

APPLICATION OF ELECTRONIC TICKET MACHINE DATA FOR ANALYSIS AND FORECASTING OF BUS TRANSPORT DEMAND

Thesis

Submitted in partial fulfilment of the requirements of the degree of
DOCTOR OF PHILOSOPHY

by

ANILA CYRIL

(Register No. 155100CV15F02)

Under the guidance of

Dr. Raviraj H. Mulangi

and

Prof. Varghese George



**DEPARTMENT OF CIVIL ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA
SURATHKAL, MANGALORE – 575025
OCTOBER 2019**

DECLARATION

I hereby declare that the Research Thesis entitled “ Application of Electronic Ticket Machine Data for Analysis and Forecasting of Bus Transport Demand” which is being submitted to the National Institute of Technology Karnataka, Surathkal in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy in Department of Civil Engineering is a bonafide report of the research work carried out by me. The material contained in this Research Thesis has not been submitted to any University or Institution for the award of any degree.

Place: NITK Surathkal

Anila Cyril

Date: 19-10-2019

Register Number: 155100CV15F02

Department of Civil Engineering

CERTIFICATE

This is to certify that the Research Thesis entitled “Application of Electronic Ticket Machine Data for Analysis and Forecasting of Bus Transport Demand” submitted by Anila Cyril (Register Number: 155100CV15F02) as the record of the research work carried out by her, is accepted as the Research Thesis submission in partial fulfillment of the requirements for the award of degree of Doctor of philosophy.

Dr. Raviraj H. Mulangi
(Research Supervisor)
Assistant Professor
Department of Civil Engineering

Dr. Varghese George
(Research Supervisor)
Professor
Department of Civil Engineering

Dr. K. Swaminathan
(Chairman-DRPC)
Head of the Department
Department of Civil Engineering

ACKNOWLEDGEMENT

The satisfaction of successful completion of any task could not be complete without acknowledging those persons whose continuous support and guidance made our efforts a successful one.

I express my deep sense of sincere gratitude to my research guides **Dr. Varghese George**, Professor and **Dr. Raviraj H. Mulangi**, Assistant Professor, Department of Civil Engineering, for the zeal with which they guided me in carrying out my thesis work. I am deeply indebted to them for their patience, systematic guidance, valuable advice, and constant motivation throughout this thesis work. The guidance of my research guides have helped me throughout in my research and in writing of this thesis.

I would like to thank the members of my Research Progress Assessment Committee- **Dr. Mrityunjay Doddamani**, Department of Mechanical Engineering and **Dr. B. Manu**, Department of Civil Engineering, for their insightful suggestions, comments and encouragement which has helped me to widen my research.

I am grateful to **Dr. K. Swaminathan**, Professor and Head, Department of Civil Engineering for providing necessary facilities for doing this work. I would like to express my gratitude to all the **Faculty Members** and **non-teaching staff** of Department of Civil Engineering for their sincere support, valuable guidance and constant encouragement throughout the course of my thesis work.

I also extend my deep sense of gratitude to the staffs of **Kerala SRTC** Transport Bhavan Trivandrum, Department of Town and Country Planning, Directorate of Census Operations and National Transportation Planning and Research Centre (NATPAC), for providing the necessary data and information for the study. Also, I would like to thank **Mrs. J. Mercykutty Amma**, Minister for Fisheries, Harbour Engineering and Cashew Industry, Government of Kerala, for her support and timely help during the data collection. I also extend my sincere gratitude to **Divisional Controller** and **staffs of Karnataka SRTC** for providing me additional data required for my study.

I would like to thank my research team, **Panditharadhya B.J.** and **Harsha M.M.**, for the enlightening discussions and suggestions which has shaped my thesis. I would like to thank my fellow Ph.D. scholars and my best friends **Anaswara S.** and **Spoorthi S.K.** for their feedback, constant motivation and above all for their friendship and fun we had in the last four years. I would also like to thank my friends **Gayatri Venkitesh**, **Amrutha G.S.**, **Nithin N.V.**, **Hina Chauhan**, **Praseetha P.**, **Alice Boruah**, **Nikhitha Zacharia** and **Aparna K.**, for their help and constant support throughout my research work.

My acknowledgement would be incomplete without thanking my biggest source of my strength, my family. The encouragement and motivation of my husband Mr. **Binu Solomon**, the love and care of my parents Mr. **Cyril A.** and Mrs. **Elizabeth Cyril**, the motivation from my mother in law Mrs. **Annie Joseph**, my brothers **Arun Cyril** and **Manu Soloman**, sister **Kabani Dinesh** and all my friends for their valuable help and encouragement throughout the period of my thesis work.

Above all, I am ever thankful to **Almighty God** for providing me the opportunity and granting me the capability to do this work.

ANILA CYRIL

ABSTRACT

Passenger demand estimation and forecasting is the fundamental process in public transport planning activity. The optimised operational planning of public transport requires the accurate prediction of the passenger flow on a periodical basis. This research tries to develop an intra-city and inter-city passenger demand estimation model using the data-driven time series approach. In this study, the source of time-series data is the Electronic Ticket Machine (ETM) data, which was not previously explored for passenger demand modelling and forecasting by the researchers. The data used for the study are mainly ETM data from Kerala State Road Transport Corporation (Kerala SRTC). During the analysis of the Kerala ETM data, it was found that the length of intra-city ETM data (for Trivandrum city) available for individual routes varies from 1.5 years to 2 years (2011 to 2013), while inter-city data were not sufficient for time-series modelling. Therefore, inter-city ETM data of Mangalore division Karnataka State Road Transport Corporation (Karnataka SRTC) was collected for a period of 2013 to 2018, which comprises of a total five-year data, to perform inter-city passenger demand modelling. The intra-city ETM data for Mangalore was not collected since the intra-city services are mainly operated by private operators. The private operators have not maintained and stored the ETM data.

Since the travel demand is affected by the performance of the public transport system and the bus transport agencies operating the services, this study evaluates the performance of the public transport system and the public transport operator. Therefore, in addition to ETM data, this study also uses data provided by various state and central government agencies. The level of service offered by the public transport system in Trivandrum city, the accessibility of public transport and the performance evaluation of Kerala SRTC was performed to evaluate the present condition of the public transport system in Trivandrum and Kerala SRTC. The performance evaluation of Karnataka SRTC was not performed since the data collected from Karnataka was only for modelling inter-city passenger demand. The Level-of-Service (LoS) was determined using the service level benchmarks set by the Ministry of Urban Development, Government of India. This approach can be used to determine the LoS at the city-level and thus provides a measure for identifying the public transport quality of an urban

area. The overall LoS is found to be 'one'. This indicates that the city has good service coverage, i.e., the public transport ply on most of the corridors in the urban area. The facilities are available wide-spread and are available to the people. This study proposed a methodology for the determination of the accessibility index, which is based on the factor that the index should measure the accessibility which comes from proximity to bus stops and land use destinations, and the proportion of the population served. It was observed that the accessibility index in the central area has a higher value and the accessibility values decrease to the periphery of the city. It can be interpreted that the availability of opportunities is higher in the central area, while the number of opportunities is less in other areas. The density of the road network is more towards the city centre, which contributes to higher accessibility values. Also, given that the concentration of population is high in the city centre and the higher number of bus stops, the accessibility index tends to be more in these areas. The performance evaluation of Kerala SRTC was performed using the methodology developed in this study. A weighted Goal Programming (GP) methodology integrated with Analytical Hierarchy Process (AHP) considering the user's and operator's perception was used. The operator cost and staff per schedule were the most important variables in regard to the operator, while the safety of travel with respect to user perception. The optimal solution indicates that increasing the accessibility, safety and regularity attract passengers to the public transport that in turn improves the load factor which influences the operators to maximise the fleet utilisation and reduce the cancellation of schedules. The proposed model solution suggests decreasing the staff per bus that will further reduce the staff cost and hence the operating cost.

The correlation analysis of ETM data revealed that the relevant temporal pattern in the time-series data (for both intra-city and inter-city) was daily time-series pattern, which was further taken into account during the modelling process. Also, in the data analysis, it was found that 85% of the intra-city data was nonstationary, while 100% of the data for inter-city was nonstationary. The nonstationary data were differenced to make it stationary for further time series analysis. Only 30% of the intra-city data was having seasonal effect while 100% of the inter-city data exhibits seasonality. It is due to the fact that the seasonal variations are not that prevalent in urban travel demand. But inter-

city travel is affected by seasonal variations. 95% and 99% of intra-city and inter-city data respectively were non-normal. Therefore, data transformation was required to make these data normal for further analysis. 90% of the intra-city routes were having nonlinearity while all the routes of the inter-city were nonlinear. Hence, nonlinear modelling was required to model it.

A methodology to develop an empirical passenger demand estimation model using the data-driven time series approach, employing the capabilities of AutoRegressive Integrated Moving Average (ARIMA) method is developed. The ARIMA model is suitable when dealing with a nonstationary time-series data. But, it cannot be applied to data sets having volatility. When the time-series data has heteroscedasticity, i.e., the variances of error terms are not equal, or some of the error terms are reasonably large, AutoRegressive Conditional Heteroscedasticity (ARCH) models are used for modelling the data sets. When the data have non-standard features like non-normal distribution, and non-linear relationship with lagged values require non-linear AutoRegressive Neural Network (ARNN) modelling methods. The methodology to improve the developed model using the nonlinear ARCH and ARNN methods are determined. It includes a comprehensive methodology for model selection and forecasting, which requires only basic econometric knowledge. The empirical results were compared with the actual value for 30 days and 7 days forecast horizon; and found that the time series model could predict the future observations with an acceptable forecast accuracy. The forecasting performance of developed models are measured using the Mean Absolute Percentage Error (MAPE), and the model goodness of fit is determined using information criteria. The lower value of MAPE represents a good forecast. Lewis (1982) describes MAPE greater than 50% as inaccuracy in forecasting, 20-50% as reasonable forecast, 10-20% as good forecast and less than 10% as highly accurate forecast. In this study, the MAPE values for the 30 days and 7 days forecast were less than 10%. This time-series methodology is useful when there is limited or no information available about the factors affecting the demand. Recommendations have been made to improve the performance of bus operating organisations.

Keywords: ETM data, AHP-GP, public transport demand, time-series, MAPE.

CONTENTS

DECLARATION.....	I
CERTIFICATE.....	II
ACKNOWLEDGEMENT.....	III
ABSTRACT.....	V
CONTENTS.....	VIII
LIST OF FIGURES	XII
LIST OF TABLES	XIV
LIST OF ABBREVIATIONS	XVI
CHAPTER 1.....	1
INTRODUCTION.....	1
1.1. General	1
1.2. Transport Sector in India-An Overview	3
1.3. Need for the Study.....	6
1.4. Objectives.....	8
1.5. Scope of the Work	9
1.6. Thesis Organisation	10
1.7. Major Contributions of this Research Work.....	11
CHAPTER 2.....	13
REVIEW OF LITERATURE.....	13
2.1. General	13
2.2. Data Collection	13
2.2.1. Manual Method or Field-Based Survey.....	15
2.2.2. Automated or Technology-Based Surveys	15
2.3. Level of Service.....	17
2.4. Accessibility of Public Transport.....	19
2.4.1. Concepts of Accessibility	20
2.4.2. Components of Accessibility	20
2.4.3. Measures of Accessibility	21
2.4.4. GIS Application to Accessibility	23
2.5. Performance Evaluation of Public Transport	24
2.6. Travel Demand Modelling.....	28

2.7. Public Transport Demand Modelling.....	31
2.7.1. Approaches to Estimate Public Transport Demand	31
2.7.2. Major Characteristics of Public Transport Demand	31
2.8. Historical Perspective of Public Transport Demand Modelling.....	32
2.8.1. Four-Step Model	33
2.8.2. Simultaneous Modelling Approach	36
2.8.3. Time-Series Techniques.....	47
2.9. Measures of Forecasting Performance.....	61
2.10. Research Gap.....	63
2.11. Summary	63
CHAPTER 3.....	67
STUDY AREA.....	67
3.1. General.....	67
3.2. Description of the Study City-Trivandrum	67
3.2.1. Trends in Land Use.....	71
3.2.2. Trends in Travel Behaviour	71
3.2.3. Trends in Public Transport.....	73
3.3. Description of the Study City-Mangalore	75
3.3.1. Trends in Land Use.....	77
3.3.2. Trends in Travel Behaviour	78
3.3.3. Trends in Public Transport.....	80
3.4. Summary	80
CHAPTER-4.....	83
METHODOLOGY AND DATA COLLECTION	83
4.1. General.....	83
4.2. Methodology	83
4.3. Data Collection	85
4.3.1. Land Use Data.....	87
4.3.2. Road Network Data.....	88
4.3.3. Bus Stop Data	90
4.3.4. Demographic Data	92
4.3.5. Physical and Financial Data of Kerala SRTC.....	92
4.3.6. Electronic Ticket Machine Data	94

4.4. Data Processing	95
4.4.1. Data from Kerala SRTC (Intra-City Data).....	95
4.4.2. Data from Karnataka SRTC (Inter-city Data).....	102
4.5. Determination of Level of Service.....	104
4.6. Determination of Accessibility Index	105
4.6.1. Network Development	105
4.6.2. Access to Nearest Bus Stop	106
4.6.3. Access from Bus Stops to the Point of Interests (POIs)	107
4.7. Performance Optimisation of Kerala SRTC	109
4.8. Determination of Variation Indices	113
4.9. Theoretical and Methodological Framework of Time-Series Modelling .	114
4.9.1. Model Identification.....	116
4.9.2. Parameter Estimation of ARIMA Model	125
4.9.3. Diagnostic Checking	126
4.9.4. AutoRegressive Neural Network Model.....	128
4.9.5. Hybrid ARIMA-ARNN Model.....	132
4.9.6. Hybrid ARIMA-GARCH Modelling.....	133
4.10. Summary	136
CHAPTER 5	139
RESULTS AND DISCUSSION	139
5.1. General	139
5.2. Level of Service.....	139
5.3. Accessibility Index.....	143
5.4. Performance Optimisation of Kerala SRTC	146
5.5. Variation Indices	156
5.5.1. Variation Indices of Intra-City Passenger Demand	156
5.5.2. Variation Indices of Inter-City Passenger Demand	159
5.6. Time-Series Analysis and Modelling for Intra-City Data	161
5.6.1. Exploratory Data Analysis	161
5.6.2. Histogram Overlaid with Kernel Density Curve	162
5.6.3. Outlier Analysis	163
5.6.4. Decomposition of Time-Series Data.....	164
5.6.5. Autocorrelation and Partial Autocorrelation Plot	165

5.6.6. Statistical Tests	166
5.6.7. ARIMA Model.....	168
5.6.8. Forecasting Bus Passenger Demand	174
5.7. Autoregressive Neural Network (ARNN) Model.....	176
5.8. Hybrid ARIMA-ARNN Model.....	178
5.9. Hybrid ARIMA-GARCH Model	180
5.9.1. Data Analysis and Model Fitting	180
5.9.2. GARCH Modelling.....	181
5.9.3. GARCH Model Forecast and Accuracy	183
5.10. Discussion on Intra-City and Inter-City Bus Passenger Demand Models	193
5.11. Summary	195
CHAPTER-6.....	199
CONCLUSION	199
6.1. General	199
6.2. Conclusion.....	199
6.2.1. Data Collection and Processing	199
6.2.2. Evaluation of Level of Service of Kerala SRTC	200
6.2.3. Determination of Accessibility Index	201
6.2.4. Performance Evaluation of Kerala SRTC.....	202
6.2.5. Determination of Variation Indices	202
6.2.6. Development of Time-Series Models	203
6.3. Recommendations	205
6.4. Scope of Future Work.....	206
APPENDIX.....	209
REFERENCES.....	261
LIST OF PUBLICATIONS	285
BIO-DATA.....	287

LIST OF FIGURES

Figure 1.1. Total Number of Registered Vehicles in India from 1951-2015 (Source: MoRTH 2015).....	4
Figure 1.2. Composition of Vehicle Population of India from 1951- 2015 (Source: MoRTH 2015).....	5
Figure 2.1. Flowchart Summarising the Literature Review.....	14
Figure 2.2. The Designed Travel Demand Forecast Method.....	34
Figure 3.1. Map of Trivandrum City	69
Figure 3.2. Road Network Map of Trivandrum City	70
Figure 3.3. Distribution of Trips According to Mode of Travel in Trivandrum.....	72
Figure 3.4. Distribution of Trips According to Trip Purpose in Trivandrum	72
Figure 3.5. Map of Mangalore City	76
Figure 3.6. Distribution of Trips According to Mode of Travel in Mangalore	79
Figure 3.7. Distribution of Trips According to Trip Purpose in Mangalore.....	79
Figure 4.1. General Process of Transit Planning (Source: Adapted and Modified from Carrel 2009)	84
Figure 4.2. Flowchart of Methodology	86
Figure 4.3. Existing Land Use Map of Trivandrum City.....	88
Figure 4.4. Road Network Map of Trivandrum City	89
Figure 4.5. Map Showing the Bus Stop Locations	91
Figure 4.6. The Raw ETM Data.....	95
Figure 4.7. Flow Chart for Data Extraction	96
Figure 4.8. Excel Converted ETM Data	98
Figure 4.9. The Procedure for Date-Stamping the Excel Files	99
Figure 4.10. Final Recompiled ETM Data.....	100
Figure 4.11. Time-Series Data Generated from Kerala SRTC ETM Data	102
Figure 4.12. Sample ETM Data from Karnataka SRTC	103
Figure 4.13. Time-Series Data Generated from Karnataka SRTC ETM Data	104
Figure 4.14. Bus Stops with 500m Buffer Overlaid on Population Density Map	106
Figure 4.15. Distribution of Point of Interests in the City	107
Figure 4.16. Performance Optimisation Methodology Integrating User and Operator Perception	110
Figure 4.17. Methodological Framework for Time-Series Analysis and Modelling.....	115
Figure 4.18. The Methodology of ARNN Model	131

Figure 4.19. The Methodology of ARIMA-GARCH Model.....	134
Figure 5.1. Accessibility Index Map of Trivandrum Municipal Corporation.....	145
Figure 5.2 Hierarchy of Performance Optimisation Process	147
Figure 5.3. Daily Variation Index of Six Depots of Trivandrum.....	156
Figure 5.4. Quarterly Variation Index of Trivandrum	157
Figure 5.5. Monthly Variation Index of Trivandrum.....	158
Figure 5.6. Weekly Variation Index of Trivandrum	158
Figure 5.7. Year-wise Daily Variation Index of Mangalore	159
Figure 5.8. Quarterly Variation Index of Mangalore	160
Figure 5.9. Monthly Variation Index of Mangalore	160
Figure 5.10. Weekly Variation Index of Mangalore.....	160
Figure 5.11 Time-Series Plot of Number of Passengers for Vizhinjam Route 137... 162	
Figure 5.12. Histogram Overlaid with Kernel Density Curve	163
Figure 5.13. Histogram Overlaid with Kernel Density Curve	164
Figure 5.14. Time-Series Plot After the Removal of the Outliers	164
Figure 5.15. Trend Decomposition Plot.....	165
Figure 5.16. Autocorrelation Plot of Time-Series Data.....	165
Figure 5.17. Partial Autocorrelation Plot of Time-Series Data	166
Figure 5.18. ACF and PACF Plot of Seasonally Differenced Time-Series Data	168
Figure 5.19. ACF and PACF Plot of Differenced Time-Series Data	168
Figure 5.20. ACF and PACF Plot of the Residual – 30 Days.....	170
Figure 5.21. ACF and PACF Plot of the Residual – 7 Days.....	170
Figure 5.22. Residual Plot for 30 Days.....	171
Figure 5.23. Residual Plot for 7 Days.....	173
Figure 5.24. Squared Residuals plot	182
Figure 5.25. ACF and PACF plot of Squared residuals.....	182
Figure 5.26. Result Summary of Intra-City (Kerala SRTC) Forecast Accuracy	190
Figure 5.27. Result Summary of Inter-City (Karnataka SRTC) Forecast Accuracy .	192

LIST OF TABLES

Table 1.1. Passenger Kilometers from 1951-2016.....	4
Table 3.1. Census Data of Trivandrum City	68
Table 3.2. Distribution of Trips According to Mode of Travel in Trivandrum	72
Table 3.3. Distribution of Trips According to Trip Purpose in Trivandrum	73
Table 3.4. Number of Private Buses and Kerala SRTC Buses in Trivandrum District in the Year 2014 and 2015	74
Table 3.5. Major Indicators Showing Operational Efficiency of Kerala SRTC	75
Table 3.6. Census Data of Mangalore City.....	75
Table 3.7. Distribution of Trips According to Mode of Travel in Mangalore.....	78
Table 3.8. Distribution of Trips According to Trip Purpose in Mangalore	79
Table 3.9. Major Indicators Showing Operational Efficiency of Karnataka SRTC for the Years 2012-13 to 2017-18.....	80
Table 4.1. District Wise Number of Private and Kerala SRTC Buses 2014 and 2015	93
Table 4.2. Speed and Fare Characteristics of Each Type of Bus.....	94
Table 4.3. Sample Origin-Destination Matrix of Trip 1 of Route 1	101
Table 4.4. Weightage Calculation Formula	108
Table 4.5. Weightage Values based on Rank Order Centroid Method.....	108
Table 4.6. Random Index Value	111
Table 5.1. Indicators to Calculate City-Wide Level-of-Service (LoS) of Public Transport Facilities	140
Table 5.2. Indicator 1. Presence of Organised Public Transport System in Urban Area	141
Table 5.3. Indicator 2. Availability of Public Transport.....	141
Table 5.4. Indicator 3. Service Coverage of Public Transport in the City.....	142
Table 5.5 Indicator 4. Average Waiting Time for Public Transport Users'	142
Table 5.6. Indicator 5. Level of Comfort in Public Transport	142
Table 5.7. Overall Level-of-Service of City-Wide Public Transport Facilities.....	143
Table 5.8. Accessibility Index Ranges and Categories	144
Table 5.9. Pairwise Comparison Matrix of the Stakeholders	148
Table 5.10. Pairwise Comparison Matrix of the User-Oriented Decision Variables.	148
Table 5.11. Pairwise Comparison Matrix of the Operator-Oriented Decision Variables	148
Table 5.12. Final Weights of Decision Variables	149

Table 5.13. Threshold Values for Decision Variables.....	149
Table 5.14. Optimal and Actual Values for Decision Variables.....	153
Table 5.15. Sensitivity Analysis	155
Table 5.16. Daily Variation Index of the Six Depots of Trivandrum	157
Table 5.17. Quarterly Variation Index of the Six Depots of Trivandrum.....	157
Table 5.18. Monthly Variation Index of Trivandrum.....	158
Table 5.19. Results of the Statistical Tests of the Data	166
Table 5.20. Best Fit ARIMA Model – 30 Days.....	169
Table 5.21. Best Fit ARIMA Model – 7 Days.....	169
Table 5.22. Statistical Test Result of 30 Days Model Residual.....	170
Table 5.23. Statistical Test Result of 7 Days Model Residual.....	172
Table 5.24. Forecast Accuracy 30 Days	174
Table 5.25. Forecasted Values for a Forecast Horizon of 30 Days	174
Table 5.26. Forecast Accuracy 7 Days	175
Table 5.27. Forecasted Values for a Forecast Horizon of 7 Days	175
Table 5.28. ARNN Forecast Summary for 30 Days Forecast Horizon	176
Table 5.29. ARNN Forecasted Values for 30 Days Forecast Horizon	176
Table 5.30. ARNN Forecast Summary for 7 Days Forecast Horizon	177
Table 5.31. ARNN Forecasted Values for 7 Days Forecast Horizon	177
Table 5.32 Hybrid Model Accuracy	178
Table 5.33. Forecasted Values for 30-Days Forecast Horizon	179
Table 5.34. Forecasted Values for 7-Days Forecast Horizon	180
Table 5.35. Best Fit ARIMA Model for Vizhinjam Route3- 30 Days	181
Table 5.36. Summary of Data and Residual Analysis of ARIMA.....	181
Table 5.37. GARCH Model Summary	182
Table 5.38. GARCH Forecast.....	183
Table 5.39. Forecast Accuracy of ARIMA and Hybrid ARIMA-GARCH	184
Table 5.40. Result Summary of Intra-City (Kerala SRTC) Forecast Accuracy	185
Table 5.41. Result Summary of Inter-City (Karnataka SRTC) Forecast Accuracy...	191

LIST OF ABBREVIATIONS

ACF	Autocorrelation Function
AD	Anderson-Darling
ADC	Automatic Data Collection
ADF	Augmented Dickey Fuller
AFC	Automatic Fare Collection
AHP	Analytical Hierarchy Process
AIC	Akaike Information Criteria
AICc	Corrected Akaike Information Criteria
ANFIS	Adaptive Network-based Fuzzy Interference System
ANN	Artificial Neural Network
APC	Automatic Passenger Count
APTA	Area Public Transport Accessibility
AR	AutoRegressive
ARCH	AutoRegressive Conditional Heteroscedasticity
ARFIMA	AutoRegressive Fractionally Integrated Moving Average
ARIMA	AutoRegressive Integrated Moving Average
ARMA	AutoRegressive Moving Average
ARNN	AutoRegressive Neural Network
AVL	Automatic Vehicle Location
BCAU	Box-Cox AutoRegressive
BIC	Bayesian Information Criteria
BP	Back Propagation
BRT	Bus Rapid Transit
CBD	Central Business District
CR	Consistency Ratio
DEA	Data Envelopment Analysis
DES	Double Exponential Smoothing
DULT	Directorate of Urban Land Transport
EGANFIS	Expert Guided Adaptive Network-based Fuzzy Interference System

EGARCH	Exponential GARCH
etc	et cetera
ETM	Electronic Ticket Machine
FFBP	Feed Forward Back Propagation
FL	Fuzzy Logic
FSM	Four-Step Model
GARCH	Generalised ARCH
GC	Generalised Cost
GIS	Geographical Information System
GJR-GARCH	Glosten-Jagannathan-Runkle GARCH
GoI	Government of India
GP	Goal Programming
GRNN	Generalised Regression Neural Network
GWR	Geographically Weighted Regression
HSR	High Speed Rail
i.e.	That is
IPT	Intermediate Public Transport
ITS	Intelligent Transportation System
JB	Jarque-Bera
Kerala SRTC	Kerala State Road Transport Corporation
Karnataka SRTC	Karnataka State Road Transport Corporation
KPSS	Kwiatkowski-Philips-Schmidt-Shin
LM	Lagrange-Multiplier
LOESS	Local Polynomial Regression
LoS	Level-of-Service
LUPTAI	Land-Use and Public Transport Accessibility Index
MA	Moving Average
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MCDM	Multi-Criteria Decision-Making
MLE	Maximum Likelihood Estimation

MoRTH	Ministry of Road Transport and Highways
MoUD	Ministry of Urban Development
MPE	Mean Percentage Error
MSE	Mean Square Error
MUDA	Mangalore Urban Development Authority
NA	Not-Applicable
NAR	Non-linear AutoRegressive
NATPAC	National Transportation Planning and Research Centre
NH	National Highway
NMA	Non-linear Moving Average
NN	Neural Network
NNAR	Autoregressive Neural Network
OD	Origin-Destination
OECD	Organisation for Economic Co-operation and Development
PACF	Partial Autocorrelation Function
PAM	Partial Adjustment Model
POI	Point of Interest
PT	Public Transport
PTAL	Public Transport Accessibility Level
RBF	Radial Basis Function
RI	Random Index
ROC	Rank Order Centroid
RP	Revealed Preference
RMSE	Root Mean Square Error
SAR	Seasonal AutoRegressive
SARIMA	Seasonal ARIMA
SES	Simple Exponential Smoothing
SH	State Highway
SMA	Seasonal Moving Average
SP	Stated Preference
SPM	Spatial Proximity Method
SRTU	State Road Transport Undertakings

STEM	Simultaneous Transportation Equilibrium Model
STL	Seasonal and Trend Decomposition
SW	Shapiro-Wilk
SWM	Spatial Weight Method
TAR	Threshold AutoRegressive
TEN	Transportation Efficiency Number
TERI	The Energy and Resources Institute
TfL	Transport for London
TSP	Time-Series Processor
UPT	Urban Public Transport
Viz.	videlicet

CHAPTER 1

INTRODUCTION

1.1. GENERAL

Sustainable transportation has been identified by the United Nations as one of the goals in achieving sustainable development. Sustainable transportation endorses environmental benefits along with economic and social development for both today's and future generations. It provides safe, accessible, affordable and efficient mobility options for all segments of society. The concept of sustainable transport has been adopted by many countries in the national transport policy to reduce environmental risks and to provide 'mobility for all'. Among the various modes of transportation, public transport has been recognized as the primary form of sustainable transport which can meet the mobility requirements of the people. An efficient and reliable public transportation system can facilitate a modal shift from the private mode of transport to public transport. Therefore, it is required to develop a planned, monitored and regulated public transport system. Thus, the strategic planning of public transport is one of the best ways to achieve sustainable development through sustainable transport.

The knowledge of passenger demand is the fundamental requirement in the public transport planning process. The problem of estimating passenger demand is a complex activity as it has to be addressed at various levels. First, there is a demand at the regional level, which is beneficial for conceptual planning. Secondly, the demand is at the district or network level for route planning. The third and foremost level is the route level, which is essential for operational planning of public transport. The individual route level demand can be further aggregated to network level or regional level based on the requirements.

Researchers over the past have formulated many approaches for passenger demand estimation and forecasting, varying from simple regression to complex agent-based modelling techniques; the data collection of which are both expensive and time-consuming and therefore are extremely infrequent. In recent years, the research on demand forecasting is focused on identifying various Automated Data Collection

(ADC) techniques and analysis, which will curb the cost and time requirements of planning procedures. The use of the ADC system (Cui 2006) for daily transport operations has generated an enormous amount of time-series data that can be used for various applications. This readily available time-series data have proved to be useful in explaining and forecasting the demand for public transit (Wijeweera et al. 2014; Doti and Adibi 1991; Odgers and Schijndel 2011). Furthermore, data-driven techniques are found to be more accurate as it rules out the bias, and the data can be updated frequently.

Recently in India, most of the public transport operators have started using Electronic Ticketing Machines (ETM) for providing tickets to the passengers and for fare collection. The ETM is a handheld Automatic Fare Collection (AFC) device that records the data when a passenger purchases a ticket (Cyril et al. 2017). The ETMs generate a substantial amount of passenger flow information, which can be utilised for passenger demand prediction, long-term and short-term based on the forecast horizon. Long-term forecasting is for strategic planning, while short-term forecasting is for operational planning and scheduling. The long-term forecasting relies on passenger data along with various other influencing factors like population characteristics, income, the attractiveness of competing modes and land-use characteristics. On the other hand, the historical data of passenger flow is processed and analysed to predict short-term forecasts. The short-term passenger forecasting using the historical data is referred as time-series forecasting method.

Time-series is a dynamic field of research in the past few decades. It focuses on the collection of the time-dependent data, rigorous data processing, using right data mining tools to examine the past observations and to develop a model which explains the characteristics of the time-series data (Adhikari and Agrawal 2013). The concept of time-series is based on mathematical foundations and probability theory. Due to the inherent characteristic of time-series data; analysing the data and forecasting the future values are drawing the attention of researchers in the field of transportation planning.

The concept of time-series was put forward by Russell (1901) in the paper “Is position in Time and Space Absolute or Relative?”; which was further explained by McTaggart (1908 and 1909). The concept was extended with mathematical explanations

simultaneously by Yule (1921, 1926 and 1927) and Slutsky (1937). The first notable study on time-series was by Yule (1927) involving the periodogram analysis of time-series data to elicit the periodicity of the data. Later, intensive studies on time-series with an application in various fields were taken up by the researchers. One such application is in the field of passenger demand modelling and forecasting, which is the interest of this study. In order to develop a reliable short-term public transport passenger demand model, this study uses the ETM based passenger demand data employing the capabilities of the time-series method.

1.2. TRANSPORT SECTOR IN INDIA-AN OVERVIEW

The transport sector in India has extensively grown since independence in terms of both infrastructure and number of vehicles. Despite the tremendous growth, it fails to meet the mobility requirements of the people and goods. Like other developing countries, India is undergoing rapid urbanisation leading to increased travel demand. The rapid urbanisation and motorisation in India triggered the transport crisis in Indian cities, characterised by congestion, air pollution, noise, greenhouse gas emission, accidents, injuries, fatalities and mobility problems. As per census 2011, the level of urbanisation is 31.16% as against 27.81% in the 2001 census, showing a growth rate of 2.76% per annum during 2001-2011 (MoUD 2016). Private vehicles are increasing at an average rate of 12% (TERI 2015). The total number of registered vehicles in India from 1951-2015 is given in Figure 1.1, and the passenger kilometre performed 1951-2016 is given in Table 1.1. There are 53 State Road Transport Undertakings (SRTU) having a total number of 1.5 lakh buses of varying fleet size carrying more than 7 crore passengers per day performing about 501 billion passenger kilometres (Planning Commission 2012). Due to the slow-paced growth of infrastructure and insufficient public transportation, the urban transport system is throttled by the flowing traffic. The Government of India (GoI) has recognized the increasing urban transport problems along with its associated problems and has developed a strategy to increase the efficiency of road space by favouring public transport.

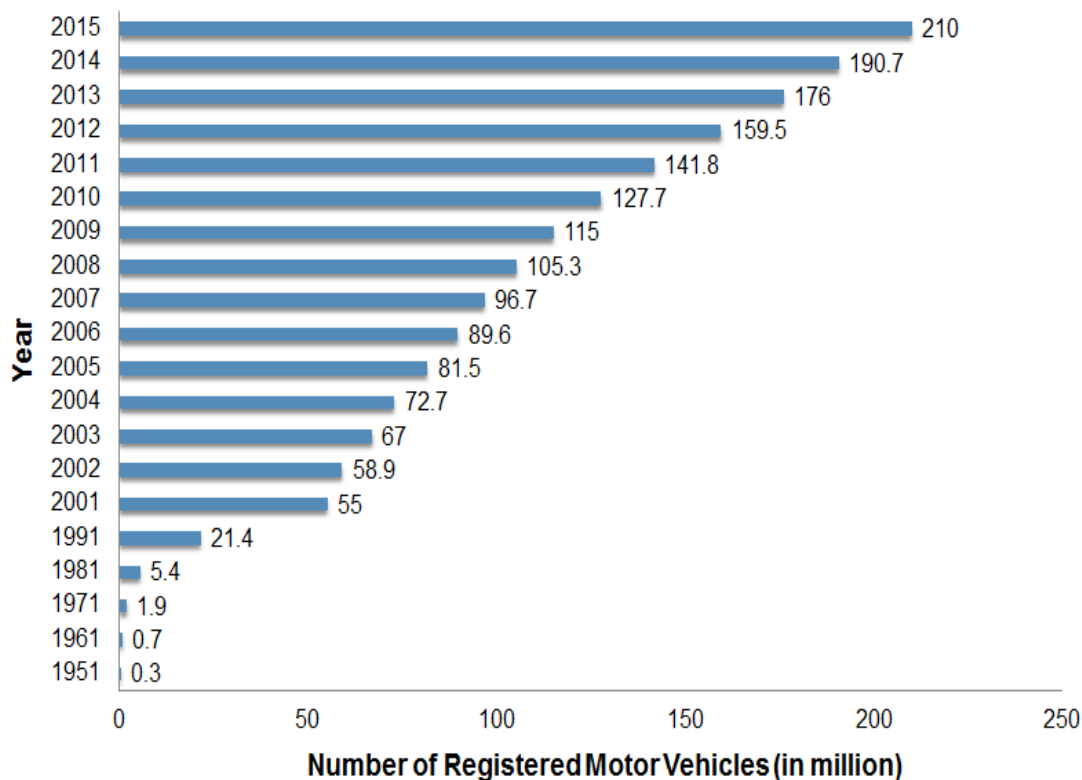


Figure 1.1. Total Number of Registered Vehicles in India from 1951-2015
(Source: MoRTH 2015)

Table 1.1. Passenger Kilometers from 1951-2016

Year	Passenger Kilometer (in billions)		
	Road	Rail	Total
1951	23.0	66.5	89.5
1961	80.9	77.7	158.6
1971	210.0	118.1	328.1
1981	541.8	208.6	750.4
1991	767.7	295.6	1063.3
2001	2075.5	457.0	2544.8
2011	9478*	1056.5*	10524.5*
2016	16950*	1158.6*	18108*

Source: Ministry of Railways (2015); *OECD (2017)

It has been estimated that the share of the road sector in total transport by road and rail are 64.5% of freight and the passenger traffic account for 85.9% in 2011-2012 as compared to 13.8% freight and 15.4% of passenger in 1950-51 (TERI 2015).

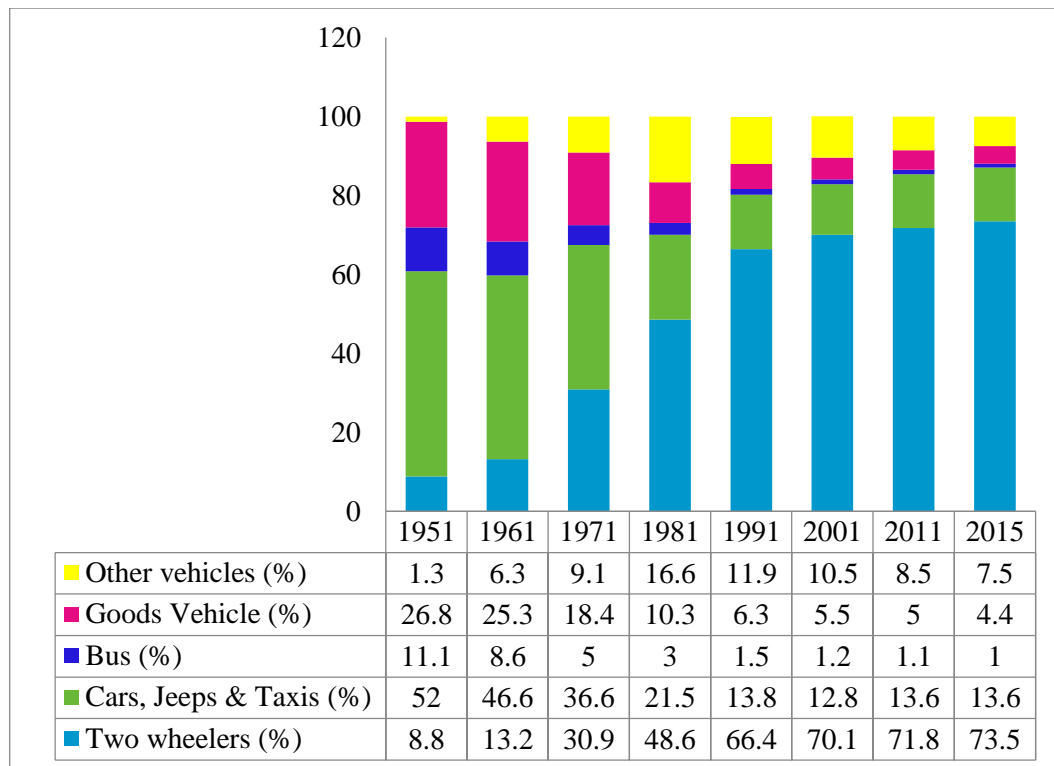


Figure 1.2. Composition of Vehicle Population of India from 1951- 2015 (Source: MoRTH 2015)

The composition of the vehicle population of India from 1951-2015 is given in Figure 1.2. The total vehicle population of India in 1951 was 0.3 million, which comprises of 8.8% two-wheelers, 52 % cars, 26.8% goods vehicles, 11.1% bus and the remaining 1.3% is other vehicles. In 2015, the total vehicle population had grown to 210 million, which comprises of 73.5% two-wheelers, 13.6 % cars, 4.4% goods vehicles, 1% bus and 7.5% other vehicles (tractors, trailers, three-wheelers/light motor vehicles and other miscellaneous vehicles which are not classified separately). It can be inferred from the data analysis that the percentage of two-wheelers have increased tremendously over the last few decades while the share of buses and goods vehicles are decreasing consistently. The percentage of cars has been decreasing till 2001, but from 2001 it is steadily increasing its share. It can be seen that there is an exponential growth of private mode of travel during this period, mainly during the last two decades. The increase in the personalized mode of transport has led to enormous problems of traffic congestion and pollution besides being a factor for the increased number of road accidents. The degeneration of the service quality of the public transport system and the rising level of

expectation of the passengers are identified as one of the reasons for the growth of the personalized mode of transport.

Public transport scenario varies from the aspects of financial issues to inefficiency and unreliability. The imprudent use of available financial resources prevents the necessary maintenance and upgrading of existing public transport led to the decline in public transport usage, especially the bus. The public transport services in most of the cities are ill-planned and uncoordinated. Also, low service frequency, low service quality, increased travel time and overcrowding are contributing to the issues of public transport. Excessively high operating cost, overstaffing, corruption and low productivity are the few institutional issues. The declining share of public transport coupled with infrastructure development that promotes car usage by building more roads and flyovers fuel the car-oriented development and thus the Indian cities are losing the required standard of living (Pucher et al. 2004).

In India, very few cities have organised, planned and regulated public transport system (Working Group of Urban Transport 2013). In the absence of sufficient public transport services, private run modes like auto rickshaws and other intermediate public transport have emerged in the cities. In cities where regular and formal public transport is absent, these intermediate modes serve the mobility needs of the people. This, in turn, has brought along the problems of increased congestion, travel time and air pollution. With declining sustainable modes like public transport and non-motorised transport, India is progressing in the wrong direction.

1.3. NEED FOR THE STUDY

Transportation has developed throughout history in substantially unique ways. From the horse-and-buggy to the invention of ships, automobiles, trains, aeroplanes, and high-speed metro-rails, humans have found ways to travel farther and faster. Since the beginning of civilization, the efficiency of the transportation system influences the economic success and sustainability of the society to a major extent. The tremendous growth in urban population and growing cities have resulted in increased travel demand. A good transportation system must be able to accommodate the transport demand. The

rapid urbanisation and motorisation triggered the transport crisis, characterised by congestion, air pollution, noise, greenhouse gas emission, accidents, injuries, fatalities and mobility problems.

Public transport is the key to solutions for major transportation problems. Estimates of public transport demand are essential to bus operators and planners. It is vital for the government as the public bus transportation is highly subsidised and should serve the society with sustainable transport. Furthermore, the Government of India has recognized the increasing urban transport problems along with its associated problems and has developed a strategy to increase the efficiency of road space by favouring public transport. Moreover, transportation planners and engineers are often required to monitor the effectiveness and efficiency of transit services based on changes in the attributes of the public transportation system and the people using the transit system. It is important to estimate the demand for public transport in order to formulate policies and strategies for operational planning. Moreover, the accurate determination of link-volume and load rate will assist in implementing the desired scheduling policies for assigning buses along selected routes. The proper estimation of link-volumes and load rates using the right data mining tools will assist in better travel demand modelling and transport planning.

Conventionally demand data is collected through the field observations like origin-destination (OD) survey, household survey, passenger surveys, vehicle occupancy surveys and boarding-alighting surveys, which is time-consuming and involves huge cost and human resources. Hence, it is practically impossible to perform the field observations on a periodical basis for the public transport where there is a need for frequent analysis and updating of the demand data for operational planning of public transport. The system-wide or single route wise OD matrix and demand data are usually not collected through traditional methods; instead, it will be collected on a zonal basis. Automated Data Collection systems (Cui 2006) are becoming increasingly common in the developed world, but the developing countries are falling behind. Recently in India, most of the public transport operators have started using Electronic Ticketing Machines for providing tickets to the passengers and for fare collection. With the use of ETM, a large amount of data related to number of passengers, fare collected, operated

kilometres, and passenger kilometres are generated and stored. This data is available for the entire public transport network and fleet. The ETM data can be used to estimate the passenger demand, improve the operational profit and for performance evaluation. The archived ETM data is commonly used for performance evaluation (Furth et al. 2006) and for improving the operational profit (Oort and Cats 2015) but the potential of the ticket data for passenger demand estimation and forecasting is not explored. Also, it is required to develop tools for extracting the data compiled using ETM. Therefore, this study tries to investigate the demand capabilities of ETM data source for passenger demand estimation and to develop inter-city and intra-city passenger demand forecasting model using ETM data by exploiting the computational power and advanced modelling techniques.

1.4. OBJECTIVES

The main aim of the study is to develop an “ETM based Bus Transport Demand Model” for analysis and forecasting of bus transport demand.

- To collect intra-city Electronic Ticket Machine (ETM) data from Kerala State Road Transport Corporation (Kerala SRTC) and inter-city ETM data from Karnataka State Road Transport Corporation (Karnataka SRTC) bus depots.
- To recompile the raw ETM data and derive ETM based Origin-Destination matrix.
- To evaluate the public transport scenario of an urban area with respect to Level-of-Service and accessibility, and to evaluate and optimise the performance of Kerala SRTC.
- To determine daily variation index, weekly variation index, monthly variation index and quarterly variation index of intra-city and inter-city time-series ETM data.
- To develop intra-city and inter-city public bus transport demand model using time-series approach.
- To recommend operational strategies for improving the performance of Kerala SRTC and Karnataka SRTC.

1.5. SCOPE OF THE WORK

The scope of the research work is to investigate the demand capabilities of ETM data source for passenger demand estimation and to develop inter-city and intra-city passenger demand forecasting model using ETM data by exploiting the computational power and advanced modelling techniques. This includes the collection of intra-city ETM data from Kerala SRTC and inter-city ETM data from Karnataka SRTC, data processing and finally, the generation of Origin-Destination (OD) matrix from the processed ETM data. This OD matrix is then used for generating time-series data. Since the travel demand is affected by the performance of the public transport system and the bus transport agencies operating the services, this study evaluates the performance of the public transport system and the public transport operator. Therefore, in addition to ETM data, this study also uses data provided by various state and central government agencies. The level of service offered by the public transport system in Trivandrum city (intra-city), the accessibility of public transport and the performance evaluation of Kerala SRTC was performed to evaluate the present condition of the public transport system in Trivandrum and Kerala SRTC. The performance evaluation of Karnataka SRTC was not performed since the data collected from Karnataka was only for modelling inter-city passenger demand. This study has the scope to develop inter-city and intra-city passenger demand estimation model using the data-driven time series approach, employing the capabilities of AutoRegressive Integrated Moving Average (ARIMA) method. The models were further improved using AutoRegressive Neural Network (ARNN) method and AutoRegressive Conditional Heteroscedasticity (ARCH) modelling methods. The data used for time-series modelling is 'big data' which is a part of fare transactions of ticket sales using ETM for the intra-city travel in Trivandrum Municipal Corporation region of India in Kerala SRTC for the period 2011-2013 and inter-city travel of Mangalore division Karnataka SRTC for the period 2013-2018. Due to the lack of availability of sufficient data, the study is limited to daily forecasting with a forecast horizon of 30 days and seven days in the case of intra-city. Monthly forecasting with a forecast horizon of five months and daily forecasting with a forecast horizon of 30 days and seven days were performed in the case of inter-city.

This time-series methodology is useful when there is limited or no information available about the factors affecting the demand.

1.6. THESIS ORGANISATION

The thesis is organised in the following manner to ensure that the project is in the right direction to achieve the objectives.

Chapter 1 of this thesis gives a general introduction to public transport and the various aspects of public transport demand modelling. It also provides a brief introduction to the transport sector in India and the need for the study. This chapter states the main aim of this work along with the objectives of the work, the scope of the work and the major contributions of this research work.

Chapter 2 reviews the literature relevant to the determination of LoS, accessibility to and from the public transport and performance evaluation of public transport. Also, this chapter reviews past studies and researches on demand modelling in the context of public transportation. The literature dealing with the various methods like four-stage modelling, simultaneous demand modelling and time-series approaches are reviewed.

Chapter 3 gives the details of the study area. Initially, Trivandrum was chosen as the study area for intra-city and inter-city passenger demand modelling. During further analysis of the data, it was found that the data was insufficient, and the lack of data available from Trivandrum led to choose a different study area. Therefore, Mangalore was chosen as an additional study city for developing inter-city demand models. The socio-demographic characteristics, trends in land use, travel behaviour and the trends in public transport are discussed in this chapter.

Chapter 4 outlines the design and methodology of the research to achieve the objectives. This chapter discusses the data collection in detail. Land-use data, road network data, bus stop data and demographic data of the study area are described here. Also, the collection of ETM data, the processing of ETM data, the determination of the OD matrix and generation of time-series data are explained in detail in this chapter.

Further, the methodology adopted for the analysis of time-series ETM data and the passenger demand modelling using time-series approach is elaborated.

Chapter 5 deals with the analysis of the data. Also, the results of the study are presented and discussed with reference to the main aim of the study, which is to develop an “ETM based Bus Transport Demand Model”. The main findings in terms of research contributions, the evaluation of the existing condition of the public transport facilities of the study area (Trivandrum) and application of the methodology to various scenarios are presented in this chapter. The level of service characteristics, accessibility and performance evaluation of Kerala SRTC are discussed. The analysis of the intra-city and inter-city time-series ETM data, development of intra-city and inter-city time-series models and the forecasting of the passenger demand for the selected forecast horizon is explained in detail.

Chapter 6 concludes the thesis with the summary of the research work. The scope of future work is given as a sub-section in this chapter. The findings of the evaluation of Level-of-Service, determination of accessibility index, performance evaluation of Kerala SRTC, intra-city and inter-city variation indices and time series modelling are summarized. A discussion on the results of intra-city and inter-city time-series models are also included. This chapter includes a recommendation section and also discusses the future scope of the work.

1.7. MAJOR CONTRIBUTIONS OF THIS RESEARCH WORK

The primary aim of the study is to develop an “ETM based Bus Transport Demand Model” for analysis and forecasting of bus transport demand. A comprehensive study of the existing literature was done to identify a suitable data-driven method for passenger demand modelling. The LoS and accessibility of public transport in the Trivandrum city was determined along with the performance evaluation of the Kerala SRTC. In addition to this, intra-city and inter-city bus passenger demand modelling was also done. The major contributions of this research work are:

- The Level-of-Service characteristics of the Trivandrum study area are determined using the service level benchmarks set by the Ministry of Urban

Development, Government of India. Also, the accessibility of public transport and the performance evaluation of Kerala SRTC were also undertaken. This was carried out to understand and evaluate the present scenario of public transport in Trivandrum city and that of Kerala SRTC.

- The study proposed a methodology for the determination of the accessibility index. The developed methodology is a Geographic Information System (GIS) based public transport accessibility index which gives the accessibility of a single point (*i.e.*, the bus stop) by considering access to the point, access from the point to various land use destinations and network impedance. This methodology is based on the factor that the index should measure the accessibility which comes from proximity to bus stops and land use destinations, and the proportion of the population served.
- The performance evaluation of Kerala SRTC was performed using the methodology developed in this study. A weighted Goal Programming (GP) methodology integrated with Analytical Hierarchy Process (AHP) considering the user's and operator's perception was used. The goal programming part of the methodology takes care of the contradictory requirements or objectives of both the user and operator.
- This study uses the ETM data which was not previously explored for passenger demand modelling and forecasting for intra-city and inter-city scenarios. Also, this study reports a univariate seasonal AutoRegressive Integrated Moving Average (ARIMA) modelling methodology for intra-city and inter-city bus passenger demand forecasting, which is further improved using AutoRegressive Neural Network (ARNN) and AutoRegressive Conditional Heteroscedasticity (ARCH) models. It includes a comprehensive methodology for model selection and forecasting, which requires only basic econometric knowledge. This methodology can be useful when there is limited or no information available about the factors affecting demand.

CHAPTER 2

REVIEW OF LITERATURE

2.1. GENERAL

Since the beginning of civilisation, the efficiency of the transportation system influences the economic success and sustainability of the society to a significant extent. The public transportation planners and engineers have to be well-versed with the transportation planning and decision making process and henceforth have to forecast the responses of travel demand due to changes in the attributes of the public transportation system and the people using the system. Travel demand modelling aids as a tool to serve this purpose; especially to analyse and forecast the travel characteristics using different modes under various socio-economic scenarios and land-use structures.

In order to position this study in the field of transportation planning, a literature review, as summarized in Figure 2.1., is performed to identify various modelling techniques which are used by the transport planners and decision makers for public transport planning with an emphasis on periodical data. Also, various methods of data collection, level of service analysis of public transport facilities, accessibility of public transport and the performance optimisation of public transport is discussed in this chapter. The purpose of this chapter is to collect and organise material from a number of sources to try to understand the similarity of findings and to find where research is still needed to resolve the issues.

2.2. DATA COLLECTION

Passenger demand estimation is the most important and tedious task in transportation planning. The knowledge of travel pattern, travel behaviour, and the preferences of the traveller helps in increasing the ridership in public transport and provide travellers with varied mobility options. Surveys are the most basic exercise adopted to gain an understanding of the traveller information. Both manual method or field-based surveys and automated or technology-based methods can be used for the data collection.

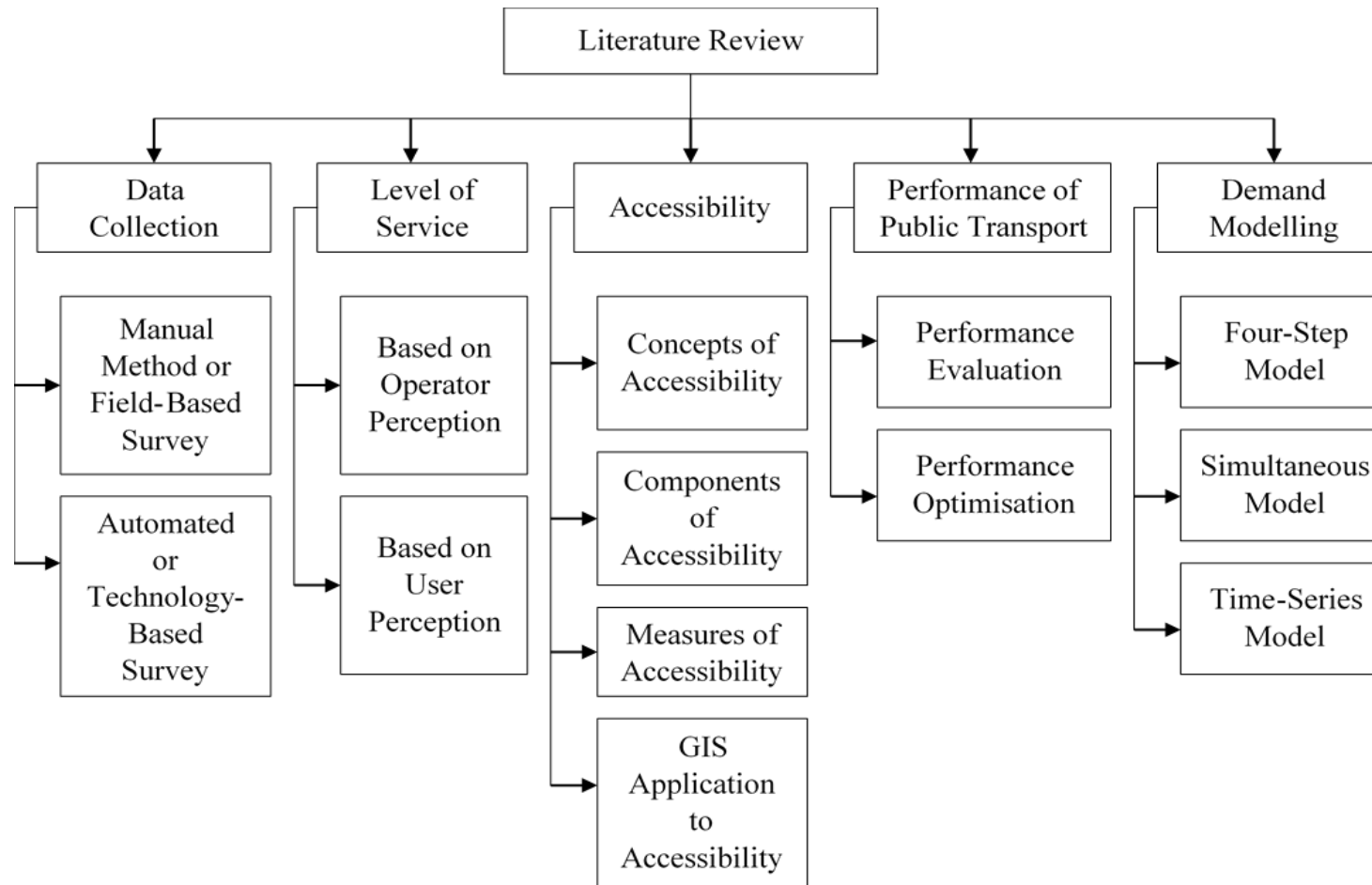


Figure 2.1. Flowchart Summarising the Literature Review

2.2.1. Manual Method or Field-Based Survey

The conventional manual method of data collection for demand estimation includes extensive household surveys or Origin-Destination (OD) surveys. Household surveys provide data on the household characteristics, individual members, their travel pattern, trip purpose, mode of travel, time of travel and preferences of travel. On the other hand, OD surveys provide detailed information about the travel pattern of people in a region. Similar to the household survey, the details of the individual trip-maker and information about the trips can be collected in OD surveys. Various methods of field-based surveys are bus occupancy survey, boarding-alighting survey and passenger interview surveys. Bus occupancy survey and boarding-alighting survey are observational surveys. The number of passengers on a bus or the load factor of the bus can be found using the bus occupancy survey. While the number of passengers boarding a bus from a bus stop and alighting at a bus stop can be determined using boarding-alighting survey. This gives a clear picture of the demand characteristics at the bus stop level. Passenger interviews can be used to study the level of service characteristics of the bus service, the willingness to pay for better mobility options and to identify the changes in the travel behaviour due to changes in the operational characteristics of the service.

The manual method of data collection through field observations are time-consuming and involves huge cost and human resources. Hence, it is practically impossible to perform the field observations on a periodical basis for the public transport where there is a need for frequent analysis and updating of the real-time traffic scenario for operational planning of public transport. The system-wide or single route wise OD matrix and demand data are usually not collected through traditional methods; instead, it will be collected on a zonal basis.

2.2.2. Automated or Technology-Based Surveys

The technology-based methods include Automatic Passenger Count (APC) system (Cui 2006), Automatic Fare Collection (AFC) system (Nunes 2016) and Automatic Vehicle Location (AVL) system (Matias and Cats 2016). These ADC systems (Cui 2006) used for daily operations have generated an enormous database that can be used for various

applications in the transport industry. The data collected from diverse sources of ADCs have distinct features and varied applicability. For example, AVL device is installed on the bus to track the location of the bus. On the other hand, APC gives the number of passengers by receiving “On or Off counts” or sometimes “On and Off counts” using sensors installed near the doors. In some cases, APC data is the by-product of fare transactions as the data is recorded when the passenger pays for the ticket (Cyril et al. 2017). When a passenger purchases a ticket, AFC data gets recorded. The AFC data gives the details regarding the origin, destination, number of passengers, fare and ticket type. But, in the proof-of-payment type AFC systems, the possibilities of fare evasion exist, and thus the passenger demand is underestimated. Hence, there is a need to establish a procedure to tackle the fare evaders to ensure that the demand data of all the passengers (Barabino et al. 2014a).

In India, the emergence of electronic fare collection system has become the fundamental system of automated data collection in the public transport industry. In the last decade, most of the transport operators have started using Electronic Ticketing Machines (ETM) for issuing tickets and for fare collection. In the case of a developing country like India, where smart card or travel card facilities are not commonly available, it is required to develop methods to use the readily available alternative data sources like ETMs. The ETMs generate a large amount of data, known as big data, pertaining to the number of passengers, time of ticket issuing, fare collected, earnings per kilometre and operated kilometre (Cyril et al. 2017). The ETM data is available on a time scale since the time of ETM implementation for the entire public transport network and the fleet. This readily available data can be used for passenger demand estimation, performance evaluation (Furth et al. 2006) improving the operational profit, decision making, planning and operation (Oort and Cats 2015).

2.2.2.1. Electronic Ticket Machine Data

Electronic Ticket Machine (ETM) is a handheld device used by the ticket collector for issuing the tickets to the passengers using the bus. The e-ticket contains details of the bus, information of the route and trip details. The machine records information such as the bus type, bus number, minimum fare, fare increment, fare stages, bus stop details,

route number, passenger type and the number of passengers (Cyril et al. 2017). The raw data from the ETM has to be processed to obtain the required passenger demand data.

The advantages of using ETM data source over other surveys are listed below (Nunes 2016; Cyril et al. 2017; Meal and Carter 1998):

- Reduced time and cost requirement
- Large sample size
- No bias in data collection
- Frequent data collection and estimation
- Data analysis can be done for any time period

The challenges of using ETM data source are as follows (Barabino et al. 2014a; Barabino et al. 2014b):

- The possibilities of fare evasion exist, and thus the passenger demand is underestimated.
- Exact origin and destination of the passengers cannot be identified.
- Matching data to the exact bus stops, data validation and anomalies in the data have to be addressed.

2.3. LEVEL OF SERVICE

The evaluation of service quality is one of the challenging task and an important practical problem encountered by service providers. The assessment of the quality of service is required to improve user satisfaction and thus to improve the demand for public transport services. Researchers have identified Level-of-Service (LoS) as an effective measure to understand the user's level of satisfaction for the service provided.

The concept of service quality measure using LoS is first introduced in Highway Capacity Manual in 1965 (Charles et al. 1965). Later various researchers have revised and modified the concept of level of service to meet the individual needs of the highway, freeway, public transport etc. Tyrinopoulos and Antoniou (2008) focused the research on LoS based on passenger perception of the performance of the transit. The

study dealt with the variability between operators and the policy implications on user perception.

Polus and Shefer (1984) evaluated the level of service of a bus line based on its operational characteristics. It is based on the average travel time using bus and passenger car on the shortest distance between the same origin and destination. Level of service is explained using the explanatory variable like the average distance between the bus stations, route length, number of signalized and unsignalised intersections.

Hensher et al. (2003) investigated various methods to quantify the service quality and compare the levels within and between the bus operators. The service quality index serves both the operator and the user. In this study, the different methods to quantify service quality and methods to compare the levels within and between bus operators are studied. Finally, the importance of establishing market segments and the need to scale the quality index for each operator is highlighted.

Das and Pandit (2013) proposed a methodology to develop an overall framework to determine the threshold LoS based on user perception. Also, they provided a measure to find the maximum number of users from different user segment that will get satisfied at these threshold LoS values. The methodology includes the identification of service quality parameters, user perception of existing and potential bus users and determines the threshold LoS, identification of the number of passengers satisfied with the different threshold LoS values and then identify the gaps in the service quality.

The Ministry of Urban Development of Government of India has set up the service level benchmarks for determining the Level-of-Service provided by the public transportation facility (MoUD 2009). The indicators for determining the LoS used in the study are the presence of organised public transport system, availability or extent of public transport supply, service coverage, average waiting time, level of comfort in public transport and the fleet availability. This approach can be used to determine the LoS at the city-level and thus provides a measure for identifying the public transport quality of an urban area.

2.4. ACCESSIBILITY OF PUBLIC TRANSPORT

Public transport is the critical component of the sustainable transportation system (Albacete et al. 2017), which can provide a systematic mobility option and reduce the transport related externalities. However, the provision of public transport will not accomplish the expected effect until the public transport system is accessible and connected to the rest of the urban systems (Tuan and Son 2015). Among the different modes of public transportation, buses are the most preferred mode in India due to the wide presence of road network and hence ease of accessibility. Thus the demand for bus transit is largely dependent on the accessibility to the transit service.

Accessibility is the measure of ease of access to and from the transit services. Some researchers explain accessibility "to" a place while few others suggest accessibility "from" a point and few researchers explain accessibility in composite terms, *i.e.* "to" and "from" a place. Quite a few authors have explored walk access to transit as a measure of accessibility (Woldeamanuel and Kent, 2015) while others explain deficiencies in transport infrastructure in evaluating accessibility (Lantseva and Ivanov 2016). Micro-accessibility and macro-accessibility were studied by Eggermond et al. (2016) to investigate the role of the built environment and accessibility on vehicle ownership. Salonen and Toivonen (2013) modelled travel time to analyse the accessibility disparity between travel modes of the car and public transport. Cheng and Chen (2015) studied perceived accessibility, mobility and connectivity of public transportation systems. For decades, accessibility has been the focus of research as it is one of the most important factors to be considered in sustainable public transportation. However, it still remains a concept that has been poorly defined and hence finding a theoretically good and operationally sound measure of accessibility is a challenging task. Historically, transport planning organisations have considered that transport problems and solutions can be determined and treated without considering the non-transport aspect of urban life. But, this perception had changed when the focus shifted from "moving vehicles" to "moving persons" (Morris et al. 1979). This has led to a comprehensive modelling of travel demand integrating spatial-land use-travel patterns for transport and urban planning.

2.4.1. Concepts of Accessibility

The concept of accessibility can be perceived as a measure of spatial separation of human activities, *i.e.* it can be defined as the ease with which activities can be reached from a location using a particular transportation system (Morris et al. 1979). Accessibility is defined in several ways and hence has different meanings. The well-known definitions include "accessibility is defined as the potential of opportunities for interaction" (Huang 2009). Accessibility accounts for land use patterns and hence the application of accessibility indicator include evaluation of transport and land use (Bhat et al. 2000b), transport policy decision making, choice modelling and evaluation of spatial structure (Morris et al. 1979). Accessibility was broadly classified into two by Ingram (1971): "relative accessibility" and "integral accessibility". Relative accessibility refers to the degree to which two places are connected while integral accessibility for a given point is defined as the degree of interconnection with all other points (Ingram 1971). In reality, accessibility indicators constitute a wide family of measures which are modified according to the researcher's perspective and objectives. The literature suggests determining accessibility as a measure "to" a point, "from" a point or composite accessibility measure, each measure having its own advantages and limitations. Gulyas and Kovacs (2016) utilised accessibility for the evaluation of transport connectivity. Focusing on passenger movement through public transport, accessibility can be defined as the ease with which passengers can reach and find public transport and arrive at the destination of interest.

2.4.2. Components of Accessibility

A number of components of accessibility can be identified from the various accessibility measures. Four main components which are required for measuring the accessibility are spatial component, temporal component, transportation component and personal component (Balya et al. 2016, Geurs and Wee 2005, Morris et al. 1979).

- The spatial component determines the spatial distribution of opportunities and its quantity in each destination, the activities of the individuals, *i.e.* the demand

for these opportunities and supply of the opportunities in the zone of individuals (*i.e.* origin zone).

- The temporal component determines how long the individual participate in an activity (work, school, shopping) and the availability of each opportunity during each time of day.
- Transportation component, also known as impedance component refers to the disutility of the transportation system. It includes the walking time to the station, waiting time at the station, cost of travel, travel time, comfort and safety factors.
- The personal component is the requirement of each category of people. The needs depend on the age, education, physical conditions, economic, social and cultural aspects of the individuals.

For an accessibility measure to be ideal, it must take into consideration all the four components. But in practical situations, accessibility measure takes into account one or more components depending on the study area, objective and perspective of the study. Depending on the study area, factors affecting accessibility must also be considered for finding the accessibility measure.

2.4.3. Measures of Accessibility

The accessibility measures defined diversely determines the connection between land use and transport. Based on the review by Bhat et al. (2000a), accessibility measures are classified into: (a) Spatial separation measure, (b) cumulative opportunity measure, (c) Gravity measure, (d) Logsum or utility measure and (e) Time and space models. The spatial separation method is the simplest accessibility measure as the distance is the only variable considered. Cumulative opportunities measure takes account of both distance and purpose of the trip. According to Bhat et al. (2000a), “this measure defines a travel time or distance threshold and uses the number of potential activities within that threshold as the accessibility for that spatial unit”. Gravity measure uses both attraction and separation factor, *i.e.*, it discusses an attractive force and the impedance of intervening space. Another approach is based on the perceived utility of traveller and is known as a utility measure. In time-space models, both spatial and temporal accessibility are determined based on spatiotemporal constraints. Bhat et al. (2000b)

have reviewed the micro-scale and macro-scale factors for inclusion in ideal accessibility measure and examined various impedance attributes, destination attributes, and traveller's attributes. Also, Bhat et al. (2002) assessed various alternative accessibility measures and proposed the aggregation of various accessibility measures over purpose, time of day, mode and space.

The potential opportunities measure was used by Hansen (1959) for developing a gravity type accessibility model in which accessibility being proportional to the size of activities in the zone and inversely to the travel time or distance. Vieira and Haddad (2012) extended Hansen's model of gravity formulation to measure accessibility to jobs in the Metropolitan Region of Sao Paulo. The opportunities considered were the number of jobs in each zone and impedance. The inverse exponential function of travel time was taken as the impedance between each pair of zones. Kalaanidhi and Gunasekaran (2013) used a composite impedance gravity measure to incorporate the effects of in-vehicle travel time, off-vehicle travel time, the stops and cost of parking in determining the accessibility. Accessibility to transit service was studied by Yatskiv et al. (2017) considering the travel distance and connectivity between the city and destinations.

On the other hand, Ingram (1971) used average squared distance as the spatial separation having a modified Gaussian functional form to determine the integral accessibility. Baxter and Lenzi (1975) proposed a method for determining the relative accessibility by considering a method to arrive at an accurate distance matrix for spatial separation by incorporating the geographical constraints. Allen et al. (1993) developed an accessibility index to calculate an area's overall accessibility level based on the spatial separation method taking travel time as the separation measure. Using this index, the overall access levels of 60 United States metropolitan areas were calculated and compared.

Another approach to determining accessibility is the logsum or utility-based measure. Sweet (1997) argued that centred logsum would be a more appropriate measure of accessibility. Bhat et al. (2005) developed a methodology having a logsum measure with a multinomial logit model to aggregate the accessibility measures over several dimensions of interest: time of day, mode, trip purpose, and space.

Kwan's (1998) study indicated that space-time index and integral accessibility index are distinctive types of accessibility measures which reflect the different dimensions of accessibility of individuals. Bok and Kwon (2016) focused on both spatial and temporal dimensions of public transport in which spatial factor of accessibility is identified as the specific distance from each node of transit, and temporal aspect is defined by the frequency of buses. Fotini (2017) examined the individual accessibility in space and time by incorporating spatial, temporal and travel data. This proposed composite accessibility measure relies on factors like individual activity space and temporal constraints.

2.4.4. GIS Application to Accessibility

In recent years, GIS technology has flourished, and various researchers have found interest in data analysis, interpretation, storing, manage and present the spatial data. Thus GIS has been applied to evaluate the accessibility and land use opportunities (Liu and Zhou 2015; Wang and Chen 2015; Wang et al. 2014). Al Mamun and Lownes (2001) developed a composite index to measure the accessibility of public transit using a weighing factor. Huang et al. (2009) developed a GIS-based accessibility modelling for estimating transit demand based on the walking distance to bus stops. Pitot et al. (2006) use a GIS-based methodology to quantify and map accessibility to various destinations by walking or public transport. The methodology discusses the land use and public transport accessibility index (LUPTAI) as a decision-making tool to optimise land use and transport integration. Saghapour et al. (2016) introduced a measure of accessibility by incorporating population density. The public transport network model was used in a GIS environment to identify the service coverage of public transport modes. Shah and Adhvaryu (2016) adapted the Public Transport Accessibility Level (PTAL) methodology developed by Transport for London (TfL) to the case study of Ahmedabad, India, taking into account average walk speed and time, distance to transit stops, and peak hour transit frequencies and GIS was used for mapping the PTAL. Spatial transit accessibility modelling was performed in GIS by Balya et al. (2016) for an Indian metropolitan city. GIS was used to develop accessibility catchments through the buffering process of various radius to identify accessibility thirst areas. Yan-Yan et al. (2016) suggested a concept of Area Public Transport

Accessibility (APTA) in GIS environment based on passenger travel behaviour, travel psychology hypothesis and service range of transit and road network which provides quantitative information for location analysis and public transport optimisation. Although researchers have developed accessibility index from various perspectives, there is still a need for integration of accurate determination of accessibility measure. This calls for the use of Geographic Information System (GIS) in determining the accessibility index as a graphical and descriptive measure of accessibility rather than a quantitative one.

2.5. PERFORMANCE EVALUATION OF PUBLIC TRANSPORT

The increased travel demand has triggered the transport crisis like congestion, pollution, and other environmental externalities. Mitigation of these transport crisis has become a challenging task in the transport industry. Development of public transport has been identified as a sustainable solution for all the major transport problems. Moreover, public transport is the primary and only means of transport for a large section of society in India. The Working Group of Urban Transport (2012) has suggested the desirable share of public transport as 60% of the motorised trips for reducing the energy needs and transport crisis. Public transport undertakings are striving to provide efficient and convenient travel. But still, it lacks to provide better travel options as the challenges in the public transport industry varies in various aspects like financial instability to incompetence and unreliability. Excessive operating cost, overstaffing, low productivity and the imprudent use of financial resources are few institutional issues while inadequate frequency, increased travel time, poor service quality and overcrowding are few of the reasons which shift the users away from public transport. The declining share of public transport has weighed down the public transport industry as a loss-incurring entity.

While the government has a complete monopoly over the rail transport sector, the road transport industry has many competing players. In the fiercely competitive environment, the state-owned public transport industry cannot operate sustainably with a mediocre performance. In this respect, a crucial question of the type of operating practices and administrative regulations that can improve the public transport industry;

and inefficiencies, bottlenecks and the potential of public transport has to be determined by the performance evaluation of current scenario. Moreover, to improve performance and efficiency, with reduced budgets, political expectations and competition between the operators, it is required to optimise the performance of the industry by optimising the available resources.

In the recent years, performance evaluation has become the focus of attention in the public transport industry as it is viewed as a method to assess the extent of the outcome of the system, which is further analysed to decide the improvement strategies. Since public transport involves multiple stakeholders, it is required to perform optimisation procedures rather than evaluation. Unfortunately, performance optimisation is an area left largely unexplored even though it facilitates the efficient and effective use of technological, financial, material and human resources. According to Perez et al. (2015), a truly optimal solution exists only if a single criterion is considered. However, in practical scenarios, several issues are to be addressed in the optimisation procedure. Firstly, the decision variables are multiple and structured in multilevel hierarchies (Wu et al. 2018). For instance, passenger transport assessment solely based on economic criteria can be too narrow as the final decision making depends on various other factors other than monetary ones (Ambrasaite et al. 2011). Secondly, there involves some level of subjective judgment in the assessment of decision variables and thus results in the use of incorrect information. Finally, the stakeholders of the public transport industry comprising users, operators and the community (Murugesan and Ramamoorthy 2001).

As previously mentioned, the performance evaluation of public transport is one of the widely investigated areas within the transit industry, and performance optimisation is largely unexplored. According to Gleason and Barnum (1982), in some instance, the term “productivity”, “efficiency” and “effectiveness” is synonymous with performance in the public transport industry while in other cases “efficiency” and “effectiveness” are considered as different aspects of “productivity”. Many authors have concentrated on the development of models for performance evaluation using parametric, non-parametric, deterministic, stochastic or soft computing techniques (Famurewa et al. 2014; Karlaftis and Tsamboulas 2012).

Gleason and Barnum (1982) examined the various performance measures of efficiency and effectiveness and pointed out that there is a lack of understanding of system analysis concepts related to criteria selection. Guidelines for bus transit performance evaluation of Washington State was given by Kelley and Rutherford (1983) considering both efficiency and effectiveness. He et al. (2016) developed a systematic network operations management system incorporating various performance assessments of the urban rail transit network. The study analysed the various features of operation, changes in travel demand, availability of urban mass transit and operational service quality. Thus, an information system based operational planning and performance evaluation using network operations management was developed. Various researchers (Chen and Yu 2016; Karlaftis 2004; Li et al. 2013; Sampaio et al. 2008; Venkatesh and Kushwaha 2018) have used data envelopment analysis (DEA), a common non-parametric approach, for performance evaluation of transit industry by estimating the efficiency scores. Karlaftis (2004) employed the DEA approach for evaluating the efficiency and effectiveness of the urban transit system using data from 256 US transit system over a five-year period. Nineteen public transport systems were analysed using DEA by Sampaio et al. (2008) for determining service efficiency based on three primary inputs: labour, fuel and capital. In the study by Lao and Liu (2009), the application of combined DEA and geographic information system (GIS) was used for performance evaluation of bus lines, considering both the operations and operating environment. In this study, GIS was used to create demographic profiles within the service corridors, and then DEA was applied to compute the operational efficiency and spatial effectiveness for comparison of operational and spatial aspects of the bus line. Vaidya (2014) evaluated the relative performance of 26 public transport undertakings in India based on 19 criteria. These criteria were grouped into operations, finance and accident based, and then evaluated the importance of these criteria using AHP. Further, DEA was used to evaluate the public transport undertakings using various criteria within each criteria group. Finally, to quantify the overall performance, a Transportation Efficiency Number (TEN) was developed. Short and long-run cost efficiency analysis of Indian public bus companies were determined using DEA by Venkatesh and Kushwaha (2018). Tsai et al. (2015) used a two-stage DEA model to identify and measure the

factors which determine technical, allocative and cost efficiency in 20 international urban rail systems.

Yeh et al. (2000) examined the use of the fuzzy multi-criteria approach for performance evaluation. The concept of optimality of each alternative with respect to each criterion was used to develop a fuzzy singleton matrix from a weighted fuzzy performance matrix. Finally, an overall performance index for each alternative was calculated by incorporating the decision maker's attitude towards risk. Sun et al. (2018) used a multi-agent based simulation method to quantify the performance improvement options of urban rail transit systems using a physical internet. The study examines the changes in the performance indicators before and after the implementation of the physical internet. The rail transit line performance was measured using a hierarchical customer satisfaction framework in Istanbul (Aydin et al. 2015). The framework for the performance evaluation of transit line used statistical analysis, fuzzy analytic hierarchy process, fuzzy sets and Choquet integrals to analyse customer satisfaction. This methodology is a multi-criteria decision-making process which can be used when the decision-making problem is complex with uncertain, subjective and vague information.

The evidence in the previous paragraph indicates that performance evaluation mainly deals only with the operational, financial and technical characteristics of the public transport industry. Assessing the public transport performance by merely using the operator's data is an unfair practice as the stakeholders involved in the public transport industry includes operators, users and society as a whole. Therefore, performance optimisation must be performed considering the various stakeholders. The optimisation is a process of maximising or minimising an objective function by systematically choosing inputs from a set of input values. According to Wey and Wu (2007), mathematical programming is an optimisation problem consisting of different models like linear programming, integer programming, goal programming, dynamic programming and game theory. Goal programming is a multi-objective optimisation technique capable of handling multi-criteria decision-making (MCDM) problems having multiple and conflicting goals. However, when goal programming is used as a decision-making tool, goals correspond to diverse and conflicting criteria (Jones and Tamiz, 2003). Performance optimisation of public transport undertaking by Murugesan

and Ramamurthy (2001) used goal programming methodology by integrating the user and operator perceptions. The methodology involved the identification of multiple objectives, prioritising these objectives by expert opinion, and determination of the optimal solution. For optimising the real-life situation more effectively, it is required to combine various techniques; and goal programming has been used as a part of mixed modelling. The integration of goal programming and AHP has been applied to a variety of areas including supplier selection (Ghodsypour and O'Brien 1998), resource planning (Lee and Kwak 1999), project selection (Lee and Kim 2000), transportation resource allocation (Wey and Wu, 2007) and location selection problems (Ho et al. 2013). Saaty (1995) illustrated the different uses of multi-criteria decision process using AHP in the field of transportation including route selection hierarchy, best mix of the route to Pittsburgh Airport, a benefits/costs hierarchy to choose a mode, a planning hierarchy for the transportation system and simple dependence with feedback cycle to choose a car. Wu et al. (2018) used AHP to identify the weights for indicators for underground rail transit station area. AHP was selected after comparing the rationality of different weight indicator scales.

2.6. TRAVEL DEMAND MODELLING

Travel demand model, also known as travel demand forecasting model, aims to establish a spatial distribution of trips and to estimate the travel behaviour and travel demand for a specific time frame based on certain assumptions. Travel demand forecasting is an attempt to predict the future travel pattern and to quantify it. There are two basic forecasting needs in transportation planning: first is travel demand studies for urban areas (intra-city) and second is inter-city travel demand. Urban travel demand forecasts require extensive database which is conventionally obtained from travel surveys (home-interview or roadside interview surveys) and observations. The information collected provides useful insight to the socio-economic characteristic of the trip maker, vehicle ownership, land use at the trip end and travel characteristics pertaining to the mode of travel, travel time, time of day, trip purpose and travel cost. The data obtained is at the individual level, which can be further aggregated at the zonal level or can be used at the disaggregate level (household or individual level) for data analysis, synthesizing and calibrating models. In the inter-city case, the data is more of

an aggregated level, such as city size, average city income, population, employment characteristics and travel time or cost between city pairs (Garber and Hoel 2014). Recently, various other sources like electronic ticket data, smart cards, mobile phone data, GPS location of vehicles and human beings, information systems, Bluetooth data and social media check-in are used as a source of travel data.

Travel forecasting as a formal process began in the early 1950s. The process that emerged came to be known as the Four-Step Method (Barnes and Davis 1999). It has been the paradigm in urban travel demand modelling for decades. Considerable refinements have been made, and the process is continuously evolving. But the general ideology behind it remains hardly untouched. The classical transport model is a sequence of four submodels and hence the name four-step model. It is also known as “traditional four-step method”. Understanding the traditional approach is greatly helpful in understanding the travel demand forecasting process. The four basic models of the classical method are Trip Generation, Trip Distribution, Modal Split and Trip Assignment. The model starts with defining the study area and then dividing it into a number of zones, known as traffic analysis zones and considering the entire transport network in the system.

Trip Generation: Trip generation is the first step in the conventional four stage transportation forecasting process. It predicts the number of trips produced in or attracted to a particular traffic analysis zone. Trip generation thus comprises of trip production and trip Attraction. Trip production deals with the number of trips originating in a particular zone. The trip attraction includes the number of trips destined in a particular zone. It determines the magnitude of total daily trips at the household level and zonal level. It is this step which translates activity based and trip based modelling.

Trip Distribution: Trip Distribution is the second stage in the sequential travel demand modelling. The trips generated in each zone are distributed to all other zones based on destination choice of the travellers. This process is called trip distribution. The underlying principle is that the zones attracting the trips are in competition with each other and thus depending upon the zonal attractiveness and the factors affecting the

choice of the trip makers. Trip distribution is generally illustrated better in a matrix form known as a trip matrix or Origin-Destination (OD) matrix. The following are the various types of approaches to trip distribution:

1. Growth factor methods

- Uniform growth factor method
- Average growth factor method
- Fratar method
- Furness method

2. Synthetic Methods

- Gravity Model
- Tanner's model
- Intervening opportunities model
- Competing opportunities model

Modal split: The third stage in the travel demand model is the modal split. The total OD matrix generated from the trip distribution phase is split according to different modes to get mode wise OD matrix. When there are several alternative modes for travel, the trip makers select their mode based on the cost of travel by each mode, availability and accessibility to alternate modes, safety, comfort and urgency of performing the trip. The modal split process can be performed at a number of stages in the transport planning process.

- Trip end modal split models- performed prior to trip distribution
- Trip interchange modal split model- performed after trip distribution
- Two-stage modal split analysis

Trip Assignment: The process of allocating the trips on the transportation network is known as the trip assignment. The fundamental aim of the trip assignment is to reproduce the traffic patterns and vehicular movements, which would be observed when the travel demand represented by the trip matrix is assigned.

2.7. PUBLIC TRANSPORT DEMAND MODELLING

Public transport is the key to solutions for major transportation problems. A good transportation system must be able to accommodate the transport demand. Estimates of public transport demand are important to bus operators and planners. It is important for the government as the public bus transportation is highly subsidized and should serve the society with sustainable transport.

2.7.1. Approaches to Estimate Public Transport Demand

There are mainly four approaches to estimate the factors influencing public transport demand (Mullen 1975).

- **Special Experiments:** This includes observing the changes which result from specific policy changes and operational characteristics.
- **Analysis of Mode Choice of the Traveller's:** This is carried out by the analysis of data obtained from a home interview survey. This involves large data and huge data collection costs. But this method will provide a good source of individual data.
- **Cross-Sectional Comparison between Cities:** This includes the comparison of public transport trip characteristics of different cities. The basic data includes trip frequency, population, vehicle ownership, etc. which are readily available, but the difference in characteristics of cities considered must be taken into account.
- **Time-Series Analysis of Bus-Operator Data:** This method can be used in aggregate level for the whole system or at an individual level for a single route if detailed information is available.

2.7.2. Major Characteristics of Public Transport Demand

Public Transport (PT) is the common carrier type of urban passenger transport having fixed routes and schedules, available for use by all persons who pay the established fare (Vuchic 2007). Public transport services are the very basic need of people's lives, especially in the developing world. It is the basic means for the mobility of a large

percentage of people, and hence, it is the driving force of economic and social life. Based on its importance, public transport service is a matter of national policy which is obliged to meet the mobility issues of the people and to reduce the environmental externalities.

The characteristics of public transport services determine the effective understanding of factors affecting the demand. Firstly, the public transport environment is dynamic and interactive. It includes a combination of various alternative modes, different types of passengers (students, workers, aged people, etc.), travelling for different purposes with different travel frequencies and travel times. In such environments, the demand will be dynamic. Secondly, the demand for public transport is highly time-dependent. It is high in the morning and evening hours because of students and workers while it is less in the remaining hours (used by shoppers, leisure or other travellers). Thirdly, the expectation of different types of travellers from the public transport service are different, based on the purpose of travel and travel time. A person using public transport for work and leisure during different hours of the day may require different levels of service (Polat 2012). Therefore, the factors affecting the demand are determined mainly by the characteristics of public transport services.

2.8. HISTORICAL PERSPECTIVE OF PUBLIC TRANSPORT DEMAND MODELLING

A large number of models have been developed to forecast travel demand, both passenger transport demand and freight transport demand. Different models are based on different assumptions, have different specifications and characteristics. In the early 1950s, the Chicago Area Transportation study was the first to use the four-step model (FSM) for transportation planning for highway-related projects (McNally 2007). In the 1960s, comprehensive urban transportation planning was performed using FSM followed by using economic theory in 1970s and then activity-based approach in 1980s. Models of all types can be classified into several groups like aggregate and disaggregate (based on the number of individuals represented), conventional and unconventional (based on data availability, *i.e.* data from direct source like surveys is conventional, and the use of readily available data is the unconventional method), stochastic and

deterministic (randomness of the outcome), static and dynamic(based on time dependence), analytical and simulation-based, univariate and multivariate (based on number of explanatory variables) and sequential and simultaneous; in which the literature may fall into various groups of classification mentioned above. The subsequent sections in this chapter discuss few families of models *viz.*, four-step method, direct demand model and time series model; which falls into the above-mentioned classification of demand modelling techniques.

2.8.1. Four-Step Model

Four-step modelling is the conventional approach used to determine the amount of traffic on the links in the transportation network dating back to the 1950s. It has been the paradigm in urban travel demand modelling for decades. Considerable refinements have been made, and the process is continuously evolving. But the general ideology behind it remains hardly untouched. The outputs of early travel demand models have faced severe criticism for being unable to predict the future travel demand with proper accuracy and its inflexibility to accommodate the rapidly changing circumstances in the urban scenario. Forecasting public transport demand is closely associated with policy formulation and urban transport planning. Most of the bus demand studies follow the conventional urban transport planning techniques: trip generation, trip distribution, mode split and assignment. In other words, most studies on public transport demand modelling use a multivariate approach, in which the travel demand is predicted based on a set of explanatory variables and model inputs. The two major multivariate forecasting methods are multi-stage modelling and direct demand model (Tsai et al. 2013); this section deals with four stage or multi-stage modelling.

Yao and Morikawa (2005) introduced a comprehensive accessibility measure known as expected maximum utility in a multi-dimensional mode/route choice and destination choice context to get the behavioural effects. An integrated travel demand model with a nested structure including trip generation model, destination choice model, mode choice model and route choice model was presented as a case study for an inter-city High-Speed Rail (HSR) in Japan. The models are built for two travel purposes: business and non-business, integrating aggregate and disaggregate data. Multiple regression

analysis was used for trip generation modelling. The explanatory variable is the expected maximum utility value, which is interpreted as the accessibility measure of the destination zone. In destination choice model, residence-based trip data is used in aggregate form along with some socio-economic factors such as Gross Domestic Product, per capita rate of working population to population and two variables indicating the degree of the attractiveness of a zone for business and non-business purpose.

In the study by Zhang and Wu (2007), a comparatively low-cost method to forecast the passenger demand for the inter-urban public transport was developed using OD information in the railway database and Revealed Preference (RP) and Stated Preference (SP) survey.

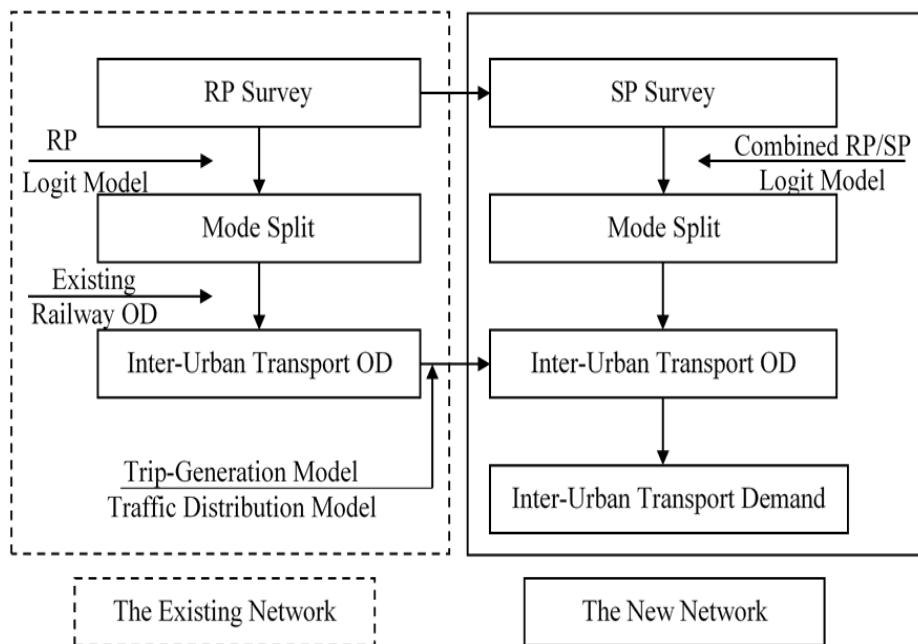


Figure 2.2. The Designed Travel Demand Forecast Method

Travel utility function was built to calculate the mode split on the existing network and then using railway OD data inter-urban public transport OD was estimated. RP Logit model was used to analyse the mode split on the existing network while combined RP/SP Logit model was used to forecast the mode split on the new network, as shown in Figure 2.2. With trip generation and distribution models and combined RP and SP data inter-urban public transport OD was forecasted.

A fuzzy-neuro travel demand model focusing on trip distribution and mode choice was developed by Yaldi et al. (2010) for improving the ability to forecast the travel demand and to simplify the model development and calibration. Three input nodes: trip production, trip attraction and trip length were used in trip distribution model, while socio-economic characteristics and characteristics of alternate modes are used for mode choice modelling, in which some of the variables are fuzzy. Seyedabrishami and Shafahi (2011) adopted an expert guided algorithm (EGANFIS) to use Adaptive Network-based Fuzzy Interference System (ANFIS) for handling uncertainties in travel demand estimation using household survey data. The study demonstrated that the traditional expert guided ANFIS increases the performance of travel demand compared to traditional ANFIS.

In a study of the capital area of Reykjavik, a four-step transport planning model was developed along with a simulation of the public bus system (Skarphendinson, 2013). The methodology starts with grouping the population based on similar travel behaviour, termed as mobility groups which are employees, unemployed, primary school students and higher education students. Furthermore, trips were grouped according to the purpose. Mobility groups, along with trip purpose, called as demand stratum, were used for the modelling process. Trip productions and attractions were developed for all demand stratum separately followed by a gravity model to distribute trip among the zones for each demand stratum. The third step was mode split of demand stratum into private and public transport modes, and finally, all the demand stratums were joined and assigned to the road network for private and public transport modes separately. Modelling was performed using VISUM. Simulations for the public bus system had less accuracy as the road user's preferences for mode choice was developed based on outdated data.

A bus demand model is created using the four-step modelling approach, including the stages of trip generation, trip distribution and assignment by Smit (2014). Mode split was not included as it is a unimodal demand model. Six generation models were developed: boarders weekday, alighters weekday, boarders Saturday, alighters Saturday, boarders Sunday and alighters Sunday. The generation models determine the number of boarders and alighters in each bus stop while the distribution model

determines the number of passengers at each bus stop pair. Three distribution models—one for weekdays, one for Saturday and one for Sunday, were developed using the top log-normal function since it is able to reproduce the observed trip length distribution. For the assignment, the shortest route (with least number of bus stops) is selected, and the passengers are assigned to it. Population characteristics, trip-specific characteristics, bus service characteristics and spatial characteristics were used. In this study, it was found that bus demand model developed for one area cannot simply be applied to another area.

Pulugurta et al. (2015) developed a fuzzy logic based travel demand modelling using subtractive clustering technique in MATLAB software. Four stage conventional models were developed to compare the efficiency of the fuzzy logic based models. The suitability of the fuzzy model was demonstrated by simulating public transport policies. For trip production modelling, the households categorized into vehicle ownership levels and trip purposes were considered, while trip purpose categorization was only considered for attraction modelling. Trip distribution model was developed using a gravity-based approach and multinomial logit model for mode choice. For private modes, user equilibrium assignment technique and for public transport, all or nothing was chosen as the traffic assignment technique.

2.8.2. Simultaneous Modelling Approach

Simultaneous travel demand model, also known as direct demand model, is a quick response simplified travel demand method. This method estimates the trip generation, distribution and mode split stages of four stage model simultaneously in which route choice may (Safwat and Magnanti 1988) or may not (Asri and Sugie 2003) be considered. Simultaneous model is further classified into simultaneous equilibrium model, direct demand model and demand elasticity model based on the principles of modelling technique. The simultaneous travel demand model is developed as an alternative to the four-stage method and can be considered as an extension of the gravity model. These models estimate the ridership of transit service rather than using mode choice resulting from the large model. The simultaneous approach is commonly used for inter-city modelling (Safwat and Magnanti 1988; Wirasinghe and Kumarage 1998;

Aldian and Taylor 2003) with few cases of urban travel model (Domencich et al. 1968; Talvitte 1973; Mullen 1975). In recent years, simultaneous demand models were developed for demand estimation at station level, corridor level or for determination of station to station transit ridership.

2.8.2.1. Simultaneous Equilibrium Model

Initially, simultaneous model was developed as a transportation equilibrium model (Beckmann et al. 1955; Safwat and Magnanti 1988) and later, economic theory of consumer demand was introduced as an elasticity approach (Domencich et al. 1968; Preston 2005) for demand estimation and determining the direct and cross-elasticities. The first simultaneous approach was found in the work by Beckmann et al. (1955), which was a case of transportation equilibrium model, which was viewed as a convex optimisation problem.

Safwat and Magnanti (1988) developed an integrated model to achieve a reasonable trade-off between behavioural and computational aspects of travel demand modelling problem. The simultaneous transportation equilibrium model (STEM) had a functional form of logit model based on random utility theory; reformulated as a convex optimisation problem with computational advantages. This model was applied to the inter-city transport network of Egypt and the urban transportation network of Austin to illustrate the effectiveness of the approach. The results indicated that the equivalent convex programming approach was effective in modelling the transportation system.

Peng et al. (1997) developed a route level transit patronage simultaneous model considering demand, supply and inter-route relationship, indicating the simultaneity between demand, supply and interrelationship among routes. The model is estimated at a route level by time of day and direction. The results show strong simultaneity among demand, supply and competing routes and suggested that the model is useful for both system-level and route-level service planning.

Tamin (1997) addressed the issues found in developing countries, which are the lack of information for long, medium and short range public transport planning due to money and time constraints. In the study, traffic counts and simple zonal planning data are used

for modelling the public transport demand. Gravity model, combined with the multinomial logit model, was developed and calibrated using nonlinear least-square estimation method.

Wirasinghe and Kumarage (1998) developed an aggregate demand model incorporating variables representing urbanisation, under-development, transfers, mode abstract cost function and intrinsic features for inter-city passenger travel for Sri Lanka. The largely data intensive disaggregate model is unviable in many cases; hence, an aggregate model is suitable in most cases for a developing country. In this study, a basic gravity model was developed as given in Equation 2.1, initially using the population of two district pairs and minimum travel time between the district pairs using any mode of public transportation.

$$V_{ij} = 1.58 \times 10^2 \frac{(P_i P_j)^{1.07}}{MT_{ij}^{2.49}} \quad (2.1)$$

where,

V_{ij} = the total person trips from i to j per day in any one direction between districts i and j by all modes of public transportation.

P_i = the population of district i .

P_j = the population of district j .

MT_{ij} = the minimum travel time between the district pairs i and j using any mode of public transportation.

The population variable in the above equation was replaced by four variables to account for travel variations due to the population growth of urban areas, the disparity between urban centres and extent of under-developed areas (UP_{ij} urban product, UD_{ij} urban difference and E_{ij} estate). The variable MT_{ij} was also replaced by a generalised cost function GC_{ij} and a variable to measure the impedance due to waiting time and transfers (R_{ij}). The recalibrated model form is as follows:

$$V_{ij} = \alpha_c \frac{(P_i P_j)^{\beta_1} \times \exp[\beta_2 UP_{ij} + \beta_3 UD_{ij} + \beta_4 E_{ij}] \times \exp \sum_k IV_k}{\exp[\gamma_1 R_{ij} + \gamma_3 GC_{ij}]} \quad (2.2)$$

where,

V_{ij} = The total person trips from i to j per day in any one direction between districts i and j by all modes of public transportation

P_i = The population of district i .

P_j = The population of district j

UP_{ij} = Product of urban population proportion

UD_{ij} = Difference of urban population proportion

E_{ij} = Estate variable

GC_{ij} = The generalised cost

R_{ij} = the variable to measure the impedance due to waiting time and transfers

IV_k = Intrinsic variable

α, β, γ = The coefficients obtained from the regression.

Aldian and Taylor (2003) proposed to combine behavioural travel demand model with a direct demand model to overcome the drawback that it cannot include qualitative variables. Fuzzy set theory was employed into the multinomial logit model to account for the uncertainties and to bring the behavioural basis into direct demand approach to predict the trip interchanges. The model developed has a disaggregate structure, but the deterministic part is at the aggregate level. In this model, the trip from the origin zone to the destination zone is a function of trip generation power of each zone (V_i) and zone attractiveness (V_{ij}). Fuzzy multi-criteria analysis is used to calculate V_i and V_{ij} . Population density and gross regional domestic product were used to determine V_i while road user cost and the number of hotel rooms determine V_{ij} . Then using the multinomial logit model, probability of travellers in the zone ' i ' chose zone ' j ' (P_{ij}) is determined.

$$P_{ij} = \frac{e^{\beta_d V_{ij}}}{\sum_{j=1}^m e^{\beta_d V_{ij}}} \quad (2.3)$$

β_d is the parameter of destination choice. Using the zone attractiveness value (V_{ij}), the expected utilities (W_i) of travelling from zone ' i ' is calculated.

$$W_i = \frac{1}{\beta} \ln \sum_{j=1}^m e^{\beta_d V_{ij}} \quad (2.4)$$

So that the probability of people of zone ‘i’ deciding to travel is equal to

$$P_i = \frac{e^{\beta_i(V_i+W_i)}}{1+e^{\beta_i(V_i+W_i)}} \quad (2.5)$$

Where β_i is the parameter of deciding to travel. The trip generation model is given by

$$T_i = f(P_i) \quad (2.6)$$

and the modelled trips between ‘i’ and ‘j’ are given by

$$T_{ij} = T_i \cdot P_{ij} \quad (2.7)$$

In a study by Asri and Sugie (2003), a simultaneous demand model for passenger travel in Indonesia was developed for inter-regional travel in which socioeconomic variables, regional conditions and existing transport systems are considered. The study used National OD survey data of 1996 for passenger travel demand analysis. The data includes passenger trips, travel cost, and service frequency of each mode. Population and per capita income (in the form of Gross Regional Domestic Product per capita, excluding oil and gas) are the socioeconomic variables included in the model. Ordinary Least-Square (OLS) regression was used to estimate the parameters of the model based on the Gauss-Newton iterative algorithm using Time Series Processor (TSP) Software version 4.3.

To enhance the accuracy of the singly constrained gravity model, Herijanto and Thorpe (2005) proposed the principles of fluid analogy theory. This method uses the iterative process to generate and allocate trips in the origin-destination matrix and then estimate the number of trips produced by and attracted to individual zones. It was found that the trip length distribution outputs were improved using this methodology.

2.8.2.2. Direct Demand Models

Recently, direct station level forecasting models have been developed which provides the fine-scale details at neighbourhood levels. Cervero (2006) suggested alternative

methods to traditional modelling of neighbourhood-scale projects to overcome the shortcomings of four scale method.

Kikuchi and Miljkovic (2001) developed ridership models at bus stop level and examined the use of fuzzy inference based model in predicting the bus ridership at individual stops, which was similar to the cross-classification approach. Transit service quality, accessibility, bus stop conditions and demographics are the data groups used in the hierarchical fuzzy inference process, which consist of two-level aggregations. The results were compared with the regression model and the ANN model.

Direct demand model proposed by Cervero et al. (2010) estimated the number of boarding (and/or alighting) passengers at a stop for a defined time period. Three sets of variables were used for demand estimation: service attributes, location and neighbourhood attributes and bus stop/site attributes. This helps in a relatively fine resolution of analysis of factors influencing the ridership. Ordinary least square regression was used for forecasting the Bus Rapid Transit (BRT) ridership, which yielded high prediction accuracy.

In the direct demand approach, models are aggregate in nature with the assumption that the relationship between the dependent and independent variable remains unvaried. But the experience is that the transit ridership and independent variables vary across space. To quantify such variations, geographically weighted regression (GWR) approach was adopted for modelling. GWR is a group of the spatially adjusted model in which weights are assigned to each observation based on its distance from a specific location or focal point.

Zhao et al. (2005) applied GWR to account for the spatial heterogeneity in the direct demand model, although the GWR model is not spatially transferrable. In this study, GWR models were used for predicting public transit use for home-based work trip purpose. A large number of potential variables for transit use was considered, including demographic, socioeconomic, land use, transit supply quality and pedestrian environment variables. The best predictors were identified through model estimation as regional accessibility to employment and percentage household with no car as global

variables and three local variables-employment density, the average number of cars in a household with children and percentage of the Black population.

Chow et al. (2006) developed a GWR model to explore the spatial variability in the strength of the relationship between transit uses for home-based work trips. The predictors included are demographics, socioeconomics, transit supply and quality and pedestrian movement. The study claimed that the GWR model has improved accuracy in predicting transit use over other linear regression models and indicated that, the effect of independent variables on transit use vary across space.

Gutierrez et al. (2011) developed a rapid response ridership model combining GIS, distance decay function and multiple linear regressions. The number of passengers boarding at each station was estimated as a function of station characteristics and service area. In this study, the distance decay function accounts for the decreasing trend of public transport usage as the distance increases.

Bus stop level ridership was developed using spatial proximity method (SPM) and spatial weight method (SWM) by Pulugurtha and Agurla (2012). Geographical Information System (GIS) was used to capture the spatial attributes of demographics, socio-economic and land use. SPM estimated bus-stop transit ridership for each buffer of widths 0.25, 0.5, 0.75, and 1 mile. The model with buffer width (D) data that has better goodness of fit was selected as the best model. In SWM weights of $1/D$, $1/D^2$, and $1/D^3$ were considered for the buffers and the weight that better estimates the ridership was considered.

Cardozo (2012) proposed the application of GWR to forecast boarding at Madrid Metro stations using network distance rather than Euclidean distance. The parameters considered for model development are the number of metro lines, workers, employment and number of suburban bus lines. The information provided by GWR regarding the spatial variation of elasticities and their statistical significance provided more realistic results than the ordinary least square model.

Tsai et al. (2012) developed a public transport demand model incorporating land use density, diversity, design and accessibility for the Sydney Greater Metropolitan Area.

geographically weighted regression method was used to identify spatial variations of the land use variables and their impact on the public transport demand at the global level and local level. The global model suggested that increasing land use density, accessibility and walkability have a positive impact on public transport demand. The local model shows that the impact of land use characteristics on public transport varies spatially, and some of the estimated parameters have different signs in different areas.

Pinedo (2014) explored the possibilities of spatial integration measures, transport network accessibility and urban environment features to constitute a direct demand model at bus stop level. Zhao et al. (2014) investigated the factors affecting rail transit ridership at station level and station-to-station level by using the multiple regression model and multiplicative model, respectively. Land use mix, intermodal connection, station context and travel impedance are the various independent factors affecting the ridership.

2.8.2.3. Demand Elasticity Models

The demand elasticities are crude and approximate measures of aggregate responses in a market condition. The main advantage of an elasticity model is that it is easily understood, empirically estimable, tested by experience and directly usable (Goodwin 1992). Demand elasticity models are commonly used for developing transport policy. The product-specific demand functions can be estimated using an elasticity model. For public transport, the elasticity model takes the form:

$$D_{PT} = f(P, S, L, A) \quad (2.8)$$

Where,

- D_{PT} = The demand for public transport,
- P = The price vector which includes price of public transport and the alternative modes,
- S = The vector of socio-economic characteristics,
- L = The vector of level of service characteristics,
- A = The activity options

For transport modes price (P) is replaced by generalised cost (GC).

$$GC = P + vJT \quad (2.9)$$

where,

GC = Generalised cost

P = The price vector which includes the price of public transport and the alternative modes

v = The value of time

JT = The journey time.

For public transport, GC can be further divided into in-vehicle time (IVT), wait time (WT), access and egress time, *i.e.*, walking time (WK) and interchanges (INT) (Preston 2015).

$$GC = P + v(IVT + WT + WK) + INT \quad (2.10)$$

Elasticities are indicators of sensitivity of the dependent variable to policy variables. The elasticity is the proportionate change in the dependent variable with respect to a given proportionate change in the relevant independent variable. The most important elasticities are the direct elasticity and cross-elasticity, first relate to the service or goods under consideration and the second to competing options or goods (Ortuzar and Wilumsen 2011). The elasticity based framework is unsuitable for use with new services since the base level of demand is zero. Oum et al. (1990) reviewed the price elasticities of demand for transport. The study showed that, since transportation is a derived demand, it tends to be inelastic. Goodwin (1992) has reviewed demand elasticities with reference to short and long-run effects of price changes.

Aggregate urban passenger travel demand model was developed by Domencich *et al.* (1968) on the basis of consumer theory; the trips were disaggregated by trip purpose and then modelled to measure the aggregate demand based on socioeconomic and land use characteristics, travel time and travel cost for specific and alternate modes. On the same basis, Talvitte (1973) developed a model for work trips considering the same variables as that of Domencich et al. (1968). Both these models had the issue of collinearity. The direct travel demand model for bus and rail modes, developed by Talvitie (1973), found that the intermodal work trips are dependent on travel time and

cost of each available mode, size and socio-economic characteristics of the workforce and the number of jobs. This model predicts the zone to zone traffic volume by travel mode (Auto, Bus and Rail) and sensitivity of travel demand with respect to transportation system characteristics. The functional form of the model is given as:

$$V (ij|k) = F \{(L_{ij}|A), (L_{ij}|B), (L_{ij}|R), (SE_i), (SE_j)\} \quad (2.11)$$

where,

$V (ij|k)$ = The number of round trips for work purposes from i to j via mode k (k = A, B, R)

$L_{ij} : k$ = vector of level-of-service characteristics for mode k; k = A, B, R

SE_i = A vector of traveller's socio-economic characteristics (origin zone i)

SE_j = A vector of socio-economic characteristics for the destination zone j

Elasticity constraints were imposed while estimating the model's coefficients to ensure that, the travel demand elasticities behave according to the economic theory of consumer behaviour. Direct elasticities to be less than or equal to zero and cross-elasticities to be greater than or equal to zero was considered. Considering the complex nature of travel behaviour, an improved set of constraints may be considered to give more accurate results.

Schmenner's (1976) study stated that the demand for bus service should be measured in seat-miles or seat-hours, which is impractical as no data exist in this regard. But data corresponding to revenue per bus-mile or bus-hour is available. This demand study related average revenue per bus-mile along a route and average revenue per bus hour along a route to the following influencing factors: average bus fare for the route or average fare per mile for the route, sum of average auto operating costs over the route and average downtown parking, peak hour of bus headway, average bus and auto speeds along the route, population within the two blocks of bus route, season for which data are derived and route-related socio-economic variables. This demand model developed is exclusively for individual bus routes within a city. Revenue estimates cannot be calculated using this model if there are many transfers or if special point to point express services are offered.

Rus (1990) in a study estimated short-term elasticities and elasticities which account for delayed demand responses. With monthly time series data, aggregate and disaggregate elasticities were estimated. The passenger trips were explained by variations in real fare, bus-kilometres run and seasonal factors. A time trend was used to account for important parameters like income, car ownership, etc., whose effect cannot be neglected because of the lack of appropriate data. The basic model proposed in this study is a double log, with an incorporation of a dynamic specification to account for delayed adjustments of demand to changes in fare and service level. The model is as follows:

$$q_{jt} = \alpha f_{1t}^{\gamma_{j1}} f_{2t}^{\gamma_{j2}} \beta_t^{\beta_j} \exp(\gamma_{j3} T_t + \sum_i \lambda_i D_i + \omega_i H_i) e_t \quad (2.12)$$

where,

- q_{jt} = Passenger trips per month with cash fare ($j = 1$) or multiple-ride ticket ($j = 2$)
- α = Constant term
- f_{1t} = Deflated price of an ordinary ticket
- f_{2t} = Deflated price of a trip with multiple ride ticket
- β_t = Vehicle-kilometers per month
- T_t = Time trend
- D_i = Dummy variables ($i= 1 \dots 11$, January to December excluding one)
- H_i = Dummy variable to account for major disruptions in service
- e_t = Error term

In this study, changes in the passenger trips are explained by variations in real fare and bus-kilometres run. Seasonal factors and time trend is used to account for an important set of variables like income, car ownership, etc., because of the lack of appropriate data.

Liu et al. (2012) performed an elasticity analysis of urban public transport (UPT) priority in Beijing to measure the elasticity of UPT price demand. It also discussed the cross price elasticity of demand among UPT, car and bicycles, they found that the low price strategy of Beijing has not been able to attract the car users but has attracted bicycle users.

Demand elasticity with respect to various policy and environmental variables that influence the demand for public bus transport in India was studied by Deb and Filippini (2013). Monetary and non-monetary variables were explored. Panel data was used for a thorough analysis of price, income and service quality as parameters for estimating demand elasticity using the log-linear model. Demand was defined as total passenger kilometer (pkm) while the regressors include public transit fare (p), per capita income (w), service quality, and other social and demographic variables like proportion of population in the labour force ($work$), literacy rate (lit) and per capita private vehicle ownership (s). The density of coverage (q) was used as the measure of service quality. The total population (pop) was included to isolate the effect of the size of the market. The model specification is the following:

$$pkm = f(p, w, q, s, pop, work, lit) \quad (2.13)$$

Daldoul et al. (2015) applied dynamic panel model with a log-linear functional form to the Tunisian regional transport companies to analyse travel behaviour of citizens and results shows that service quality, the mean price and active population are the most significant variables affecting public transport demand. It was found that the price elasticity in the short-run was -0.4 and -0.8 in the long-run. They suggested that reducing the price will induce passengers in public transport.

Sumetha and Mathew (2016) analysed the demand for public transport work trips through behavioural mode choice models for Tiruchirappalli city. Binomial logit model and multinomial logit model were used for disaggregate behavioural model formulation. The direct and cross elasticities for various factors affecting the demand for public transport are determined.

2.8.3. Time-Series Techniques

Time-series is a chronological sequence of observations recorded over an equal interval of time. These observations can be continuous over time or a discrete set of time points (Chatfield 2000). Thus, the time-series represents a quantitative measure of a variable y_t recorded at a time t . It is mathematically expressed as y_t , where $t = 0, 1, 2, 3, \dots$. The observation made upon a single variable of a process forms a univariate time-series

(Fathima and Jothiprakash 2014), *i.e.*, it does not deal with relationships of other variables; while, multivariate time-series model is an extension of the univariate model and involves multiple variables of interest. Recently, application of time-series has been applied in various fields of transportation engineering like traffic flow prediction (Kumar and Vanajakshi 2015), for modelling traffic fatalities (Yannis et al. 2011), demand forecasting of public transport (Chidlovskii 2017), and estimating bus dwell time (Rashidi 2015).

According to Montgomery et al. (2015), time-series models employ the statistical properties of past data to build a model and then calibrate this model by estimating the unknown parameters, generally by least squares. Time-series model thus includes four steps in the process of model development. The first step is the analysis of data, which comprises of using right data mining tools to observe, understand and examine the historical data to identify the underlying pattern in the data. The second step is the model selection and calibration; where calibration is the estimation of unknown model parameters. The next step is model validation, *i.e.*, evaluating the performance of the model. Finally, the model is deployed for forecasting application and monitored periodically to ensure that the model is performing satisfactorily. The time-series analysis has been performed predominantly using parametric methods, but, recently soft computing techniques and non-parametric approaches like fuzzy logic (FL), artificial neural network (ANN) and other machine learning techniques are becoming increasingly used by the researchers.

2.8.3.1. Parametric Time-Series Model

Parametric time-series models are well-structured and have a finite number of parameters. The modelling process includes model identification, parameter estimation, diagnostic checking and forecasting (Chen et al. 1997). The parametric models postulate global structures for the underlying time-series data, which has high interpretability. Hence, the classic parametric models are still widely used in the analysis and modelling of time-series data. The parametric time-series modelling methods include exponential smoothing techniques, the family of AutoRegressive (AR) models, and AutoRegressive Conditional Heteroscedasticity (ARCH) models.

Exponential Smoothing Techniques: Exponential Smoothing Techniques are simple tools used for smoothening the time-series data and subsequently forecasted to get future observations. The time-series data is smoothened using a smoothening parameter by eliminating the noise in the data to get the pattern in the series. The weighted averages of the past observations are used for forecasting (Hyndman and Athanasopoulos 2018). When new data points are available, the weights of the older data points decay exponentially. Newer the data, the higher the weightage and older the data lower is the weight. The weights assigned to the observations are determined by the smoothing parameters. Exponential smoothing techniques consist of different variants like simple exponential smoothing (SES), double exponential smoothing (DES) and Holt-Winters' method. These variants of exponential smoothing are intended for different scenarios.

Simple exponential smoothing can be used when the data has no significant trend and seasonality. In double exponential smoothing, a trend variable is incorporated to account for the slope or trend in the time-series data. The SES has a single smoothening equation for level while the DES has two smoothening equations for level and trend. When the data has both trend and seasonality, seasonal Holt-Winters model is used. Holt-Winters model has three smoothening equation- for level, trend and seasonality. Holt-Winters model has two variants- additive model and multiplicative model. The additive model is used when the seasonality is almost constant throughout, while the multiplicative model is used when the seasonality changes proportionally to the level of the series.

Holt-Winters Exponential Smoothing Methods: Holt-Winters forecasting model consists of four equations- one forecast equation ($\hat{y}_{t+h|t}$) and three smoothening equations, one each for level (l_t), trend (b_t) and seasonality (s_t). These four equations are the component form representation of Holt-Winters exponential smoothing model. Holt-Winters additive model and multiplicative model with and without damping are explained in the following paragraphs.

Holt-Winters Additive Method: The formulation of Holt-Winters' additive model is given below.

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)} \quad (2.14)$$

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (2.15)$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1} \quad (2.16)$$

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad (2.17)$$

Where,

$\hat{y}_{t+h|t}$ = The forecast equation. It is the estimate of y_{t+h} based on the data y_1, y_2, \dots, y_t

l_t = The level (or smoothed value) of series at time t ,

b_t = The estimate of trend or slope of the series at time t ,

s_t = The seasonal estimate of the time series at time t ,

α = The smoothing parameter for level,

β^* = The smoothing parameter for trend,

γ = The smoothing parameter for seasonality,

h = The forecast horizon

m = The seasonality in the data

k = The integer part of $(h - 1)/m$ for confirming that the seasonal indices come from the latest observations.

Holt-Winters Multiplicative Method: The components of Holt-Winters Multiplicative Model, having the parameters, as explained above, is given in the following equations.

$$\hat{y}_{t+h|t} = (l_t + hb_t)s_{t+h-m(k+1)} \quad (2.18)$$

$$l_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (2.19)$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1} \quad (2.20)$$

$$s_t = \gamma \frac{y_t}{l_{t-1} + b_{t-1}} + (1 - \gamma)s_{t-m} \quad (2.21)$$

Holt-Winters Damped Additive Method: The trend damping can be performed on both additive and multiplicative models. The damping parameter used in the equation is given by ϕ . The component equations of the damped additive model are as follows:

$$\hat{y}_{t+h|t} = l_t + (\emptyset + \emptyset^2 + \dots + \emptyset^h)b_t + s_{t+h-m(k+1)} \quad (2.22)$$

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + \emptyset b_{t-1}) \quad (2.23)$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)\emptyset b_{t-1} \quad (2.24)$$

$$s_t = \gamma(y_t - l_{t-1} - \emptyset b_{t-1}) + (1 - \gamma)s_{t-m} \quad (2.25)$$

Holt-Winters Damped Multiplicative Method: The damped multiplicative model form is given below.

$$\hat{y}_{t+h|t} = l_t + (\emptyset + \emptyset^2 + \dots + \emptyset^h)b_t + s_{t+h-m(k+1)} \quad (2.26)$$

$$l_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(l_{t-1} + \emptyset b_{t-1}) \quad (2.27)$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)\emptyset b_{t-1} \quad (2.28)$$

$$s_t = \gamma \frac{y_t}{l_{t-1} + \emptyset b_{t-1}} + (1 - \gamma)s_{t-m} \quad (2.29)$$

Holt-Winters models are used in the various forecasting problems such as air transport demand (Bermúdez et al. 2007; Dantas et al. 2017), electricity demand (Taylor, 2003), automobile petrol demand (Li et al. 2010), agricultural commodities (Brandt and Bessler 1983), and modelling of traffic flow (Ghosh et al. 2005). Li et al. (2010) considered various forecasting techniques like linear trend model, exponential trend model, quadratic trend model, Holt's linear model, Holt-Winters model, partial adjustment model and AutoRegressive Integrated Moving Average (ARIMA) model to evaluate the automobile fuel demand forecasting performance empirically. It has been identified that for both seasonal and non-seasonal data, quadratic trend model gave optimum forecasts. Ghosh et al. (2005) used three different time series models, viz. random walk model, Holt-Winters' exponential smoothing technique and seasonal ARIMA model for modelling traffic flow in Dublin. The data used for the study was obtained from loop detectors in the city centre of Dublin. It was found that data had a strong daily seasonal pattern and a stable trend. The forecasts of seasonal ARIMA and Holt-Winters Model was highly competitive with the Holt-Winters method having slightly better forecasts. Brandt and Bessler (1983) compared the performance of seven

forecasting models by examining their performance over 24 quarter. The methods include various exponential smoothing methods, ARIMA model, expert judgement, econometric model and simple average models. The performance of the models was evaluated using the Mean Absolute Percentage Error (MAPE) and Mean Square Error (MSE). The data having a quarterly seasonality was forecasted accurately by the ARIMA models. Taylor (2003) studied the electricity demand forecasting using data having double seasonality. The electricity demand data recorded at half-hourly intervals contains both intra-day and daily seasonality. To capture the double seasonality in the data, a double seasonal Holt-Winters method was used. It was found that, the double seasonal method outperformed the traditional Holt-Winters method.

AutoRegressive Models: The next class of parametric time-series models are the family of autoregressive models like AutoRegressive (AR), Moving Average (MA), AutoRegressive Moving Average (ARMA), AutoRegressive Integrated Moving Average (ARIMA) Seasonal ARIMA (SARIMA), AutoRegressive Fractionally Integrated Moving Average (ARFIMA), AutoRegressive Conditional Heteroscedasticity (ARCH), Generalised ARCH (GARCH), Exponential GARCH (EGARCH), Threshold AutoRegressive (TAR), Non-linear AutoRegressive (NAR) and Non-linear Moving Average (NMA). These classical approaches of time-series forecasting are based on the regression analysis.

The common approach for modelling univariate time-series is the AutoRegressive (AR) model. AutoRegressive model, also known as conditional model or Markov model, is a linear predictive regression model, generally denoted as $AR(p)$. $AR(p)$ model establishes that a realization at time t is a linear combination of p previous realizations and an error term (white noise). In other words, $AR(p)$ models use specific lagged values of y_t as predictor variables with error term having zero mean, finite variance, and independent and identically distributed (iid) random variable. Moreover, the past values of y_t have a linear influence on the conditional mean of y_t . The model formulation of the $AR(p)$ model is given in Equation 2.30.

$$\hat{y}_t = \sum_{n=1}^p \phi_n y_{t-n} + \varepsilon_t \quad (2.30)$$

Where,

\hat{y}_t = The estimate of y_t based on the data y_1, y_2, \dots, y_{t-1}

p = The order of AR term

ϕ_n = The coefficient of AR term

ε_t = The error term

The AR models are employed for forecasting stationary processes; the parameters can be estimated using least squares, Kalman filtering, Yule-Walker, expectation maximisation, burg, forward back algorithm, etc.

The Moving Average (MA) model of order q or $MA(q)$ establishes the correlated noise structure in the data, *i.e.*, the lagged previous error term or white noise is used to forecast the future values (Chatfield 2000). $MA(q)$ model goes beyond the traditional assumption that the errors are iid. The model formulation of $MA(q)$ model is given in Equation 2.31.

$$\hat{y}_t = \theta_q(B)\varepsilon_t \quad (2.31)$$

Where,

\hat{y}_t = The estimate of y_t

B = The backshift operator given by $By_t = y_{t-1}$

θ = Coefficient of MA term

q = Order of MA term

ε_t = The error term

The AR process requires relatively long series and have infinite autocorrelations which decay with lag while MA can be used for even very short series. The estimation of model parameters of MA is more complicated than AR as the white noise is not observable.

The combined representation of the $AR(p)$ and $MA(q)$ model gives the ARMA (p, q) model of order p and q having the structure as given in Equation 2.32 and parameters as given above.

$$\phi_p(B)\hat{y}_t = \theta_q(B)\varepsilon_t \quad (2.32)$$

ARMA has a flexible stationary stochastic process having the properties of both AR and MA methods. AR(p) and MA(q) models are restricted to stationary data. But, the data is non-stationary in many of the practical scenarios, and it has to be differenced to make it stationary. The ARIMA (p, d, q) models or commonly known as Box-Jenkins models having an AR term, MA term and the differencing operation for making the data stationary by eliminating trend. Since the ARIMA model is a combined model; the forecast accuracy is considered to be better than the other stochastic time-series model. The ARIMA model can be extended to seasonal data to include a seasonal AR term and seasonal MA term and is represented as ARIMA (p, d, q) (P, D, Q) s or SARIMA, where the order in lower case indicates the non-seasonal part, and the order in upper case indicates seasonal component with 's' being the number of periods per season. The general forecasting equation of ARIMA (p, d, q) (P, D, Q) s model is defined by Equation 2.33 (Hyndman and Athanasopoulos 2018).

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D(y_t - ct) = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (2.33)$$

Where,

B = The backshift operator ($By_t = y_{t-1}$)

c = The drift parameter

s = The periodicity or seasonal period

d = The degree of non-seasonal differencing

D = The degree of seasonal differencing

p = Order of AR term

q = Order of MA term

P = Order of seasonal AR term

Q = Order of Seasonal MA term

ϕ = The coefficient of AR term

θ = The coefficient of MA term

Φ = The coefficient of seasonal AR term

Θ = The coefficient of seasonal MA term

ε_t = The random error at time 't'

y_t = The prediction result of the ARIMA model at time 't'.

According to Chatfield (2000), for the effective fitting of ARIMA model requires a minimum of 50 observations. This method is suitable when dealing with a nonstationary shorter time-series data. The models like AR, MA, ARMA and ARIMA can be effectively used when the time-series data have standard features like a normal distribution, the linear relationship among lagged variables, invertibility etc. Moreover, the limitation of the family of AR models is that it cannot be applied to data sets having volatility. Also, minimum of 60 data points are required for effectively modelling using time-series data.

When the time-series suffer from heteroscedasticity, *i.e.*, the variances of error terms are not equal, or some of the error terms are reasonably large, autoregressive conditional heteroscedasticity (ARCH) models, later expanded to generalised autoregressive conditional heteroscedasticity (GARCH) by Bollerslev (1988), are used for modelling the data sets. This technique model the variances in error term (heteroscedasticity) and prediction is for the variances of each error term. Moreover, nonstandard features like time irreversibility, non-normal distribution, multimodality, and non-linear relationship with lagged values, which are commonly found in practical problems, requires non-linear modelling approaches. ARCH models for non-constant conditional variance and Threshold Autoregressive models for the conditional mean are two approaches of non-linear parametric time-series modelling.

An ARCH model is a stationary non-linear model used to describe a changing volatile variance. This model is used when the conditional variance of y_k evolves according to an Autoregressive process. The generalised ARCH model, known as GARCH, is the most commonly used variance model. The GARCH (p, q) model is defined by the following Equations.

$$y_t = \sigma_t \varepsilon_t \quad (2.34)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (2.35)$$

Where,

y_t = The observation at time t
 σ_t = The volatility or standard deviation
 σ_t^2 = The conditional variance of y_t given the information upto $t-1$
 p, q = Order of the GARCH model
 $\omega > 0, \alpha_i \geq 0, \beta_j \geq 0$ and $\varepsilon_k \sim iid(0,1)$.

The severe limitation of these models is the pre-assumed linear form of the associated time-series, which becomes inadequate in many practical situations. To overcome this drawback, various non-linear non-parametric models have been proposed in the literature; however, from the implementation point of view, these are not so straightforward and simple as the ARIMA models. Recently, artificial neural networks (ANNs) have attracted increasing attention in the domain of time-series forecasting. Although initially biologically inspired, but later on, ANNs have been successfully applied in many different areas, especially for forecasting and classification purposes.

The readily available monthly data was used by Gaudry (1975) to estimate the demand for public transit using multivariate time-series approach. The aggregate demand for public transit trips during period t was explained by the prices of trips by the various modes and of non-transportation goods, time and comfort characteristics of relevant modes, income and levels of activities. The Box-Jenkins procedure, along with linear regression techniques, are used to obtain the estimates. Similarly, the time series model proposed by Mullen (1975) utilises fare and service elasticities to explain passenger demand. The main disadvantage of using this model is that the results explain the past, but it does not necessarily predict the future because of the assumptions of constant elasticities. Frankena (1978) used the annual citywide time-series data for demand estimation using two-stage least square method. Geographic characteristics (spatial structure, climate and size of urban area) of the urban area were also considered for modelling the demand. Constant elasticity was considered for all variables except for fare.

In the study by Tegner and Holmberg (1998), nonlinear two-step model (Quasi- Direct Format model) was developed for a coordinated choice of ticket type and the number of public transport trips using aggregate time-series monthly data for Stockholm region.

Three models were tested to find the best fit model: one linear model, one Box-Cox transformed model and one Box-Cox AutoRegressive model (BCAU) and found that BCAU model performs substantially better than the other two models.

The study by Andreoni and Postorino's (2006) developed both univariate and multivariate ARIMA model to forecast air transport demand. The time-series data analysis was performed by the Box-Jenkins approach. The univariate model, followed by a multivariate model considering two explanatory variables (per capita income and the yearly number of passengers), was developed. It was found that the univariate model fits better than the multivariate model. Time series econometric method, together with annual data from 1983-2008, was used by Wijeweera et al. (2014) to study the rail passenger demand in Perth. The number of boarding passengers at the time 't' was regressed over the explanatory variables such as fare, per-capita income, fuel price, population, the number of kilometres run annually and fatality which is the number of accidental deaths relating to the Australian rail sector.

A method of statistical modelling of passenger demand by using univariate time-series theory was proposed by Cyprich et al. (2013). In this study, three different approaches—multiple regression, exponential smoothing, and Box-Jenkins were used in order to determine the most accurate and reliable model. It was found that the ARIMA model represents a more precise and reliable passenger demand forecasting method in spite of the fact that this model does not conform to the requirements of normality of probability distribution of its standardized residuals. The non-normality of the probability distribution of residuals of a stochastic linear model is a common phenomenon which can be ignored for economic time series model.

A methodology was proposed by Tsai et al. (2013) for forecasting public transport demand for the Sydney Greater Metropolitan Area using a Partial Adjustment Model (PAM) and an Autoregressive Integrated Moving Average (ARIMA) model. ARIMA model uses the monthly bus and train boarding data while PAM model estimates demand elasticities such as fare, land use characteristics, socio-demographics of public transport users and level of public transport service. The study suggested that the

ARIMA model have better prediction accuracy in short-term forecasting, while PAM is used to determine the demand response to various policy scenarios.

Williams and Hoel (2003) presented the theoretical basis for modelling traffic flow as a univariate seasonal ARIMA process. It was based on Wold decomposition theorem and asserted that a one-week lagged first seasonal difference applied to discrete interval traffic data will yield weakly stationary transformation. An empirical study using the actual intelligent transportation system data was found to be consistent with the theoretical hypothesis. The Mean Absolute Percentage Error (MAPE) was found to be 8.9%. Kumar and Vanajakshi (2015) used limited traffic data for empirically validating the seasonal ARIMA method by short-term (24 hours) traffic forecasting. The traffic flow data of a three-lane arterial highway for three consecutive days were used for modelling. The developed model was validated by performing a 24-hour forecast and comparing with actual values. MAPE for the models were between 4-10%, which is acceptable in all ITS applications.

A time series forecasting framework based on the Box-Jenkins method is developed for the public transportation system by Anvari et al. (2016). The proposed methodology is completely automated forecasting tool which reduces the dependency on the econometric knowledge requirement. The algorithm starts with original data checking for stationarity and then performs the related transformations. Maximum likelihood estimation (MLE) is used to estimate the parameters of AR and MA terms. To find the best model, adjusted Root Mean Square Error (RMSE) was used.

Xue et al. (2015) developed a short term bus passenger demand prediction model based on a time-series approach. Passenger demand aggregated in 15 min interval, of a bus route over four months, were used to generate time-series. Three time-series were developed, named weekly, daily, and 15 min time-series. The prediction models ARMA, SARIMA, and ARIMA were developed, specific to capture different features of time-series. The GARCH model is applied to improve the accuracy of the ARIMA model prediction.

Gavirangaswamy et al. (2013) used ARIMA method for developing short-term and long-term road traffic volume forecasting. Further GARCH method was used to improve the ARIMA model to account for volatility. It was observed that SARIMA and ARIMA-GARCH exhibit similar traffic prediction. However, ARIMA-GARCH had slightly better forecast performance in terms of RMSE.

Zhang et al. (2014) proposed a methodology to forecasts traffic flow by decomposing the data into three modelling components: an intra-day or periodic trend by introducing the spectral analysis technique, a deterministic part modelled by the ARIMA model, and the volatility estimated by the GJR-GARCH model. It was observed that treating the periodicity and volatility individually can generate better forecasting accuracy.

As a promising alternative to the statistical modelling of time series predictions, soft computing techniques have been proposed. The application of soft computing techniques in the field of transportation is due to their ability to handle the data of qualitative and quantitative nature and to solve the problems involving complexity, uncertainty and imprecision. Researches have been carried out in the application of soft computing techniques in the field of transportation, which includes traffic control and management, transportation planning, logistics of freight, design and construction of transport facilities.

2.8.3.2. Artificial Neural Network Method

Artificial Neural Network (ANN) is a method which simulates the functioning of a human brain. The main component of ANN is the “Artificial Neurons” which are the building block of neural network. Thus, a neural network consists of a set of interconnected layers of neurons. Each neuron of a layer is connected to each neuron of the next layer. These networks mimic the functioning of a human brain by understanding the patterns in the data, analysing the relationship between the variables, learning from the experience (training) and then provide results based on the knowledge acquired. In other words, ANN follows a self-learning algorithm and an auto-correction technique.

In the study by Guan et al. (2002), ANN method was employed as a forecasting tool to forecast the ridership in Beijing public transportation from 2001-2005 based on the past 20 years' historical data. This method combines the qualitative and quantitative data in which ANN handles the quantitative knowledge while SCENARIO method of systems engineering handles the qualitative information. For SCENARIO method, variables considered are service levels of public transportation, alternative modes, population, income and expenditure levels. A standard three-layered backpropagation network was used for the learning process.

To model the number of bus tickets sold in public transport, the neural network method was used in a study by Bošnjak et al. (2005). In this study, the number of sold bus tickets were divided into two main categories: tickets sold by full price and tickets sold by a reduced price. The input variables were the type of the day, price of gasoline/ price of tickets, earnings/ expenditures, air pressure, lowest night temperature, air moisture, the direction of the wind, speed of the wind, cloudiness, general weather condition, rainfalls, condition of the ground and height of snow. Two models: a linear and a non-linear model were formulated. Linear model used multiple linear regression, and feedforward backpropagation neural network was used as the non-linear model. When compared to the linear model, the non-linear neural network model was more successful, having a smaller prediction error.

Celikoglu and Akad (2005) estimated the public transport trips using feedforward backpropagation (FFBP) and compared it with the AutoRegressive (AR) model. Time series data from AKBIL (an electronic ticket machine data integrated to all PT modes) was used to obtain the average half hourly PT data. It was found that FFBP had a disadvantage of generating negative values after training. To overcome this, Celikoglu and Cigizoglu (2006a) used two different ANN algorithms, FFBP and radial basis function (RBF) for the purpose of trip flow forecasting and compared with the conventional AR model and found that RBF NN provides better prediction results than FFBP by generating non negative values and closer estimates while AR model underperformed. Subsequently, another study was carried out by Celikoglu and Cigizoglu (2006b) to check the usefulness of generalised regression neural networks

(GRNN). The study pointed out that GRNN did not provide a negative forecast, and the local minima problem of FFBP was not encountered in GRNN.

Panapakidis et al. (2014) developed a hybrid forecasting model that involves the combination of unsupervised clustering algorithm and supervised ANN algorithms using historical travel data along with the weather data for the day-ahead forecasting. The results indicate that the model leads to increased accuracy and is able to simulate nonlinearities in the bus load. The model was developed using historical load data of the period 2006-2009 and tested on the 2010 load data of a suburban area of Northern Greece.

Gong (2010) proposed ARMA-GRNN technique to capture both the linear and non-linear characteristics of inter-city passenger demand. The flight demand data from 1991 to 2008 in the Beijing-Shanghai corridor was selected for the study. The results revealed that the linear part modelled using ARMA and non-linear part modelling using GRNN can effectively forecast the passenger demand with a maximum error of 1.08%.

In the study by Tsai et al. (2005), two dynamic neural networks were developed to forecast short-term railway passenger demand. The first network is similar to the nonlinear AutoRegressive model, while the second network is an extension of the first network to integrate the recurrent pattern. The results show that both the networks can perform well with a MAPE of less than 20%.

2.9. MEASURES OF FORECASTING PERFORMANCE

Producing a good forecast is a crucial requirement of any good model. The residuals of model fitting cannot be considered as a reliable source of forecast errors. The forecast performance must be measured by a new data set, which was not used for the model fitting process. In the process of modelling, the data set is divided into two sets- calibration or training data and validation or the testing data. The training data is used for estimating the parameters of the model by the process of model fitting, while the testing data is used to check the forecast accuracy of the model. Generally, 75% of the total data available is used for calibration and 25% for validation. Different combinations of ratios of training data to testing data can be used (90:10, 80:20, 75:25,

70:30, 60:40 and 50:50). Also, as a rule of thumb, the forecast horizon can be chosen as the length of the test data. In the case of time-series modelling, the model development is mainly based on the seasonal component, level component, trend and frequency. Therefore, the length of the test data is to be taken as the forecast horizon, and remaining data points are assigned as the training set data. Data analysis is done prior to model development, and complete data is analysed to identify the above-mentioned components in the data. Hence the forecasts will be based on these components. So, as the length of training set increases, forecast quality will also increase as the forecasts will be in line with the observed parameters. Commonly, in time-series modelling, the length of the test data can be the size of the forecast horizon (Hyndman and Athanasopoulos 2018).

There are various measures for checking the forecast accuracy of a model *viz*, Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Deviation (MAD). The commonly used forecast accuracy measure is MAPE, due to the advantage of being unit free. It is also used to compare the forecast performance between data sets (Hyndman and Athanasopoulos 2018). The MAPE is calculated as given in Equation 2.36:

$$MAPE = \frac{1}{n} \sum_{t=t+1}^{t+n} \frac{y_t - \hat{y}_t}{y_t} \quad (2.36)$$

Where,

y_t = the actual value at time ‘t’

\hat{y}_t = the forecasted value at time ‘t’

n = Number of fitted points.

The lower value of MAPE represents a good forecast. Lewis (1982) describes MAPE greater than 50% as inaccuracy in forecasting, 20-50% as reasonable forecast, 10-20% as good forecast and less than 10% as highly accurate forecast. Also, according to Horowitz et al. (2014), a forecast accuracy of + 7.5% and + 12.5% is acceptable for a forecast horizon of next day and one year, respectively.

2.10. RESEARCH GAP

- In India, Electronic Ticket Machine is one such source of ADC system, the use of which has generated an enormous database which can be used for various applications in the transport industry. There is no documented literature on ETM based public transport demand modelling.
- The accuracy and reliability of the available data are a serious issue in the case of most of the developing countries. Therefore, it is often challenging to study the land use-transportation scenario and measure the accessibility in developing countries by the researchers. In light of this, there is a need to develop an alternative methodology for examining the transport scenario and measure the accessibility using the available data. The methodology considering the population proportion served, proximity to and from the bus stop using GIS methods are not considered thoroughly in the literature.
- To evaluate and optimise the performance of the transit agency effectively, it is required to integrate various techniques. Also, user perception is not considered in the literature to a desirable extent.

2.11. SUMMARY

- The knowledge of the demand for travel and the pattern of travel is essential for public transport planning. There is a need for frequent analysis and updating of the travel demand for operational planning of public transport. The demand data is generally collected through manual methods or automated methods. The manual method of data collection is time-consuming and requires huge financial and human resources. Hence, it is practically impossible to perform manual data collection on a periodical basis for public transport planning. Therefore, technology-based or automated data collection technique can be used. In India, Electronic Ticket Machine is one such source of ADC system, the use of which has generated an enormous database which can be used for various applications in the transport industry. Also, there is no documented literature on ETM based public transport demand modelling.

- The evaluation of service quality is an important practical problem encountered by public transport service providers. The assessment of the quality of service is required to improve user satisfaction and thus to improve the demand for public transport services. The Level-of-Service is an effective measure to understand the user's level of satisfaction for the service provided. The MoUD has developed a service level benchmark to determine the LoS which can be used in this study for determining the city-wide LoS.
- Accessibility is another indicator used for evaluating the public transport scenario. Accessibility to the transit service determines the service attractiveness, and hence, better accessibility increases the demand. Although accessibility has been the focus of research in the past few decades, it still remains a concept that has been poorly defined. The availability of the data as the data-sharing tradition is absent. The accuracy and reliability of the available data are a serious issue in the case of most of the developing countries. Therefore, it is often challenging to study the land use-transportation scenario and measure the accessibility in developing countries by the researchers. In light of this, there is a need to develop an alternative methodology for examining the transport scenario and measure the accessibility using the available data. This calls for the use of Geographic Information System (GIS) in determining the accessibility index as a graphical and descriptive measure of accessibility rather than a quantitative one.
- Additionally, to know the present condition of the transit agency or the service provider, performance evaluation has to be performed. Various methods like DEA, integrated GIS-DEA, AHP, goal programming and fuzzy logic are used by the researchers. To evaluate and optimise the performance of the transit agency effectively, it is required to integrate various techniques. Also, user perception is not considered in the literature to a desirable extent. This is even more so in the Indian context. In this study, we have to address these gaps by performing an integrated AHP-GP procedure taking into account both user and operators perception.

- Further, to achieve the main aim of the study, a literature review was performed to identify various modelling techniques which are used by the transport planners and decision makers for public transport planning with an emphasis on periodical data. Various methods like four-step models, simultaneous demand models and time-series methods were reviewed.
- The initial purpose of the four-step model in the 1950s was traffic forecasting, mainly on a regional scale. Thus, the four-step model is less effective for forecasting the journeys using public transport on more detailed scales (station-level).
- Direct models are an alternate approach in estimating ridership. Such models are relatively rapid and less expensive alternative to the four-step model. Direct station-level ridership forecasting models are normally multiple regression analysis either with linear or multiplicative functional form. It is assumed that the explanatory power in regression models (measured by the coefficient of determination) is the same for the whole set of observations; the problem of parametric instability is ignored. The main drawback of the traditional direct demand model is parametric instability.
- Another approach is the time-series method, which can be used to determine the demand for public transport using Automatic Data Collection system data as it is the best (hypothetically) way to deal with temporal effects.
- The models like AR, MA, ARMA and ARIMA can be effectively used when the time-series data have standard features like a normal distribution, the linear relationship among lagged variables, invertibility etc. The ARIMA model is suitable when dealing with a nonstationary shorter time-series data. The limitation of the family of AR models is that it cannot be applied to data sets having volatility. Also, minimum of 60 data points are required for effectively modelling using time-series data.
- When the time-series suffer from heteroscedasticity, i.e., the variances of error terms are not equal, or some of the error terms are reasonably large,

autoregressive conditional heteroscedasticity (ARCH) models are used for modelling the data sets.

- Non-standard features like time irreversibility, non-normal distribution, multimodality, and non-linear relationship with lagged values, which are commonly found in practical problems require non-linear modelling approaches.
- The excellent feature of ANNs, when applied to time-series forecasting problems is their inherent capability of non-linear modelling, without any presumption about the statistical distribution followed by the observations.
- Both parametric time-series models (ARIMA, GARCH, etc.) and ARNN models give good results when used for linear and non-linear modelling, respectively.
- Even though ARIMA-GARCH can model nonlinear problems; it cannot be applied to complex nonlinear problems. Therefore, a hybrid model which can combine the capabilities of both linear and nonlinear model can be of better use to forecast the real-life problem. Hence, the ARIMA method improved using ARNN and ARCH can be adopted for the development of public transport demand model using time-series ETM data.
- The forecast performance of the models developed can be found using MAPE due the advantage of being unit free, and it is also used to compare the forecast performance between data sets. Also, the forecast performance is measured by a new data set (testing data) which was not used for the model fitting process. Different combinations of ratios of training data to testing data can be used (90:10, 80:20, 75:25, 70:30, 60:40 and 50:50). Also, as a rule of thumb, the forecast horizon can be chosen as the length of the testing data. Moreover, the characteristics of the data are the determining factor for the selection of the ratio of training and testing data.

CHAPTER 3

STUDY AREA

3.1. GENERAL

This chapter deals with the study area chosen for the study. Initially, Trivandrum was chosen as the study area for intra-city and inter-city passenger demand modelling. The data used for the study are mainly ETM data from Kerala State Road Transport Corporation (Kerala SRTC). During the analysis of the Kerala ETM data, it was found that the length of intra-city ETM data available for individual routes varies from 1.5 years to 2 years (2011 to 2013), while inter-city data were not sufficient for time-series modelling. Therefore, inter-city ETM data of Mangalore division Karnataka State Road Transport Corporation (Karnataka SRTC) was collected for a period of 2013 to 2018, which comprises of a total five-year data, to perform inter-city passenger demand modelling. The intra-city ETM data for Mangalore was not collected since the intra-city services are mainly operated by private operators. The private operators have not maintained and stored the ETM data. Therefore, Mangalore was chosen as an additional study city for the comparison of the demand models developed. The land use characteristics, travel characteristics, distribution of trip by the purpose of travel and public transport scenario are discussed. The performance indicators of Kerala SRTC and Karnataka SRTC are also mentioned.

3.2. DESCRIPTION OF THE STUDY CITY-TRIVANDRUM

Thiruvananthapuram, also known as Trivandrum, is the capital of Indian state Kerala and is an emerging metropolitan city in the southernmost part of India. Trivandrum district is situated between North latitudes $8^{\circ}17'$ and $8^{\circ}54'$ and East longitudes $76^{\circ}41'$ and $77^{\circ}17'$. The city has an area of 214.86 km^2 with a population of 10,32,292 as per census 2011, making it the most populous city in Kerala. The average population density of Trivandrum is 4454 persons/ km^2 (Thiruvananthapuram Corporation 2012). The details of census data are as given in Table 3.1.

Table 3.1. Census Data of Trivandrum City

Thiruvananthapuram city	Total	Male	Female
City + Out Growths	7,88,271	3,84,004	4,04,267
City Population	7,43,691	3,61,994	3,81,697
Literates	6,45,863	3,17,665	3,28,198
Children (0-6)	64,566	32,831	31,735
Average Literacy (%)	95.10%	96.51%	93.78%
Sex ratio (Female/Male)	1.054		
Child Sex ratio (Female/Male)	0.967		

Trivandrum city corporation area is divided into 100 wards which mark the study boundary. The map of Trivandrum city along with the ward names and numbering is shown in Figure 3.1. The Trivandrum city is served by three dominant modes of transport viz., road, rail and air. Of these, the road network has a wide presence throughout the city. Rail network has a limited but dominating influence with air network having a supplementary role. As per Trivandrum Master Plan 2012, the analysis of the road network shows that a well-defined hierarchy of roads is totally absent in the city. However, the city can be considered to have a partial ring and radial road network with NH-66 as a trunk road, as shown in Figure 3.2.

The city is home to a number of central and state government offices, private sector organisations, educational institutions, health care centres and research institutions. It is a major educational hub of Kerala with many academic and research institutions. Indian Institute of Space Science and Technology, Indian Institute of Science Education and Research and National Institute of Interdisciplinary Science and Technology are few among them. Also, the city is a major destination for tourism. It is a strategically important city in view of military logistics and civil aviation as it is the headquarters of Southern Air Command and a base for the Military. Furthermore, the Vikram Sarabhai Space Centre and Indian Space Research Organisation are also located in the outskirts of the city. The mega Deep Water Container Trans-shipment Port is being constructed at Vizhinjam, having international shipping lanes between Europe and also with major ports like Colombo, Kochi and Tuticorin in close proximity (Thiruvananthapuram Corporation 2012). Thiruvananthapuram has an immense potential to develop into an important city in South India.

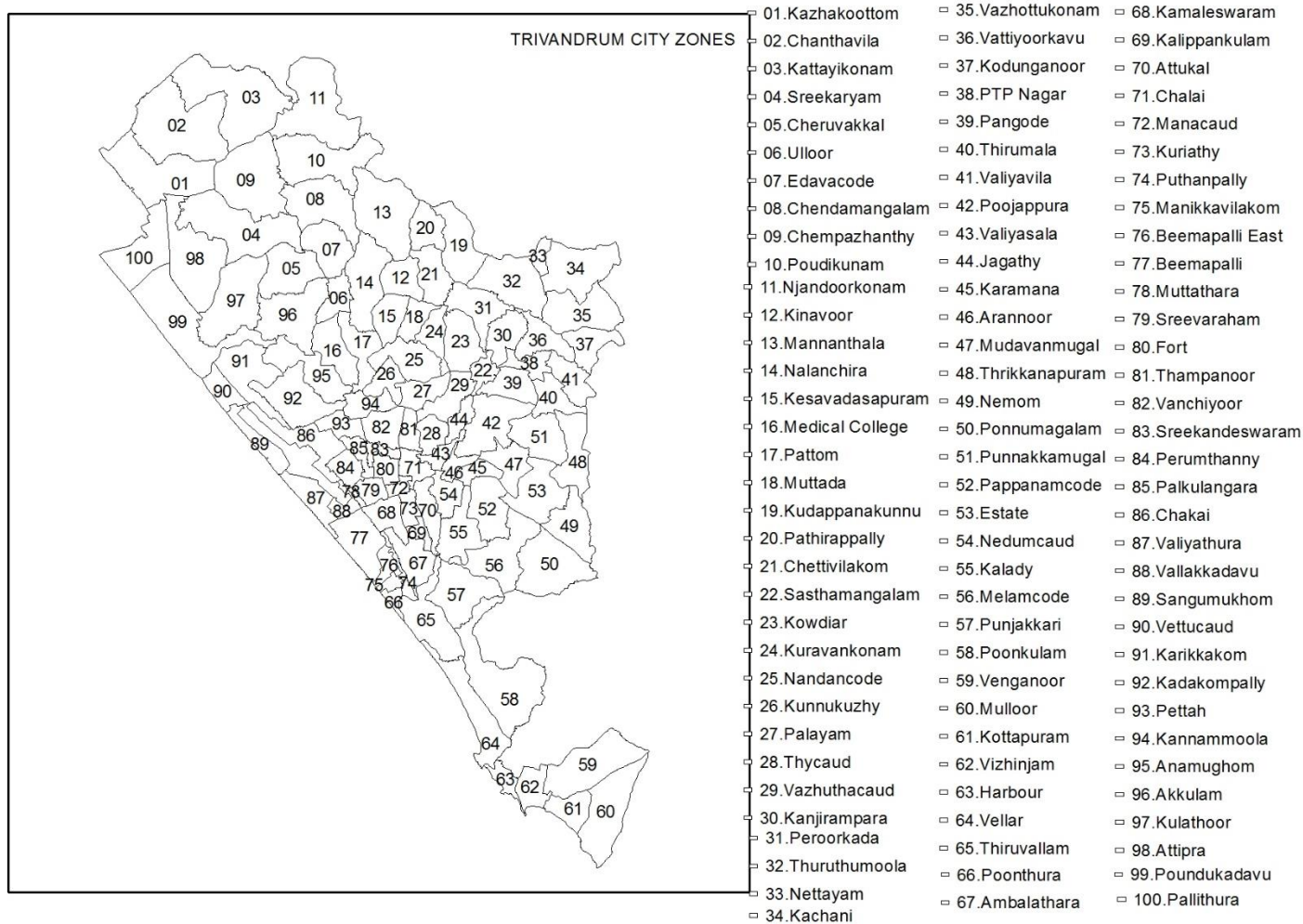


Figure 3.1. Map of Trivandrum City



Figure 3.2. Road Network Map of Trivandrum City

3.2.1. Trends in Land Use

Land use has a vital role in determining the urban conditions of any city. Land use analysis helps to understand the economic activities of an area and the direction in which development is oriented. The total area of the corporation was 74.86 km² (Thiruvananthapuram Corporation, 2012) in 1966, with the major share of land use for residential purpose (65%) followed by agriculture (11%). Being the capital city, 6% of the land was for public and semi-public use, and vacant land of 6% was also maintained within the corporation. The corporation had an area of 5% for transportation and 2% as restricted and water bodies.

The total area of the city grew up to 141.74 km². in 1990 with 59% residential, 10% agricultural, 9% public and semi-public land, followed by 8% for parks and open spaces. The city had roads and streets covering up to 7% of the total area. Studies have shown that the core area has more concentration of commercial activities with the periphery serving as the raw material centre.

In 2012, the total area of the city increased to 214.86 km². The land use for residential purpose is found to be 56%, followed by agricultural (22%). There is a considerable increase in the public and semi-public land use (13%), while the transportation area declined to 3%. Water bodies account for 3%, and the industrial and commercial area comprises of 1% each. Comparing the land use of various years, it can be concluded that there is a decrease in the residential area while the public and semi-public land use has increased. Furthermore, there is a considerable reduction in the area of the transport sector (Thiruvananthapuram Corporation, 2012).

3.2.2. Trends in Travel Behaviour

The city depends on all modes of transport-road, rail, water and air- for passenger and goods movement. The inter-city passenger transport is predominantly met by roads through Kerala State Road Transport Corporation (Kerala SRTC) bus, private bus, minibuses, car and two-wheelers. The mode share of each category is given in Table 3.2. and Figure 3.3. The distribution of trips by trip purpose is given in Table 3.3 and Figure 3.4.

Table 3.2. Distribution of Trips According to Mode of Travel in Trivandrum

Mode	% of trips	Mode	% of trips
Walk	5.49%	Mini-bus/ tempo	0.35%
Bi-cycle	1.29%	Car/ van/ jeep	9.36%
Bus- Kerala SRTC	32.67%	Two-wheeler	35.69%
Bus- private	0.71%	Taxi	0.14%
Bus - school	8.47%	Auto-rickshaw	4.10%
Bus- Institutional	1.29%	Train	0.43%
Source: NATPAC 2013		Total	100.00%

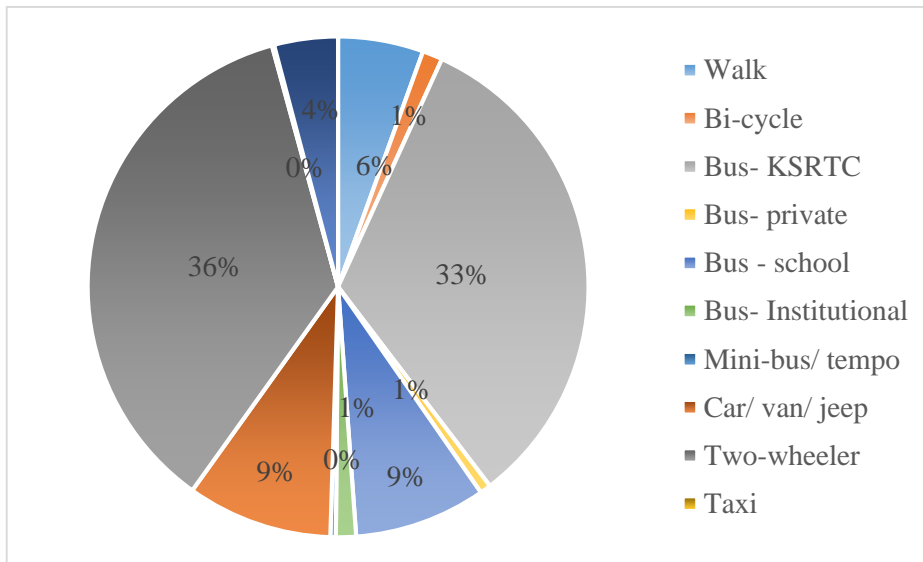


Figure 3.3. Distribution of Trips According to Mode of Travel in Trivandrum

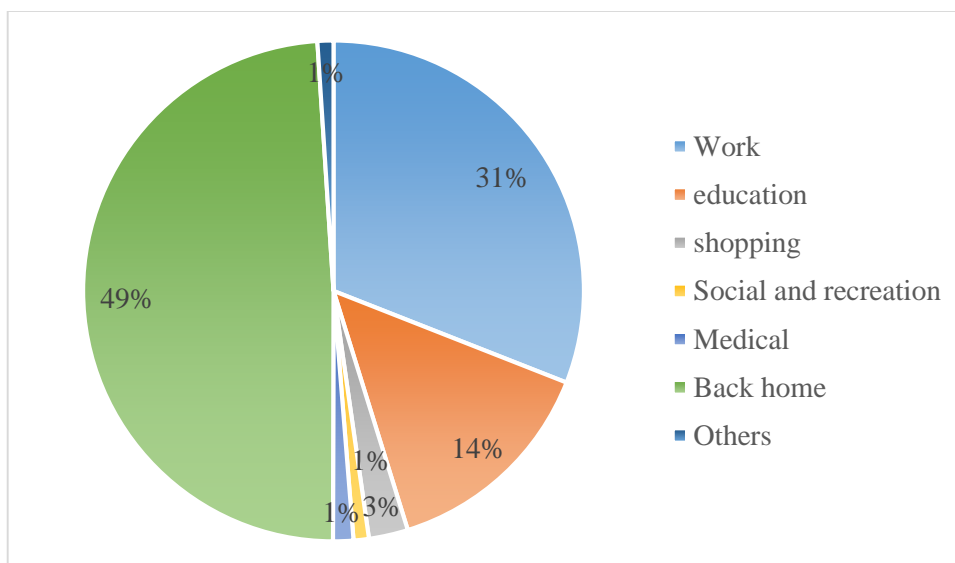


Figure 3.4. Distribution of Trips According to Trip Purpose in Trivandrum

Table 3.3. Distribution of Trips According to Trip Purpose in Trivandrum

Purpose	% of trips
Work	30.96%
Education	14.24%
Shopping	2.53%
Social and Recreation	0.99%
Medical	1.31%
Back home	48.93%
Others	1.03%
Total	100.00%
Source: NATPAC 2013	

The study by Ashalatha et al. (2012) has shown that low and middle-income people prefer public transport system. The preference for bus decreases as the income increases. Also, higher age groups and persons with two-wheelers preferred state-owned bus over the private vehicle. The persons owning both two-wheeler and car gave the least preference to public transport, and their common mode of travel is by car. Gender influences the mode choice, *i.e.* females prefer bus while males use two-wheelers.

3.2.3. Trends in Public Transport

The city bus is the most important mode of public transport in any city. In Thiruvananthapuram city, the public transport is dominated by state-owned Kerala State Road Transport Corporation (Kerala SRTC) buses. Auto rickshaws and taxi cabs are the major intermediate public transport (IPT). Private buses ply on few selected routes but are limited in number. The Kerala SRTC is the single largest public sector undertaking, which carries out passenger transport operations in the State. With 28 number of main depots and 45 number of sub-depots spread across the 14 districts of Kerala state, Kerala SRTC caters for the transport of passengers. Kerala SRTC owns a total of 6304 buses with the highest share of it being city ordinary buses, *i.e.* 3725. According to 2016 data, 1095 number of buses are of age 10 years or above. The number of passengers carried in the year 2015-2016 is 10437.8 lakhs.

The Kerala SRTC have 20 depots in Trivandrum district, out of which 6 are within the city limits to perform intra-city public bus transport operations. The Trivandrum central depot which operates the inter-city services is also located at the centre of the city. The depots which operate intra-city buses are Trivandrum city, Vikas Bhavan, Vizhinjam, Peroorkada, Pappanamcode and Kaniyapuram.

The major intra-city routes served by Kerala SRTC are:

- NH bypass between Kazhakkootam and Vizhinjam
- Medical College-Pattom
- Pattam-Kowdiar
- Mannanthala-Peroorkada-Vattiyoorkavu-Pangode-Karamana
- Overbridge-Pallimukku
- Pallimukku-Medical College-Ulloor

The longest intra-city bus route is Trivandrum-Attingal Route which is about 31 km. This route connects the most important bus stops such as East Fort, Pattom, Medical College and Kazhakkootam. This route covers major hospitals, institutions, and offices. Hence, it is the most important and busiest route in Trivandrum with only Kerala SRTC buses plying in this route. The shortest route covered by Kerala SRTC is East Fort to Peroorkada via Palayam and Kowdiar, covering around 16 km. The route plys along the major residential localities of Trivandrum city connecting the central area of Trivandrum.

The number of private buses and Kerala SRTC buses in Trivandrum during the year 2014 and 2015 is given in Table 3.4. The indicators of operational efficiency are given in Table 3.5.

Table 3.4. Number of Private Buses and Kerala SRTC Buses in Trivandrum District in the Year 2014 and 2015

Year	2014	2015
Private Buses	431	450
Kerala SRTC Buses	1461	1539

Source: Economic Review 2015

Table 3.5. Major Indicators Showing Operational Efficiency of Kerala SRTC

Sl. No	Items	2014-15	2015-16	Increase / Decrease during last year
1	Fleet Strength (as on March 31st)	5629	5682	(+) 53
2	Gross revenue earnings (Crore)	1923.82	2165.16	(+) 241.39
3	Gross Revenue expenditure (Crore)	2541.1	2778.3	(+)237.20
4	Gross operating loss (Crore)	617.33	613.14	(-) 4.19
5	Number of schedules operated as on 31st March	4602	4522	(-) 80
6	Average earning per vehicle on the road per day (INR)	10928	11191	(+) 263
7	Average earning per kilometre of buses operated (Paise)	3268	3309	(+) 41
8	Average earning per passenger (Paise)	1636	1741	(+) 105
9	Average route length (kms)	63.78	60.76	(-) 3.02
10	Average kilometre run per bus per day	329.16	332.54	(+) 3.38
11	Average number of buses held daily (Numbers)	5691	5636	(-) 55
12	Passengers Carried (Lakh)	11059.4	10437.8	(-) 621.65

Source: Kerala SRTC

3.3. DESCRIPTION OF THE STUDY CITY-MANGALORE

Mangalore city, also known as Mangaluru, is located in the southernmost part of Karnataka state of India. It is the main port city of India which is situated between the Western Ghats and the Arabian Sea. Mangalore is selected for inter-city passenger demand modelling. As per the Census of India 2011, Mangalore city has a population of 488968 with a population of 623841 in its urban area (Census 2011). The average population density of Mangalore city is 2876 persons/ km². The details of census data are given in Table 3.6.

Table 3.6. Census Data of Mangalore City

Mangalore city	Total	Male	Female
City + Out Growths	4,99,487	2,47,903	2,51,584
City Population	4,88,968	2,42,512	2,46,456
Literates	4,17,649	2,11,572	2,06,077
Children (0-6)	43,340	22,323	21,017
Average Literacy (%)	93.72%	96.09%	91.41%
Sex ratio (Female/Male)	1.016		
Child Sex ratio (Female/Male)	0.941		

The town is located at the 12⁰⁵²' North latitude and 74⁰⁵³' East longitude. The Mangalore corporation limit has an area of 170 km². The Mangalore corporation is divided into 60 wards. The map of Mangalore city is given in Figure 3.5. Mangalore city is well connected by all the three modes of transport-road, rail and air, in addition to being the port city. Three national highways, namely NH66, NH48 and NH13, pass through the city. Also, two state highways, namely SH 66A and SH 64A, also pass through the town.

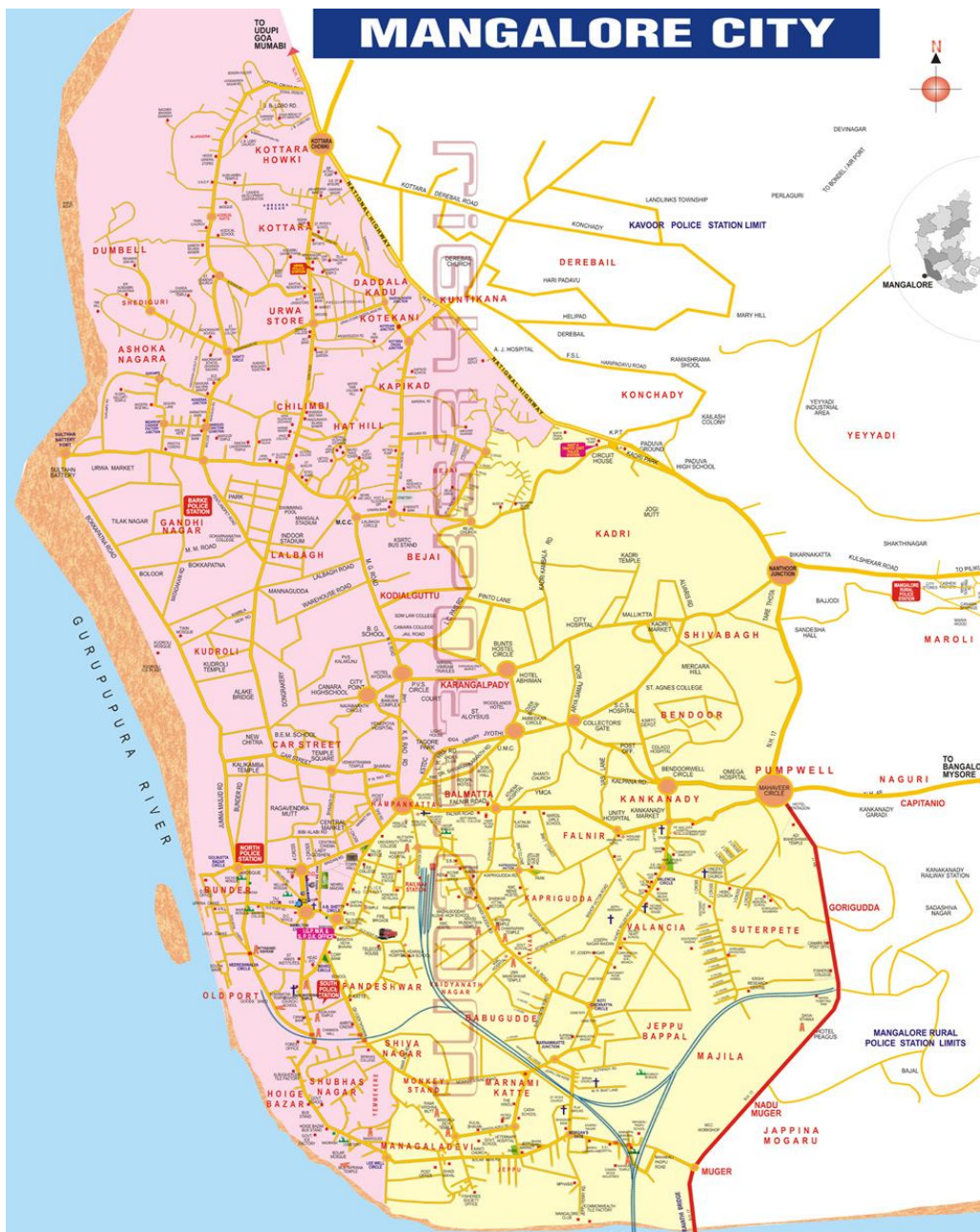


Figure 3.5. Map of Mangalore City

The major inter-city road services operated by Karnataka SRTC from Mangalore are as follows:

- Route Number 1: Mangalore-Bangalore
- Route Number 14: Mangalore-Birur
- Route Number 34: Mangalore-Kayarthadka
- Route Number 40: Mangalore-Uppinangady
- Route Number 112: Mangalore-Pajeer
- Route Number 175: State bank-Kallugudde
- Route Number 378: Mangalore-Subramanya
- Route Number 455: Mangalore-Muddebihala

The major inter-city road services operated by Karnataka SRTC from Udupi are as follows:

- Route Number 2: Udupi –Bagalkote
- Route Number 3: Udupi –Sirsi- Belagavi
- Route Number 4: Udupi-Karwar-Belagavi
- Route Number 5: Udupi-Lakshmeswar
- Route Number 6: Udupi-Kalaburagi
- Route Number 12: Manipal-Mysuru
- Route Number 18: Manipal-Mysuru-Bengaluru
- Route Number 28: Udupi-Indiranagar
- Route Number 87: CR Nagara-Mysuru

3.3.1. Trends in Land Use

Mangalore is one of the cosmopolitan non-metro cities of India. It is the largest coastal city in Karnataka. Also, it is an industrial city, commercial, educational and healthcare hub in the southern part of Karnataka. The growth of the city is towards the eastern part, and northern part since the south is bounded by the administrative boundary of Kerala while the western side is the Arabian Sea. The total area of Mangalore district in 1991 was 210 km² with the major share of land use for water bodies (36.7%) followed by a

green belt (30.04%). The residential area comprises of 16.27 %, and industrial area is 4.03%. Being the capital city, 3.01% of the land is for public and semi-public use. The playgrounds and open land is 0.97%. The district had an area of 6.83% for transportation, and other public utilities were 0.10%. The commercial area has a share of 1.61% of the total land use (MUDA 2011).

In 2001, the area of land under Mangalore district had increased to 306 km². The residential area increased to 19.06%. The land share for the commercial area is 1.15%, the industrial area is 3.56%, the transport area is 2.18%, and water bodies constitute of 6.81%. The public and semi-public area is about 1.49 %, while playgrounds and open spaces are 0.11%. The total developed area (agricultural land) was found to be 59.35% (MUDA 2011).

3.3.2. Trends in Travel Behaviour

The city depends on all modes of transport-road, rail, water and air- for passenger and goods movement. The inter-city passenger transport is predominantly met by roads through Karnataka SRTC bus, private bus, minibus, car and two-wheelers. The mode share of each category is given in Table 3.7. and Figure 3.6. The distribution of trips by trip purpose is given in Table 3.8 and Figure 3.7. It can be observed that buses are the most preferred mode of travel followed by two-wheelers. One of the interesting facts is that about 15% of the trip is made by walk. About 40% of the trip purpose is for work, followed by education (18%). Social and recreational purpose travel account for 13% of travel while 10% of people travel for business purpose. Shopping and religious purpose account for 9% each.

Table 3.7. Distribution of Trips According to Mode of Travel in Mangalore

Mode	% of trips
Walk	15%
Bi-cycle	1%
Bus	58%
Auto-rickshaw	2%
Two-wheeler	15%
Four-wheeler	9%
Total	100.00%
Source: DULT 2011	

Table 3.8. Distribution of Trips According to Trip Purpose in Mangalore

Purpose	% of trips
Work	41%
Education	18%
Shopping	9%
Social and Recreation	13%
Business	10%
Religious	9%
Total	100.00%
Source: DULT 2011	

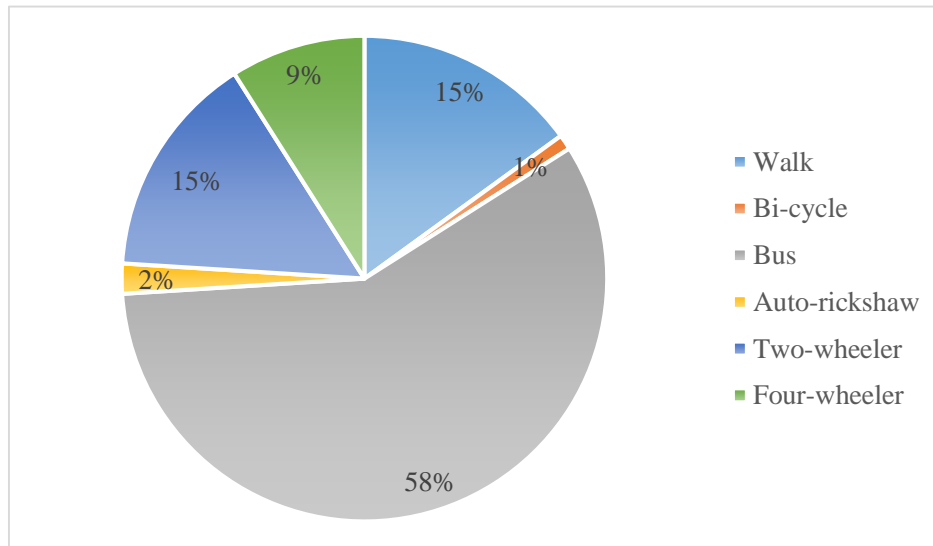


Figure 3.6. Distribution of Trips According to Mode of Travel in Mangalore

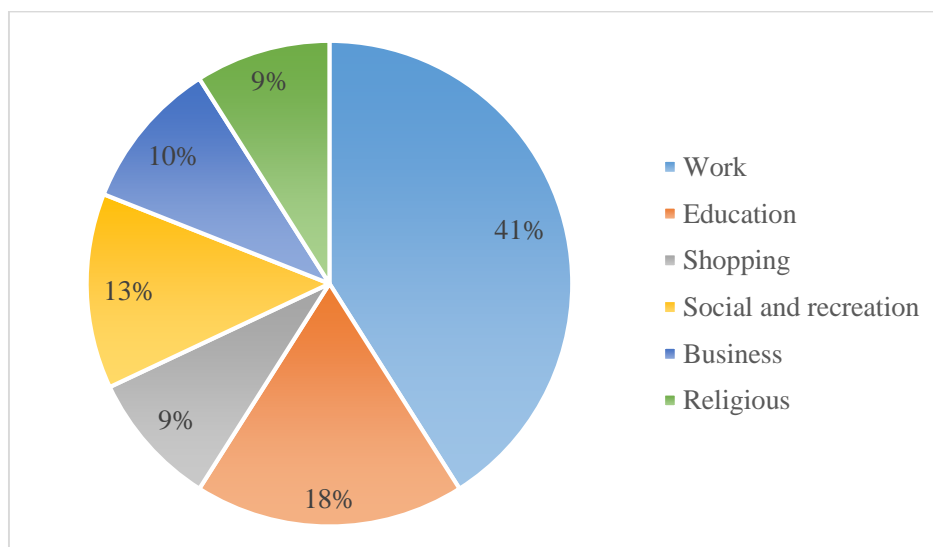


Figure 3.7. Distribution of Trips According to Trip Purpose in Mangalore

3.3.3. Trends in Public Transport

The city bus is the most important mode of public transport in any city. In Mangalore city, public transport is dominated by private buses, but for inter-city travel, Karnataka SRTC buses are in significant numbers. The Mangalore division Karnataka SRTC operates its buses from five depots namely, Mangalore1, Mangalore2, Mangalore3, Udipi and Kundapura. Table 3.9. gives the indicators of operational efficiency details of Mangalore division Karnataka SRTC buses. It can be observed that the fleet strength had been increased from 502 to 607 through the years. Also, it can be noted that the passengers carried has increased by 22% in the year 2017-18 compared to 2013-14. Even though the passengers have increased, the Karnataka SRTC of Mangalore division operates under loss.

Table 3.9. Major Indicators Showing Operational Efficiency of Karnataka SRTC for the Years 2012-13 to 2017-18

Sl. No.	Indicator	Performance from 2012-13 to 2017-18				
		2013-14	2014-15	2015-16	2016-17	2017-18
1	Fleet strength (Number)	502	496	523	545	607
2	Gross revenue earnings (Lakh)	20839.12	22743.20	22374.16	21740.63	23393.63
3	Gross revenue expenditure (Lakh)	22356.54	23470.41	22360.71	23861.94	25645.82
4	Gross operating loss (Lakh)	1517.42	727.21	13.45	2121.31	2252.19
5	Number of schedules operated	451	448	467	510	539
6	Average earnings per day (Lakhs)	54.34	59.14	57.64	55.97	59.30
7	Average earnings per kilometre of buses operated	2995.0	3295.9	3206.3	3062.2	2945.4
8	Passengers carried (Lakh)	366.38	348.62	328.30	378.49	448.64

3.4. SUMMARY

In this chapter, the description of the study areas- Trivandrum and Mangalore city are given. The land use characteristics, travel characteristics, distribution of trip by the purpose of travel and public transport scenario are discussed.

- Trivandrum city was selected for intra-city passenger demand modelling while Mangalore was selected for inter-city passenger demand modelling. The Trivandrum city has an area of 214.86 km² and Mangalore city has an area of 170 km².
- The land use characteristics of Trivandrum city suggests that more than 50% of land use is for residential purpose. In the case of Mangalore district, agricultural land accounts for a major share of land use followed by the residential use.
- About 40% of travel is by bus in Trivandrum, while the Mangalore has 60% of travel by bus. In both the cities, the purpose of travel is mainly for work.
- The performance indicators of Kerala SRTC and Karnataka SRTC are discussed in this chapter. It can be observed that in Trivandrum, the demand for public transport is decreasing (-6%) while in Mangalore it is increasing (18%). The public transport (Kerala SRTC and Karnataka SRTC) operates in a loss in both the cities.

CHAPTER-4

METHODOLOGY AND DATA COLLECTION

4.1. GENERAL

This chapter outlines the methodology and data collection to achieve the objectives of this research. The procedure adopted for data collection and data processing is also dealt in detail in this chapter. The data used for the study are mainly ETM data from Kerala SRTC for a period 2011 to 2013. In addition to ETM data provided by Kerala SRTC, this study also uses data provided by various state and central government agencies. The methodology to determine the level of service offered by the public transport system in Trivandrum, the accessibility of public transport and the methodology for performance evaluation of Kerala SRTC to evaluate the present condition of the public transport system in Trivandrum and Kerala SRTC are discussed in detail. Further, a methodology to develop intra-city and inter-city passenger demand estimation model using the data-driven time series approach, employing the capabilities of ARIMA method is described. The methodology to improve the developed model using the nonlinear ARCH and ARNN methods are explained. During the analysis of the Kerala ETM data, it was found that the length of data available for individual routes for intra-city varies from 1.5 years to 2 years, while inter-city data were not sufficient for time-series modelling. Therefore, inter-city ETM data of Mangalore division Karnataka SRTC was collected for a period of 2013 to 2018, which comprises of a total five-year data, to perform inter-city passenger demand modelling. The intra-city ETM data for Mangalore was not collected, since the intra-city services are mainly operated by private operators. The private operators have not maintained and stored the ETM data.

4.2. METHODOLOGY

The public transport planning comprises of mainly four activities: (1) network/ route design and planning, (2) headway determination and timetable development (3) vehicle scheduling and (4) crew scheduling (Ceder 2016). The general process involved in public transport planning is given in Figure 4.1.

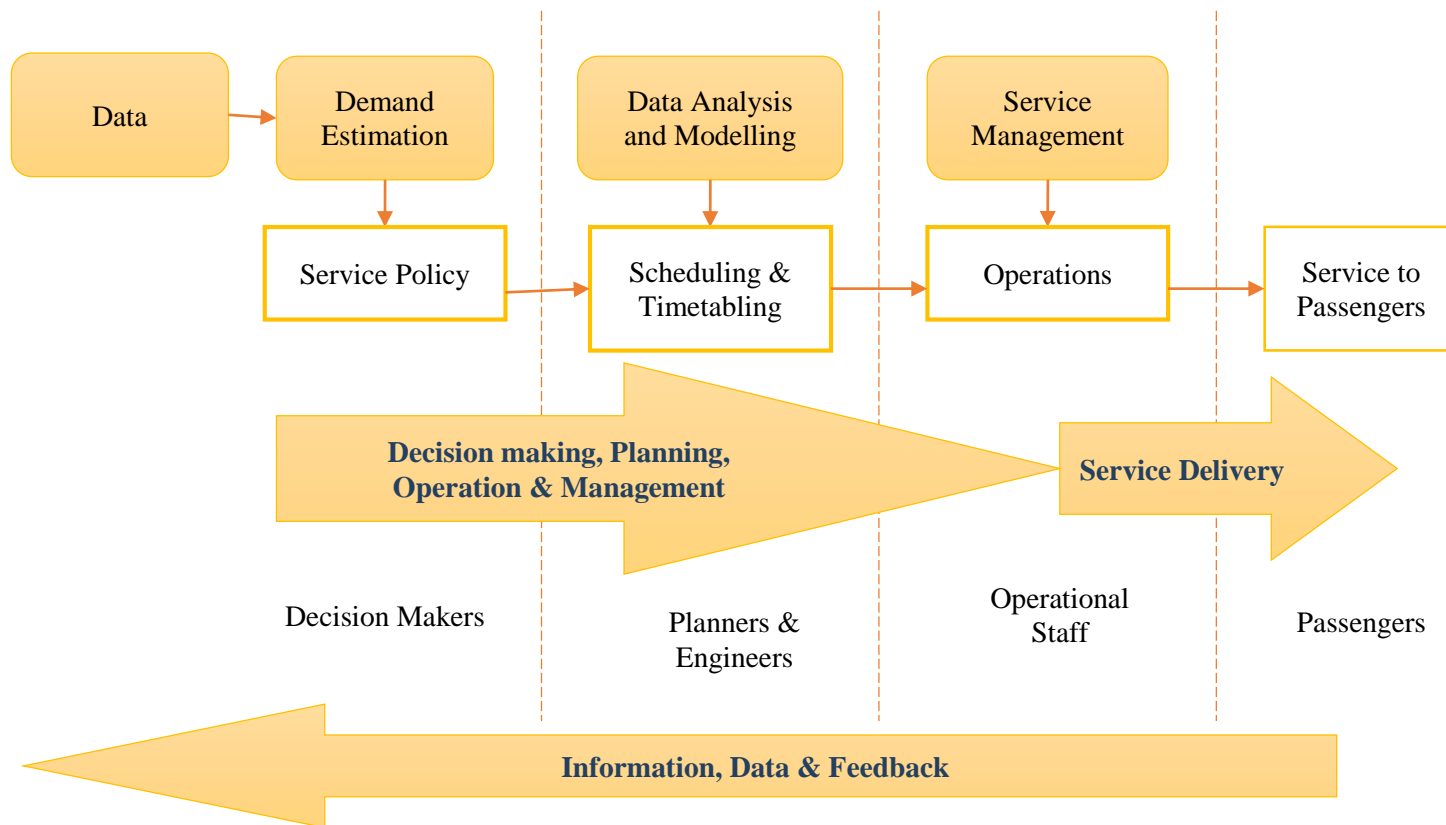


Figure 4.1. General Process of Transit Planning (Source: Adapted and Modified from Carrel 2009)

The demand estimation process triggers the planning process. The actual or expected demand is the basis of service policy decision making such as, whether a service is required or not, the route on which the service has to be operated, etc. Therefore, the first step in the public transport planning process is the estimation of passenger demand and identification of the travel pattern of the passengers. The passenger demand is further analysed during the data analysis stage of the planning process, and then models are developed for future demand estimation. This future demand will be further used for developing vehicle schedules and timetable. The planners and engineers use the demand data for developing models for route design, vehicle scheduling, crew scheduling, headway determination and timetable development. The developed time tables and schedules will be implemented on the public transport network. Once the bus service is delivered to the passengers, information and feedback is generated and archived which is used for rescheduling (if necessary) and for day to day operational planning. Thus, the public transport planning process is continuous in nature, which is triggered by the demand data and is to be kept moving using the newly generated data set. If no further monitoring and analysis of the public transport scenario is done, the public transport service will deteriorate, leading to mode shift of passengers. Therefore, there is a need for frequent updating and analysis of demand scenario for day to day operational planning.

Therefore, based on the literature reviewed, the following methodology has been adopted for the development of ETM based public bus transport demand model. The flow chart given in Figure 4.2 presents a brief idea of the procedure involved in the process.

4.3. DATA COLLECTION

This study investigates various aspects related to demand modelling based on data elicited from the Electronic Ticket Machine (ETM). The ETM data has not been explored thoroughly for transportation planning, although it is now-a-days collected and compiled by public transport undertakings on a regular basis. The data used in the study is part of transactions on ticket sales by Kerala State Road Transport Corporation

(Kerala SRTC) maintained at 6 inter-zonal bus depots in Trivandrum city. The Kerala SRTC operates its services for Trivandrum city from mainly 6 depots namely Trivandrum City Depot, Vikas Bhavan, Peroorkada, Pappanamcode, Vizhinjam and Kaniyapuram. The ticketing details of all the buses and routes were collected from the respective depots for the period between 2011 and 2013 (Cyril et al. 2018). The number of buses, bus stops, timetables and schedules were also provided by Kerala SRTC. Also, the physical and financial performance data was obtained from Kerala SRTC in the form of Annual Audit Reports. Economic Review published by State Planning Board, Kerala was also collected for getting the financial performance data of Kerala SRTC.

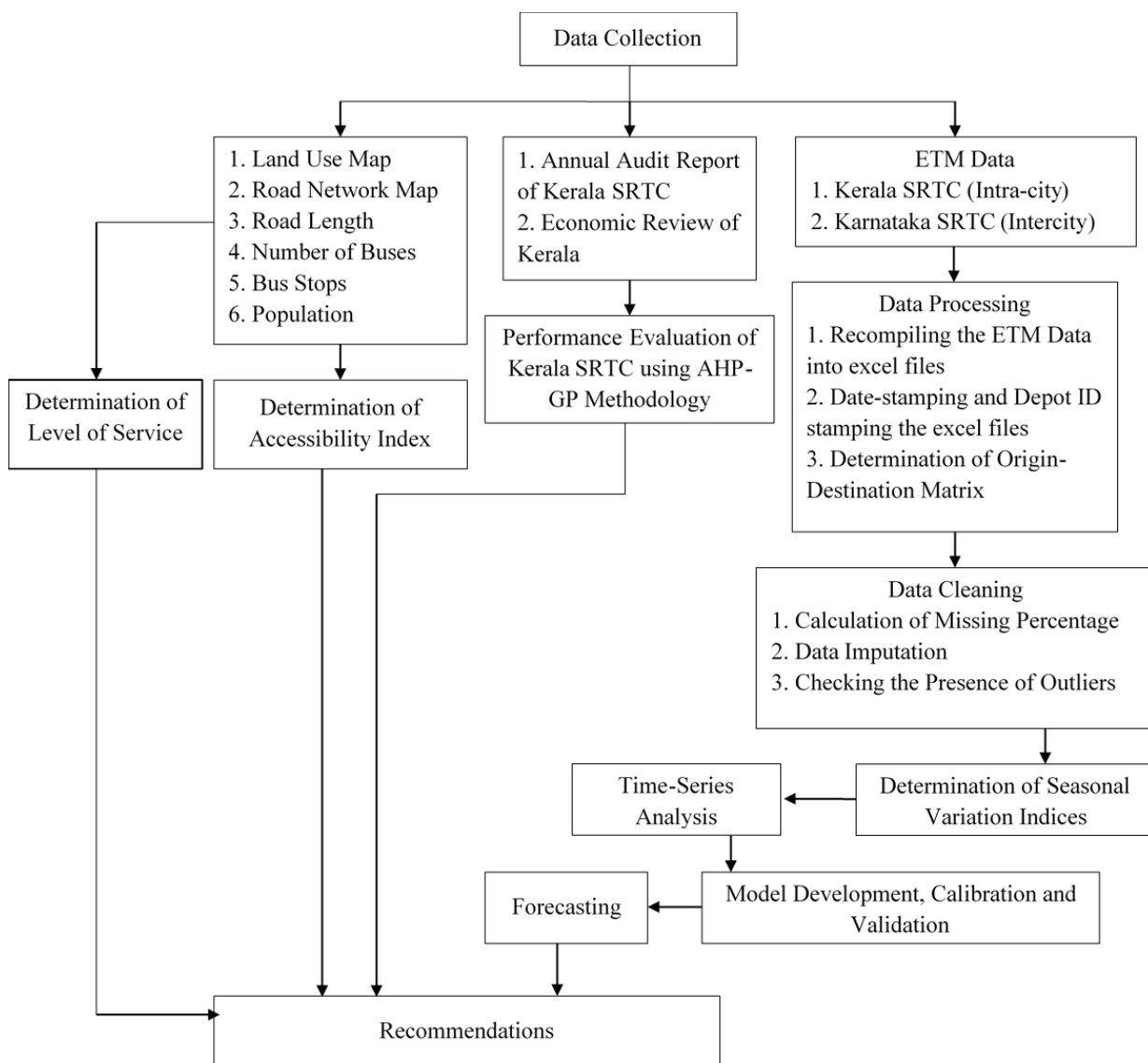


Figure 4.2. Flowchart of Methodology

The level of service offered by the public transport system in Trivandrum city, the accessibility of public transport and the performance evaluation of Kerala SRTC was performed to evaluate the present condition of the public transport system in Trivandrum and Kerala SRTC. Therefore, in addition to ETM data provided by Kerala SRTC, this study also uses secondary data provided by National Transportation Planning and Research Centre (NATPAC), Department of Town and Country Planning and Directorate of Census Operations. Demographic and socioeconomic data for the Trivandrum city, which includes population, number of households and population density are obtained from Census 2011 data of Trivandrum city and household survey data collected by NATPAC, Trivandrum. The road network map and land use map was collected from the Department of Town and Country Planning, Kerala.

During the analysis of the Kerala ETM data, it was found that the length of data available for individual routes for intra-city varies from 1.5 years to two years, while inter-city data were not sufficient for time-series modelling. Therefore, inter-city ETM data of Mangalore division Karnataka SRTC was collected for a period of 2013 to 2018, which comprises of a total five-year data, to perform inter-city passenger demand modelling. The intra-city ETM data for Mangalore was not collected since the intra-city services are mainly operated by private operators. The private operators have not maintained and stored the ETM data.

4.3.1. Land Use Data

The land use and road network data were obtained from the Department of Town and Country Planning, Trivandrum. Land use analysis helps to understand the economic activities of an area and the direction in which development is oriented. In 2012, the total area of the Trivandrum city was 214.86 km². The land use for residential purpose is found to be 56%, followed by agricultural (22%). The public and semi-public land use is 13% while the transportation area is found to be 3%. Water bodies account for 3%, and the industrial and commercial area comprises of 1% each. Figure 4.3. shows the existing land use. The scale of collected land use map is 1:35000.

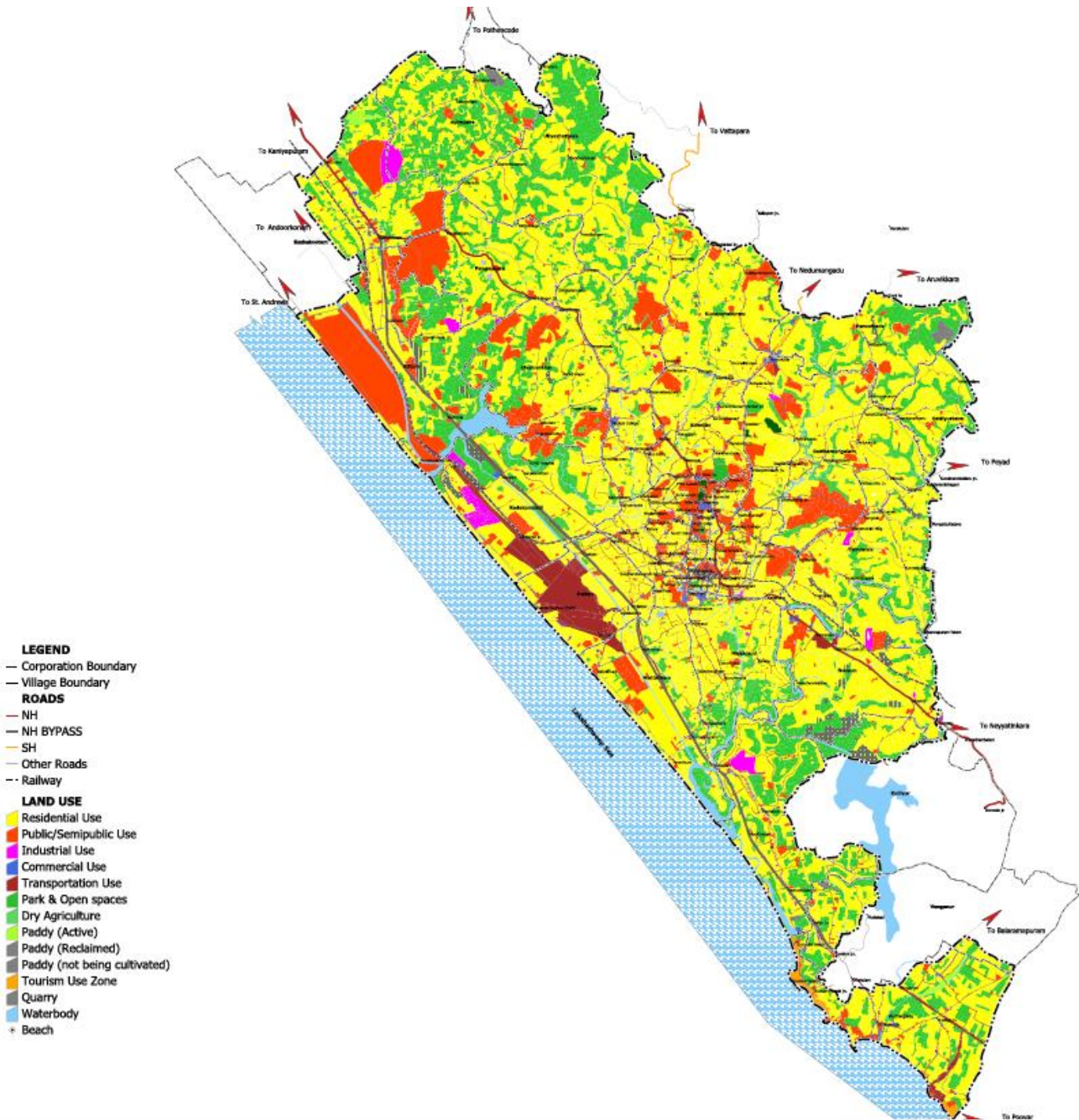


Figure 4.3. Existing Land Use Map of Trivandrum City

4.3.2. Road Network Data

The road network data was obtained from NATPAC (2013) and Department of Town and Country Planning, Trivandrum. The road network map was also collected and is shown in Figure 4.4. It can be seen that the road network has a wide presence throughout the city. The analysis of the road network shows that a well-defined hierarchy of roads is totally absent in the city. However, the city can be considered to have a partial ring

and radial road network with NH-66 as a trunk road. The details of the road network were collected from NATPAC. The total length of road in the Trivandrum Municipal Corporation region is 2586 km (CMP 2014).

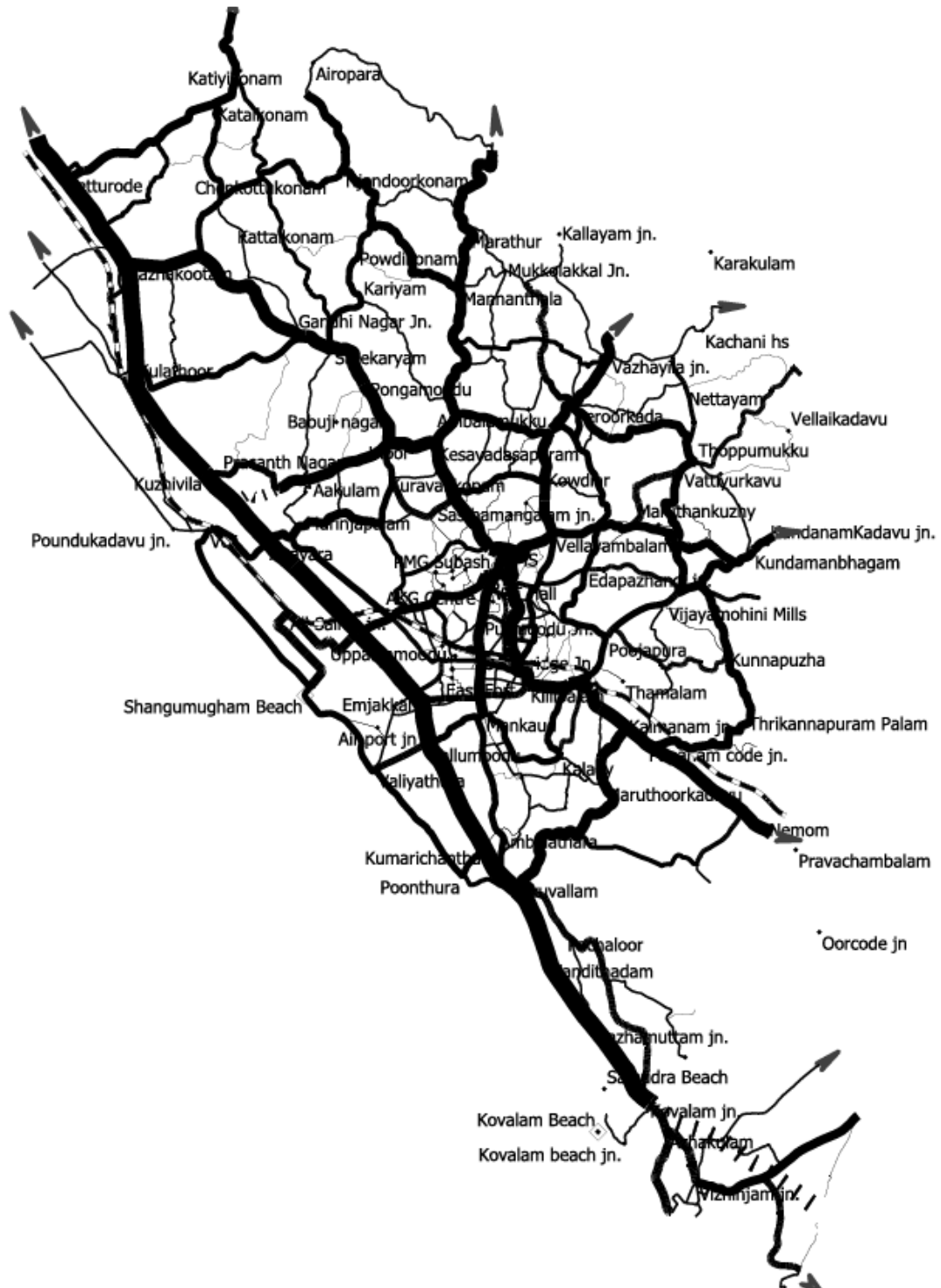


Figure 4.4. Road Network Map of Trivandrum City

The major radial arterial roads which serve as the inter-city link roads are:

- National Highway 66- It passes through the heart of the city and covers important traffic generating points like Pappanamcode, Karamana, Kilippalam, MG Road, Pattam, Kesavadasapuram, Sreekaryam and Kazhakkootam.
- Main Central Road (MC Road): Originating from Kesavadasapuram, passing through Nalanchira and Mannanthala and exiting on North towards Kottayam.
- Trivandrum-Shencottah Road (TS Road): Originating from PMG junction on NH 66, passing through Peroorkada and Nedumangad and exiting towards Shencottah in Tamil Nadu.
- Thiruvananthapuram-Kattakada Road: Originating from Thampanoor Junction on NH 66 moving towards east via Thirumala and Malayinkeezh.
- Thiruvananthapuram- Thiruvallam-Poovar Road: Originating from East Fort

The intra-city link roads which serve as partial ring roads are:

- NH bypass between Kazhakkootam and Kovalam
- Palayam-Airport Road
- Medical College-Pattom Road via Pottakuzhy
- Pattam-Kowdiar Road via Kuravankonam
- Mannanthala-Peroorkada-Vattiyoorkavu-Pangode-Karamana Road.
- Overbridge-Pallimukku Road
- Pallimukku-Medical College-Ulloor Road

4.3.3. Bus Stop Data

The data pertaining to buses and bus stops are collected from Kerala SRTC. The number of Kerala SRTC buses operating in Trivandrum is 446. The length of road on which the Kerala SRTC buses ply is 651.35 km. Since, the data (ETM data) for the study is fare stage wise, 192 fare stages within the city limits were selected as the bus stops. These fare stage bus stops were obtained from Kerala SRTC. The selected bus stop location is given in Figure 4.5.

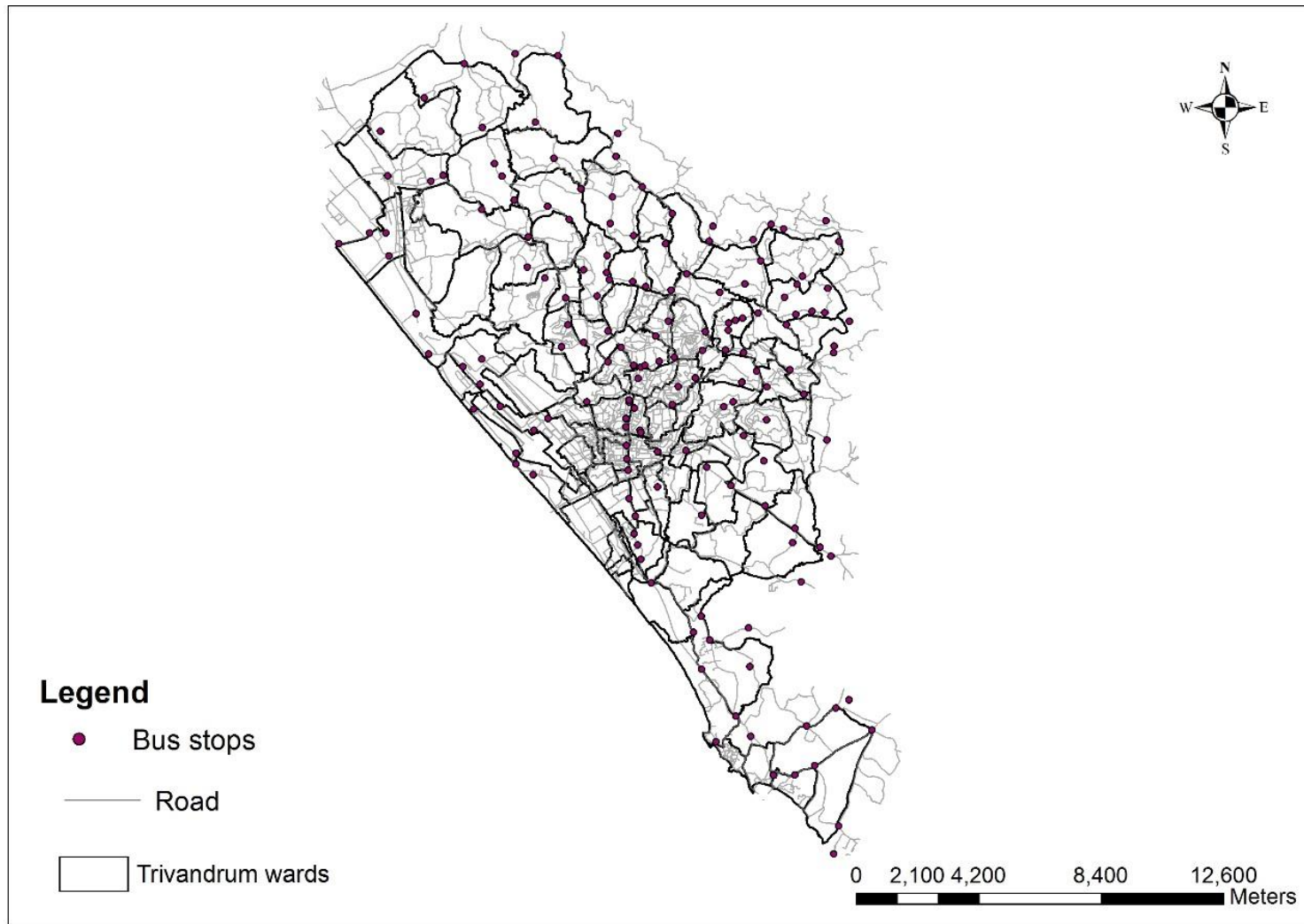


Figure 4.5. Map Showing the Bus Stop Locations

The important bus stops of Trivandrum city are as follows:

- Thampanoor- It is the major transit point for inter-city and intra-city travel. It is located near the railway station.
- East Fort- It is the bus stop near the major shopping centres. Also, the famous religious centre- Sri Padmanabaswami Temple is accessible by walk from East Fort.
- Pattom- It is located near to many government offices, commercial buildings, shopping centres and schools. Also, buses to both inter-city and intra-city travel are available from Pattom.
- Palayam- It is located inside the Central business district (CBD) area.
- Medical College- As the name suggests, this bus stop is near to Trivandrum Medical College Hospital, regional cancer centre, and various other government and private hospitals. Also, many schools and commercial establishments can be reached by walk from this bus stop.
- Kazhakkuttam- It is located near the major IT hub of Kerala.

4.3.4. Demographic Data

The population data for the year 2011 was collected from Directorate of Census Operations, Trivandrum. The population data for the year 2013 was obtained by the extrapolation of 2011 data using a growth factor of 2.1%. The wardwise population of details of Trivandrum is given in Appendix A.1. The city has 100 wards of which, the ward Beemapally East is the smallest with a population of 8315 people, while the ward Medical College is the biggest with a population of 10528. Moreover, the Beemapally ward has low-income population who mainly rely on fishing for their livelihood. On the other hand, the residents of Medical College ward are mainly government officials and doctors.

4.3.5. Physical and Financial Data of Kerala SRTC

The physical and financial data pertaining to the performance of Kerala SRTC was collected as audit reports provided by Kerala SRTC and Economic review report from the State Planning Board. The Kerala SRTC is the single largest public sector

undertaking, which carries out passenger transport operations in the State. With 28 number of main depots and 45 number of sub-depots spread across the 14 districts of Kerala state, Kerala SRTC caters for the transport of passengers. Kerala SRTC owns a total of 6304 buses with the highest share of it being city ordinary buses, *i.e.* 3725. According to 2016 data, 1095 number of buses are of age 10 years or above. The number of passengers carried in the year 2015-2016 is 10437.8 lakhs. The number of buses and speed and fare characteristics is given in Table 4.1. and Table 4.2. respectively.

Table 4.1. District Wise Number of Private and Kerala SRTC Buses 2014 and 2015

District	Private Buses		Kerala SRTC buses	
	2014	2015	2014	2015
Trivandrum	431	450	1461	1539
Kollam	928	1022	631	615
Pathanamthitta	945	1042	308	313
Alappuzha	955	1053	467	431
Kottayam	1627	1793	445	458
Idukki	415	457	230	235
Ernakulam	2535	2793	567	546
Trichur	1956	2156	386	365
Palakkad	1170	1290	208	208
Malappuram	2118	2335	213	215
Kozhikode	1816	2002	240	237
Wayanad	484	533	270	252
Kannur	1944	2142	269	270
Kasargode	478	527	153	154
	17802	19145	5848 + 320 = 6168	5660 + 424 = 6084

Source: Economic Review 2015

In Table 4.1., it can be seen that every district other than Trivandrum has private buses account for major share of buses. In Trivandrum, Kerala SRTC dominates public transport travel. Another important trend to be noted is that there is a decline in the number of Kerala SRTC buses (by 1.4%) in the year 2015 compared to 2014, while the number of private buses has increased by 7.5%. Even though the overall number of

Kerala SRTC buses has decreased in the Kerala state, the number of Kerala SRTC buses in Trivandrum has increased by 5%.

Table 4.2. Speed and Fare Characteristics of Each Type of Bus

Type of bus	Speed (min/km)	Speed (km/hr)	Minimum Fare
Volvo	1	60.0	40
Super Deluxe	1.1	54.5	28
Super Express	1.3	46.2	20
Superfast Passenger	1.5	40.0	13
Limited Stop Fast Passenger	1.6	37.5	10
Fast Passenger	1.7	35.3	10
Ordinary	2.5	24.0	6
Town to Town Ordinary	1.8	33.3	6
Limited Stop Ordinary	2	30.0	6
City Ordinary	3	20.0	6
City Fast	2.5	24.0	6

Source: Kerala SRTC

From Table 4.2, it is observed that the Volvo buses are the fastest among all the buses as they are used for inter-city and inter-state travel. The slowest is the city ordinary buses with a speed of 20kmph as it runs within the city limits.

4.3.6. Electronic Ticket Machine Data

This study investigates various aspects related to demand modelling based on data elicited from the Electronic Ticket Machine (ETM). The data used in the study is part of transactions on ticket sales by Kerala SRTC maintained at 6 interzonal bus depots in Trivandrum city. The Kerala SRTC operates its services for Trivandrum city from mainly 6 depots namely Trivandrum City Depot, Vikas Bhavan, Peroorkada, Pappanamcode, Vizhinjam and Kaniyapuram (Cyril et al. 2018). In the data collection stage, the intra-city and inter-city ETM data for 10 years were requested from Kerala SRTC. Due to unavailability of 10-year ETM data, the ticketing details of all the buses and routes of intra-city were collected from the respective depots for the period between 2011 and 2013. The database contains trip tables along with the details of fare stages

(Cyril et al., 2018). During the analysis of the Kerala ETM data, it was found that the length of intra-city ETM data (for Trivandrum city) available for individual routes varies from 1.5 years to 2 years (2011 to 2013), while inter-city data were not sufficient for time-series modelling. Therefore, inter-city ETM data of Mangalore division Karnataka SRTC was collected for a period of 2013 to 2018, which comprises of a total five-year data, to perform inter-city passenger demand modelling. The intra-city ETM data for Mangalore was not collected since the intra-city services are mainly operated by private operators. The private operators have not maintained and stored the ETM data. The ETM data must be processed to facilitate it to use for further analysis.

4.4. DATA PROCESSING

The data collected from the Kerala SRTC and Karnataka SRTC have to be processed for further analysis.

4.4.1. Data from Kerala SRTC (Intra-City Data)

The data used in the study is a set of transactions in the Kerala SRTC six city intra-city depots for the year 2011-2013. The database contains trip tables along with the details of fare stages. A single ETM data text file represents all the trips made by a single bus in a day (Cyril et al., 2018). One such ETM text file is given in Appendix A.2. This data has to be processed to get single files for each fare stages (route) for all depots. These files are then used for the OD matrix generation.

```
(0281100001 01/01/11 280235 00 . 22580908
00000000 000190150

01 105 00549687 01 02 01 00 00 00 04:18 0000400 T
01 105 00549688 02 05 01 00 00 00 04:22 0000500 T
01 105 00549689 03 05 01 00 00 00 04:28 0000450 T
01 105 00549690 03 05 01 00 00 00 04:30 0000450 T
01 105 00549691 03 05 01 00 00 00 04:32 0000450 T
```

Figure 4.6. The Raw ETM Data

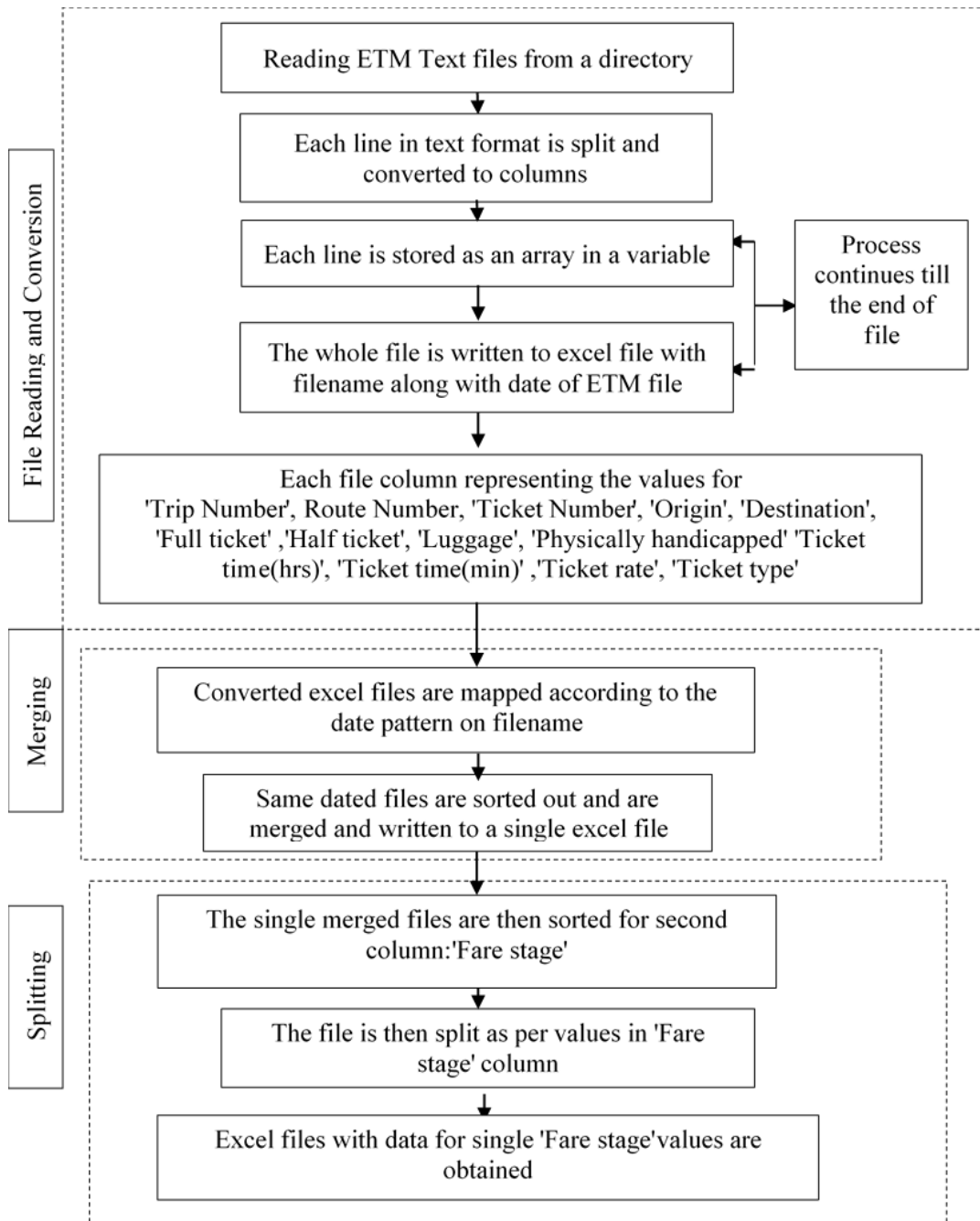


Figure 4.7. Flow Chart for Data Extraction

1. Recompiling ETM Data into Excel Files: The intra-city ETM data collected from Kerala SRTC was in the form of text file, as shown in Figure 4.6. A single ETM data text file represents all the trips made by a single bus in a day. The first two rows in the text file represent waybill number, waybill date and details related to the depot. From the third row gives the fare transaction details. Each column in the text file corresponds to trip number, route number or fare stage number, ticket number, origin, destination,

full ticket, half ticket, luggage, physically handicapped, ticket time, ticket rate and ticket type respectively. The raw ETM data is in text format and is converted into Microsoft Excel format. The converted Excel files are then merged based on the date. The merged file is split into different sheets on the basis of route code which gives the data according to routes. The procedure for data extraction is shown in Figure 4.7.

The text files are read from the file directory. The data values in each line of the file shown in Figure 4.6. is recompiled and entered into the relevant columns. Each column in the text file corresponds to trip number, route number or fare stage number, ticket number, origin, destination, full ticket, half ticket, luggage, physically handicapped, ticket time, ticket rate and ticket type respectively. The time column of the text file is split into time in hours and time in minutes when converted to Excel file. The data in each file is stored as a separate array. The file name for each array will include details on ticket collection. The array is then stored as an excel file, as shown in Figure 4.8. The name of the excel file is such that it includes the name of text file along with the date of ticket collection. The MATLAB code used to read the excel file from the directory and then convert it into Excel file is given in Appendix A.3.

Then, the above converted excel files will be recompiled based on the trip date. The Excel files having the same date of ticket collection is merged together to get a single Excel file for a single day. The MATLAB code used to read the Excel file from the directory and merge the Excel files having the same file name date is given in Appendix A.4. Using these recompiled files, data for each route will be filtered, and then these Excel files are further split based on the route number. Thus a single Excel file is the ETM data recordings for a day for a single route. The MATLAB code used to read the excel file from the directory and then split the date wise aggregated Excel files based on the route number is given in Appendix A.5. Alternatively, a Microsoft Excel Macro was also used for splitting the Excel files based on the route number. The Excel Macro used to split the files based on route number is given in Appendix A.6. Finally, the Excel files having the same route number and same date is again merged to get a single Excel file for a single route for a single date. The MATLAB code used for merging the files based on the route number and date is given in Appendix A.7. Thus, the final recompiled Excel file given in Appendix A.8. is route-wise ETM data for each day.

Trip Number	Route Number	Ticket Number	Origin	Destination	Full Ticket	Half Ticket	Luggage	Physically Challenged	Time (hrs)	Time (min)	Fare	Ticket Type
1	105	549687	1	2	1	0	0	0	4	18	400	T
1	105	549688	2	5	1	0	0	0	4	22	500	T
1	105	549689	3	5	1	0	0	0	4	28	450	T
1	105	549690	3	5	1	0	0	0	4	30	450	T
1	105	549691	3	5	1	0	0	0	4	32	450	T
1	105	549692	3	5	1	0	0	0	4	33	450	T
1	105	549693	3	5	2	0	0	0	4	34	900	T
1	105	549694	3	5	1	0	0	0	4	35	450	T
1	105	549695	4	5	1	0	0	0	4	38	400	T
2	105	549696	1	3	1	0	0	0	5	0	450	T
2	105	549697	1	2	1	0	0	0	5	1	400	T
2	105	549698	1	2	1	0	0	0	5	1	400	T
2	105	549699	1	2	1	0	0	0	5	1	400	T
2	105	549700	2	5	1	0	0	0	5	6	500	T
2	105	549701	2	5	1	0	0	0	5	7	500	T
2	105	549702	3	5	1	0	0	0	5	13	450	T
2	105	549703	3	5	1	0	0	0	5	14	450	T
2	105	549704	3	5	1	0	0	0	5	14	450	T
2	105	549705	3	5	2	0	0	0	5	17	900	T
2	105	549706	3	5	2	0	0	0	5	17	900	T
2	105	549707	3	5	1	0	0	0	5	19	450	T

Figure 4.8. Excel Converted ETM Data

2. Date-Stamping the Excel Files: The recompiled Excel files for each date from the previous data processing step is to be date-stamped. The procedure of date-stamping the Excel files is given in Figure 4.9.

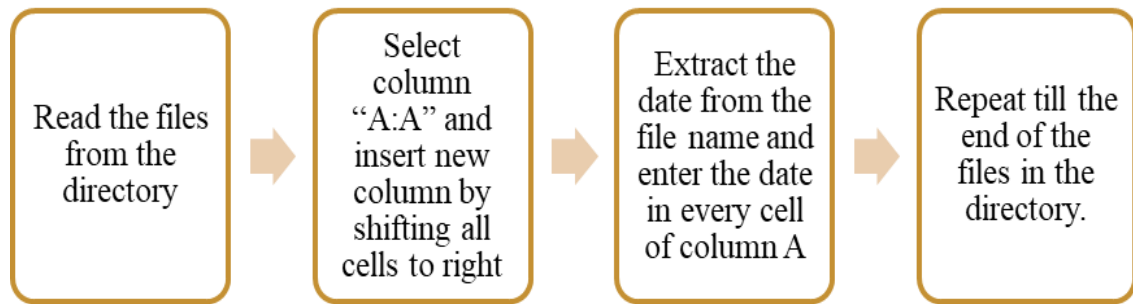


Figure 4.9. The Procedure for Date-Stamping the Excel Files

Date-stamped Excel files are merged month-wise, and then the depot ID is entered on every sheet. The procedure for depot ID stamping is the same as that of date-stamping. The Excel Macro used for date-stamping and depot ID stamping is given in Appendix A.9 and A.10, respectively. The final Excel sheet is as shown in Figure 4.10.

3. Renaming the Final Recompiled ETM Data File Based on Depot ID: In further data analysis and processing, it is favourable to have the file name of the recompiled Excel data based on the depot ID and date. The file name of the recompiled ETM data now has only the date of ticket collection (For example, the recompiled ETM data has the file name 01-01-11_combined). A Microsoft Excel Macro given in Appendix A.11 is used to rename the recompiled excel file. The macro fetches the old file name of all the files in a directory to an Excel file and lists the file name in column A of the Excel file. Using the function concatenate, depot ID is inserted in the file name of column A and list it in the second column. And the program renames all the files in the folder based on the new file name listed in column B. For example, the old file name for recompiled ETM data of 1st January 2011 for Trivandrum City depot (Depot ID=28) is 01-01-11_combined. The new file name will be 028-01-01-11_combined.

Depot_ID	Date	Trip Number	Route Number	Ticket Number	Origin	Destination	Full Ticket	Half Ticket	Luggage	Physically Challenged	Time (hrs)	Time (min)	Fare	Ticket Type
28	01-01-2012	1	105	179048	2	1	21	0	0	0	3	9	10500	T
28	01-01-2012	1	105	179049	2	1	2	0	0	0	3	12	1000	T
28	01-01-2012	1	105	179050	2	1	4	0	0	0	3	12	2000	T
28	01-01-2012	1	105	179051	2	1	1	0	0	0	3	13	500	T
28	01-01-2012	1	105	179052	2	1	1	0	0	0	3	13	500	T
28	01-01-2012	1	105	179053	2	1	1	0	0	0	3	13	500	T
28	01-01-2012	2	105	179054	1	2	4	0	0	0	5	20	2000	T
28	01-01-2012	2	105	179055	1	2	1	0	0	0	5	20	500	T
28	01-01-2012	3	105	179056	2	1	2	0	0	0	7	46	1000	T
28	01-01-2012	3	105	179057	2	1	2	0	0	0	7	47	1000	T
28	01-01-2012	3	105	179058	2	1	1	0	0	0	7	47	500	T
28	01-01-2012	3	105	179059	2	1	7	0	0	0	7	47	3500	T
28	01-01-2012	4	105	179060	1	3	1	0	0	0	8	57	500	T
28	01-01-2012	4	105	179061	1	3	2	0	0	0	8	58	1000	T
28	01-01-2012	4	105	179062	1	3	3	0	0	0	8	58	1500	T
28	01-01-2012	4	105	179063	1	3	1	0	0	0	9	0	500	T
28	01-01-2012	5	105	179064	3	1	3	0	0	0	9	20	1500	T
28	01-01-2012	5	105	179065	3	1	33	0	0	0	9	21	16500	T
28	01-01-2012	5	105	179066	3	1	4	0	0	0	9	22	2000	T
28	01-01-2012	5	105	179067	3	1	1	0	0	0	9	23	500	T

Figure 4.10. Final Recompiled ETM Data

4. Generation of Origin-Destination Matrix: The excel files obtained from the previous step of data processing is further processed for creating the origin-destination (OD) matrix. The data for each route is filtered, and then the columns full ticket, half ticket and physically challenged are summed up to get the total passengers travelling in a single ticket. Then for each trip number of a route, the unique values in the origin column are identified by removing all the duplicate values and entered in the first column of the spreadsheet, and the unique values of the destination column are selected and entered in the first row excluding the first cell. The source data is filtered, and the sum of the passengers from one origin to destination is entered in the respective cells to obtain the OD matrix from the recomplied ETM data as given in Table 4.3. The route wise OD matrix gives the number of passengers boarding and alighting the bus at various stops. The origin-destination (OD) matrix is to be developed for each route for each trip number. The python (version 2.7) code given in Appendix A.12 is used to develop the OD matrix for reach depot, for each route and for each trip number.

Table 4.3. Sample Origin-Destination Matrix of Trip 1 of Route 1

Origin\ Destination	1	2	3	4	5	6	Total number of boarding passengers
1	0	7	14	14	7	11	53
2	0	0	0	1	0	2	3
3	0	0	0	4	3	1	8
4	0	0	0	0	0	4	4
5	0	0	0	0	0	1	1
6	0	0	0	0	0	0	0
Total number of alighting passengers	0	7	14	19	10	19	69

5. Generation of Time-Series Data from OD Matrix: The OD matrix generated in the previous step is further processed to generate time-series data for each trip of each route. This can be further aggregated at route level and depot level to get the route wise and depot wise time-series data of the number of passengers respectively. The python (version 2.7) code given in Appendix A.13 is used to generate the time-series data shown in Figure 4.11.

This time-series data can be further analysed and modelled for passenger demand estimation and forecasting.

Date	Number of Passengers
03-05-2012	400913
04-05-2012	410328
05-05-2012	345138
06-05-2012	306734
07-05-2012	462823
08-05-2012	424813
09-05-2012	442310
10-05-2012	445534
11-05-2012	426479
12-05-2012	385168
13-05-2012	344152
14-05-2012	438862
15-05-2012	420686
16-05-2012	452632
17-05-2012	360762
18-05-2012	434738
19-05-2012	400579
20-05-2012	330967

Figure 4.11. Time-Series Data Generated from Kerala SRTC ETM Data

4.4.2. Data from Karnataka SRTC (Inter-city Data)

The ETM data from Kerala SRTC was for a period from 2011 to 2013. During the analysis of the Kerala ETM data, it was found that the length of intra-city ETM data available for individual routes varies from 1.5 years to 2 years (2011 to 2013), while inter-city data were not sufficient for time-series modelling. Therefore, inter-city ETM data of Mangalore division Karnataka SRTC was collected for a period of 2013 to 2018, which comprises of a total five-year data, to perform inter-city passenger demand modelling. The intra-city ETM data for Mangalore was not collected since the intra-city services are mainly operated by private operators. The private operators have not maintained and stored the ETM data. The columns of Karnataka SRTC ETM data includes waybill number, trip number, route number, route type, ticket number, ticket type code, description of the ticket type (whether it is passenger or toll gate), origin code, destination code, origin name, destination name, number of adults, number of children, fare, date of ticket issued and time of ticket issued, respectively. The Karnataka SRTC ticket data is more detailed and systematic. This data requires only limited data processing, which requires simple Excel operations. The following are the steps involved in data processing.

ETD_WAYBILL_NO	ETD_TRIP_NO	ETD_ROUTE_NO	ETD_ROUTE_TYPE	ETD_TICKET_NO	ETD_TICKET_TYPE	ETD_TICKET_TYPE_DESCR	ETD_CUR_STOP_CODE	ETD_DST_STOP_CODE	ETD_CUR_STOP_NAME	ETD_DST_STOP_NAME	ETD_ADULTS	ETD_CHILD	ETD_AMOUNT	ETD_DATE	ETD_TD_TIME
31452	2	000001	1	066505	01	PASSENGER	041	014	PERIYAPATNA	B C ROAD	02	01	578	01-01-2016 00:00	00:08:52
31452	2	000001	1	066506	09	TOLL GATE	004	004	KAPU	KAPU	00	00	75	01-01-2016 00:00	05:43:26
31452	2	000001	1	066507	44	-----	001	001	MANIPAL	MANIPAL	00	00	0	01-01-2016 00:00	06:47:24
31457	2	000001	1	064900	01	PASSENGER	038	011	KUSHLANAGAR	MANGALURU	01	00	236	01-01-2016 00:00	01:01:54
31457	2	000001	1	064901	01	PASSENGER	019	014	PUTTUR	B C ROAD	04	00	172	01-01-2016 00:00	03:49:57
31457	2	000001	1	064902	01	PASSENGER	011	002	MANGALURU	UDUPI	02	00	132	01-01-2016 00:00	05:18:14
31457	2	000001	1	064903	01	PASSENGER	011	002	MANGALURU	UDUPI	01	00	66	01-01-2016 00:00	05:18:31
31457	2	000001	1	064904	01	PASSENGER	011	008	MANGALURU	SURATKAL	01	00	26	01-01-2016 00:00	05:19:43
31457	2	000001	1	064905	09	TOLL GATE	009	008	SURATKAL	SURATKAL	00	00	75	01-01-2016 00:00	05:38:58
31457	2	000001	1	064906	44	-----	002	002	UDUPI	UDUPI	00	00	0	01-01-2016 00:00	06:53:27
31485	1	000001	1	071408	01	PASSENGER	041	045	PERIYAPATNA	HUNSUR	01	00	35	01-01-2016 00:00	03:46:42
31485	1	000001	1	071409	33	-----	045	052	HUNSUR	MYSURU	01	00	45	01-01-2016 00:00	04:29:03
31485	1	000001	1	071410	33	-----	045	052	HUNSUR	MYSURU	01	00	45	01-01-2016 00:00	04:29:37
31485	1	000001	1	071411	09	TOLL GATE	052	052	MYSURU	MYSURU	00	00	75	01-01-2016 00:00	17:38:12
31515	1	000001	1	051949	33	-----	002	052	UDUPI	MYSURU	01	00	332	01-01-2016 00:00	19:57:00
31515	1	000001	1	051950	01	PASSENGER	011	052	MANGALURU	MYSURU	01	00	356	01-01-2016 00:00	21:47:56
31515	1	000001	1	051951	01	PASSENGER	011	052	MANGALURU	MYSURU	02	00	712	01-01-2016 00:00	21:52:16
31515	1	000001	1	051952	01	PASSENGER	011	052	MANGALURU	MYSURU	01	00	356	01-01-2016 00:00	21:55:27
31515	1	000001	1	051953	01	PASSENGER	011	052	MANGALURU	MYSURU	01	00	356	01-01-2016 00:00	22:00:29
31515	1	000001	1	051954	01	PASSENGER	011	052	MANGALURU	MYSURU	01	00	356	01-01-2016 00:00	22:01:43
31515	1	000001	1	051955	01	PASSENGER	011	052	MANGALURU	MYSURU	01	00	356	01-01-2016 00:00	22:28:43
31515	1	000001	1	051956	01	PASSENGER	011	019	MANGALURU	PUTTUR	02	00	150	01-01-2016 00:00	22:46:15
31515	1	000001	1	051957	01	PASSENGER	011	052	MANGALURU	MYSURU	01	00	356	01-01-2016 00:00	22:46:48
31515	1	000001	1	051958	01	PASSENGER	011	019	MANGALURU	PUTTUR	01	00	75	01-01-2016 00:00	22:47:38
31515	1	000001	1	051959	01	PASSENGER	011	019	MANGALURU	PUTTUR	01	00	75	01-01-2016 00:00	22:48:46
31515	1	000001	1	051960	01	PASSENGER	014	019	B C ROAD	PUTTUR	01	00	43	01-01-2016 00:00	23:14:24
31515	1	000001	1	051961	01	PASSENGER	014	052	B C ROAD	MYSURU	01	00	325	01-01-2016 00:00	23:15:05
31515	1	000001	1	051962	01	PASSENGER	014	038	B C ROAD	KUSHLANAGAR	02	00	412	01-01-2016 00:00	23:16:14
31515	1	000001	1	051963	01	PASSENGER	025	052	SULYA	MYSURU	01	00	231	02-01-2016 00:00	00:30:53

Figure 4.12. Sample ETM Data from Karnataka SRTC

1. Splitting the ETM Data According to Route: The first step in the data processing was to split the data according to the route. A Microsoft Excel Macro was used for splitting the Excel files based on the route number. The Excel Macro used to split the files based on route number is given in Appendix A.6. Finally, the Excel files having the same route number is again merged to get a single Excel file for a single route using the Excel Macro given in Appendix A.14. Thus, the final recompiled Excel file is route wise ETM data.

2. Inserting the Missing Dates and Generating Time-Series ETM Data: Once the route-wise ETM data is obtained, it is checked manually for any missing dates. If there are missing dates, it is inserted using an Excel Macro given in Appendix A.15. Now, a single Excel file contains the data for all the dates from June 2013 to May 2018 for a route. Next step is to aggregate the data day wise within the Excel file using simple “sumif” function to obtain the final time-series ETM data, as shown in Figure 4.13.

Date	Number of Passengers
01-06-2013	318
02-06-2013	357
03-06-2013	401
04-06-2013	302
05-06-2013	343
06-06-2013	355
07-06-2013	336
08-06-2013	282
09-06-2013	331
10-06-2013	406

Figure 4.13. Time-Series Data Generated from Karnataka SRTC ETM Data

This time-series data can be further analysed and modelled for passenger demand estimation and forecasting.

4.5. DETERMINATION OF LEVEL OF SERVICE

The Level-of-Service provided by the public transportation facility is determined using the service level benchmarks set up by the Ministry of Urban Development (2009). The indicators for determining the Level-of-Service used in the study are the presence of organised public transport system, availability or extent of public transport supply,

service coverage, average waiting time, level of comfort in public transport and the fleet availability. This approach can be used to determine the LoS at the city-level and thus provides a measure for identifying the public transport quality of an urban area. The calculation of various indicators of LoS is as follows:

1. Presence of Organised Public Transport System in Urban Area = $\frac{\text{The total number of buses owned by State Transport Undertakings}}{\text{Total number of buses operating in the city}}$
2. Availability of Public Transport = $\frac{\text{Number of buses available in a city}}{\text{Total population of the city}}$
3. Service Coverage of Public Transport in the City = $\frac{\text{Total length in road kilometres of the corridors on which public transport systems ply in the city}}{\text{Area of the urban limits of the city}}$
4. Average Waiting Time for Public Transport = Half of the average headway
5. Level of Comfort in Public Transport = Load factor

The overall city-wide LoS of public transport facilities is the sum of LoS of the indicators from one to five.

4.6. DETERMINATION OF ACCESSIBILITY INDEX

The proposed public transport accessibility index is a less data-intensive method incorporating the network-based distance to the bus stops, population, travel time and opportunities. The methodology is based on the factor that the index should measure the accessibility which comes from proximity to bus stops, proximity to land use destinations and the proportion of the population served. The accessibility index determination uses a few basic calculations and fundamental commands of Geographical Information System (GIS). The methodology for acquiring and compiling the data are:

4.6.1. Network Development

ArcGIS 10.4 software was used to create the public transport network. Travel time was taken as the network impedance for the public transport network; the speed and delay

data for which were obtained from the National Transportation Planning and Research Centre (NATPAC), Trivandrum. NATPAC has carried out speed and delay study using moving car observer method on nearly 125 km of road stretch in Trivandrum city. The transit data, including bus stops, routes and schedules were collected from Kerala SRTC. A walking speed of 1.2 m/s (IRC 93 1985) was used to calculate the walking time. To calculate the walking distance, the city was divided into grids of 30m x 30m, which is the approximate area per household. In GIS, the average distance from the centroids of the 30m x 30m grids within 500m search radius (Huang 2009) is calculated, which gives the walking distance to the bus stop.

4.6.2. Access to Nearest Bus Stop

Access to the nearest bus stop is the service coverage of the stop and hence known as the stop access coverage. Access to bus stops measures the walking accessibility of the population. A 500-meter buffer is taken around each bus stop, which is, in general, the minimum walkable distance and is overlaid in population density map, as shown in Figure 4.14.

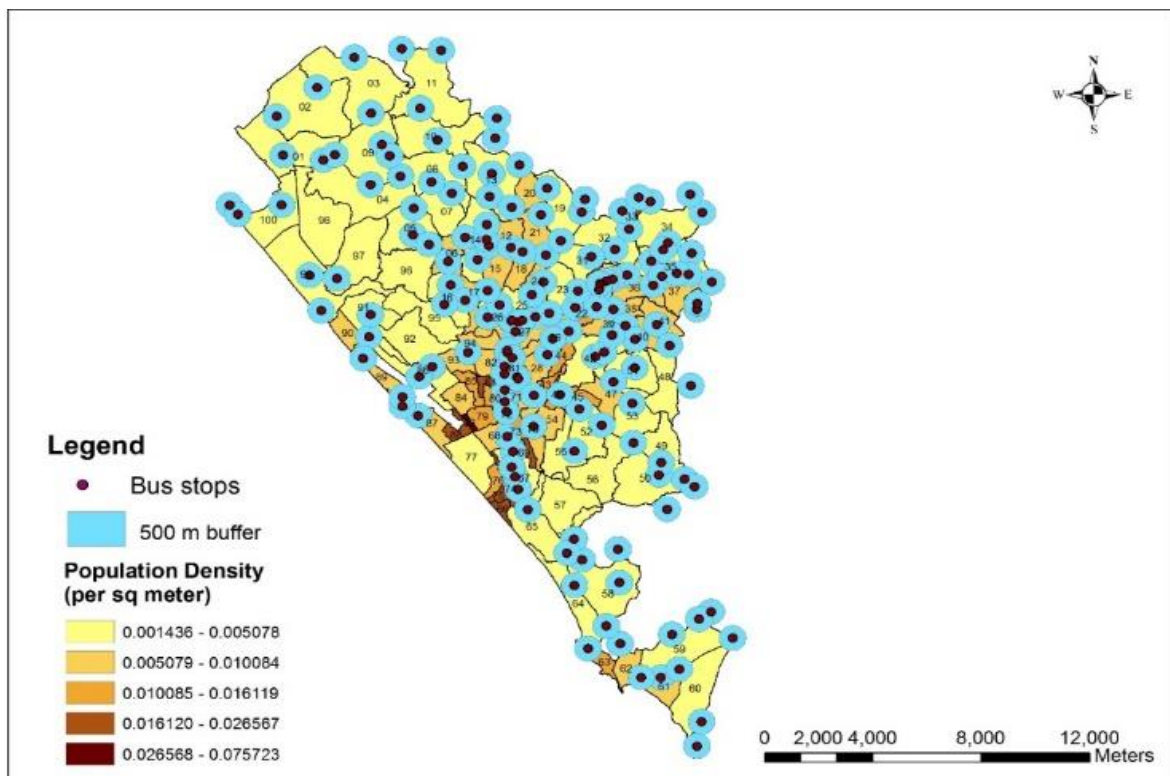


Figure 4.14. Bus Stops with 500m Buffer Overlaid on Population Density Map

The population in the buffer area is taken as the population served and the proportion of the population is calculated as the buffer population divided by the total population of the zones under the influence area of the buffer. It can be seen that only a portion of the urban area is covered in the 500m buffer. The area with low population density is devoid of public transport access.

4.6.3. Access from Bus Stops to the Point of Interests (POIs)

This is a measure of service coverage from each bus stop that gives the area covered in a given time interval and the number of POIs accessible in these areas. Data for Point of Interests (alternatively known as land use destinations) were obtained from various readily available sources as these are the common daily destinations in the study area. The POIs include educational institutions, administrative centres, commercial centres, hospitals, banks and places of worship. Figure 4.15 shows the distribution of POIs in the city, and it can be seen that POIs are concentrated in the central area of the city.

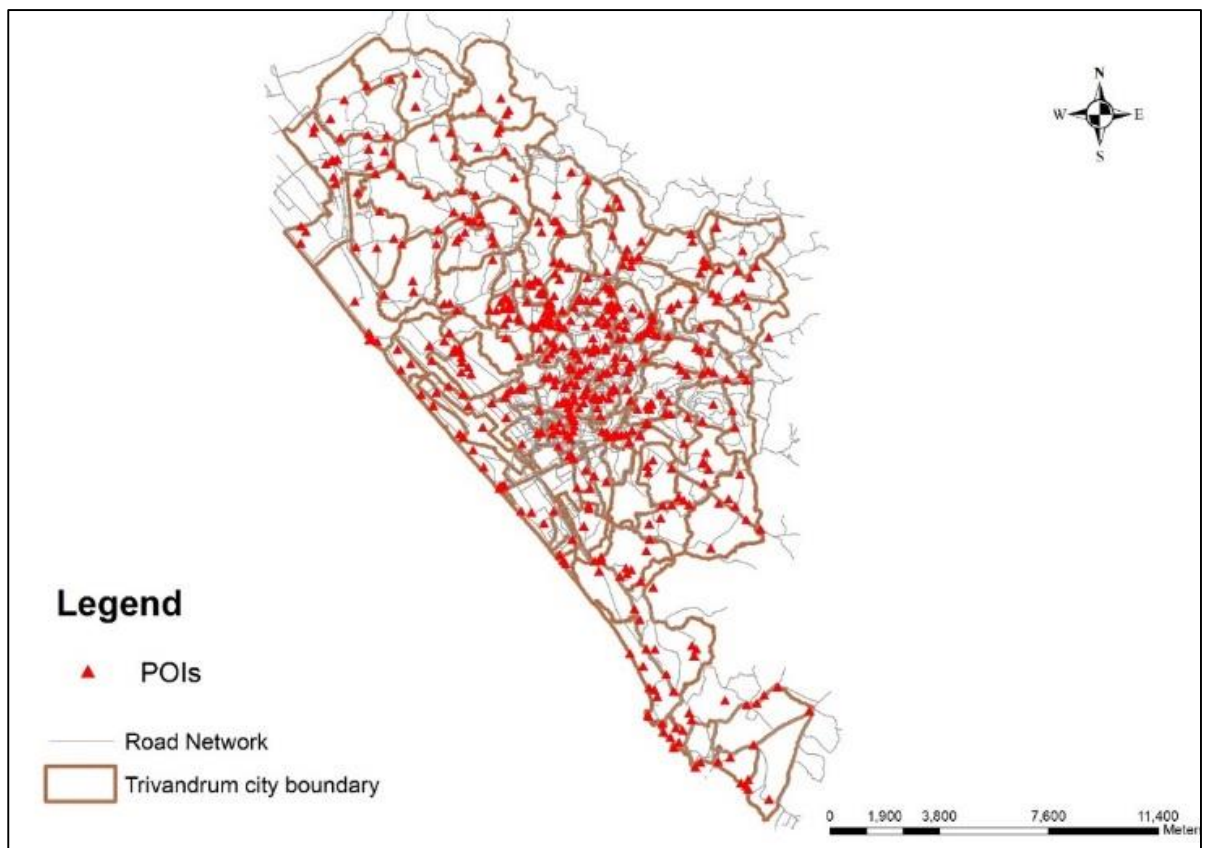


Figure 4.15. Distribution of Point of Interests in the City

The POI shapefile is overlaid on the transportation analysis zones, and the walking distance from the bus stop to the nearest POI is calculated in ArcGIS. The service area for each time interval is created using the Network Analyst Tool. The service area is created as ring polygons corresponding to the time intervals of 15, 30, 45 and 60 minutes respectively. The service area is assigned a weight based on the time interval. The service area having minimum time is given maximum weightage. According to Barron and Barret (1996), the weights derived by Rank Order Centroid (ROC) method results in high performance as the process involves a more systematic analysis of information embedded in the ranks. The weightage calculation formula and the weightage values based on ROC method for a given bus stop are given in Table 4.4. and Table 4.5. respectively.

Table 4.4. Weightage Calculation Formula

Rank	1	2	3	4
Weightage	$(1+1/2+1/3+1/4)/4$	$(1/2+1/3+1/4)/4$	$(1/3+1/4)/4$	$(1/4)/4$

Table 4.5. Weightage Values based on Rank Order Centroid Method

Weightage factors for each service area (Rank Order Centroid method)			
0-15	15-30	30-45	45-60
0.52	0.27	0.15	0.06

The access measure for the bus stop is calculated using the weightage factor for the service area, and percentage of POIs served. Access measure for bus stop 'i' developed in this study is given by Equation 4.1.

$$A_i = \sum_{j=1}^n w_j \times p_j \quad (4.1)$$

where,

A_i = Access measure for bus stop i

w_j = The weights assigned to the service area j

p_j = The proportion of POIs in the corresponding service area.

The final stage in creating accessibility index involves combining the two access measures: access to bus stops and access from bus stops. The total accessibility measure for a bus stop $A_{i(t)}$ given by Equation 4.2, is thus obtained by the summation of

population proportion served, which gives the accessibility to the bus stop and the access measure from service coverage, which provides accessibility from the bus stop.

$$A_{i(t)} = A_i + P_{i(500)} \quad (4.2)$$

where,

$A_{i(t)}$ = The total accessibility measure for bus stop 'i'

A_i = The access measure for bus stop 'i'

$P_{i(500)}$ = The proportion of the population served by bus service calculated as the 500m buffer population divided by the total population of the zones under the influence area of the buffer.

4.7. PERFORMANCE OPTIMISATION OF KERALA SRTC

The performance evaluation and optimisation of Kerala SRTC was performed using an integrated Analytical Hierarchy Process (AHP) and Goal programming (GP) method. In the integrated AHP-GP model, the priorities of the decision variables are established using AHP in the forms of weighted deviations. The proposed model requires the evaluation of the decision variables with regard to how much these will affect the performance of the public transport. Figure 4.16. presents the methodology proposed in this study for performance optimisation of public transport undertakings using integrated AHP-GP method.

The methodology starts with the selection of various performance indicators considering both operator's and user's perception. The decision variables are then categorised as user and operator. Analytical Hierarchy Process (AHP) is used to evaluate these parameters and calculate their weights. AHP is a multi-criteria decision-making approach in which the process hierarchises the system involving multiple objectives, criteria or factors. Saaty (1994) provided a theoretical formulation for AHP, which explains it as a decision support tool for solving complex decision problems. The AHP breaks down the decision problem into simpler elements based on their characteristics and organises them into a hierarchy to incorporate significant quantities of information. The uppermost level of the hierarchy is the "goal" of the problem; intermediate levels are the criteria and sub-criteria while the lowest level is the decision

variables. After the process of hierarchisation, AHP proceeds to priority setting. This involves a pairwise comparison of elements at each level of the hierarchy and assigning a weight to each element to denote the priority based on the intensity of importance as described by Saaty (1994). The judgement matrix of pairwise comparison of the decision variables as per AHP rating scale was collected from experts associated with bus transportation and academia in the designed proforma.

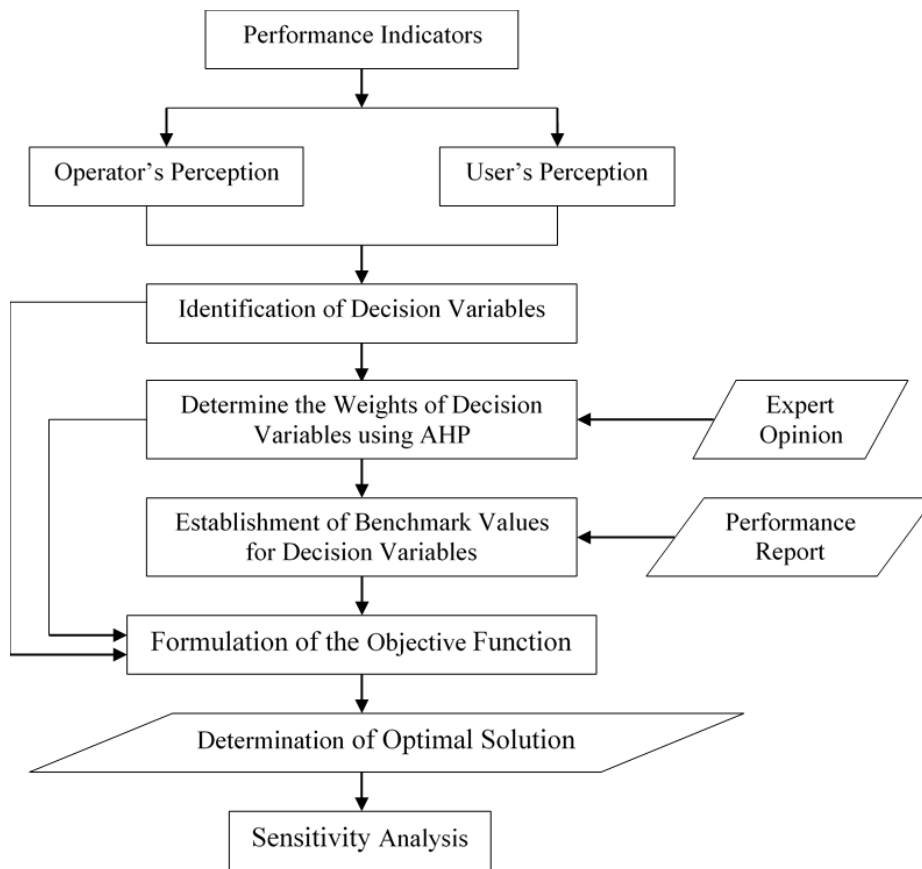


Figure 4.16. Performance Optimisation Methodology Integrating User and Operator Perception

The priorities of these elements can be estimated by determining the Principal Eigenvector ω of the judgment matrix A using Equation 4.3 (Saaty 1995).

$$A \cdot \omega = \lambda_{max} \cdot \omega \quad (4.3)$$

Where,

A = The judgment matrix

ω = The principal Eigenvector

λ_{max} = The largest Eigenvalue of the matrix A

The normalised vector ω gives the vector of priorities of elements with respect to the other elements. In the final step of AHP, logical consistency is checked as pairwise comparisons, and they are based on subjective judgement. To test the consistency of the judgements systematically, first, the consistency index (CI) needs to be estimated using Equation 4.4 (Saaty 1995).

$$CI = (\lambda_{max} - n)/(n - 1) \quad (4.4)$$

Where,

CI = The consistency index

λ_{max} = The largest Eigenvalue of the matrix A

n = The order of the matrix

Then the consistency ratio (CR) is obtained by dividing the consistency index value by random index (RI) values (Saaty 1980) given in Table 4.6.

Table 4.6. Random Index Value

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

The pairwise comparisons in a judgement matrix are considered to be satisfactorily consistent if the corresponding consistency ratio (CR) is less than 0.1. Considering the overall weightage of operator and user, the final weightage for all the decision variables are obtained. These weights are given as penalties for the decision variables in the mathematical model using goal programming (GP) methodology developed for performance optimisation.

Goal Programming (GP) is a multi-objective optimisation used for the multi-criteria decision-making process. This optimisation technique is used to handle multiple objectives which are conflicting in nature. The GP methodology establishes a specific numeric goal for each objective which is to be achieved, assigns weights (obtained from AHP) to these goals based on the relative importance and finds a solution that minimises the weighted sum of deviations from the goals. The general form of the

weighted GP model is given in Equation 4.5, having ‘ m ’ goals, ‘ n ’ decision variables and ‘ p ’ system constraints.

$$\text{Minimise: } Z = \sum_{i=1}^n w_i(p_i + n_i) \quad (4.5)$$

Subjected to,

$$\text{Constraints: } \sum_{j=1}^n a_{ij}x_j - p_i + n_i = b_i, \quad \forall i = 1, \dots, m \quad (4.6)$$

$$p_i, n_i, x_j \geq 0, \quad \forall i = 1, \dots, m; j = 1, \dots, n \quad (4.7)$$

where:

Z = the objective function which is the summation of all weighted deviations,

w_i = the weight assigned to i^{th} goal constraint,

a_{ij} = the coefficient of variable j in the i^{th} goal,

x_j = the j^{th} decision variable,

b_i = the associated right-hand side value,

p_i = positive deviation (overachievement) and

n_i = the negative deviation (under-achievement).

Application of integrated AHP-GP methodology requires solving two problems: AHP and the linear goal program. The main goal of the study is to optimise the performance of Kerala SRTC by considering both the stakeholders: the operator and the user. The problem consists of prioritising the operator and user first and then prioritise the decision variables using AHP. The elements- user and operator are further broken down into various factors or decision variables. Twelve decision variables are identified taking into account both user and operator perception: controllable costs (X_1), non-controllable costs (X_2), taxes (X_3), staff per bus ratio (fleet operated) (X_4), safety (X_5), accessibility (X_6), regularity (X_7), load factor (X_8), fleet utilisation (X_9), percentage of dead kilometre to effective kilometre (X_{10}), journey speed (X_{11}), percentage of cancelled kilometre to scheduled kilometre (X_{12}). Safety, accessibility, reliability and journey speed corresponds to user perception, and all other factors are operator oriented.

Once the hierarchy is established, the pairwise comparison in each level was performed to determine the weights to be used in the objective function of the goal program. The comparison matrices are:

- Level 1: Goal of the study: optimise the performance of Kerala SRTC
- Level 2: The stakeholders of the public transport industry (Kerala SRTC)- user and operator: 2 x 2 matrix
- Level 3: Decision variables: 7 x 7 matrix for the operator and 4x4 matrix for user

Further, the decision variables divided into user-oriented and operator oriented were analysed for determining their priorities. Once the weights for the decision variables are obtained from the AHP method, proceed to the optimisation process using goal programming.

The aim is to minimise the sum of deviations so that the values obtained from the optimisation gives the values close to the goal values. The decision variables with higher weight reach the goal value prior to attainment of goal values for the variables with a lower weight. The objective function is the minimisation of the deviations. The set of constraints formulated takes into account the higher and the lower bounds of the decision variables and are inequality constraints. The constraints are converted to standard equality constraints by assigning positive and negative deviations. The problem is solved using the software LINGO 17.0, a product of LINDO Systems INC, which is optimisation modelling software for linear and non-linear integer programming.

4.8. DETERMINATION OF VARIATION INDICES

The demand for public transport is subjected to seasonal variations and fluctuations during different points of time, such as day, week, month, quarter and year. The useful indication of the seasonal variation is the seasonal index. The 'season' here does not refer to seasons of a year. It refers to constant periods of highs and lows in the demand. This index is an estimate of how much the demand during the period will be above or below the average demand for travel. In other words, the seasonal variation index

relates the average demand in a particular period to the average demand of all the periods as given in Equation 4.8 (Pillai 2008).

$$\text{Seasonal Index} = \frac{\text{Average demand during the seasonal period}}{\text{Average demand for all the periods}} \quad (4.8)$$

The seasonal periods are daily, weekly, monthly, quarterly, or yearly. The average demand for all the period is a value which averages out the seasonality. This is also known as de-seasonalised demand. It could be noted that the sum of all seasonal indices equals the number of periods. The seasonal index can be used to forecast the seasonal demand using any quantitative methods like moving average or exponential smoothing. This forecast does not consider random variations as it uses average demand.

$$\text{Seasonal demand} = \text{seasonal index} * \text{deseasonalised demand} \quad (4.9)$$

For forecasting using seasonality, only de-seasonalised data can be used, which yields de-seasonalised demand forecasts. The seasonal forecast values can be obtained by applying the seasonal index to this base forecast. For calculating variation indices following formulas are used:

- Seasonal Variation Index= Average Seasonal data / Overall average
- Quarterly Index = Quarterly Average / Overall quarterly average
- Daily Variation Index = Daily Average / Overall daily average
- Weekly Variation Index =Weekly Average / Overall weekly average
- Monthly Variation Index = Monthly Average / Overall monthly average

If the overall average is higher or lower than the seasonal average, the demand has to be adjusted down or up to account for seasonality.

4.9. THEORETICAL AND METHODOLOGICAL FRAMEWORK OF TIME-SERIES MODELLING

In this study, Seasonal ARIMA modelling technique is used to estimate the future public transport demand or the number of passengers using PT. For the available data, it is observed that the model required heteroscedasticity methods and Artificial Neural Network (ANN) methods to improve the ARIMA models. Therefore, for public

transport demand modelling the ARIMA models can be improved using heteroscedasticity models and ANN. Once the historical time-series data is collected, the SARIMA modelling can be performed following the methodology developed, which is given in Figure 4.17.

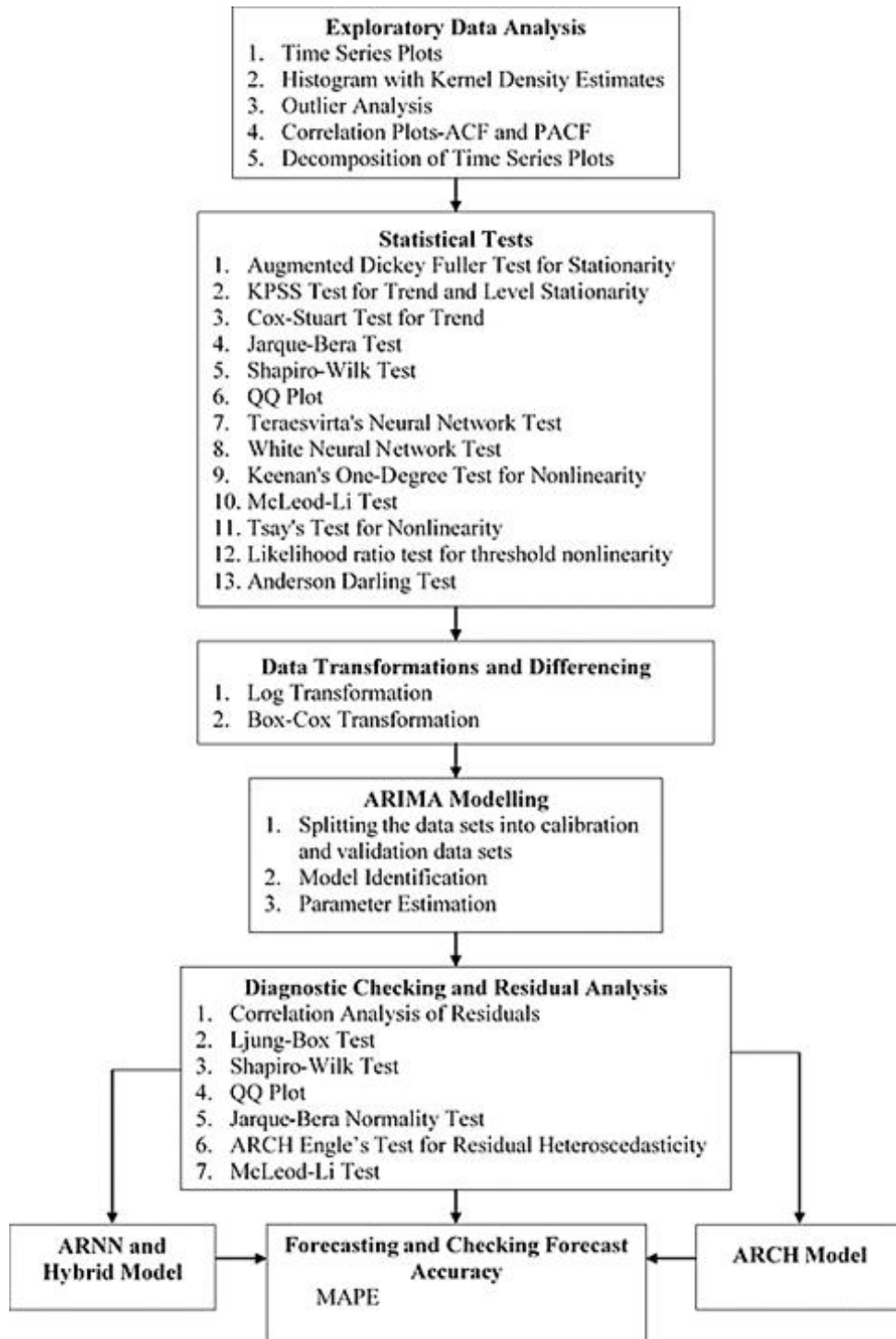


Figure 4.17. Methodological Framework for Time-Series Analysis and Modelling

The model development framework involves three steps- model identification, parameter estimation and diagnostic checking. In model identification, the possible combinations and structure of the model are identified from the data analysis. Then the parameters of the defined model structure are estimated in the parameter estimation stage. The residuals of the developed models are then further analysed in the diagnostic checking step to ensure the adequacy of the model. If the residuals have non-linearity, then AutoRegressive Neural Network (ARNN) modelling is performed. AutoRegressive Conditional Heteroscedasticity (ARCH) modelling is done if the residuals have heteroscedasticity. Finally, the forecasting performance of the model is measured using MAPE. The various steps involved in model development is explained in the following subsections.

4.9.1. Model Identification

The first step in model identification is to identify the patterns in the data, determine the seasonality, identification of the trend and the correlation between the lagged variables. It comprises of plotting the time-series data, correlation analysis, data transformations and differencing. The procedure involved in the analysis of time-series data are as follows:

4.9.1.1. Exploratory Data Analysis: Exploratory data analysis is performed to understand the data. It is the initial critical investigation of the data to find the patterns in the data, unusual observations or outliers, trends and seasonality, generally using graphical representations. These features of the data have to be incorporated into the forecasting model. Time-series plot and univariate distribution analysis using histogram are used in this study.

- (a) Time-series Plot: The time-series plot is a graph of observations versus the time period. It reveals some interesting features of the data like trend, seasonality, outliers, etc.
- (b) Univariate Distribution Analysis: Histogram overlaid with kernel density curve is used to determine the underlying distribution in the data. Kernel density estimate is a nonparametric way to estimate the probability density

function (Zambom and Dias 2012) of the random variable. The kernel density function is calculated using Equation 4.10.

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K \left[\frac{(x-x_i)}{h} \right] \quad (4.10)$$

Where,

$f(x)$ = Kernel density function

K = The chosen kernel function (generally, Gaussian function) of the variable ‘ X ’ (with n observations x_1, x_2, \dots, x_n)

h = The bandwidth or the smoothing parameter

n = The number of data points or observations.

4.9.1.2. Calculation of Missing Percentage: The amount of missing data can lead to biased parameter estimates, incorrect model identification and high standard error. The proportion of missing data is directly related to the quality of the model developed and its statistical parameters. There is no established threshold in the literature regarding the acceptable value of the missing percentage. According to Bennett (2010), statistical analysis is likely to be biased if the missing percentage is more than 10%. In this study, the missing percentage is found out by dividing the available count of the time-series data by the total number of days (length of data range).

4.9.1.3. Data Imputation: Data imputation is the process of replacing the missing values based on the available information with a “likely” value. Since the time-series method is designed to handle complete data set, it is necessary to impute the missing data. Various methods of data imputations found in the literature are mean value imputation, hot deck, cold deck, regression, interpolation and extrapolation. Moritz et al. (2015) suggest that the linear interpolation technique is the best method for problems with all degrees of complexities. Therefore, in this study, linear interpolation is used for data imputation.

4.9.1.4. Checking the Presence of Outliers: The data have to be checked for outliers and subsequently removed to ensure the model to perform well in forecasting. Chen and Liu (1993) procedure of outlier detection is used in this study for the detection of outliers in the time-series data. The method involves the identification of residuals.

Residuals are identified by fitting a Local Polynomial Regression (LOESS) curve for non-seasonal data and through a periodic seasonal and trend decomposition (STL decomposition) for seasonal data. Local polynomial regression method is a nonparametric method used to fit a smooth curve on the variable of interest. The method is different from the conventional regression as the linearity assumption of conventional regression method is not considered in this method. The method is called local regression as the fitting at point 'x' depends only on the neighbouring values. The number of neighbouring points considered for regression is controlled by the span setting 'α'. If the span is large, then the fitted curve will be over-smooth and resulting in loss of information. In STL decomposition, the seasonal time-series data is decomposed into seasonal, trend and remainder components. The seasonal components are estimated iteratively using local polynomial regression fitting. The trend component is computed as the last iteration. Once the curve fitting is done for the data, the residuals are labelled as outliers if it lies outside the range $\pm 2(q_{0.9} - q_{0.1})$, where q_p is the p -quantile of the residuals. Quantiles are the points which divide the probability distribution curve of the variable of interest into continuous intervals with equal probabilities.

4.9.1.5. Correlation Analysis: The correlation analysis of time-series is performed by determining the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Analysis of ACF and PACF reveals exciting features of the time-series data. Autocorrelation measures the linear relationship between the lagged values of the time-series data. The ACF at lag k denoted by ρ_k is given in Equation 4.11. Partial autocorrelation is the correlation which results after removing the linear dependencies of lagged values. The PACF at lag k denoted by φ_{kk} given in Equation 4.12.

$$\rho_k = \frac{\text{cov}(y_t, y_{t+k})}{\sqrt{\text{var } y_t} \sqrt{\text{var } y_{t-k}}} = \frac{\gamma_k}{\gamma_0} \quad (4.11)$$

$$\varphi_{kk} = \frac{\text{cov}(y_t - \hat{y}_t, y_{t+k} - \hat{y}_{t+k})}{\sqrt{\text{var}(y_t - \hat{y}_t)} \sqrt{\text{var}(y_{t+k} - \hat{y}_{t+k})}} \quad (4.12)$$

Where,

ρ_k = ACF at lag k ,

- φ_{kk} = The PACF at lag k ,
- y_t = The observation at time t ,
- y_{t+k} = The observation at time $t+k$,
- γ_k = The autocovariance at lag k ,
- γ_0 = The autocovariance at $k=0$,
- \hat{y}_t = The best linear estimate of y_t
- \hat{y}_{t+k} = The best linear estimate of y_{t+k}

This correlation analysis helps in the identification of the significant time lags, seasonality, stationarity and the order of the model (Cunha and Soares 1999).

4.9.1.6. Decomposition of Time-Series Data: A time-series data y_t consists of three components- seasonal component (S_t), trend-cyclic component (T_t), and irregular or random component (r_t). The time-series can, therefore, be the sum or product of these components. Multiple seasonal decomposition is used to decompose the seasonal time-series data into seasonal, trend and remainder components. The seasonal components are estimated iteratively using local polynomial regression fitting. The trend component is computed as the last iteration. If the data is non-seasonal, then it is decomposed into trend and irregular component. Friedman's super smoother technique is used to decompose the non-seasonal data. This technique can incorporate variable span as opposed to loess, thus accounting for the changes in the variances in the data. An increase in variance calls for an increase in span and vice versa. Friedman's super smoother technique is a non-parametric regression smoother based on local linear regression with adaptive bandwidths or varying spans. In this method, the number of fixed bandwidth curves are estimated by local linear regression. The residuals of these initial estimates are then smoothed using a constant bandwidth. At each data point, the best bandwidths or spans of the initial estimates are chosen based on the smoothed residuals. Those selected best local spans at each data point are then smoothed using a constant bandwidth. The estimates from the initial estimates closest with the smoothed bandwidths are selected. Then, the smoothed outcomes are interpolated. The interpolated outcomes are smoothed with a constant span to get the final estimate.

4.9.1.7. Trend and Stationarity Analysis: The next step in the time-series analysis is to check the trend and stationarity of the data. Cox-Stuart test and Augmented Dickey-Fuller Test is used to check trend and stationarity, respectively. Additionally, the KPSS test is also used as a double check for stationarity analysis.

(a) Augmented Dickey-Fuller Test: A time-series data is considered to be stationary if the statistical properties like mean, variance and autocovariance do not vary with time. Various sources of nonstationarity like unit root, multicollinearity, time-varying variance or structural breaks, have to be tested based on the preliminary data analysis. The most common cause of nonstationarity in the data is the presence of unit root. Augmented Dickey-Fuller (ADF) test is a unit root test to check the stationarity of the time-series data. Null Hypothesis (H_0) states that unit root is present in the data, making the time-series non-stationary. The alternate hypothesis (H_1) indicates that there is no unit root for the series. The test statistics based on the serial correlation of the observations is given in Equation 4.13.

$$DF_t = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \quad (4.13)$$

Where,

DF_t = Dickey-Fuller statistic

$\hat{\gamma}$ = The serial correlations of subsequent observations

SE = The standard error.

Dickey-Fuller test statistics gives negative values. If the computed Dickey-Fuller test statistics is less than the critical value, the stronger the evidence to reject the null hypothesis. Also, a p -value is used as an indicator to reject the null hypothesis. The p -value can be compared with the chosen alpha value (significance level value) to decide whether the null hypothesis should be rejected or not. Smaller the p -value, more evident that the null hypothesis can be rejected. This test must be used with caution as it has a disadvantage of relatively high Type-I error, *i.e.*, incorrect rejection of the true null hypothesis.

(b) Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test: KPSS test is a stationarity test, the null hypothesis stating that the data is level (or trend) stationary and alternate hypothesis is data is not stationary. KPSS test follows a linear regression by breaking the time-series (y_t) into a deterministic trend (β_t), random walk (r_t) and a stationary error (ε_t) (Kwiatkowski et al. 1992) as given in Equation 4.14.

$$y_t = \beta_t + r_t + \varepsilon_t \quad (4.14)$$

Here the random walk r_t is given $r_t = r_{t-1} + u_t$.

Where,

y_t = The observation at time t

β_t = Deterministic trend

r_t = Random walk

ε_t = Stationary error

r_{t-1} = Random walk at time $t - 1$

u_t = The stochastic component or white noise (independent and identically distributed with mean “0” and variance “ σ_u^2 ”).

The stationarity hypothesis is $\sigma_u^2 = 0$, and one-sided Lagrange Multiplier (LM) test statistic is used. If the LM statistic is greater than the critical value, then the null hypothesis is rejected. The p -value can be used as an indicator for accepting or rejecting the hypothesis as in the case of the ADF test.

(c) Cox-Stuart Test: The Cox-Stuart test is a nonparametric test used to identify the trend in the series. This test is a sign test for paired values and measures the number of positive signs. The null hypothesis is that the data is random, *i.e.* it exhibits no trend. The possible values for the alternate hypothesis are two-sided (an upward or downward trend exists), left-sided (upward trend exists), and right-sided (downward trend exists). Let y_1, \dots, y_t be the time-series data. In the testing procedure if t is even (if not, remove the middle value from the series), divide the series into two equal sets: $y_j = y_1, \dots, y_{\frac{t}{2}}$ and $y_{j+\frac{t}{2}} = y_{\frac{t}{2}+1}, \dots, y_t$ (Rutkowska 2015). The test statistic T is given by Equation 4.15.

$$T = \sum_{j=1}^{\frac{t}{2}} \text{sign}(y_{j+\frac{t}{2}} - y_j) , \forall j = 1, \dots, \frac{t}{2} \quad (4.15)$$

If the null hypothesis is true, the statistic T follows a binomial distribution ($T \sim B(\frac{t}{2}, \frac{1}{2})$).

4.9.1.8. Test for Normality: The next statistical test in the analysis of time-series data is the test for normality. This tests whether the data are normally distributed or not. The data must be normally distributed so that the order selection of the model using Akaike Information Criteria (AIC) and the prediction intervals are calculated correctly. The normality tests used in this study are the Jarque-Bera test and Shapiro-Wilk test. Jarque-Bera test is a moment test while Shapiro-Wilk is based on regression and correlation.

(a) Jarque-Bera Normality Test: The Jarque-Bera (JB) test for normality is a type of Lagrange-Multiplier test which tests whether the data have skewness and kurtosis matching the normal distribution. This tests the null hypothesis that the data has a normal distribution as against non-normal distribution. The JB test statistic is given by Equation 4.16.

$$JB = T \left[\frac{S^2}{6} + \frac{(K-3)^2}{24} \right] \quad (4.16)$$

Where,

- JB = Jarque-Bera test statistic
- T = The length of time-series
- S = The skewness coefficient
- K = The kurtosis coefficient

(b) Shapiro-Wilk Test: The Shapiro-Wilk (SW) tests the null hypothesis that the time-series data are i.i.d. (independent and identically distributed) and normal, *i.e.* $N(\mu, \sigma^2)$. The test computes the correlation between the quantiles of standard normal distribution and the ordered data points of time-series data (Machiwel and Jha 2012). The test statistic is given as in Equation 4.17.

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (4.17)$$

Where,

W = Shapiro-Wilk test statistic

$x_{(i)}$ = The ordered sample

x_i = The i^{th} the observation

\bar{x} = The sample mean

a_i = The constants generated from the means, variances and covariances of the order statistics of a sample of size n from a normal distribution (Machiwel and Jha 2012).

If the test statistic has a higher value, then the data comes from a normal distribution. Also, the p -value can be used as an indicator for accepting or rejecting the hypothesis as in the case of the ADF test.

4.9.1.9. Tests for Non-Linearity: The non-linearity test is performed to rule out the possibility of linearity in the data. The various tests that are used for testing the presence of nonlinearity in the data are as follows:

(a) White Neural Network Test: The White Neural Network test for neglected nonlinearity is based on neural network models. In this method, a feedforward single hidden layer neural network is built. The input data (y_t) is attenuated by weights (γ_{ij}) in the links from the input unit to hidden unit. The hidden unit receives the attenuated signals, processes them, produces an output by means of the logistic activation function ($\Phi(\tilde{Y}', \gamma_{ij})$) where Φ is the logistic function and \tilde{Y}' is the input values) and passes the output to the output layer. The null hypothesis of linearity of white neural network test is that the optimal weights of the network are equal to zero (Bisaglia and Gerolimetto 2014). In other words, if the optimal weights of the neural network is zero, then the time-series data is linear.

(b) Teraesvirta's Neural Network Test: The test is similar to White neural network test. The intercept in the power of logistic function chosen as the activation function in White neural network test can affect the test result. Therefore, this test is a modified version of the white test, which, uses a Taylor series expansion of the activation function to determine the weight parameters.

(c) Keenan's One-Degree Test for Nonlinearity: The null hypothesis of Keenan's one-degree test for nonlinearity is that the time-series follows some AR process. Let y_t be the time-series. The number of lags (p) to be used for the linear regression model fitting of y_t is selected first. Then the data y_t is fitted to a linear regression model upto lag p to obtain the fitted value \hat{y}_t . The estimated residuals are $\hat{e}_t = y_t - \hat{y}_t$. The Keenan's test checks whether the squared forecast has any additional forecasting ability by finding the correlation between the residuals and squared forecasts by regressing the squared forecast on the lagged values.

(d) Tsay's Test for Nonlinearity: The Tsay's test for nonlinearity is similar to Keenan's test. The first step of Keenan's step is kept unchanged in Tsay's test. The difference in Tsays' test is that, instead of squared residuals, the products $y_{t-i}y_{t-j}$, $i, j = 1, 2, \dots, p$ are regressed on the lagged values.

(e) McLeod-Li Test: McLeod-Li test is a Portmanteau test of serial independence by examining the Q test statistic of the squared residual of ARIMA process. This tests the Autoregressive Conditional Heteroscedasticity (ARCH) effect based on the Ljung-Box Test. The null hypothesis states that the squared residuals are independent versus the alternate hypothesis of serial dependency. The Q statistic defined by Equation 4.18 tests whether the first m autocorrelations of the squared residuals are significant or not.

$$Q = T(T + 2) \sum_k^m \frac{\hat{r}_k^2}{T-k} \quad (4.18)$$

Where,

Q = The test statistic

T = The length of the series

\hat{r}_k^2 = The squared sample autocorrelation of squared residuals at lag k .

(f) Likelihood-Ratio Test for Threshold Nonlinearity: The likelihood ratio test for determining the presence of Threshold AutoRegressive (TAR) model was proposed by Chan and Tong (1986). The presence of the TAR model indicates that there exists hidden nonlinearity in the data. The null hypothesis of the data follows the AR process against the alternative hypothesis of the TAR model.

(f) Anderson-Darling Test: The Anderson-Darling (AD) test is a test for normality. The rationale of using this test for nonlinearity is that if the data is not normally distributed, then it is nonlinear. It is a goodness of fit test, which measures how good the data fit a normal distribution. The AD test statistic is given in Equation 4.19.

$$AD = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) [\log F(y_i) + \log(1 - F(y_{n-i+1}))] \quad (4.19)$$

Where,

AD = Anderson-Darling test statistic

n = The size of the time-series

$F(.)$ = The cumulative distribution of normal distribution.

4.9.1.10. Data Transformations and Differencing: Data transformations and differencing operations of time-series data are necessary before using the modelling techniques. The fundamental assumption of most stochastic time-series modelling is that data is stationary, *i.e.* the mean, variance and covariance do not change over time. Data transformations help in stabilizing the variance while differencing stabilizes the mean. The data with a linear trend has to be differenced to make it stationary while the curvilinear trend in the data requires a log transformation. If the data is not normally distributed, Box-Cox transformation is necessary to make the data into normality.

4.9.2. Parameter Estimation of ARIMA Model

The parameter estimation includes the determination of the coefficients of the AR term (\emptyset), MA term (θ), SAR term (Φ) and SMA term (Θ) in the seasonal ARIMA model. The Hyndman-Khandakar algorithm which combines unit root tests, minimisation of Akaike Information Criteria (AIC), Corrected AIC (AICc) and Bayesian Information Criterion (BIC) values to obtain the best model is used to identify the model. The parameters are estimated using the maximum likelihood estimation (MLE) method in R (version 3.5) software. The R code used for model development is given in Appendix A.16. The AIC, given by Equation 4.20, is a technique to estimate the likelihood of a model to predict future observation.

$$AIC = -2\log L + 2(p + q + k + 1) \quad (4.20)$$

Second order AIC given by Equation 4.21 known as corrected AIC is used to correct the bias in the Kullnack-Liebler divergence (Sen et al. 2016) in the ARIMA model.

$$AICc = -2\log L + 2(p + q + k + 1) \left[1 + \frac{p+q+k+2}{T-p-q-k-2} \right] \quad (4.21)$$

Another criterion for model selection is the Bayesian Information Criteria (BIC), which measures the trade-off between the model fit and complexity of the model. BIC is calculated using Equation 4.22.

$$BIC = -2\log L + 2(p + q + k + 1) + \log(T)(p + q + k - 1) \quad (4.22)$$

Where,

AIC = Akaike Information Criteria

AICc = Corrected AIC

BIC = Bayesian Information Criterion

L = The likelihood value

k = The number of estimated parameters

T = The sample size

p and *q* = The model parameter counts.

The model with minimum AIC, AICc and BIC value is selected as the best model.

4.9.3. Diagnostic Checking

After the model identification and parameter estimation of the best model, diagnostic checking of the selected best model is performed by the residual analysis. This is to check the adequacy of the model chosen. A good model should have the residuals with the following characteristics (Hyndman and Athanasopoulos 2018):

(a) The residuals are not serially correlated. If the residuals exhibit serial autocorrelation, then it implies that there is information which can be extracted from the residuals to modify the model further.

(b) The residuals have mean “zero”. A non-zero mean of the residuals indicates that the forecasts are biased.

4.9.3.1. Correlation Analysis of the Residuals: The ACF and PACF of the residuals are plotted to check the presence of significant correlation values. If there are significant values of correlation in the ACF and PACF plot, then further information can be extracted to modify the developed model. The requisite of the model to be good is that there are no significant correlations in the residuals.

4.9.3.2. Residual Analysis: Ljung-Box test for autocorrelation of residuals and Jarque-Bera normality test is performed for residual analysis.

(a) Ljung-Box Test: Ljung-Box test is a Portmanteau test to check that the residuals of the fitted time-series model resemble white noise. It is based on the test statistic given in Equation 4.23 and checks the absence of serial autocorrelation. In other words, it tests the hypothesis of random and independently distributed errors versus serially correlated errors.

$$Q^* = T(T + 2) \sum_k^m \frac{\hat{r}_k^2}{T-k} \quad (4.23)$$

Where,

Q^* = Ljung-Box test statistic

T = The length of the data

m = The number of lags being tested

\hat{r}_k = The sample autocorrelation of residuals at lag k .

The Q statistic asymptotically follows $\chi_{(m)}^2$ under H_0 . For a significance level α , the critical region for rejecting the null hypothesis is $Q^* > \chi_{1-\alpha, m}^2$, where $\chi_{1-\alpha, m}^2$ is the $1 - \alpha$ quantile of the χ^2 -distribution with m degrees of freedom. A large value of Q^* suggests that the autocorrelation is not from white noise. In the case of checking the residuals of the ARIMA model, the degrees of freedom is adjusted to reflect the parameter estimation. For example, the ARIMA (p,d,q) model will have a degree of freedom $m-p-q$. The Ljung-Box test is only asymptotically valid and may perform poorly for a smaller sample size.

(b) Test for Normality: The normality tests used are the Jarque-Bera test and Shapiro-Wilk test as explained in the. Section 4.9.1.8 Jarque-Bera test is a moment test while

Shapiro-Wilk is based on regression and correlation. The normality tests are performed on the residuals to check whether it is normally distributed.

(c) Tests for Non-Linearity: The non-linearity test is used to check whether there exist any non-linearity in the residuals. The tests include Teraesvirta's Neural Network Test, White Neural Network Test, Keenan's One-Degree Test for non-linearity, McLeod-Li Test, Tsay's Test for non-linearity and Anderson-Darling Test. The procedure of these tests are as explained in section 4.9.1.9.

(d) Heteroscedasticity Analysis: ARCH Engle's test for residual heteroscedasticity is used in this study to check the heteroscedasticity of the residuals. It performs a Portmanteau Q and Lagrange Multiplier test for the null hypothesis that the residuals of the ARIMA model are homoscedastic (Qui 2015). The ARCH test consists of two parts: a Portmanteau Q test similar to the Ljung-Box to test whether the residuals are uncorrelated and Lagrange Multiplier (LM) test. The Portmanteau Q test is identical to the Ljung-Box test on squared residuals, which checks whether the residuals are uncorrelated. On the other hand, the LM test fits a linear regression model on the squared residuals and examine whether the model fit is good.

4.9.4. AutoRegressive Neural Network Model

The basic time-series modelling is based on the linear stochastic process such as ARIMA, where the predictions are a linear function of past observations and past residuals. It is difficult to produce accurate predictions if the data has non-linearity. AutoRegressive Neural Network (ARNN) is a semiparametric method to handle non-linearity in the data.

A process is called an AutoRegressive process, if the observation at time ' t ' can be explained by the observation at time ' $t-1$ ' or ' $t-n$ ', i.e., it can be expressed as

$$y_t = f(y_{t-n}) + \varepsilon_t \quad (4.24)$$

If the $f(y_{t-n})$ is linear, then the process is a linear AR model, and if the function is nonlinear, then it is a nonlinear AR model. Hence, nonlinear modelling has to be used.

In this study, ANN is used to model the nonlinear AR model and hence known as the AutoRegressive Neural Network (ARNN) model.

ARNN is a feed-forward neural network, which provides a general approximation; hence, any hidden non-linearity will be taken care of. ARNN consists of three layers (input, hidden and output layers) with a bounded non-linear function in the hidden layer which handles the non-linearity in the data.

4.9.4.1. ARNN Architecture

Linear Autoregressive function if expressed as a neural network consists of two feed forward layers – Input layer (Independent variable) and an Output layer (Dependent variable). The constant term in AR is decomposed into a bias neuron with value 1 and bias parameter α_0 . Sometimes the linear part is not able to suffice the need. Hence, a nonlinear part, known as the hidden layer, is introduced between the input and output layer. The hidden layer consists of h hidden neurons which transform the input data attenuated by the weighing parameters γ_{ij} and the bias parameter γ_{0j} . Inside the non-linear part, a nonlinear function is introduced, which transforms the data. The hidden layer in Neural Network is also known as the black box, as it contributes to the result, but the transformation process is unknown. The transformation in the hidden layer takes place based on the activation function used. The activation function should be a bounded function that is integrable and non-polynomial. The boundness of the activation function is a necessary condition to maintain the stationarity in the data. The best-known activation function is the sigmoidal function. The sigmoidal function reduces the effect of outliers as they compress the data to the high and low ends. A hidden neuron in the hidden layer is denoted by Equation 4.25.

$$\psi (\gamma_{0j} + \sum_{i=1}^n \gamma_{ij} y_{t-i}) \quad (4.25)$$

Where,

- ψ = The activation function
- i = The number of lags
- j = The number of hidden layers
- γ_{0j} = The bias parameter

γ_{ij} = Weighing parameter for a link from input node i to hidden node j
 y_{t-i} = The observation at lag i

Further, each hidden neuron is attenuated by a weighing parameter β_j in the links between the hidden and output layer. The weights are selected in the neural network framework using a learning algorithm which minimises the cost function. The cost function is a measure of how bad the model is in terms of its ability to estimate the relationship between y_t and y_{t-n} . In this study, the cost function is the Mean Squared Error. The mathematical representation of the full neural network is as given in Equation 4.26.

$$y_t = \sum_{i=1}^n \alpha_i y_{t-i} + \sum_{j=1}^h \psi (\gamma_{0j} + \sum_{i=1}^n \gamma_{ij} y_{t-i}) \beta_j + \varepsilon_t \quad (4.26)$$

Where,

α_i = the coefficient of linear AR term
 y_{t-i} = the observation at lag i
 ψ = the activation function
 γ_{0j} = the bias parameter
 γ_{ij} = the weighing parameter for a link from input node i to hidden node j
 β_j = the weighing parameter from hidden node j
 ε_t = the error term.

Figure 4.18. shows the steps involved in ARNN modelling. The steps involved in the ARNN model are:

- Check for non-linearity in the data: - Statistical tests for non-linearity are conducted to check for non-linearity in the data. If non-linearity exists, then ARNN method is to be used.
- Neural Network is designed based on the lagged inputs as obtained from the PACF plot of the model. The PACF plot gives the order of the AR model and also gives the lag up to which correlations for AR part exists. Lag is considered based on this correlation and the lag at which AIC is minimum is considered.

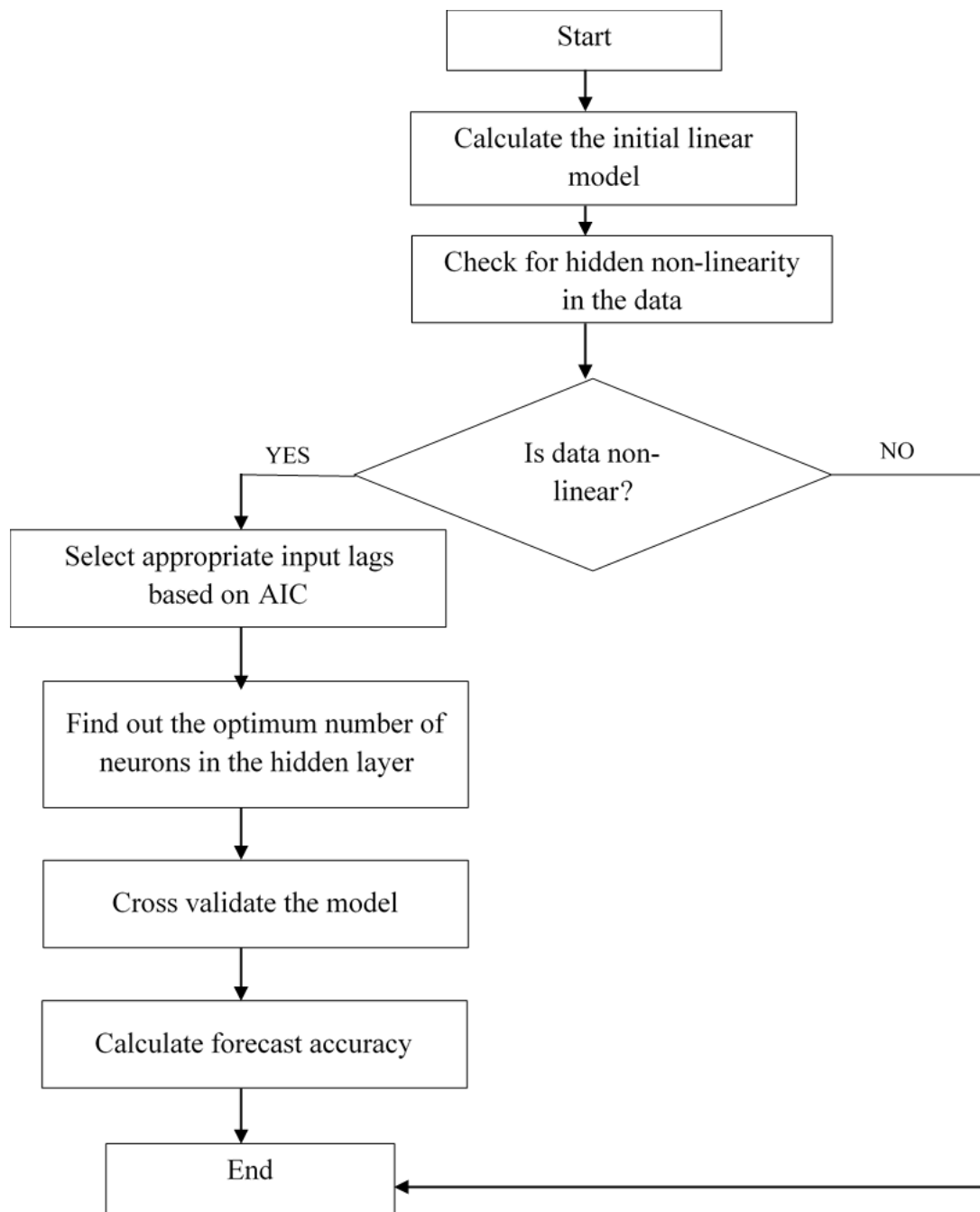


Figure 4.18. The Methodology of ARNN Model

- ARNN is generally represented as $ARNN(p,k,P)$ where p is the non-seasonal lagged inputs, k is the number of neurons in the hidden layer, and P is the seasonal lags if present.
- Calculating the prediction interval- Neural Networks are not based on the well-defined stochastic model. Hence cannot specify the prediction intervals. Instead, the prediction intervals are calculated by simulation. By an iterative process, the

distribution of the data based on the neural network can be identified. This step is done in the forecast process.

- 10-fold cross-validation of the fitted model is carried out, and the credibility of the fit is assessed based on the Ljung-Box test result in cross-validation.

4.9.5. Hybrid ARIMA-ARNN Model

From the literature, it was found that ARIMA can be used to model linear data while other modelling techniques have to be used for the nonlinear part. If the nonlinearity is due to the presence of nonstationarity in the data, then the data can be transformed to remove the nonstationary nonlinear part. This transformed data, which is now linear, can be modelled using linear ARIMA model. Further, if the nonlinearity is due to the presence of conditional heteroscedasticity in the data, then the linear part of the data can be modelled using ARIMA, and the conditional heteroscedasticity can be modelled using ARCH. But if there exist complex nonlinearities, then both ARIMA and ARCH modelling techniques are incapable of modelling these characteristics. Therefore, methods like Artificial Neural Networks can be used. The ANNs used for modelling linear problems have mixed results (Zhang 2003). In the practical scenarios, the characteristics of the data are completely unknown. Therefore, a hybrid model which incorporates the capabilities of both the linear ARIMA model and nonlinear ARNN is considered a good approach for practical problems.

The methodology adopted for modelling the hybrid ARIMA-ARNN model is as follows. The time-series data is considered to have a linear part (L_t) and nonlinear part N_t .

$$y_t = L_t + N_t \quad (4.27)$$

The first step is to model the linear part using a linear ARIMA model. Now, the residuals of the ARIMA model is supposed to contain only the remaining nonlinear characteristics of the data. Let e_t be the residuals of the linear ARIMA model at time 't' expressed as in Equation 4.28.

$$e_t = y_t - \hat{L}_t \quad (4.28)$$

Where,

e_t = The residuals of the linear ARIMA model at time 't'

y_t = The observation at time t

\widehat{L}_t = The forecast value using ARIMA for time 't'

L_t = The linear model

N_t = The nonlinear model

The residuals are checked for any remaining linearity and nonlinearity. If linearity still exists in the residuals, the ARIMA model is further modified to incorporate the residual linearity. If the data has nonlinearity, it must be further analysed to rule out the possibilities of heteroscedasticity. If the nonlinearity is due to the presence of heteroscedasticity, then the model can be improved using ARCH modelling technique. If the residuals have hidden nonlinearity, then it can be modelled using ARNN modelling technique. The methodology for ARNN model is as explained in section 4.8.4. The thing to be noted is that the neural network is applied to the residuals of the ARIMA model. The nonlinear ARNN model is denoted as \widehat{N}_t . Thus, the combined ARIMA-ARNN model forecast is given by Equation 4.29.

$$\widehat{y}_t = \widehat{L}_t + \widehat{N}_t \quad (4.29)$$

Thus, the hybrid method exploits the capabilities of both ARIMA and neural network to improve forecasting performance.

4.9.6. Hybrid ARIMA-GARCH Modelling

From the literature, it was found that ARIMA can be used to model linear data while other modelling techniques have to be used for the nonlinear part. If the nonlinearity is due to the presence of nonstationarity in the data, then the data can be transformed to remove the nonstationary nonlinear part. This transformed data, which is now linear, can be modelled using linear ARIMA model. Further, if the nonlinearity is due to the presence of conditional heteroscedasticity in the data, then the linear part of the data can be modelled using ARIMA, and the conditional heteroscedasticity can be modelled using ARCH. Thus the ARCH/ GARCH is a hybrid model which can be used to model

the linear mean and conditional variance. The flowchart showing the methodology of ARIMA-GARCH modelling is shown in Figure 4.19.

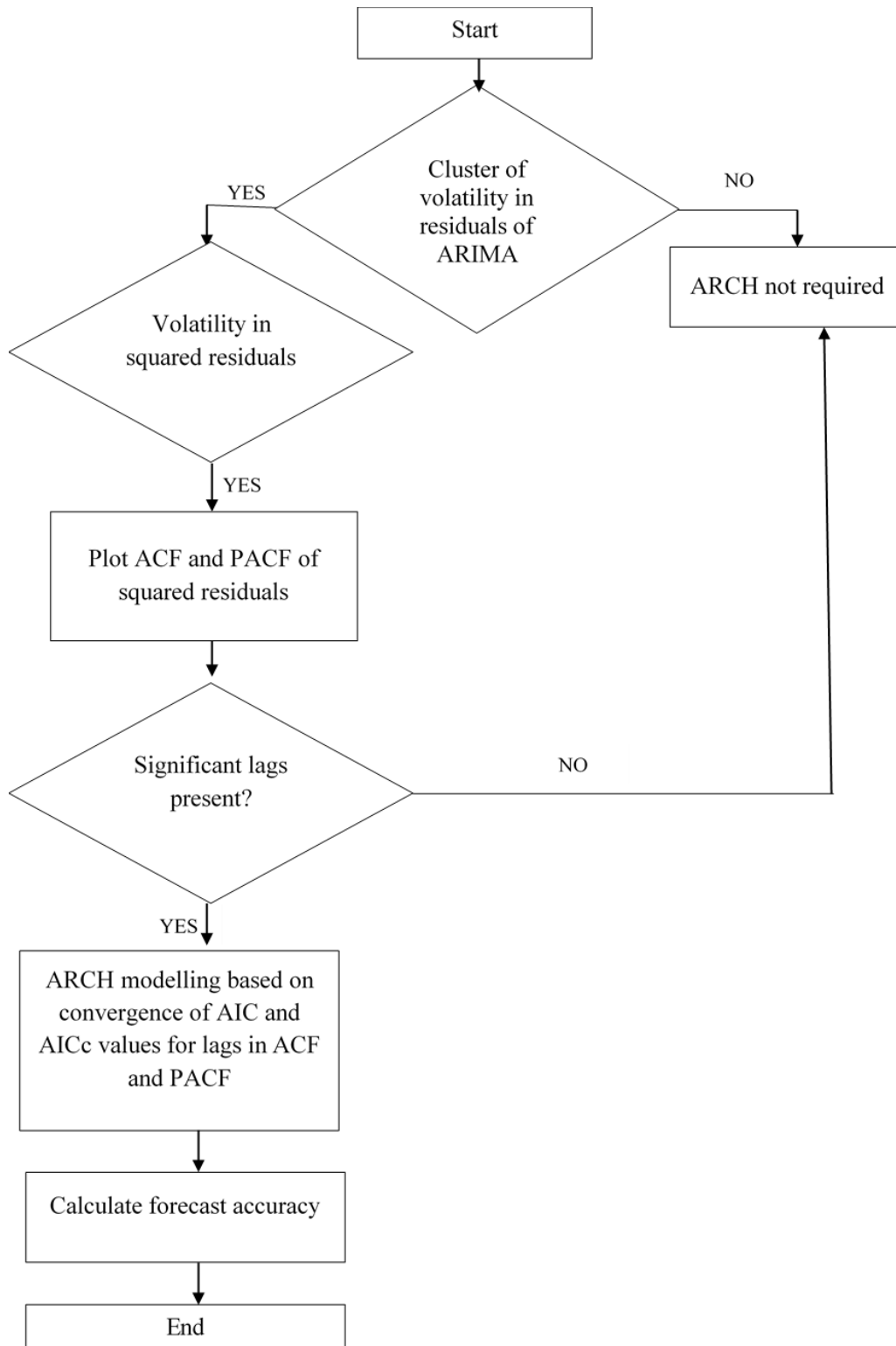


Figure 4.19. The Methodology of ARIMA-GARCH Model

The methodology adopted for modelling the hybrid ARIMA-GARCH model is as follows. The time-series data is considered to have a linear part (L_t) and nonlinear part N_t .

$$y_t = L_t + N_t \quad (4.30)$$

The first step is to model the linear part using a linear ARIMA model. Now, the residuals of the ARIMA model is supposed to contain only the remaining nonlinear characteristics of the data. Let e_t be the residuals of the linear ARIMA model at time 't' expressed as in Equation 4.31.

$$e_t = y_t - \hat{L}_t \quad (4.31)$$

Where,

- e_t = The residuals of the linear ARIMA model at time 't'
- y_t = The observation at time t
- \hat{L}_t = The forecast value using ARIMA for time 't'
- L_t = The linear model
- N_t = The nonlinear model

The residuals are checked for any remaining linearity and nonlinearity. If linearity still exists in the residuals, the ARIMA model is further modified to incorporated the residual linearity. If the data has nonlinearity, it must be further analysed to rule out the possibilities of heteroscedasticity. If the nonlinearity is due to the presence of heteroscedasticity, then the model can be improved using ARCH modelling technique. If the residuals have hidden nonlinearity, then it can be modelled using ARNN modelling technique.

Once the ARCH effects are determined using heteroscedasticity tests explained in section 4.7.3, then the next step involves identifying the proper orders for the GARCH(P,Q) model. The squared residuals are plotted to see the presence of conditional volatility. The AIC and BIC, along with ACF and the PACF of squared return of the residuals, are used to select the appropriate model order. The ACF and PACF of the residuals and the squared residuals are plotted to check the presence of

significant spikes or significant correlation in the residuals. The lag at which there is a significant correlation in the ACF plot is taken as the order P of the GARCH model. The significant correlation in the PACF plot gives the order Q . In general, a desirable model is one that minimises the AIC or the BIC on the significance tests for each parameter. Parameter estimation of the model is using Maximum Likelihood Estimation (MLE) since the GARCH model is non-linear. Thus the hybrid ARIMA-GARCH model is the summation of the forecasts from the ARIMA model and GARCH model in which linear mean is modelled using ARIMA, and conditional variance is modelled using GARCH.

4.10. SUMMARY

This chapter describes the data collection and methodology to achieve the objectives of this research, along with the procedure adopted for data processing. The summary of this chapter is as below.

- The data used for the study are mainly intra-city Electronic Ticket Machine data from Kerala SRTC. The number of buses, bus stops, timetables and schedules were also collected from Kerala SRTC. Also, the physical and financial performance data were obtained from Kerala SRTC in the form of Annual Audit Reports. Economic Review published by State Planning Board, Kerala was also collected for getting the financial performance data of Kerala SRTC.
- The methodology for LoS calculation, performance evaluation and determination of accessibility of public transport for Trivandrum study area is explained.
- During the analysis of the Kerala ETM data, it was found that the length of intra-city ETM data available for individual routes varies from 1.5 years to 2 years (2011 to 2013), while inter-city data were not sufficient for time-series modelling. Therefore, inter-city ETM data of Mangalore division Karnataka SRTC was collected for a period of 2013 to 2018, which comprises of a total five-year data, to perform inter-city passenger demand modelling. The intra-city ETM data for Mangalore was not collected since the intra-city services are

mainly operated by private operators. The private operators have not maintained and stored the ETM data. The performance evaluation of Karnataka SRTC was not performed since the data collected from Karnataka was only for modelling inter-city passenger demand.

- The Level-of-Service can be used to determine the city-wide Level-of-Service provided by the public transport system using the service level benchmarks. This approach can be used to determine the LoS at the city-level and thus provides a measure for identifying the public transport quality of an urban area.
- The methodology used in the study to evaluate the accessibility is based on the factor that the accessibility index should measure the accessibility which comes from proximity to bus stops and land use destinations. The GIS-based public transport accessibility index developed gives the accessibility of a single point (i.e., the bus stop) by considering access to the point, access from the point to various land use destinations and network impedance.
- The performance evaluation of Kerala SRTC can be performed using a weighted goal programming methodology integrated with analytical hierarchy process. The user's and operator's perception are included in the study. These decision variables based on both operator and user were prioritised using AHP and are modelled as a goal program.
- Integration of user and operator ensures that the optimisation of the public transport industry is executed by safeguarding the interest of both the user and the operator.
- Further, time-series modelling methodology for intra-city and inter-city bus passenger demand forecasting is explained. It includes a comprehensive methodology for ARIMA model selection and forecasting, which requires only basic econometric knowledge.

- Also, the method to improve the forecasting performance of the ARIMA model by incorporating the characteristics of nonlinear ARNN and ARCH model are elaborated in this chapter.
- This methodology is useful when there is limited or no information available about the factors affecting demand.

CHAPTER 5

RESULTS AND DISCUSSION

5.1. GENERAL

In this chapter, the results of the study are presented and discussed with reference to the main aim of the study, which is to develop an “ETM based Bus Transport Demand Model”. The main findings in terms of research contributions, the evaluation of the existing condition of the public transport facilities of the study area and application of the methodology to various scenarios are presented in this chapter. Since the travel demand is affected by the performance of the public transport system and the bus transport agencies operating the services, this study evaluates the performance of the public transport system and the public transport operator. The level of service characteristics, accessibility and performance evaluation of Kerala SRTC are discussed. The performance evaluation of Karnataka SRTC was not performed since the data collected from Karnataka was only for modelling inter-city passenger demand. The application of time series methodology to various scenarios (intra-city and inter-city) is the important aspect of this study, and hence, it is the main focus of attention in this chapter. The analysis of the time-series ETM data, development of various models and the forecasting of the passenger demand for the selected forecast horizon are explained in detail.

5.2. LEVEL OF SERVICE

The Level-of-Service (LoS) provided by the public transportation facility is determined using the service level benchmarks set up by the Ministry of Urban Development (2009). The indicators for determining the LoS used in the study are the presence of organised public transport system, availability or extent of public transport supply, service coverage, average waiting time, level of comfort in public transport and the fleet availability. This approach can be used to determine the LoS at the city-level and thus provides a measure for identifying the public transport quality of an urban area. The determination of LoS is performed only for Trivandrum study area since the data

collected from Karnataka was only for modelling inter-city passenger demand. The indicators for LoS calculations and their ranges as per MoUD is given in Table 5.1.

Table 5.1. Indicators to Calculate City-Wide Level-of-Service (LoS) of Public Transport Facilities

Level of Service	1. Presence of Organised Public Transport System in Urban Area (%)	2. Extent of Supply/ Availability of Public Transport	3. Service Coverage of Public Transport in the city	4. Average waiting time for Public Transport users (minutes)	5. Level of Comfort in Public Transport
1	≥ 60	≥ 0.6	≥ 1	≤ 4	≤ 1.5
2	40 – 60	0.4 – 0.6	0.7- 1	4 – 6	1.5 – 2.0
3	20 – 40	0.2 – 0.4	0.3 – 0.7	6 – 10	2.0 – 2.5
4	< 20	< 0.2	< 0.3	> 10	> 2.5

The first indicator is the presence of organised public transport system in the urban area expressed as a percentage. If the percentage is greater than or equal to 60, then the LoS is given a value ‘one’. If the value is 40-60% then, the LoS is two. A 20-40% presence of public transport means the LoS is three, while less than 20% means LoS is four. The second indicator used in the study is the availability of public transport. It is the ratio of the number of buses per thousand populations. The value ≥ 0.6 corresponds to LoS1, 0.4-0.6 means LoS2, 0.2-0.4 means LoS3 and < 0.2 corresponds to LoS4. The third indicator is the service coverage of public transport in the city. It is the ratio of the length of road kilometres in which public transport is operating to the area of the urban limits of the city. The value ≥ 1 corresponds to LoS1, 0.7-1 means LoS2, 0.3-0.7 means LoS3 and < 0.3 corresponds to LoS4. The fourth indicator is the average waiting time, taken as half of the average headway. The value ≤ 4 corresponds to LoS1, 4-6 means LoS2, 6-10 means LoS3 and > 10 corresponds to LoS4. The fifth indicator is the level of comfort in public transport determined based on the load factor or crowding in the bus. The value ≤ 1.5 corresponds to LoS1, 1.5-2.0 means LoS2, 2-2.5 means LoS3 and > 2.5 corresponds to LoS4.

Table 5.2. shows the calculation of the first indicator- the presence of organised public transport system in the urban area. The presence of organised public transport system is determined as the percentage of State Transport Undertaking (STU) buses operating

in the city to the total number of buses in the city. The inter-city buses are not included as a part of urban public transport buses. It was found that there was 80% presence of organised public transport in the urban area. Therefore, the LoS for indicator 1 is found to be ‘one’.

Table 5.2. Indicator 1. Presence of Organised Public Transport System in Urban Area

(a)	The total number of buses operating in the city (B)	Kerala SRTC	446	699
		Private buses	103	
		JnNURM	150	
(b)	The total number of buses under the ownership of State Transport Undertakings (A)	560		
(c)	Presence of Public Transport System in Urban Area (%) = $[A / B] \times 100$	80		
	LoS1 as per Table 5.1	1		

The calculation of the LoS for indicator 2- the availability of public transport is given in Table 5.3. The extent of supply, or the availability of public transport is determined as the ratio of the number of buses available in the city to the total population of the city. The LoS for the second indicator was found to be ‘2’.

Table 5.3. Indicator 2. Availability of Public Transport

(a)	Number of buses available in a city on any day	446		
(b)	Total population of the city	Census,2011	957675	
		Growth rate (Economic Review,2013)	2.1%	
		Current Population	997897.4	
(c)	Availability of public transport = $[\text{No of Buses available in a city on any day} / \text{Total population of the city}]$	0.45		
	LoS2 as per Table 5.1	2		

The LoS for third indicator-the service coverage of public transport in the city is given in Table 5.4. The third indicator is the service coverage of public transport in the city. It is the ratio of the total length in road kilometres on which the public transport operates to the urban area limits of the city. The total length of public transport corridor within the urban limit is considered. The LoS of indicator three is found to be ‘1’.

Table 5.4. Indicator 3. Service Coverage of Public Transport in the City

(a)	Total length in road kilometres of the corridors on which public transport systems ply in the city (Road km) = A	651.35
(b)	Area of the urban limits of the city (Area in sq.km) = B	211.76
(c)	Service Coverage (road km/ sq. km) = [A / B]	3.08
	LoS3 as per Table 5.1	1

The average waiting time for public transport users is the fourth indicator and is determined, as shown in Table 5.5. The fourth indicator is the average waiting time for public transport, determined as half of the average headway. The headway details were obtained from Kerala SRTC schedule data. The LoS of the fourth indicator is calculated as ‘2’.

Table 5.5 Indicator 4. Average Waiting Time for Public Transport Users’

(a)	Average headway of buses/route Calculated for peak hours at the bus stops Morning: 8:00 – 10:00 am Evening: 4:00 – 6:00pm	10 minutes (Source: Kerala SRTC schedule data)
(b)	Waiting time /route	Half of the average headway = 0.5×10 =5minutes
	LoS4 as per Table 5.1	2

The LoS which comes from the level of comfort in public transport is given in Table 5.6. The level of comfort is based on the crowding in the buses. The fifth indicator is the level of comfort in public transport. This is based on the load factor of the bus. Since the load factor is 0.8, the LoS of the fifth indicator is ‘1’.

Table 5.6. Indicator 5. Level of Comfort in Public Transport

Passenger comfort level measured based on Load factor (passengers per seat)	
Load Factor	0.8 (Source: Kerala SRTC)
Los5 as per Table 5.1	1

The overall city-wide LoS of public transport facilities as per MoUD (2009) is the sum of LoS of the indicators from one to five. The overall LoS calculated as the sum of LoS of indicators one to five is tabulated in Table 5.7.

$$\begin{aligned} \text{The overall LoS} &= \text{LoS1} + \text{LoS2} + \text{LoS3} + \text{LoS4} + \text{LoS5} \\ &= 1 + 2 + 1 + 2 + 1 = 7 \end{aligned}$$

Table 5.7. Overall Level-of-Service of City-Wide Public Transport Facilities

Overall LOS	Calculated LOS
1	< 12
2	12-16
3	17-20
4	21-24

(Source: MoUD 2009)

Quality of service, characterized by the LoS, is identified as the major factor determining the demand for public transport. This study determines the city-wide LoS of Trivandrum City provided by the public transport system using the service level benchmarks of MoUD. The indicators for determining the Level-of-Service used in the study are the presence of organised public transport system, availability or extent of public transport supply, service coverage, average waiting time, level of comfort in public transport and the fleet availability. This approach can be used to determine the LoS at the city-level and thus provides a measure for identifying the public transport quality of an urban area. The LoS of the indicators are calculated separately and then overall LoS of the city-wide public transport is calculated as the sum of all the indicators LoS. The sum of all the LoS indicators is 'seven'. Therefore, the overall LoS is considered as 'one' from Table 5.7. The presence of organised public transport system in the urban area is found to be 80% while the availability or extent of public transport is moderate. The city has good service coverage, *i.e.*, the public transport ply on most of the corridors in the urban area. But the average waiting time is more than 5 minutes making the LoS as 2, which is a moderate value. However, since the load factor is very less (with a value of 0.8) in public transport, the level of comfort provided is good. It can be concluded that the city has a good public transport system. The facilities are available wide-spread and are available to the people. Thus, the city has a comfortable public transport system. However, if the number of buses can be increased by changing the size of buses and reduce the waiting time, then LoS can be further improved, which will attract more passengers to public transport.

5.3. ACCESSIBILITY INDEX

The proposed public transport accessibility index is a less data-intensive method incorporating the network-based distance to the bus stops, population, travel time and

opportunities. The methodology is based on the factor that the index should measure the accessibility which comes from proximity to bus stops and land use destinations and the proportion of the population served. The accessibility index determination uses a few basic calculations and fundamental commands of GIS. Table 5.8. presents the ranges and categories of the accessibility index developed in this study. The index was grouped into 5 categories including very poor, poor, moderate, good and very good. The accessibility of about 6.55% of bus stops are very poor, which serves about 2.36% of the population. 1.19% bus stops serving 1.47% population have very good accessibility. Overall, around 45% of the population has moderate accessibility.

Table 5.8. Accessibility Index Ranges and Categories

Range	Accessibility Index Category	Bus Stops (%)	Population (%)
0.1- 0.26	Very Poor	6.55	2.36
0.26- 0.41	Poor	36.90	20.91
0.41- 0.57	Moderate	36.31	44.38
0.57-0.72	Good	19.05	30.88
0.72-0.88	Very Good	1.19	1.47

Figure 5.1 illustrates the bus stop wise accessibility index of Trivandrum Municipal Corporation. As demonstrated in Figure 5.1, it can be observed that the areas with very poor and poor accessibility lie in the periphery of the study area while moderate and good accessible areas are the central area of the city. In other words, bus stops located in the central area of the city have high accessible values; the accessibility values decrease as the distance from the city centre increases.

Correlation analysis was used to investigate whether there is any relation between developed accessibility index and the trips performed using public transport. The fare transaction of ticket sales using Electronic Ticketing Machine provided the number of passengers boarding and alighting each bus stops. This was used to determine the trips (in precise, the number of passengers) performed using public transport. The value of correlation coefficient ranges between -1 to 1 representing negative to a positive relationship between the variables under consideration. The results show that the accessibility index developed in the study and the trips performed using public transport are having a positive correlation with a value of 0.67. This is in association with the literature (Morris et al. 1979) that accessibility to public transport is one of the

determining factors of ridership and hence it can be justified that the accessibility index developed in this study has an acceptable level of performance.

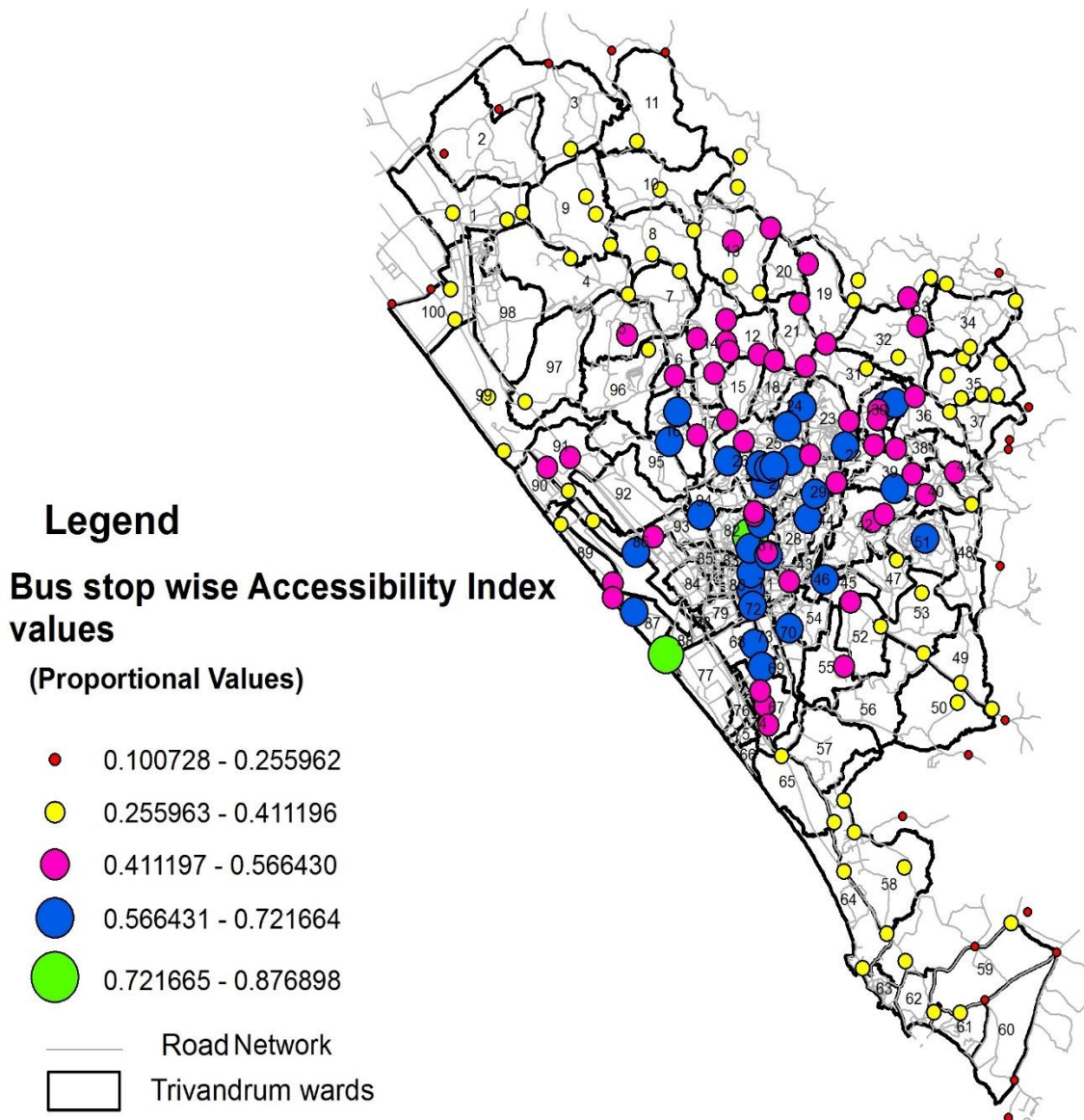


Figure 5.1. Accessibility Index Map of Trivandrum Municipal Corporation

Moreover, this accessibility index helps in identifying the extent of remoteness in the city of Trivandrum. The proposed public transport accessibility index takes into account the network-based distance to the bus stops, population, travel time and opportunities or the land use destinations. It was observed that the accessibility index in the central

area has a higher value and the accessibility values decrease to the periphery of the city. It can be interpreted that the availability of opportunities is higher in the central area, while the number of opportunities is less in other areas. The density of the road network is more towards the city centre, which contributes to higher accessibility values. Also, given that the concentration of population is high in the city centre and the higher number of bus stops, the accessibility index tends to be more in these areas. This method is relatively cost-effective and fast to use and can be used for transport and land use planning exercises, giving emphasis on public transport. The accessibility index proposed in the study uses a few basic calculations and fundamental commands of GIS. The visual representation of the composite public transport accessibility helps in identifying the extend remoteness in Trivandrum city. Thus, it is useful in the route planning by introducing new buses in the high-density zones, add missing links in the network and to increase the accessibility in the city outskirts.

The GIS-based public transport accessibility index gives the accessibility of a single point (*i.e.*, the bus stop) by considering access to the point, access from the point to various land use destinations and network impedance. Since Trivandrum city has a predominant share of the public transportation system, it is required for the planners to evaluate the public transport accessibility and ensure moderate accessibility. Given the ease of calculating and updating the accessibility index and its operationalisation, overall, this methodology has the potential to become a useful decision support tool for transit operators and planners.

5.4. PERFORMANCE OPTIMISATION OF KERALA SRTC

Application of integrated Analytical Hierarchy Process (AHP)-Goal Programming (GP) methodology requires to solve two problems: AHP and the linear goal program. The main goal of the study is to optimise the performance of Kerala SRTC by considering both the stakeholders: the operator and the user. The problem consists of prioritising the operator and user first and then prioritise the decision variables using AHP. Applying the AHP involves setting the hierarchy of decision process, as shown in Figure 5.2.

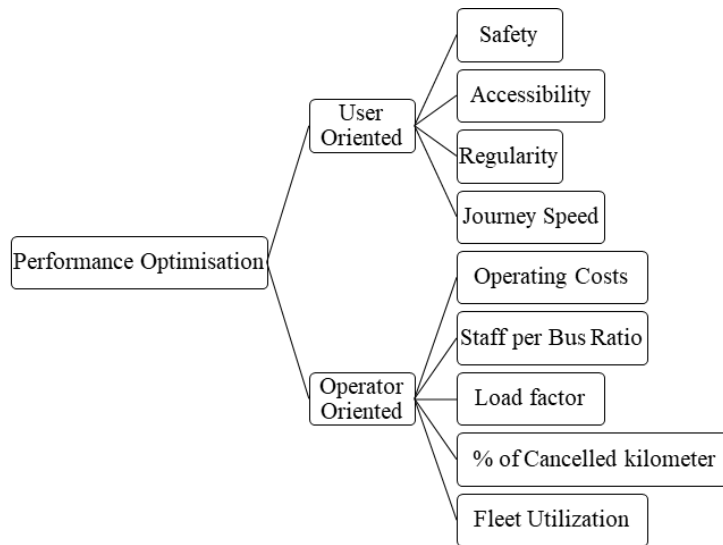


Figure 5.2 Hierarchy of Performance Optimisation Process

The elements- user and operator are further broken down into various factors or decision variables. Twelve decision variables are identified taking into account both user and operator perception: controllable costs (X_1), non-controllable costs (X_2), taxes (X_3), staff per bus ratio (fleet operated) (X_4), safety (X_5), accessibility (X_6), regularity (X_7), load factor (X_8), fleet utilisation (X_9), percentage of dead kilometre to effective kilometre (X_{10}), journey speed (X_{11}), percentage of cancelled kilometre to scheduled kilometre (X_{12}). Safety, accessibility, reliability and journey speed corresponds to user perception, and all other factors are operator oriented. Once the hierarchy is established, the pairwise comparison in each level was performed to determine the weights to be used in the objective function of the goal program. The comparison matrices are:

- Level 1: Goal of the study: optimise the performance of Kerala SRTC
- Level 2: The stakeholders of the public transport industry (Kerala SRTC)- user and operator: 2 x 2 matrix
- Level 3: Decision variables: 7 x 7 matrix for the operator and 4x4 matrix for user

The judgement matrix of pairwise comparison of level 2 elements (the stakeholders of Kerala SRTC) based on AHP rating scale was given by experts associated with bus transportation and academia in the designed pro forma. Table 5.9 gives the pairwise comparison of the two elements- operator and user.

Table 5.9. Pairwise Comparison Matrix of the Stakeholders

Factors	Operator	User	Priority Vector
Operator	1	2	0.667
User	1/2	1	0.333
Sum	1.5	3	1.000

Further, the decision variables are divided into user-oriented and operator oriented; they are analysed for determining their priorities. Table 5.10 and 5.11 show a sample pairwise comparison matrix of user-oriented decision variables and operator-oriented decision variables, respectively.

Table 5.10. Pairwise Comparison Matrix of the User-Oriented Decision Variables

Decision Variables	Regularity	Safety	Accessibility	Journey Speed / Travel Time	Priority Vector
Regularity	1	1/9	3	1	0.11
Safety	9	1	9	9	0.72
Accessibility	1/3	1/9	1	1/4	0.05
Journey Speed / Travel Time	1	1/9	4	1	0.12

Consistency Ratio (CR): 0.08

Table 5.11. Pairwise Comparison Matrix of the Operator-Oriented Decision Variables

Decision Variables	Operating Cost	Staff per bus ratio	Load Factor	% of cancelled kilometre to scheduled kilometre	Fleet Utilisation	% of dead kilometre to effective kilometre	Priority Vector
Operating Cost	1	2	4	2	3	3	0.339
Staff per bus ratio	1/2	1	3	4	4	4	0.287
Load Factor	1/4	1/3	1	3	2	3	0.143
% of cancelled kilometre to scheduled kilometre	1/2	1/4	1/3	1	2	3	0.104
Fleet Utilisation	1/3	1/4	1/2	1/2	1	2	0.071
% of dead kilometre to effective kilometre	1/3	1/4	1/3	1/3	1/2	1	0.056

Consistency Ratio (CR): 0.09

The pairwise comparison of level 2 elements (the stakeholders of Kerala SRTC) based on AHP rating scale was given by experts associated with bus transportation and

academia in the designed pro forma. The overall weights for user and operator were obtained as 33% and 67% respectively. Further, the decision variables divided into user-oriented and operator oriented were analysed for determining their priorities. Taking the overall weight of the operator as 67% and that of the user as 33%, the final weights for the decision variables are obtained as given in Table 5.12.

Table 5.12. Final Weights of Decision Variables

Decision Variables	Notation	Average Priority Vector	Weights(%)
Operator: 67%			
Operating cost	X ₁	0.411	27.389
Staff per schedule	X ₄	0.252	16.771
Load factor	X ₈	0.147	9.803
Fleet utilisation	X ₉	0.058	3.870
% of dead to effective kilometre	X ₁₀	0.044	2.937
% of cancelled to scheduled kilometre	X ₁₂	0.088	5.897
User: 33%			
Safety	X ₅	0.637	21.222
Accessibility	X ₆	0.130	4.335
Regularity	X ₇	0.102	3.409
Journey Speed/travel time	X ₁₁	0.131	4.367

Once the weights for the decision variables are obtained from the AHP method, we proceed to the optimisation process using goal programming. The threshold values for the decision variables for GP optimisation are obtained from the performance annual report of Kerala SRTC. The values (both actual and goal value) given in Table 5.13. are considered for the performance optimisation using the proposed methodology.

Table 5.13. Threshold Values for Decision Variables

Variable	Parameter	Actual Value	Goal value
X ₁	Controllable costs	3973	3377-3400
X ₂	Non-controllable costs	654.42	654.42
X ₃	Taxes	159.02	159.02
X ₄	Staff per bus ratio (fleet operated)	8.85	4.5-5
X ₅	Safety (%)	80.09	99-99.99
X ₆	Accessibility (%)	42.5	60-75
X ₇	Regularity (%)	80.5	90-95
X ₈	Load Factor (%)	80.09	115-125
X ₉	Fleet utilisation (%)	80.99	90-95
X ₁₀	% of dead kilometre to effective kilometre	6.9	04-05
X ₁₁	Journey speed (kmph)	25	30-35
X ₁₂	% of cancelled kilometre to scheduled kilometre	20.68	05-08

The aim is to minimise the sum of deviations so that the values obtained from the optimisation gives the values close to the goal values. The formulated GP model is as follows.

Objective Function:

$$\begin{aligned} \text{Min } & 27.39n_1 + 27.39p_1 + 27.39n_2 + 27.39p_2 + 27.39n_3 + 27.39n_4 + 21.22n_5 + \\ & 21.22p_5 + 16.77n_6 + 16.77p_6 + 10.32n_7 + 10.32p_7 + 9.8n_8 + 9.8p_8 + 9.66n_9 + \\ & 9.66p_9 + 7.85n_{10} + 7.85p_{10} + 6.84n_{11} + 6.84p_{11} + 5.9n_{12} + 5.9p_{12} + 4.65n_{13} + \\ & 4.65p_{13} + 4.37n_{14} + 4.37p_{14} + 4.34n_{15} + 4.34p_{15} + 3.87n_{16} + 3.87p_{16} + \\ & 3.64n_{17} + 3.64p_{17} + 3.41n_{18} + 3.41p_{18} + 2.94n_{19} + 2.94p_{19} \end{aligned} \quad (5.1)$$

Constraints:

Priority 1: Optimise the financial resources adequately

$$X_1 + X_2 + X_3 - n_1 = 4190.44 \quad (5.2)$$

$$X_1 + X_2 + X_3 + p_1 = 4214 \quad (5.3)$$

$$X_1 - n_2 = 3377 \quad (5.4)$$

$$X_1 + p_2 = 3400 \quad (5.5)$$

$$X_2 - n_3 = 654.42 \quad (5.6)$$

$$X_3 - n_4 = 159.02 \quad (5.7)$$

Priority 2: Maximise the safety of the users

$$X_5 - n_5 = 99 \quad (5.8)$$

$$X_5 + p_5 = 99.99 \quad (5.9)$$

Priority 3: Reduce the staff per bus ratio

$$X_4 - n_6 = 4.5 \quad (5.10)$$

$$X_4 + p_6 = 5 \quad (5.11)$$

Priority 4: Maximise the productivity by decreasing staff per bus ratio and increasing fleet utilisation

$$X_4 + X_9 - n_7 = 94.5 \quad (5.12)$$

$$X_4 + X_9 + p_7 = 100 \quad (5.13)$$

Priority 5: Maximise the utilisation of public transport by increasing load factor

$$X_8 - n_8 = 115 \quad (5.14)$$

$$X_8 + p_8 = 125 \quad (5.15)$$

Priority 6: Improve Safety, accessibility, and regularity

$$0.33X_5 + 0.33X_6 + 0.33X_7 - n_9 = 83 \quad (5.16)$$

$$0.33X_5 + 0.33X_6 + 0.33X_7 + p_9 = 90 \quad (5.17)$$

Priority 7: Increase the load factor and reduce % of cancelled kilometre to scheduled kilometre

$$X_8 + X_{12} - n_{10} = 120 \quad (5.18)$$

$$X_8 + X_{12} + p_{10} = 133 \quad (5.19)$$

Priority 8: Increase load factor and fleet utilisation

$$X_8 + X_9 - n_{11} = 205 \quad (5.20)$$

$$X_8 + X_9 + p_{11} = 220 \quad (5.21)$$

Priority 9: Improve regularity by reducing % of cancelled kilometre to scheduled kilometre

$$X_7 + X_{12} - n_{12} = 95 \quad (5.22)$$

$$X_7 + X_{12} + p_{12} = 103 \quad (5.23)$$

Priority 10: Reduce % of cancelled kilometre to scheduled kilometre

$$X_{12} - n_{13} = 5 \quad (5.24)$$

$$X_{12} + p_{13} = 8 \quad (5.25)$$

Priority 11: Improve journey speed

$$X_{11} - n_{14} = 30 \quad (5.26)$$

$$X_{11} + p_{14} = 35 \quad (5.27)$$

Priority 12: Increase accessibility

$$X_6 - n_{15} = 60 \quad (5.28)$$

$$X_6 + p_{15} = 75 \quad (5.29)$$

Priority 13: Maximise fleet utilisation

$$X_9 - n_{16} = 90 \quad (5.30)$$

$$X_9 + p_{16} = 95 \quad (5.31)$$

Priority 14: Improve accessibility and reduce % of dead kilometre to effective kilometre

$$X_6 + X_{10} - n_{17} = 64 \quad (5.32)$$

$$X_6 + X_{10} + p_{17} = 80 \quad (5.33)$$

Priority 15: Improve regularity

$$X_7 - n_{18} = 90 \quad (5.34)$$

$$X_7 + p_{18} = 95 \quad (5.35)$$

Priority 16: Minimise % of dead to effective kilometre

$$X_{10} - n_{19} = 4 \quad (5.36)$$

$$X_{10} + p_{19} = 5 \quad (5.37)$$

Where, X_1, X_2, \dots, X_{12} are the various decision variables considered; n and p are negative and positive deviations, respectively.

5.4.1. Optimal Solution and Sensitivity Analysis

To get the optimal values, the linear weighted integrated AHP-GP problem was solved using LINGO 17.2. The optimal values of the decision variables, along with actual values, are presented in Table 5.14.

The optimal solution indicates that decision variables having higher penalty values tend more towards goal values than variables having lower penalties. The optimal value of operator cost, staff per schedule and safety having higher penalties attached to it are closer to the goal values, while % of dead km to effective km, journey speed etc. that have low penalties attached to it have optimal values a little farther from goal values. From Table 5.14, it can be noted that an increase of 43.59% of load factor along with 11.12% of fleet utilisation and a reduction of 42.03% of the percentage of dead km to effective km helps in achieving the goal values with optimal value for the objective function. Reducing the controllable costs by 15%, staff per bus ratio by 49.15% and the percentage of cancelled kilometre to scheduled kilometre by 75.82% can improve the optimal value of the objective function. The model solution suggests

decreasing the staff per bus that will further reduce the staff cost and hence the operating cost. This can be achieved by implementing a single duty system, i.e., eight-hour duty in a single shift. Thus only one driver and conductor is required in an intra-city bus. Also, the report of sub-group on SRTU's (Planning Commission 2012) suggests that SRTU's should employ both mechanical and operational staff on contractual basis to reduce the staff cost. The bus-staff ratio can be reduced by introducing technological improvements like Intelligent Transport System (ITS), computerized staff scheduling, passenger information system, Electronic Ticketing Machine (ETM), Global Positioning System (GPS) based ETM and Automated Ticketing System. The staff per bus is reduced to the national average of 4.5 from 8.85. by reducing the staff per bus from 8.85 to the national average of 4.5, safety of the passengers will not be affected by the staff reduction. In view of the user, increasing the accessibility by 44.76%, safety by 24.85% and improvement in regularity by 11.80% attracts passengers to the public transport that in turn improves the load factor and thus enhance the performance of Kerala SRTC.

Table 5.14. Optimal and Actual Values for Decision Variables

Variable	Parameter	Actual Value	Optimal Value	Percentage increase/decrease
X ₁	Controllable costs	3973	3377	-15.00
X ₂	Non-controllable costs	654.42	654.42	0
X ₃	Taxes	159.02	159.02	0
X ₄	Staff per bus ratio (fleet operated)	8.85	4.5	-49.15
X ₅	Safety (%)	80.09	99.99	24.85
X ₆	Accessibility (%)	42.5	61.525	44.76
X ₇	Regularity (%)	80.5	90	11.80
X ₈	Load Factor (%)	80.09	115	43.59
X ₉	Fleet utilisation (%)	80.99	90	11.12
X ₁₀	% of dead kilometre to effective kilometre	6.9	4	-42.03
X ₁₁	Journey speed (kmph)	25	30	20
X ₁₂	% of cancelled km to scheduled km	20.68	5	-75.82

The data used for goal programming may be subject to change, and therefore, it is essential to check the sensitivity of the optimal solution to variations in the data. The sensitivity analysis identifies the impact of variations in the decision variables has on the performance of the system. The values of the optimal solution and degree of goal

attainment under different combinations of input are listed in Table 5.15. From the values obtained, the best condition is with: controllable costs: ₹3400 (in lakh), non-controllable costs: ₹654.42 (in lakh), taxes: ₹159.02 (in lakh), staff per bus ratio (fleet operated): 4.5, safety: 99%, accessibility: 60%, regularity: 95%, load factor: 115%, fleet utilisation: 90%, percentage of dead kilometre to effective kilometre: 4, journey speed: 35kmph, and percentage of cancelled kilometre to scheduled kilometre: 5%. The worst condition is with controllable costs: ₹3377(in lakh), non-controllable costs: ₹654.42 (in lakh), taxes: ₹159.02 (in lakh), staff per bus ratio (fleet operated): 4.5, safety: 95%, accessibility: 61.525%, regularity: 95%, load factor: 115%, fleet utilisation: 95%, percentage of dead kilometre to effective kilometre: 4, journey speed: 30kmph, and percentage of cancelled kilometre to scheduled kilometre: 5.

A weighted goal programming methodology integrated with analytical hierarchy process was used for performance optimisation of Kerala SRTC. The following are the findings from the study.

- The operator cost and staff per schedule were the most important variables in regard to the operator, while the safety of travel with respect to user perception has the highest weightage.
- The optimal solution indicates that increasing the accessibility, safety and regularity attract passengers to the public transport that in turn improves the load factor which influences the operators to maximise the fleet utilisation and reduce the cancellation of schedules.
- The proposed model solution suggests decreasing the staff per bus that will further reduce the staff cost and hence the operating cost.

The proposed method can be adopted for solving the performance optimisation problems of public transport undertakings. With AHP, the planners can consistently integrate the opinions of all the decision makers and effectively evaluate the scenarios and optimise the resources. Further works include the application of the proposed methodology to the various other fields of public transport like urban rail transit, metro rail system, and integrated public transport system.

Table 5.15. Sensitivity Analysis

Decision Variable	Actual Values	Optimal Values	$X_4 \geq 4.8,$ $X_7 = 93$	$X_1 = 3400,$ $X_7 = 95,$ $X_{11} = 35$	$X_5 = 95$	$X_4 \geq 4.8,$ $X_{12} \geq 6$	$X_6 = 65$	$X_8 = 120,$ $X_{10} = 4.5$	$X_6 \geq 70,$ $X_8 = 117,$ $X_9 = 92$	$X_5 = 99,$ $X_{10} = 5$	$X_7 = 92$
Objective Function Value		1971.131	1970.9	1302.26	1973.63	1965	1871.12	1848.64	1936.22	1969	1895
X_1	3973	3377 (100%)	3377 (100%)	3400 (100%)	3377 (100%)	3377 (100%)	3377 (100%)	3377 (100%)	3377 (100%)	3377 (100%)	3377 (100%)
X_2	654.42	654.42 (100%)	654.42 (100%)	654.42 (100%)	654.42 (100%)	654.42 (100%)	654.42 (100%)	654.42 (100%)	654.42 (100%)	654.42 (100%)	654.42 (100%)
X_3	159.02	159.02 (100%)	159.02 (100%)	159.02 (100%)	159.02 (100%)	159.02 (100%)	159.02 (100%)	159.02 (100%)	159.02 (100%)	159.02 (100%)	159.02 (100%)
X_4	8.85	4.5 (100%)	4.8 (93.33%)	4.5 (100%)	4.5 (100%)	4.8 (93.33%)	4.5 (100%)	4.5 (100%)	4.5 (100%)	4.5 (100%)	4.5 (100%)
X_5	80.09	99.99 (99%)	99 (100%)	99 (100%)	95 (94.73%)	99 (100%)	99 (100%)	99 (100%)	99 (100%)	99 (100%)	99.515 (99.48%)
X_6	42.5	61.525 (97.46%)	60 (100%)	60 (100%)	61.525 (97.46%)	60 (100%)	65 (91.67%)	70 (83.33%)	70 (83.33%)	62.515 (95.81%)	60 (100%)
X_7	80.5	90 (100%)	93 (96.67%)	95 (100%)	90 (100%)	93 (96.67%)	90 (100%)	90 (100%)	90 (100%)	90 (100%)	92 (97.78%)
X_8	80.09	115 (100%)	115 (100%)	115 (100%)	115 (100%)	115 (100%)	115 (100%)	120 (95.65%)	117 (98.26%)	115 (100%)	115 (100%)
X_9	80.99	90 (100%)	90 (100%)	90 (100%)	95 (100%)	90 (100%)	90 (100%)	90 (100%)	92 (97.78%)	90 (100%)	90 (100%)
X_{10}	6.9	4 (100%)	4 (100%)	4 (100%)	4 (100%)	4 (100%)	4 (100%)	4.5 (87.5%)	4 (100%)	5 (100%)	4 (100%)
X_{11}	25	30 (100%)	30 (100%)	35 (100%)	30 (100%)	30 (100%)	30 (100%)	30 (100%)	30 (100%)	30 (100%)	30 (100%)
X_{12}	20.68	5 (100%)	5 (100%)	5 (100%)	5 (100%)	6 (80%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)	5 (100%)

Note: Values in the bracket indicate the degree of goal attainment with different input values.

5.5. VARIATION INDICES

The demand for public transport is subjected to seasonal variations and fluctuations during different points of time, such as day, week, month, quarter and year. The useful indication of the seasonal variation is the seasonal index. The seasonal variation index relates the average demand in a particular period to the average demand of all the periods as given in Equation 5.38 (Pillai 2008).

$$\text{Seasonal Index} = \frac{\text{Average demand during the seasonal period}}{\text{Average demand for all the periods}} \quad (5.38)$$

The daily, weekly, monthly and quarterly variation indices are determined for all the routes and depots of Trivandrum and Mangalore.

5.5.1. Variation Indices of Intra-City Passenger Demand

The daily, weekly, monthly and quarterly variation indices for intra-city (Trivandrum division Kerala SRTC) passenger demand is determined. The daily index, quarterly and monthly variation indices of Trivandrum depots are given in Table 5.16. , 5.17. and 5.18., respectively. Figure 5.3., 5.4., 5.5. and 5.6. shows the variation in daily, quarterly, monthly and weekly index respectively. The weekly variation indices are given in Appendix A.17.

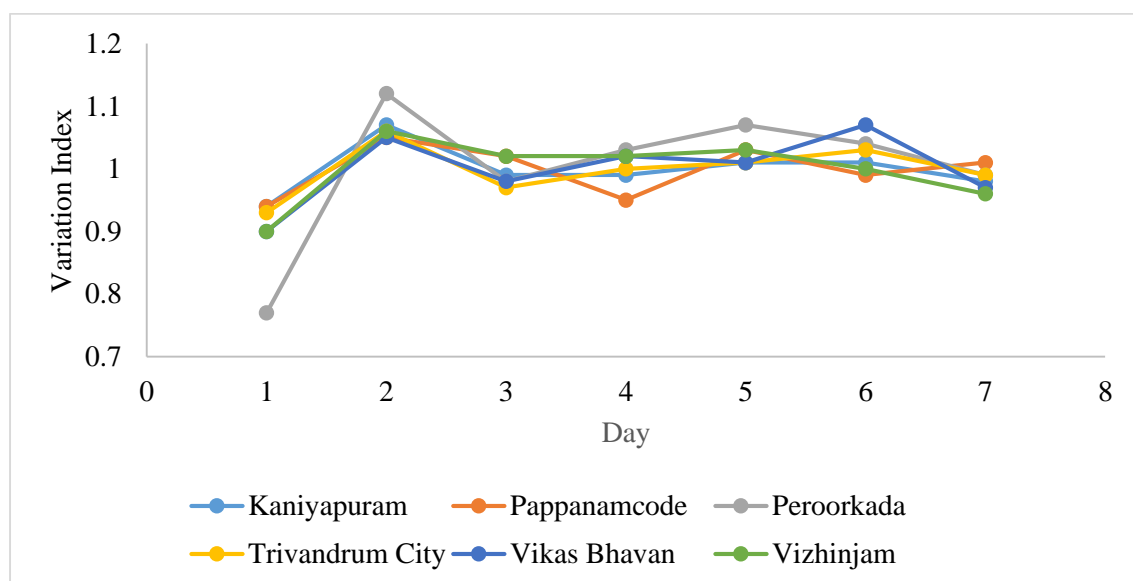


Figure 5.3. Daily Variation Index of Six Depots of Trivandrum

Table 5.16. Daily Variation Index of the Six Depots of Trivandrum

Day	Kaniyapuram	Pappanamcode	Peroorkada	Trivandrum City	Vikas Bhavan	Vizhinjam	Average Index
Sunday	0.94	0.94	0.77	0.93	0.90	0.90	0.90
Monday	1.07	1.05	1.12	1.06	1.05	1.06	1.07
Tuesday	0.99	1.02	0.98	0.97	0.98	1.02	0.99
Wednesday	0.99	0.95	1.03	1.00	1.02	1.02	1.00
Thursday	1.01	1.03	1.07	1.01	1.01	1.03	1.03
Friday	1.01	0.99	1.04	1.03	1.07	1.00	1.02
Saturday	0.98	1.01	0.99	0.99	0.97	0.96	0.98

Table 5.17. Quarterly Variation Index of the Six Depots of Trivandrum

Quarter	Kaniyapuram	Pappanamcode	Peroorkada	Trivandrum City	Vikas Bhavan	Vizhinjam	Average Index
Q1	1.02	0.98	0.93	1.03	0.89	1.08	0.99
Q2	1.02	1.03	0.98	1.02	0.98	1.03	1.01
Q3	1.02	1.10	1.15	0.97	1.01	0.95	1.03
Q4	0.94	0.88	1.01	0.97	1.11	0.94	0.98

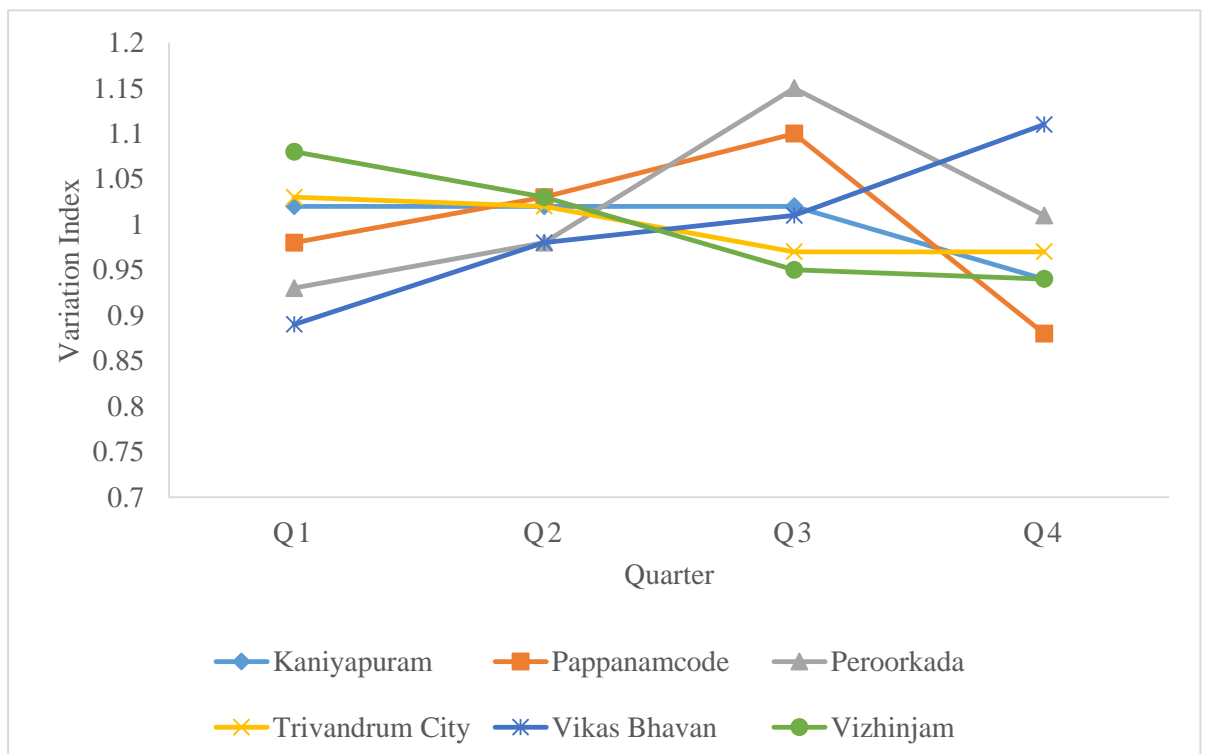


Figure 5.4. Quarterly Variation Index of Trivandrum

Table 5.18. Monthly Variation Index of Trivandrum

Month	Kaniyapuram	Pappanamcode	Peroorkada	Trivandrum City	Vikas Bhavan	Vizhinjam
1	0.71	0.96	1.08	1.07	1.02	1.08
2	0.96	0.95	1.00	1.18	1.07	1.10
3	1.11	0.86	0.87	1.12	0.81	1.10
4	1.10	0.92	0.88	0.95	0.83	1.09
5	1.13	0.77	0.86	1.02	1.08	1.09
6	1.10	1.24	1.18	1.04	0.98	1.04
7	1.07	1.31	1.18	1.00	1.05	0.97
8	1.04	1.17	1.05	0.90	1.00	0.90
9	1.05	1.10	1.03	0.88	0.93	0.92
10	1.00	1.09	1.02	1.04	1.18	0.93
11	0.90	0.88	0.96	0.93	1.06	0.84
12	0.82	0.73	0.89	0.88	0.99	0.97

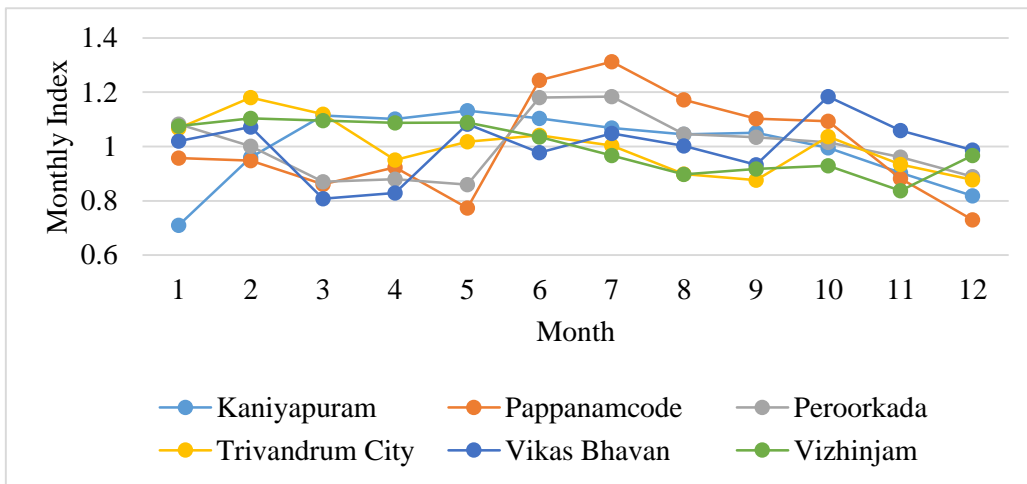


Figure 5.5. Monthly Variation Index of Trivandrum

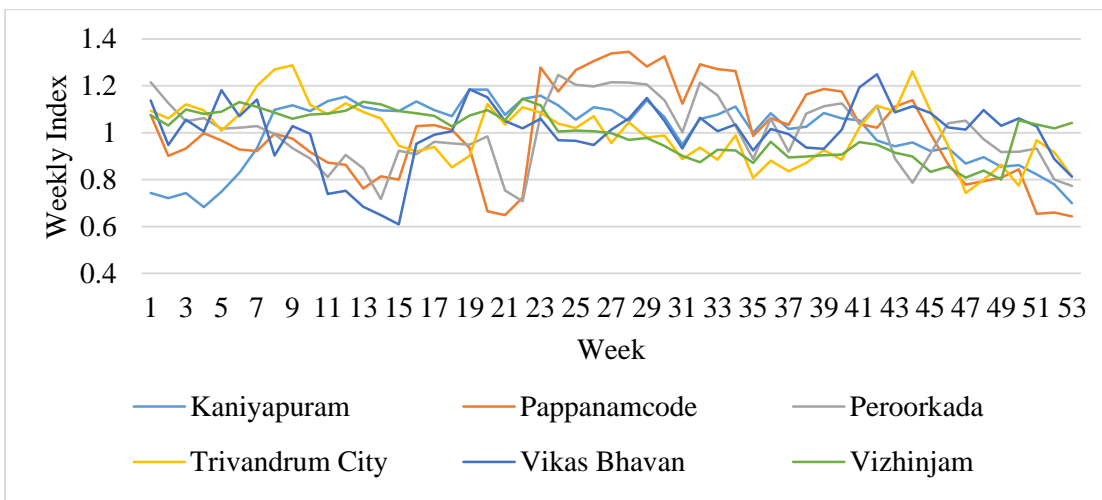


Figure 5.6. Weekly Variation Index of Trivandrum

The daily variation indices of all the six depots are as given in Table 5.16 and Figure 5.3. It can be observed that Sunday corresponds to demand below the average demand while the variation index for all other days remains almost the same with Monday and Friday been slightly on a higher side. Figure 5.4. shows the quarterly variation index. It can be seen that no pattern is followed by any of the depots. The monthly variation index given in Figure 5.5. and Table 5.18. shows a clear pattern in the demand. It can be seen that during the school working months, the demand is high as against the vacation days. The weekly variation index given in Figure 5.6. does not show any significant pattern in the demand.

5.5.2. Variation Indices of Inter-City Passenger Demand

The daily variation index of the inter-city passenger data for the year from 2013 to 2018 is shown in Figure 5.7. It can be seen that the variation is having a regular pattern of fluctuation with Sunday being the lowest. Figure 5.8 shows the quarterly variation index of Mangalore passenger demand. The third quarter shows a peak in travel with the second quarter being the lowest. The monthly variation index of passenger demand reveals that the demand is almost constant throughout the various months. But it can be seen that there is a sudden increase in the demand in the month of September due to festivals and good climate. Figure 5.10 shows the weekly variation index of Mangalore, the first week in the Figure is the first week of January. It can be seen that the weekly variations are roughly constant with a higher value for the weeks of September.

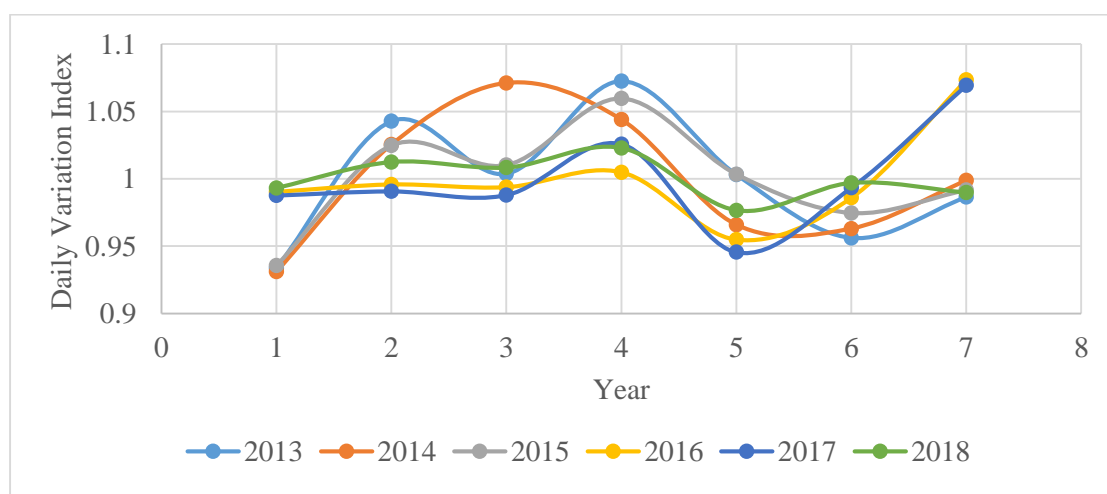


Figure 5.7. Year-wise Daily Variation Index of Mangalore

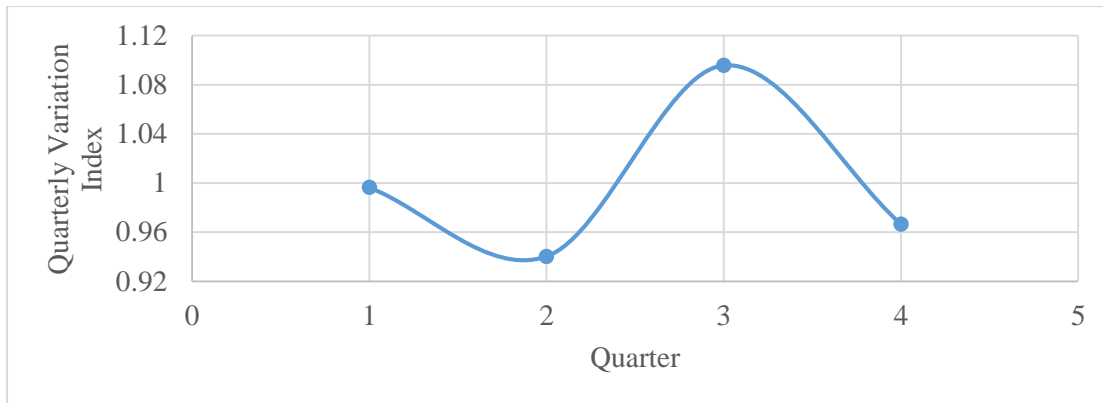


Figure 5.8. Quarterly Variation Index of Mangalore

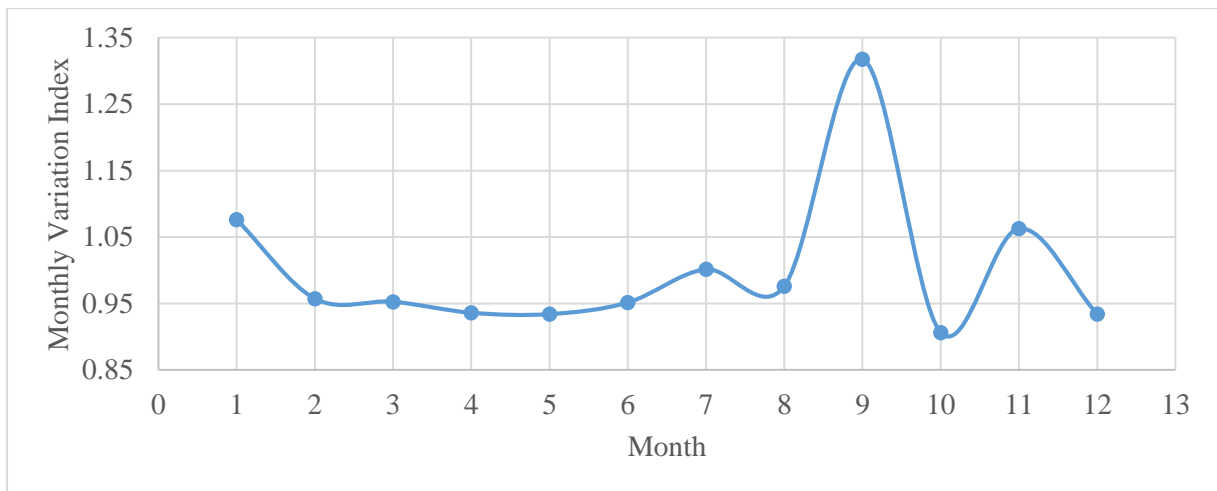


Figure 5.9. Monthly Variation Index of Mangalore

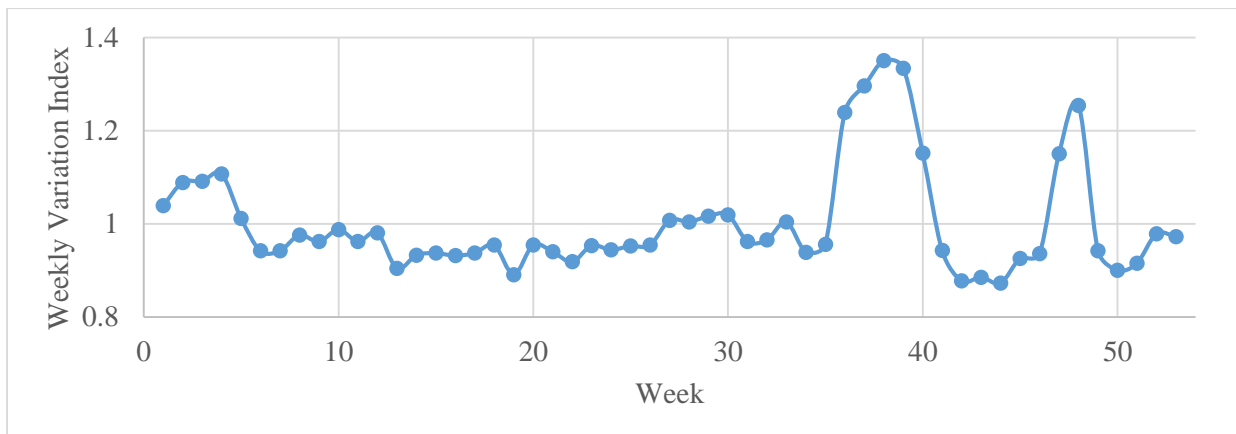


Figure 5.10. Weekly Variation Index of Mangalore

The variation indices of both Trivandrum and Mangalore revealed a few interesting features of these study areas. The data from Trivandrum is intra-city bus passenger demand data, while that of Mangalore was inter-city demand. The daily variation index

of Trivandrum shows that the demand the Monday's corresponds to more travel with Sunday being the lowest. In the case of Mangalore, the daily variation is having a regular pattern of fluctuation with Sunday being the lowest. This suggests that the demand for travel during the weekend is less compared to weekdays. The quarterly variation index of Trivandrum was almost constant without any pattern revealing that the demand for intra-city travel remains unaltered due to quarter changes. But in the inter-city Mangalore data, third quarter shows a rise in the demand. This due to the fact that the third quarter contains festivals in the state urging the people to travel to home town and visiting relatives. Also, the weather in the third quarter is good (not so hot and not rainy). The monthly variation index of Trivandrum shows that the school working month have slightly more demand compared to vacations. This slight change is mainly because of school going children using the Kerala SRTC buses during school days. The monthly pattern of travel in Mangalore is similar to the quarterly pattern of Mangalore with September on the higher side. This is due to the festival season in the state. The weekly travel variation is not predominant in the case of Trivandrum. But in case of Mangalore, it can be seen that there is a steady rise in travel from week 36 to week 40 due to festivities in the state (Week 01 is the first week of January).

5.6. TIME-SERIES ANALYSIS AND MODELLING FOR INTRA-CITY DATA

Time-series technique was applied to develop passenger demand forecasting model using ETM based bus passenger data for intra-city (Trivandrum City) and inter-city (Mangalore). The forecast accuracy was calculated for the models developed. The analysis was carried out both at route level and at the depot level. This section explains the result summary and discussion for a sample route of intra-city. The steps involved in time-series analysis are discussed in the following subsection.

5.6.1. Exploratory Data Analysis

The sample route selected for intra-city passenger demand modelling is from Trivandrum is route number 137 of Vizhinjam depot. The route starts from Vizhinjam depot and terminates in the Trivandrum city. The data is for a period of 01/01/2011 to 07/12/2013 with 1072 data points.

Time-Series Plot: Time-series models are explicitly based on the historical characteristics of data. A time-series plot is used to identify the various patterns in the data, missing observation and outliers. Figure 5.11. shows a line plot of a sample time-series data.

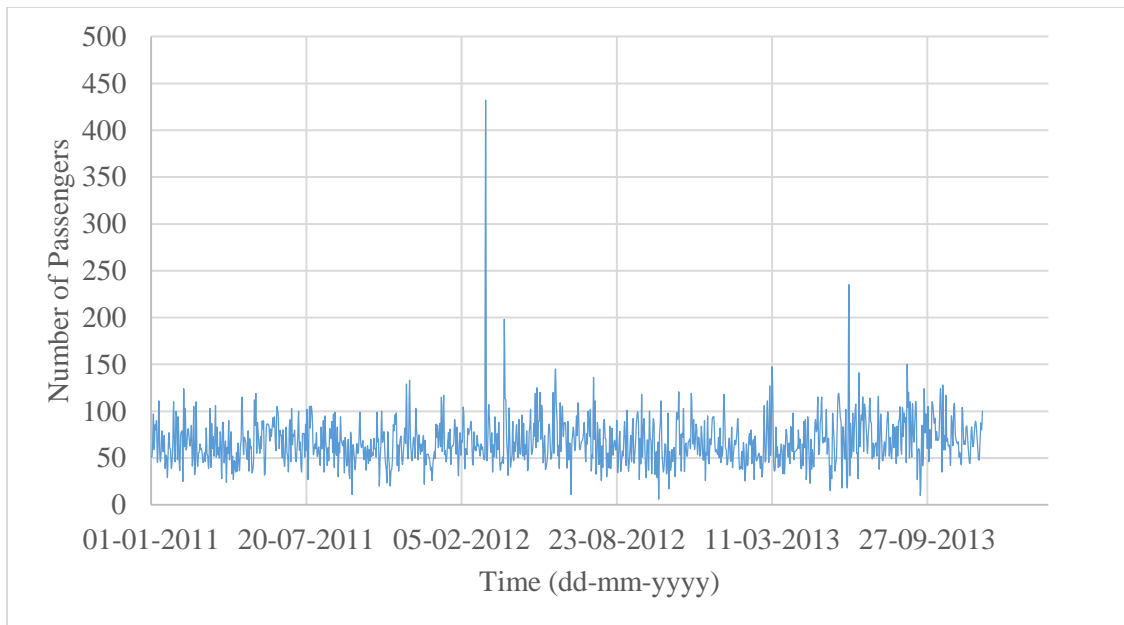


Figure 5.11 Time-Series Plot of Number of Passengers for Vizhinjam Route 137

The plot shows that there are some unusual observations at specific time points. Due to this unusual observations, the pattern and trend in the data cannot be visually identified. It is observed from the time-series plot that the data has some missing observations. Also, out of 1072 data points, 55 values were found to be missing. This account for about 5.13% of the data, which is less than 10% Bennett (2010). Hence, this data set is considered for further analysis. The missing values were replaced with values using linear interpolation technique. The results of the detailed analysis are given below.

5.6.2. Histogram Overlaid with Kernel Density Curve

To identify the distribution of the data points, a histogram overlaid with Kernel Density curve was plotted, as shown in Figure 5.12. The histogram smoothed by a kernel density curve shows that the data follows a right-skewed distribution. Therefore, it cannot be clearly stated that the data follows a normal distribution. Hence suitable

transformation (preferably Box-Cox transformation) has to be applied to transform the data into a normal distribution.

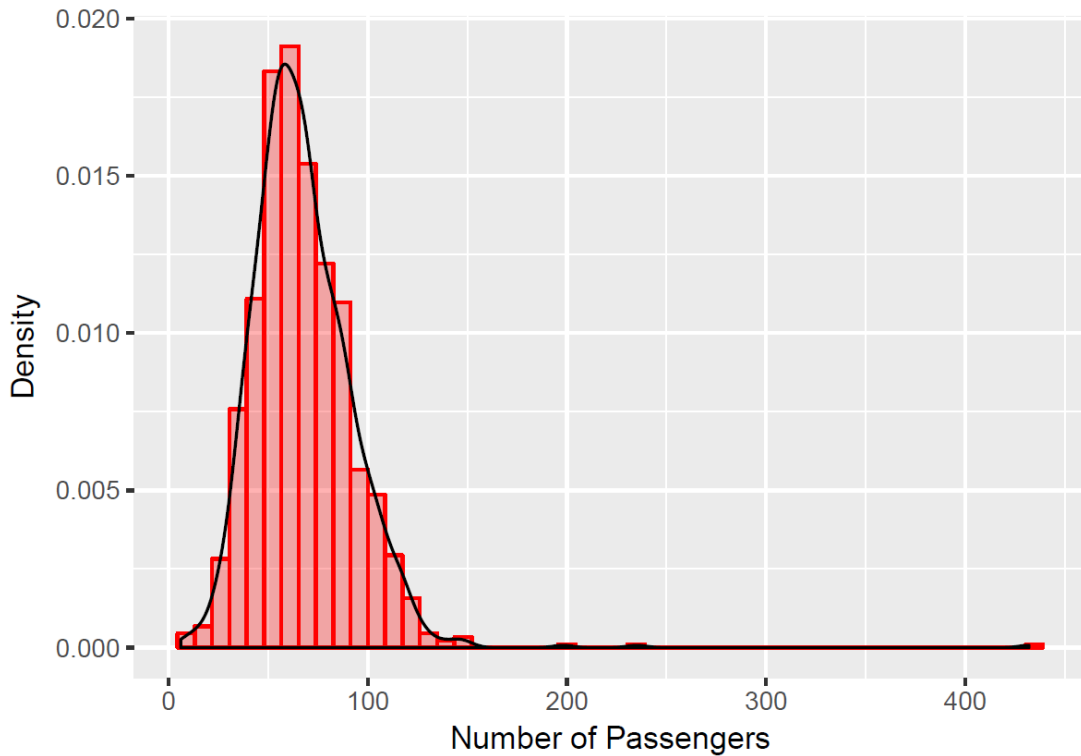


Figure 5.12. Histogram Overlaid with Kernel Density Curve

5.6.3. Outlier Analysis

Outlier detection and analysis were carried out using Chen and Liu's algorithm. In Figure 5.11, it can be observed that there are some unusually high values causing a disturbance in the overall distribution in the data. Such observations termed as outliers need to be removed. Figure 5.13 shows the location of the outlier and adjusted time-series plot after outlier removal.

Three outliers were detected in this sample data of the selected route. These outliers were removed and were replaced using the linear interpolation method. Figure 5.14 shows the time-series plot after the removal of the outliers. It can be observed that there exists seasonality in the data. The data has a slightly increasing trend but have fluctuations.

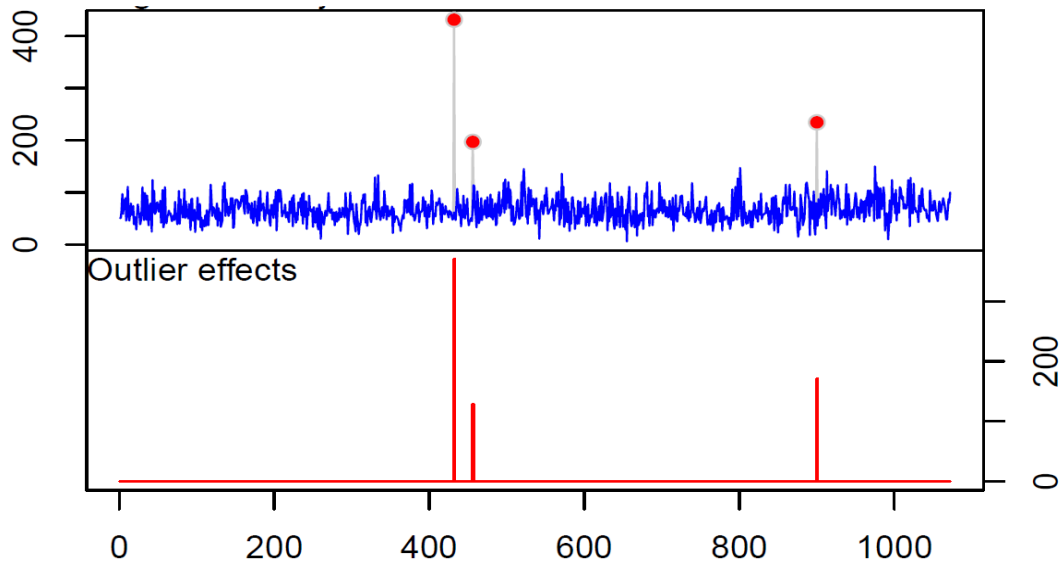


Figure 5.13. Histogram Overlaid with Kernel Density Curve

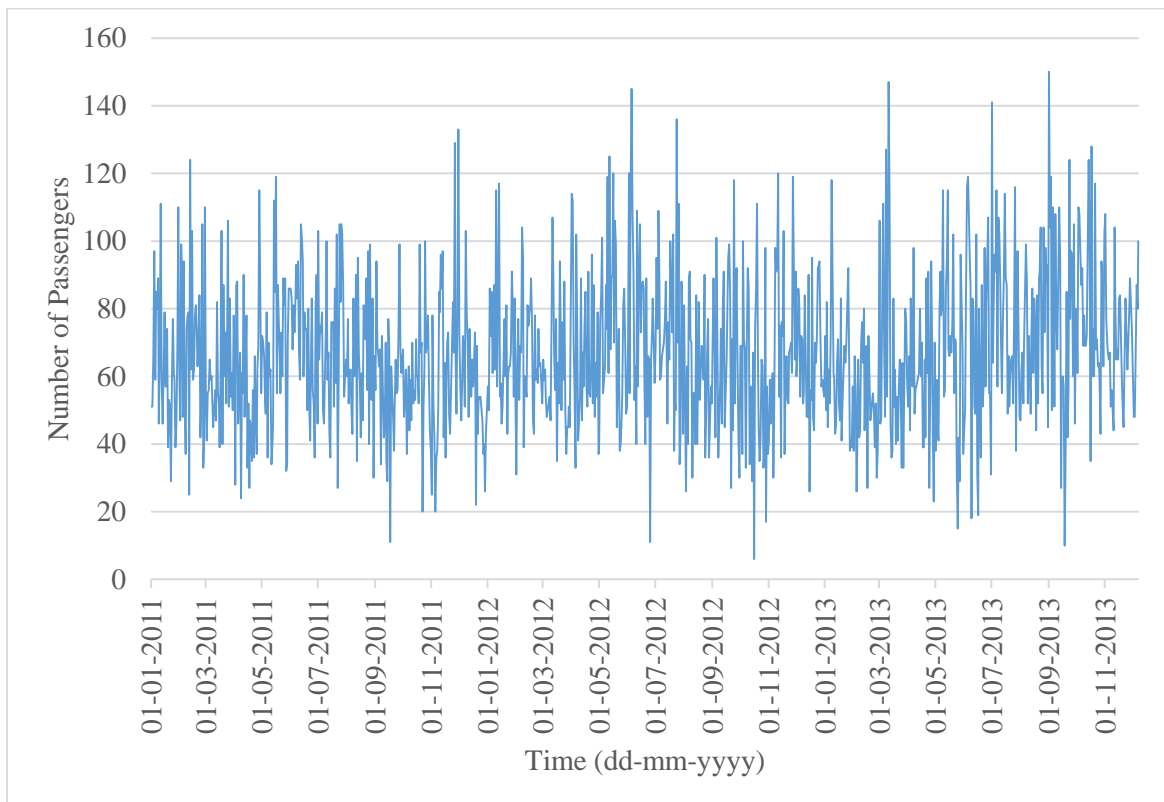


Figure 5.14. Time-Series Plot After the Removal of the Outliers

5.6.4. Decomposition of Time-Series Data

A trend decomposition was plotted as in Figure 5.15., to identify the underlying trend and irregularities. The trend decomposition line shows that the data has an overall increasing trend with non-linear variations.

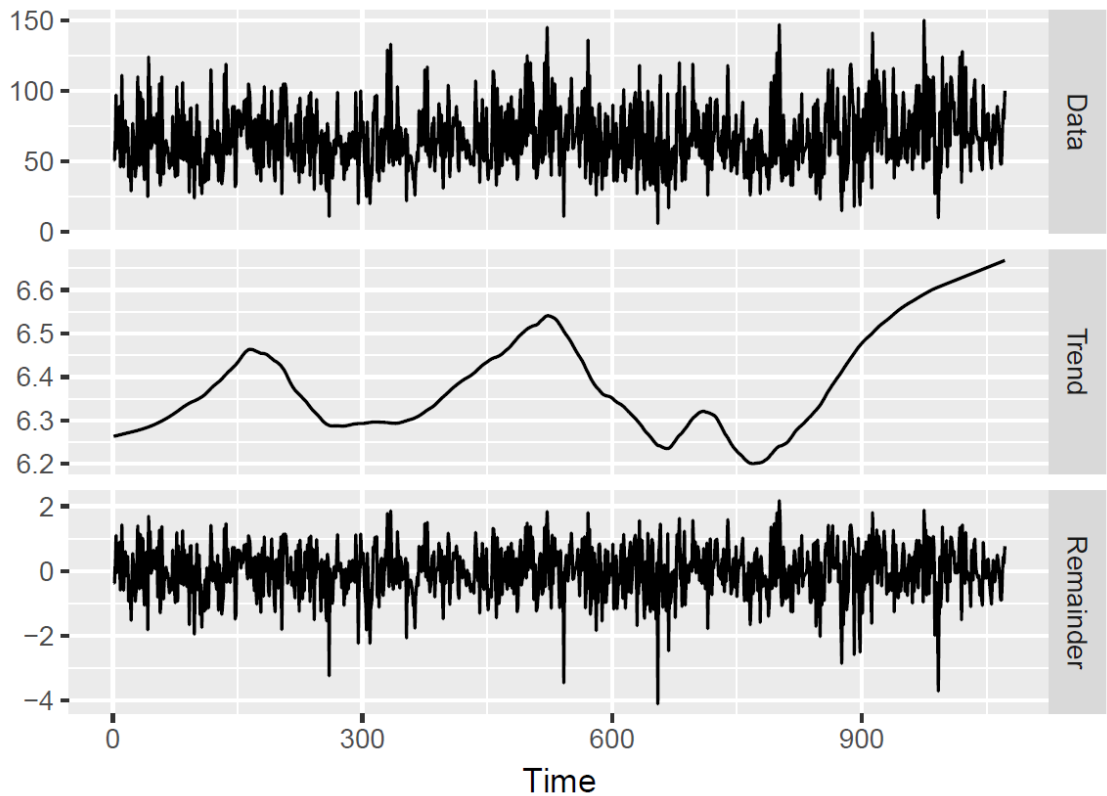


Figure 5.15. Trend Decomposition Plot

5.6.5. Autocorrelation and Partial Autocorrelation Plot

The autocorrelation plot of the time-series data in Figure 5.16 shows significant correlations upto lag four. At lag seven, a significant spike can be noted. Significant spikes are not visible at points other than the points mentioned; hence, no inference could be drawn about the seasonality and trend in the data.

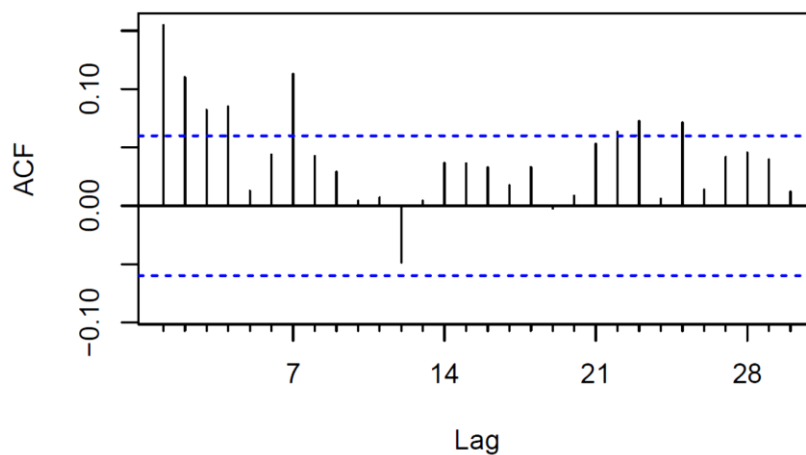


Figure 5.16. Autocorrelation Plot of Time-Series Data

The partial autocorrelation plot shown in Figure 5.17 is obtained after removing the indirect correlations in the data. The PACF plot also shows significant spikes at 1,2, and 7. Since the ACF and PACF plot shows a sudden decay. The lags in the seasonality are not consistent in the data. Hence, further analysis is required to confirm the presence of seasonality. Also, significant correlation exists in the lag seven. Hence a seasonal AR term can also be included in the model along with one or two MA terms and one non-seasonal AR term.

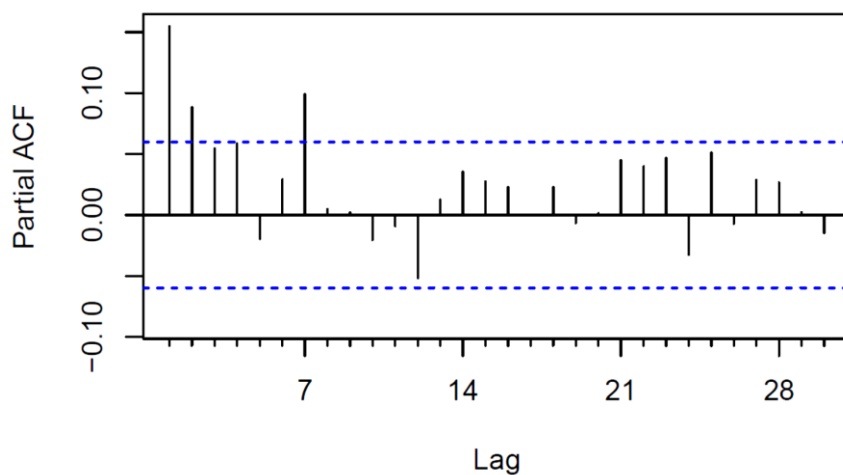


Figure 5.17. Partial Autocorrelation Plot of Time-Series Data

5.6.6. Statistical Tests

To confirm the visual interpretations of the data, statistical tests were conducted in R software (version 3.5), the results of which are tabulated Table 5.19.

Table 5.19. Results of the Statistical Tests of the Data

Statistical Tests	Hypothesis	Test statistic value / p -value
Test for Stationarity		
ADF test	H_0 : Presence of unit root	Dickey-Fuller = -8.86
	H_1 : Data is stationary	p -value = 0.01
KPSS test	H_0 : Data is stationary with respect to a deterministic trend and level	KPSS = 0.19
	H_1 : Presence of unit root (Non Stationary)	p -value = 0.02
Test for Normality		
Shapiro-Wilk Test	H_0 : Data is normally distributed	W = 0.98
	H_1 : Data is non normal	p -value = 4.13×10^{-9}

Statistical Tests	Hypothesis	Test statistic value / <i>p</i> -value
Test for Normality		
Jarque-Bera Test	H ₀ : Skewness and Kurtosis of data does not match the normal distribution	X-squared = 45.02
	H ₁ : Data is non normal	<i>p</i> -value = 1.65 x10 ⁻¹⁰
Test for Non-Linearity		
Anderson-Darling Test	H ₀ : Data is normal hence linear	A value = 4.83
	H ₁ : Data is non-linear	<i>p</i> -value = 5.99x10 ⁻¹²
Terasvirta Neural Network Test	H ₀ : Data is linear with respect to mean	X-squared = 0.28
	H ₁ : Data is non-linear in mean	<i>p</i> -value = 0.89
White Neural Network Test	H ₀ : Data is linear in mean	X-squared = 0.07
	H ₁ : Data is non-linear in mean	<i>p</i> -value = 0.97
Keenan Test	H ₀ : Data follows AR process	Keenans' Test stat = 86.51
	H ₁ : The time-series plot does not have AR element	<i>p</i> -value = 7.77 x10 ⁻²⁰
McLeod Li Test	H ₀ : Data follows the ARIMA process	-
	H ₁ : Data does not have ARIMA element	<i>p</i> -value = 7.14x10 ⁻⁷
Tsays Test statistic	H ₀ : Data follows AR process	Tsays Test stat = 1.52
	H ₁ : Data does not follow AR process	<i>p</i> -value = 0.04

The ADF test result conducted at 5% level of significance yielded a *p*-value of 0.01. Hence, the null hypothesis (presence of unit root in the data) is rejected. Therefore, the data is stationary with respect to the unit root. The KPSS trend and level test conducted yielded a *p*-value of 0.01. Hence, the null hypothesis (Data stationary with respect to trend and level) can be rejected. This means that the data is non-stationary with respect to trend and level. Also, the test for normality was conducted on the data. Jarque-Bera test and Shapiro-Wilk test yielded a *p*-value less than 0.05. Hence, the null hypothesis that the data comes from a normal distribution was rejected. The non-normality can be visually interpreted from Figure 5.12. From the linearity tests, it is concluded that the data is nonlinear with respect to mean but does not have hidden AR process.

From the data analysis, it is concluded that the data is non-stationary. Hence, it requires differencing. Also, a seasonal component was observed in the ACF plot. But strong seasonality can be neglected as the seasonal component was not visible in PACF plot. The need for seasonal differencing can be analysed from seasonally differenced ACF

and PACF plot. Also, the data is non-normal. Hence, Box-Cox transformation is required to normalise the data to perform a further modelling. Figure 5.18 represents the ACF and PACF plot of seasonally differenced (lag =7) data.

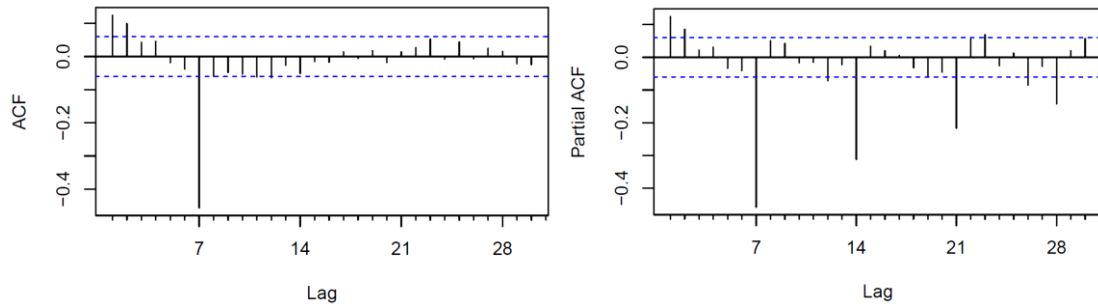


Figure 5.18. ACF and PACF Plot of Seasonally Differenced Time-Series Data

Fig 5.18 shows the ACF and PACF plot of seasonally differenced time-series data. The negative partial autocorrelation at the lag seven denotes the existence of a seasonal AutoRegressive (SAR) component in the model.

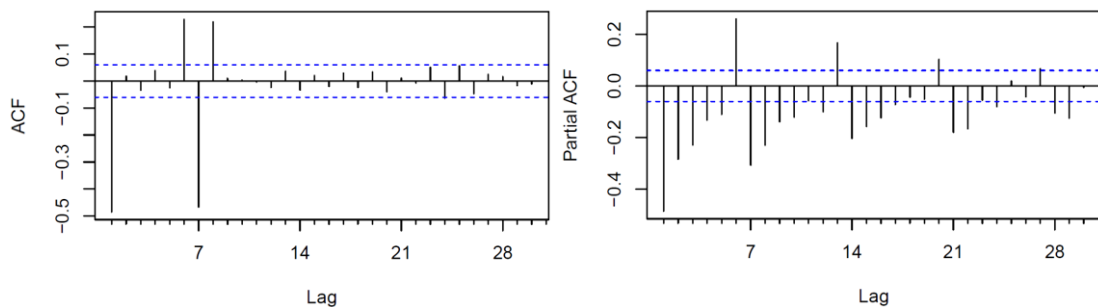


Figure 5.19. ACF and PACF Plot of Differenced Time-Series Data

Figure 5.19 gives the ACF and PACF plot of the differenced time-series data. The ACF correlation coefficient of differenced data is almost equal to -0.5. This indicates that the normal differencing of the data is sufficient to handle both non-stationary and seasonal components in the data. Seasonal differencing is not required as seasonal differencing leads to no correlation. Hence an ARIMA model with a seasonal component would be best to model for the data under consideration.

5.6.7. ARIMA Model

The Hyndman-Khandakar algorithm which combines unit root tests, minimisation of Akaike Information Criteria (AIC), Corrected AIC (AICc) and Bayesian Information

Criterion (BIC) values to obtain the best model are used to identify the model using R software (version 3.5). The parameters are estimated as discussed in the methodology. Different models were tested out of which the details of the best fit model for 30 days and 7 days forecast are given in Table 5.20 and 5.21 respectively.

Table 5.20. Best Fit ARIMA Model – 30 Days

ARIMA (1,1,2)(1,0,0)[7]				
	AR1	MA1	MA2	SAR1
Coefficients	0.67	-1.55	0.56	0.09
Standard Error	0.12	0.13	0.12	0.03
Sigma ² estimated = 0.61	log likelihood = -1223.51	AIC = 2457.03	AICc = 2457.09	BIC = 2481.77

The ARIMA model has one AR term, two MA terms and one Seasonal AR terms. The standard errors of all the coefficients are small, indicating that the estimates of the coefficients have a small standard deviation. It has a normal differencing to make the data stationary.

Table 5.21. Best Fit ARIMA Model – 7 Days

ARIMA (1,1,2)(1,0,0)[7]				
	AR1	MA1	MA2	SAR1
Coefficients	0.66	-1.54	0.54	0.09
Standard Error	0.12	0.17	0.13	0.03
sigma ² estimated = 0.61	log likelihood = -1243.72	AIC = 2497.44	AICc = 2497.5	BIC = 2522.29

The ARIMA model fitted for 7 days forecast has one AR term, two MA term and one seasonal AR term. This model also has a differencing to make the data stationary. The standard errors of all the coefficients are small, indicating that the estimates of the coefficients have a small standard deviation.

After the model identification and parameter estimation of the best model, diagnostic checking of the fitted model is performed by the residual analysis. The sample ACF and PACF of the residuals of the 30 days model and 7 days model are plotted as shown in Figure 5.20 and 5.21 respectively to check whether significant correlations exist in the residuals that can be extracted to improve the model further.

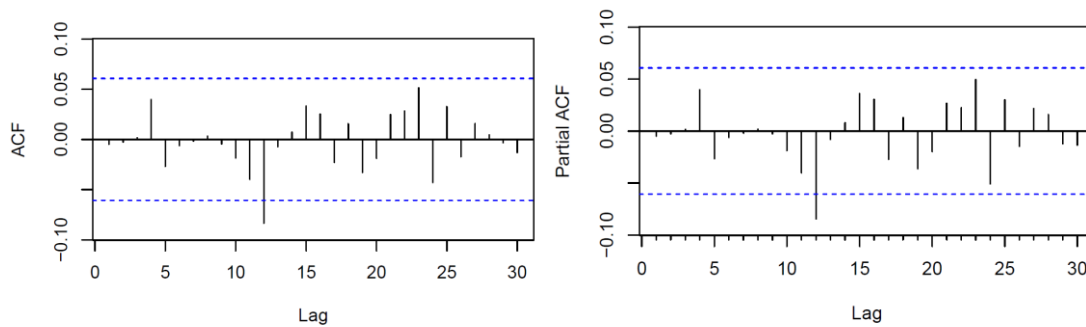


Figure 5.20. ACF and PACF Plot of the Residual – 30 Days

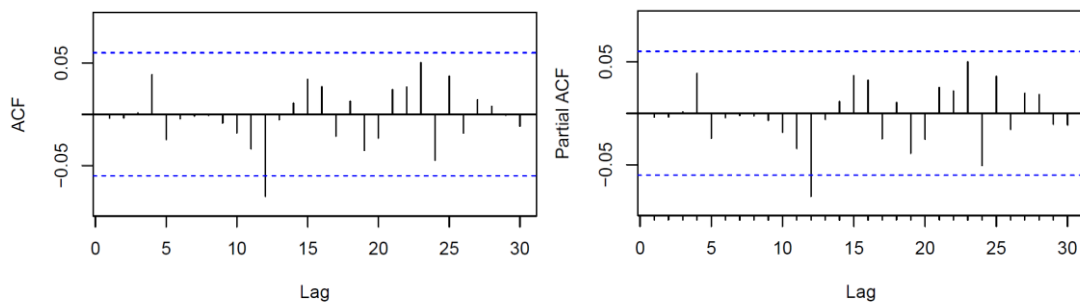


Figure 5.21. ACF and PACF Plot of the Residual – 7 Days

It can be observed that the ACF and PACF of the residuals for both 30 days (Figure 5.20) and 7 days (Figure 5.21) does not significantly differ from zero which indicates that there is no further information that can be extracted to improve the model.

The statistical tests were carried out to check the normality of the residuals, ARCH and the goodness of fit of the model. The results of the statistical test are tabulated below.

Table 5.22. Statistical Test Result of 30 Days Model Residual

Statistical Tests	Hypothesis	Test Statistic value/ <i>p</i> -value
Normality Test		
Shapiro-Wilk Test	H_0 : Data is normally distributed	$W = 0.98$
	H_1 : Data is non normal	p -value = 2.804×10^{-10}
Jarque-Bera Test	H_0 : Skewness and Kurtosis of data does not match the normal distribution	X-squared = 45.62
	H_1 : Data is non normal	p -value = 1.24×10^{-10}
Goodness of Fit Test		
Ljung-Box Test	H_0 : Data is serially independent and act as white noise	Q statistic = 11.99
	H_1 : Data is serially correlated, information can be extracted	p -value = 0.27

Statistical Tests	Hypothesis	Test Statistic value/ <i>p</i> -value
Heteroscedasticity Test		
Engles' ARCH Test	H ₀ : Data is homoscedastic	-
	H ₁ : Data is heteroscedastic	<i>p</i> -value (avg) = 0.14
Non-Linearity Test		
Terasvirta Neural Network Test	H ₀ : Data is linear with respect to mean	X-Squared value = 0.75
	H ₁ : Data is nonlinear in mean	<i>p</i> -value = 0.69
White Neural Network Test	H ₀ : Data is linear in mean	X-squared = 1.38
	H ₁ : Data is nonlinear in mean	<i>p</i> -value = 0.5
Keenan's Test	H ₀ : Data follows AR process	Keenans' Test stat = 0.60
	H ₁ : The time series plot does not have an AR element	<i>p</i> -value = 0.44
McLeod Li Test	H ₀ : Data follows the ARIMA process	-
	H ₁ : Data does not have ARIMA element	<i>p</i> -value = 0.37
Tsay's Test	H ₀ : Data follows AR process	Tsays' Test stat = 2.72
	H ₁ : Data does not follow AR process	<i>p</i> value = 0.003

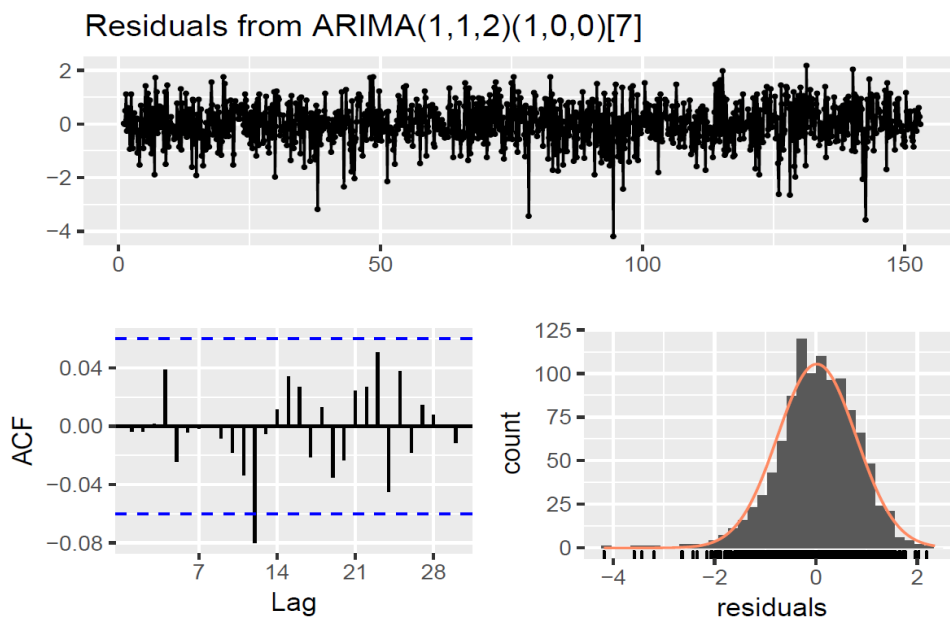


Figure 5.22. Residual Plot for 30 Days

The statistical tests for normal distribution shows that the residuals are normally distributed (p -value < 0.05). Also, the Ljung-Box test result reveals that the residuals are serially independent, *i.e.*, the model fit is good. The p -value of Engles' ARCH test

is greater than 0.05, indicating that the residuals are homoscedastic. The non-linearity test reveals that the residuals are linear with respect to mean, but it still has hidden AR or ARIMA process. Further, residuals were plotted (Figure 5.22) to check whether the residuals are white noise, normally distributed and independent.

From the statistical tests, it can be confirmed that the residuals of 30 days' forecasting model are serially independent, linear with respect to mean and have no ARCH effect. The residuals have some hidden non-linearity with hidden AR effect. The same can be visualized from Figure 5.22. Therefore, further improvements in the ARIMA model can be made by utilising the hidden AR component of the residuals.

Table 5.23. Statistical Test Result of 7 Days Model Residual

Statistical Tests	Hypothesis	Test Statistic value/ <i>p</i> -value
Normality Test		
Shapiro-Wilk Test	H ₀ : Data is normally distributed	W = 0.98
	H ₁ : Data is non normal	<i>p</i> -value = 2.10x10 ⁻¹⁰
Jarque-Bera Test	H ₀ : Skewness and Kurtosis of data does not match the normal distribution	X-squared = 163.78
	H ₁ : Data is non normal	<i>p</i> -value = 2.2x 10 ⁻¹⁶
Goodness of Fit Test		
Ljung-Box Test	H ₀ : Data is serially independent and act as white noise	Q statistic = 11.00
	H ₁ : Data is serially correlated, information can be extracted	<i>p</i> -value = 0.36
Heteroscedasticity Test		
Engles' ARCH Test	H ₀ : Data is homoscedastic	
	H ₁ : Data is heteroscedastic	<i>p</i> -value (avg) = 0.3
Non-Linearity Test		
Terasvirta Neural Network Test	H ₀ : Data is linear with respect to mean	X-Squared value = 2.05
	H ₁ : Data is nonlinear in mean	<i>p</i> -value = 0.36
White Neural Network Test	H ₀ : Data is linear in mean	X-squared = 3.02
	H ₁ : Data is nonlinear in mean	<i>p</i> -value = 0.22
Keenan's Test	H ₀ : Data follows AR process	Keenans' Test stat = 2.06
	H ₁ : The time series plot does not have an AR element	<i>p</i> -value = 0.15
McLeod Li Test	H ₀ : Data follows the ARIMA process	
	H ₁ : Data does not have ARIMA element	<i>p</i> -value = 0.27
Tsay's Test	H ₀ : Data follows AR process	Tsay's Test stat = 0.46
	H ₁ : Data does not follow AR process	<i>p</i> value = 0.71

The statistical tests for normal distribution shows that the residuals are normally distributed (p -value < 0.05). Also, the Ljung-Box test result reveals that the residuals are serially independent, *i.e.*, the model fit is good. The p -value of Engles' ARCH test is greater than 0.05, indicating that the residuals are homoscedastic. The non-linearity test reveals that the residuals are linear with respect to mean, but it still has hidden AR or ARIMA process. Further, residuals were plotted (Figure 5.23) to check whether the residuals are white noise, normally distributed and independent.

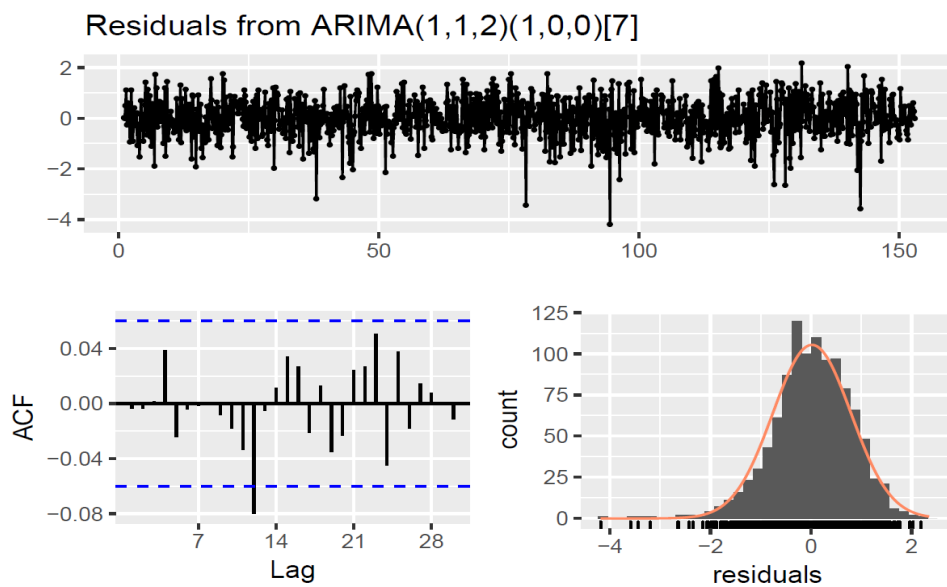


Figure 5.23. Residual Plot for 7 Days

The Ljung-Box test is used to test the goodness of fit of the model by checking whether the residuals are white noise (mean is zero, constant variance and uncorrelated observations) or not. The Ljung-Box test p -value for 30 days model is 0.27 and p -value for 7 days model is 0.36, both values are greater than 0.05. Hence, the null hypothesis is accepted, indicating that the residuals are serially independently distributed and behave like white noise. Therefore, the diagnostic checking confirms that the model exhibits a good fit. Additionally, the test for normality of the residual errors was carried out by the Jarque-Bera test and Shapiro-Wilk test, which yielded a p -value of 1.24×10^{-10} and 2.804×10^{-10} respectively for 30 days and 2.2×10^{-16} and 2.10×10^{-10} for 7 days forecast, which confirms that the residuals are normally distributed. Finally, residual heteroscedasticity is tested using ARCH Engle's test and found that the squared residuals are homoscedastic. Further, the nonlinearity analysis indicates that the

residuals are linear with respect to mean but has hidden AR or ARIMA process. This indicates that the developed model can be further improved by utilising the residual AR or ARIMA components.

5.6.8. Forecasting Bus Passenger Demand

Using the ARIMA model developed for 30 days and 7 days forecast, one step ahead passenger demand was forecasted. Mean Absolute Percentage Error (MAPE) was used to estimate the forecast accuracy of the models. Tables 5.24 and 5.25 gives the forecast accuracy and forecasted values for 30 days forecast. Table 5.26 and 5.27 gives the forecast accuracy and forecasted values for 7 days forecast.

Table 5.24. Forecast Accuracy 30 Days

	MAPE (%)
Training Set	20.54
Test Set	13.47

Table 5.25. Forecasted Values for a Forecast Horizon of 30 Days

Actual	ARIMA Forecasts	ARIMA Accuracy (MAPE)
56	70.30	25.53
50	68.99	37.99
52	69.02	32.73
55	69.17	25.76
60	69.27	15.45
65	69.65	7.15
68	68.13	0.19
66	70.17	6.32
65	70.11	7.86
62	70.15	13.15
71	70.19	1.14
58	70.22	21.07
51.5	70.27	36.44
60	70.13	16.89
64	70.33	9.89
83	70.33	15.27
72.5	70.33	2.99
62	70.34	13.45
71	70.34	0.93
80	70.35	12.07
65	70.33	8.21
81.5	70.35	13.68

Actual	ARIMA Forecasts	ARIMA Accuracy (MAPE)
74	70.35	4.93
61	70.35	15.34
67	70.35	5.01
69	70.35	1.96
67.5	70.35	4.23
87	70.35	19.13
80	70.35	12.06
85	70.35	17.23

The forecast accuracy of the model is 13.47%. As per literature, forecast accuracy for passenger demand of public transports should be less than 10%. Hence the requirement is not satisfied; this may be due to the fitting of the model along the mean value. About 56% of the values have MAPE greater than 10%.

Table 5.26. Forecast Accuracy 7 Days

	MAPE (%)
Training Set	18.86
Test Set	11.07

Table 5.27. Forecasted Values for a Forecast Horizon of 7 Days

Actual	ARIMA Forecasts	ARIMA Accuracy (MAPE)
74	72.18	2.46
61	70.58	15.70
67	71.00	5.97
69	71.48	3.59
67.5	71.99	6.65
87	71.30	18.05
80	70.61	11.73

The forecast accuracy of the model is 11.07%. As per literature, forecast accuracy for passenger demand of public transports should be less than 10% (Kumar and Vanajakshi 2013). From Table 5.27, it can be observed that about 28% of the values are greater than 10%, indicating that the ARIMA model is better than the conventional Naïve Forecast model. Since the residuals have hidden AR or ARIMA component, the developed model can be improved using those residual data. As the residuals do not exhibit ARCH effect, ARCH modelling is not required. The model can be improved by Neural Network Method.

5.7. AUTOREGRESSIVE NEURAL NETWORK (ARNN) MODEL

Since the actual data exhibits non-linearity and follows the AR process, ARNN can be used to model the data. The number of time lags to be considered for the input layer are decided based on the correlation plots (ACF and PACF). For the data used, from the correlation plot, it was decided that a time lag of up to 15 is to be considered. The number of neurons in the hidden layer was taken as 10 based on trial and error. The data are split into training data and testing data. The length of testing data equals the forecast horizon (Hyndman and Athanasopoulos 2018) and remaining was given as training data. ARNN is fitted for the training data, and forecast accuracy was checked against the testing data. The result of ARNN model for 30 days is summarised in Table 5.28.

Table 5.28. ARNN Forecast Summary for 30 Days Forecast Horizon

ARNN (15-1-10), $\sigma^2 = 0.003$, Box-Cox Lambda = 0.139	
Forecast Accuracy	
	MAPE (%)
Training Data	3.27
Testing Data	14.87

Table 5.29. ARNN Forecasted Values for 30 Days Forecast Horizon

Actual Value	ARNN Forecasts	ARNN Accuracy (MAPE)
56	73.23	30.77
50	70.52	41.04
52	48.82	6.11
55	72.82	32.40
60	75.87	26.45
65	63.13	2.87
68	62.89	7.51
66	75.64	14.60
65	58.15	10.53
62	71.89	15.94
71	52.31	26.33
58	65.14	12.32
51.5	56.17	9.06
60	61.23	2.05
64	68.76	7.44
83	68.82	17.08
72.5	64.00	11.72

Actual Value	ARNN Forecasts	ARNN Accuracy (MAPE)
62	67.98	9.64
71	65.21	8.16
80	71.79	10.25
65	69.77	7.34
81.5	65.87	19.18
74	66.63	9.96
61	63.11	3.47
67	49.97	25.45
69	69.91	1.32
67.5	62.87	6.87
87	70.99	18.39
80	58.34	27.08
85	64.01	24.69

The forecast accuracy of ARNN model for 30 days' forecast horizon is 14.87%, higher than 10%. Hence the forecast accuracy is not satisfactory. The forecast accuracy is even less than that of the ARIMA model. Since the ARNN model is a feed-forward model, the training will be proper, but testing may not yield a better result. From Table 5.28, it is observed that the fitting/training accuracy of ARNN model has tremendously increased in comparison with the ARIMA model. In Table 5.29, around 57% of the individual forecast exceeds the threshold MAPE (10%).

Table 5.30. ARNN Forecast Summary for 7 Days Forecast Horizon

ARNN (15-1-10), $\sigma^2 = 0.003$, Box-Cox Lambda = 0.139	
Forecast Accuracy	
	MAPE (%)
Training Data	6.68
Testing Data	15.35

Table 5.31. ARNN Forecasted Values for 7 Days Forecast Horizon

Actual Value	ARNN Forecasts	ARNN Accuracy
61	62.98	3.24
67	60.22	10.12
69	68.78	0.32
67.5	68.73	1.82
87	51.31	41.02
80	66.20	17.25
85	56.36	33.69

From Table 5.30 and 5.31, it can be seen that the forecast accuracy of ARNN model for 7 days forecast is not satisfactory. Also, around 57% of the one-step forecasts are greater than the threshold limit. The ARNN model does not exhibit any lack of fit, and the residuals behave like white noise as the Ljung-Box test p -value is greater than 0.05. The forecast result of ARIMA was better than ARNN model, but the fitting properties of ARNN model have tremendously increased. Hence a Hybrid ARIMA-ARNN model can give better fitting as well as forecast results.

5.8. HYBRID ARIMA-ARNN MODEL

From the results of the residual analysis, it was observed that there exists a hidden non-linearity in the residuals, which can be modelled using ARNN. Hence in the hybrid model, the residuals of the ARIMA model fitted are modelled using ARNN. Thus, the linearity in the data is handled by ARIMA and nonlinearity in the residuals are handled by a neural network model. For the development of a hybrid model, the data is first fitted to (1,1,2)(1,0,0)[7] ARIMA model using the ARIMA modelling methodology as discussed in section 5.6. The ARIMA model fitted is having one AR term, one normal differencing, two MA term and one SAR term. The values '0', '0' in the model represents the seasonal differencing and mseasonal MA terms. The model is having a seasonality of seven. The residuals of this ARIMA model are then extracted. The residual data is split into training data and testing data. The length of test data equals the forecast horizon and remaining data was given as train data. Then the training data is fitted to the neural network (NN) with structure NN (15-1-10) model. The neural network uses 15 lagged values for building the model with one hidden layer and 10 hidden neurons. The residual forecasts obtained from the NN model is then added to the ARIMA forecasts to obtain the Hybrid ARIMA-ARNN forecasts. The forecast accuracy of the hybrid forecasts is then calculated to evaluate the performance of the model. The data is split into train data and test data. The 30 days and 7 days forecast accuracy for the hybrid model are given in Tables 5.32.

Table 5.32 Hybrid Model Accuracy

HYBRID MODEL- ARIMA(1,1,2)(1,0,0)+ARNN(15-1-10)		
	Forecast Accuracy - 30 days	Forecast Accuracy - 7 days
MAPE (%)	9.53	9.72

Table 5.33 and 5.34 gives the accuracy and the forecasted values of ARIMA, ARNN and hybrid ARIMA-ARNN model of 30 days forecast horizon and 7 days forecast horizon.

Table 5.33. Forecasted Values for 30-Days Forecast Horizon

Actual Values	ARIMA Forecast	ARNN Forecast	Hybrid Forecast	ARIMA Accuracy (MAPE)	ARNN Accuracy (MAPE)	Hybrid Accuracy (MAPE)
56	70.29	73.23	45.84	25.53	30.77	8.64
50	68.99	70.52	56.09	37.98	41.04	4.18
52	69.02	48.82	54.74	32.73	6.11	5.27
55	69.17	72.82	57.69	25.76	32.40	4.89
60	69.27	75.87	62.43	15.45	26.45	4.05
65	69.65	63.13	63.21	7.15	2.87	2.75
68	68.13	62.89	62.07	0.19	7.51	8.72
66	70.17	75.64	64.91	6.32	14.60	1.65
65	70.11	58.15	65.18	7.86	10.53	0.28
62	70.15	71.89	64.63	13.15	15.94	4.24
71	70.19	52.31	64.39	1.14	26.33	9.31
58	70.22	65.14	63.68	21.07	12.32	9.79
51.5	70.27	56.17	63.04	36.44	9.06	2.99
60	70.13	61.23	63.82	16.89	2.05	6.37
64	70.33	68.76	64.59	9.89	7.44	0.92
83	70.33	68.82	64.27	15.27	17.08	9.31
72.5	70.33	64.00	61.14	2.99	11.72	6.01
62	70.34	67.98	63.67	13.45	9.64	2.69
71	70.34	65.21	61.65	0.93	8.16	7.54
80	70.35	71.79	65.07	12.07	10.25	6.16
65	70.33	69.77	66.37	8.21	7.34	2.11
81.5	70.35	65.87	66.08	13.68	19.18	11.56
74	70.35	66.63	65.62	4.93	9.96	11.32
61	70.35	63.11	66.19	15.33	3.47	8.51
67	70.35	49.97	67.51	5.01	25.42	0.76
69	70.35	69.91	68.24	1.96	1.32	1.10
67.5	70.35	62.87	67.62	4.23	6.87	0.18
87	70.35	70.99	67.02	19.13	18.39	11.47
80	70.35	58.34	68.31	12.06	27.07	12.11
85	70.35	64.01	66.68	17.23	24.69	9.79

Table 5.34. Forecasted Values for 7-Days Forecast Horizon

Actual Values	ARIMA Forecast	ARNN Forecast	Hybrid Forecast	ARIMA Accuracy (MAPE)	ARNN Accuracy (MAPE)	Hybrid Accuracy (MAPE)
61	72.18	62.98	70.77	18.33	3.24	9.01
67	70.58	60.22	68.66	5.34	10.12	2.48
69	71.00	68.78	70.99	2.90	0.32	2.89
67.5	71.49	68.73	71.82	5.90	1.82	6.39
87	71.99	51.31	71.99	17.24	41.02	9.24
80	71.30	66.19	71.31	10.87	17.25	9.87
85	70.61	56.36	74.69	16.93	33.69	8.13

The forecast accuracy for 30 days and 7 days are satisfactory as they have MAPE values less than 10%. Out of the models created the hybrid model gave satisfactory results in terms of MAPE and model fitting properties. From the results, it can be concluded that the Hybrid ARIMA-ARNN model is best suited for this route. This is evident from the results of forecast accuracy. ARNN was attempted as the model does not exhibit the ARCH effect. If the model exhibits ARCH effect, then ARCH has to be modelled suitably using ARCH/GARCH models.

5.9. HYBRID ARIMA-GARCH MODEL

If the residuals of ARIMA exhibit ARCH effect, then the ARCH effect (conditional volatility) has to be modelled using ARCH/GARCH model, which is a stochastic model. The ARIMA model handles the linear variations of the time-varying mean for a constant variance. ARCH refers to time-varying variances, which cannot be modelled by ARIMA. This time-varying variance is modelled using ARCH/GARCH modelling. A sample data analysis and result for ARCH modelling are given below.

5.9.1. Data Analysis and Model Fitting

Data analysis, model fitting and residual analysis were carried out on Vizhinjam Route Number 3. The best fit ARIMA model details are given in Table 5.35 and results are summarised in Table 5.36. The data is fitted to ARIMA (2,1,1) (2,0,0) [7] model using the ARIMA modelling methodology explained above (section 5.6). The ARIMA model fitted is having two AR term, one normal differencing, one MA term and two SMA term. The model is having a seasonality of seven.

Table 5.35. Best Fit ARIMA Model for Vizhinjam Route3- 30 Days

ARIMA (2,1,1) (2,0,0) [7]					
Coefficients	AR1	AR2	MA1	SAR1	SAR2
	0.1241	0.1459	-0.9724	0.0353	0.0856
s.e.	0.0365	0.0359	0.0092	0.0359	0.0361
Sigma ² estimated: 290.9	Log likelihood=-3496.58		AIC=7005.16	AICc=7005.27	BIC=7033.43

Table 5.36. Summary of Data and Residual Analysis of ARIMA

Data Analysis and Model Fitting	
Stationarity	Non-stationary due to trend
Seasonality	Non-significant seasonality
Trend	Increasing Trend
Normality	Non-Normal Data
Linearity	Non-Linear data; has an AR component
ARIMA Model Fitted	(2,1,1)(2,0,0)[7]
Model Accuracy	
Forecast MAPE (%)	12.53%
Residual Analysis	
Normality	Non -Normal Residuals
Linearity	N.A.
Serial Independency	Serially independent
ARCH effect	Residuals Exhibit ARCH effect

5.9.2. GARCH Modelling

From ARIMA residual analysis, it was observed that the residuals exhibit ARCH effect. On ARIMA modelling, the data is modelled only with respect to the conditional mean. To deal with the conditional variance, GARCH modelling is to be done. The ARCH test was conducted on the residuals as a preliminary analysis. The plot on squared residuals and correlation of squared residuals gives conclusive evidence for the presence of ARCH effect. Figure 5.24 and 5.25 gives the plot of squared residuals and its ACF and PACF plots, respectively.

From the squared residuals plot, a cluster of varying variances can be observed changing over time. Also, significant spikes can be observed both in ACF and PACF plot of squared residuals. Hence the presence of ARCH is confirmed, and a GARCH model should be used to model the ARCH effect. The order of the GARCH model was

obtained from ACF and PACF of squared residuals. For the data used, the AIC and AICc values converged at GARCH (2,2). The model summary of GARCH is shown in Table 5.37.

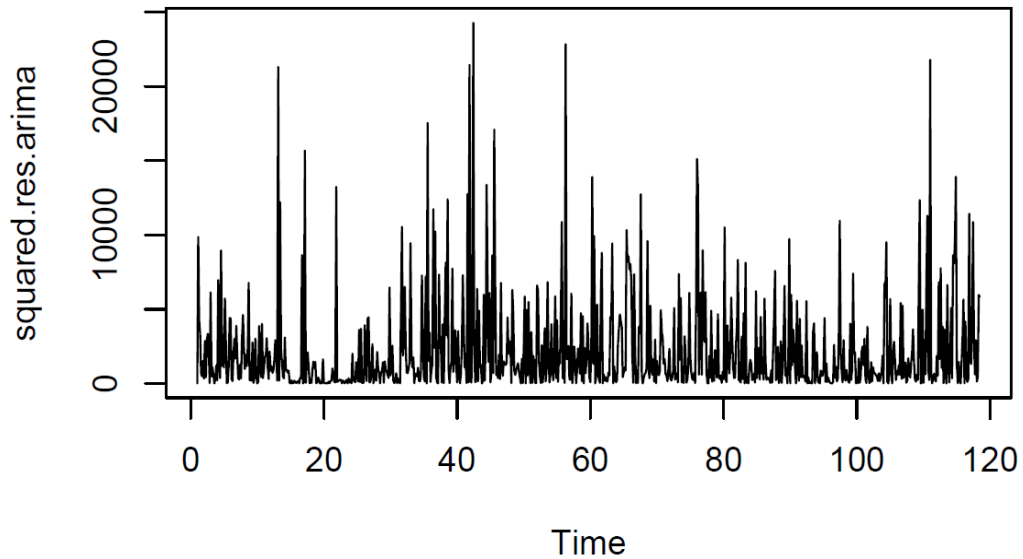


Figure 5.24. Squared Residuals plot

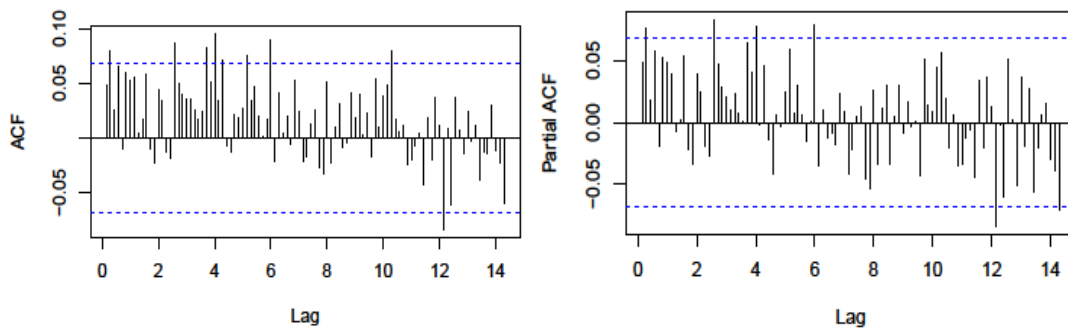


Figure 5.25. ACF and PACF plot of Squared residuals

Table 5.37. GARCH Model Summary

GARCH (2,2)		
Coefficients	Estimate	Standard Error
a ₀	1.744e+03	7.749e+02
a ₁	5.594e-02	4.285e-02
a ₂	8.126e-02	4.406e-02
b ₁	2.718e-11	3.956e-01
b ₂	6.800e-02	2.401e-01

5.9.3. GARCH Model Forecast and Accuracy

The conditional variance modelled for each time step is shown in Table 5.38.

Table 5.38. GARCH Forecast

Time Step	Forecasted Variance
T+1	54.63
T+2	54.38
T+3	54.34
T+4	54.26
T+5	54.18
T+6	54.11
T+7	54.03
T+8	53.96
T+9	53.89
T+10	53.82
T+11	53.75
T+12	53.68
T+13	53.61
T+14	53.54
T+15	53.48
T+16	53.41
T+17	53.35
T+18	53.28
T+19	53.22
T+20	53.15
T+21	53.09
T+22	53.03
T+23	52.97
T+24	52.91
T+25	52.85
T+26	52.79
T+27	52.73
T+28	52.68
T+29	52.62
T+30	52.56

This forecast is added with ARIMA forecast to obtain a hybrid ARIMA-GARCH model. This ARIMA-GARCH model is a complete stochastic model modelled for

varying mean and variance. Table 5.39 gives the forecast accuracy of ARIMA and ARIMA-GARCH model.

Table 5.39. Forecast Accuracy of ARIMA and Hybrid ARIMA-GARCH

ARIMA Forecast Accuracy		
Forecast Horizon	MAPE	Ljung-Box p -Value
30 Days	12.53	0.20
7 Days	9.2	0.23
ARIMA-GARCH Forecast Accuracy		
Forecast Horizon	MAPE	Ljung-Box p -Value
30 Days	9.81	0.80
7 Days	8.8	0.75

From the results, it can be observed that ARIMA-GARCH modelling improves the forecast accuracy.

Table 5.40 and Figure 5.26 shows the results of intra-city forecast accuracy in the form of MAPE. The average MAPE value of ARIMA model for 30-days forecast horizon is 12.77%, and that of 7-days forecast horizon is 12.73. The ARNN model has an average MAPE value of 14.18 and 12.82 for 30-days and 7-days forecast horizon, respectively. In the case of the hybrid-ARIMA-ARNN model, the forecast accuracy has been increased with an average MAPE value of 9.16 and 7.98 for 30-days and 7-days forecast horizon, respectively. It can be seen that the MAPE values of the ARIMA models are improved using hybrid ARIMA-ARNN model. On the data having ARCH effect, hybrid ARIMA-ARNN model was used. The forecast accuracy was found to be 9.40% for 30 days forecast horizon and 7.50% for 7 days forecast horizon.

Table 5.40. Result Summary of Intra-City (Kerala SRTC) Forecast Accuracy

Route	30 Days Forecast Horizon				7 Days Forecast Horizon			
	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH
	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE
Vizhinjam Depot								
50	11.64	19.45	9.69	N.A	11.1	16.1	7.97	N.A.
68	10.07	11.38	9.61	N.A.	10.58	12.36	7.37	N.A.
72	10.07	11.38	9.61	N.A.	22.8	11.88	8.26	N.A.
90	11.21	15.05	9.15	N.A.	11.21	15.05	9.15	N.A.
91	12.37	10.14	9.44	N.A.	18.98	11.65	5.58	N.A.
137	13.47	14.87	9.53	N.A.	11.07	15.35	9.72	N.A.
3	8.027			7.132	10.15			9.3
56	11.34			Strict White Noise	12.15			Strict White Noise
66	19.37			Strict White Noise	18.36			Strict White Noise
74	20.36			Strict White Noise	19.99			Strict White Noise
78	17.24			Strict White Noise	16.29			Strict White Noise
84	7.75			Strict White Noise	10.2			Strict White Noise
85	9.59			9.55	9.88			9.68
142	14.5			Strict White Noise	15.44			Strict White Noise
Pappanamcode Depot								
29	15.11	19.42	9.26	N.A.	14.56	13.47	6.82	N.A.
76	11.22	20.51	8.71	N.A.	16.65	13.42	6.11	N.A.
409	12.39	15.07	8.99	N.A.	11.13	6.23	9.59	N.A.
410	10.2	7.48	8.11	N.A.	10	13.44	7.79	N.A.

Route	30 Days Forecast Horizon				7 Days Forecast Horizon			
	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH
	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE
426	12.34	12.09	8.66	N.A.	8.28	8.43	3.87	N.A.
427	12.21	18.54	8.66	N.A.	13.23	12.32	8.83	N.A.
710	8.62	10.42	7.6	N.A.	8.1	5.53	7.56	N.A.
713	9.47	12.81	8.94	N.A.	7.94	9.13	6.34	N.A.
424	10.21			10.09	6.21			6.21
93	23.03			Strict White Noise	23.35			Strict White Noise
132	16.98			Strict White Noise	20.04			Strict White Noise
419	17.92			Strict White Noise	18.2			Strict White Noise
				Strict White Noise				Strict White Noise
Trivandrum City Depot								
3	12.31	19.89	9.64	NA	11.56	17.37	9.88	NA
4	11.62	12.72	11.62	NA	8.43	9.29	8.17	NA
15	9.92	11.62	9.71	NA	11.16	12.88	8.22	NA
52	11.58	11.36	9.84	NA	13.89	14.96	9.59	NA
90	7.21	12.6	6.6	NA	15.99	7.12	6.36	NA
95	12.96	17.56	9.92	NA	13.94	16.77	7.76	NA
133	10.79	8.89	13.78	NA	8.94	8.35	8.39	NA
137	10.79	12.23	9.72	NA	13.53	12.41	7.53	NA
204	10.29	14.41	9.96	NA	10.4	13.4	9.65	NA
205	7.54	7.48	14.61	NA	9.33	13.55	7.34	NA
142	12.19			Strict White Noise	19.18			Strict White Noise
203	18.07			Strict White Noise	13.03			Strict White Noise

Route	30 Days Forecast Horizon				7 Days Forecast Horizon			
	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH
	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE
206	13.79			Strict White Noise	9.64			Strict White Noise
208	15.47			Strict White Noise	15.47			Strict White Noise
2	13.93			Strict White Noise	11.86			Strict White Noise
21	18.93			Strict White Noise	12.89			Strict White Noise
Vikas Bhavan Depot								
4	15.61	17.89	9.53	NA	20.13	15.07	9.11	NA
11	11.2	13.06	9.51	NA	12.89	9.3	9.52	NA
13	10.84	17.28	9.97	NA	8.5	14.62	6.89	NA
17	10.1	13.89	9.85	NA	15.11	10.53	9.21	NA
19	13.33	13.06	9.84	NA	16.57	11.09	7.46	NA
27	10.93	15.64	8.97	NA	10.51	8.34	8.44	NA
56	7.5	9.5	6.3	NA	10.31	11.43	7.51	NA
65	9	10.48	8.85	NA	11.89	9.89	9.67	NA
71	12.6	17.09	9.65	NA	10.89	11.71	8.78	NA
73	13.61	13.78	9.64	NA	9.87	19.73	7.34	NA
94	11.62	12.44	9.97	NA	11.51	26.61	8.03	NA
130	14.04	19.87	9.45	NA	17.44	16.15	9.59	NA
141	12.19	13.16	8.5	NA	9.76	13.59	8.86	NA
143	14.46	16.63	9.01	NA	9.92	13.96	8.27	NA
2	11.08			9.83	10.69			6.01
55	10.15			9.23	15.72			7.36
54	14.052			Strict White Noise	10.03			Strict White Noise

Route	30 Days Forecast Horizon				7 Days Forecast Horizon			
	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH
	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE
Kaniyapuram Depot								
11	11.43	16.49	6.66	NA	6.301	14.29	3.987	NA
12	8.96	9.6	7.55	NA	8.4	7.96	6.03	NA
72	12.38	15.68	8.89	NA	9.38	12.68	7.32	NA
10	10.25			9.26	10.94			9.59
51	10.77			9.89	2.68			2.44
64	10.03			9.45	12.87			7.89
92	9.93			9.76	6.59			3.97
53	11.53			8.87	10.38			9.74
52	12.68			Strict White Noise	9.83			Strict White Noise
55	15.42			Strict White Noise	18.103			Strict White Noise
60	14.53			Strict White Noise	19.62			Strict White Noise
720	20.93			Strict White Noise	13.27			Strict White Noise
Peroorkada Depot								
3	12.53			9.81	9.15			8.87
7	10.2			9.89	9.24			8.89
101	17.97			Strict White Noise	14.45			Strict White Noise
1	12.36	16.35	8.65	NA	13.65	14.59	8.98	NA
8	11.63	12.36	7.68	NA	10.25	11.65	7.85	NA
17	12.58	15.65	7.85	NA	11.62	12.66	6.65	NA
96	13.65	18.36	6.58	NA	15.36	13.39	9.85	NA
103	18.65	15.67	9.68	NA	14.36	18.36	8.68	NA

Route	30 Days Forecast Horizon				7 Days Forecast Horizon			
	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH
	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE
118	14.36	12.58	8.66	NA	16.85	12.68	7.68	NA
134	15.78	18.65	7.58	NA	17.58	11.65	9.68	NA
145	14.75	12.36	8.56	NA	16.45	15.58	7.77	NA
DEPOT WISE								
Network Level	6.21	8.76	5.31	NA	7.44	8.91	6.72	NA
Pappanamcode Depot	7.25			7.21	8.69			9
Peroorkada Depot	10.13			Strict White Noise	14.14			Strict White Noise
Trivandrum City Depot	8.58			8.23	8.56			7.94
Kaniyapuram Depot	13.29			Strict White Noise	14.14			Strict White Noise
Vizhinjam Depot	10.24			9.71	8.25			8.23
Vikas Bhavan Depot	16.31			Strict White Noise	9.38			Strict White Noise

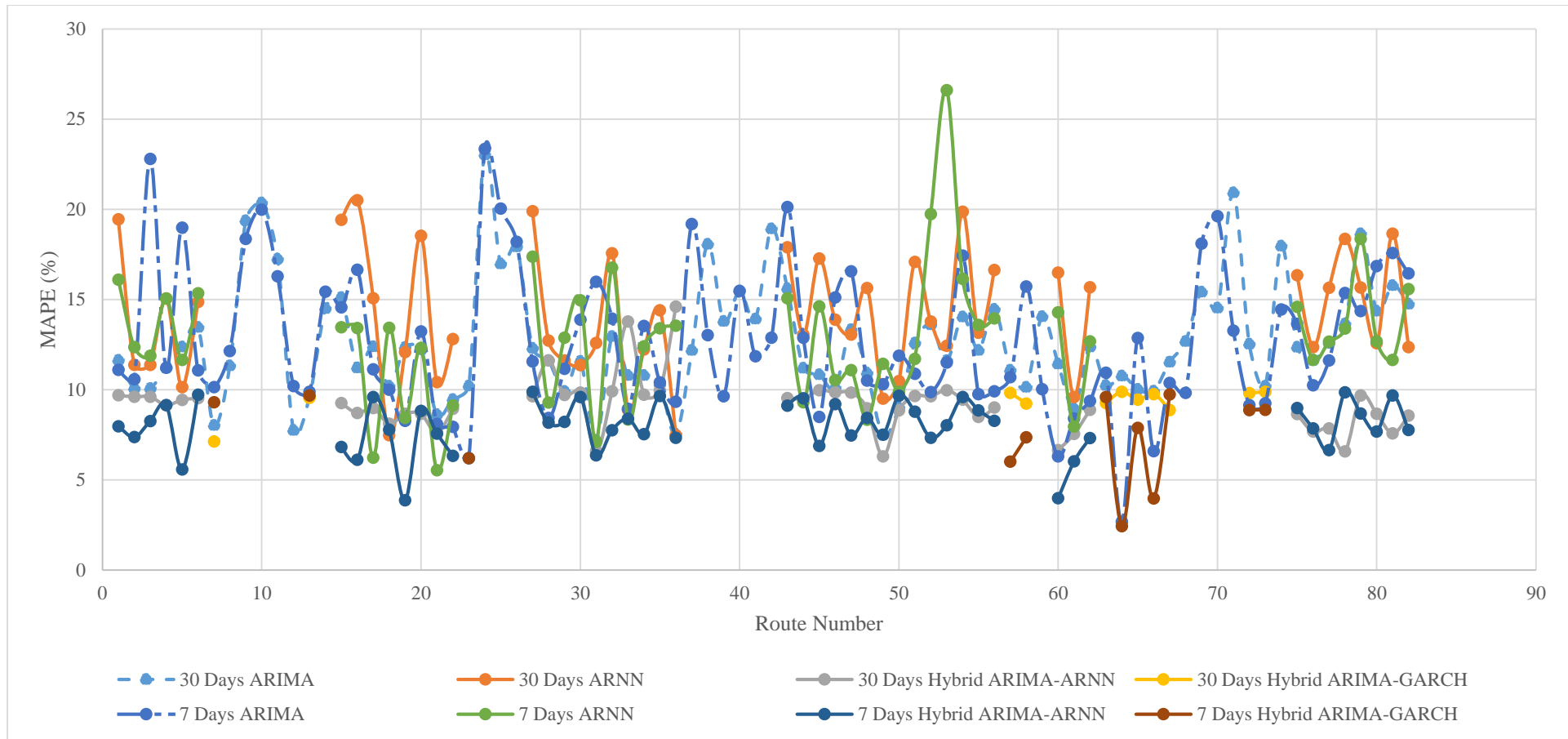


Figure 5.26. Result Summary of Intra-City (Kerala SRTC) Forecast Accuracy

Table 5.41. Result Summary of Inter-City (Karnataka SRTC) Forecast Accuracy

UDUPI DEPOT																
Route	30 Days Forecast Horizon				7 Days Forecast Horizon				Monthly				Quarterly			
	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH	ARIMA	ARNN	Hybrid ARIMA-ARNN	Hybrid ARIMA-GARCH
	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE	MAPE
2	11.73	NA	NA	9.73	15.69	NA	NA	9.05	11.96	12.5	9.91	NA	13.95	13.68	14.82	NA
3	14.27	NA	NA	9.51	11.26	NA	NA	7.74	22.18	12.94	8.82	NA	11.19	16.17	12.3	NA
4	11.56	NA	NA	8.18	11.39	NA	NA	8.84	11.64	13.37	8.25	NA	19.57	15.91	20	NA
5	13.57	NA	NA	9.16	10.55	NA	NA	9.15	159.14	58.1	NA	NA	21.1	9.45	21.64	NA
6	9.7	NA	NA	9.15	8.45	NA	NA	8.43	77.36	11.38	9.92	NA	7.27	9.928	7.925	NA
12	12.57	NA	NA	9.2	5.77	NA	NA	5.16	163.76	59.79	NA	NA	61.11	15.47	59.72	NA
18	16.21	NA	NA	7.88	10.84	NA	NA	8.01	75.44	18.38	8.02	NA	61.53	14.96	62.95	NA
28	15.65	NA	NA	9.45	11.68	NA	NA	7.26	62.53	7.55	9.65	NA	8.33	8.02	9.023	NA
87	27.24	NA	NA	15.24	38.9	NA	NA	22.21	51.07	94.4	NA	NA	49.17	40.81	43.17	NA
MANGALORE DEPOT																
1	18.53	NA	NA	9.64	13.25	NA	NA	5.45	21.4	6.96	9.75	NA	41.28	35.12	41.05	NA
14	20.5	NA	NA	9.83	9.91	NA	NA	9.69	17.53	7.89	9.26	NA	13.47	10.88	13.77	NA
34	17.2	NA	NA	9.7	21.4	NA	NA	9.65	14.9	16.75	9.81	NA	21.45	19.75	20.61	NA
40	25.16	NA	NA	13.24	24.25	NA	NA	12.35	26.01	20.32	14.28	NA	19.97	22.12	19.63	NA
112	22.24	NA	NA	10.05	16.7	NA	NA	7.65	35.38	18.77	12.36	NA	30.49	18.62	31.4	NA
175	13.26	NA	NA	9.04	17.92	NA	NA	9.22	14.63	16.41	9.52	NA	123.12	96.64	29.24	NA
378	35.3	NA	NA	18.25	142.42	NA	NA	86.22	32.78	21.45	15.24	NA	17.21	10.45	17.2	NA
455	15.91	NA	NA	9.92	7.76	NA	NA	7.62	12.62	12.1	6.63	NA	11.22	14.73	11.13	NA

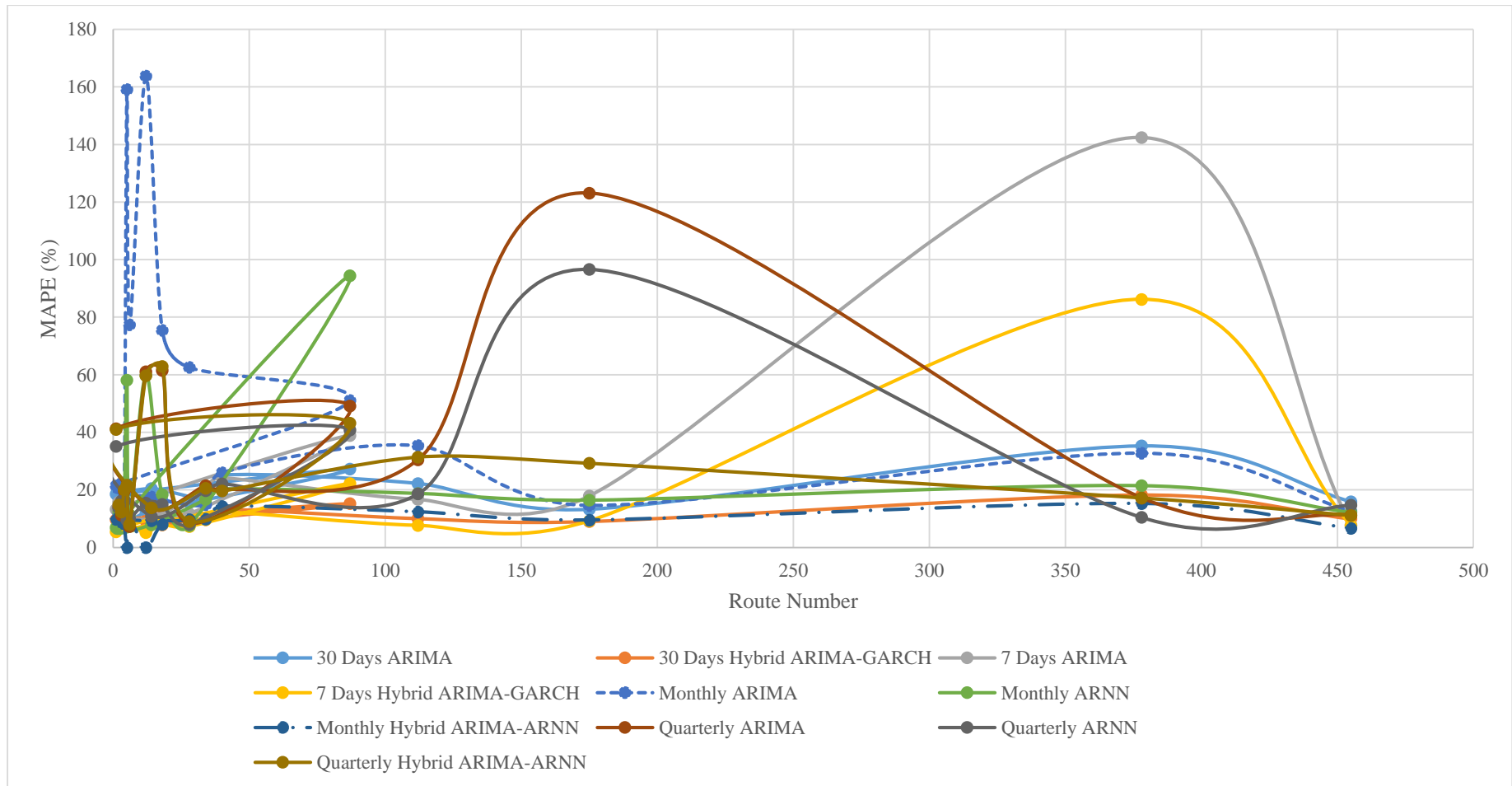


Figure 5.27. Result Summary of Inter-City (Karnataka SRTC) Forecast Accuracy

During the analysis of the Kerala ETM data, it was found that the length of intra-city ETM data available for individual routes varies from 1.5 years to 2 years (2011 to 2013), while inter-city data were not sufficient for time-series modelling. Therefore, inter-city ETM data of Mangalore division Karnataka SRTC was collected for a period of 2013 to 2018, which comprises of a total five-year data, to perform inter-city passenger demand modelling. The intra-city ETM data for Mangalore was not collected since the intra-city services are mainly operated by private operators. The private operators have not maintained and stored the ETM data. Seventeen inter-city routes from two depots of Mangalore division Karnataka SRTC depots were selected in this study. From Table 5.41 and Figure 5.27, it is observed that the average MAPE value of 30 days ARIMA forecast 17.68% while for 7 days forecast it is 22.24%. All the routes of the depots aggregated at daily level (daily data) had ARCH effect. So hybrid ARIMA-GARCH model was used to model the conditional variance. It was found that the MAPE was reduced to 8.42% and 7.75% for 30 days and 7 days forecast horizon respectively. The monthly forecasts of the ARIMA model have an average MAPE value of 48.49 while the average MAPE value of quarterly forecast is 31.26. In the case of monthly data, ARNN model gave an average MAPE of 24.06% while the hybrid ARIMA-ARNN model had an average MAPE value as 10.10%. For quarterly data, ARNN model gave 21.92% MAPE value while hybrid ARIMA-ARNN had 25.62% as average MAPE. The monthly and quarterly forecasts yield poor results due to lack of sufficient data points for modelling.

5.10. DISCUSSION ON INTRA-CITY AND INTER-CITY BUS PASSENGER DEMAND MODELS

Electronic Ticket Machine data was collected from both Kerala SRTC and Karnataka SRTC for intra-city and inter-city, respectively. The data analysis yielded the following features.

- The Kerala SRTC data yielded intra-city travel bus passenger demand. The passenger demand within the urban limits of Trivandrum city was examined using this data. On the other hand, the Karnataka SRTC data was for inter-city

passenger demand. Therefore, the inter-city bus passenger travel characteristics were studied using the Karnataka SRTC data.

- The average length of the data available for intra-city was two years, and that of inter-city was 5 years. Therefore, intra-city data was aggregated on a daily basis while the inter-city data was aggregated to get daily, monthly and quarterly data.
- In the data analysis, it was found that 67% of the intra-city data was nonstationary, while 100% of the data for inter-city was nonstationary. The nonstationary data were differenced to make it stationary for further time series analysis.
- Only 57% of the intra-city data was having seasonal effect while 100% of the inter-city data exhibits seasonality. It is due to the fact that the seasonal variations are not that prevalent in urban travel demand. But inter-city travel is affected by seasonal variations.
- 85% and 99% of intra-city and inter-city data respectively were non-normal. Therefore data transformation was required to make these data normal for further analysis.
- 90% of the intra-city routes were having nonlinearity while all the routes of the inter-city were nonlinear. Hence, nonlinear modelling was required to model it.
- The average MAPE value of ARIMA model for 30-days forecast horizon is 12.77%, and that of 7-days forecast horizon is 12.73% for intra-city data. In the case of inter-city data, the average MAPE value of 30 days ARIMA forecast was 17.68% while for 7 days forecast it is 22.24%. Despite the fact that the inter-city data contains three times more data points than the intra-city data, the average MAPE value of intra-city data was comparatively on a lower side than the inter-city data. It can be attributed to the existence of more nonlinearity in the inter-city data which have to be modelled using nonlinear modelling techniques.
- Also, for inter-city data, the monthly forecasts of the ARIMA model have an average MAPE value of 48.49% while the average MAPE value of quarterly forecast is 31.26%. The high value of the monthly and quarterly data points suggest that the 40 data points taken for monthly ARIMA model and 15 data

points taken for quarterly modelling were insufficient to capture the patterns in the data. Due to the lack of enough length of intra-city data, only daily level aggregation was done for intra-city demand modelling.

- The ARNN model has an average MAPE value of 14.18% and 12.82% for 30-days and 7-days forecast horizon respectively for intra-city data. Since the inter-city daily data had conditional variance (ARCH effect), ARNN was not performed on it. In the case of monthly data of inter-city, ARNN model gave an average MAPE of 24.06% while for quarterly data, ARNN model gave an average of 21.92% MAPE value.
- In the case of hybrid ARIMA-ARNN model for intra-city, the forecast accuracy has been increased with an average MAPE value of 9.16% and 7.98% for 30-days and 7-days forecast horizon respectively. For inter-city data, the hybrid ARIMA-ARNN model had average MAPE value as 10.10% for monthly model while the quarterly model had 25.62% as average MAPE. The lower MAPE values of the hybrid model in the case of inter-city is due to the lack of enough data points to capture the patterns and characteristics of time-series data.
- The hybrid ARIMA-GARCH model for inter-city yielded a MAPE value of 8.42% and 7.75% for 30 days and 7 days forecast horizon, respectively. The forecast accuracy for hybrid ARIMA-GARCH model for intra-city was found to be 9.40% for 30 days forecast horizon and 7.50% for 7 days forecast horizon. Here also, it can be noted that despite having three times more data points intra-city, the ARIMA models perform better than the inter-city models. It is due to the reason that the 66% of the ARIMA model exhibit lack of fit of ARIMA model in the case of inter-city data, *i.e.*, further information could be extracted from the residuals to improve the model. Therefore, GARCH was used to extract the information from the residuals.

5.11. SUMMARY

This chapter explains the results and finding of the study with reference to the main aim of the study, which is to develop an “ETM based Bus Transport Demand Model”. The level of service characteristics, accessibility and performance evaluation of Kerala SRTC are discussed. The analysis of the intra-city and inter-city time-series ETM data,

development of various models and the forecasting of the passenger demand for the selected forecast horizon are explained in detail.

- In Trivandrum, the presence of organised public transport system in the urban area is found to be 80% while the availability or extent of public transport is moderate.
- The Trivandrum city has good service coverage, *i.e.*, the public transport ply on most of the corridors in the urban area. But the average waiting time is more than 5 minutes.
- It can be concluded that the city has a good public transport system. The facilities are available wide-spread and are available to the people. Thus, the city has a comfortable public transport system.
- A methodology to determine public transport accessibility was developed in this study. The accessibility index, which measures the public transport accessibility takes into account the network-based distance to the bus stops, population, travel time and opportunities or the land use destinations. The GIS-based public transport accessibility index gives the accessibility of a single point (*i.e.*, the bus stop) by considering access to the point, access from the point to various land use destinations and network impedance.
- The study presented a weighted goal programming methodology integrated with analytical hierarchy process for performance optimisation of public transport undertakings. The user's and operator's perception was included in the study.
- Various time-series techniques were used to model the passenger demand of intra-city (Kerala SRTC) and inter-city (Karnataka SRTC).
- ARIMA, ARNN, Hybrid ARIMA-ARNN ana Hybrid ARIMA-GARCH methods were used.
- The linear conditional mean was modelled using ARIMA while the conditional variance in the residuals of ARIMA was modelled using GARCH modelling technique.
- The nonlinearity in the data was modelled using ARNN while the nonlinearity in the residuals of the ARIMA model was modelled using hybrid ARIMA-ARNN.

- The average MAPE value of ARIMA model for 30-days forecast horizon is 12.77%, and that of 7-days forecast horizon is 12.73% for intra-city data. The ARNN model has an average MAPE value of 14.18% and 12.82% for 30-days and 7-days forecast horizon respectively for intra-city data. In the case of hybrid ARIMA-ARNN model for intra-city, the forecast accuracy has been increased with an average MAPE value of 9.16% and 7.98% for 30-days and 7-days forecast horizon respectively. The forecast accuracy for hybrid ARIMA-GARCH model for intra-city was found to be 9.40% for 30 days forecast horizon and 7.50% for 7 days forecast horizon.

In the case of inter-city data, the average MAPE value of 30 days ARIMA forecast was 17.68% while for 7 days forecast it is 22.24%. Also, the monthly forecasts of the ARIMA model have an average MAPE value of 48.49% while the average MAPE value of quarterly forecast is 31.26%. In the case of monthly data, ARNN model gave an average MAPE of 24.06%, while for quarterly data, ARNN model gave an average of 21.92% MAPE value. The hybrid ARIMA-GARCH model for inter-city data yielded a MAPE value of 8.42% and 7.75% for 30 days and 7 days forecast horizon, respectively.

CHAPTER-6

CONCLUSION

6.1. GENERAL

This chapter summarises the research. The data sources of the data, data collection methodology, data processing and data analysis are summarised. The findings of the evaluation of Level-of-Service of Trivandrum city, determination of accessibility index for Trivandrum public bus transport, performance evaluation of Kerala SRTC, variation indices and time series modelling using intra-city ETM data of Kerala SRTC and inter-city Karnataka SRTC are stated in the following sections.

6.2. CONCLUSION

For Kerala SRTC, performance evaluation and public bus transport demand modelling was planned. During the analysis of the Kerala ETM data, it was found that the length of intra-city ETM data available for individual routes varies from 1.5 years to 2 years (2011 to 2013), while inter-city data were not sufficient (3 months to 5 months data was provided) for time-series modelling. Therefore, inter-city ETM data of Mangalore division Karnataka SRTC was collected for a period of 2013 to 2018, which comprises of a total five-year data, to perform inter-city passenger demand modelling. The intra-city ETM data for Mangalore was not collected, since the intra-city services are mainly operated by private operators. The private operators have not maintained and stored the ETM data. The conclusions drawn from this research work are listed below. The conclusion from data collection and processing, LoS of Trivandrum city public transport, accessibility of public transport in Trivandrum city, performance evaluation of Kerala SRTC, variation indices and demand modelling using intra-city ETM data of Kerala SRTC and inter-city Karnataka SRTC are listed below.

6.2.1. Data Collection and Processing

- The data used for in this study are mainly ETM data from Kerala SRTC for a period 2011 to 2013. In addition to ETM data provided by Kerala SRTC, this

study also uses data provided by various state and central government agencies. In the data collection stage, the intra-city and inter-city ETM data for 10 years were requested from Kerala SRTC. Due to unavailability of 10-year ETM data, the ticketing details of all the buses and routes of intra-city were collected from the respective depots for the period between 2011 and 2013.

- During the analysis of the Kerala ETM data, it was found that the length of data available for individual routes varies from 1.5 years to two years, which was barely sufficient for time-series modelling. Therefore, ETM data from Karnataka SRTC was collected for a period of 2013 to 2018, which comprises of a total five-year data. This was also due to unavailability of 10-year ETM data. Also, the intra-city ETM data for Mangalore was not considered since the share of private buses is more in Mangalore compared to Karnataka SRTC buses. The methodology employed for Kerala SRTC passenger demand estimation is applied to the Karnataka SRTC ETM data to analyse the intra-city and inter-city passenger demand.
- The intra-city ETM data collected from Kerala SRTC was in the form of text file. A single ETM data text file represents all the trips made by a single bus in a day. The raw ETM data in text format was converted into Microsoft Excel format. The converted Excel files were then merged based on the date. The merged file is split into different sheets on the basis of route code which gives the data according to routes for a single day.
- The inter-city data obtained from Karnataka SRTC was in an Excel Format. The Karnataka SRTC ETM data is more detailed and systematic. This data requires only limited data processing, which requires simple Excel operations.

6.2.2. Evaluation of Level of Service of Kerala SRTC

- The Level-of-Service provided by the public transportation facility was determined using the service level benchmarks set up the Ministry of Urban Development (2009). The indicators for determining the Level-of-Service used in the study are the presence of organised public transport system, availability

or extent of public transport supply, service coverage, average waiting time, level of comfort in public transport and the fleet availability.

- The presence of organised public transport system in the urban area is found to be 80% while the availability or extent of public transport is moderate.
- The Trivandrum city has an LoS of '1' which suggests that it has a good service coverage, *i.e.*, the public transport ply on most of the corridors in the urban area. But the average waiting time is more than 5 minutes.
- It can be concluded that Trivandrum city has a good public transport system. The facilities are available wide-spread and are available to the people. Thus, the city has a comfortable public transport system.

6.2.3. Determination of Accessibility Index

- The proposed public transport accessibility index takes into account the network-based distance to the bus stops, population, travel time and opportunities or the land use destinations.
- The methodology was based on the factor that the index should measure the accessibility which comes from proximity to bus stops and land use destinations and the proportion of the population served. This GIS-based methodology for public transport accessibility index gives the accessibility of a single point (*i.e.*, the bus stop) by considering access to the point, access from the point to various land use destinations and network impedance.
- The accessibility of about 6.55% of bus stops are very poor, which serves about 2.36% of the population. 1.19% bus stops serving 1.47% population have very good accessibility. Overall, around 45% of the population has moderate accessibility.
- It was observed that the accessibility index in the central area has a higher value and the accessibility values decrease to the periphery of the city. It can be interpreted that the availability of opportunities is higher in the central area, while the number of opportunities is less in other areas. Moreover, the density of the road network is more towards the city centre, which contributes to higher accessibility values. Also, given that the concentration of population is high in

the city centre and the higher number of bus stops; the accessibility index tends to be more in these areas.

- The visual representation of the composite public transport accessibility helps in identifying the extend remoteness in Trivandrum city. Thus, it is useful in the route planning by introducing new buses in the high-density zones, add missing links in the network and to increase the accessibility in the city outskirts.
- Given the ease of calculating and updating the accessibility index. Overall, this methodology has the potential to become a useful decision support tool for transit operators and planners.

6.2.4. Performance Evaluation of Kerala SRTC

- The study presented a weighted goal programming methodology integrated with analytical hierarchy process for performance optimisation of public transport undertakings. The user's and operator's perception was included in the study.
- The operator cost and staff per schedule were the most important variables in regard to the operator while the safety of travel with respect to user perception has the highest weightage.
- The optimal solution indicates that increasing the accessibility, safety and regularity attract passengers to the public transport that in turn improves the load factor which influences the operators to maximise the fleet utilisation and reduce the cancellation of schedules.
- The proposed model solution suggests decreasing the staff per bus ratio will further reduce the staff cost and hence the operating cost.
- With AHP, the planners can consistently integrate the opinions of all the decision makers and effectively evaluate the scenarios and optimise the resources.

6.2.5. Determination of Variation Indices

- The variation indices of both Trivandrum and Mangalore revealed a few interesting features of these study areas. The data from Trivandrum is intra-city bus passenger demand data while that of Mangalore was inter-city demand.

- The daily variation index of Trivandrum shows that the demand for the Mondays corresponds to more travel with Sunday being the lowest. In the case of Mangalore, the daily variation is having a regular pattern of fluctuation with Sunday being the lowest. This suggests that the demand for travel during the weekend is less compared to weekdays.
- The quarterly variation index of Trivandrum was almost constant without any pattern revealing that the demand for intra-city travel remains unaltered due to quarter changes. But in the inter-city Mangalore data, third quarter shows a rise in the demand. This is due to the fact that the third quarter contains festivals in the state urging the people to travel to home town and visiting relatives. Also, the weather conditions are good which is favourable for travel in bus.
- The monthly variation index of Trivandrum shows that the school working months are having slightly more demand compared to vacations. This slight change is mainly because of school going children using the Kerala SRTC buses during school days. The monthly pattern of travel in Mangalore is similar to the quarterly pattern of Mangalore with September on the higher side. This is due to the festival season in the state.
- The weekly travel variation is not predominant in the case of Trivandrum. But in the case of Mangalore, it can be seen that there is a steady rise in travel from week 36 to week 40 due to festivities in the state (Week 01 is first week of January).

6.2.6. Development of Time-Series Models

- Intra-city and inter-city bus passenger demand modelling using time-series methods were performed using Kerala SRTC (for Trivandrum) and Karnataka SRTC (Mangalore) ETM data, respectively. The intra-city passenger demand modelling was not performed for Mangalore study area as the share of private buses in Mangalore is more than the Karnataka SRTC buses.
- Various time-series techniques were used to model the passenger demand. ARIMA, ARNN, Hybrid ARIMA-ARNN ana Hybrid ARIMA-GARCH methods were used.

- It was observed that one-week lagged seasonal differencing when applied to the time series data was an appropriate modelling technique to be used for forecasting bus passenger demand. In other words, a seasonal differencing of 7 was found to be appropriate for daily passenger demand modelling. The temporal effect, such as daily seasonality was considered in the model when the seven-day lag period was considered.
- The linear conditional mean was modelled using ARIMA while the conditional variance in the residuals of ARIMA was modelled using GARCH modelling technique.
- The nonlinearity in the data was modelled using ARNN while the nonlinearity in the residuals of the ARIMA model was modelled using hybrid ARIMA-ARNN.
- The average MAPE value of ARIMA model for 30-days forecast horizon is 12.77%, and that of 7-days forecast horizon is 12.73% for intra-city data. In the case of inter-city data, the average MAPE value of 30 days ARIMA forecast was 17.68% while for 7 days forecast it is 22.24%. Despite the fact that the inter-city data contains three times more data points than the intra-city data, the average MAPE value of intra-city data was comparatively on a lower side than the inter-city data. It can be attributed to the existence of more nonlinearity in the inter-city data which have to be modelled using nonlinear modelling techniques.
- Also, for inter-city data, the monthly forecasts of the ARIMA model have an average MAPE value of 48.49% while the average MAPE value of quarterly forecast is 31.26%. The high value of the monthly and quarterly data points suggests that the 40 data points taken for monthly ARIMA model and 15 data points taken for quarterly modelling were insufficient to capture the patterns in the data. Due to the lack of enough length of intra-city data, only daily level aggregation was done for intra-city demand modelling.
- The ARNN model has an average MAPE value of 14.18% and 12.82% for 30-days and 7-days forecast horizon respectively for intra-city data. Since the inter-city daily data had conditional variance (ARCH effect), ARNN was not performed on it. In the case of monthly data of inter-city, ARNN model gave an

average MAPE of 24.06% while for quarterly data, ARNN model gave an average of 21.92% MAPE value.

- In the case of hybrid ARIMA-ARNN model for intra-city, the forecast accuracy has been increased with an average MAPE value of 9.16% and 7.98% for 30-days and 7-days forecast horizon respectively. For inter-city data, the hybrid ARIMA-ARNN model had average MAPE value as 10.10% for monthly model while the quarterly model had 25.62% as average MAPE. The lower MAPE values of the hybrid model in the case of inter-city is due to the lack of enough data points to capture the patterns and characteristics of time-series data.
- The hybrid ARIMA-GARCH model for inter-city yielded a MAPE value of 8.42% and 7.75% for 30 days and 7 days forecast horizon, respectively. The forecast accuracy for hybrid ARIMA-GARCH model for intra-city was found to be 9.40% for 30 days forecast horizon and 7.50% for 7 days forecast horizon. Here also, it can be noted that despite having three times more data points intra-city, the ARIMA models perform better than the inter-city models. It is due to the reason that the 66% of the ARIMA model exhibit lack of fit of ARIMA model in the case of inter-city data, i.e., further information could be extracted from the residuals to improve the model. Therefore, GARCH was used to extract the information from the residuals.
- The present study has established the effectiveness of using ETM data for public bus transport demand modelling.
- The univariate time-series ARIMA models improved using ARNN or GARCH methods prove to be good in case of short term demand forecasting, which can be used for operational planning of public transport. This method can be used for the whole system at the aggregate level or for a single route at an individual level depending on the level of information available.

6.3. RECOMMENDATIONS

- The maintenance and storage of the ETM data are highly recommended. This data contains a lot of information that can be extracted for understanding the existing scenario of public transport operations. Analysing the data helps in

various policy making for improving public transport operations. Also, this data is required for doing research in the field of public transport.

- Also, it is recommended that the 53 SRTUs should maintain the ETM data in a common format which can be managed by Association of State Road Transport Undertakings. This will help in the demand analysis for bus transport in a more systematic way.
- Increase the number of buses by changing the size of buses and reduce the waiting time. This will further improve the LoS of public transport, which will attract more passengers to public transport and maintain a good load factor on the buses.
- Increasing the accessibility to public transport, and by maintaining the schedules properly will attract passengers to public transport.
- Also, it is recommended to maximise the fleet utilisation and reduce the cancellation of schedules in Kerala SRTC to improve the LoS, which will thus attract the passengers and helps to improve the profitability of the operator.
- The demand for Kerala SRTC is declining due to poor operational strategies. The percentage of cancelled kilometre to scheduled kilometre was 20% even though the staff per bus ratio is 8.85. Also, the regularity of services was found to be 80% with 80% fleet utilization. This is despite the fact that Kerala SRTC is dominating the public transport services in Trivandrum city. These factors lead to the degeneration in the quality of service offered by Kerala SRTC, thus causes the decline in the demand for Kerala SRTC buses and encourages the use of the private mode of transport. Therefore, measures have to be taken to reduce the cancelled kilometre and increase the fleet utilisation and thus increase the reliability and regularity of Kerala SRTC buses.

6.4. SCOPE OF FUTURE WORK

- The quarterly, monthly and yearly forecasting can be performed exploring the historical ETM data.
- Multivariate time-series modelling can be explored to determine the effect of exogenous variables on public transport demand.

- The time-series methodology can be applied to model the passenger demand for metro rail, bus rapid transit system and railways, which are having more systematic data.
- The ARNN model used in the study is a feedforward neural network which has limitations as they pass information in the forward direction only. This can be improved by using Recurrent Neural Network (RNN) models. RNN is a promising field of time-series research which can be used for passenger demand modelling using ETM data. Also, the further studies can be done using the fuzzy time-series for modeling public transport demand.
- Suitable data analysis and modelling interfaces are to be developed for helping decision-makers.

APPENDIX

A.1. Table Showing the Ward-wise Population of Trivandrum City

Ward Number	Ward name	Population 2011	Population 2013
1	Kazhakuttom	10528	21498.18
2	Chanthavila	9351	19094.74
3	Kaattayikonam	9229	18845.62
4	Sreekariyam	10455	21349.11
5	Cheruvaikkal	9525	19450.05
6	Ulloor	10528	21498.18
7	Edavacode	8659	17681.68
8	Chellamangalam	10036	20493.51
9	Chempazhanthy	10184	20795.73
10	Powdikonam	9676	19758.39
11	Njandoorkonam	8501	17359.04
12	Kinavoor	10545	21532.89
13	Mannanthala	9357	19106.99
14	Nalanchira	8777	17922.63
15	Kesavadasapuram	9102	18586.28
16	Medical college	10545	21532.89
17	Pattom	9761	19931.96
18	Muttada	9856	20125.95
19	Kudappanakunnu	9512	19423.5
20	Pathirappally	9951	20319.94
21	Chettivilakam	8659	17681.68
22	Sasthamangalam	10496	21432.83
23	Kowdiar	10020	20460.84
24	Kuravankonam	9840	20093.28
25	Nanthancode	9676	19758.39
26	Kunnukuzhi	10399	21234.76
27	Palayam	9498	19394.92
28	Thycaud	9184	18753.73
29	Vazhuthacaud	9512	19423.5
30	Kanjirampara	10069	20560.9
31	Peroorkada	10490	21420.58
32	Thuruthumoola	10027	20475.13
33	Nettayam	10509	21459.38
34	Kachani	9328	19047.78
35	Vazhottukonam	10371	21177.58
36	Vattiyoorkavu	9807	20025.89
37	Kodunganoor	10348	21130.62
38	PTP Nagar	10250	20930.5
39	Pangode	9184	18753.73

40	Thirumala	10004	20428.17
41	Valiyavila	10200	20828.4
42	Poojappura	9905	20226.01
43	Valiyasala	9348	19088.62
44	Jagathy	8644	17651.05
45	Karamana	8774	17916.51
46	Arannoor	9872	20158.62
47	Mudavanmugal	10004	20428.17
48	Thrikkannapuram	9020	18418.84
49	Nemom	9856	20125.95
50	Ponnumangalam	9950	20317.9
51	Punnakkamugal	10007	20434.29
52	Pappanamcode	10168	20763.06
53	Estate	9872	20158.62
54	Nedumcaud	9856	20125.95
55	Kalady	10528	21498.18
56	Melamcode	8651	17665.34
57	Punchakkari	10384	21204.13
58	Poonkulam	10168	20763.06
59	Venganoor	10250	20930.5
60	Mulloor	9696	19799.23
61	Kottappuram	9778	19966.68
62	Vizhinjam	10610	21665.62
63	Harbar	8724	17814.41
64	Vellar	8692	17749.06
65	Thiruvallam	9758	19925.84
66	Poonthura	10332	21097.94
67	Ambalathara	8651	17665.34
68	Kamaleswaram	8659	17681.68
69	Kalippankulam	8675	17714.35
70	Attukal	8708	17781.74
71	Chala	8699	17763.36
72	Manacaud	8692	17749.06
73	Kuryathi	8660	17683.72
74	Puthenpally	9266	18921.17
75	Manikyavilakam	8692	17749.06
76	Beemapally east	8315	16979.23
77	Beemapally	8642	17646.96
78	Muttathara	8790	17949.18
79	Sreevaraham	10528	21498.18
80	Fort	8659	17681.68
81	Thampanoor	8888	18149.3
82	Vanchiyur	8856	18083.95
83	Sreekandeswaram	9790	19991.18
84	Perumthanni	9676	19758.39

85	Palkulangara	8941	18257.52
86	Chackai	9754	19917.67
87	Valiyathura	8856	18083.95
88	Vallakkadavu	8790	17949.18
89	Sanghumugham	10171	20769.18
90	Vettucaud	10351	21136.74
91	Karikkakom	9761	19931.96
92	Kadakampally	9908	20232.14
93	Pettah	10089	20601.74
94	Kannammoola	9955	20328.11
95	Anamugham	9872	20158.62
96	Akkulam	8856	18083.95
97	Kulathoor	8938	18251.4
98	Attipra	8741	17849.12
99	Poundkadavu	9004	18386.17
100	Pallithura	10046	20513.93

A.2. Sample ETM Text Data

(0281100001 01/01/11 280235 00 . 22580908
00000000 000190150

01 105 00549687 01 02 01 00 00 00 04:18 0000400 T
01 105 00549688 02 05 01 00 00 00 04:22 0000500 T
01 105 00549689 03 05 01 00 00 00 04:28 0000450 T
01 105 00549690 03 05 01 00 00 00 04:30 0000450 T
01 105 00549691 03 05 01 00 00 00 04:32 0000450 T
01 105 00549692 03 05 01 00 00 00 04:33 0000450 T
01 105 00549693 03 05 02 00 00 00 04:34 0000900 T
01 105 00549694 03 05 01 00 00 00 04:35 0000450 T
01 105 00549695 04 05 01 00 00 00 04:38 0000400 T
02 105 00549696 01 03 01 00 00 00 05:00 0000450 T
02 105 00549697 01 02 01 00 00 00 05:01 0000400 T
02 105 00549698 01 02 01 00 00 00 05:01 0000400 T
02 105 00549699 01 02 01 00 00 00 05:01 0000400 T
02 105 00549700 02 05 01 00 00 00 05:06 0000500 T
02 105 00549701 02 05 01 00 00 00 05:07 0000500 T
02 105 00549702 03 05 01 00 00 00 05:13 0000450 T
02 105 00549703 03 05 01 00 00 00 05:14 0000450 T
02 105 00549704 03 05 01 00 00 00 05:14 0000450 T
02 105 00549705 03 05 02 00 00 00 05:17 0000900 T
02 105 00549706 03 05 02 00 00 00 05:17 0000900 T
02 105 00549707 03 05 01 00 00 00 05:19 0000450 T
02 105 00549708 03 05 01 00 00 00 05:19 0000450 T
02 105 00549709 03 05 01 00 00 00 05:19 0000450 T
02 105 00549710 03 05 01 00 00 00 05:21 0000450 T

02 105 00549711 03 05 01 00 00 00 05:22 0000450 T
02 105 00549712 03 05 01 00 00 00 05:22 0000450 T
02 105 00549713 04 05 01 00 00 00 05:22 0000400 T
03 105 00549714 01 03 01 00 00 00 05:42 0000450 T
03 105 00549715 01 02 01 00 00 00 05:43 0000400 T
03 105 00549716 01 03 01 00 00 00 05:43 0000450 T
03 105 00549717 01 02 01 00 00 00 05:44 0000400 T
03 105 00549718 01 03 01 00 00 00 05:46 0000450 T
03 105 00549719 03 05 01 00 00 00 05:57 0000450 T
03 105 00549720 03 05 01 00 00 00 05:57 0000450 T
03 105 00549721 03 05 02 00 00 00 05:57 0000900 T
03 105 00549722 03 05 01 00 00 00 05:58 0000450 T
03 105 00549723 03 05 01 00 00 00 05:58 0000450 T
03 105 00549724 03 05 01 00 00 00 05:58 0000450 T
03 105 00549725 03 05 01 00 00 00 05:59 0000450 T
03 105 00549726 03 05 01 00 00 00 05:59 0000450 T
03 105 00549727 03 05 01 00 00 00 05:59 0000450 T
03 105 00549728 03 05 01 00 00 00 06:00 0000450 T
03 105 00549729 03 05 01 00 00 00 06:00 0000450 T
03 105 00549730 03 05 02 00 00 00 06:00 0000900 T
03 105 00549731 03 05 01 00 00 00 06:00 0000450 T
03 105 00549732 03 05 01 00 00 00 06:02 0000450 T
03 105 00549733 03 05 01 00 00 00 06:03 0000450 T
03 105 00549734 03 05 01 00 00 00 06:03 0000450 T
03 105 00549735 03 05 01 00 00 00 06:04 0000450 T
03 105 00549736 03 05 01 00 00 00 06:04 0000450 T

03 105 00549737 03 05 01 00 00 00 06:04 0000450 T
03 105 00549738 03 05 01 00 00 00 06:04 0000450 T
03 105 00549739 03 05 01 00 00 00 06:05 0000450 T
03 105 00549740 04 05 01 00 00 00 06:06 0000400 T
03 105 00549741 04 05 01 00 00 00 06:06 0000400 T
03 105 00549742 04 05 01 00 00 00 06:07 0000400 T
03 105 00549743 04 05 01 00 00 00 06:09 0000400 T
03 105 00549744 04 05 01 00 00 00 06:09 0000400 T
04 105 00549745 01 02 01 00 00 00 06:36 0000400 T
04 105 00549746 01 02 01 00 00 00 06:37 0000400 T
04 105 00549747 01 02 04 00 00 00 06:38 0001600 T
04 105 00549748 01 02 01 00 00 00 06:38 0000400 T
04 105 00549749 01 02 01 00 00 00 06:39 0000400 T
04 105 00549750 01 03 01 00 00 00 06:40 0000450 T
04 105 00549751 01 02 01 00 00 00 06:40 0000400 T
04 105 00549752 01 03 01 00 00 00 06:41 0000450 T
04 105 00549753 01 02 01 00 00 00 06:41 0000400 T
04 105 00549754 01 02 02 00 00 00 06:42 0000800 T
04 105 00549755 01 02 01 00 00 00 06:42 0000400 T
04 105 00549756 01 03 01 00 00 00 06:42 0000450 T
04 105 00549757 01 02 02 00 00 00 06:43 0000800 T
04 105 00549758 01 02 01 00 00 00 06:43 0000400 T
04 105 00549759 01 02 01 00 00 00 06:44 0000400 T
04 105 00549760 01 02 01 00 00 00 06:44 0000400 T
04 105 00549761 01 02 01 00 00 00 06:44 0000400 T
04 105 00549762 01 02 02 00 00 00 06:45 0000800 T
04 105 00549763 01 04 01 00 00 00 06:45 0000500 T

04 105 00549764 01 02 01 00 00 00 06:46 0000400 T
04 105 00549765 01 03 01 00 00 00 06:47 0000450 T
04 105 00549766 02 03 01 00 00 00 06:53 0000400 T
04 105 00549767 03 05 03 00 00 00 06:58 0001350 T
04 105 00549768 03 05 01 00 00 00 06:59 0000450 T
04 105 00549769 03 05 01 00 00 00 06:59 0000450 T
04 105 00549770 03 05 03 00 00 00 06:59 0001350 T
04 105 00549771 03 05 02 00 00 00 07:00 0000900 T
04 105 00549772 03 05 02 00 00 00 07:00 0000900 T
04 105 00549773 03 05 01 00 00 00 07:00 0000450 T
04 105 00549774 03 05 01 00 00 00 07:00 0000450 T
04 105 00549775 03 05 01 00 00 00 07:00 0000450 T
04 105 00549776 03 05 01 00 00 00 07:01 0000450 T
04 105 00549777 03 05 01 00 00 00 07:01 0000450 T
04 105 00549778 03 05 01 00 00 00 07:01 0000450 T
04 105 00549779 03 05 01 00 00 00 07:02 0000450 T
04 105 00549780 03 05 01 00 00 00 07:02 0000450 T
04 105 00549781 03 05 01 00 00 00 07:02 0000450 T
04 105 00549782 03 05 02 00 00 00 07:03 0000900 T
04 105 00549783 03 05 01 00 00 00 07:03 0000450 T
04 105 00549784 03 05 01 00 00 00 07:04 0000450 T
04 105 00549785 03 05 01 00 00 00 07:05 0000450 T
04 105 00549786 03 05 01 00 00 00 07:05 0000450 T
04 105 00549787 03 05 01 00 00 00 07:05 0000450 T
04 105 00549788 03 05 01 00 00 00 07:06 0000450 T
04 105 00549789 03 05 01 00 00 00 07:06 0000450 T

04 105 00549790 03 05 01 00 00 00 07:06 0000450 T
04 105 00549791 03 05 01 00 00 00 07:07 0000450 T
04 105 00549792 03 05 01 00 00 00 07:07 0000450 T
04 105 00549793 03 05 01 00 00 00 07:07 0000450 T
04 105 00549794 03 05 01 00 00 00 07:08 0000450 T
04 105 00549795 04 05 01 00 00 00 07:09 0000400 T
04 105 00549796 04 05 01 00 00 00 07:11 0000400 T
04 105 00549797 04 05 01 00 00 00 07:16 0000400 T
05 105 00549798 01 03 01 00 00 00 07:21 0000450 T
05 105 00549799 01 02 02 00 00 00 07:21 0000800 T
05 105 00549800 01 03 01 00 00 00 07:22 0000450 T
05 105 00549801 01 02 01 00 00 00 07:23 0000400 T
05 105 00549802 01 02 01 00 00 00 07:23 0000400 T
05 105 00549803 01 02 01 00 00 00 07:23 0000400 T
05 105 00549804 01 02 01 00 00 00 07:23 0000400 T
05 105 00549805 01 02 01 00 00 00 07:24 0000400 T
05 105 00549806 01 02 01 00 00 00 07:24 0000400 T
05 105 00549807 01 02 02 00 00 00 07:24 0000800 T
05 105 00549808 01 03 01 00 00 00 07:25 0000450 T
05 105 00549809 01 02 01 00 00 00 07:25 0000400 T
05 105 00549810 01 02 01 00 00 00 07:26 0000400 T
05 105 00549811 01 02 01 00 00 00 07:27 0000400 T
05 105 00549812 01 02 02 00 00 00 07:27 0000800 T
05 105 00549813 01 02 09 00 00 00 07:27 0003600 T
05 105 00549814 01 02 09 00 00 00 07:27 0003600 T
05 105 00549815 01 02 02 00 00 00 07:27 0000800 T
05 105 00549816 01 02 01 00 00 00 07:28 0000400 T

05 105 00549817 01 02 01 00 00 00 07:28 0000400 T
05 105 00549818 01 02 03 00 00 00 07:28 0001200 T
05 105 00549819 01 02 01 00 00 00 07:29 0000400 T
05 105 00549820 01 02 04 00 00 00 07:29 0001600 T
05 105 00549821 03 05 01 00 00 00 07:42 0000450 T
05 105 00549822 03 05 01 00 00 00 07:43 0000450 T
05 105 00549823 03 05 01 00 00 00 07:43 0000450 T
05 105 00549824 03 05 01 00 00 00 07:44 0000450 T
05 105 00549825 03 04 01 00 00 00 07:44 0000400 T
05 105 00549826 03 05 01 00 00 00 07:44 0000450 T
05 105 00549827 03 05 01 00 00 00 07:45 0000450 T
05 105 00549828 03 05 01 00 00 00 07:45 0000450 T
05 105 00549829 03 05 01 00 00 00 07:45 0000450 T
05 105 00549830 03 05 01 00 00 00 07:45 0000450 T
05 105 00549831 03 05 01 00 00 00 07:45 0000450 T
05 105 00549832 03 05 01 00 00 00 07:45 0000450 T
05 105 00549833 03 05 01 00 00 00 07:46 0000450 T
05 105 00549834 03 05 01 00 00 00 07:47 0000450 T
05 105 00549835 03 05 01 00 00 00 07:47 0000450 T
05 105 00549836 03 05 01 00 00 00 07:47 0000450 T
05 105 00549837 03 05 01 00 00 00 07:48 0000450 T
05 105 00549838 03 05 01 00 00 00 07:48 0000450 T
05 105 00549839 03 05 01 00 00 00 07:48 0000450 T
05 105 00549840 03 05 01 00 00 00 07:49 0000450 T
05 105 00549841 03 05 01 00 00 00 07:50 0000450 T
05 105 00549842 03 05 01 00 00 00 07:50 0000450 T

05 105 00549843 03 05 01 00 00 00 07:50 0000450 T
05 105 00549844 03 05 01 00 00 00 07:50 0000450 T
05 105 00549845 03 05 01 00 00 00 07:51 0000450 T
05 105 00549846 03 05 01 00 00 00 07:51 0000450 T
05 105 00549847 03 05 01 00 00 00 07:51 0000450 T
05 105 00549848 04 05 01 00 00 00 07:51 0000400 T
05 105 00549849 04 05 01 00 00 00 07:55 0000400 T
06 105 00549850 01 02 02 00 00 00 08:03 0000800 T
06 105 00549851 01 03 01 00 00 00 08:04 0000450 T
06 105 00549852 01 03 01 00 00 00 08:04 0000450 T
06 105 00549853 01 02 01 00 00 00 08:06 0000400 T
06 105 00549854 01 02 04 00 00 00 08:07 0001600 T
06 105 00549855 01 02 01 00 00 00 08:07 0000400 T
06 105 00549856 01 02 02 00 00 00 08:08 0000800 T
06 105 00549857 01 02 01 00 00 00 08:08 0000400 T
06 105 00549858 01 03 01 00 00 00 08:08 0000450 T
06 105 00549859 01 03 01 00 00 00 08:08 0000450 T
06 105 00549860 01 02 01 00 00 00 08:10 0000400 T
06 105 00549861 01 02 02 00 00 00 08:10 0000800 T
06 105 00549862 01 02 01 00 00 00 08:10 0000400 T
06 105 00549863 01 02 01 00 00 00 08:10 0000400 T
06 105 00549864 01 02 01 00 00 00 08:11 0000400 T
06 105 00549865 01 02 01 00 00 00 08:11 0000400 T
06 105 00549866 01 02 01 00 00 00 08:11 0000400 T
06 105 00549867 01 04 01 00 00 00 08:12 0000500 T
06 105 00549868 01 02 01 00 00 00 08:12 0000400 T
06 105 00549869 01 02 01 00 00 00 08:13 0000400 T

06 105 00549870 01 03 01 00 00 00 08:17 0000450 T
06 105 00549871 03 05 01 00 00 00 08:25 0000450 T
06 105 00549872 03 05 01 00 00 00 08:26 0000450 T
06 105 00549873 03 05 01 00 00 00 08:26 0000450 T
06 105 00549874 03 05 05 00 00 00 08:26 0002250 T
06 105 00549875 03 05 01 00 00 00 08:27 0000450 T
06 105 00549876 03 05 01 00 00 00 08:27 0000450 T
06 105 00549877 03 05 01 00 00 00 08:28 0000450 T
06 105 00549878 03 04 01 00 00 00 08:28 0000400 T
06 105 00549879 03 05 01 00 00 00 08:29 0000450 T
06 105 00549880 03 05 01 00 00 00 08:29 0000450 T
06 105 00549881 03 05 01 00 00 00 08:30 0000450 T
06 105 00549882 03 05 01 00 00 00 08:30 0000450 T
06 105 00549883 03 05 01 00 00 00 08:30 0000450 T
06 105 00549884 03 05 02 00 00 00 08:31 0000900 T
06 105 00549885 03 05 01 00 00 00 08:33 0000450 T
06 105 00549886 03 05 01 00 00 00 08:33 0000450 T
06 105 00549887 03 05 01 00 00 00 08:33 0000450 T
06 105 00549888 03 05 01 00 00 00 08:33 0000450 T
06 105 00549889 03 05 01 00 00 00 08:34 0000450 T
06 105 00549890 03 05 01 00 00 00 08:34 0000450 T
06 105 00549891 03 05 01 00 00 00 08:35 0000450 T
06 105 00549892 03 05 01 00 00 00 08:35 0000450 T
06 105 00549893 03 05 01 00 00 00 08:36 0000450 T
06 105 00549894 03 05 01 00 00 00 08:36 0000450 T
06 105 00549895 03 05 01 00 00 00 08:36 0000450 T

06 105 00549896 03 05 01 00 00 00 08:37 0000450 T
06 105 00549897 03 05 02 00 00 00 08:37 0000900 T
06 105 00549898 03 05 01 00 00 00 08:37 0000450 T
06 105 00549899 03 05 01 00 00 00 08:37 0000450 T
06 105 00549900 03 05 01 00 00 00 08:37 0000450 T
06 105 00549901 03 05 01 00 00 00 08:38 0000450 T
06 105 00549902 03 05 01 00 00 00 08:38 0000450 T
06 105 00549903 03 05 01 00 00 00 08:38 0000450 T
06 105 00549904 03 05 01 00 00 00 08:38 0000450 T
06 105 00549905 03 05 01 00 00 00 08:38 0000450 T
06 105 00549906 03 05 01 00 00 00 08:39 0000450 T
06 105 00549907 03 05 01 00 00 00 08:39 0000450 T
06 105 00549908 04 05 01 00 00 00 08:39 0000400 T
06 105 00549909 04 05 01 00 00 00 08:39 0000400 T
06 105 00549910 04 05 02 00 00 00 08:39 0000800 T
06 105 00549911 04 05 01 00 00 00 08:42 0000400 T
06 105 00549912 04 05 01 00 00 00 08:43 0000400 T
06 105 00549913 04 05 01 00 00 00 08:48 0000400 T
06 105 00549914 04 05 01 00 00 00 08:48 0000400 T
06 105 00549915 04 05 01 00 00 00 08:48 0000400 T
07 156 00549916 01 02 01 00 00 00 08:52 0000400 T
07 156 00549917 01 03 01 00 00 00 08:53 0000450 T
07 156 00549918 01 03 01 00 00 00 08:53 0000450 T
07 156 00549919 01 02 02 00 00 00 08:54 0000800 T
07 156 00549920 01 02 01 00 00 00 08:56 0000400 T
07 156 00549921 01 03 01 00 00 00 08:56 0000450 T
07 156 00549922 01 02 01 00 00 00 08:57 0000400 T

07 156 00549923 01 03 01 00 00 00 08:57 0000450 T
07 156 00549924 01 02 01 00 00 00 08:58 0000400 T
07 156 00549925 01 03 01 00 00 00 08:58 0000450 T
07 156 00549926 01 02 01 00 00 00 08:58 0000400 T
07 156 00549927 01 02 01 00 00 00 08:58 0000400 T
07 156 00549928 01 02 01 00 00 00 08:59 0000400 T
07 156 00549929 01 02 01 00 00 00 08:59 0000400 T
07 156 00549930 01 02 01 00 00 00 08:59 0000400 T
07 156 00549931 01 02 01 00 00 00 08:59 0000400 T
07 156 00549932 01 04 01 00 00 00 08:59 0000500 T
07 156 00549933 01 02 01 00 00 00 09:00 0000400 T
07 156 00549934 01 03 01 00 00 00 09:00 0000450 T
07 156 00549935 01 03 02 00 00 00 09:01 0000900 T
07 156 00549936 01 03 01 00 00 00 09:01 0000450 T
07 156 00549937 01 03 01 00 00 00 09:02 0000450 T
07 156 00549938 01 03 01 00 00 00 09:02 0000450 T
07 156 00549939 01 02 03 00 00 00 09:02 0001200 T
07 156 00549940 01 02 01 00 00 00 09:04 0000400 T
07 156 00549941 01 03 01 00 00 00 09:04 0000450 T
07 156 00549942 01 02 01 00 00 00 09:05 0000400 T
07 156 00549943 01 02 01 00 00 00 09:05 0000400 T
07 156 00549944 01 02 01 00 00 00 09:05 0000400 T
07 156 00549945 01 02 02 00 00 00 09:06 0000800 T
07 156 00549946 01 03 01 00 00 00 09:06 0000450 T
07 156 00549947 01 03 01 00 00 00 09:06 0000450 T
07 156 00549948 01 03 01 00 00 00 09:06 0000450 T

07 156 00549949 01 03 01 00 00 00 09:07 0000450 T
07 156 00549950 01 03 01 00 00 00 09:07 0000450 T
07 156 00549951 01 02 01 00 00 00 09:08 0000400 T
07 156 00549952 01 02 01 00 00 00 09:08 0000400 T
07 156 00549953 01 03 01 00 00 00 09:08 0000450 T
07 156 00549954 01 03 01 00 00 00 09:13 0000450 T
07 156 00549955 01 03 01 00 00 00 09:14 0000450 T
07 156 00549956 01 03 02 00 00 00 09:14 0000900 T
07 156 00549957 01 03 01 00 00 00 09:14 0000450 T
07 156 00549958 01 03 01 00 00 00 09:15 0000450 T
07 156 00549959 02 03 01 00 00 00 09:16 0000400 T
07 156 00549960 03 05 01 00 00 00 09:22 0000450 T
07 156 00549961 03 04 01 00 00 00 09:23 0000400 T
07 156 00549962 03 05 01 00 00 00 09:23 0000450 T
07 156 00549963 03 05 01 00 00 00 09:23 0000450 T
07 156 00549964 03 07 01 00 00 00 09:23 0000550 T
07 156 00549965 03 06 01 00 00 00 09:23 0000500 T
07 156 00549966 03 07 01 00 00 00 09:24 0000550 T
07 156 00549967 03 06 01 00 00 00 09:24 0000500 T
07 156 00549968 03 06 01 00 00 00 09:24 0000500 T
07 156 00549969 03 08 01 00 00 00 09:24 0000700 T
07 156 00549970 03 05 01 00 00 00 09:24 0000450 T
07 156 00549971 03 05 01 00 00 00 09:25 0000450 T
07 156 00549972 03 06 01 00 00 00 09:25 0000500 T
07 156 00549973 03 05 01 00 00 00 09:25 0000450 T
07 156 00549974 03 04 01 00 00 00 09:25 0000400 T
07 156 00549975 03 06 01 00 00 00 09:27 0000500 T

07 156 00549976 03 05 01 00 00 00 09:28 0000450 T
07 156 00549977 03 07 01 00 00 00 09:28 0000550 T
07 156 00549978 03 05 01 00 00 00 09:29 0000450 T
07 156 00549979 03 05 02 00 00 00 09:30 0000900 T
07 156 00549980 03 05 01 00 00 00 09:30 0000450 T
07 156 00549981 03 08 01 00 00 00 09:31 0000700 T
07 156 00549982 03 05 01 00 00 00 09:31 0000450 T
07 156 00549983 03 05 01 00 00 00 09:32 0000450 T
07 156 00549984 03 05 01 00 00 00 09:32 0000450 T
07 156 00549985 03 06 01 00 00 00 09:33 0000500 T
07 156 00549986 03 05 01 00 00 00 09:33 0000450 T
07 156 00549987 03 05 01 00 00 00 09:34 0000450 T
07 156 00549988 03 05 01 00 00 00 09:34 0000450 T
07 156 00549989 03 05 01 00 00 00 09:34 0000450 T
07 156 00549990 03 05 02 00 00 00 09:35 0000900 T
07 156 00549991 03 05 01 00 00 00 09:35 0000450 T
07 156 00549992 03 05 01 00 00 00 09:35 0000450 T
07 156 00549993 03 05 02 00 00 00 09:36 0000900 T
07 156 00549994 03 06 02 00 00 00 09:36 0001000 T
07 156 00549995 03 05 01 00 00 00 09:36 0000450 T
07 156 00549996 03 06 01 00 00 00 09:36 0000500 T
07 156 00549997 03 07 01 00 00 00 09:36 0000550 T
07 156 00549998 03 06 01 00 00 00 09:37 0000500 T
07 156 00549999 03 06 01 00 00 00 09:37 0000500 T
07 156 00550000 04 05 01 00 00 00 09:39 0000400 T
07 156 00550001 04 06 02 00 00 00 09:42 0000900 T

07 156 00550002 04 05 01 00 00 00 09:42 0000400 T
07 156 00550003 04 06 01 00 00 00 09:42 0000450 T
07 156 00550004 04 06 01 00 00 00 09:43 0000450 T
07 156 00550005 04 05 01 00 00 00 09:43 0000400 T
07 156 00550006 04 07 01 00 00 00 09:44 0000500 T
07 156 00550007 04 06 01 00 00 00 09:44 0000450 T
07 156 00550008 04 05 01 00 00 00 09:44 0000400 T
07 156 00550009 04 05 01 00 00 00 09:45 0000400 T
07 156 00550010 04 07 01 00 00 00 09:45 0000500 T
07 156 00550011 04 05 01 00 00 00 09:46 0000400 T
07 156 00550012 05 06 01 00 00 00 09:50 0000400 T
07 156 00550013 05 06 01 00 00 00 09:51 0000400 T
07 156 00550014 05 06 01 00 00 00 09:51 0000400 T
07 156 00550015 05 06 01 00 00 00 09:51 0000400 T
07 156 00550016 05 07 01 00 00 00 09:54 0000450 T
07 156 00550017 06 08 01 00 00 00 10:00 0000450 T
07 156 00550018 06 08 01 00 00 00 10:01 0000450 T
07 156 00550019 06 07 01 00 00 00 10:04 0000400 T
08 017 00550020 04 01 03 00 00 00 10:12 0001500 T
08 017 00550021 04 02 01 00 00 00 10:12 0000450 T
08 017 00550022 04 02 01 00 00 00 10:12 0000450 T
08 017 00550023 04 02 01 00 00 00 10:13 0000450 T
08 017 00550024 04 02 00 01 00 00 10:13 0000400 T
08 017 00550025 04 01 01 00 00 00 10:13 0000500 T
08 017 00550026 04 01 01 00 00 00 10:13 0000500 T
08 017 00550027 04 02 01 00 00 00 10:14 0000450 T
08 017 00550028 04 01 01 00 00 00 10:14 0000500 T

08 017 00550029 04 02 01 00 00 00 10:16 0000450 T
08 017 00550030 04 01 01 00 00 00 10:16 0000500 T
08 017 00550031 04 02 01 00 00 00 10:17 0000450 T
08 017 00550032 04 02 01 00 00 00 10:17 0000450 T
08 017 00550033 04 01 01 00 00 00 10:17 0000500 T
08 017 00550034 04 02 01 00 00 00 10:17 0000450 T
08 017 00550035 03 02 02 00 00 00 10:19 0000800 T
08 017 00550036 03 02 01 00 00 00 10:19 0000400 T
08 017 00550037 03 01 01 00 00 00 10:19 0000450 T
08 017 00550038 03 01 00 01 00 00 10:19 0000400 T
08 017 00550039 03 01 00 01 00 00 10:20 0000400 T
08 017 00550040 03 01 02 00 00 00 10:20 0000900 T
08 017 00550041 03 01 01 00 00 00 10:20 0000450 T
08 017 00550042 03 01 01 00 00 00 10:21 0000450 T
08 017 00550043 03 02 01 00 00 00 10:21 0000400 T
08 017 00550044 03 01 01 00 00 00 10:22 0000450 T
08 017 00550045 03 02 01 00 00 00 10:22 0000400 T
08 017 00550046 03 01 01 00 00 00 10:22 0000450 T
08 017 00550047 03 01 01 00 00 00 10:23 0000450 T
08 017 00550048 03 01 01 00 00 00 10:23 0000450 T
08 017 00550049 03 02 01 00 00 00 10:23 0000400 T
08 017 00550050 03 01 01 00 00 00 10:23 0000450 T
08 017 00550051 02 01 01 00 00 00 10:29 0000400 T
)

A.3. MATLAB Program to read ETM data text file and concerting it to Microsoft

Excel

```
folder = 'G:\Intrazonal Data\Aryanad\TEXT FILES\250 Files'; % Give the folder
which contains all the text files
write_folder = 'G:\Intrazonal Data\Aryanad\TEXT FILES\250 Files'; % where you
want to write the results
myFiles = dir(fullfile(folder, '*.txt'));
pattern = '\d{2}(\s)*\/(\s)*\d{2}(\s)*\/(\s)*\d{2}'
for k=1:length(myFiles)
    clearvars c;
    fileName = myFiles(k).name;

    fullFileName = fullfile(folder, fileName);

    display(fullFileName);
    fid = fopen(fullFileName);
    data_line = fgetl(fid);
    date = regexp(data_line, pattern, 'match');
    i = 2;
    c(1, :) = {'col1' 'col2' 'col3' 'col4' 'col5' 'col6' 'col7' 'col8' 'col9' 'col10(hrs)'
'col11(min)' 'col12' 'col13'};
    tline = fgetl(fid);
    while ischar(tline)
        elem = strsplit(tline, '\t');
        if size(elem, 2) ~= 12
            % try to split with space
            elem = strsplit(tline, ' ');
            if size(elem, 2) > 12
                elem = elem(1:12);
            end
        end
        if size(elem, 2) == 12
            display(elem);
            counter = 1;
            for index=1:size(elem, 2)
                if index == 10
                    time = elem{10};
                    details = strsplit(time, ':');
                    c(i, 10) = details(1);
                    counter = counter + 1;
                    c(i, 11) = details(2);
                else
                    c(i, counter) = cellstr(elem(index));
                end
                counter = counter + 1;
            end
        end
    end
end
```



```

        i = i + 1;
    end
    tline = fgetl(fid);
end
k = strsplit(fileName, '.');
date = date{1};
date = strrep(date, '/', '-');
date = date(find(~isspace(date)));
excel_filename = strcat(k(1), '-', date, '.xlsx');
%excel_filename = excel_filename(find(~isspace(excel_filename))), ;
full_path = fullfile(write_folder, excel_filename);
xlswrite(full_path{1}, c);
fclose(fid);
clearvars c;
end

```

A.4. MATLAB Program to Merge ETM Data Files of Same Date

```
% Read all the files from a folder and combine the files with the same date on it
folder = 'F:\project\Intercity Data\TVM12TXT\'; % Give the folder which contains all
the text files
write_folder = 'F:\project\Intercity Data\TVM12TXT\'; % where you want to write the
results
pattern = '\d{2}-\d{2}-\d{2}\.';
myFiles = dir(fullfile(folder, '*.xlsx'));
c = containers.Map;
% merge the files
for k=1:length(myFiles)
    file = myFiles(k).name;
    %display(file);
    date = regexp(file, pattern, 'match');
    if isempty(date)
        continue
    end
    date = date{1};
    date = date(1:length(date) - 1);
    if c.isKey(date)
        m = c(date);
    else
        c(date) = containers.Map;
        m = c(date);
    end
    m(file) = 1;
end
display(c(date).keys());
keys = c.keys();
for i=1:length(keys)
    files = c(keys{i}).keys();
    % merge to one file
    %header(1, :) = {'col1' 'col2' 'col3' 'col4' 'col5' 'col6' 'col7' 'col8' 'col9' 'col10(hrs)'
'col11(min)' 'col12' 'col13'};
    first_time = true;
    for j=1:length(files)
        file_name = files(j);
        file_name = file_name{1};
        fullFileName = fullfile(folder, file_name);
        [num, text, raw] = xlsread(fullFileName);
        if first_time
            res = raw
            first_time = false;
        else
            new_raw = raw(2:length(raw), :); % Removing the headers
            res = [res; new_raw];
        end
    end
end
```

```
    end
end
% Now write to a single file
excel_filename = strcat(keys{i}, '_combined', '.xlsx');
full_path = fullfile(write_folder, excel_filename);
xlswrite(full_path, res);
clearvars res;
end
```

A.5. MATLAB Program to Split the Merged Files According to Route Number

```
clear all;
folder='G:\Intrazonal Data\Aryanad\Daily Excel - Copy';
write_folder='G:\Intrazonal Data\Aryanad\New folder (2)';
myFiles = dir(fullfile(folder,'*.xlsx'));

for k=1:length(myFiles)

    filename=myFiles(k).name;
    display(filename);

    fullfilename = fullfile(folder,filename);
    [num,txtdata,raw] = xlsread(fullfilename);%display(txtdata);

    fare_stage=num(:,2);

    N=unique(fare_stage);
    [x,y]=hist(fare_stage,N);
    display(y);
    display(x);
    for i=1:length(N)

        for m=1:x(i)

            for n=1:length(fare_stage)

                if N(i)==fare_stage(n)
                    res(m,:)=raw(n,:);
                end

            end

        end

        display(res);
        %firstrow(1,:)={'Trip no' 'Fare stage' 'Ticket no' 'Origin' 'Destination' 'Full ticket'
'Half ticket' 'Luggage' 'Physically handicapped' 'Ticket time(hrs)' 'Ticket time(min)'
'Ticket rate' 'Ticket type'};
        %new_res=[firstrow;res];
        s=num2str(N(i));
        write_filename=strcat(s,'-',filename); display(write_filename);
        fullpath=fullfile(write_folder,write_filename);
        xlswrite(fullpath,res);
    end

end

end
```

A.6. Microsoft Excel Macro To Split Worksheets According to the Route Number

Sub Copy_To_Worksheets()

'Note: This macro use the function LastRow

Dim My_Range As Range

Dim FieldNum As Long

Dim CalcMode As Long

Dim ViewMode As Long

Dim ws2 As Worksheet

Dim Lrow As Long

Dim cell As Range

Dim CCount As Long

Dim WSNew As Worksheet

Dim ErrNum As Long

'Set filter range on ActiveSheet: A1 is the top left cell of your filter range

'and the header of the first column, D is the last column in the filter range.

'You can also add the sheet name to the code like this :

'Worksheets("Sheet1").Range("A1:D" & LastRow(Worksheets("Sheet1")))

'No need that the sheet is active then when you run the macro when you use this.

Set My_Range = Range("A1:DA" & LastRow(ActiveSheet))

My_Range.Parent.Select

If ActiveWorkbook.ProtectStructure = True Or _

My_Range.Parent.ProtectContents = True Then

MsgBox "Sorry, not working when the workbook or worksheet is protected", _
vbOKOnly, "Copy to new worksheet"

Exit Sub

End If

'This example filters on the first column in the range(change the field if needed)

'In this case the range starts in A so Field:=1 is column A, 2 = column B,

FieldNum = 3

'Turn off AutoFilter

My_Range.Parent.AutoFilterMode = False

'Change ScreenUpdating, Calculation, EnableEvents,

With Application

CalcMode = .Calculation

.Calculation = xlCalculationManual

.ScreenUpdating = False

.EnableEvents = False

End With

ViewMode = ActiveWindow.View

ActiveWindow.View = xlNormalView

ActiveSheet.DisplayPageBreaks = False

'Add a worksheet to copy the a unique list and add the CriteriaRange

Set ws2 = Worksheets.Add

With ws2

```

'first we copy the Unique data from the filter field to ws2
My_Range.Columns(FieldNum).AdvancedFilter _
    Action:=xlFilterCopy, _
    CopyToRange:=.Range("A1"), Unique:=True
'loop through the unique list in ws2 and filter/copy to a new sheet
Lrow = .Cells(Rows.Count, "A").End(xlUp).Row
For Each cell In .Range("A2:A" & Lrow)
    'Filter the range
    My_Range.AutoFilter Field:=FieldNum, Criteria1:="=" & _
    Replace(Replace(Replace(cell.Value, "~", "~~"), "*", "~*"), "?", "~?")
    'Check if there are no more then 8192 areas(limit of areas)
    CCount = 0
    On Error Resume Next
    CCount = My_Range.Columns(1).SpecialCells(xlCellTypeVisible) _
    .Areas(1).Cells.Count
    On Error GoTo 0
    If CCount = 0 Then
        MsgBox "There are more than 8192 areas for the value : " & cell.Value _
        & vbNewLine & "It is not possible to copy the visible data." _
        & vbNewLine & "Tip: Sort your data before you use this macro.", _
        vbOKOnly, "Split in worksheets"
    Else
        'Add a new worksheet
        Set WSNew = Worksheets.Add(After:=Sheets(Sheets.Count))
        On Error Resume Next
        WSNew.Name = cell.Value
        If Err.Number > 0 Then
            ErrNum = ErrNum + 1
            WSNew.Name = "Error_" & Format(ErrNum, "0000")
            Err.Clear
        End If
        On Error GoTo 0
        'Copy the visible data to the new worksheet
        My_Range.SpecialCells(xlCellTypeVisible).Copy
        With WSNew.Range("A1")
            ' Paste:=8 will copy the columnwidth in Excel 2000 and higher
            ' Remove this line if you use Excel 97
            .PasteSpecial Paste:=8
            .PasteSpecial xlPasteValues
            .PasteSpecial xlPasteFormats
            Application.CutCopyMode = False
            .Select
        End With
    End If
    'Show all data in the range
    My_Range.AutoFilter Field:=FieldNum
Next cell

```

```

'Delete the ws2 sheet
On Error Resume Next
Application.DisplayAlerts = False
.Delete
Application.DisplayAlerts = True
On Error GoTo 0
End With
'Turn off AutoFilter
My_Range.Parent.AutoFilterMode = False
If ErrNum > 0 Then
    MsgBox "Rename every WorkSheet name that start with ""Error_"" manually" _
        & vbNewLine & "There are characters in the name that are not allowed" _
        & vbNewLine & "in a sheet name or the worksheet already exist."
End If
'Restore ScreenUpdating, Calculation, EnableEvents, ....
My_Range.Parent.Select
ActiveWindow.View = ViewMode
With Application
    .ScreenUpdating = True
    .EnableEvents = True
    .Calculation = CalcMode
End With
End Sub
Function LastRow(sh As Worksheet)
    On Error Resume Next
    LastRow = sh.Cells.Find(What:="*", _
        After:=sh.Range("A1"), _
        Lookat:=xlPart, _
        LookIn:=xlValues, _
        SearchOrder:=xlByRows, _
        SearchDirection:=xlPrevious, _
        MatchCase:=False).Row
    On Error GoTo 0
End Function

```

A.7. MATLAB Program to Merge the Excel Files According to the Route Number

```
clear all;
folder = 'F:\City 2012 Converted\New folder\New folder\';
write_folder = 'F:\City 2012 Converted\New folder\New folder\';
Files = dir(fullfile(folder,'*.xlsx'));

for k=1:length(Files)
    filename = Files(k).name; display(filename);
    fullFileName = fullfile(folder,filename);
    if k==1
        [~,~,row]=xlsread(fullfile(fullFileName));
        res = row(1:size(row, 1), :);
    end
    [~,~,row]=xlsread(fullfile(fullFileName));
    new_row = row(1:size(row, 1), :);
    res= [res;new_row];
    clearvars new_row;
end
first_row(1,:) = {'Trip no' 'Fare stage' 'Ticket no' 'Origin' 'Destination' 'Full ticket'
'Half ticket' 'Luggage' 'Physically handicaped' 'Ticket time(hrs)' 'Ticket time(min)'
'Ticket rate' 'Ticket type'};
new_res=res(2:size(res, 1), :);
full_file=[first_row;new_res];
%display(full_file);
excel_filename = strcat('996', '.xlsx');
full_path = fullfile(write_folder,excel_filename);
xlswrite(full_path,full_file);
```


A.8. The Recompiled Route-wise ETM Data

Trip Number	Route Number	Ticket Number	Origin	Destination	Full Ticket	Half Ticket	Luggage	Physically Challenged	Time (hrs)	Time (min)	Fare	Ticket Type
1	105	549687	1	2	1	0	0	0	4	18	400	T
1	105	549688	2	5	1	0	0	0	4	22	500	T
1	105	549689	3	5	1	0	0	0	4	28	450	T
1	105	549690	3	5	1	0	0	0	4	30	450	T
1	105	549691	3	5	1	0	0	0	4	32	450	T
1	105	549692	3	5	1	0	0	0	4	33	450	T
1	105	549693	3	5	2	0	0	0	4	34	900	T
1	105	549694	3	5	1	0	0	0	4	35	450	T
1	105	549695	4	5	1	0	0	0	4	38	400	T
2	105	549696	1	3	1	0	0	0	5	0	450	T
2	105	549697	1	2	1	0	0	0	5	1	400	T
2	105	549698	1	2	1	0	0	0	5	1	400	T
2	105	549699	1	2	1	0	0	0	5	1	400	T
2	105	549700	2	5	1	0	0	0	5	6	500	T
2	105	549701	2	5	1	0	0	0	5	7	500	T
2	105	549702	3	5	1	0	0	0	5	13	450	T
2	105	549703	3	5	1	0	0	0	5	14	450	T
2	105	549704	3	5	1	0	0	0	5	14	450	T
2	105	549705	3	5	2	0	0	0	5	17	900	T
2	105	549706	3	5	2	0	0	0	5	17	900	T
2	105	549707	3	5	1	0	0	0	5	19	450	T

A.9. Excel Macro for Date Stamping the Excel Files

```
Sub datemacro()  
Dim File_Name As String  
  
For i = 1 To 31  
If Len(i) = 1 Then  
File_Name = "C:\Users\Anila Cyril\Desktop\New folder\" & "0" & i & "-01-  
13_combined.xlsx"  
Else  
File_Name = "C:\Users\Anila Cyril\Desktop\New folder\" & i & "-01-  
13_combined.xlsx"  
End If  
Workbooks.Open Filename:= _  
File_Name  
Columns("A:A").Select  
Selection.Insert Shift:=xlToRight, CopyOrigin:=xlFormatFromLeftOrAbove  
Range("A1").Select  
ActiveCell.FormulaR1C1 = Mid(File_Name, InStrRev(File_Name, "\") + 1, 8)  
Range("B1").Select  
Selection.End(xlDown).Select  
Selection.End(xlToLeft).Select  
Range(Selection, Selection.End(xlUp)).Select  
Selection.FillDown  
ActiveCell.FormulaR1C1 = "Date"  
ActiveWorkbook.Save  
ActiveWindow.Close  
Next  
End Sub
```

A.10. Excel Macro for Depot ID Stamping the Excel Files

```
Sub insertdepotid()
Dim MyFile As String, Str As String, MyDir As String, Wb As Workbook

Set Wb = ThisWorkbook
MyDir = "E:\Trial data\"
MyFile = Dir(MyDir & "*.xlsx")
ChDir (MyDir)
Do While MyFile <> ""
    Workbooks.Open Filename:=MyDir & "\" & MyFile
        'Inserting a Column at Column A
        Columns("A:A").Select
        Selection.Insert Shift:=xlToRight, CopyOrigin:=xlFormatFromLeftOrAbove
        Range("A1").Select
        ActiveCell.FormulaR1C1 = 029
        Range("B1").Select
        Selection.End(xlDown).Select
        Selection.End(xlToLeft).Select
        Range(Selection, Selection.End(xlUp)).Select
        Selection.FillDown
        ActiveCell.FormulaR1C1 = "Depot_id"
        ActiveWorkbook.Save
        ActiveWindow.Close

    MyFile = Dir()
Loop

End Sub
```

A.11. Microsoft Excel Macro for Renaming the Files in a Directory

```
Sub Fetchname()  
Dim myPath As String  
Dim myFile As String  
myPath = "E:\Ph.D\Data\Trivandrum City\Rename\  
myFile = Dir(myPath & "*.xlsx")  
r = 2  
Do While myFile <> ""  
    Cells(r, 1).Value = myFile  
    r = r + 1  
    myFile = Dir  
Loop  
End Sub  
=====
```

```
Sub Renamefiles()  
Dim myPath As String  
myPath = "E:\Ph.D\Data\Trivandrum City\Rename\  
  
r = 2  
Do Until IsEmpty(Cells(r, 1)) And IsEmpty(Cells(r, 2))  
Name myPath & Cells(r, 1).Value As myPath & Cells(r, 2).Value  
r = r + 1  
Loop  
  
End Sub
```

A.12. Python Code for Origin-Destination Matrix Generation

```
import os
import sys
import xlrd
import xlwt

import pandas as pd
import xlswriter
import openpyxl
from xlrd import open_workbook
from xlutils.copy import copy

def list_files(dir_name):
    list_of_files=os.listdir('/home/Desktop/'+dir_name)
    return list_of_files

def read_file(dirname,filename):

    path='/home/Desktop/'
    print "processing file"+filename;
    _file=pd.read_excel(path+dirname+'/'+filename,sheetname='Sheet1')
    temp_list=[]
    dird = filename.replace("_combined.xlsx","")
    directory = '/home/Desktop/transport12new/'+dird
    if (os.path.exists(directory) == False):
        os.mkdir(directory)
    else:
        return
    for index, row in _file.iterrows():
        depotid=row['Depot_id']
        try:
            route=row['col2']
        except:
            route=row['Fare stage']
        try:
            trip = row['col1']
        except:
            trip=row['Trip no']
        try:
            depotid += 0;
            route += 0;
            trip += 0;
        except TypeError:
            continue;
```

```

fname=directory+'/'+str(depotid)+'_'+str(route)+'_'+str(trip)+':'+dird+'.xlsx'
if os.path.isfile(fname):
    x=open_workbook(fname)
    book = copy(x)
else:
    y=xlswriter.Workbook(fname)
    d=y.add_worksheet()
    for i in range(0,100):
        for j in range(0,100):
            if i==99 and j==0:
                d.write(i,j,'col_sum')
            elif i==0 and j==99:
                d.write(i,j,'row_sum')
            elif i==0 and j!=0:
                d.write(i,j,j)
            elif j==0 and i!=0:
                d.write(i,j,i)
            elif i==0 and j==0:
                d.write(i,j,'org/des')
            else:
                d.write(i,j,0)
        y.close()
    x=open_workbook(fname)
    book = copy(x)
sheet=book.get_sheet(0)
temp=x.sheet_by_index(0)
try:
    origin=row['col4']
except:
    origin=row['Origin']
try:
    dest=row['col5']
except:
    dest=row['Destination']
try:
    prev=temp.cell(origin+1, dest+1).value
except:
    continue;
try:
    r2=row['col6']
    r3=row['col7']
    r4=row['col9']
except:
    r2=row['Full ticket']
    r3=row['Half ticket']
    r4=row['Physically handicaped']

```

```

    try:
        value=prev+r2+r3+r4
    except:
        continue;
    if origin>98 or dest>98:
        continue
    sheet.write(origin+1,dest+1,value)
    prev_row=temp.cell(origin+1,99).value
    sheet.write(origin+1,99,prev_row+value-prev)
    prev_col=temp.cell(99,dest+1).value
    sheet.write(99,dest+1,prev_col+value-prev)
    prev_sum=temp.cell(99,99).value
    sheet.write(99,99,prev_sum+value-prev)
    book.save(fname)

```

```

if __name__ == "__main__":
    if len(sys.argv) !=2:
        print "enter folder name"
        sys.exit(1)
    list_of_directories=list_files(sys.argv[1])
    if (os.path.exists('/home/Desktop/transport12new/') == False):
        os.mkdir('/home/Desktop/transport12new/')
    for dirname in list_of_directories:
        list_of_files=list_files(sys.argv[1]+'/'+dirname);
        for filename in list_of_files:
            read_file(sys.argv[1]+'/'+dirname,filename)

```

A.13. Python Code for Generation of Time-Series Data from OD Matrix

```
import os
import sys
import xlrd
import xlwt

import pandas as pd
import xlswriter
import openpyxl
from xlrd import open_workbook
from xlutils.copy import copy

def list_files(dir_name):
    list_of_files=os.listdir('/home/Desktop/'+dir_name)
    return list_of_files

def read_file(dirname,filename):

    path='/home/student/Desktop/'
    print "processing file "+filename;
    _file=pd.read_excel(path+dirname+'/'+filename,sheetname='Sheet1')
    temp_list=[]
    dird = filename.replace("_combined.xlsx","")
    directory = '/home/Desktop/transport1/'+dird
    os.mkdir(directory)
    for index, row in _file.iterrows():
        depotid=row['Depot_id']
        route=row['col2']
        trip = row['col1']

    fname=directory+'/'+str(depotid)+'_'+str(route)+'_'+str(trip)+':'+dird+'.xlsx'
    if os.path.isfile(fname):
        x=open_workbook(fname)
        book = copy(x)
    else:
        y=xlswriter.Workbook(fname)
        d=y.add_worksheet()
        for i in range(0,101):
            for j in range(0,101):
                if i==100 and j==0:
                    d.write(i,j,'col_sum')
                elif i==0 and j==100:
                    d.write(i,j,'row_sum')
                elif i==0 and j!=0:
                    d.write(i,j,j)
                elif j==0 and i!=0:
```



```

        d.write(i,j,i)
    elif i==0 and j==0:
        d.write(i,j,'org/des')
    else:
        d.write(i,j,0)

    y.close()
    x=open_workbook(fname)
    book = copy(x)
    sheet=book.get_sheet(0)
    temp=x.sheet_by_index(0)
    origin=row['col4']
    dest=row['col5']
    prev=temp.cell(origin+1, dest+1).value
    value=prev+row['col6']+row['col7']+row['col9']
    sheet.write(origin+1,dest+1,value)
    prev_row=temp.cell(origin+1,100).value
    sheet.write(origin+1,100,prev_row+value-prev)
    prev_col=temp.cell(100,dest+1).value
    sheet.write(100,dest+1,prev_col+value-prev)
    prev_sum=temp.cell(100,100).value
    sheet.write(100,100,prev_sum+value-prev)
    book.save(fname)

```

```

if __name__ == "__main__":
    if len(sys.argv) !=2:
        print "enter folder name"
        sys.exit(1)
    list_of_directories=list_files(sys.argv[1]);
    os.mkdir('/home/Desktop/transport1/')
    for dirname in list_of_directories:
        list_of_files=list_files(sys.argv[1]+'/'+dirname);
        for filename in list_of_files:
            read_file(sys.argv[1]+'/'+dirname,filename)

```

A.14. Macro for Merging Excel Sheets

```
Sub simpleXlsMerger()  
Dim bookList As Workbook  
Dim mergeObj As Object, dirObj As Object, filesObj As Object, everyObj As Object  
Application.ScreenUpdating = False  
Set mergeObj = CreateObject("Scripting.FileSystemObject")  
  
'change folder path of excel files here  
Set dirObj = mergeObj.Getfolder("D:\change\to\excel\files\path\  
Set filesObj = dirObj.Files  
For Each everyObj In filesObj  
Set bookList = Workbooks.Open(everyObj)  
  
'change "A2" with cell reference of start point for every files here  
'for example "B3:IV" to merge all files start from columns B and rows 3  
'If you're files using more than IV column, change it to the latest column  
'Also change "A" column on "A65536" to the same column as start point  
Range("A2:IV" & Range("A65536").End(xlUp).Row).Copy  
ThisWorkbook.Worksheets(1).Activate  
  
'Do not change the following column. It's not the same column as above  
Range("A65536").End(xlUp).Offset(1, 0).PasteSpecial  
Application.CutCopyMode = False  
bookList.Close  
Next  
End Sub
```

A.15. Macro for Inserting Missing Dates in Excel

```
Sub loop_through_all_worksheets()

Dim ws As Worksheet
Dim starting_ws As Worksheet
Set starting_ws = ActiveSheet 'remember which worksheet is active in the beginning

For Each ws In ThisWorkbook.Worksheets
    ws.Activate

    i = 2

    Do
        'Use less than instead of <> so it doesn't flag duplicate cells
        If Cells(i, 1) + 1 < Cells(i + 1, 1) Then
            Rows(i + 1).Insert
            Cells(i + 1, 1) = Cells(i, 1) + 1
        End If
        'Second check to add value if the next row is blank
        If (Cells(i + 1, 1) = "") Then
            Cells(i + 1, 1) = Cells(i, 1) + 1
        End If

        i = i + 1
    'Changed the loop function from cells(i+1,1) to cells(i,1) since you already
    'incremented i
    'Also made the date check slightly more robust with dateserial
    Loop Until Cells(i, 1).Value >= DateSerial(2013, 12, 31)
    ws.Cells(1, 1) = "Date" 'this sets cell A1 of each sheet to "1"
    Next

    starting_ws.Activate 'activate the worksheet that was originally active

End Sub
```

A.16. R Code for Time-Series Model Development

```
# Load libraries and set global parameters
packages = c("fGarch", "rugarch", "FinTS", "bimixt", "quantmod", "xlsx",
"tstools", "astsa", "tsoutliers", "TSstudio", "imputeTS", "forecast", "aTSA",
"lubridate", "tidyquant", "timetk", "ggplot2", "dplyr", "tidyr", "data.table", "corrplot",
"gridExtra", "tseries", "TSA", "tibble", "TTR", "xts", "dygraphs",
"assertthat", "neuralnet", "nonlinearTseries", "nortest")

my.install <- function(pkg, ...){
  if (!(pkg %in% installed.packages()[,1])) {
    install.packages(pkg)
  }
  return (library(pkg, ...))
}

purrr::walk(packages, my.install, character.only = TRUE, warn.conflicts = FALSE)

#####Initializing and loading the data#####
#Change this variable to appropriate data directory
data_dir <- "G:/phase 3 ARCH/KANIYAPURAM/Route 10"
setwd(data_dir)

#loading route 100 data to R
Data <- read.csv("10-IT.csv", stringsAsFactors = F)

#formatting date to standard date format
Data$Date <- as.Date(Data$Date, "%d-%m-%Y")

#Finding max and min dates to enumerate all the dates in between.
max_date <- max(Data$Date)
min_date <- min(Data$Date)

#Generating all dates from max to min
all_dates <- seq(min_date, max_date, by = "days")

#Saving all the dates in a dataframe
all_dates_df <- data.frame(Date = all_dates, stringsAsFactors = F)

#Merging route100 data with all dates-trip combination
data_df <- merge(all_dates_df, Data, all.x = T)

#####Data Conversion (Data Frame to xts)#####
#converting data frame object to time series object
```

```

data_xts <- xts(data_df[,-1], order.by = data_df[,1]) #this is to include the column
heading
colnames(data_xts) <- c("Total")
str(data_xts) #to check the structure of the trip1_xts

#Statistics about missing data
missing_statistics = statsNA(data_xts, printOnly = T)

#Missing data imputation using linear interpolation
imputed_trip <- na.interpolation(data_xts, option = "linear")

#OUTLIER ANALYSIS
outliers <- tso(ts(imputed_trip), tsmethod = c("auto.arima"), args.tsmethod =
list(allowdrift=TRUE, stepwise=FALSE, approximation=FALSE))
print(outliers)

#imputed_trip["2013-12-05"]<-NA
imputed_trip <- na.interpolation(imputed_trip, option = "linear")

#ARCH TEST
h1 <- ArchTest(imputed_trip, demean = FALSE) #default lag =12
print(h1)
sink("original Arch.txt")
print(h1)
sink()

# BOX COX TRANSFORMATION
lambda <- BoxCox.lambda(imputed_trip)
print(lambda)

#SPLITTING TRANSFORMED DATA
imputed_trip.split <- floor(0.955*length(imputed_trip))
traindata <- imputed_trip[1:imputed_trip.split]
testdata <- imputed_trip[(imputed_trip.split+1):length(imputed_trip)]
h <- length(testdata)

#ARIMA MODELLING
arima.fit1 <- auto.arima(ts(traindata, frequency=7), d=NA, D =NA, stepwise = FALSE,
trace= TRUE, approximation = FALSE, lambda = "auto")
print(arima.fit1)
sink("arima.fit1.txt")
arima.fit1
sink()
var1 <- var(arima.fit1$residuals)
print(var1)
boxcox.inv(var1, 0.2031358)

```

```

#ARIMA FORECAST
arima.fc1<- forecast::forecast(arima.fit1, h, level=c(90,95))
write.xlsx(arima.fc1, "arima.fc1.xlsx")
plot(arima.fc1)
dev.print(pdf, 'forecast plot.pdf')
test_forecast(actual =imputed_trip, forecast.obj = arima.fc1, test = testdata)
checkresiduals(arima.fit1)
sink("residuals.txt")
checkresiduals(arima.fit1)
sink()
dev.print(pdf, 'Rplot20.pdf')
summary(arima.fit1)
sink("summary_arima.fit1.txt")
summary(arima.fit1)
sink()
accuracy(arima.fc1,testdata)
sink("accuracy_arima.fit1.txt")
accuracy(arima.fc1,testdata)
sink()

#ARIMA VALIDATION
validate_arima1 <- Arima(testdata, model = arima.fit1)
print(validate_arima1)
sink("validate_arima1.txt")
validate_arima1
sink()

accuracy(validate_arima1)
sink("accuracy_validate_arima1.txt")
accuracy(validate_arima1)
sink()

# BOX TESTS AND RESIDUAL ACF AND PACF
tsdisplay(residuals(arima.fit1))
dev.print(pdf, 'RESIDUAL PLOT.pdf')
Box.test(arima.fit1$residuals, type = "Ljung-Box")
sink("box.test.1.txt")
Box.test(arima.fit1$residuals, type = "Ljung-Box")
sink()

#ACF AND PACF OF RESIDUALS
Acf(ts(arima.fit1$residuals),main='ACF Residual')
dev.print(pdf, 'ACF(r).pdf')
Pacf(ts(arima.fit1$residuals),main='PACF Residual')
dev.print(pdf, 'PACF(r).pdf')

```

#RESIDUAL ANALYSIS-CHECKING HETEROSCEDASTICITY IN FITTED ARIMA MODEL

```
arima.residuals<-residuals(arima.fit1)
write.xlsx(arima.residuals,"ARIMARESIDUALS.xlsx")
jarque.bera.test(arima.residuals)
sink("jarqueres.txt")
jarque.bera.test(arima.residuals)
sink()
shapiro.test(as.numeric(arima.residuals))
sink("shapiroresiduals.txt")
shapiro.test(as.numeric(arima.residuals))
sink()
```

```
qqnorm(arima.residuals,main='RESIDUAL NORMALITY')
qqline(arima.residuals)
dev.print(pdf,'qqline.pdf')
```

```
McLeod.Li.test(arima.fit1,imputed_trip,plot=TRUE)
dev.print(pdf,'mcleoad1.pdf')
```

```
archtest<-
arch.test(arima(imputed_trip,order=c(1,1,1),seasonal=list(order=c(2,0,0),period=7)),o
utput=TRUE)
summary(archtest)
sink("archtest.txt")
archtest<-
arch.test(arima(imputed_trip,order=c(1,1,1),seasonal=list(order=c(2,0,0),period=7)),o
utput=TRUE)
summary(archtest)
sink()
```

#ARIMA MODELLING

```
arima.fit2<-auto.arima(ts(traindata, frequency=7),d=NA, D =1, stepwise = FALSE,
trace= TRUE, approximation = FALSE, lambda = "auto")
print(arima.fit2)
sink("arima.fit2.txt")
arima.fit2
sink()
```

#ARIMA FORECAST

```
arima.fc2<- forecast::forecast(arima.fit2, h, level=c(90,95))
write.xlsx(arima.fc2, "arima.fc2.xlsx")
plot(arima.fc2)
dev.print(pdf, 'forecast plot2.pdf')
test_forecast(actual =imputed_trip, forecast.obj = arima.fc2, test = testdata)
```

```

checkresiduals(arima.fit2)
sink("residuals2.txt")
checkresiduals(arima.fit2)
sink()
dev.print(pdf, 'Rplot202.pdf')
summary(arima.fit2)
sink("summary_arima.fit2.txt")
summary(arima.fit2)
sink()
accuracy(arima.fc2,testdata)
sink("accuracy_arima.fit2.txt")
accuracy(arima.fc2,testdata)
sink()

#ARIMA VALIDATION
validate_arima2 <- Arima(testdata, model = arima.fit2)
print(validate_arima2)
sink("validate_arima2.txt")
validate_arima2
sink()

accuracy(validate_arima2)
sink("accuracy_validate_arima2.txt")
accuracy(validate_arima2)
sink()

# BOX TESTS AND RESIDUAL ACF AND PACF
tsdisplay(residuals(arima.fit2))
dev.print(pdf, 'RESIDUAL PLOT2.pdf')
Box.test(arima.fit2$residuals, type = "Ljung-Box")
sink("box.test.2.txt")
Box.test(arima.fit2$residuals, type = "Ljung-Box")
sink()

#ACF AND PACF OF RESIDUALS
Acf(ts(arima.fit2$residuals),main='ACF Residual')
dev.print(pdf, 'ACF(r)2.pdf')
Pacf(ts(arima.fit2$residuals),main='PACF Residual')
dev.print(pdf, 'PACF(r)2.pdf')

#RESIDUAL ANALYSIS-CHECKING HETEROSCEDASTICITY IN FITTED
ARIMA MODEL

arima.residuals2<-residuals(arima.fit2)
write.csv(arima.residuals2,"ARIMARESIDUALS2.csv")
jarque.bera.test(arima.residuals2)
sink("jarque2.txt")

```



```

jarque.bera.test(arima.residuals2)
sink()
shapiro.test(as.numeric(arima.residuals2))
sink("shapiroresiduals2.txt")
shapiro.test(as.numeric(arima.residuals2))
sink()

qqnorm(arima.residuals2,main='RESIDUAL NORMALITY')
qqline(arima.residuals2)
dev.print(pdf,'qqline2.pdf')

McLeod.Li.test(arima.fit2,imputed_trip,plot=TRUE)
dev.print(pdf,'mcleoad2.pdf')

archtest2<-
arch.test(arima(imputed_trip,order=c(4,0,0),seasonal=list(order=c(0,1,1),period=7),
include.mean = TRUE),output=TRUE)
summary(archtest2)
sink("archtest2.txt")
archtest2<-
arch.test(arima(imputed_trip,order=c(4,0,0),seasonal=list(order=c(0,1,1),period=7),
include.mean = TRUE),output=TRUE)
summary(archtest2)
sink()

#ARIMA MODELLING
arima.fit3<-auto.arima(ts(traindata, frequency=7),d=1, D =1, stepwise = FALSE,
trace= TRUE, approximation = FALSE, lambda = "auto")
print(arima.fit3)
sink("arima.fit3.txt")
arima.fit3
sink()

#ARIMA FORECAST
arima.fc3<- forecast::forecast(arima.fit3, h, level=c(90,95))
write.xlsx(arima.fc3, "arima.fc3.xlsx")
plot(arima.fc3)
dev.print(pdf, 'forecast plot3.pdf')
test_forecast(actual =imputed_trip, forecast.obj = arima.fc3, test = testdata)
checkresiduals(arima.fit3)
sink("residuals3.txt")
checkresiduals(arima.fit3)
sink()
dev.print(pdf, 'Rplot203.pdf')
summary(arima.fit3)
sink("summary_arima.fit3.txt")
summary(arima.fit3)

```

```

sink()
accuracy(arima.fc3,testdata)
sink("accuracy_arima.fit3.txt")
accuracy(arima.fc3,testdata)
sink()

#ARIMA VALIDATION
validate_arima3 <- Arima(testdata, model = arima.fit3)
print(validate_arima3)
sink("validate_arima3.txt")
validate_arima3
sink()

accuracy(validate_arima3)
sink("accuracy_validate_arima3.txt")
accuracy(validate_arima3)
sink()

# BOX TESTS AND RESIDUAL ACF AND PACF
tsdisplay(residuals(arima.fit3))
dev.print(pdf, 'RESIDUAL PLOT3.pdf')
Box.test(arima.fit3$residuals, type = "Ljung-Box")
sink("box.test.3.txt")
Box.test(arima.fit3$residuals, type = "Ljung-Box")
sink()

#ACF AND PACF OF RESIDUALS
Acf(ts(arima.fit3$residuals),main='ACF Residual')
dev.print(pdf, 'ACF(r)3.pdf')
Pacf(ts(arima.fit3$residuals),main='PACF Residual')
dev.print(pdf, 'PACF(r)3.pdf')

#RESIDUAL ANALYSIS-CHECKING HETEROSCEDASTICITY IN FITTED
ARIMA MODEL

arima.residuals3<-residuals(arima.fit3)
write.csv(arima.residuals2,"ARIMARESIDUALS3.csv")

jarque.bera.test(arima.residuals3)
sink("jarque3.txt")
jarque.bera.test(arima.residuals3)
sink()
shapiro.test(as.numeric(arima.residuals3))
sink("shapiroresiduals3.txt")
shapiro.test(as.numeric(arima.residuals3))
sink()

```

```

qqnorm(arima.residuals3,main='RESIDUAL NORMALITY')
qqline(arima.residuals3)
dev.print(pdf,'qqline3.pdf')

McLeod.Li.test(arima.fit3,imputed_trip,plot=TRUE)
dev.print(pdf,'mcleoad3.pdf')

archtest3<-
arch.test(arima(imputed_trip,order=c(1,1,1),seasonal=list(order=c(0,1,1),period=7)),o
utput=TRUE)
summary(archtest3)
sink("archtest3.txt")
archtest3<-
arch.test(arima(imputed_trip,order=c(1,1,1),seasonal=list(order=c(0,1,1),period=7)),o
utput=TRUE)
summary(archtest3)
sink()

#ARCH MODELLING

data_dir <- "G:/phase 3 ARCH/ARCH/DATA"
setwd(data_dir)

#loading route 100 data to R
Data <- read.csv("93-sqr.csv", stringsAsFactors = F)

#formatting date to standard date format
Data$Date <- as.Date(Data$Date, "%d-%m-%Y")

#Finding max and min dates to enumerate all the dates in between.
max_date <- max(Data$Date)
min_date <- min(Data$Date)

#Generating all dates from max to min
all_dates <- seq(min_date, max_date, by = "days")

#Saving all the dates in a dataframe
all_dates_df <- data.frame(Date = all_dates, stringsAsFactors = F)

#Merging route100 data with all dates-trip combination
data_df <- merge(all_dates_df, Data, all.x = T)

#Line plot for trip1
plot(data_df$Total, type = "l", xlab = "Time Index", ylab = "Number of Passengers")
dev.print(pdf, 'Rplotdata.pdf')

```

```

#####Data Conversion (Data Frame to xts)#####
#converting data frame object to time series object
data_xts <-xts(data_df[,-1], order.by = data_df[,1]) #this is to include the column
heading
colnames(data_xts) <- c("Total")
str(data_xts) #to check the structure of the trip1_xts

missing_statistics = statsNA(data_xts, printOnly = T)
sink("missing.txt")
missing_statistics = statsNA(data_xts, printOnly = T)
sink()

#Missing data imputation using linear interpolation
imputed_tripsqr<-na.interpolation(data_xts, option = "linear")

#ARCH TEST
ArchTest(imputed_tripsqr)

#ACF AND PACF OF BOXCOX TRANSFORMED
Autocorrelationbox<-Acf(imputed_tripsqr, lag.max = NULL, type = c("correlation",
"covariance", "partial"),
plot = TRUE, na.action = na.contiguous, demean = TRUE)
dev.print(pdf, 'ACFsqr.pdf')

#PACF
Partialautocorrelationbox<-Pacf(imputed_tripsqr, lag.max = NULL, plot = TRUE,
na.action = na.contiguous,
demean = TRUE)
dev.print(pdf, 'PACFsqr.pdf')

#GARCH MODELLING
garch.fit1<-garch(imputed_tripsqr, order = c(4,4))
print(garch.fit1)
sink(garch.fit1,"garch.txt")
garch.fit1
sink()

garch.fc<-forecast::forecast(garch.fit1, h = 30)

#GARCH WITH RUGARCH
model.garch = ugarchspec(variance.model = list(garchOrder = c(0,1)),mean.model =
list(armaOrder = c(0,0),include.mean = F), distribution.model = "std")
model.garch.fit <- ugarchfit(spec = model.garch, solver = "hybrid", data =
squared.res.arima)
model.forecast <- ugarchforecast(n.ahead = 30, model.garch.fit )

```

```

print(model.forecast)

#ARCH MODELLING METHOD -3
res.arma=arma.fit2$residuals
write.csv(res.arma,"arma residuals.csv")
y1<-ArchTest(res.arma,lags = 12, demean = F)
print(y1)
sink("archtest.txt")
print(y1)
sink()

squared.res.arma=res.arma^2
write.csv(squared.res.arma,"squared residuals.csv")
y2 <- ArchTest(squared.res.arma,lags = 12, demean = F)
print(y2)
sink("archtestsqr.txt")
print(y2)
sink()

plot(squared.res.arma,main='Squared Residuals')
dev.print(pdf,'squared plot.pdf')

acf.squared=acf(squared.res.arma,main='ACF Squared Residuals',lag.max=100)
dev.print(pdf,'acf squared.pdf')

pacf.squared=pacf(squared.res.arma,main='PACF Squared Residuals',lag.max=100)
dev.print(pdf,'pacf squared.pdf')

# arch01=garch(res.arma,order=c(0,16),trace=F)
# loglik01=logLik(arch01)
# print(loglik01)
# sink("loglik.txt")
# print(loglik01)
# sink()

#for (p in 1:10){

# for (q in 1:10){

arch01=garch(res.arma,order=c(0,2),trace=F)
loglik01=logLik(arch01)

AICc=(-2*loglik01 + (2*((0+2)+1)*(661/(661-(2)-2)))) #no constant
print(AICc)

```

```

AICc1=(-2*loglik01 + (2*((2)+2)*(661/(661-(2)-3)))) #with constant
print(AICc1)

#AICc<-(-2*loglik01 + (2*(q+1)*(N/(N-q-2)))) #no constant
# AICc=(-2*loglik01 + (2*(16+1)*(661/(661-16-2)))) #no constant
# AICc
# sink("AICc.txt")
# AICc
# sink()

#AICc<-(-2*loglik01 + (2*(q+2)*(N/(N-q-3)))) #with constant
# AICc1=(-2*loglik01 + (2*(16+2)*(661/(661-16-3)))) #with constant
# AICc1
# sink("AICc1.txt")
# AICc1
# sink()
summary(arch01)

#CHECKING (METHOD 3)
ht.arch01=arch01$fit[,1]^2
plot(ht.arch01,main='Conditional variances')
dev.print(pdf,'conditional variance.pdf')

archres=res.arma/sqrt(ht.arch01)
qqnorm(archres,main='ARIMA-ARCH Residuals')
qqline(archres)
dev.print(pdf,'qq arch.pdf')

fit=arma.fit2$fitted
low=fit-1.96*sqrt(ht.arch01)
high=fit+1.96*sqrt(ht.arch01)
plot(fit,type='s',main="Trip,Low,High")
lines(low,col='red')
lines(high,col='blue')
dev.print(pdf,"Trip,Low,high.pdf")

#Modelling with fGarch Package
Fgarch.fit <- garchFit(formula = ~garch(2,2),data= squared.res.arma,cond.dist =
"norm",include.mean = TRUE)
h.t<-Fgarch.fit@h.t
eps<-Fgarch.fit@residuals
B<-coef(Fgarch.fit)
my.predict<-function(k) {
  h.predict<-
B[2]+h.t[length(squared.res.arma)]*B[4]+(eps[length(squared.res.arma)]^2)*B[3]
  i<-2

```

```
while(i==k){  
  h.predict[i]<-B[2]+h.predict[i-1]*(B[4]+B[3])  
  print(h.predict)  
  i<-i+1  
}  
return(h.predict)  
}
```

```
fcast<-predict(Fgarch.fit, n.ahead = 30)  
print(fcast)  
sink("fgarch forecast.txt")  
print(fcast)  
sink()
```

A.17. Weekly Variation Index of Trivandrum Depots

Week	Kaniyapuram	Pappanamcode	Peroorkada	Trivandrum City	Vikas Bhavan	Vizhinjam
1	0.74	1.08	1.22	1.09	1.14	1.07
2	0.72	0.90	1.13	1.06	0.95	1.03
3	0.74	0.93	1.05	1.12	1.06	1.10
4	0.68	1.00	1.06	1.10	1.00	1.08
5	0.75	0.97	1.02	1.01	1.18	1.09
6	0.83	0.93	1.02	1.08	1.07	1.13
7	0.93	0.92	1.03	1.20	1.14	1.11
8	1.10	1.00	0.99	1.27	0.90	1.09
9	1.12	0.97	0.93	1.29	1.03	1.06
10	1.09	0.92	0.89	1.12	1.00	1.08
11	1.13	0.87	0.81	1.08	0.74	1.08
12	1.15	0.86	0.91	1.12	0.75	1.09
13	1.11	0.76	0.85	1.09	0.68	1.13
14	1.10	0.81	0.72	1.06	0.65	1.12
15	1.09	0.80	0.92	0.94	0.61	1.09
16	1.13	1.03	0.91	0.92	0.95	1.08
17	1.10	1.03	0.96	0.94	0.99	1.07
18	1.07	1.01	0.95	0.85	1.01	1.03
19	1.18	0.94	0.95	0.90	1.18	1.07
20	1.18	0.67	0.98	1.12	1.15	1.10
21	1.08	0.65	0.75	1.03	1.05	1.05
22	1.14	0.72	0.71	1.11	1.02	1.14
23	1.16	1.28	1.08	1.09	1.06	1.12
24	1.12	1.18	1.25	1.04	0.97	1.01
25	1.06	1.27	1.20	1.02	0.97	1.01
26	1.11	1.30	1.20	1.07	0.95	1.01
27	1.10	1.34	1.22	0.96	1.01	1.00
28	1.05	1.34	1.21	1.04	1.06	0.97
29	1.14	1.28	1.21	0.98	1.15	0.98
30	1.07	1.33	1.14	0.99	1.05	0.94
31	0.95	1.12	1.00	0.89	0.93	0.90
32	1.06	1.29	1.21	0.94	1.06	0.87
33	1.08	1.27	1.16	0.88	1.01	0.93
34	1.11	1.26	1.03	0.99	1.04	0.92
35	1.00	0.99	0.89	0.81	0.92	0.87
36	1.08	1.06	1.06	0.88	1.02	0.96
37	1.02	1.03	0.92	0.84	0.99	0.89
38	1.02	1.16	1.08	0.87	0.94	0.90
39	1.08	1.19	1.11	0.93	0.93	0.90
40	1.06	1.18	1.13	0.88	1.01	0.91

41	1.05	1.04	1.04	1.02	1.19	0.96
42	0.97	1.02	1.12	1.11	1.25	0.95
43	0.94	1.11	0.89	1.09	1.09	0.91
44	0.96	1.14	0.79	1.26	1.11	0.90
45	0.92	1.00	0.91	1.09	1.08	0.83
46	0.94	0.87	1.04	0.95	1.02	0.85
47	0.87	0.78	1.05	0.74	1.01	0.81
48	0.90	0.79	0.97	0.80	1.10	0.84
49	0.85	0.81	0.92	0.86	1.03	0.80
50	0.86	0.84	0.92	0.77	1.06	1.05
51	0.82	0.65	0.93	0.97	1.03	1.03
52	0.78	0.66	0.80	0.92	0.89	1.02
53	0.70	0.64	0.77	0.81	0.81	1.04

REFERENCES

- [1] Adhikari, R., and Agrawal, R. K. (2013). “An introductory study on time series modeling and forecasting.” *arXiv Prepr. arXiv1302.6613*, 1302.6613, 1–68.
- [2] Al Mamun, M. and Lownes, N.E. (2011). “A composite index of public transit accessibility.” *Journal of Public Transportation*, 14(2), 69-87.
- [3] Albacete, X., Olaru, D., Paül, V., and Biermann, S. (2017). “Measuring the accessibility of public transport: a critical comparison between methods in Helsinki.” *Applied Spatial Analysis and Policy*, 10(2),161-188.
- [4] Aldian, A., and Taylor, M. A. P. (2003). “Fuzzy Multicriteria Analysis for Intercity Travel Demand Modeling.” *Journal of Eastern Asia Society for Transportation Studies*, 5, 1294-1307.
- [5] Allen, W.B., Liu, D. and Singer, S. (1993). “Accessibility measures of US metropolitan areas.” *Transportation Research Part B: Methodological*, 27 (6), 439-449.
- [6] Ambrasaite, I., Barfod, M. B., and Salling, K. B. (2011). “MCDA and risk analysis in transport infrastructure appraisals: The Rail Baltica case.” *Procedia-Social and Behavioral Sciences*, 20, 944-953.
- [7] Andreoni, A., and Postorino, M. N. (2006). “A multivariate ARIMA model to forecast air transport demand.” *Proceedings of the Association for European Transport and Contributors*, 1-14.
- [8] Annual Audit Report 2014-15. (2015). “*Kerala State Road Transport Corporation.*” Thiruvananthapuram, Kerala.
- [9] Annual Audit Report 2015-16. (2016). “*Kerala State Road Transport Corporation.*” Thiruvananthapuram, Kerala.

- [10] Anvari, S., Tuna, S., Canci, M., and Turkey, M. (2016). "Automated Box-Jenkins forecasting tool with an application for passenger demand in urban rail systems." *Journal of Advanced Transportation*, 50(1), 25–49.
- [11] Ashalatha, R., Manju, V. S., and Zacharia, A. B. (2012). "Mode choice behavior of commuters in Thiruvananthapuram City." *Journal of Transportation Engineering*, 139(5), 494-502.
- [12] Asri, D. U., and Sugie, Y. (2003). "Simultaneous demand model for passenger travel: A case study of Indonesia." *Journal of Eastern Asia Society for Transportation Studies*, 4, 869-884.
- [13] Aydin, N., Celik, E., and Gumus, A. T. (2015). "A hierarchical customer satisfaction framework for evaluating rail transit systems of Istanbul." *Transportation Research Part A: Policy and Practice*, 77, 61-81.
- [14] Balya, M. I., Katti, B. K., and Saw, K. (2016). "Spatial transit accessibility modeling of Indian metropolitan city in GIS environment." *International Journal for Traffic and Transport Engineering*, 6(1), 51-62.
- [15] Barabino, B., Salis, S., and Useli, B. (2014). "Fare Evasion In Proof-Of-Payment Transit Systems: Deriving The Optimum Inspection Level." *Transportation Research Part B: Methodological*, 70, 1-17.
- [16] Barabino, B., Di Francesco, M., and Mozzoni, S. (2014). "An Offline Framework for Handling Automatic Passenger Counting Raw Data." *IEEE Transactions on Intelligent Transportation Systems*, 15 (6), 2443-2456.
- [17] Barnes, G., and Davis, G. A. (1999). "Understanding urban travel demand: problems, solutions, and the role of forecasting." Report 2 in the Series: Transportation and Regional Growth Study CTS 99-02, Center for Transportation Studies, University of Minnesota
- [18] Barron, F.H., and Barrett, B.E. (1996). "Decision quality using ranked attribute weights." *Management Science*, 42(11),1515-1523.

- [19] Baxter, R. S., and Lenzi, G. (1975). "The measurement of relative accessibility." *Regional Studies*, 9(1), 15-26.
- [20] Beckmann, M. J., McGuire, C. B., and Winsten, C. B. (1955). "*Studies in the Economics of Transportation.*" Yale University Press, New Haven, Connecticut; also published as Rand-RM-1488-PR, Rand Corporation, Santa Monica, CA, May 12, 1955.
- [21] Bennett, D. A. (2001). "How can I deal with missing data in my study?." *Australian and New Zealand Journal of Public Health*, 25(5), 464-469.
- [22] Bermúdez, J. D., Segura, J. V., and Vercher, E. (2007). "Holt-Winters Forecasting: An Alternative Formulation Applied to UK Air Passenger Data." *Journal of Applied Statistics*, 34 (9), 1075-1090.
- [23] Bhat, C., Handy, S., Kockelman, K., Mahmassani, H., Chen, Q., and Weston, L. (2000). "*Urban accessibility index: literature review.*" Center of Transportation Research, University of Texas at Austin, Springfield. Work, 7, 4938-1.
- [24] Bhat, C., Handy, S., Kockelman, K., Mahmassani, H., Chen, Q., and Weston, L. (2000). "*Accessibility measures: formulation considerations and current applications.*" Center of Transportation Research, University of Texas at Austin, Springfield. Work, 7, 4938-2.
- [25] Bhat, C., Handy, S., Kockelman, K., Mahmassani, H., Gopal, A., Srour, I., and Weston, L. (2002). "*Development of an urban accessibility index: formulations, aggregation, and application.*" Center of Transportation Research, University of Texas at Austin, Springfield. Work 7, 4938-4.
- [26] Bhat, C. R., Guo, J. Y., Sen, S., and Weston, L. (2005). "*Measuring access to public transportation services: Review of customer-oriented transit performance measures and methods of transit submarket identification.*" Center of Transportation Research, University of Texas at Austin, Springfield. Report No. 0-5178-1.

- [27] Bisaglia, L., and Gerolimetto, M. (2014). “*Testing for (non)linearity in economic time series: a Monte-Carlo comparison.*” Working Paper Series, 3.
- [28] Bok, J., and Kwon, Y. (2016). “Comparable measures of accessibility to public transport using the general transit feed specification.” *Sustainability*, 8(3), 224.
- [29] Bollerslev, T. (1986). “Generalized autoregressive conditional heteroscedasticity.” *Journal of Econometrics*, 31(3), 307-327.
- [30] Bošnjak, Z., Stojković, M., and Bošnjak, S. (2005). “Processing Time Series Data by Means of Forecasting Models.” *Proceedings of Third Serbian-Hungarian Joint Symposium on Intelligent Systems, SISY 2005-Subotica, Serbia.*
- [31] Brandt, J.A., and Bessler, D.A. (1983). “Price Forecasting and Evaluation: An Application in Agriculture.” *Journal of Forecasting*, 2 (3), 237-248.
- [32] Cardozo, O. D., García-Palomares, J. C., and Gutiérrez, J. (2012). “Application of geographically weighted regression to the direct forecasting of transit ridership at station-level.” *Applied Geography*, 34, 548-558.
- [33] Carrel, A. (2009). “*Diagnosis and assessment of operations control interventions: Framework and applications to a high frequency metro line.*” Doctoral dissertation, Massachusetts Institute of Technology.
- [34] Ceder, A. (2016). “*Public transit planning and operation: modeling, practice and behaviour.*” 2nd edition, CRC press.
- [35] Çelikoğlu, H. B., and Akad, M. (2005). “Estimation of public transport trips by feed forward back propagation artificial neural networks: A case study for Istanbul.” In *Soft Computing: Methodologies and Applications*, 27-36, Springer, Berlin, Heidelberg.
- [36] Celikoglu, H.B., and Cigizoglu, H. K. (2006). “Modeling public transport trips by radial basis function neural networks.” *Mathematical and Computer Modeling*, 45(3-4), 480-489.

- [37] Celikoglu, H. B., and Cigizoglu, H. K. (2006). "Public transportation trip flow modeling with generalized regression neural networks." *Advances in Engineering Software*, 38(2).
- [38] Census. (2011). "*Trivandrum District Census Handbook*." Registrar General of India, Ministry of Home Affairs, Government of India.
- [39] Census. (2011). "*Mangalore Census Handbook*." Registrar General of India, Ministry of Home Affairs, Government of India.
- [40] Cervero, R. (2006). "Alternative approaches to modeling the travel-demand impacts of smart growth." *Journal of the American Planning Association*, 72(3), 285-295.
- [41] Cervero, R., Murakami, J., and Miller, M. (2010). "Direct ridership model of bus rapid transit in Los Angeles County, California." *Transportation Research Record: Journal of the Transportation Research Board*, 2145, 1-7.
- [42] Charles, J. K., Charles, P., and Donald, R. D. (1965). "Highway capacity the level of service concept." *35th Annual Meeting of the Institute of Traffic Engineers*, A world traffic engineering conference, Boston.
- [43] Chatfield, C. (2000). "*Time-series forecasting*." CRC Press.
- [44] Chen, C., and Liu, L. M. (1993). "Joint estimation of model parameters and outlier effects in time series." *Journal of the American Statistical Association*, 88(421), 284-297.
- [45] Chen, G., Abraham, B., and Bennett, G. W. (1997). "Parametric and non-parametric modelling of time series—an empirical study." *Environmetrics: The official Journal of the International Environmetrics Society*, 8(1), 63-74.
- [46] Chen, L., and Yu, R. (2016). "Research on performance evaluation of urban public transport based on DEA competing cross efficiency." *Advances in Engineering Research*, 72, 465-470.

- [47] Cheng, Y. H., and Chen, S.Y. (2015). “Perceived accessibility, mobility, and connectivity of public transportation systems.” *Transportation Research Part A: Policy and Practice*, 77, 386-403.
- [48] Chidlovskii, B. (2017). “Multi-task learning of time-series and its application to the travel demand.” *arXiv:1712.08164v1 [cs.LG]*.
- [49] Chow, L. F., Zhao, F., Liu, X., Li, M. T., and Ubaka, I. (2006). “Transit ridership model based on geographically weighted regression.” *Transportation Research Record: Journal of the Transportation Research Board*, 1972, 105-114.
- [50] *Comprehensive Mobility Plan*. (2014). Government of Kerala. Trivandrum.
- [51] Cyprich, O., Konečný, V., and Kilianova, K. (2013). “Short-term passenger demand forecasting using univariate time series theory.” *PROMET-Traffic and Transportation*, 25(6), 533-541.
- [52] Cyril, A., George, V., and Mulangi, R. H. (2017). “Electronic ticket machine data analytics for public bus transport planning.” In 2017 *International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, Chennai, 3917-3922, IEEE.
- [53] Cyril, A., Mulangi, R. H., and George, V. (2018). “Data processing methodology for big data in public transportation.” *International Journal of Pure and Applied Mathematics*, Special Issue, 120 (6), 649-660.
- [54] Cui, A. (2006). “Bus passenger origin-destination matrix estimation using automated data collection systems.” M.S. Dissertation, Massachusetts Institute of Technology, Cambridge.
- [55] Cunha, C., and Soares, C. G. (1999). “On the choice of data transformation for modelling time series of significant wave height.” *Ocean Engineering*, 26(6), 489-506.
- [56] Daldoul, M., Jarboui, S., and Dakhlaoui, A. (2015). “Public transport demand: dynamic panel model analysis.” *Transportation*, 491–505.

- [57] Dantas, T.M., Oliveira, F. L. C., and Repolho, H.M.V. (2017). “Air Transportation Demand Forecast Through Bagging Holt-Winters Methods.” *Journal of Air Transport Management*, 59, 116-123.
- [58] Das, S., and Pandit, D. (2013). “Methodology to determine level of service for bus transit in a developing country like India.” *In 13th International Conference on Computers in Urban Planning and Urban Management*, Faculty of Geosciences, Utrecht University, Utrecht, 2-5.
- [59] Deb, K., and Filippini, M. (2013). “Public bus transport demand elasticities in India.” *Journal of Transport Economics and Policy*, 47(3), 419–436.
- [60] Domencich, T. A., Kraft, G., and Valette, J. P. (1968). “Estimation of urban passenger travel behavior: An economic demand model.” *Highway Research Record*, 238, 64-78.
- [61] Doti, J, and Adibi, E. (1991). “A model for forecasting public transit.” *Transportation Research Part A: General*, 25(6), 443-450.
- [62] Directorate of Urban Land Transport. (2011). “*Comprehensive transport and traffic plan for Mangalore City- Interim Report.*” Government of Karnataka.
- [63] Economic Review. (2013). “*Final Report.*” State Planning Board, Government of Kerala, Thiruvananthapuram, Kerala.
- [64] Economic Review. (2015). “*Final Report.*” State Planning Board, Government of Kerala, Thiruvananthapuram, Kerala.
- [65] Eggermond, M. A. V., Erath, A., and Axhausen, K.W. (2016). “Vehicle ownership and usage in Switzerland: the role of micro and macro-accessibility.” *In 95th Annual Meeting of the Transportation Research Board*, Washington DC.
- [66] Famurewa, S. M., Stenström, C., Asplund, M., Galar, D., and Kumar, U. (2014). “Composite indicator for railway infrastructure management.” *Journal of Modern Transportation*, 22, 214-224.

- [67] Fathima, T. A., and Jothiprakash, V. (2014). "Behavioural analysis of a time series—A chaotic approach." *Sadhana*, 39(3), 659-676.
- [68] Fotini, M. (2017). "A new anthropocentric approach in accessibility analysis: the activity space and the accessibility measures." *Transportation Research Procedia*, 24, 491-498.
- [69] Frankena, M. W. (1978). "The demand for urban bus transit in Canada." *Journal of Transport Economics and Policy*, 280-303.
- [70] Furth, Peter G., Hemily, B., Muller, T H., and Strathman, J. G. (2006). "Using archived AVL-APC data to improve transit performance and management." Transit Cooperative Research Program (TCRP) Report 113, published by Transportation Research Board, Washington, 2006.
- [71] Garber, N. J., and Hoel, L. A. (2014). "*Traffic and highway engineering*." Cengage Learning.
- [72] Gaudry, M. (1975). "An aggregate time-series analysis of urban transit demand: The Montreal case." *Transportation Research*, 9(4), 249–258.
- [73] Gavirangaswamy, V. B., Gupta, G., Gupta, A., and Agrawal, R. (2013). "Assessment of ARIMA-based prediction techniques for road-traffic volume." *In Proceedings of the fifth international conference on management of emergent digital EcoSystems*, 246-251, ACM.
- [74] Geurs, K.T., and Wee, B. V. (2004). "Accessibility evaluation of land-use and transport strategies: review and research directions." *Journal of Transport Geography*, 12(2), 127-140.
- [75] Ghodsypour, S. H., and O'Brien, C. (1998). "A decision support system for supplier selection using an integrated analytic hierarchy process and linear programming." *International Journal of Production Economics*, 56, 199-212.

- [76] Ghosh, B., Basu, B., and O'Mahony, M. (2005). "Time-Series Modelling for Forecasting Vehicular Traffic Flow in Dublin." *In 84th Annual Meeting of the Transportation Research Board*, Washington, DC.
- [77] Gleason, J. M., and Barnum, D. T. (1982). "Toward valid measures of public sector productivity: Performance measures in urban transit." *Management Science*, 28, 379-386.
- [78] Gong, W. (2010). "ARMA-GRNN for passenger demand forecasting." *Sixth International Conference on Natural Computation*.
- [79] Goodwin, P. B. (1992). "A review of new demand elasticities with special reference to short and long run effects of price changes." *Journal of Transport Economics and Policy*, 155-169.
- [80] Guan, W., Shen, J., and Wang, W. (2002). "Using S-ANN method to forecast the ridership of Beijing public transportation." *Traffic And Transportation Studies*, 877-882.
- [81] Gulyas, A., and Kovács, Á. (2016). "Assessment of transport connections based on accessibility." *Transportation Research Procedia*, 14, 1723-1732.
- [82] Gutiérrez, J., Cardozo, O. D., and García-Palomares, J. C. (2011). "Transit ridership forecasting at station level: an approach based on distance-decay weighted regression." *Journal of Transport Geography*, 19(6), 1081-1092.
- [83] Hansen, W. G. (1959). "How accessibility shapes land use." *Journal of the American Institute of Planners*, 25(2), 73-76.
- [84] He, L., Liang, Q., and Fang, S. (2016). "Challenges and innovative solutions in urban rail transit network operations and management: China's Guangzhou metro experience." *Urban Rail Transit*, 2, 33-45.
- [85] Hensher, D. A., Stopher, P., and Bullock, P. (2003). "Service quality—developing a service quality index in the provision of commercial bus contracts." *Transportation Research Part A: Policy and Practice*, 37 (6), 499-517.

- [86] Herijanto, W., and Thorpe, N. (2005). "Developing the singly constrained gravity model for application in developing countries." *Journal of the Eastern Asia Society for Transportation Studies*, 6, 1708-1723.
- [87] Ho, H. P., Chang, C. T., and Ku, C. Y. (2013). "On the location selection problem using analytic hierarchy process and multi-choice goal programming." *International Journal of Systems Science*, 44, 94-108.
- [88] Horowitz, A., Creasey, T., Pendyala, R., and Chen, M. (2014). "Analytical Travel Forecasting Approaches for Project-Level Planning and Design." Project 08-83, Transportation Research Board.
- [89] Huang, Z., Ding, Y., and Li, J. (2009). "A GIS-based accessibility modeling process for estimating transit travel demand." *In International Conference on Management and Service Science, MASS'09, IEEE*, 1-4.
- [90] Hyndman, R. J., and Athanasopoulos, G. (2018). "Forecasting: principles and practice." OTexts.
- [91] Ingram, D. R. (1971). "The concept of accessibility: a search for an operational form." *Regional Studies*, 5(2), 101-107.
- [92] IRC-93. (1985). "Guidelines on design and installation of road traffic signals." New Dehi: The Indian Roads Congress.
- [93] Jones, D. F., and Tamiz, M. (2003). "Goal programming in the period 1990–2000." *In Multiple Criteria Optimization: State of the art annotated bibliographic surveys*, Springer, Boston, 129-170.
- [94] Kalaanidhi, S., and Gunasekaran, K. (2013). "Estimation of bus transport ridership accounting accessibility." *Procedia-Social and Behavioral Sciences*, 104, 885-893.
- [95] Karlaftis, M. G. (2004). "A DEA approach for evaluating the efficiency and effectiveness of urban transit systems." *European Journal of Operational Research*, 152, 354-364.

- [96] Karlaftis, M. G., and Tsamboulas, D. (2012). "Efficiency measurement in public transport: are findings specification sensitive?." *Transportation Research Part A: Policy and Practice*, 46, 392-402.
- [97] Kelley, W. J., and Rutherford, G. S. (1983). "Transit system performance evaluation methodology for Washington State." No. WA-RD-57.1, Final Report.
- [98] Kikuchi, S., and Miljkovic, D. (2001). "Use of fuzzy inference for modeling prediction of transit ridership at individual stops." *Transportation Research Record: Journal of the Transportation Research Board*, 25-35.
- [99] Kumar, S. V., and Vanajakshi, L. (2015). "Short-term traffic flow prediction using seasonal ARIMA model with limited input data." *European Transport Research Review*, 7(3), 21.
- [100] Kwan, M. P. (1998). "Space-time and integral measures of individual accessibility: a comparative analysis using a point-based framework." *Geographical Analysis*, 30(3), 191-216.
- [101] Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?." *Journal of Econometrics*, 54(1-3), 159-178.
- [102] Lantseva, A. A., and Ivanov, S. V. (2016). "Modeling transport accessibility with open data: Case study of St. Petersburg." *Procedia Computer Science*, 101:197-206.
- [103] Lao, Y., and Liu, L. (2009). "Performance evaluation of bus lines with data envelopment analysis and geographic information systems." *Computers, Environment and Urban Systems*, 33, 247-255.
- [104] Lee, C. W., and Kwak, N. K. (1999). "Information resource planning for a health-care system using an AHP-based goal programming method." *Journal of the Operational Research Society*, 50, 1191-1198.

- [105] Lee, J. W., and Kim, S. H. (2000). "Using analytic network process and goal programming for interdependent information system project selection." *Computers and Operations Research*, 27, 367-382.
- [106] Lewis, C. D. (1982). "*International and business forecasting methods.*" London: Butter-Worths.
- [107] Li, J., Chen, X., Li, X., and Guo, X. (2013). "Evaluation of public transportation operation based on data envelopment analysis." *Procedia-Social and Behavioral Sciences*, 96, 148-155.
- [108] Li, Z., Rose, J.M., and Hensher, D.A. (2010). "Forecasting Automobile Petrol Demand in Australia: An Evaluation of Empirical Models." *Transportation Research Part A: Policy and Practice*, 44 (1), 16-38.
- [109] Liu, M., Jiang, Y., Yang, L., and Zhang, Z. (2012). "An Elastic Analysis on Urban Public Transport Priority in Beijing." *Procedia - Social and Behavioral Sciences*, 43, 79–85.
- [110] Liu, X., and Zhou, J. (2015). "Spatial pattern of land use and its implications for mode-based accessibility: Case study of Nanjing, China." *Journal of Urban Planning and Development*, 141(2), 05014012.
- [111] Machiwele, D., and Jha, M.K. (2012). "*Hydrologic Time Series Analysis.*" Capital Publishing Company.
- [112] Mangalore Urban Development Authority. (2011). "*Mangalore Master Plan.*" Government of Karnataka
- [113] Matias, L. M., and Cats, O. (2016). "Toward a Demand Estimation Model Based on Automated Vehicle Location." *Transportation Research Record: Journal of the Transportation Research Board*, 2544, 141-149.
- [114] McNally, M. G. (2007). "*The four-step model.*" In Handbook of Transport Modelling: 2nd Edition, Emerald Group Publishing Limited, 35-53.

- [115] McTaggart, J. E. (1908). "The unreality of time." *Mind, New Series*, 17(68), 457–474.
- [116] McTaggart, J. E. (1909). "The relation of time and eternity." *Mind, New Series*, 18(71), 343–362.
- [117] Meal, J., and Carter, D. (1998). "Use of Electronic Ticket Machine Data in Transport Planning Models." *Proceedings of Seminar D Held at AET European Transport Conference*, Loughborough.
- [118] Ministry of Railways. (2015). "*Indian railways lifeline of the nation (A white paper)*." New Delhi.
- [119] Ministry of Road Transport and Highways. (2015). "*Road transport year book*." New Delhi.
- [120] Ministry of Urban Development (MoUD). (2009). "*Service level benchmarks for urban transport*." New Delhi: Ministry of Urban Development, Government of India.
- [121] Ministry of Urban Development. (2016). "*Handbook of Urban Statistics*." New Delhi.
- [122] Montgomery, D. C., Jennings, C. L., and Kulahci, M. (2015). "*Introduction to time series analysis and forecasting*." John Wiley and Sons.
- [123] Moritz, S., Sardá, A., Bartz-Beielstein, T., Zaefferer, M., and Stork, J. (2015). "Comparison of different methods for univariate time series imputation in R." *arXiv preprint arXiv:1510.03924*.
- [124] Morris, J. M., Dumble, P. L., and Wigan, M. R. (1979). "Accessibility indicators for transport planning." *Transportation Research Part A: General*, 13(2), 91–109.
- [125] Mullen, P. (1975). "Estimating the demand for urban bus travel." *Transportation*, 4(3), 231–252.

- [126] Murugesan, R., and Ramamoorthy, N. V. (2001). "Performance optimization of state transport undertaking integrating the user and operator perceptions." *Indian Highways*, 29, 11-20.
- [127] NATPAC. (2013). "*Traffic and transportation studies for 23 towns in Kerala state, Thiruvananthapuram city, Final report.*" Kerala.
- [128] Nunes, A. A., Dias T. G., and Cunha, J. F. (2016). "Passenger Journey Destination Estimation from Automated Fare Collection System Data Using Spatial Validation." *IEEE Transactions on Intelligent Transportation Systems*, 17 (1).
- [129] Odgers, J. F., and Schijndel, L. A. V. (2011). "Forecasting annual train boardings in Melbourne using time series data." *34th Australian Transport Research Forum*, Adelaide, 1–20.
- [130] OECD. (2017). "*Passenger transport (indicator).*" doi:10.1787/463da4d1-en (Accessed on 04 December 2017)
- [131] Oort, N. V., and Cats, O. (2015). "Improving public transport decision making, planning and operations by using Big Data." *IEEE 18th International Conference on Intelligent Transportation Systems (ITSC)*, 19-24.
- [132] Ortuzar, J., and Willumsen, L. G. (2011). "Modelling Transport." John Wiley and Sons, Ltd. <https://doi.org/10.1002/9781119993308>
- [133] Oum, T. H., Waters, W. G., and Yong, J. S. (1990). "*A survey of recent estimates of price elasticities of demand for transport.*" World Bank Working Paper, WP359, Washington, DC.
- [134] Panapakidis, I. P., Papagiannis, G. K., and Christoforidis, G. C. (2014). "Bus load forecasting via a combination of machine learning algorithms." *In Proceedings of the Universities Power Engineering Conference*. IEEE Computer Society.

- [135] Peng, Z. R., Dueker, K. J., Strathman, J., and Hopper, J. (1997). "A simultaneous route-level transit patronage model: demand, supply, and inter-route relationship." *Transportation*, 24(2), 159-181.
- [136] Perez, J. C., Carrillo, M. H., and Montoya-Torres, J. R. (2015). "Multi-criteria approaches for urban passenger transport systems: a literature review." *Annals of Operations Research*, 226, 69-87.
- [137] Pillai, R. S. N. (2008). "*Statistics (Theory and Practice)*." S. Chand Publishing.
- [138] Pinedo, J. (2014). "Urban Bus Demand Forecast at Stop Level: Space Syntax and Other Built Environment Factors. Evidence from Madrid." *Procedia-Social and Behavioral Sciences*, 160, 205-214.
- [139] Pitot, M., Yigitcanlar, T., Sipe, N., and Evans, R. (2006). "Land use and public transport accessibility index (LUPTAI) tool: the development and pilot application of LUPTAI for the Gold Coast." In *ATRF06 Forum Papers*, Planning and Transport Research Centre (PATREC).
- [140] Planning Commission. (2012). "*Report of the Sub-Group On State Road Transport Undertakings formed under the Working Group on Road Transport Constituted by Planning Commission (2012 – 2017)*." Government of India, 1–51.
- [141] Polat, C. (2012). "The demand determinants for urban public transport services: A review of literature." *Journal of Applied Sciences*, 12 (12), 1211-1231.
- [142] Polus, A., and Shefer, D. (1984). "Evaluation of a public transportation level of service concept," *Journal of Advanced Transportation*, 18(2), 135-144.
- [143] Preston, J. (2015). "*Public transport demand*." In *Handbook of research methods and applications in transport economics and policy*, Cheltenham, UK: Edward Elgar Publishing.

- [144] Pucher, J., Korattyswaroopam, N., and Ittyerah, N. (2004). “The crisis of public transport in India: Overwhelming needs but limited resources.” *Journal of Public Transportation*, 7(4), 1–20.
- [145] Pulugurta, S., Madhu, E., and Kayitha, R. (2015). “Fuzzy logic–based travel demand model to simulate public transport policies.” *Journal of Urban Planning and Development*, 141(4), 4014044.
- [146] Pulugurtha, S. S., and Agurla, M. (2012). “Assessment of models to estimate bus-stop level transit ridership using spatial modelling methods.” *Journal of Public Transportation*, 15(1), 3.
- [147] Qui, D. (2015). “aSTSA: Alternative Time Series Analysis (Version 3.1.2).” Retrieved from <https://cran.r-project.org/web/packages/aTSA/aTSA.pdf>
- [148] Rashidi, S., and Ranjithkar, P. (2015). “Estimation of bus dwell time using univariate time series models.” *Journal of Advanced Transportation*, 49, 139–152.
- [149] Rus, G. (1990). “Public transport demand elasticities in Spain.” *Journal of Transport Economics and Policy*, 24(2), 189–201.
- [150] Russell, B. (1901). “Is position in time and space absolute or relative?.” *Mind, New Series*, 10(39), 293–317.
- [151] Rutkowska, A. (2015). “Properties of the Cox–Stuart test for trend in application to hydrological series: the simulation study.” *Communications in Statistics-Simulation and Computation*, 44(3), 565–579.
- [152] Saaty, T.L. (1980). “The analytic hierarchy process.” New York, McGraw-Hill.
- [153] Saaty, T.L. (1994). “How to make a decision: the analytic hierarchy process.” *Interfaces* 24, 19–43.

- [154] Saaty, T. L. (1995). "Transport planning with multiple criteria: the analytic hierarchy process applications and progress review." *Journal of Advanced Transportation*, 29, 81-126.
- [155] Safwat, K. N., and Magnanti, T. L. (1988). "A combined trip generation, trip distribution, modal split, and trip assignment model." *Transportation Science*, 22(1), 14-30.
- [156] Saghapour, T., Moridpour, S., and Thompson, R.G. (2016). "Public transport accessibility in metropolitan areas: A new approach incorporating population density." *Journal of Transport Geography*, 54, 273-285.
- [157] Salonen, M., and Toivonen, T. (2013). "Modelling travel time in urban networks: comparable measures for private car and public transport." *Journal of Transport Geography*, 31, 143-153.
- [158] Sampaio, B. R., Neto, O. L., and Sampaio, Y. (2008). "Efficiency analysis of public transport systems: Lessons for institutional planning." *Transportation research part A: Policy and Practice*, 42, 445-454.
- [159] Schmenner, R. W. (1976). "The Demand for Urban Bus Transit: A Route-by-Route Analysis." *Journal of Transport Economics and Policy*, 10(1), 68–86.
- [160] Sen, P., Roy, M., and Pal, P. (2016). "Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization." *Energy*, 116, 1031-1038.
- [161] Seyedabrishami, S., and Shafahi, Y. (2011). "Expert knowledge-guided travel demand estimation: Neuro-fuzzy approach." *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 15(1), 13–27.
- [162] Shah, J. S., and Adhvaryu, B. (2016). "Public transport accessibility levels for Ahmedabad, India." *Journal of Public Transportation*, 19(3), 19-35.
- [163] Skarphedinsson, A., (2013). "Evaluating a simplified process for developing a four-step transport planning model in VISUM –Application on the capital area

of Reykjavik.” Lund, Lunds University, LTH, Institutionen för Teknik och samhälle. Trafik och väg 2013. Thesis. 238.

- [164] Slutsky, E. (1937). “The summation of random causes as the source of cyclic processes.” *Econometrica*, 5(2), 105–146.
- [165] Smit, M., (2014). “*Exploring the demand for bus transport.*” Master Thesis, University of Twente.
- [166] Sumetha, R., and Mathew, S (2016). “Analyzing the Demand for Public Transport through Behavioral Mode Choice Models for Work Trips in Tiruchirappalli City.” *12th TPDMC*.
- [167] Sun, Y., Zhang, C., Dong, K., and Lang, M. (2018). “Multi-agent modelling and simulation of a physical internet enabled rail-road intermodal transport system.” *Urban Rail Transit*, 4, 141-154.
- [168] Sweet, R. J. (1997). “An aggregate measure of travel utility.” *Transportation Research Part B: Methodological*, 31(5), 403-416.
- [169] Talvitie, A. (1973). “A direct demand model for downtown work trips.” *Transportation*, 2(2), 121–152.
- [170] Tamin, O. Z. (1997). “Public transport demand estimation by calibrating a trip distribution-mode choice (TDMC) model from passenger counts: A case study in Bandung, Indonesia.” *Journal of Advanced Transportation*, 31(1), 5-18.
- [171] Taylor, J.W. (2003). “Short-Term Electricity Demand Forecasting using Double Seasonal Exponential Smoothing.” *Journal of the Operational Research Society*, 54 (8), 799-805.
- [172] Tegner, G. and Holmber, I. (1998). “Public transport demand analysis-a non-linear time-series model for Stockholm region.” *In transportation planning methods. Proceedings of seminar held at AET European Transport Conference*, Loughborough University, UK, 14-18 September 1998.

- [173] TERI. (2015). “*Green growth and transport in India.*” New Delhi.
- [174] Thiruvananthapuram Corporation. (2012). “*Thiruvananthapuram Master Plan (Draft).*” Government of Kerala, Trivandrum.
- [175] Tsai, C. H. P., Mulley, C., and Clifton, G. (2012). “The spatial interactions between public transport demand and land use characteristics in the Sydney Greater Metropolitan Area.” *Road and Transport Research: A Journal of Australian and New Zealand Research and Practice*, 21(4), 62.
- [176] Tsai, C. H. P., Mulley, C., and Clifton, G. (2013). “Forecasting public transport demand for the Sydney Greater Metropolitan Area: A comparison of univariate and multivariate methods.” *Australasian Transport Research Forum 2013 Proceedings*, 2 - 4 October 2013, Brisbane, Australia.
- [177] Tsai, C. H. P., Mulley, C., and Merkert, R. (2015). “Measuring the cost efficiency of urban rail systems an international comparison using DEA and tobit models.” *Journal of Transport Economics and Policy*, 49, 17-34.
- [178] Tsai, T.H., Lee, C. K., and Wei, C.H. (2005). “Design of dynamic neural networks to forecast short-term railway passenger demand.” *Journal of Eastern Asia Society for Transportation Studies*, 6, 1651-1666.
- [179] Tuan, V.A., and Son, D.T. (2015). “Accessibility to public transport and developing cities- an empirical study in Ho Chi Mihn City Vietnam.” *Journal of the Eastern Asia Society for Transportation Studies*, 11,1240-1258.
- [180] Tyrinopoulos, Y., and Antoniou, C. (2008). “Public transit user satisfaction: Variability and policy implications.” *Transport Policy*, 15 (4), 260-272.
- [181] Vaidya, O. S. (2014). “Evaluating the performance of public urban transportation systems in India.” *Journal of Public Transportation*, 17, 174-191.
- [182] Venkatesh, A., and Kushwaha, S. (2018). “Short and long-run cost efficiency in Indian public bus companies using Data Envelopment Analysis.” *Socio-Economic Planning Sciences*, 61, 29-36.

- [183] Vieira, R.S., and Haddad, E. A. (2012). “An accessibility index for the metropolitan region of São Paulo.” *The Rise of the City: Spatial Dynamics in the Urban Century*, 242.
- [184] Vuchic, V. R. (2007). “*Urban transit systems and technology*.” John Wiley and Sons.
- [185] Wang, C.H., and Chen, N.A. (2015). “A GIS-based spatial statistical approach to modelling job accessibility by transportation mode: case study of Columbus, Ohio.” *Journal of Transport Geography*, 45,1-11.
- [186] Wang, H., Shen, Q., and Tang, B.S. (2014). “GIS-based framework for supporting land use planning in urban renewal: Case study in Hong Kong.” *Journal of Urban Planning and Development*, 141(3), 05014015.
- [187] Wey, W. M., and Wu, K. Y. (2007). “Using ANP priorities with goal programming in resource allocation in transportation.” *Mathematical and Computer Modelling*, 46, 985-1000.
- [188] Wijeweera, A., To, H., Charles, M. B., and Sloan, K. (2014). “A time series analysis of passenger rail demand in major Australian cities.” *Economic Analysis and Policy*, 44(3), 301–309.
- [189] Williams, B. M., and Hoel, L. A. (2003). “Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results.” *Journal of Transportation Engineering*, 129(6), 664-672.
- [190] Wirasinghe, S. C., and Kumarage, A. S. (1998). “An aggregate demand model for intercity passenger travel in Sri Lanka.” *Transportation*, 25(1), 77–98.
- [191] Woldeamanuel, M., and Kent, A. (2015). ”Measuring walk access to transit in terms of sidewalk availability, quality, and connectivity.” *Journal of Urban Planning and Development*, 142(2), 04015019.

- [192] Working Group on Urban Transport (2012). “*Final Report.*” National Transport Development Policy Committee. New Delhi. Ministry of Urban Development, Government of India.
- [193] Working group of Urban Transport. (2013). “*Section 5-Urban Transport- India Transport Report Moving India to 2032.*” National Transport Development Policy Committee, Ministry of Urban Development, Government of India.
- [194] Wu, L., Xia, H., Cao, X., Zhang, C., and Dai, C. (2018). “Research on quantitative demand of underground space development for urban rail transit station areas: a case study of metro line 1 in Xuzhou, China.” *Urban Rail Transit*, 4, 257-273.
- [195] Xue, R., Sun, D. J., and Chen, S. (2015). “Short-term bus passenger demand prediction based on time series model and interactive multiple model approach.” *Discrete Dynamics in Nature and Society*, 2015.
- [196] Yaldi, G., Taylor, M.A.P., and Yue, W.L. (2010). “Developing fuzzy-neuro model for travel demand modeling.” *Journal of Eastern Asia Society for Transportation Studies*, 8.
- [197] Yannis, G., Antoniou, C., and Papadimitriou, E. (2011). “Autoregressive nonlinear time-series modeling of traffic fatalities in Europe.” *European Transport Research Review*, 3, 113-127.
- [198] Yan-Yan, C., Pan-Yi, W., Jian-Hui, L., Guo-Chen, F., Xin, L., and Yi, G. (2016). “An evaluating method of public transit accessibility for urban areas based on GIS.” *Procedia Engineering*, 137,132-140.
- [199] Yao, E., and Morikawa, T. (2005). “A study of on integrated intercity travel demand model.” *Transportation Research Part A: Policy and Practice*, 39(4), 367–381.

- [200] Yatskiv, I., Budilovich, E., and Gromule, V. (2017). "Accessibility to Riga public transport services for transit passengers." *Procedia Engineering*, 187, 82-88.
- [201] Yeh, C. H., Deng, H., Chang, Y. H. (2000). "Fuzzy multicriteria analysis for performance evaluation of bus companies." *European Journal of Operational Research*, 126, 459-473.
- [202] Yule, G. U. (1921). "On the time-correlation problem , with especial reference to the variate- difference correlation method." *Journal of the Royal Statistical Society*, 84(4), 497–537.
- [203] Yule, G. U. (1927). "On a method of investigating periodicities in disturbed series , with special reference to wolfer ' s sunspot numbers." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 226, 267–298.
- [204] Yule G. U. (1926). "Why do we sometimes get nonsense-correlations between time-series ? --A study in sampling and the nature of time-series." *Journal of the Royal Statistical Society*, 89(1), 1–63.
- [205] Zambom, A. Z., and Dias, R. (2012). "A review of kernel density estimation with applications to econometrics." *arXiv preprint arXiv:1212.2812*.
- [206] Zhang, G.P. (2003). "Time series forecasting using a hybrid ARIMA and neural network model." *Neurocomputing*, 50, 159-175.
- [207] Zhang, R., and Wu, X. L. (2007). "Study On The Demand Forecast Method For The Interurban Public Transport Under The High-Speed Railways In Shanghai-Nanjing Corridor." *International Conference Series on Competition and Ownership in Land Passenger Transport*, Hamilton Island, Queensland, Australia, 2007.

- [208] Zhang, Y., Zhang, Y., and Haghani, A. (2014). "A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model." *Transportation Research Part C: Emerging Technologies*, 43, 65-78.
- [209] Zhao, F., Chow, L. F., Li, M., and Liu, X. (2005). "A transit ridership model based on geographically weighted regression and service quality variables." Lehman Center for Transportation Research, Florida International University, Miami, Florida.
- [210] Zhao, J., Deng, W., Song, Y., and Zhu, Y. (2014). "Analysis of Metro ridership at station level and station-to-station level in Nanjing: an approach based on direct demand models." *Transportation*, 41(1), 133-155.

LIST OF PUBLICATIONS

Journal:

1. Anila Cyril, Raviraj H. Mulangi and Varghese George (2018). "Data Processing Methodology for Big Data in Public Transportation". International Journal of Pure and Applied Mathematics, Special Issue, Vol. 120, No.6, 649-660.
2. Anila Cyril, Raviraj H. Mulangi and Varghese George (2019). "Bus Passenger Demand Modelling Using Time-series Techniques- Big Data Analytics". The Open Transportation Journal, Vol. 13, 41-47.
3. Anila Cyril, Raviraj H. Mulangi and Varghese George (2019). "Performance Optimization of Public Transport Using Integrated AHP-GP Methodology". Urban Rail Transit, Vol. 5, No. 2, 133-144.
4. Anila Cyril, Raviraj H. Mulangi and Varghese George (2019). "Level of Service Analysis of Public Transport Facilities: A Case Study of Trivandrum City". International Journal of Recent Technology and Engineering (IJRTE), Vol. 8, No. 2S3, 1019-1022.
5. Anila Cyril, Raviraj H. Mulangi and Varghese George (2019). "Development of a GIS-Based Composite Public Transport Accessibility Index". Journal of Urban and Environmental Engineering (Accepted for Publication)
6. Anila Cyril, Raviraj H. Mulangi and Varghese George (2019). "A Time Series Approach for Modelling Public Bus Passenger Demand". Transpometrica A: Transportation Science (Communicated)

International Conferences:

1. Anila Cyril, Varghese George and Raviraj H. Mulangi. (2017). "Electronic Ticket Machine Data Analytics for Public Bus Transport Planning". IEEE International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS 2017), 1-2 August, 2017, Chennai. (Published in IEEE Xplore Digital Library)

2. Anila Cyril, Raviraj H. Mulangi and Varghese George (2018). “Modelling and Forecasting Bus Passenger Demand Using Time Series Method”. IEEE 7th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO 2018), 29-31 August, 2018, Amity University, Noida (Published in IEEE Xplore Digital Library)
3. Anila Cyril, Raviraj H. Mulangi and Varghese George (2019). “Short Term Bus Passenger Demand Estimation Using Time Series Method” International Engineering Symposium (IES 2019), 13-15 March 2019, Japan
4. Anila Cyril, Raviraj H. Mulangi and Varghese George (2019). “Demand-Based Model for Line Planning in Public Transport”. World Conference on Transport Research - WCTR 2019, IIT Mumbai, 26-31 May 2019 (Accepted for publishing in Transport Research Procedia- Elsevier)

BIO-DATA

Name : Anila Cyril
Department : Civil Engineering
Specialization : Transportation Engineering
Gender : Female
Date of Birth : 30/06/1989
Email Id : anilacyril110@gmail.com
Mobile Number : +91-9400220928
Address : Puthenpuraickal House
Punnapra P.O.
Alappuzha, Kerala
Pin: 688004

Academic Profile:

Degree	Year of Completion	Institution	Board	Percentage/CGPA
Ph.D. (Civil Engineering)	2019	National Institute of Technology Karnataka	Deemed University	8.5
MTECH (Transportation Engineering)	2013	Government Rajiv Gandhi Institute of Technology Kottayam	MG University	9.01
BTECH (Civil Engineering)	2011	Government Engineering College Trichur	University of Calicut	73.94 %
XII (PCMB)	2007	Fr Thomas Porukara Central School Champakulam Alappuzha	CBSE	85 %
X	2005	St Aloysius Secondary School Punnapra Alappuzha	CBSE	79.6 %