{1_4} * 0.18) + (y_{1_5} * -0.24) + (y_{1_6} * -1.17) + (y_{1_7} * 0.39) + (y_{1_8} * 0.69) + (y_{1_9} * 0.69) + (y_{1_10} * 2.38))$$

Equations for Quasi-Newton (Two Perceptron Layers) Mid-Day Model

$$\text{scaled_}X_1 = \frac{X_1 - 3.94}{2.7}$$

$$\text{scaled_}X_2 = \frac{X_2 - 6734.2}{5663.62}$$

$$\text{scaled_}X_3 = \frac{X_1 - 12.96}{6.57}$$

$$\text{scaled_}X_4 = \frac{X_4 - 16.08}{12.09}$$

$$\text{scaled_}X_5 = \frac{X_5 - 16.71}{10.88}$$

$$y_{1_1} = \tanh(-0.09 + (\text{scaled_}X_1 * -0.90) + (\text{scaled_}X_2 * 0.77) + (\text{scaled_}X_3 * -0.21) + (\text{scaled_}X_4 * -0.01) + (\text{scaled_}X_5 * 0.31))$$

$$y_{1_2} = \tanh(1.09 + (\text{scaled_}X_1 * 0.34) + (\text{scaled_}X_2 * -0.35) + (\text{scaled_}X_3 * 0.05) + (\text{scaled_}X_4 * -0.02) + (\text{scaled_}X_5 * -0.01))$$

$$y_{1_3} = \tanh(1.68 + (\text{scaled_}X_1 * 0.65) + (\text{scaled_}X_2 * -0.78) + (\text{scaled_}X_3 * -0.01) + (\text{scaled_}X_4 * -0.02) + (\text{scaled_}X_5 * 0.04))$$

$$y_{1_4} = \tanh(-0.012 + (\text{scaled_}X_1 * 0.05) + (\text{scaled_}X_2 * 0.07) + (\text{scaled_}X_3 * 0.01) + (\text{scaled_}X_4 * -0.13) + (\text{scaled_}X_5 * 0.03))$$

$$\begin{aligned} \text{Scaled_}Y_{Mid-Day} &= (-0.21 + (y_{1_1} * -0.16) + (y_{1_1} * -1.66) + (y_{1_3} * 0.80) \\ &+ (y_{1_4} * 1.14)) \end{aligned}$$

Equations for Quasi-Newton (Two Perceptron Layers) PM Model

$$\text{scaled_}X_1 = \frac{X_1 - 3.94}{2.71}$$

$$\text{scaled_}X_4 = \frac{X_4 - 17.54}{17.71}$$

$$\text{scaled_}X_2 = \frac{X_2 - 6016.89}{5186.32}$$

$$\text{scaled_}X_5 = \frac{X_5 - 15.66}{10.39}$$

$$\text{scaled_}X_3 = \frac{X_1 - 16.92}{9.71}$$

The scaled values of X_1 through X_5 for the PM Peak periods are same for all analysis (with all layers).

$$y_{1_1} = \tanh (-0.13 + (\text{scaled}X_1 * -0.33) + (\text{scaled}_X2 * 1.98) + (\text{scaled}_X3 * -0.21) + (\text{scaled}_X4 * 0.13) + (\text{scaled}_X5 * -1.18))$$

$$y_{1_2} = \tanh (0.43 + (\text{scaled}X_1 * -0.11) + (\text{scaled}_X2 * 1.63) + (\text{scaled}_X3 * -0.06) + (\text{scaled}_X4 * 0.08) + (\text{scaled}_X5 * -1.27))$$

$$y_{1_3} = \tanh (0.89 + (\text{scaled}X_1 * -0.14) + (\text{scaled}_X2 * 0.11) + (\text{scaled}_X3 * -0.11) + (\text{scaled}_X4 * -0.02) + (\text{scaled}_X5 * -0.47))$$

$$y_{1_4} = \tanh (0.34 + (\text{scaled}X_1 * -0.05) + (\text{scaled}_X2 * -0.78) + (\text{scaled}_X3 * -0.09) + (\text{scaled}_X4 * -0.03) + (\text{scaled}_X5 * 0.84))$$

$$y_{1_5} = \tanh (-1.14 + (\text{scaled}_X1 * 0.16) + (\text{scaled}_X2 * -2.72) + (\text{scaled}_X3 * 0.24) + (\text{scaled}_X4 * -0.11) + (\text{scaled}_X5 * 1.99))$$

$$y_{1_6} = \tanh (1.35 + (\text{scaled}_X1 * 0.07) + (\text{scaled}_X2 * -1.51) + (\text{scaled}_X3 * -0.03) + (\text{scaled}_X4 * -0.09) + (\text{scaled}_X5 * 1.63))$$

$$\text{Scaled}_{Y_{PM}} = (-0.11 + (y_{1_1} * 1.09) + (y_{1_2} * -2.69) - (y_{1_3} * -0.55) + (y_{1_4} * -2.14) + (y_{1_5} * -0.94) + (y_{1_6} * -1.21))$$

Equations for Quasi-Newton (Three Perceptron Layers) AM Model

The scaled values of X_1 to X_5 used in the AM Peak period calculations for the Quasi-Newton method with three perceptron layers were the same as the scaled values used for AM Peak period calculations with two perceptron layers.

$$y_{1_1} = \tanh (-0.53 + (\text{scaled}_X1 * 0.33) + (\text{scaled}_X2 * 0.27) + (\text{scaled}_X3 * -0.16) + (\text{scaled}_X4 * -0.02) + (\text{scaled}_X5 * -0.33))$$

$$y_{1_2} = \tanh (0.31 + (\text{scaled}X_1 * -0.09) + (\text{scaled}_X2 * -0.49) + (\text{scaled}_X3 * -0.01) + (\text{scaled}_X4 * -0.02) + (\text{scaled}_X5 * 0.07))$$

$$y_{1_3} = \tanh (-0.29 + (\text{scaled}X_1 * 0.23) + (\text{scaled}_X2 * -1.69) + (\text{scaled}_X3 * 0.01) + (\text{scaled}_X4 * 0.03) + (\text{scaled}_X5 * 1.52))$$

$$y_{2_1} = \tanh (-0.54 + (y_{1_1} * 1.51) + (y_{1_2} * 0.76) + (y_{1_3} * -1.32))$$

$$y_{2_2} = \tanh (-1.94 + (y_{1_1} * -0.39) + (y_{1_2} * -1.71) + (y_{1_3} * 0.29))$$

$$y_{2_3} = \tanh (1.43 + (y_{1_1} * 0.70) + (y_{1_2} * 0.25) + (y_{1_3} * 1.04))$$

$$y_{2_4} = \tanh (0.05 + (y_{1_1} * -0.77) + (y_{1_2} * -0.51) + (y_{1_3} * -0.25))$$

$$y_{2_5} = \tanh (0.76 + (y_{1_1} * -0.15) + (y_{1_2} * 0.82) + (y_{1_3} * 1.99))$$

$$y_{2_6} = \tanh (1.55 + (y_{1_1} * 0.12) + (y_{1_2} * -0.33) + (y_{1_3} * -1.35))$$

$$y_{2_7} = \tanh (0.02 + (y_{1_1} * 0.37) + (y_{1_2} * 0.33) + (y_{1_3} * 0.18))$$

$$y_{2_8} = \tanh (-0.13 + (y_{1_1} * -0.33) + (y_{1_2} * -0.26) + (y_{1_3} * 0.59))$$

$$y_{2_9} = \tanh (-1.89 + (y_{1_1} * -1.69) + (y_{1_2} * 0.33) + (y_{1_3} * 1.17))$$

$$\begin{aligned} Scaled_{Y_{AM}} = & (0.22 + (y_{2_1} * -0.98) + (y_{2_2} * 1.23) + (y_{2_3} * 1.12) + (y_{2_4} * -1.6) \\ & + (y_{2_5} * -1.05) + (y_{2_6} * 1.34) + (y_{2_7} * 0.27) + (y_{2_8} * -0.35) \\ & + (y_{2_9} * 1.2) \end{aligned}$$

Equations for Quasi-Newton (Three Perceptron Layers) Mid-Day Model

The scaled values of X_1 to X_5 used in the Mid-Day Peak period calculations for the Quasi-Newton method with three perceptron layers were the same as those of Quasi-Newton method with two perceptron layers.

$$y_{1_1} = \tanh (1.26 + (scaled_{X_1} * 0.34) + (scaled_{X_2} * -0.39) + (scaled_{X_3} * 0.03) + (scaled_{X_4} * -0.02) + (scaled_{X_5} * 0.07))$$

$$y_{1_2} = \tanh (0.32 + (scaled_{X_1} * 0.14) + (scaled_{X_2} * 0.03) + (scaled_{X_3} * 0.02) + (scaled_{X_4} * -0.01) + (scaled_{X_5} * 0.02))$$

$$y_{1_3} = \tanh (-1.69 + (scaled_{X_1} * -0.57) + (scaled_{X_2} * 0.8) + (scaled_{X_3} * 0.07) + (scaled_{X_4} * 0.04) + (scaled_{X_5} * -0.18))$$

$$y_{2_1} = \tanh (-0.44 + (y_{1_1} * -1.12) + (y_{1_2} * 0.87) + (y_{1_3} * -0.54))$$

$$Scaled_{Y_{Mid-Day}} = (0.1 + (y_{2_1} * 1.62))$$

Equations for Quasi-Newton (Three Perceptron Layers) PM Model

$$y_{1_1} = \tanh (-0.25 + (scaled_{X_1} * 0.26) + (scaled_{X_2} * -1.1) + (scaled_{X_3} * 0.12) + (scaled_{X_4} * -0.09) + (scaled_{X_5} * 0.8))$$

$$y_{1_2} = \tanh (0.15 + (scaled_{X_1} * -1.19) + (scaled_{X_2} * -0.08) + (scaled_{X_3} * -0.89) + (scaled_{X_4} * 0.08) + (scaled_{X_5} * 0.68))$$

$$y_{1_3} = \tanh (0.03 + (scaled_{X_1} * 0.28) + (scaled_{X_2} * -0.19) + (scaled_{X_3} * 0.14) + (scaled_{X_4} * -0.02) + (scaled_{X_5} * 0.41))$$

$$y_{2_1} = \tanh (0.45 + (y_{1_1} * 2.69) + (y_{1_2} * 0.03) + (y_{1_3} * -0.82))$$

$$y_{2_2} = \tanh (-0.11 + (y_{1_1} * -0.71) + (y_{1_2} * 0.41) + (y_{1_3} * 1.17))$$

$$y_{2_3} = \tanh (-0.32 + (y_{1_1} * -1.98) + (y_{1_2} * 0.31) + (y_{1_3} * 1.17))$$

$$y_{2_4} = \tanh (-0.59 + (y_{1_1} * -1.84) + (y_{1_2} * -0.59) + (y_{1_3} * -1.14))$$

$$y_{2_5} = \tanh (2.47 + (y_{1_1} * -0.1) + (y_{1_2} * 0.04) + (y_{1_3} * -2.22))$$

$$y_{2_6} = \tanh (0.55 + (y_{1_1} * 1.09) + (y_{1_2} * 0.69) + (y_{1_3} * 1.52))$$

$$y_{2_7} = \tanh (1.44 + (y_{1_1} * -0.14) + (y_{1_2} * -1.19) + (y_{1_3} * -1.27))$$

$$y_{2_8} = \tanh (-2.5 + (y_{1_1} * -2.04) + (y_{1_2} * 1.92) + (y_{1_3} * 2.29))$$

$$y_{2_9} = \tanh (-0.28 + (y_{1_1} * -1.63) + (y_{1_2} * 1.43) + (y_{1_3} * 1.34))$$

$$y_{2_10} = \tanh(-2.36 + (y_{1_1} * -2.65) + (y_{1_2} * 1.57) + (y_{1_3} * 2.56))$$

$$\begin{aligned} Scaled_{Y_{PM}} = & (0.68 + (y_{2_1} * 1.35) + (y_{2_2} * -1.29) + (y_{2_3} * 2.40) + (y_{2_4} * 1.27) \\ & + (y_{2_5} * -1.08) + (y_{2_6} * 1.47) + (y_{2_7} * -0.89) + (y_{2_8} * -1.32) \\ & + (y_{2_9} * -0.49) + (y_{2_10} * 1.01)) \end{aligned}$$

Equations for Quasi-Newton (Five Perceptron Layers) AM Model

The scaled values of X_1 to X_5 used in the AM Peak period calculations for the Quasi-Newton methods with five perceptron layers were same as the scaled values used for AM Peak period calculations with two and three perceptron layers.

$$y_{1_1} = \tanh(-0.99 + (scaled_{X_1} * 0.17) + (scaled_{X_2} * -0.05) + (scaled_{X_3} * -0.29) + (scaled_{X_4} * -0.05) + (scaled_{X_5} * -0.52))$$

$$y_{1_2} = \tanh(0.05 + (scaled_{X_1} * -0.01) + (scaled_{X_2} * -0.09) + (scaled_{X_3} * -0.04) + (scaled_{X_4} * -0.01) + (scaled_{X_5} * -0.07))$$

$$y_{1_3} = \tanh(-1.9 + (scaled_{X_1} * -0.84) + (scaled_{X_2} * -2.99) + (scaled_{X_3} * 0.31) + (scaled_{X_4} * 0.01) + (scaled_{X_5} * -2.37))$$

$$y_{2_1} = \tanh(0.28 + (y_{1_1} * 1.31) + (y_{1_2} * -0.69) + (y_{1_3} * -0.35))$$

$$y_{2_2} = \tanh(-0.55 + (y_{1_1} * 1.04) + (y_{1_2} * 0.41) + (y_{1_3} * 1.17))$$

$$y_{2_3} = \tanh(0.29 + (y_{1_1} * 0.02) + (y_{1_2} * 0.65) + (y_{1_3} * 0.01))$$

$$y_{3_1} = \tanh(0.01 + (y_{2_1} * -0.09) + (y_{2_2} * -0.48) + (y_{2_3} * 0.45))$$

$$y_{3_2} = \tanh(0.16 + (y_{2_1} * -1.15) + (y_{2_2} * 0.63) + (y_{2_3} * 0.59))$$

$$y_{3_3} = \tanh(0.13 + (y_{2_1} * -0.01) + (y_{2_2} * 0.14) + (y_{2_3} * 0.32))$$

$$y_{4_1} = \tanh(-0.05 + (y_{3_1} * 0.9) + (y_{3_2} * 1.32) + (y_{3_3} * 0.29))$$

Equations for Quasi-Newton (Five Perceptron Layers) Mid-Day Model

$$scaled_{X_1} = \frac{X_1 - 3.91}{2.70}$$

$$scaled_{X_4} = \frac{X_4 - 15.49}{10.96}$$

$$scaled_{X_2} = \frac{X_2 - 6344.63}{4255.46}$$

$$scaled_{X_5} = \frac{X_5 - 16.62}{10.82}$$

$$scaled_{X_3} = \frac{X_3 - 12.78}{6.12}$$

$$\begin{aligned}
y_{1_1} &= \tanh(-0.84 + (\text{scaled_}X_1 * -0.27) + (\text{scaled_}X_2) + (\text{scaled_}X_3 * 0.09) \\
&\quad + (\text{scaled_}X_4 * -0.06) + (\text{scaled_}X_5 * -0.79)) \\
y_{1_2} &= \tanh(1.65 + (\text{scaled_}X_1 * 0.38) + (\text{scaled_}X_2 * -1.54) + (\text{scaled_}X_3 * -0.4) \\
&\quad + (\text{scaled_}X_4 * 0.16) + (\text{scaled_}X_5 * 1.66)) \\
y_{1_3} &= \tanh(0.38 + (\text{scaled_}X_1 * 0.02) + (\text{scaled_}X_2 * 0.19) + (\text{scaled_}X_3 * 0.01) \\
&\quad + (\text{scaled_}X_4 * -0.01) + (\text{scaled_}X_5 * -0.08)) \\
y_{2_1} &= \tanh(0.6 + (y_{1_1} * 1.61) + (y_{1_2} * 1.76) + (y_{1_3} * -0.87)) \\
y_{2_2} &= \tanh(0.2 + (y_{1_1} * 1.03) + (y_{1_2} * 0.04) + (y_{1_3} * -0.19)) \\
y_{2_3} &= \tanh(0.42 + (y_{1_1} * 0.29) + (y_{1_2} * -0.53) + (y_{1_3} * -0.62)) \\
y_{3_1} &= \tanh(0.41 + (y_{2_1} * -1.37) + (y_{2_2} * 0.22) + (y_{2_3} * 0.42)) \\
y_{3_2} &= \tanh(0.21 + (y_{2_1} * -1.42) + (y_{2_2} * -0.14) + (y_{2_3} * 0.36)) \\
y_{3_3} &= \tanh(0.12 + (y_{2_1} * 0.88) + (y_{2_2} * -0.69) + (y_{2_3} * 1.36)) \\
y_{4_1} &= \tanh(0.12 + (y_{3_1} * 1.27) + (y_{3_2} * -1.67) + (y_{3_3} * -1.68))
\end{aligned}$$

Equations for Quasi-Newton (Five Perceptron Layers) PM Model

$$\begin{aligned}
y_{1_1} &= \tanh(0.13 + (\text{scaled_}X_1 * 0.05) + (\text{scaled_}X_2 * 0.05) + (\text{scaled_}X_3 * 0.02) \\
&\quad + (\text{scaled_}X_4 * 0.01) + (\text{scaled_}X_5 * 0.11)) \\
y_{1_2} &= \tanh(3.06 + (\text{scaled_}X_1 * 2.38) + (\text{scaled_}X_2 * -0.75) + (\text{scaled_}X_3 * 1.67) \\
&\quad + (\text{scaled_}X_4 * -0.15) + (\text{scaled_}X_5 * -2.21)) \\
y_{1_3} &= \tanh(0.39 + (\text{scaled_}X_1 * -0.05) + (\text{scaled_}X_2 * 0.79) + (\text{scaled_}X_3 * 0.07) \\
&\quad + (\text{scaled_}X_4 * 0.09) + (\text{scaled_}X_5 * -0.59)) \\
y_{2_1} &= \tanh(-0.38 + (y_{1_1} * -0.99) + (y_{1_2} * -0.92) + (y_{1_3} * 2.04)) \\
y_{2_2} &= \tanh(-0.09 + (y_{1_1} * -0.39) + (y_{1_2} * -0.99) + (y_{1_3} * 3.53)) \\
y_{2_3} &= \tanh(0.09 + (y_{1_1} * -0.53) + (y_{1_2} * -0.04) + (y_{1_3} * 0.13)) \\
y_{3_1} &= \tanh(-1.19 + (y_{2_1} * -3.39) + (y_{2_2} * 2) + (y_{2_3} * -0.6)) \\
y_{3_2} &= \tanh(-1.02 + (y_{2_1} * -1.22) + (y_{2_2} * 0.85) + (y_{2_3} * 0.97)) \\
y_{3_3} &= \tanh(-1.03 + (y_{2_1} * -0.07) + (y_{2_2} * 0.38) + (y_{2_3} * -1.04)) \\
y_{4_1} &= \tanh(-0.03 + (y_{3_1} * 0.64) + (y_{3_2} * 0.48) + (y_{3_3} * -0.02)) \\
y_{4_2} &= \tanh(0.35 + (y_{3_1} * 0.59) + (y_{3_2} * -0.85) + (y_{3_3} * 2.01))
\end{aligned}$$

Equations for Levenberg-Marquardt (Three Perceptron Layers) AM Model

The scaled values of X_1 to X_5 used in the AM Peak period calculations for the Levenberg-Marquardt method were same as the scaled values used for AM Peak period calculations using the Quasi-Newton methods.

$$y_{1_1} = \tanh (1.03 + (\text{scaled_}X_1 * -0.66) + (\text{scaled_}X_2 * 3.64) + (\text{scaled_}X_3 * 0.75) + (\text{scaled_}X_4 * -0.15) + (\text{scaled_}X_5 * -1.13))$$

$$y_{1_2} = \tanh (0.35 + (\text{scaled_}X_1 * -0.28) + (\text{scaled_}X_2 * 0.21) + (\text{scaled_}X_3 * 1.98) + (\text{scaled_}X_4 * -0.44) + (\text{scaled_}X_5 * 0.1))$$

$$y_{1_3} = \tanh (-0.15 + (\text{scaled_}X_1 * 0.1) + (\text{scaled_}X_2 * 0.01) + (\text{scaled_}X_3 * 0.01) + (\text{scaled_}X_4 * 0.01 + (\text{scaled_}X_5 * 0.04)))$$

$$y_{2_1} = \tanh (0.53 + (y_{1_1} * -0.04) + (y_{1_2} * -0.01) + (y_{1_3} * 1.48))$$

$$y_{2_2} = \tanh (3.39 + (y_{1_1} * -0.25) + (y_{1_2} * 0.06) + (y_{1_3} * 0.98))$$

$$y_{2_3} = \tanh (-0.96 + (y_{1_1} * 0.39) + (y_{1_2} * -0.09) + (y_{1_3} * 1.48))$$

$$y_{2_4} = \tanh (1.71 + (y_{1_1} * -2.98) + (y_{1_2} * 3.37) + (y_{1_3} * 0.57))$$

Equations for Levenberg-Marquardt (Three Perceptron Layers) Mid-Day Model

The scaled values of the inputs X_1 through X_5 were the same as the scaled inputs used for the Mid-Day Peak period Quasi-Newton analysis with five perceptron layers.

$$y_{1_1} = \tanh (-0.12 + (\text{scaled_}X_1 * 0.11) + (\text{scaled_}X_2 * 0.12) + (\text{scaled_}X_3 * 0.01) + (\text{scaled_}X_4 * -0.01) + (\text{scaled_}X_5 * 0.05))$$

Equations for Levenberg-Marquardt (Three Perceptron Layers) PM Model

$$y_{1_1} = \tanh (-0.09 + (\text{scaled_}X_1 * 0.49) + (\text{scaled_}X_2 * -1.78) + (\text{scaled_}X_3 * -0.81) + (\text{scaled_}X_4 * -0.08) + (\text{scaled_}X_5 * 0.18))$$

$$y_{1_2} = \tanh (-0.15 + (\text{scaled_}X_1 * -1.27) + (\text{scaled_}X_2 * 2.79) + (\text{scaled_}X_3 * 0.02) + (\text{scaled_}X_4 * -0.06) + (\text{scaled_}X_5 * -0.39))$$

$$y_{1_3} = \tanh (-0.09 + (\text{scaled_}X_1 * 0.06) + (\text{scaled_}X_2 * 0.05) + (\text{scaled_}X_3 * 0.14) + (\text{scaled_}X_4 * 0.07 + (\text{scaled_}X_5 * 0.29)))$$

$$y_{1_4} = \tanh (0.61 + (\text{scaled_}X_1 * 1.59) + (\text{scaled_}X_2 * -0.53) + (\text{scaled_}X_3 * -0.3) + (\text{scaled_}X_4 * -0.17) + (\text{scaled_}X_5 * -0.97))$$

$$y_{1_5} = \tanh (2.58 + (\text{scaled_}X_1 * 2.04) + (\text{scaled_}X_2 * 1.04) + (\text{scaled_}X_3 * -0.78) + (\text{scaled_}X_4 * 0.48 + (\text{scaled_}X_5 * 1.87)))$$

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Founder, Honorable Norman Mineta*
Secretary (ret.),
US Department of Transportation

Chair, Abbas Mohaddes
President & COO
Econolite Group Inc.

Vice Chair, Will Kempton
Retired Transportation Executive

Executive Director, Karen Philbrick, PhD*
Mineta Transportation Institute
San José State University

Winsome Bowen
Chief Regional Transportation
Strategy
Facebook

David Castagnetti
Co-Founder
Mehlman Castagnetti
Rosen & Thomas

Maria Cino
Vice President
America & U.S. Government
Relations Hewlett-Packard Enterprise

Grace Crunican**
Owner
Crunican LLC

Donna DeMartino
Managing Director
Los Angeles-San Diego-San Luis
Obispo Rail Corridor Agency

John Flaherty
Senior Fellow
Silicon Valley American
Leadership Form

William Flynn *
President & CEO
Amtrak

Rose Guilbault
Board Member
Peninsula Corridor
Joint Powers Board

Ian Jefferies*
President & CEO
Association of American Railroads

Diane Woodend Jones
Principal & Chair of Board
Lea + Elliott, Inc.

David S. Kim*
Secretary
California State Transportation
Agency (CALSTA)

Therese McMillan
Executive Director
Metropolitan Transportation
Commission (MTC)

Jeff Morales
Managing Principal
InfraStrategies, LLC

Dan Moshavi, PhD*
Dean, Lucas College and
Graduate School of Business
San José State University

Toks Omishakin*
Director
California Department of
Transportation (Caltrans)

Takayoshi Oshima
Chairman & CEO
Allied Telesis, Inc.

Paul Skoutelas*
President & CEO
American Public Transportation
Association (APTA)

Beverley Swaim-Staley
President
Union Station Redevelopment
Corporation

Jim Tymon*
Executive Director
American Association of
State Highway and Transportation
Officials (AASHTO)

* = Ex-Officio

** = Past Chair, Board of Trustees

Directors

Karen Philbrick, PhD
Executive Director

Hilary Nixon, PhD
Deputy Executive Director

Asha Weinstein Agrawal, PhD
Education Director
National Transportation Finance
Center Director

Brian Michael Jenkins
National Transportation Security
Center Director

