

LB/TH/01/2026

TH6101

**USE OF HIGH FREQUENCY DATA IN TRANSPORT
AND LAND USE INTEGRATED MODEL**

Thivya Precilla Amalan

(198120N)

Degree of Doctor of Philosophy

Department of Transport Management and Logistics Engineering

University of Moratuwa

Sri Lanka

October 2024

USE OF HIGH FREQUENCY DATA IN TRANSPORT AND LAND USE INTEGRATED MODEL

Thivya Precilla Amalan

(198120N)

Thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Department of Transport Management and Logistics Engineering

University of Moratuwa
Sri Lanka

October 2024

CERTIFICATION

I, Thivya Precilla Amalan, declare that this thesis submitted in fulfilment of the requirements for the conferral of the degree Doctor of Philosophy, from the University of Moratuwa, is wholly my work unless otherwise referenced or acknowledged. This document has not been submitted for qualifications at any other academic institution.

Thivya Precilla Amalan
Department of Transport Management and
Logistics Engineering
University of Moratuwa
Sri Lanka
31st December 2024

The above candidate has carried out research for the PhD thesis under our supervision.

	21/02/2025
Supervisor	
Prof Amal S Kumarege	Date
Department of Transport Management	
and Logistics Engineering	
University of Moratuwa	
Sri Lanka	

	21/02/2025
Supervisor	
Dr Bo Du	Date
Department of Business Strategy and	
Innovation	
Griffith University	
Australia	

	21/02/2025
Supervisor	
Dr Rohan Wickramasuriya	Date
Endeavor Energy	
Australia	

LIST OF PUBLICATIONS

1. Amalan, T. P., Du, B., Kumarage, A. S., Wickramasuriya, R., & Perez, P. (2023). High frequency data in Land use and transport integrated model: A review of sources and application. *Asian Transport Studies*, 9, 100092. doi: <https://doi.org/10.1016/j.eastsj.2022.100092>
2. Amalan, T. P., Du, B., Kumarage, A. S (2023). A Fuzzy Logic based Trip Generation Approach using High-Frequency Data. *Travel Behaviour and Society*, Under Review
3. Amalan, T. P., & Wickramasuriya, R., (2024). Developing Trip Generation and Attraction Model using High Frequency Data. *JSALT*, Under Review

ABSTRACT

Efficient urban planning is crucial for creating liveable and sustainable cities. This involves designing urban areas that maximize economic, social, and environmental viability while minimizing negative impacts. Land use and transport development play a significant role in shaping transport patterns and infrastructure needs. However, applying Land Use and Transport Integrated (LUTI) models in developing countries can be challenging due to limited resources and data availability. This dissertation aims to explore freely, or low-cost high-frequency (HF) data related to land use and transport and develop an integrated model using available data in Sri Lanka as an example for developing countries.

The complexity of the urbanisation process, particularly in developing countries, demands sophisticated decision support tools such as the Land Use and Transport Integrated (LUTI) models to assist in the balanced and sustainable development of transport and land use. However, the high cost and time-consuming nature of LUTI models have limited their application in developing countries due to limited resources and funding.

The advent of information and communication technology (ICT) has revolutionised data collection and analysis, providing more extensive and high frequency (HF) data, such as real-time traffic data collected from sensors and mobile devices. This data can be used for various purposes, including traffic management, urban planning, public health and safety, and research and innovation. The availability of high-frequency data presents both opportunities and challenges for urban planning and development. However, based on the literature review, application of HF data in developing countries are very limited due to the data availability such as high resolution satellite data, technological constraints, lack of understanding of potential benefits of HF data, data sensitivity and privacy concerns and resource constraints including financial and human capacity. For these cases, researchers and planners can use HF proxy data as a substitute for HF data.

The study aims to extend the use of LUTI models by incorporating HF proxy data in Western Province, Sri Lanka, and demonstrates the approach's effectiveness. It utilizes electricity consumption data, GPS point data of customers, and satellite imagery to develop the model. The resulting trip generation model achieves high accuracy in predicting travel demand of 79% in 2013. Additionally, the study uses

satellite imagery and supervised classification to obtain land use area information. Separate trip attraction models for general and industrial areas are developed, achieving accuracy of 70% and 79.5% for general and industrial purpose related trip in predicting attracted trips, respectively. A gravity model is then employed to obtain the final trip distribution, highlighting the value of incorporating household electricity consumption data and satellite imagery in travel demand modelling. The research involves assessing the impact of land use and transport on each other based on the developed model. Different scenarios are tested, revealing insights into various transportation metrics such as work trips, passenger kilometres, average trip length, passenger hours, average speed, and costs. The comparison underscores the benefits of land use improvements and a developed transport network in optimising travel patterns, increasing passenger kilometres, and enhancing system efficiency. This emphasizes the importance of careful analysis in transportation and land use decision-making.

Finally, the research also focuses on developing a disaggregated trip generation model using HF proxy data and assessing its accuracy. By utilizing a fuzzy model based on electricity consumption data and satellite imagery, the study overcomes the limitations of traditional models that rely on costly and challenging data acquisition. The developed disaggregated model achieves high accuracy of 88.7% calculating trip rates compared to the aggregated model.

Overall, this dissertation showcases the potential of leveraging HF data for efficient urban planning in developing countries. It highlights the importance of incorporating household electricity consumption data and satellite imagery in travel demand modelling, and the benefits of land use improvements and a developed transport network in optimising travel patterns and enhancing system efficiency. The research findings contribute to the field of urban planning and provide valuable insights for planners and decision-makers in developing sustainable and liveable cities.

ACKNOWLEDGMENTS

It is my great pleasure to present my thesis, which could not have been fulfilled without acknowledging the great people who were the backbone of this successful research.

First and foremost, I would like to express my deepest gratitude to my supervisor Dr. Bo Du, Prof. Amal.S.Kumarage, Dr. Rohan Wickramasuriya, and Prof. Pascal Perez for their tremendous guidance, assistance, motivational continuous encouragement, and continuous assessment and evaluation, which put me on the right path to success. Moreover, I am grateful for their dedicated support and optimum patience throughout my research journey.

I would also like to extend my sincere gratitude to all academic heads of the University of Wollongong, Australia, and the University of Moratuwa, Sri Lanka, for the facilitation provided to make this thesis successful.

Furthermore, it is inevitable to acknowledge my gratefulness towards the Ceylon Electricity Board (CEB) and Lanka Electricity Company (Private) Limited (LECO) for providing access for data collection within their busy working hours. My thesis would not have been possible without their support.

I would like to express my sincere appreciation to Accelerating Higher Education Expansion and Development (AHEAD), which is a World Bank-funded Sri Lankan government operation to support the higher education sector, for the generous scholarship they awarded me during the completion of this thesis. This financial assistance has been instrumental in enabling me to focus on my research and successfully complete my academic journey.

It is my obligation to convey my appreciation to all academic and non-academic staff of the University of Wollongong, Australia, and the University of Moratuwa, Sri Lanka, for the assistance given to achieve the research objective.

Table of Contents

Abstract	1
Acknowledgments	5
Certification.....	1
List of Names or Abbreviations	13
List of Figures	6
List of tables.....	12
Chapter 1	13
Introduction	1
1.1 Overview	1
1.2 Research Problem.....	5
1.3 Research Aim	6
1.4 Research Objectives	7
1.5 Overall Methodology	8
1.6 Structure of the Dissertation.....	11
1.7 Significance of the study	13
Chapter 2	15
A Review of High Frequency Data use in LUTI Models	15
2.1 Introduction	15
2.2 Literature Search (Overview/Summary)	18
2.3 Theoretical Background of the LUTI Framework.....	19
2.3.1 Land-use Modelling Techniques.....	19
2.3.2 Transport Modelling Techniques	20
2.4 LUTI Model Evolution.....	22
2.5 Conventional LUTI Models and Relevant Data Sources	25
2.5.1 Analysing of the Inputs of Comprehensive Transport Models and Land-Use Models 25	
2.5.2 Data Used in Existing LUTI Models	26
2.6 High Frequency Data.....	29
2.6.1 Mobile CDR Data	30
2.6.2 Satellite Imagery Data.....	31
2.6.3 Smart Card Data.....	34
2.6.4 GPS Data.....	34
2.7 Use of HFD in Developing Countries	35
2.7.1 Application and Readiness of the Data	35
2.7.2 Application of HF Proxy Data	36
2.8 Adopting HF data into the LUTI Model	37

2.9	Summary and Conclusion	38
Chapter 3		40
Application of HF Data Sources for LUTI model: A Case Study on Western Province, Sri Lanka.....		40
3.1	Overview of the Chapter	40
3.2	Overview of Study Area.....	41
3.3	Transport Pattern (Home-based Work) of Western Province	44
3.4	Overview of HF Data in Western Province.....	47
3.5	Analysis of Electricity Consumption GPS Data.....	47
3.5.1	Overview of Electricity Consumptions in Sri Lanka	47
3.5.2	Analysis of Satellite Data.....	54
3.6	Summary and Conclusion	61
Chapter 4.....		63
Developing the Trip Generation and Attraction Models using High-Frequency Proxy Data		63
4.1	Introduction	63
4.2	Data set and Methodology.....	63
4.3	Overview of Trip Generation Models	65
4.4	Regression Model for Trip Generation.....	66
4.4.1	Factors affecting Home based Work Trip Generation	67
4.5	Relationship between Income and Electricity Consumption.....	67
4.5.1	Relationship of Electricity Consumption based on HVS Survey 2013.....	71
4.6	Developing Trip Generation Model and Validation.....	72
4.7	Developing Trip Attraction Model and Validation	75
4.7.1	Trip Attraction Model for the General Purpose and the Industrial Area ..	76
4.8	Trip Distribution Model	81
4.8.1	Growth Factor Methods	81
4.8.2	Synthetic or Gravity Models	82
4.8.3	Entropy Maximisation Model	83
4.9	Application of Trip Distribution Model	83
4.9.1	Calibration of Model Parameters based on CoMTrans Study 2013	83
4.9.2	Model Assessment	85
4.10	Sensitivity Analysis.....	85
4.11	Summary and Conclusion	87
Chapter 5.....		88
Development of the LUTI Model with Identified HFD and Scenario Testing.....		88
5.1	Introduction	88
5.2	Methodology Used to Develop NRT-LUTI Model.....	90

5.2.1	Overview of Trip Assignment.....	91
5.2.2	Overview of JICA STRADA	92
5.2.3	Basic Assumptions and Inputs for the Model Development.....	96
5.2.4	Origin and Destination Matrix of 2013 and 2019	101
5.3	Traffic Assignment Results and Validation of Scenario 1 (2013) and Scenario 2 (2019)	102
5.4	Scenario Testing for LUTI Integration.....	106
5.4.1	Overview	106
5.4.2	Scenario 3: Do Nothing (2030).....	106
5.4.3	Scenario 4: Land use Improvement (2030).....	111
5.4.4	Scenario 5: Land use and Transport Network Development	117
5.5	Cost Analysis of Each Scenario	119
5.5.1	Value of Time (VOT).....	120
5.5.2	Vehicle Operating Cost (VOC).....	120
5.5.3	Accident Cost	120
5.5.4	Emission Cost	121
5.5.5	Estimation of Results	121
5.6	Summary and Conclusion	124
Chapter 6		125
Developing A Fuzzy Rule-Based Trip Generation using High Frequency Data		125
6.1	Introduction	125
6.2	Background Study	127
6.3	Methodology	130
6.3.1	Overall Methodology of the Study.....	130
6.3.1	Preparation of Satellite Image of the Study Area.....	131
6.3.2	Detecting Building Footprint Using Deep Learning.....	132
6.3.3	Overview of the Fuzzy Logic Model	135
6.4	Data Analysis	136
6.4.1	Description of the study area.....	136
6.4.2	Using the Deep Learning Model.....	137
6.4.3	Extraction of Building Footprint.....	140
6.5	Primary Analysis of Data Collection.....	141
6.5.1	Relationship between Income and Electricity Consumption	141
6.6	Validation of Electricity Domestic Customers' GPS Location Data	142
6.7	Results of Fuzzy Logic Model for Trip Generation	143
6.8	Summary and Conclusion	148
6.8.1	Detecting Building Height from the Satellite Image.....	150

Chapter 7	156
Conclusion and Recommendation.....	156
7.1 General	156
7.2 Summary of Key Findings	157
7.3 Recommendation and Future Research	161
Bibliography or List of References	165
Appendices.....	173
Appendix 1	173
Distance Matrix.....	173
Appendix 2.....	174
2013 OD Matrix	174
Appendix 3.....	175
RDA Projects	175
Appendix 4 2019 OD Matrix	177
Appendix 5.....	178
2030 OD Matrix (Scenario 3)	178
Appendix 6.....	179
Matrix (Scenario 4)Land use Development	179
Appendix 7.....	180
Road Network 2013	180
Appendix 8.....	181
Road Network 2019 and Capacity	181
Appendix 9.....	182
Deep Learning Model- USA.....	182
Building Footprint-USA	182
Appendix 10.....	183
Deep Learning Model- Africa.....	183
Building Footprint-Africa	183
Appendix 11	184
Deep Learning Model- 30 Samples.....	184
Building Footprint- Fine-tuned model with 30 Samples	184
Appendix 12.....	185
Case 5: Retraining USA model- Pre trained	185
Case 5: Retraining Africa- Pre trained	186
Appendix 13.....	187
Road Network 2030	187

LIST OF FIGURES

Figure 1.1: Land use Feedback Cycle	2
Figure 1.2: Conceptual Approach	10
Figure 2.1: Application of LUTI (Source: Scopus Database).....	16
Figure 2.2: Schedule Vs Trips Vs Tours.....	21
Figure 2.3: Evolution of LUTI Models	23
Figure 2.4: HF Data in Transport and Land use.....	29
Figure 2.5: Smart Card application in Transport	35
Figure 2.6: GPS application in Transport	35
Figure 2.7: CDR data in Transport.....	35
Figure 2.8: RS in Satellite	35
Figure 2.9: HF Data Vs HF Proxy Data.....	36
Figure 3.1: Overall Study Area (Western Province, Sri Lanka)	41
Figure 3.2 Average Land Price in Western Province (2019)	43
Figure 3.3: Land Assessment Index in Western Province (2019).....	44
Figure 3.4: Home to Work Trip Density in Western Province (2013).....	45
Figure 3.5: Hourly Fluctuation of Work Related Trips in Western Province (2013)	46
Figure 3.6: Percentage Use by Energy Sources in Western Province by Year.....	48
Figure 3.7: Tax Classification of Electricity Providers.....	48
Figure 3.8: Census Vs GPS Electricity Data- 2013	50
Figure 3.9: Households Growth Rate (2013 and 2019)	51
Figure 3.10: General Purpose Growth rate (2013 and 2019).....	52
Figure 3.11: Industrial Buildings Growth rate (2013 and 2019).....	53
Figure 3.12: Spatial and Temporal Resolution of Satellite Images	54
Figure 3.13: Journal Papers (Open access) Published using Landsat Data (2000 to 2022)	56
Figure 3.14: Landsat Satellite Images (2013 and 2019)	58
Figure 4.1: Area Classification of CEB	64
Figure 4.2: No of Households vs Total Trips Generated	65
Figure 4.3: GDP Vs Electricity Growth Rate- Sri Lanka	69
Figure 4.4: Relationship between Per Capita Income vs Per Capita Average Electricity Consumption	69
Figure 4.5: Trend of GDP and Domestic Electricity Sales of Western Province	70
Figure 4.6: Average Electricity Consumption by TAZ (2013).....	71
Figure 4.7: Households' Electricity Consumption vs Income (2013).....	72
Figure 4.8: Home to Work Trip Generation (Actual vs Predicted)- Western Province (2013).....	74
Figure 4.9: Home to Work Trip Generation in Western Province (2013 and 2019)	74
Figure 4.10: Methodology used for Calculating Area of General and Industrial	75
Figure 4.11: Trip Attraction (General Purpose) Actual vs Predicted - 2013	79
Figure 4.12: Trip Attraction (Industrial) Actual vs Predicted - 2013	80
Figure 4.13: Summary of Trip Attraction between 2013 and 2019	80
Figure 4.14: Selected TAZs for Gravity Model Application	84
Figure 4.15: Distance Vs Cumulative % Trips	85
Figure 4.16: Annual Electricity Consumption (Western Province)	86
Figure 5.1: Overall Framework of NRT-LUTI model.....	90
Figure 5.2: Flow Diagram for JICA STRADA Application.....	93
Figure 5.3: Overall Methodology of JICA STARDA Model Development.....	96
Figure 5.4: Traffic Analysis Zones (TAZ) of Study Area (Western Province)	97

Figure 5.5: Household Income Distribution of Western Province (2013).....	98
Figure 5.6: Household Income Distribution of Western Province (2019).....	99
Figure 5.7: Road Projects Considered between 2013 and 2019 in Western Province .	101
Figure 5.8: Desire Map of 2013 and 2019	102
Figure 5.9: Survey Points Considered for the Comparison at CMC Boundary	103
Figure 5.10: Volume/ Capacity Ratio (VCR) of Scenario 1 (2013)	104
Figure 5.11: Volume/ Capacity Ratio of Scenario 2 (2019)	105
Figure 5.12: Trip Generation 2013, 2019 and 2030.....	108
Figure 5.13: Trip Attraction 2013, 2019 and 2030	110
Figure 5.14: Volume/ Capacity Ratio of Scenario 3 (DNS 2030)	111
Figure 5.15: Land use Development Plan DSDs 2030	112
Figure 5.16: Trip Generation based on Land use Development Scenario 4	115
Figure 5.17: Trip Attraction based on Land use Development Scenario 4.....	116
Figure 5.18: Volume/ Capacity Ratio Scenario 4 (2030).....	117
Figure 5.19: Volume/ Capacity Ratio of Scenario 5 (2030)	119
Figure 6.1: Overall Methodology.....	131
Figure 6.2: Mosaicked Image of Thimbirigasyaya	132
Figure 6.3: Methodology of Building Footprint Extraction and Map Matching of GPS	135
Figure 6.4: Study Area and TAZs of Thimbirigasyaya DSD.....	137
Figure 6.5: Methodology used Transfer Learning	138
Figure 6.6: Building Detected and Samples.....	138
Figure 6.7: Sample of Detected Buildings using DL	140
Figure 6.8: Electricity Consumption according to the Income (Thimbirigasyaya DSD)	142
Figure 6.9: Decision Tree (Trip Generation vs Electricity Consumption and Area of Households)	143
6.10a: Electricity Membership Function.....	146
6.10c: Output Variable (TG).....	146
6.10b: Area Membership Function	146
6.10d: Surface View of Fuzzy Model	146
Figure 6.10: Fuzzy Logis Model Input and Output.....	146
Figure 6.11: Work based Trip Generation for 2013 and 2019.....	148
Figure 6.12: No of Condominium Units in Colombo	150
Figure 6.13: Methodology for Extracting Building Height from Google Satellite Images	151
Figure 6.14: Methodology for Sample Apartment Building from the Street View and the Counting of Floors	152
Figure 6.15: Building Height Based on the Image Shadow.....	153
Figure 6.16: Actual Height vs Height from Satellite	155

LIST OF TABLES

Table 2.1: Comparison of Activity-Based vs Trip-Based Model	21
Table 2.2: Comparison of Components in Different LUTI Models.....	24
Table 2.3: Summary of Data Sources of Selected LUTI Models	28
Table 2.4: Summary of Data Sources Used for Land Use Classifications	33
Table 2.5: LUTI Models with Key Requirements	38
Table 3.1: Census Vs GPS Counts of Households in Western Province (2013)	49
Table 3.2: Growth rate of Households in Western Province (2013 and 2019).....	50
Table 3.3 Land use Classification by USGS.....	56
Table 3.4: Land use Classification Considered for Studies	57
Table 3.5: Supervised Classification of Landsat images for Thimbirigasyaya DSD.....	60
Table 3.6: Actual vs Supervised Land use Classifications in 2013	60
Table 4.1: Annual Growth Rate	65
Table 4.2: Literature on Income vs Electricity Consumption.....	68
Table 4.3: Descriptive Statistics of Income vs Electricity Consumption.....	72
Table 4.4: Regression Analysis (Trip Generation)- 2013	73
Table 4.5: Area of General Purpose and Industrial Buildings in 2019	76
Table 4.6: Regression Model (General Purpose)- 2013.....	77
Table 4.7: Regression Model (Industrial)- 2013.....	77
Table 4.8: Trip Attraction (General Purpose Buildings 2013).....	78
Table 4.9: Trip Attraction (Industrial)- 2013	79
Table 4.10: Annual Electricity Consumption (Western Province)	86
Table 4.11: Sensitivity Test for Trip Attraction.....	86
Table 5.1 Description of Scenario Cases	88
Table 5.2: Time Value- 2013	98
Table 5.3: Time Value - 2019	99
Table 5.4: Comparison of 2013 Assignment Output with Survey Data	103
Table 5.5: Trip Generation 2030.....	107
Table 5.6: Trip Attraction 2030	109
Table 5.7: Land use Development Plan 2030	113
Table 5.8: Road Development Plan (Scenario 5- 2030)	118
Table 5.9: Vehicle Operating Cost.....	120
Table 5.10: KPI of Scenario Cases	123
Table 6.1: Cases Considered for Transfer Learning	139
Table 6.2: Detected HHS vs GPS counts (2013 and 2019).....	141
Table 6.3: Descriptive Statistics of Income vs Electricity Consumption.....	142
Table 6.4: Households Census vs GPS data 2013.....	143
Table 6.5: Brief Description of Mean and Standard Deviation (SD).....	144
Table 6.6: Fuzzy Model Output for 2013 and 2019.....	147
Table 6.7: Total Trip Generation for Households (HHS) Based on Fuzzy Model	147
Table 6.8: Standard Height of the Building Applied in Sri Lanka.....	151
Table 6.9: Samples of Buildings Information	153
Table 6.10: Relationship between Actual Height and Shadow Height.....	154

LIST OF NAMES OR ABBREVIATIONS

LUTI	: Land use and Transport Integrated Model
ICT	: Information and Communication Technology
RS	: Remote Sensing
CDR	: Call Details Record
GPS	: Global Positioning System
GDP	: Gross Domestic Product
USD	: United States Dollars
DSD	: Divisional Secretariat Division
GND	: Grama Niladari Division
NRT-LUTI	: Near Real Time- Land use and Transport Integrated Model
HVS	: Home Visit Survey
CoMTrans	: Urban Transport System Development Project for Colombo Metropolitan Region and Suburbs
UDA	: Urban Development Authority
MLR	: Multiple Linear Regression
OD	: Origin- Destination matrix
UOM	: University of Moratuwa
JICA	: Japan International Cooperation Agency
CVC	: Classified Vehicle Count
LUS	: Land use Survey
FSM	: Four step model
ABM	: Activity Based Model
USGS	: United States Geological Survey
VGI	: Volunteered Geographic Information
RFID	: Radio Frequency Identification Technology
CMC	: Colombo Municipal Council
LVI	: Land Assessment Index
CBD	: Commercial Business District
CEB	: Ceylon Electricity Board
LECO	: Lanka Electricity Company (Private) Limited
SH	: Single Household
AJH	: Adjoining Households

EPZ	: Export Processing Zone
IE	: Industrial Estates
HBW	: Home based Work
HBNW	: Home based Non Work
NHB	: Non Home based
DOT	: Department of Transport
LKR	: Sri Lankan Rupees
kWh	: Kilowatt hour
TAZ	: Traffic Analysis Zone
UE	: User Equilibrium
SO	: System Optimum Assignment
RDA	: Road Development Authority
SLS	: Screen Line Survey
VCR	: Velocity Capacity Ratio
ERP	: Electronic Road Pricing
RMSE	: Root Mean Square Error
VOT	: Value of Time
KPI	: Key Performance Indicator

CHAPTER 1

INTRODUCTION

1.1 Overview

In urban areas, transportation issues are widely reverberated, affecting not only daily life, but also the economic vitality and environmental sustainability of cities. Several factors contribute to the urban transportation conundrum, including congestion, high costs, and environmental impacts. According to (Small and Verhoef 2007), congestion affects commuters' productivity and quality of life by imposing a substantial time cost. In addition, the financial burden of maintaining and improving infrastructure is enormous, often resulting in disputes over resources (Todd 2019). As a result of vehicular emissions, urban transportation poses serious public health and global concerns due to air pollution and climate change (Chester and Horvath 2009).

There is an inextricable link between urban transport and land use. A poor urban planning strategy can exacerbate transportation problems in cities. For instance, in Beijing, uncontrolled urban expansion has led to a surge in demand for transportation, resulting in severe traffic congestion and high levels of air pollution (Wang et al. 2023). Similarly, in Manila, Philippines, the absence of effective land use planning has led to over-concentration of activities in the city center, leading to a transport system that is plagued with extreme traffic congestion. Similarly, poor transport planning and infrastructure decisions can lead to unintended land use consequences such as urban sprawl. Atlanta, Georgia, in the United States serves as an example where lack of public transport and reliance on personal vehicles has contributed to widespread urban sprawl, leading to environmental and public health issues (Frumkin 2002). In the Australian city of Perth, investment in freeway infrastructure over public transportation options has promoted low-density suburban development, further driving the issue of urban sprawl (Giles-Corti et al. 2013). As a result of the lack of integration between transport modes and traffic congestion, the sustainable development of urban transport systems and the quality of life and environment have been threatened.

Land use and transportation are two interconnected systems that dynamically interact within an urban landscape, and their sustainability is critical for the overall system's sustainability. Extensive research has been conducted to investigate various aspects of the complex interplay between land use and transportation in

This framework provides a foundation for understanding the dynamic feedback loops between transportation infrastructure and land use patterns. It highlights that transportation accessibility can shape land use decisions and, in turn, land use choices can impact transportation demand and infrastructure requirements. By integrating land use and transportation planning, policymakers can better coordinate transportation investments with land development decisions, promoting sustainable urban growth, and reducing reliance on private vehicles. This research has significantly contributed to the understanding of the interdependencies between land use and transportation, emphasizing the importance of their integration for achieving sustainable and liveable urban environments.

With rapidly increasing global urbanisation, evidence-based methodologies and tools are needed to analyse balanced, integrated, and sustainable urban transport development. Critical economic, social, and environmental sustainability attributes, focusing on the dynamic interactions between different urban landscape systems arising from land use and transport systems, become essential to ensure sustainability. Theoretical formulations, including those proposed by Wegener, have been translated into practical tools known as Land Use Transport Integrated (LUTI) models. The integrated land use and transport policy has been identified as a practical approach to address transport problems (Silva and Pinho 2006). These models integrate data on land use, transportation networks, and travel behaviour to simulate and predict the interactions between land use and transportation systems over time. One notable example of a LUTI model is the UrbanSim model (Waddell 2002), which has been widely used in urban planning and policy analysis. UrbanSim integrates land use, transportation, and economic data to simulate and visualise the potential impacts of various policy interventions and land development scenarios. It allows decision-makers to assess the effects of transportation investments, zoning regulations, and land development patterns on travel behaviour, accessibility, and environmental sustainability. The model provides valuable insights into the potential consequences of different planning decisions and helps in evaluating trade-offs between competing urban development objectives. Another example is the TRANSIMS (Transportation Analysis and Simulation System) model (Wegener 2004), which focuses on simulating the behaviour of individual travellers and their interactions with the transportation

system. TRANSIMS incorporates detailed data on demographics, land use, and transportation networks to simulate travel patterns and predict the impacts of transportation policies on land use.

LUTI models have found extensive application in developed countries, but their utilisation in developing countries has been limited. Yet, such models could have profound positive impact in developing countries if used in decision making, given the rapid and complex nature of urban growth in such geographies. A common problem in developing countries is the need for more data and financial stability to survey and collect the relevant information to develop that model. Developing countries often face challenges in collecting sufficient data and resources to develop robust land use and transportation models. This can limit their ability to effectively plan and manage urban transportation systems, leading to traffic congestion and inadequate infrastructure development.

The emergence of information and communication technology (ICT) has revolutionised data collection and analysis, offering extensive and High Frequency (HF) data even in developing countries. These innovative data sources have the potential to bridge the data gaps in developing countries discussed earlier. Even some models already calibrated for developed countries cannot be directly used for developing countries due to their unsuitable functional forms and data inadequacies. Therefore, seeking alternative data sources that are less expensive and frequently available is necessary to substitute the data required to develop a dynamic and sustainable LUTI model. For instance, mobile phone data has become a valuable source of information in urban planning and transportation studies. Call detail records (CDRs) from mobile networks can provide valuable insights into human mobility patterns, travel behaviour, and transportation network usage (Mamei et al. 2019). Such data can be anonymised and aggregated to maintain privacy while still offering valuable inputs for LUTI models. Additionally, data from global positioning systems (GPS) devices, smart cards, satellite images, and social media platforms can also contribute to a more comprehensive understanding of travel behaviour and land use dynamics in developing countries (Shield and Blythe 1997, Ge and Fukuda 2016, Guo et al. 2020).

There are few notable applications of HF data in LUTI models such as real time travel behaviour analysis, mobility pattern analysis and microsimulation modelling. Data from HF sources, such as GPS on smartphones or vehicle tracking devices,

can be used to provide real-time information about travel patterns, such as trip origins and destinations, travel speeds, and route choices. These data can be integrated into LUTI models to capture dynamic changes in transportation demand and simulate traffic conditions in real time. Land use and transport policies and interventions can be predicted and understood more accurately by doing this. Data from HF sources such as smart cards and mobile phones can provide insights into mobility patterns. A LUTI model can capture travel behaviour variations based on demographics, time of day, and geographic location by analysing this data. As a result, transportation planning and policy decisions can be made with more nuanced modelling of traffic demand and land use interactions. The HF data sources provide detailed information on individual travellers and their interactions with transportation systems. This allows LUTI frameworks to develop microsimulation models based on real-world travel behaviour data, which can be used for developing microsimulations. In response to changes in land use and transportation conditions, microsimulation models capture the dynamics of travel choices, including mode selection, route selection, and departure time. Using fine-grained modelling, complex interactions between land use and transportation can be accurately represented.

1.2 Research Problem

Transport demand modelling often focuses on travel behaviour while neglecting important aspects such as active participation, travel needs, and scheduling. This oversight has led to the overlooked impacts of transport on urban development, potentially resulting in uncontrolled growth and environmental degradation. Urban sprawl, characterised by the unplanned expansion of urban areas beyond their original boundaries, is a common consequence. It leads to the development of low-density, car-dependent suburbs and exurbs featuring single-family homes on large lots, shopping centres, office parks, and transportation infrastructure like highways and roads that connect these areas. Urban sprawl has been linked to numerous negative effects on the environment, public health, and social equity. These include increased traffic congestion, air and water pollution, greenhouse gas emissions, and reduced access to public transportation and essential services for low-income and minority communities. Additionally, transportation systems can support urban sprawl, and changes in transportation significantly influence land use patterns, population distribution, and land prices.

While internationally used, most land-use transport integrated models have yet to be calibrated for use in developing countries due to the required data and associated costs of collection. It is crucial to have a model that can support balanced transport and land use development. The LUTI model consist of two components: land use and transport. These models can be developed at aggregated and disaggregated levels, requiring primary and secondary data in the appropriate format for calibration and validation. For instance, household income data is an essential variable for developing trip generation in the transport model, but it is only occasionally available, even at the aggregated level. Recently developed land use-transport interaction (LUTI) models are based on microsimulation land use models, with many utilising an activity-based approach to analyse transport demand. Disaggregated microsimulation-based models for land use and transport demand necessitate high-quality datasets, as conventional survey data may not suffice. While most land use model studies have used HF data such as Landsat and Remote Sensing (RS), limited research has tested transport models with HF data, such as CDR and GPS.

This research is specifically focused on addressing the following research questions:

1. How can transport and land use models be effectively adapted to account for data limitations, particularly in developing countries where comprehensive datasets are often scarce?
2. To what extent can high-frequency data sources, such as CDR and data from GPS devices, enhance the accuracy and reliability of transport models, particularly in predicting travel behaviour and understanding the dynamics of urban sprawl?

1.3 Research Aim

The rapid urbanisation of cities and the complexities associated with it call for sophisticated decision support tools, such as LUTI models, to facilitate balanced and sustainable development of transport and land use. However, the extensive data requirements of most LUTI models have made them expensive, time-consuming, and challenging to implement, particularly in developing countries. Fortunately, the increasing availability of HF data, facilitated by advances in information and communication technology, offers a promising opportunity for data collection at a

disaggregated level and at minimal or no cost.

The literature highlights the need to bridge the gap between research utilising HF data and its practical application in real-world contexts. To address this, the research aims to develop an interactive model specifically focused on home-based work trips, incorporating existing component models that have already been utilised in regional planning efforts in Sri Lanka's Western Province and other regions. This comprehensive approach involves sourcing and aligning suitable inputs and outputs related to mobility, land use, employment, and economic changes within the LUTI model.

A key aspect of this research is the formulation of an integrated model where the outputs from each component model dynamically feed into the inputs of the others. This interconnectedness allows for a holistic assessment of the proposed development options. To evaluate the expected benefits of each option, the researchers have identified Key Performance Indicators (KPIs) that measure the impacts on land use, environment, and mobility.

By adhering to these principles, the research aims to address the challenges of data requirements and model integration, leveraging the availability of HF data to enhance the applicability and effectiveness of LUTI models. Through this approach, decision-makers can make informed choices regarding transport and land use development, fostering more sustainable and balanced urban growth.

1.4 Research Objectives

This research aims to develop a model that can accurately predict the number and characteristics of home-based work trips based on various factors such as distance, transportation mode, and demographics and with some other surrogate variables such as electricity consumption and area of the households. By doing so, the model can be used to evaluate the impact of different land use and transportation policies on commuting behaviour and travel demand. More specifically, the following four main objectives need to be addressed:

1. Identification of suitable HF data inputs for the study area, which can support the development of LUTI models based on previous studies.
2. Developing a methodological framework for the Land Use and Transport Integrated Model based on the identified vital HF data inputs.
3. Assessing the land use and transport impact based on a developed aggregated model by testing with different scenarios.

4. Developing a disaggregated Trip Generation Model based on the available HF data and assessing the accuracy with the aggregated model.

1.5 Overall Methodology

The Western Province of Sri Lanka is facing numerous challenges such as rapid population growth, income disparities, overcrowding, mismatched housing-to-jobs areas, deteriorating environmental conditions, and economic losses from excessive traffic. These transportation-related costs, including travel time, vehicle operating costs, and accident costs, hinder the economic development of the region. To alleviate this problem, better coordination between transport and land use planning is necessary. While many studies and collaborations have focused on land use and transport integration over the past 60 years, they have yet to be applied or developed in Sri Lanka. Land use and transport models in developing countries are limited by the extensive data requirements and associated acquisition costs. However, novel datasets of high-frequency data have been discovered at the aggregated and disaggregated level, addressing the issue of data availability and its impact due to rapid land use change. This study seeks to bridge the gap between research with this type of data and its subsequent application in a real-world context to develop an efficient, near-real-time LUTI (NRT-LUTI) model.

The overall methodology for achieving these objectives is illustrated in Figure 1.2. Objective 1 will be based on exploring the insights and findings from existing research. Based on prior studies, the current effort leverages the cumulative understanding of the field by applying the knowledge and methodologies developed in previous studies. To accomplish Objective 2, methodological frameworks will be developed based on high-frequency (HF) data inputs, which have been identified as crucial. For an accurate representation of urban dynamics, data from the most relevant and influential sources should be leveraged. The third objective will be to examine how land use and transportation interact. This objective describes a research step that explores the reciprocal effects between the two. To understand how changes in one aspect influence changes in another, a model is developed and tested across diverse scenarios. Finally, objective 4 will be based on creating a disaggregated trip generation model by utilizing the existing HF data and subsequently evaluating its accuracy in comparison to the aggregated model.

Despite numerous sources of HF data for land use and transportation studies,

medium resolution Landsat data and electricity consumption data from GPS locations of customers combined with standard classifications practiced by electricity providers in Sri Lanka show promise for developing a model for the Western Province. The main advantage of these datasets is their frequent availability with low and no cost. While Landsat's ability to produce medium-resolution satellite images does not provide a broad view of land use, electricity consumption data can provide a disaggregated level of information, including GPS location of each building, electricity consumption, and category. Using these datasets, an aggregated level of Trip Generation and Trip Attraction can be derived. The above linear models were validated using the Home Visit Survey (HVS), collected in 2013 by the Colombo Metropolitan Region Transport Study (CoMTrans) Study.

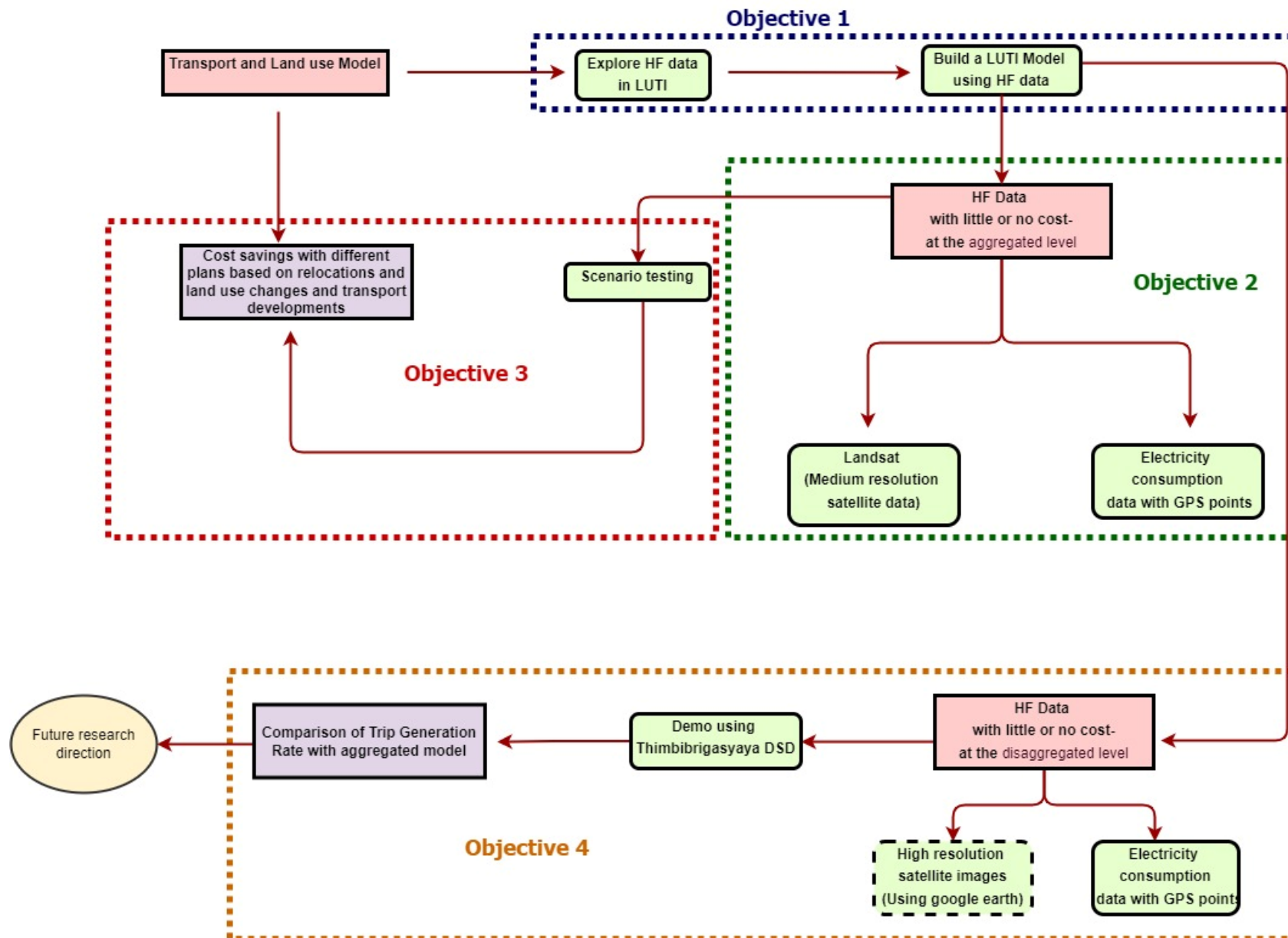


Figure 1.2: Conceptual Approach

A trip generation model was developed based on the number of households and each zone's average monthly electricity consumption. Landsat satellite data were used to develop the trip attraction model for commercial and industrial purposes. The overall model was developed for 2013 as a validation year and 2019 as a forecasting year. The developed model was also tested with different scenarios based on electricity consumption patterns and land use development plans suggested by the government and other related institutions such as the Urban Development Authority (UDA).

As previously mentioned, despite the availability of disaggregated electricity consumption data, the resolution of satellite imagery limits the ability to develop the model at disaggregated levels. On the other hand, Google Earth provides high-resolution satellite images that can be obtained for free. However, due to the challenges of extracting images and the time-consuming process of extracting images for the entire study area, this study has been expanded to focus on the analysis of the smallest area. Thimbirigasyaya DSD was chosen for this study due to its mixed land use pattern. The trip generation and trip attraction models were developed using satellite images and electricity consumption data obtained from Google Earth. The accuracy of the trip generation and attraction models was then compared to the values obtained from the aggregated linear model.

1.6 Structure of the Dissertation

The first objective was addressed through an extensive literature search based on HF data (Chapter 2). The chapter focuses on three main areas: (1) summarizing existing LUTI models and their corresponding data requirements; (2) explaining the sources of HF data in LUTI modelling; and (3) discussing the applications and challenges in implementing such data in LUTI modelling. This review identifies recent technological developments and the availability of HF data to fill the data availability gap for LUTI, as discussed in the literature. Furthermore, this novel data source can increase the potential of the LUTI model, particularly in developing countries where land use and transport patterns are changing rapidly, and traditional forms of data are expensive to collect.

To address the second objective, Chapter 3 analyses the HF data sources available in the study area. This chapter has two parts. Part 1 explains the study area and

analyses the available HF data sources. Moreover, those identified sources are further analysed and validated with existing data for the particular purpose considered in this study. The second part explains the development framework of the LUTI model for the study area using the identified HF data. This part explains the overall methodology employed to develop a transport and land use integrated model using the available HF data.

The third objective is addressed by developing a transport and land use model using available HF data. This is explained in two chapters (Chapters 4 and 5). Chapter 4 explains the development of the trip distribution model. This chapter starts with developing trip attraction and generation models, which play a crucial role in transport planning and forecasting. Multiple Linear Regression (MLR) is the most popular method for calculating trip attractions and generations, which can be used to forecast trips with updated values of independent variables. However, surveys require updating the independent variables, making them expensive and time-consuming. This study aims to develop an MLR model for trip attraction and generation based on available survey data from 2013 and updates those independent variables using HF proxy data for the predicted year (2019). Data from HF proxy sources, such as electricity consumption and GPS customer points, and Landsat satellite imagery data have been used to update the independent variables of trip generation and attraction for 2019. Initially, data from the HVS and land use data from the Western Province of Sri Lanka conducted in 2013 were used to develop the MLR model for trip generation and attraction. A correlation and regression analysis are performed using these surveyed data, and the model is validated with 2013 survey data and would be helpful for real-time estimation of each zone's trip generation and attraction. The gravity model trip distribution for 2013 and 2019 was developed based on the calculated trip generation and attraction.

Chapter 5 addresses developing scenarios to test the developed model. The five scenarios include a validation scenario for 2013, a do-nothing scenario for 2019, a future scenario for 2030 based on electricity consumption rate, a scenario based on improvement plans proposed by the Urban Development Authority of Sri Lanka for 2030, and a scenario based on improvement plans proposed by the University of Moratuwa for the revised CoMTrans 2013 for 2030. This chapter has two parts. The

first part is based on the development and preparation of input data for the model run. Among several transport modelling software, JICA STRADA software has been used to update the output of each scenario. This software has been developed by the Japan International Cooperation Agency (JICA) and has been used for several transport projects and studies in Sri Lanka, including the CoMTrans 2013 study. The second part of this section focuses on the outcomes of each scenario, such as passenger km, passenger hours, average and peak network speed, and average trip length. The comparison of each case is based on the overall system cost considered for development.

The fourth research objective is to develop a disaggregated model using available HF data for the study area. Chapter 6 explains how fuzzy logic is used to develop a disaggregated trip generation model and compares its accuracy with the aggregated model. Calculating trips from each traffic zone is a critical step in the four-step model, and the most popular method for this is MLR, which is also used in the aggregated model. However, the main limitation of this method is its reliance on independent variables related to the zone, and its future predictions are subject to questions of accuracy due to the assumptions made in the method. Updating these independent variables requires additional time and resources, such as conducting surveys like HVS. In this chapter, HF data, which is freely available and frequently updated, is used to estimate trip generation using fuzzy logic to fill in the gap. The fuzzy rule-based model was created using 2013 HVS data and was updated with 2013 data for validation purposes and 2019 data as a prediction year. Thimbirigasyaya DSD in Western Province, Sri Lanka, was chosen as the research area for this purpose. According to the results of this study, a fuzzy rule-based model can be used when no exact data is available, and the available high-frequency data shows a non-linear relationship with the dependent variable.

Chapter 7 concludes the thesis by summarising the findings of each chapter and highlighting the importance of data in developing the LUTI model. Furthermore, this chapter further analyses the application of other HFD sources to improve the model's accuracy, with suggestions for adding new variables to the existing models.

1.7 Significance of the study

The objective of this research is to develop a methodology for the Land Use and Transport Integration (LUTI) model that can be applied to urban areas in Sri Lanka and similar countries. By combining the expertise of various professionals, including urban planners, transport planners, economists, data scientists, and environmentalists, an integrated model will be created as a collaborative platform. This model utilises new data sources, particularly HF data, to enhance the accuracy and effectiveness of predictions.

Traditional LUTI models often rely on survey data, which can be outdated or incomplete, leading to inaccuracies in predicting traffic flow and travel behaviour. However, by incorporating HF data, the research can calibrate and validate the models more effectively, using real-time information on transportation conditions. This data enables policymakers to make well-informed decisions regarding transportation policies and investments, as it provides reliable and up-to-date insights.

Moreover, HF data allows for the identification of emerging transportation trends and patterns, such as changes in travel behaviour and the adoption of new technologies. These trends provide valuable information for policymakers to anticipate and plan for future transportation needs, thus improving the planning process.

The integration of HF data into LUTI models promises to provide accurate and current information on transportation conditions and trends. This, in turn, can lead to improved transportation planning and policymaking. The outcome of this research will expand the existing methodology for the LUTI model, incorporating various socioeconomic attributes to guide the competitive, sustainable, and liveable development of urban regions. The significance of this research extends beyond its potential for international publication value, as it holds practical implementation value for planning institutions and consultants.

Furthermore, the planning tools developed through this research are expected to be utilised in the proposed urban development programs in Sri Lanka, leveraging available HF data related to land use and transport. Additionally, this research provides a promising direction for developing countries to achieve sustainable land use and transport integration by harnessing the potential of high-frequency data.

CHAPTER 2

A REVIEW OF HIGH FREQUENCY DATA USE IN LUTI MODELS

[This chapter has been published in the Journal and details are below

Amalan, T. P., Du, B., Kumarage, A. S., Wickramasuriya, R., & Perez, P. (2023). High frequency data in Land use and transport integrated model: A review of sources and application. Asian Transport Studies, 9, 100092. doi: <https://doi.org/10.1016/j.eastsj.2022.100092>]

2.1 Introduction

Land use and transport systems interact at a fundamental level in deciding the evolution of the urban fabric, which leads to the necessity of developing the LUTI Models. Over 60 years, numerous transdisciplinary studies and collaborations have focused on land use and transport integration. The development trend of LUTI models can be traced from the aggregate to the disaggregate level, as well as from static (or cross-sectional) to dynamic (or longitudinal) models (Miller and Salvini 1998). Due to this model advancement, integrating and calibrating these two models can face a more significant challenge regarding the dataset's quality and availability. Conversely, LUTI models require extensive data to produce valid and precise results based on the literature review. The most essential and primary step before developing or selecting a LUTI model is based on the problem definition and the solution.

Transportation and land use challenges in developing countries differ significantly from those in wealthier, advanced countries. The rapid growth of populations, income disparities, overcrowded environments, mismatched housing-to-jobs areas, deteriorating environmental conditions, and economic losses from excessive traffic are among the more vexing challenges facing developing cities that can be alleviated through better coordination between transportation and urban development (Cervero 2013). Land use and transportation challenges in developing countries are often intertwined and require innovative solutions that address their specific characteristics. There are a variety of solutions available, including improvements to transportation and sustainable land use planning, as well as developing new technologies. Aside from considering social, economic, and cultural factors, climate change and other environmental factors play an important role in influencing transportation and land use in these countries. To address these

challenges, governments, private sector organizations, and civil society groups must collaborate and be willing to experiment with new approaches and think creatively.

Traditionally, the data related to transport and land use relies on manual surveys such as HVS, Classified Vehicle Counts (CVC), Land use Surveys (LUS), etc. However, these methods are costly, labour-intensive, and, more importantly, involve a small sample size. Additionally, these data are prone to measurement, acquisition, and reporting errors, such as sample bias and imprecision, which could cause difficulties measuring the impact of land use and transport policies and lead to inconsistency in land use and transportation planning operational decisions. Also, collecting extensive data using surveys may require much time and money, challenging developing countries, predominantly low-income countries. Due to the extensive data requirement and associated data acquisition costs, the application of LUTI models is mainly limited to a few regions like the United States and Western Europe (including the UK) (Miller 2018) (Figure 2.1).

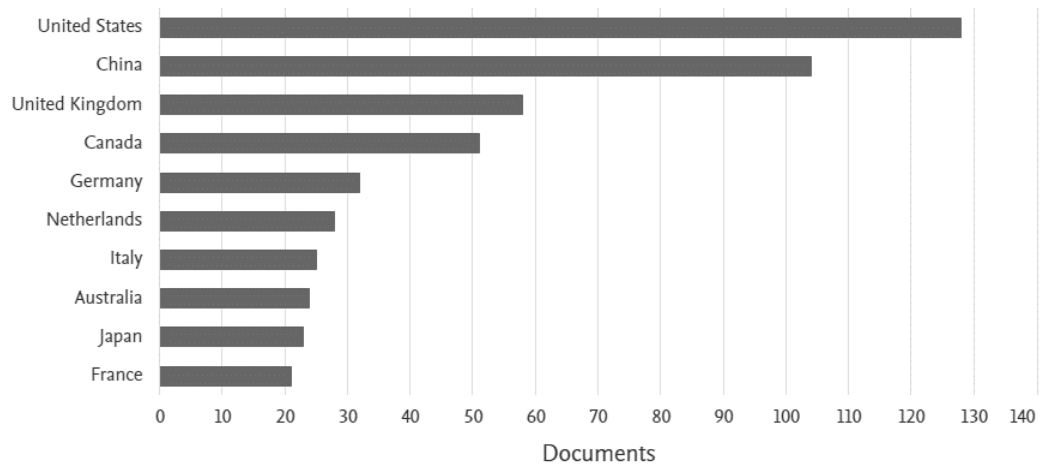


Figure 2.1: Application of LUTI (Source: Scopus Database)

Conversely, the availability of HF data is fast expanding because of the advancement of ICT. Among the different definitions of HF data in various fields, this paper defines HF data in transport and land use as follows: "Data contains diverse sets of information related to land use and transport which can be collected more rapidly than traditional data". These include dynamic data or information related to land use and transport from different sources, such as sensors, satellites,

devices, third-party data, web applications, etc., at higher frequencies. In addition, it included both big data and traditional data, which are constantly changing and evolving. Compared with traditional data sources, these HF data show many unique attributes and advantages concerning land use and transport.

- a. HF data sources contain updated and near or real-time spatial and temporal information that is impossible to collect through traditional travel surveys (Lefebvre, Sannier, and Corpetti 2016).
- b. They contain an enormous set of disaggregated/ individual-level data with higher accuracy at minimal or no cost. Some of these data can be linked with supplementary data such as land use (Figliozi and Stoll 2018).
- c. HF data can complement limited data in developing countries and enable the development and application of LUTI models as much-needed decision-support tools (Joubert, Murawski, and Bick 2021).

Moreover, such models developed for rapidly urbanising areas like cities in developing countries must be regularly updated to ensure relevance. HF data can fulfil this requirement as well. Despite these opportunities, challenges are associated with collecting, processing, and analysing HF data, which need to be addressed while employing these data in land use and transport planning. As evidenced in the extensive literature, the application of HF data and big data that can update itself can frequently be observed in developing transport models, primarily in developing activity-based models and land use classification using satellite imagery data. Despite this, there are no attempts to integrate these two separate models based on such data.

The central aim of this systematic review-based study is to explore the opportunities for extending the LUTI models to use HF data and demonstrate the approach using data availability. The developed LUTI model using HF data serves two purposes:

- a. Complementing the existing traditionally collected data with HF data for setting up, calibration and validation of the model, and
- b. Regularly update the model using HF data to keep pace with rapid urbanisation and stay relevant.

This section is organised in the following order: provides the methodology adopted for the literature review, discusses the theoretical background of LUTI models and

summarises the existing LUTI models published over the last decade. It begins with a discussion of the theories applied in LUTI models, followed by an analysis of input data required to set up NRT-LUTI models. This section explores sources of obtaining the HF data, highlighting their availability and limitations in a broader picture. In contrast, the last part discusses the challenges and benefits of actual implementation in LUTI models. Based on these, the areas needing further research are discussed.

2.2 Literature Search (Overview/Summary)

In order to conduct the review, two online searchable databases were consulted: Web of Science and Scopus. The review of the literature identifies, assesses, and interprets all research results to provide answers to the research focus areas shown below.

- a. Summarising existing LUTI models and their corresponding data requirements.
- b. Assessing the different sources of HF data availability for near-real-time forecasting.
- c. Discussing the challenges and benefits of the actual implementation of such data on LUTI modelling

All available research on the topic published from 2000 to 2021 was searched using search terms in English. The following keywords were used to start the database searches "Land use and Transport Integrated Model or LUTI", "High frequency data and Transport", and "High frequency data and Land use". Due to the limited number of studies found using HF data, the search string expanded to the following keywords: "Big data and Transport" and "Big data and Land use". In addition, the snowball approach was used to enrich additional publications related to the above focus areas. This literature review compiled information from 58 academic and professional publications.

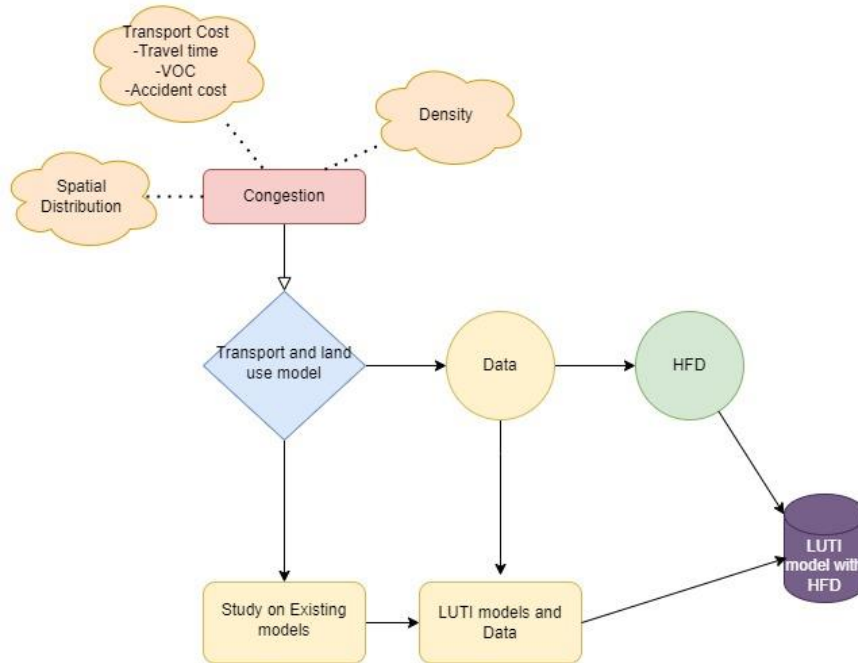


Figure 2.2: Conceptual Framework of Literature

2.3 Theoretical Background of the LUTI Framework

2.3.1 Land-use Modelling Techniques

2.3.1.1 Spatial interaction Based Model

The spatial interaction model explains that the interaction between any two zones is proportional to the number of activities in each zone and inversely proportional to the friction impeding movement between both zones (Niu, Wang, and Chen 2019). It is an aggregated model where space and traveller characteristics are grouped into more significant categories. In the 1970s, Wilson introduced the maximum entropy method, which was used as a framework for the spatial interaction model. It can be utilised for many interaction models, such as residential location decision models, shopping decision models, or the doubly constrained gravity model.

2.3.1.2 Utility-Based Model

Utility-Based Models are based on the utility model of personal-level socio-demographics and activity travel environment (Eboli, Forciniti, and Mazzulla 2012). Most utility-based models are based on discrete choice and random utility

theory. It can be estimated using aggregated data or disaggregated (individual) data. Most of the LUTI models used aggregated data to express the characteristics of the different zones within the study area.

Utility models are based on two assumptions,

- Individuals belong to a homogeneous class from a behavioural point of view
- Each individual chooses to maximise their utility from different alternatives.
-

2.3.1.3 *Microsimulation Model*

Micro-Simulation Model is an advanced model of a utility-based model that makes it possible to discover model components, such as individuals, households, jobs, and dwellings, over time to observe the modelled process of change at a level of detail (Gulhan and Ceylan 2016).

2.3.2 Transport Modelling Techniques

2.3.2.1 Traditional Four-Step Model

The four-step (FSM) approach focuses on "trips" without clearly identifying the reason for the trips and travel. It represents travel as a mere collection of "trips". Each trip is considered a separation from other trips without considering the inter-relationship of the travel. FSM formulated the process as a sequential four-step model. The first step is trip generation which measures trip frequency between zones in a study area. Next is trip distribution and production, accommodating trip tables of trip demands of a person. The third step is mode choice, where trip tables are proportionated according to the alternative modes (Car vs Bus etc.). Finally, the route choice model distributes the modelled trips into the specified network.

2.3.2.1.1 Activity-Based Model

The traditional trip-based model (FSM) of travel demand forecasting has always needed more validation of individuals' travel behaviour. During the 1970s, the Activity-Based Model (ABM) was first structured in-depth to address the fundamental changes in urban, environmental, and energy policies. Transport demand derives from people's needs and desires according to their participation in activities. Based on activity-based approaches, routes can be determined by predicting the location, time, duration, participants, transport mode, and mode of

conduct of activities. Figure 2.2 shows the clear difference between the schedule, tours, and trips.

The following factors are considered when making activity and travel scheduling decisions:

- Continually influenced by longer-term outcomes
- Engages in transportation interactions
- Intra-household interactions influence the outcome
- A dynamic process that is influenced by past events as well as anticipated events in the future

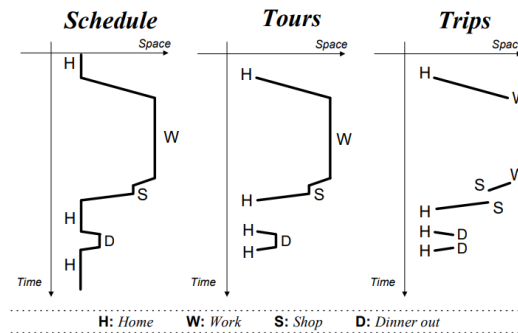


Figure 2.2: Schedule Vs Trips Vs Tours

Source: Ben Akiva, 2008

It mainly focuses on “activity participation behaviour”, disqualifies individual mode-trip chains, and uses “tours” as the essential elements to represent the model travel patterns. Individuals have 24 hours in a day, and they decide how to spend that time on activities and travel based on their sociodemographic, spatial, temporal, transportation system, and other contextual constraints.

Table 2.1: Comparison of Activity-Based vs Trip-Based Model

Parameters	Trip Based	Activity-based
Spatial/ Temporal Details	Low- Moderate	Moderate-High
Person/HHS details	Moderate	High
Police Sensitivity	Moderate	Moderate-High
Run Time	Moderate	Moderate

Cost	Moderate	Moderate
------	----------	----------

2.4 LUTI Model Evolution

In the broad spectrum of LUTI studies, it is crucial to understand the current state-of-the-art models and theories being applied. According to available data and information from previous studies, thirty-two (32) LUTI models were selected and categorised according to the modelling approach adopted in land use and transport (Anderstig and Mattsson 1998, Bain, Hensher, and Li 2011, Brandi et al. 2014, Johnston, Shabazian, and Gao 2003, Lierop et al. 2017, Miyamoto et al. 2007,

Moeckel et al. 2018, Salvini and Miller 2005, Silva, Reis, and Pinho 2014,

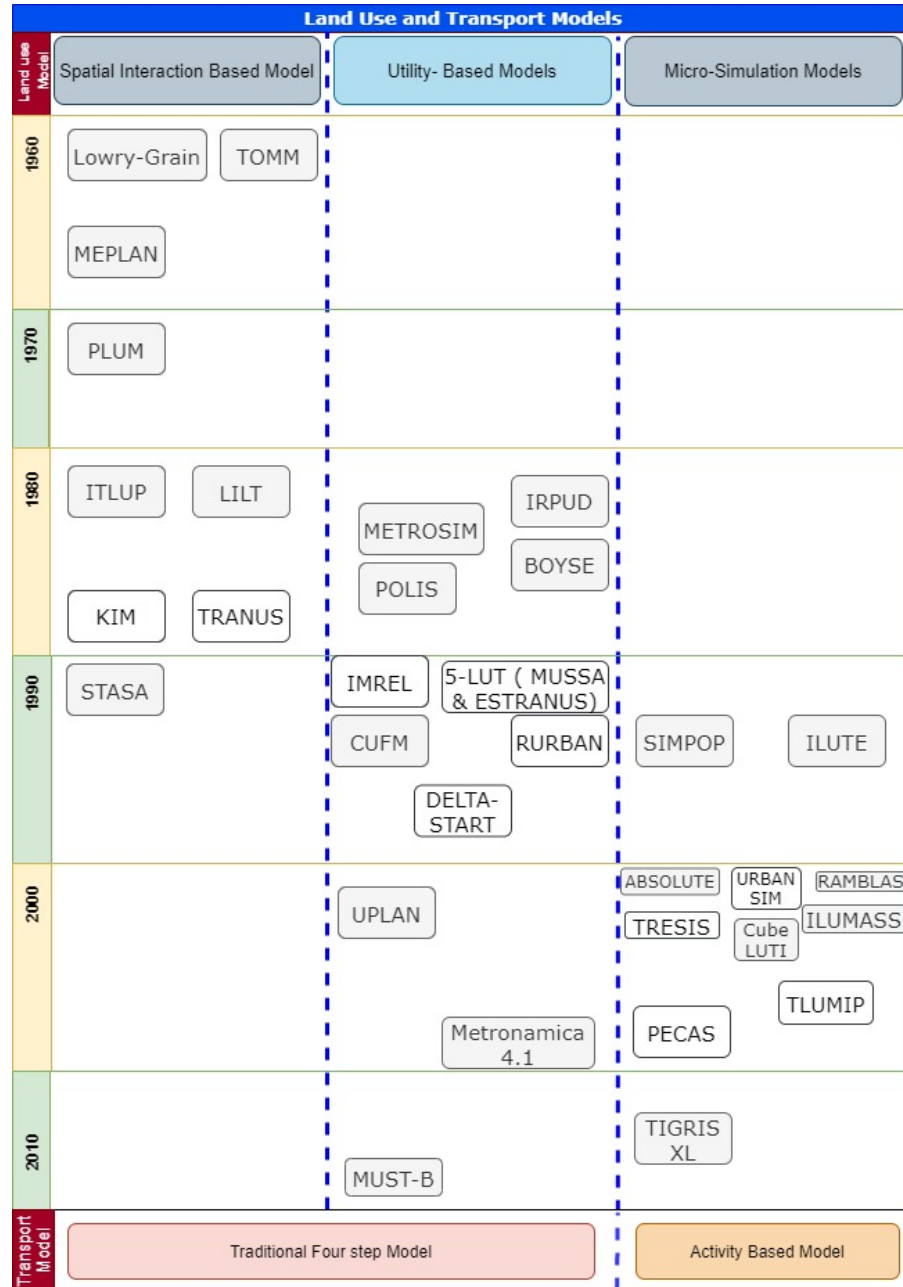


Figure 2.3: Evolution of LUTI Models

Veldhuisen, Timmermans, and Kapoen 2000, Waddell 2002, Waddell et al. 2006, Wagner and Wegener 2007, Zhong, Hunt, and Abraham 2007) (Figure 2.3).

As seen in Figure 2.3, the land use subsystem has been modelled in these LUTI models using one of three distinct approaches: spatial interaction, utility theory or micro-simulation. In comparison, the transport subsystem has used two approaches: the traditional four-step and activity-based models. Traditional LUTI models are

focused on zonal-based outcomes, whereas micro-simulation models describe land use and transport at the disaggregated level of individuals; and discover components such as individuals, households, jobs, dwellings etc., over a period to observe the modelled process of change at a level of detail (Acheampong and Silva 2015). LUTI models use the micro-simulation approach to model land use and the activity-based approach (Acheampong and Silva 2015) (Zhu et al. 2018).

Furthermore, each model was contrasted according to the policies adaptation such as environmental, economic, land use regulations, and transport improvement policies. Figure 4 compares LUTI models according to some subsystems and policy considerations. According to the comparison of subsystems and policies of recently developed LUTI models, almost all the LUTI models considered transport networks, land use, employment, housing, and population as their key components. Consideration of the environmental impacts in existing LUTI models is still minimal. In the transport model, consideration of goods transport impact on LUTI models is also minimal. According to Table 2.2, almost all the LUTI models considered transport network, land use, employment, housing, and population as their key components. Consideration of the environmental impacts in existing LUTI models is still minimal. In the transport model, consideration of goods transport impact on LUTI models is also very limited.

Table 2.2: Comparison of Components in Different LUTI Models

LU Modelling Techniques	LUTI Models	Sub Systems								Policies Modelled				Country Applied
		Network	Land Use	Workplaces	Housing	Employment	Population	Goods Transport	Access ibility	Enviro nment	Economic	LU Regulations	Transport Improve ments	
Spatial Interaction Models	Lowry-Grain	Y	Y		Y	Y	Y		Y					USA
	TOMM	Y	Y		Y	Y	Y		Y					USA
	MEPLAN	Y	Y	Y	Y	Y	Y	Y	Y		Y (Market)	Y	Y	UK
	PLUM	Y	Y		Y	Y	Y		Y					USA
	ITLUP	Y	Y			Y	Y					Y	Y	USA
	KIM	Y	Y			Y	Y	Y			Y (Market)		Y	USA
	LILT	Y	Y	Y	Y	Y	Y					Y	Y	UK
	TRANUS	Y	Y	Y	Y	Y	Y	Y	Y		Y (Market)	Y	Y	Netherlands
	STASA	Y	Y	Y	Y	Y	Y	Y	Y					Germany
Utility Based Model	METROSIM	Y	Y	Y	Y	Y	Y				Y (Market)		Y	USA
	IRPUD	Y	Y	Y	Y	Y	Y				Y (Housing)	Y	Y	Germany
	POLIS	Y	Y			Y	Y					Y	Y	California
	BOYSE	Y	Y			Y	Y						Y	USA
	IMREL	Y	Y	Y	Y	Y	Y		Y		Y (Market)			Sweden
	5-LUT (MUSSA and ESTRUSUS)	Y	Y		Y	Y	Y				Y (Market)		Y	Chile
	RURBAN	Y	Y			Y	Y				Y (Market)	Y	Y	Japan
	CUFM	Y	Y	Y	Y	Y	Y			Y		Y	Y	USA
	DELTS-START	Y	Y	Y	Y	Y	Y				Y (Market)			UK
	UPLAN	Y	Y	Y	Y	Y	Y				Y (Housing)	Y	Y	USA
Multi-Modal	MUST-B	Y	Y	Y	Y	Y	Y			Y		Y	Y	France
	SIMPOP	Y	Y		Y	Y	Y			Y				Europe

	ILUTE	Y	Y	Y	Y	Y	Y			Y		Y	Y	Canada
	URBAN SIM	Y	Y	Y	Y	Y	Y		Y		Y (Housing)			USA
	RAMBLASS	Y	Y	Y	Y							Y	Y	Netherlands
	ILUMASS	Y	Y	Y	Y	Y	Y	Y	Y	Y				Germany
	TRESIS	Y	Y	Y	Y	Y	Y		Y	Y	Y (Market)			Australia
	PECAS	Y	Y	Y	Y	Y	Y	Y	Y		Y (Market)			North America
	TLUMIP	Y	Y	Y	Y	Y	Y	Y	Y		Y (Market)			USA
	TIGRIS XL	Y	Y	Y	Y	Y	Y		Y		Y	Y	Y	Netherlands

2.5 Conventional LUTI Models and Relevant Data Sources

Earlier LUTI models that used the spatial interaction model assume that the interaction between any zones (geographic areas) is proportional to the number of activities in each zone and inversely proportional to the friction impeding movement between both zones. It is an aggregated model where space and individual characteristics are grouped into common categories. In the 1970s, Wilson introduced the maximum entropy method, which was used as a framework for the spatial interaction model (Griffith and Chun 2018, Wilson 2010). It can be utilised for many interaction models, such as residential location decision models, shopping decision models, or the doubly constrained gravity model.

In the mid-1980s, most LUTI models employed utility-based models that explained the utility of personal-level socio-demographics and activity travel environment (Eboli, Forciniti, and Mazzulla 2012). Most of the utility-based models adopt discrete choice and random utility theory. The utility-based models can be estimated using aggregated or disaggregated (individual) data. Most of the LUTI models used aggregated data to express the characteristics of different zones within a study area (Acheampong and Silva 2015).

2.5.1 Analysing of the Inputs of Comprehensive Transport Models and Land-Use Models

According to LUTI literature, input data are generally classified into four categories: sociodemographic, economic, real estate, and transport. Data from these four categories are critical inputs to LUTI models, as they provide information about how land use and transportation interact and affect the built environment and travel behaviour. As a result of analysing these inputs, LUTI models can help policymakers and planners make more informed decisions about transportation and land use planning and develop more effective and efficient interventions. Decision-makers can use LUTI models to understand the dynamics of urban systems, identify

the most effective strategies for promoting sustainable mobility and land use, and assess the relative merits of different policy options by analysing these inputs.

2.5.1.1 Socio-demographic Data

Socio-demographic data includes information such as age, income, and employment-based distribution of the population. Some micro-level LUTI models require a disaggregated level of such data. The data also can be obtained from other sources such as Home Visit Survey (HVS) with other information such as travel diary of everyone. Some level of information can be collected from the census data as well.

2.5.1.2 Economic Data

Economic activity data can be collected from the census. However, the model needs more information related to the industry classification, employment location data, different facility locations of a company such as a warehouse, storage, and production plant, etc., which can be collected from different sources such as company directories.

2.5.1.3 Real-estate Data

Information related to property classifications and the market price is necessary to model the impacts of transport. This data type is challenging to collect and requires specific surveys or secondary sources. In some instances, online property buys and sells portals publish land price reports annually or quarterly based on the prices user's demand. A perfect example of such a portal is Lanka property in Sri Lanka, Makaan in India etc. These portals can be used for research purposes, and the information is gathered frequently.

2.5.1.4 Mobility and Activity Data

Transportation modelling requires both supply and demand information related to the study area. Supply-side data are easy to collect and most available in the public domain, for example, census data such as population, income, road network, public transport network, timetable, and vehicle ownership. The demand-side data requirement, such as trips by personal and commercial vehicles, are traditionally collected via travel surveys which is an expensive and resource-intensive process.

2.5.2 Data Used in Existing LUTI Models

Transportation modelling requires both supply and demand information related to the study area. Supply-side data such as census data such as population, income

etc., road network, public transport network and timetable and vehicle ownership are easy to collect and most available in the public domain. The demand-side data requirement, such as trips by personal and commercial vehicles, are traditionally collected via travel surveys which is an expensive and resource-intensive process. In general, the literature on LUTI classifies input data under four significant components: the socio-demographic model, economic model, real estate model, and transport model and adopts the theories above and models based on the data availability. Based on the widespread literature review, the twelve most recently developed LUTI models were selected from the thirty-two models and analysed according to the input data sources used for the abovementioned components (Table 2.3). Most of those twelve models were developed or modified after 2000 except ILUTE. The ILUTE was selected based on several research studies' popularity and frequent usage.

Table 2.3 depicts that the socio-demographic sub-model requires socio-economic data; however, according to each model, the necessary level of detail in data may vary. For example, URBANSIM needs detailed demographic data, such as factors that affect the relocation decisions of households. The economic sub-model requires employment data related to the workplace and job classification. PECAS, for example, needs detailed economic data based on their activity and land allocations to micro-level simulation results for land use and transport. Some models require real-estate data to assess the land value impact on accessibility. For instance, a model like Cube Land is based on bid rent (MUSSE model) theory and requires detailed land value data. Nearly all the LUTI models adopt the four-step modelling approach for the transportation component. Most transport-related data are collected from transport and traffic surveys and secondary sources to update network and infrastructure data.

The results of Table 2.3 indicate that HFDs are rarely used, and most rely on survey data and publicly available information. It is also evident that LUTI models from the perspective of many countries, especially for developing countries where long-term, well-curated, and publicly accessible data needed to run these models are not readily available.

Table 2.3: Summary of Data Sources of Selected LUTI Models

LU Modelling Techniques	LUTI Models	Socio-demographic	Economic	Real estate	Transport Model
Utility-Based Model	UPLAN (Johnston, Shabazian, and Gao 2003)	Census data (Population, location of a residential development)	Industrial Location Classification (LU classification Data)	Vacant Lands (LU classification Data)	Infrastructure data (SACMET modal), income, trip properties, and mode choice modal
	MUST-B (Bouanan 2017)	INSEE, 2012 data (Population, Households (HHS) location data, taxes, and duties and HHS energy bill)	INSEE data, 2012 (Workplace, location data, land price, and taxes & duties)	Land plot available for future development (Price of land property, business and/or residential), land price	Infrastructure data and flow matrix
	Metronamica (Navarro Cerrillo et al. 2020)	Population and employment data	Economic sectors	Not considered	Road Network, zones trip distribution (car and PT), city centres
Microsimulation Model	ILUTE (Salvini and Miller 2005)	Census data, residential search, behaviour survey, HHS activity scheduling behaviour, energy consumption, and HHS auto ownership	Location of workplace, firm evolution, and energy consumption	GTS housing market (Sales and housing data)	Travel surveys and Goods transport surveys, residential mobility
	URBAN SIM (Waddell 2002)	Census (Person, workers, children, age of HHS head, income), Total HHS units	Business Establishment file (Sector classification)	Vacant HHS unit and land value	HHS and Jobs, mobility pattern and travel characteristics
	RAMBLASS (Veldhuisen, Timmermans, and Kapoen 2000)	Population and HHS classification	Employment data/zone, employment classification	Not considered	Road system (MOT), activity data, time-budget survey (Description of activities/hour for a week), continuing mobility survey (validation)
	ILUMASS (Wagner and Wegener 2007)	Housing, HHS survey, persons, and car ownership	Floor space, firms, and jobs	Land value (Commercial and residential)	Roads and public transport
	TRESIS (Acheampong and Silva 2015, Bain, Hensher, and Li 2011)	Census data	Information not available	Information not available	HVS, Road network (GMA, 2008), Google map (OD pair of time and distance), Railway (Country link-fare structure, Online booking system), Aircraft movements (Airline flight schedule, MOT, Bureau of Infrastructure, regional transport economics), Domestic tourism data (National visitor survey and tourism research Australia)
	PECAS (Hunt and Abraham 2003)	Census data (Income)	Activity allocation model (Choice of location for activity, production and consumption model, choice of where to exchange) Space allocation model (Land allocation)	Price of floor space from different entities	Activity-based transport model-related data collected using survey
	TLUMIP (Donnelly 2018)	Census and HVS	Employment, Imports, and exports	Land price data from different sources	Truck surveys (Tonnage and truckloads, Freight Analysis framework data for truck OD), HVS, Long distance travel data, Carload waybill data (Rail and intermodal traffic patterns)
	TIGRIS XL (Zondag et al. 2015)	Annual housing market survey, HHS location, and population data	Jobs by sectors (Based on historical data)	Land market (Housing and real estate)	Transport network and infrastructure
	Cube Land (Voorraa 2004)	Average income and demographic data	Economic data through surveys and existing information	Real estate supply and rents data from different sources	Trip generation/Attraction, transport time and distance, and transport cost

2.6 High Frequency Data

This section explains the availability of HF data and its application to the LUTI model. These include dynamic data or information related to land use and transport from different sources, such as sensors such as GPS data, satellites, devices, third-party data, web applications, etc., at high frequencies (Fig 2.4).

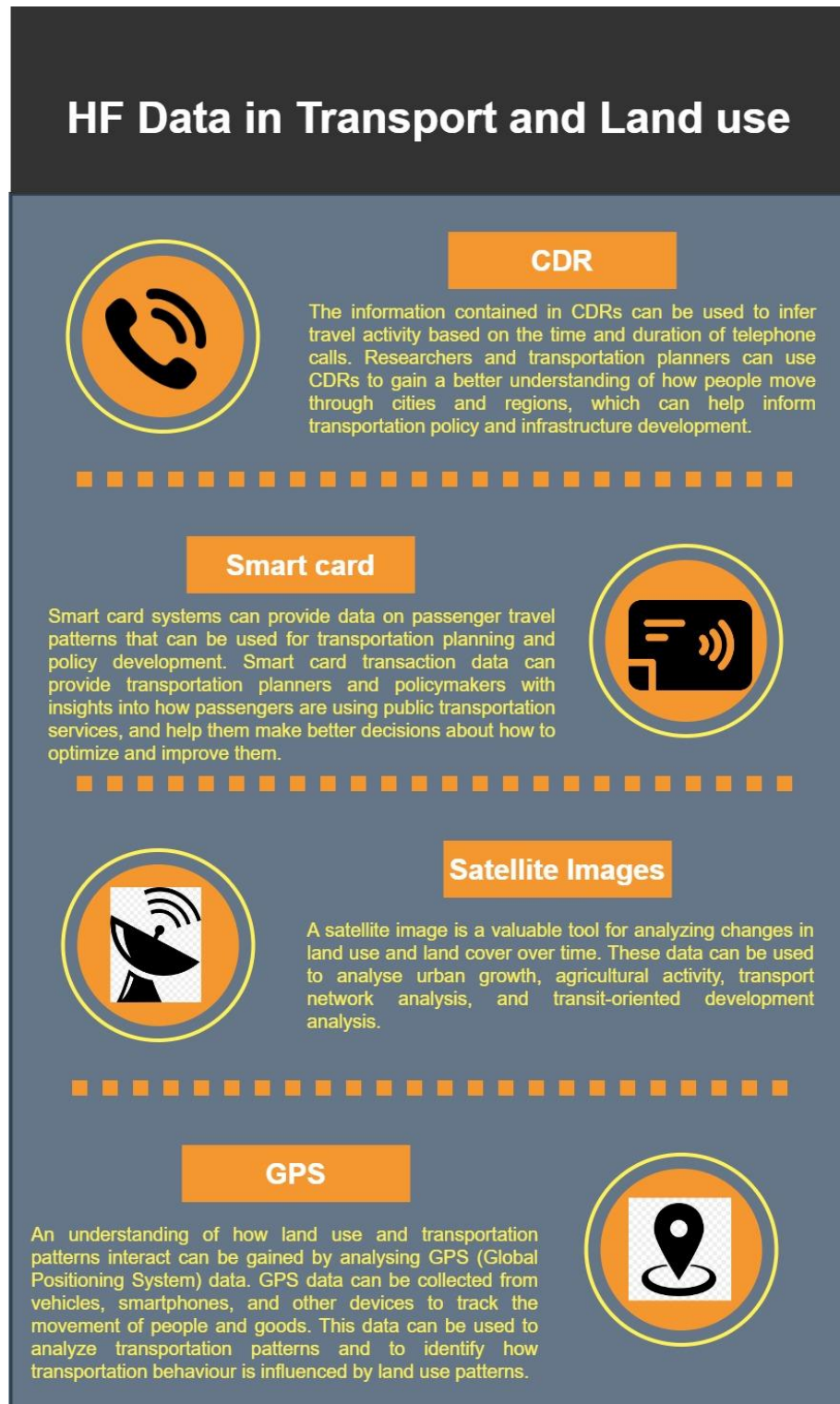


Figure 2.4: HF Data in Transport and Land use

The sources of HF data for land use and transport were identified based on the review of input data for each component and explained together with the availability and limitation.

2.6.1 Mobile CDR Data

As explained earlier, the transport subsystem has used two different approaches: the traditional four-step model and the Activity-Based Model (ABM), whereas several review articles explained the limitation of using the four-step traditional model and emphasise the importance of using ABM for detailed analysis of transport policies and demand (Rasouli and Timmermans 2014) (Hwang et al. 2017). For ABM to be effective, it must rely on accurate data because the model considers activity engagement, trips, network environments, individual characteristics, and socio-spatial features (Rasouli and Timmermans 2014). Currently, the household visit survey (HVS) is the only available large-scale data set that can be used to estimate activity-based simulations of individual travel behaviour. However, this survey is not always accurate.

Call Detail Records (CDR) maintained by mobile network operators can identify human mobility patterns. CDR can supplement traditional data sampling methods like HVS, resulting in cost and time savings and remarkable growth in the volume of data available for analysis. These data can be used to detect the population's detailed and updated mobility patterns. Several studies used different methodologies such as k-mean algorithms, non-linear programming, and ordered response logit (ORL) to derive origin and destination matrix from CDR data (Bwambale, Choudhury, and Hess 2019, Ge and Fukuda 2016, Pourmoradnasseri et al. 2019, Waddell 2002). However, there are significant areas for improvement in the data, such as the absence of interpersonal attributes that can affect travel, limitations in sample size to estimate travel demand, and mode inference and route estimation not being considered. Because of the above issues, CDR data can only be considered a valuable HF data source to estimate OD distribution, provided that all of its limitations and issues are properly addressed and resolved (Iqbal et al. 2014, Mamei et al. 2019). Adapting CDR data offers an excellent opportunity to compensate for the context-specific constraints associated with developing countries (Tsumura et al. 2022).

2.6.2 Satellite Imagery Data

Rapid urbanisation is driving unprecedented change in land use, particularly in developing countries. Urban planners and policymakers need accurate information on current and historical snapshots of land use to assess the degree of change in land patterns and its impact on cities' and regions' socio-economic and ecological fabric. Land use classification is based on land use and land cover. The fundamental difference between land use and land cover is that land use typically shows the landscape use by people, and land cover indicates the physical coverage of land, such as forests, rivers, etc.

Land use data can be differentiated by source and acquired by primary or secondary sources. For the accurate calibration of land use models, land use surveys need specific requirements on sampling, such as a high concentration of sample collection (nearly 100 % sampling) to get detailed information (Dong et al. 2015). Secondary data such as satellite images provide wide-ranging information on land cover data, which can collect at a meagre cost compared to field data collection. Various satellite image sources are available, and each source varies based on the spatial resolution and the acquisition frequency (revisiting time).

Among various satellite data sources, Landsat and Sentinel-2 satellite data is well known for the free dimension with extensive data (Alam, Bhat, and Maheen 2020). Landsat satellites have operated since 1972, providing information about the Earth's land surface (Lee, Cardille, and Coe 2018). Since 2008, the imagery has been available at no cost through the United States Geological Survey (USGS). Based on the survey conducted by USGS in 2013, most land use studies have used Landsat data as their primary source. Landsat data for land use and land cover supports various other applications such as environmental sciences, planning, and development. Moreover, several studies explored the potentiality of this data for forecasting, such as population density distribution. Moreover, Sentinel-2 imagery stands out as the only free and space-borne imagery available at the finer resolution of 10m (Malinowski et al. 2020).

However, the relatively low spatial resolution of satellite data such as Landsat and Sentinel 2 is limited to identifying broad land cover categories such as built-up areas, vegetation, open space, etc. (Hu et al. 2013). Most high-resolution

satellite instruments (resolution less than 5m) aboard such platforms as SPOT, Quickbird and Worldview can be used for land use classification, precisely to extract the building blocks footprint of an area, an essential input to nearly all LUTI models. However, acquiring simultaneous high spatial and temporal resolution remote sensing images is almost impossible due to technical and financial constraints (Borgogno-Mondino et al. 2018, Yan, Schultz, and Zipf 2019). On the other hand, to provide free and open data, Google Earth's high spatial resolution images have supported traditional land use and cover mapping. However, only some studies considered Google Earth for the detailed land use classification (Hu et al. 2013). Some studies have used Google Earth imagery as a reference dataset for the random forest-supervised classification accuracy calculation (Nitze, Barrett, and Cawkwell 2015). Despite the fact that Google Earth provides free access to high-quality satellite imagery, there has been relatively little research that has specifically examined this aspect. This may be due to restricting the acquisition and calibration of those images from the Google Earth Platform.

2.6.2.1 Land use classification using other HFD sources

Several studies are conducted to classify land use using sources such as open street maps using volunteered geographic information (VGI), Flickr photos and satellite images, and foursquare place data. The summary of such fundamental studies is provided below in Table 2.4. Even though these datasets can be collected freely and updated frequently, the information is only available for some developing countries. For example, data such as Flickr provides more reliable information to the tourism industry, not for the land use classification, as all the land use places, such as households and some non-famous places, need to be captured and uploaded. On the other hand, several challenges exist in voluntarily contributed data such as open street maps, Flickr etc. Some fundamental challenges include incorrect geo-referencing information, overlapping information, and spatial and temporal are not homogeneous (Yan, Schultz, and Zipf 2019).

Table 2.4: Summary of Data Sources Used for Land Use Classifications

Year	Paper Title	Authors	Parameters Considered	Model Used for Simulation	Findings
2013	Exploiting volunteered geographic information to ease land use mapping of an urban landscape (Jokar Arsanjani, Helbich, and Bakillah 2013)	Arsanjani et al	volunteered geographic information (VGI) contributions to the OpenStreetMap (OSM) project as an alternative data source instead of gathering training sites through on-site visits and to evaluate how accurate land use patterns can be mapped	High resolution imagery of RapidEye with 5-meter spatial resolution is selected to derive land use patterns of Koblenz, Germany through a maximum likelihood classification technique	The achieved land use map is compared with the Global Monitoring for Environment and Security Urban Atlas (GMESUA) and a Kappa Index of 89% is achieved.
2014	Comparative study of Land Use/Cover classification using Flickr photos, satellite imagery and Corine Land Cover database. (Estima, Fonte, and Painho 2014)	Estima et al	classification obtained for selected photo locations , against the classification obtained from high resolution satellite (30cm resolution) imagery for the same locations.	Based on the image content interpretation, an expert classification of Flickr photos was done. Flickr photo locations were overlaid with the high-resolution satellite imagery, and a land cover class was assigned to each location based on the imagery interpretation.	The flickr source cannot be used alone for the purpose of Land Use/Cover classification but we also believe that it might contain helpful information if combined with other sources.
2016	Using Foursquare place data for estimating building block use (Spyratos et al. 2016b)	Spyratos et al	Foursquare data and official buildings dataset and the non-residential use dataset of Amsterdam . All building classification except residential	evaluation metric is the kappa coefficient, which determines if the classification results are significantly better than a random guess result.	highest kappa coefficient values for the land use categories “ hotels, restaurants and cafes ” (0.76) and “ retail ” (0.65). The lowest kappa coefficients were found for the land use categories “industries” and “storage and unclear” . Foursquare place data can be trusted only for the estimation of land use categories.
2017	Enhanced Land Cover and Land Use Information Generation from Satellite Imagery and Foursquare Data (Jayanetti, Wickramapala, et al. 2017)	Jayanetti et al	Foursquare data for inferring land use information from land cover classification of satellite imagery (Google satellite imagery (96 DPI)) of Colombo, Sri Lanka .	Land cover is derived from satellite imagery by the application of supervised and unsupervised classification techniques . It is then integrated with Foursquare data, which comprises of land use information.	use of Foursquare dataset as the only additional data source for inferring complete land use information is not a viable approach

2.6.3 Smart Card Data

Researchers and professionals have found data such as smart card systems of great value, as they can be collected frequently. Geospatial, time-stamped fare data from smart cards is typically more reliable than other data (Soltani et al. 2015). There are many well-known examples of the implementation of these systems, such as the Octopus card in Hong Kong, Navigo in Paris, Compass card in Vancouver, Oyster in London, Bip! in Santiago or OV-Chip in the Netherlands. Smart card comprises data of the public transport users at the disaggregated level in a specific area. These kinds of data can be analysed at different territorial and temporal scales, and all the travels can be reported in a particular time-stamped with geo-referenced. Passengers must tap their cards during entry and exit to use these cards based on radio-frequency identification technology (RFID) (Chen and Fan 2020, Lei et al. 2020). Using this to on and off information, smart card data has also been used to study complex issues such as trip chaining and origin-destination travel by public transport users, bus commuters' jobs-housing balance, and bus bunching identification and prediction (Du and Dublanche 2018, Gong et al. 2020, Li et al. 2020, Nassir et al. 2011).

While gathering and analysing such data has many advantages, it also has some disadvantages, including that analysing these kinds of big data may require additional effort due to technological advancement, which can pose a major challenge for most developing countries even though they have access to these kinds of data. Over the last decade, many studies have used big data technologies to manage smart card systems. A variety of database management systems are available that can handle these large volumes of information (e.g., Hadoop Distributed File System (HDFS), Hbase, Apache Cassandra, Redis, or distributed object-relational databases such as Postgres, to name a few) (Zaragozí et al. 2021).

2.6.4 GPS Data

GPS has received considerable attention recently as standalone technologies and mobile devices embedded with GPSs. Vehicles with GPS technology can collect time, location, and service status data (Chen, Wang, and Liu 2020, Guo et al. 2020). The GPS tracking technology permits tracking the movements (routes) of travellers. Still, it cannot provide location details or the purpose of travel, which are essential inputs to transport models (Bellemans et al. 2008). GPS can be used for many applications, such as car routine mapping, taxi service measurements, goods

tracking and commercial vehicle management.

2.7 Use of HFD in Developing Countries

2.7.1 Application and Readiness of the Data

Through technological and scientific advances in developed nations, LUTI planning has evolved to represent the integration of multiple subsystems within city planning. A greater understanding of the complexities of human behaviour, urban-space issues, and planning methods has accompanied this. The situation is almost the opposite in developing countries (Lopes et al. 2020). The issue poses a challenge since planning acts must be extracted from documents that are not always readily available and lack standardization. In addition to that, though there are various sources of HF data, methods to integrate into different perspectives do not exist and are very limited in developing countries specially to collect and analyse HF big data.

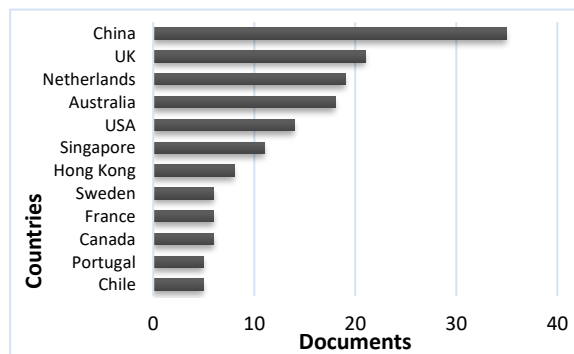


Figure 2.5: Smart Card application in Transport

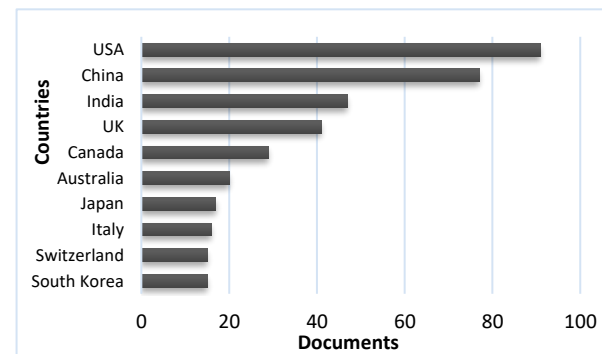


Figure 2.6: GPS application in Transport

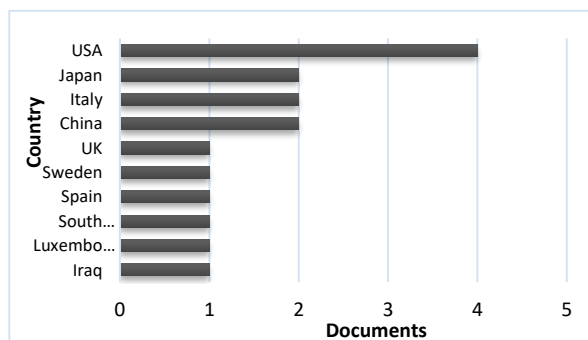


Figure 2.7: CDR data in Transport

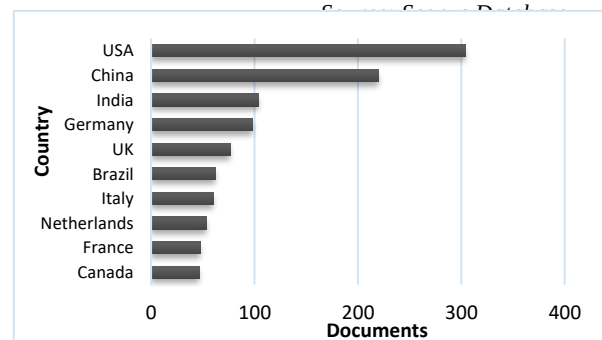


Figure 2.8: RS in Satellite

Figures 2.5, 2.6, 2.7, and 2.8 show several studies based on applying some identified critical HF big data in transport and land use. Results show that such applications are prominently used in high-income, upper-middle-income countries

like the USA, China, England, Australia, and some European regions. Minimal studies are conducted in developing countries, especially in low-income countries. Another critical factor is significant data readiness, which means preparing and exploiting the data (Joubert, Murawski, and Bick 2021). Developing countries face this issue because of limited human and technological resources, infrastructure shortages, financial resources, and institutional barriers due to privacy concerns. Numerous studies have shown that researchers worldwide lack the infrastructure, software, skills, networks, training, and support to access, interpret, and share big data, particularly in developing countries.

2.7.2 Application of HF Proxy Data

Many developing countries, however, lack this type of data or can't collect it because it is prohibitively expensive. To substitute HF data for these cases, researchers and planners may turn to high frequency proxy data. A proxy data source might come from social media, remote sensing, or another publicly available resource.

Figure 2.9 illustrates the differences between HF data and HF proxy data in the context of transport and land use:

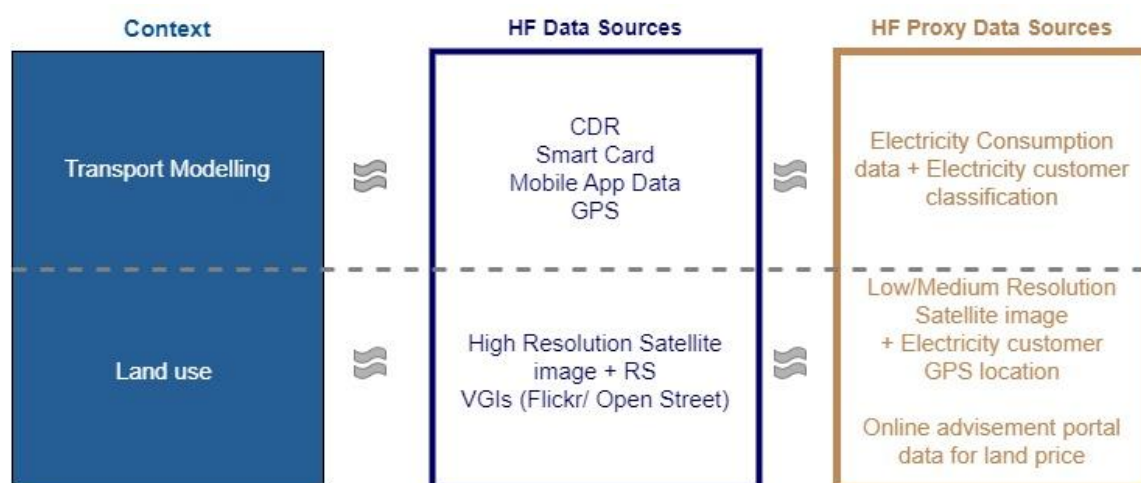


Figure 2.9: HF Data Vs HF Proxy Data

As shown in the Figure 2.9, HF data is highly detailed and accurate, but difficult to collect or costly. Comparatively, HF proxy data may be easier to access and afford but may not be as accurate or detailed as the HF data. Monthly electricity consumption, for example, can be used as a surrogate for income, which is an important factor in determining travel patterns. In many developing countries, electricity taxes may vary based on different purposes, and this data can be used to

further analyse land use patterns and changes. Researchers and planners in developing countries who do not have access to more advanced data sources can still use high frequency proxy data, even though it may not provide the same level of detail and accuracy as HF data.

2.8 Adopting HF data into the LUTI Model

Moreover, there remain challenges to accommodating HF data into the LUTI model in practice. Based on the previous studies and spatial characteristics of HF data, a few critical challenges have been identified and discussed in this section. First, selecting a suitable LUTI model that can accommodate HF data needs to meet essential requirements is shown in Table 2.5. Spatial resolution refers to the size of the image's pixel, whereas temporal resolution refers to how frequently data can be obtained for the same pixel/area. Another essential feature of the NRT-LUTI model is the frequency of updates of the outputs. The main objective of the model is to provide frequent updates on land use and transport integration. Several HF data have been identified as inputs into the model to get periodic updates.

At the end of the study, the selected model was tested using HF data substituting the traditional data the latter is time-consuming and expensive. Based on the detailed analysis of LUTI models with specified vital requirements, it is evident that most models are unsuitable for feeding the HF data. Few models are available in an open-source platform, such as UPLAN, URBANSIM, and PECAS. However, these models can be updated annually, and the spatial resolution will not support the satellite imagery retrieved from medium-resolution sources like Landsat. In contrast, ILUTE and CUBE Land can be used as a base model, but ILUTE is not yet operational practice (Salvini and Miller 2005).

Table 2.5: LUTI Models with Key Requirements

LUTI Models	Land use model	Transport Model	Key Requirements		
			Spatial resolution	Availability	Frequency of update
UPLAN	Utility-based	Four-Step	50x50m to 200x200m	Available Open Source	Based on users' assumptions
MUST-B	Utility-based	Four-Step	N/A	N/A	N/A
ILUTE	Microsimulation	Activity-based	Grid Cells and Zones	None/ Probably Discontinued	Flexible
URBAN SIM	Microsimulation	Four-Step	150x150m	Available Open Source	Annually
RAMBLASS	Microsimulation	Activity-based	500mx500m	N/A	Based on users' assumptions
ILUMASS	Microsimulation	Activity-based	100mx100m	Proprietary	Land use: month to year Synthetic Population: Annually
TRESIS	Microsimulation	N/A	N/A	Proprietary	N/A
PECAS	Microsimulation	Activity-based	N/A	Available Open Source	Annually
TLUMIP	Microsimulation	Activity-based	N/A	N/A	Economic annually and transport model once in three years
TIGRIS XL	Microsimulation	Discrete Choice Model	N/A	N/A	Annually
CUBE Land	Bid rent	Four-step and Activity-based	Parcels	Proprietary	Flexible
Metronamica	Microsimulation (Cellular Automata)	Four-Step	25m to 100m	Proprietary	Annually

2.9 Summary and Conclusion

This literature review-based study critically analysed the current application of HF data in LUTI. It was found that the high potential and usefulness of HF data in LUTI models are not widely conceded. The general findings of these types of novel source data can be used to develop a cheaper and more dynamic LUTI model than those involving traditional data. Miller (2018) pointed out the impact of short-term decisions on long-term results in determining the location of the households and job decisions and explained the integration of the short and long run in the same LUTI model (Miller 2018). However, most reviewed studies have focused on developing passively collected data sources and lack the focus to establish dynamic LUTI models to capture short-term impacts. Moreover, from the previous studies, the application of LUTI is predominant in the context of developed countries and very limited in developing countries.

A limited number of studies have explored the possibility of combining high-frequency (HF) big data with small data in Land Use and Transport Interaction (LUTI) models, such as integrating land values and satellite data. Similarly, exploring new data sources becomes critical to achieving desired model outputs when dealing with model complexity. To enhance LUTI models, it is vital to explore new methods and techniques to effectively leverage HF datasets.

Additionally, only a handful of studies have addressed land use and transport modelling despite the availability of several high-frequency data sources. Datasets like Call Detail Records (CDR) and smart card data include extensive individual mobility information, serving as valuable sources for microsimulation and complex transport models. In developing countries, and other nations with a low income, the availability of such data is often limited due to additional effort required for data provisioning. Similarly, satellite imagery data can be used to classify land use in a detailed manner, but low-cost or free satellite data are not available in developing countries.

In LUTI models, the development trends can be traced from the aggregate to the disaggregate level and from static (or cross-sectional) to dynamic (or longitudinal) models (Kii, Moeckel, and Thill 2019). Moreover, many articles have pointed out the general issues related to the LUTI models, such as the need for micro-level data and the complexity of the models, which provide a challenge for the HF data application. As explained, most recently developed LUTI models move towards activity-based models for transport analysis, considering the disaggregated level simulation (Individual). Even though the micro-level simulation might provide accurate LUTI model calibration, exploring relevant HF data to get detailed land use and transport information is still challenging. Considering the availability of each kind of HF data, the next chapter explains the availability of such data for the study area considered for the model development.

CHAPTER 3

APPLICATION OF HF DATA SOURCES FOR LUTI MODEL: A CASE STUDY ON WESTERN PROVINCE, SRI LANKA

3.1 Overview of the Chapter

The Western Province of Sri Lanka is the country's smallest province in terms of land area, but it is highly populated. The rapid urbanization of the province, especially in urban areas, is a debatable issue that has been addressed by several articles from different perspectives. However, despite the urgent need for effective land use and transport integration (LUTI) models to guide the province's development, most existing models have not been calibrated for use in Sri Lanka due to data collection and budget limitations (Bandara and Jayasinghe 2013).

This section seeks to bridge the gap in LUTI modelling and planning in Sri Lanka by leveraging the novel set of data collection and application facilitated by current technology advancements. By developing a methodology for the LUTI model that incorporates high-frequency data, such as data from mobile phones and other sensors, this research aims to provide a platform for more accurate and up-to-date information on transportation conditions and trends in the province.

The importance of developing an effective LUTI model for Western Province cannot be overstated. The province is a hub of economic activity in Sri Lanka, and its development has significant implications for the country's overall economic growth and competitiveness. However, the province is also facing significant challenges in terms of traffic congestion, air pollution, and other environmental issues. A well-calibrated and validated LUTI model can help policymakers and planners make informed decisions about transportation policies and investments, which can ultimately lead to more sustainable and liveable urban areas.

Moreover, this section focuses on Sri Lanka as an example of a developing country highlights the potential for using advanced data collection and analysis techniques to improve LUTI modelling and planning in other similar contexts. The methodology developed in this thesis can be adapted and applied to other urban areas in Sri Lanka and other countries with similar demographic, social, and economic conditions, providing a valuable tool for sustainable and equitable urban development. Due to the data availability Western Province in Sri Lanka

has been selected to develop the model as an example for rapidly urbanised area in developing countries.

3.2 Overview of Study Area

The Western Province of Sri Lanka, despite being the country's smallest province, is highly populated and has experienced rapid urbanization over the past decade. This urbanization has mainly affected the urban areas and has highlighted the need for an integrated model to address the resulting complexities. The region is the economic hub of Sri Lanka, contributing almost 60% to the country's Gross Domestic Product (GDP) and over 30% to industry. In 2013, the province recorded a share of 3643.2 billion rupees out of 8674.2 billion rupees, and in 2012, it also recorded the highest per capita income of United States Dollars (USD) of 2,922 (approximately Rs. 372,814). However, the continuous traffic congestion on urban roads and the decline in public transport services have negatively impacted the region's economic performance, environment, and quality of life.

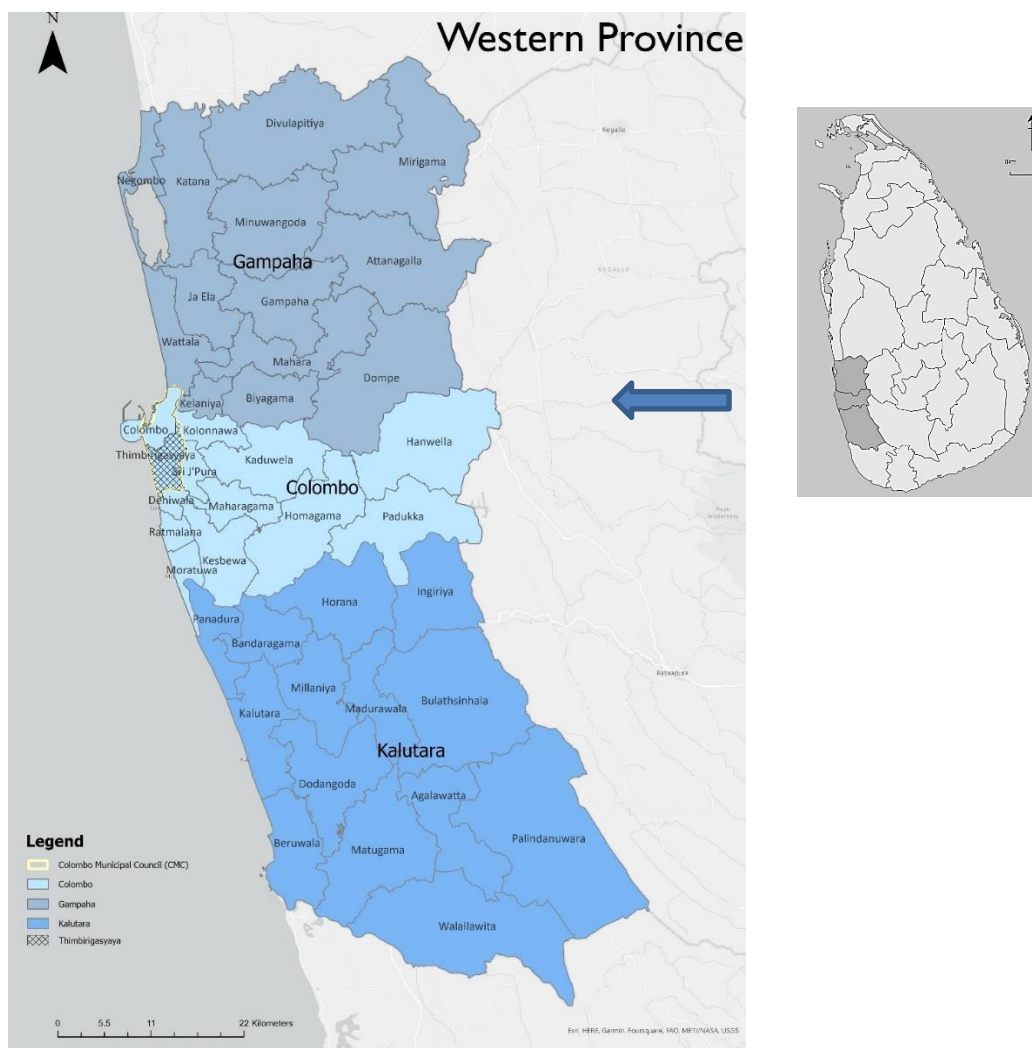


Figure 3.1: Overall Study Area (Western Province, Sri Lanka)

Colombo, Gampaha, and Kalutara are the three administrative districts that make up the Western Province in Sri Lanka, which is the smallest province but has a high population density, with approximately 5.8 million people in 2013 in a land area of 3,684 sq. km. By 2025, the population is estimated to reach 7.8 million, and by 2035, it is expected to increase to 9.1 million due to economic growth and the development anticipated in the Megapolis Structure Plan (CoMTrans, 2013). The Western Province is the country's economic hub, contributing almost 60% of the country's Gross Domestic Product (GDP) and over 30% of the industry, with a recorded per capita income of \$2,922 in 2012. In the context of Sri Lanka, A Divisional Secretariat Division (DSD) is an administrative division at the district level and a Grama Niladhari Division (GND) is a smaller administrative unit within a DSD. Colombo, Gampaha, and Kalutara are further divided into 33 DSDs, and there are 2,505 GNDs in the Western Province. Colombo covers an area of 699 sq. km, with 13 DSDs and 566 GNDs, while the Gampaha district covers an area of 1,387 sq. km with 13 DSDs and 1,177 GNDs, and the Kalutara district covers an area of 1,598 sq. km with 14 DSDs and 762 GNDs. In Sri Lanka, Colombo has the highest proportion of the urban population at 54.6%, followed by Gampaha with 14.6% and Kalutara with 10.6% (Figure 3.1). With the nation's busiest seaport and airport situated in the Western Province, and the expansion of such facilities in the pipeline, introducing structural changes to the transport system is crucial to make the metropolitan area modern, liveable, and investor friendly. As a result, residential and commercial land is scarcer and more expensive in major cities, particularly in urban areas, than in suburban areas.

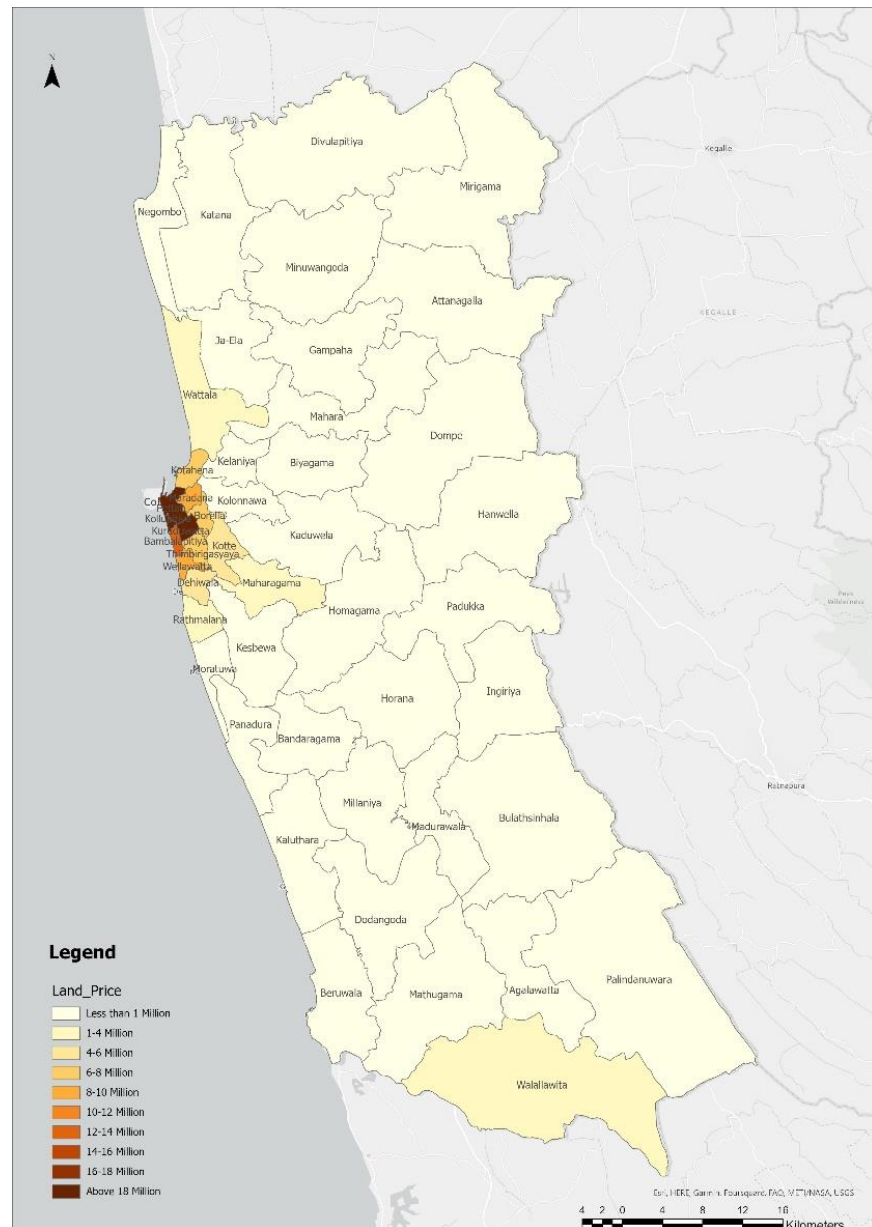
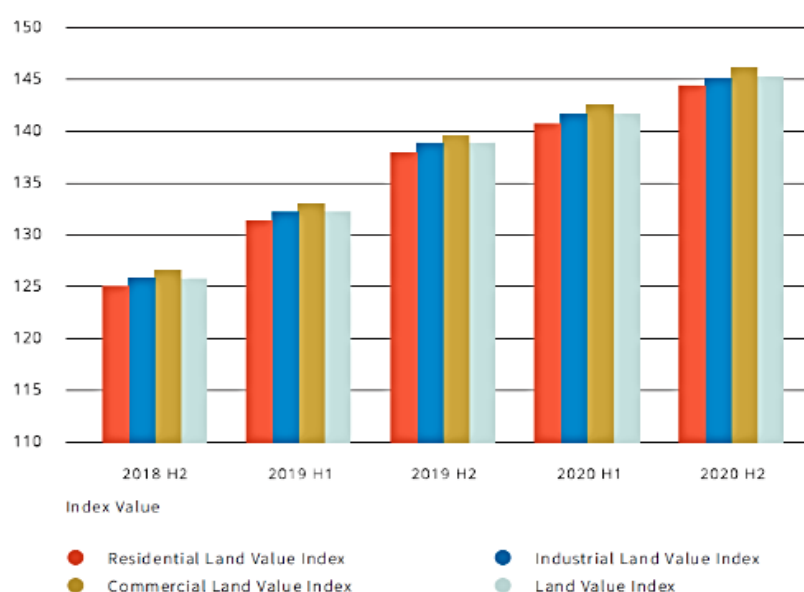


Figure 3.2 shows the average land prices (per perch) of major cities in the Western Province, where the information is retrieved from the Lanka Property web portal. Some of the suburban areas' data are missing due to the lack of information to calculate the average cost of each land plot. Based on the results, Colombo's land price has demonstrated a significant growth rate of 13.6%. It is also evident that the land price is higher in the urban area, especially in the Colombo Municipal Council (CMC) area. Among the cities, Nawala, a suburb located 6 km south of Colombo city, has recorded the highest land price value. Dehiwala, an adjacent suburb to Colombo city, shows the second-highest land price value, which has also recorded a considerable increase in land demand

from 2018 (nearly 64% increase in land value of 2018). Notably, due to the increasing demand inside the CMC and scarcity of land, most of the residential areas are developed in suburbs adjacent to the CMC, such as Piliyandala, Athurugiriya, and Pannipitiya, which have shown an increase in demand of 55%, 30%, and 29% from 2018, respectively. These suburbs are located around 15-20 km from the Colombo city center.

Figure 3.3 shows the Land Assessment Index (LVI) calculated by Central Bank for residential lands, commercial and industrial. The growth rate shows 1.28%, 14.9% and 13.2% for residential, industrial, and commercial, respectively, to 2019. Due to the covid-19 pandemic, the growth rate was slightly reduced to 10.5% for all the classifications.



Source: (CBSL, 2021)

Figure 3.3: Land Assessment Index in Western Province (2019)

3.3 Transport Pattern (Home-based Work) of Western Province

The national road network in Sri Lanka is centred around the commercial business district of Colombo. The Western Province of Sri Lanka is connected to Colombo through several major corridors, including Negombo Road (A3), Kandy Road (A9), Low-Level Road (B435), Malabe Road (A0-B240), High-Level Road (A4), Horana Road (B084), and Galle Road (A2). These corridors are vital for economic activities and are well-connected to other major cities in Sri Lanka.

To prepare a comprehensive long-term transportation plan for the Colombo Metropolitan Region and Suburbs, a Household Travel Survey (HVS) was

conducted in 2013 under the Urban Transport System Development Project for the Western Province of Sri Lanka (CoMTrans). The survey collected travel activity information from 38,500 households and socioeconomic data on the households and individuals. The data contains information about every trip undertaken by every household member, including the purpose of the trip, origin and destination, mode of transport, transfer point, departure time, and arrival time. The survey covers individuals of all ages, including children aged five or above, and includes all three districts of the Western Province (Colombo, Gampaha, and Kalutara).

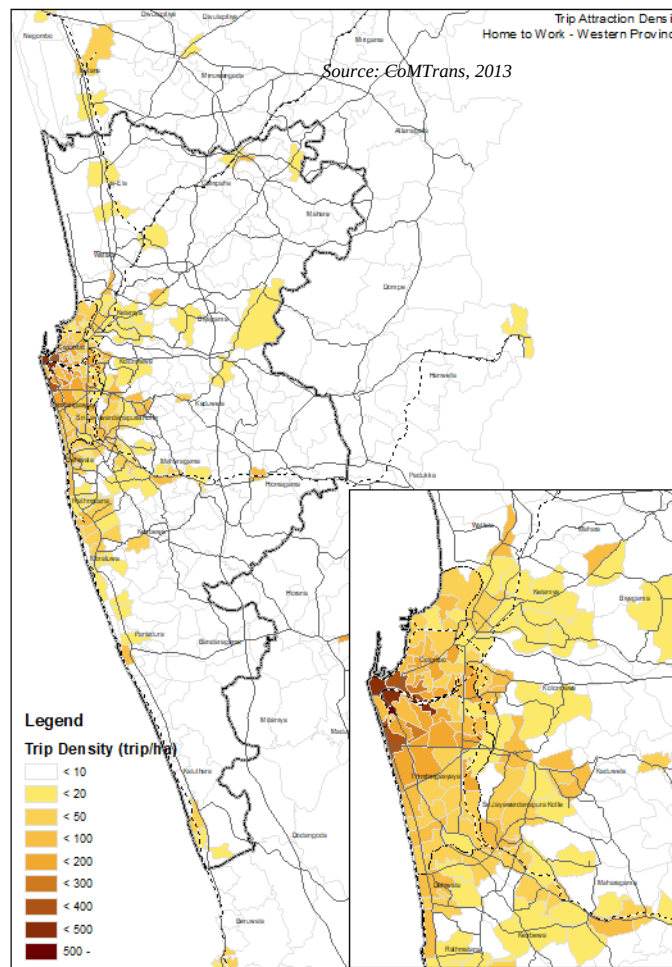


Figure 3.4: Home to Work Trip Density in Western Province (2013)

The Western Province of Sri Lanka has a significant working population, with 28% of the population engaged in full-time employment and 8.6% in part-time employment according to HVS data from 2013. The study area of 1.46 million households produces an estimated 10 million daily trips. Of these trips, approximately 1.8 million originate from the Colombo Municipal Council (CMC) (Figure 3.4), with 1.1 million being intra-city trips. The remaining 1.4 million trips are made between Colombo and the rest of the Western Province,

with most of these trips being between CMC and the surrounding areas.

Home-based work accounts for nearly 26% of total trips in Western province, with the CMC area attracting many residents from outside the city for this purpose. Commuting to work in CMC accounts for around 400 thousand trips daily, with over 70% of these trips coming from outside the city. The CMC is a popular destination for outside trips as most workplaces, educational institutions, and other facilities are located there.

Buses are the most dominant mode of transportation for getting to work, accounting for nearly 45% of trips, followed by cars, motorcycles, and non-motorized modes. Three peak periods can be detected within a day based on hourly fluctuations of trips. The morning peak lasts for about one hour between 7 A.M. and 8 A.M., accounting for nearly 24% of daily trips. In the afternoon and evening, peak periods last about two hours each, from 1 P.M. to 3 P.M. and from 5 P.M. to 7 P.M., accounting for 19% and 12% of daily trips, respectively.

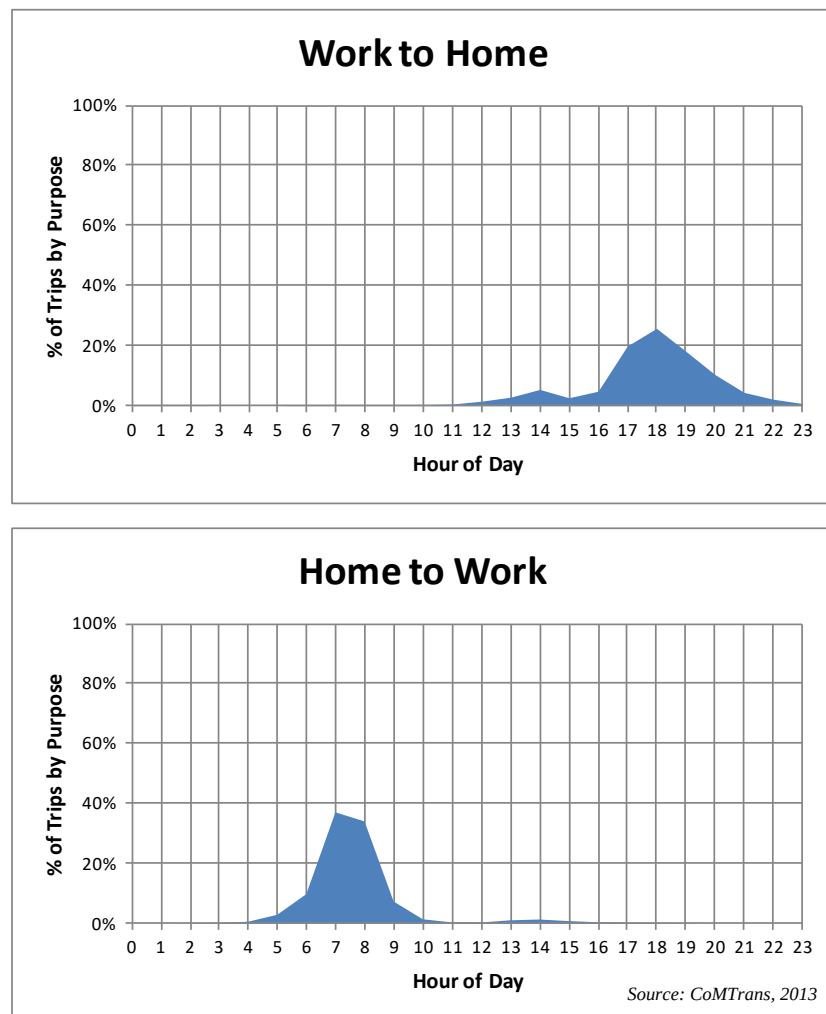


Figure 3.5: Hourly Fluctuation of Work Related Trips in Western Province (2013)

In the morning, peak hours are dominated by home-to-work and school-to-home

trips, while the afternoon and evening peak hours are dominated by school-to-home and work-to-home trips. Figure 3.5 shows the time distribution between home to work and work home. Additionally, bus shares are highest during the three peak times.

3.4 Overview of HF Data in Western Province

The study employed a methodology for the home-based work transport model by updating variables using HF proxy data from the recent year. The transport model was developed using transport and travel survey data in 2013 due to the unavailability and limitations of HF data such as CDR or smart cards in Sri Lanka. Among the different variables related to the trip generation model, monthly electricity consumption and the floor area of households are considered valuable variables since they can be derived from several sources and updated frequently. Electricity consumption data is used as a surrogate variable for household income classification, which can be updated monthly at disaggregated levels of households. The variables of land space of offices, shops, and industries are identified as influential factors to model trip attractions, which can be updated based on satellite imagery and GPS point data on electricity usage.

As explained earlier, the essential parameters of the LUTI model are land use and transport model development. The critical input for each model mentioned above is the extraction and classification of the built-up regions. Classifying these regions from any freely available medium-resolution satellite imagery, such as Landsat, requires more methodological work.

3.5 Analysis of Electricity Consumption GPS Data

3.5.1 Overview of Electricity Consumptions in Sri Lanka

In Sri Lanka, there are only two electricity providers, the Ceylon Electricity Board (CEB) and Lanka Electricity Company (Pvt) Ltd (LECO). CEB was established in 1969 and controls the island's electricity consumption. On the other hand, LECO began operations on 1st June 1984 and serves over 376,000 customers in the western and southern coastal belts between Negombo and Galle. The remaining electricity users are served by CEB, which is owned and operated by the Sri Lankan government. In 2019, electricity was the primary energy source for lighting in Western Province, accounting for nearly 99.1% of energy consumption. Other sources such as kerosene, solar energy, and others

were used by very few users, accounting for 0.5%, 0.3%, and 0.1%, respectively (CBSL 2018).

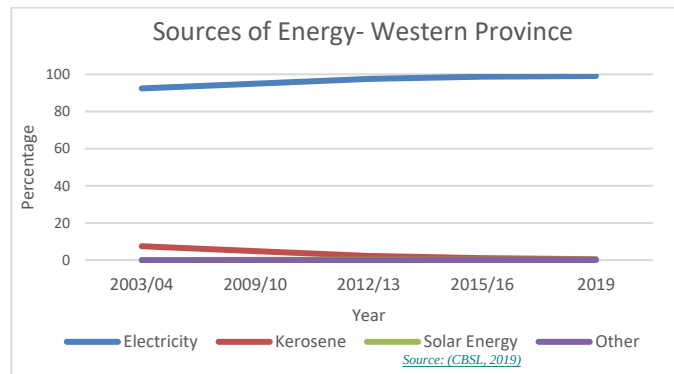


Figure 3.6: Percentage Use by Energy Sources in Western Province by Year

Figure 3.6 shows the primary sources of energy consumption in Western Province over time. The study also utilized GPS data of building classifications applied by CEB and LECO, the electricity providers in Sri Lanka. These providers classify the built-up area for tariff calculation into several categories, including domestic (apartments, adjoining houses, and single houses), general purpose (offices, shops, etc.), industrial, hotels, religious places, and others (schools, etc.) (Figure 3.7).

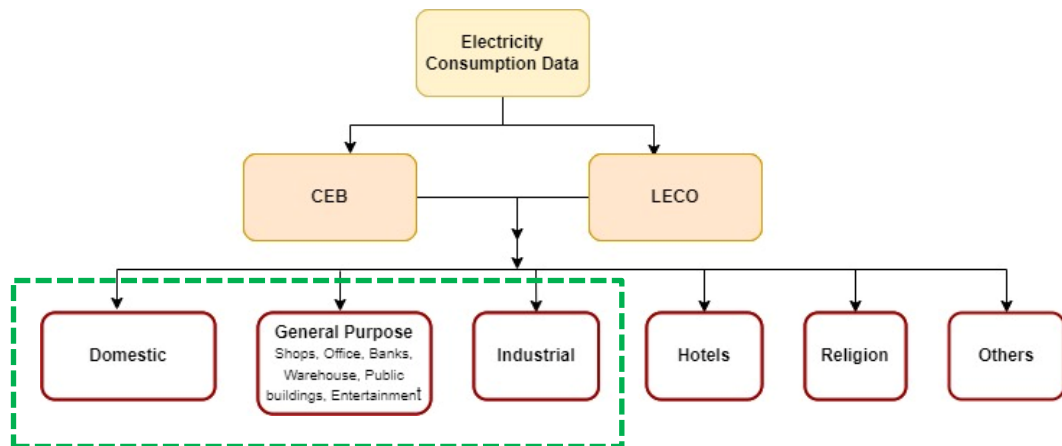


Figure 3.7: Tax Classification of Electricity Providers

The study collected secondary data of electricity customer data for the Western Province from the respective entities for 2013 and 2019. The primary objective of this research is to develop a model to test home-based work trips, and the data used for the model's development included the classification of domestic, general purpose, and industrial data. The data was converted from an electronic format to Microsoft Access files to sort it into zones, and it contains the customer code, GPS location of the customers, the type of consumer, and their monthly electricity consumption. A commercial ArcGIS program with a

database containing GND information in Sri Lanka was used to analyse the data further.

3.5.1.1 Analysis of Domestic GPS Points Data and Validation

The study has selected 2013 as the base year for model development and 2019 as the predicted year. The choice of 2013 as the base year could be attributed to the availability of survey data specifically pertaining to the Western Province area for that year. Conversely, 2019 has been designated as the predicted year, indicating the researchers' objective to construct a model capable of estimating or forecasting particular outcomes or variables for the Western Province area in 2019. According to that, the GPS point data of customers and their monthly electricity consumption were collected for October in both 2013 and 2019. The GPS point data for 2013 was analysed and validated against the latest census data in 2013, based on district (Table 3.1). The results showed an overall accuracy of 94% when compared to the census data with GPS counts, with 97%, 89%, and 97% accuracy for Colombo, Kalutara, and Gampaha districts, respectively.

Table 3.1: Census Vs GPS Counts of Households in Western Province (2013)

Districts	Census	GPS Count	%
Colombo	572,475	553,713	97
Gampaha	604,009	538,028	89
Kalutara	305,737	296,100	97
Total	1,482,221	1,387,841	94

Furthermore, Figure 3.8 shows a r^2 value of 91.6 with statistical significance (p value less than or equal to 0.05). Based on the r^2 value it can be concluded that variability in household GPS counts based on the electricity data can be explained by the census data based on the DSDs-wise analysis (Eq 3.1).

$$Total\ HHS_{census} = 1761.6 + 1.003 * Total\ HHS_{GPS}$$

(3.1)

Total HHS_{census} : Total households counted by Census in DSDs

Total HHS_{GPS} : Total households by GPS data from Electricity

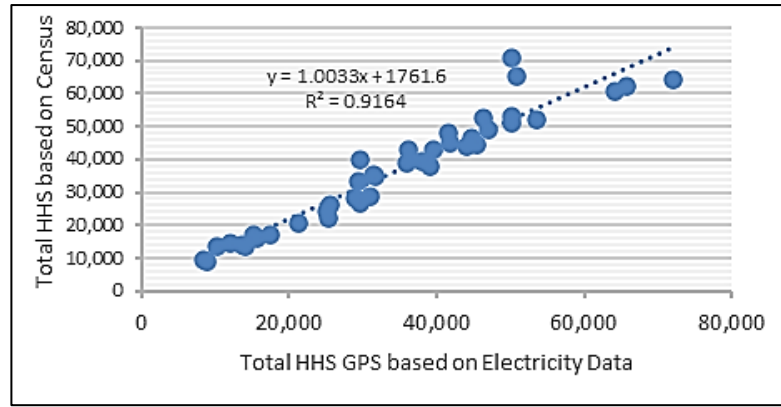


Figure 3.8: Census Vs GPS Electricity Data- 2013

Moreover, the relationship between these two variables also indicates a positive correlation coefficient of 0.95, further validating the strong positive relationship, where one variable tends to increase as the other variable increases. Overall accuracy and DSD-based accuracy best fit against the census data carried out in 2013. This further validates that the GPS counts of electricity customers can be used to get the total number of households based on the considered study area. Furthermore, it can also be used for the analysis of the electricity consumption of each household as a surrogate variable for the income in disaggregated levels based on the exact customer locations. The same analysis has been carried out between 2019 electricity data, and Table 3.2 shows the total number of GPS points detected in 2019 against 2013. The annual growth rate of households has been calculated based on the 2019 and 2013 household GPS points. The overall annual growth rate is 2.6% for the entire study area, and no data is available to validate this calculated value.

Table 3.2: Growth rate of Households in Western Province (2013 and 2019)

Districts	Households 2013	Households 2019	GF/Annum
Colombo	553,713	644,599	2.6%
Gampaha	538,028	627,191	2.6%
Kalutara	296,100	351,842	2.9%
Total	1,387,841	1,623,631	2.6%

Moreover, the analysis was further extended to classifying the households into the different categories such as single households, apartments and adjoining households for the purpose of developing disaggregated based model for trip generation using the GPS points sharing details. The below assumption has been made to classify the households further into single households, adjoining

households and apartment types.

- Single Unit (SH): Only a GPS point identified
- Adjoined House (AJH): more than one to three customers shared a GPS point
- Apartments/ Housing Scheme: More than three customers shared a GPS point

Based on the above assumption, the households' GPS points are further analysed based on the above categorisation, and the comparison between the 2013 and 2019 growth rates is shown in Figure 3.9.

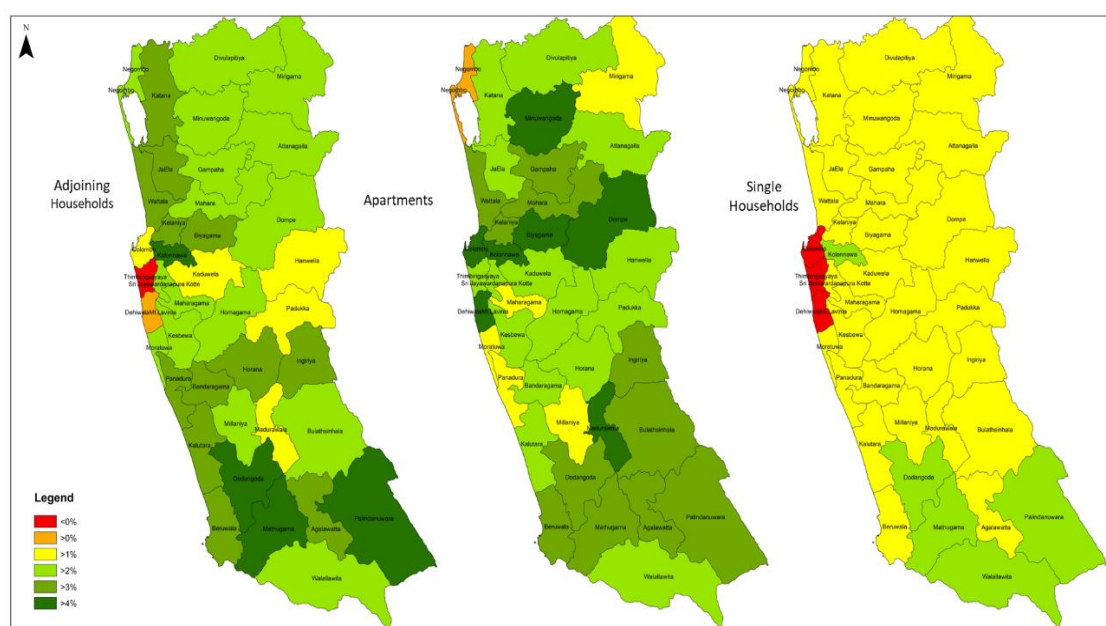


Figure 3.9: Households Growth Rate (2013 and 2019)

As per Figure 3.9, the rate of apartment-type households is increased in the urban area, and there is a considerable decrease in the single households. This is because of the lack of land space for further development inside the metropolitan area in Western Province for the residential area, and Government proposed more apartment type of buildings. It has also been noted that the average apartment development between 2013 and 2019 was about 4.02% inside the CMC area. In contrast, the single households and adjoining households' trends are about -0.54% and 2.04%, respectively. CMC area is the most congested area in Western Province, and due to the lack of land space, there is a considerable decrease in single households. Apartment types of residential buildings are increasing rapidly. It also can be noticed that single-household development and adjoining household development are growing in the suburban area, creating more trips than in urban areas.

3.5.1.2 GPS data of General Purpose

Customers' GPS data of general-purpose customers were also analysed to investigate the trend between 2013 and 2019. General-purpose categories include shops, banks, offices, and public buildings. In 2013 and 2019, 147,556 and 205,779 general buildings were detected, respectively, showing an annual growth rate of around 6.09%. Figure 3.10 displays the DSDs based on the growth rate of general-purpose buildings between 2013 and 2019. The highest number of general-purpose buildings was found in Colombo DSD, where 11,029 and 13,539 buildings' GPS were observed in 2013 and 2019, respectively.

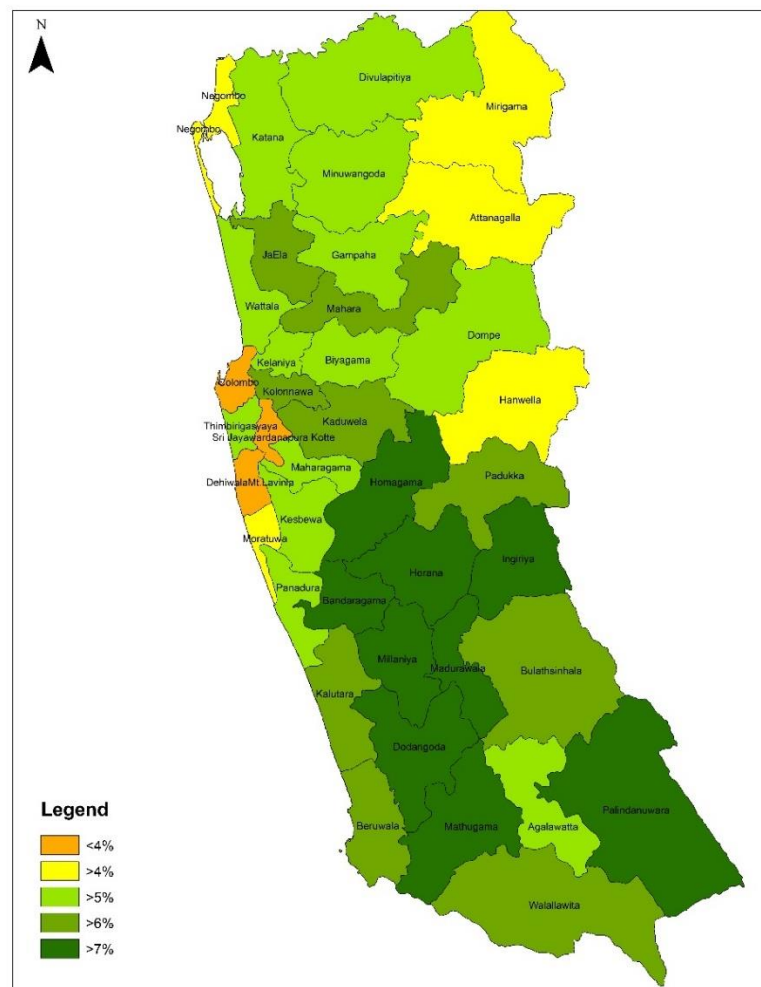


Figure 3.10: General Purpose Growth rate (2013 and 2019)

Based on Figure 3.10, the general-purpose development trend shows the same pattern as the household trend. The number of GPS points in general-purpose buildings increased rapidly in the suburban area, primarily in the southern part of the province (nearly 8%). There was also a gradual increase in urban areas,

especially in the CMC area (almost 4.19%), where most public business entities moved out of the CMC area and relocated to suburban areas such as Battaramulla and Malabe.

3.5.1.3 GPS data of Industrial Customers

There were 5,704 and 6,106 industrial customers whose GPS were detected in 2013 and 2019, respectively, with an annual growth rate of 1.45%. The graphical representation of the growth rate of industrial building between 2013 and 2019

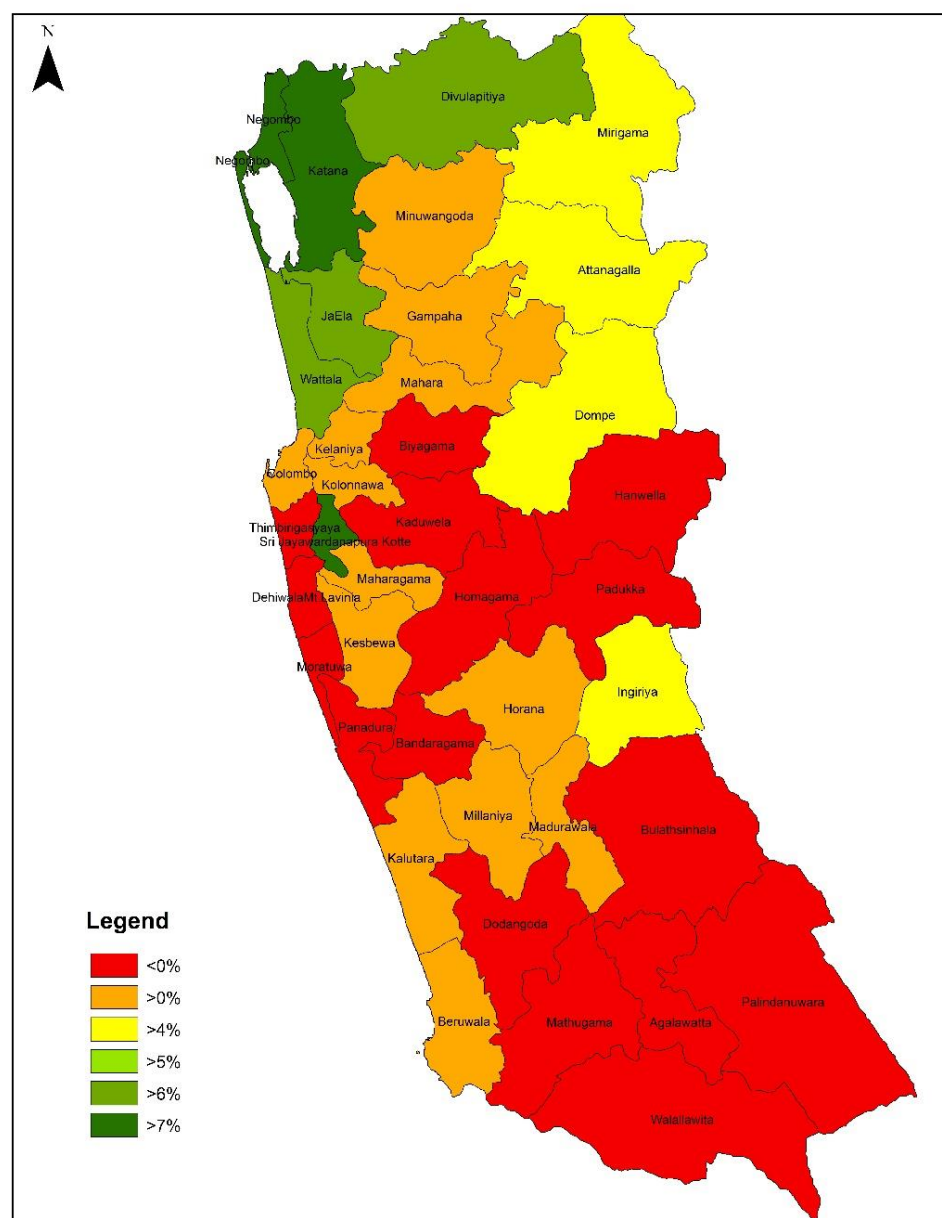


Figure 3.11: Industrial Buildings Growth rate (2013 and 2019)

is shown in figure 3.11. The highest GPS number of industrial buildings is detected in Moratuwa DSDs, where 856 and 701 GPS were seen in 2013 and 2019, respectively. Most of the export processing zones (EPZ) and industrial estates (IE) are in the highly detected area.

3.5.2 Analysis of Satellite Data

Spatial resolution refers to the size of the image's pixel, whereas temporal resolution refers to how frequently data can be obtained for the same pixel/area. The comparison of spatial and temporal resolution of different types of satellite images is shown in Figure 3.12.

Land use data can be differentiated by source and acquired by primary or secondary sources. Collecting preliminary data, such as land use surveys, requires specific requirements for sampling. On the other hand, detailed information must be collected at a high concentration (100% sampling) to expedite the more accurate calibration and validation (Agarwal et al. 2002). Secondary data such as satellite images provide wide-ranging information on land cover data, which can be collected at a meagre cost compared to manual collection. Various satellite image sources are available. Each source can be assessed based on the spatial resolution and frequency of acquisition with time (revisiting time) which are a crucial part of this research (Figure 3.12).

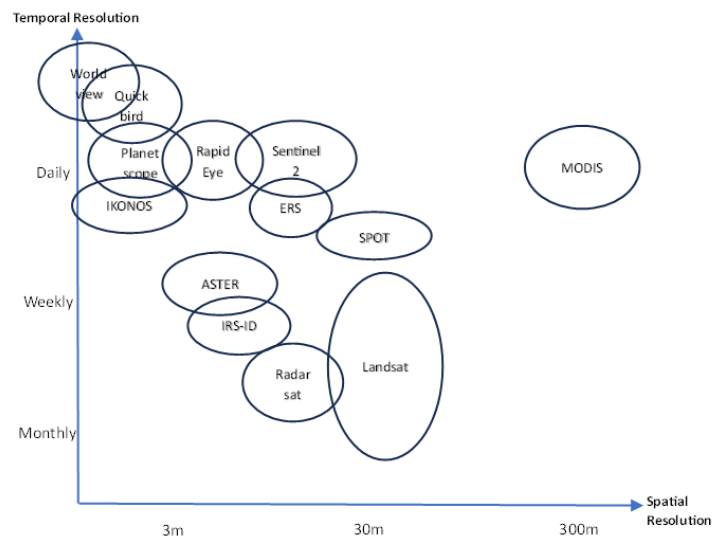


Figure 3.12: Spatial and Temporal Resolution of Satellite Images

There is a wide range of spatial resolutions and revisiting times available with the satellite systems listed. Several options are available for higher-resolution imagery, including SPOT, IRS-1D, RadarSat, ERS, PlanetScope, Rapid Eye, ASTER, IKONOS, QuickBird, and WorldView. Spatial resolutions of SPOT and IRS-1D are 20 meters and 5 meters, respectively, with revisiting times of 1-5 days and 22 days, respectively. Both RadarSat and ERS have a spatial resolution of 12.5 meters, with RadarSat revisiting every 24 days and ERS every 1 to 8 days. In addition to PlanetScope and Rapid Eye, WorldView offers a

revisiting interval of 1 to 5.4 days, whereas PlanetScope revisits daily. Options such as Sentinel 2 and MODIS are available for data with lower resolution and more frequent revisiting. Compared to MODIS, Sentinel 2 has a resolution of 1 kilometre, but revisits every 1 to 2 days, while Sentinel 2 has a resolution of 10 meters and a revisiting time of 10 days. AVHRR offers daily revisiting in 1-kilometer resolution, which is ideal for monitoring weather conditions. Depending on the application, the type of satellite system to use depends on detail levels and revisiting frequencies.

3.5.2.1 Overview of Landsat Data

Recently, changes in land use have been progressing due to rapid urbanization. Accurate information on current land use coverage and how land-use changes occur is crucial for urban planners and policymakers to assess the significant impact of land use changes on the urban system and the trend of land use patterns.

Most high-resolution satellites, such as SPOT, Quickbird, Worldview, etc., are expensive. Landsat 7 data has a spatial resolution ranging from 15 to 60 meters and a temporal resolution of 16-30 days. It is essential to evaluate the spatial and temporal resolutions of the LUTI models to assess the adaptability of Landsat images. Due to the limitation and cost of acquiring high-resolution satellite imagery for Sri Lanka, the study uses freely available Landsat images as these data are available for 2013 and 2019. The spatial resolution of Landsat and the revisiting time can satisfy the essential requirement of the frequent update of the LUTI model.

Landsat satellites have been operating since 1972, providing information about the Earth's land surface. Since 2008, the imagery has been available at no cost through the USGS. Based on the survey conducted by USGS in 2013, most land use researchers have used Landsat data as their primary source.

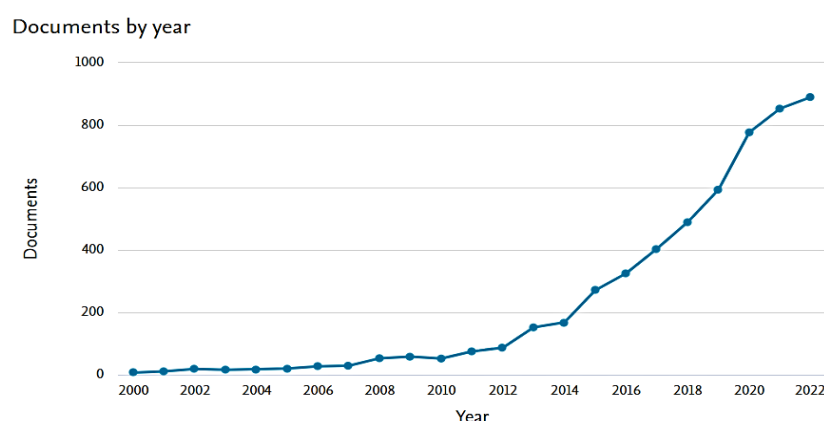


Figure 3.13: Journal Papers (Open access) Published using Landsat Data (2000 to 2022)

Figure 3.13 depicts the number of open-access journal papers published between 2002 and 2022 based on Landsat satellite information. The information was retrieved from the Scopus and Web of Science databases. Due to its free availability, this data has been used as a source for most studies. Landsat data for land use and land cover supports various other applications such as environmental sciences, planning and development, earth science etc. Landsat data can also be used in various models related to LUTI, such as population density distribution models.

3.5.2.2 Land Use Maps Classification Based on Landsat Data

Land use classification is based on land use and land cover. The fundamental difference between land use and land cover is that land use typically shows people's landscape use. The land cover indicates the physical type of land, such as forest, river, etc. The USGS has developed major land use classifications based on multiple levels (Table 3.3)

Table 3.3 Land use Classification by USGS

Level 1	Level 11
1. Urban or Built-up Land	11. Residential
	12. Commercial and Services
	13. Industrial
	14. Transportation, Communications and Utilities
	15. Industrial and Commercial Complexes
	16. Mixed Urban or Built-up Land
	17. Other Urban or Built-up land
2. Agricultural Land	21. Cropland and pasture
	22. Orchards, Groves, vineyards, Nurseries and Ornamental Horticultural Areas
	23. Confined Feeding Operations
	24. Other Agricultural Land

3. Rangeland	31. Herbaceous Rangeland
	32. Shrub and Brush Rangeland
	33. Mixed Rangeland
4. Forest land	41. Deciduous Forest Land
	42. Evergreen Forest Land
	43. Mixed Forest Land
5. Water	51. Streams and canals
	52. Lakes
	53. Reservoirs
	54. Bays and Estuaries
6. Wetland	61. Forested Wetland
	62. Non forested Wetland
7. Barren Land	71. Dry Salt Flats
	72. Beaches
	73. Sandy Areas Other than Beaches
	74. Bare Exposed Rock
	75. Strip Mines, Quarries, and Gravel Pits
	76. Transitional Areas
	77. Mixed Barren Land
8. Tundra	81. Shrub Brush Tundra
	82. Herbaceous Tundra
	83. Bare Ground
	84. Wet Tundra
	85. Mixed Tundra
9. Perennial Snow or Ice	91. Perennial Snowfields
	92. Glaciers

Several land-use classifications adopted by LUTI models, and the most common classification are provided in Table 3.4, which is adopted for this study as well. This classification is divided into two main categories based on the critical area of LUTI models.

Table 3.4: Land use Classification Considered for Studies

Transport	Highways
	Interchanges
	Major Arterials
	Minor Arterials
	Railway stations
	Airport
	Port
Other	Residential
	Commercial
	Industrial

	Mixed (Residential/ Commercial)
	Government
	Vacant developable lands
	Vacant undevelopable lands

3.5.2.3 Developing a Machine Learning Algorithms for Western Province to Establish the Land Use Trends

As the first step, all available Landsat TM, ETM+ and OLI surface reflectance products from 1 January to 31 January (next year) for 2013 and 2019 were obtained from the USGS data archive. The suitable explicit imagery was acquired by applying a cloud coverage filter. The Landsat image for 2013 was acquired on 20th Jan 2014 as there were no suitable clear images available during 2013, and 2019 imagery was obtained on 10th Jan 2019. Figure 3.14 shows the retrieved original images for 2013 and 2019 from the Landsat database. These images are pre-processed using Arc GIS pro software before starting the classification.



Figure 3.14: Landsat Satellite Images (2013 and 2019)

The three most common remote sensing classification methods from satellite images can be used for land cover classification: unsupervised classification, supervised classification, and object-oriented image analysis. Many studies have focused on land-use change detection analysis based on various techniques. Recently, the application of machine learning algorithms for land use mapping has been drawing considerable attention (Talukdar et al. 2020). Machine learning techniques can be categorized into two sub-techniques: supervised and unsupervised.

Supervised classification techniques include

- Support vector machine (SVM)
- Random forest (RF)
- Spectral angle mapper (SAM)
- Fuzzy adaptive resonance theory-supervised predictive mapping (Fuzzy ARTMAP),
- Mahalanobis distance (MD)
- Radial basis function (RBF),
- Decision tree (DT),
- Multilayer perception (MLP),
- Naive Bayes (NB),
- Maximum likelihood classifier (MLC), and
- Fuzzy logic

Unsupervised classification techniques include

- Affinity Propagation (AP)
- Cluster algorithm
- Fuzzy c-means algorithms
- K-means algorithm
- ISODATA (iterative self-organizing data) etc.

Source: (Talukdar et al. 2020)

Supervised classification involves selecting training samples and classifying the image based on those samples. Each pixel in the overall image will inherit a class based on the training samples the user selects. In unsupervised classification, clusters are generated using similar spectral characteristics. In the next step, each cluster is classified without any training samples. Supervised and unsupervised classification assigns land cover class based on per pixel, and those pixels are the same size and shape and have no concept of neighbours. However, image segments are grouped into small pixels with vector objects in object-oriented classification. Instead of pixel-basis classification, segment digitizes the classification as an image.

3.5.2.4 Obtaining the area of the building using supervised classification

The study employed supervised classification of satellite data using ArcGIS Pro and 2013 land use classification data to obtain the area of general purpose and industrial buildings. The model was tested in a small area to validate the supervised classification, and Thimbrigasyaya DSD was selected for testing as it is known for mixed land use in the Western Province.

A sufficient number of samples were taken from the satellite images of 2013 based on the following land covers: water, barren land, developed area, and

vegetation. The results of the performed supervised classification for the selected testing area are shown in Table 3.5. The results indicate that a trained, supervised model can be used as it matches the actual land use survey of 2013, which was conducted for the CoMTrans study in 2013.

Table 3.5: Supervised Classification of Landsat images for Thimbirigasyaya DSD

Land cover	Detected		Actual
Water	291,145	1%	1%
Developed	19,230,340	86%	90%
Barren	2,080,679	9%	8%
Vegetation	755,899	3%	1%
Total	22,358,063	100%	100%

The developed supervised model was applied to other DSD areas in Western Province, and the result is shown in Table 3.6.

Table 3.6: Actual vs Supervised Land use Classifications in 2013

Districts	Land Use Classes	Actual			Supervised Classification			Developed Area (Actual vs Supervised)
		Water	Developed	Barren	Water	Developed	Barren	
Kalutara	Bandaragama DSD	6.42	34.77	31.73	2.02	37.75	7.89	109%
	Beruwala DSD	3.10	45.52	30.01	2.11	32.62	8.77	72%
	Dodangoda DSD	No data			2.34	66.03	11.19	No data
	Horana DSD				0.01	91.62	9.45	
	Kaluthara DSD	5.74	39.33	37.02	3.68	40.75	16.84	104%
	Madurawala DSD	No data			0.93	42.34	1.88	No data
	Agalawatta DSD				0.44	38.03	4.03	
	Bulathsinhala DSD				3.03	107.72	11.65	
	Palindanuwara DSD				2.88	99.53	7.65	
	Walallawita DSD				1.33	39.72	3.62	
	Ingiriya DSD				0.20	59.42	1.05	
	Mathugama DSD				2.20	61.89	13.49	
	Millaniya DSD				0.95	61.71	4.17	
	Panadura DSD	12.11	31.37	14.02	5.50	21.84	9.31	70%
Gampaha	Attanagalla DSD	No data			0.17	93.55	2.13	No data
	Biyagama DSD	1.44	51.61	16.94	0.25	53.24	0.82	103%
	Divulapitiya DSD	No data			1.10	110.41	1.48	No data
	Dompe DSD				0.29	118.96	1.51	
	Gampaha DSD	1.35	71.24	54.67	0.20	76.16	3.57	107%
	Ja-Ela DSD	3.59	50.94	33.34	0.21	52.08	1.39	102%
	Katana DSD	3.18	86.49	34.45	17.34	83.74	3.16	97%
	Kelaniya DSD	2.89	21.95	4.48	0.30	19.60	0.36	89%
	Mahara DSD	0.20	73.37	48.74	0.09	69.86	0.60	95%
	Minuwangoda DSD	No data			0.05	96.65	1.07	No data
	Mirigama DSD				0.36	97.03	1.29	
	Negombo DSD	1.53	27.17	8.33	17.87	23.05	2.21	85%
	Wattala DSD	7.10	36.21	24.76	2.20	38.93	5.11	108%
Colombo	Colombo DSD	2.48	16.92	1.08	1.52	21.41	0.96	126%
	Dehiwala DSD	0.25	9.81	0.45	0.21	7.69	0.32	78%
	Hanwella DSD	No data			1.71	95.36	1.62	No data
	Homagama DSD	2.89	76.18	61.65	0.56	95.87	10.89	126%
	Kaduwela DSD	6.42	67.97	25.91	1.70	77.50	0.90	114%
	Kesbewa DSD	7.49	45.52	21.55	1.76	40.98	14.40	90%
	Kolonnawa DSD	4.23	23.93	3.37	0.85	22.85	0.54	95%
	Maharagama DSD	1.37	35.43	9.76	0.42	34.77	0.61	98%
	Moratuwa DSD	8.34	16.09	0.79	2.09	14.25	1.56	89%
	Padukka DSD	No data			0.90	67.36	0.46	No data
	Rathmalana DSD	0.48	12.96	1.90	0.19	9.96	2.23	77%
	Sri J'pura Kotte DSD	4.90	17.90	0.77	0.86	13.99	0.53	78%

	Thimbirigasyaya DSD	1.08	22.37	1.93	0.29	19.23	2.08	86%
	Summary							96%

The actual value in Table 3.7 is from the land use survey conducted for the CoMTrans study in 2013. The values for all DSDs are not available, and the comparison of the supervised classification vs actual has been conducted based on the available DSDs results. The validation of each DSD is performed according to the district's basis, and the total developed area comparison has been made based only on the available DSDs. Accordingly, the accuracy of each district, Kalutara, Gampaha, and Colombo, is 88%, 99%, and 104%, respectively, with an overall accuracy of 96% for the entire study area. The supervised classification values are applied directly for the DSDs that do not have the actual values. According to the derived developed area from the supervised classification, the area of general purpose is calculated based on the in-detail land use classification available with the CoMTrans study in 2013.

3.6 Summary and Conclusion

Due to the vast development in modern technological aspects, numerous inventions related to electromagnetic spectra and wireless communication facilitates the traffic data collection to be non-intrusive manner at any location in road networks (Shariff et al.2016). Most developing countries, including Sri Lanka, face challenges due to extensive data requirements and high acquisition costs. High frequency datasets that provide aggregated and disaggregated information can help address data availability issues and the impacts of rapid land use changes. However, many developing countries, such as Sri Lanka, do not have access to such data and require additional effort to acquire it. Exploratory efforts have been undertaken in Sri Lanka to substitute traditional travel data, such as Call Detail Records (CDR). However, it's crucial to note that this research is still in the exploratory stage and requires a more robust methodological approach for the analysis of such significant big data sources (Kumarage, Maldeniya, and Lokanathan 2015).

Remote sensing technologies can be used to analyse satellite imagery data and obtain detailed land use classifications. However, it is rare for developing countries to obtain free or low-cost satellite data with high resolution. Despite the numerous sources of high frequency data for land use and transportation studies, a promising direction for developing a model for the Western Province

of Sri Lanka is to use medium resolution Landsat data and electricity consumption data from GPS locations of customers, combined with standard classifications practised by electricity providers. The main advantage of these kinds of datasets is their frequent availability with low or no cost. Electricity consumption data with distinct classification based on the tariff classification can provide a good source of information to develop a model for transport and land use. Most of the developing countries has similar classification like Sri Lanka such as households/ domestic, industry and commercial (Siyambalapitiya 2018)

This preliminary analysis and background of the electricity and satellite imagery data are helpful for an excellent start to model development. Calibration of the model requires an understanding of the general behaviour of the variables. Therefore, a good understanding and validation of these data could be helpful for the development of land use and transport models, as discussed in chapter 4.

CHAPTER 4

DEVELOPING THE TRIP GENERATION AND ATTRACTION MODELS USING HIGH-FREQUENCY PROXY DATA

4.1 Introduction

Travel demand forecasting is a critical aspect of transportation planning, and various methods and models have been developed to estimate travel demand in urban areas. One widely used approach is the four-step demand model, which consists of trip generation, trip distribution, mode choice, and trip assignment. Among these steps, trip generation and attraction play a crucial role in predicting travel demand. Trip generation estimates the total number of trips generated by households or other land uses, whereas trip attraction identifies the number of trips attracted to a particular area based on land use activities. Accurately estimating trip generation and attraction is essential for designing transport facilities and developing sustainable transportation and land use planning policies. Thus, a comprehensive understanding of the factors influencing trip generation and attraction is essential for developing reliable travel demand forecasting models (Mukherjee and Raghuram Kadali 2022).

However, many studies have pointed out several constraints related to land use-based trip generation and attraction models, especially in developing countries, due to the lack of updated land use data or any other data related to travel demand (Ahmed et al. 2020, Jayasinghe, Sano, and Rattanaporn 2017). TP Amalan et al. discussed the data requirement of building a transport model, especially in developing countries. They introduced a new concept of data called “High Frequency Data” to solve the current issues (Amalan et al. 2023).

Against this backdrop, employing different methods to estimate trip attractions and generation models can solve the aforementioned data constraints in developing countries. Accordingly, the study focuses on developing a home-based work trip generation and attraction model based on freely available HF data such as electricity consumption and satellite images. The study area is based on Western Province, Sri Lanka, the smallest province but highly urbanized in Sri Lanka. Numerous articles address the rapid urbanization of Western Province, particularly in urban areas, in several dimensions.

4.2 Data set and Methodology

This study uses electricity consumption data from the CEB and LECO, the

primary electricity providers in Western Province, Sri Lanka (Fig 4.1). The electricity consumption data and GPS location of each customer of Western Province have been gathered for 2013 and 2019. The customer classification used by the electricity providers is given in Figure 4.1. Since the study focuses only on home-based work trips, the classification of domestic, ger and industrial customers details were used for modelling purposes. The comparison between the number of households calculated from the GPS points of electricity consumption data and the census and statistics data of 2013 shows the validation of 102%.

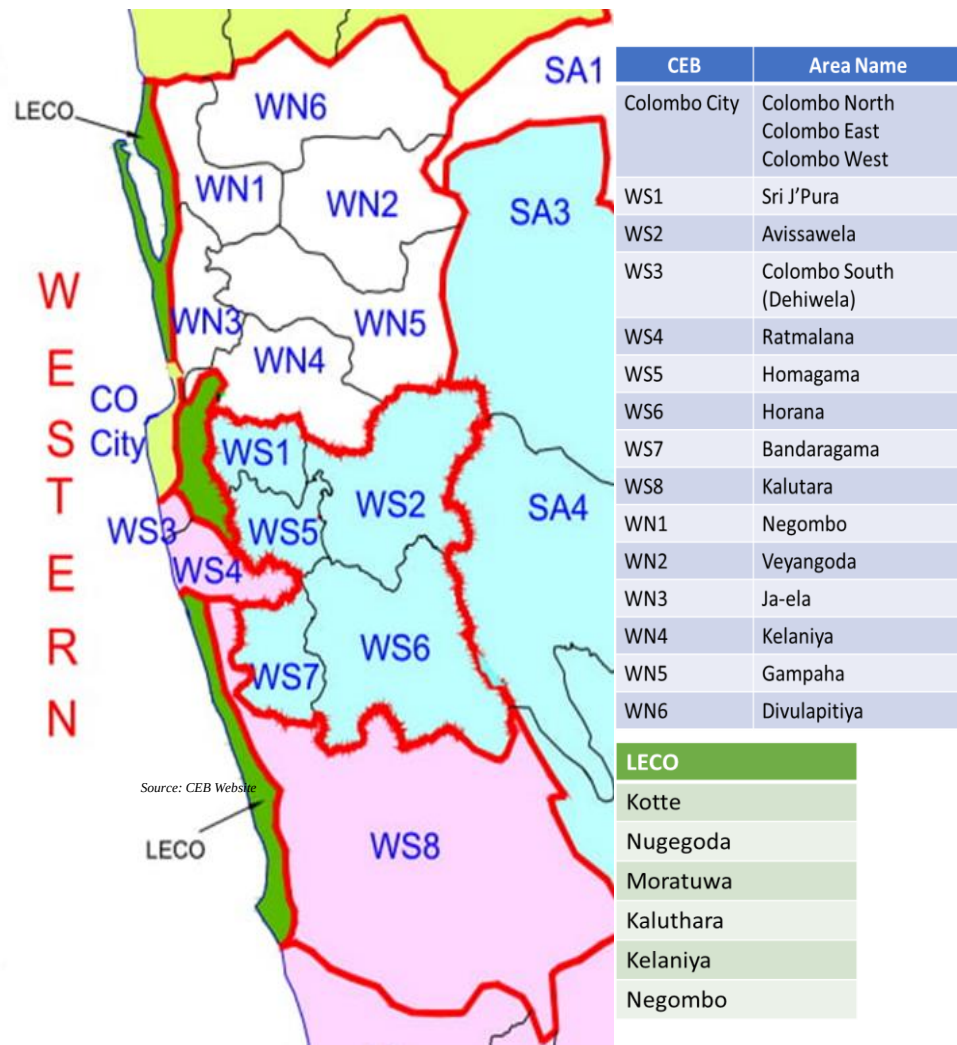


Figure 4.1: Area Classification of CEB

Furthermore, the GPS data was analysed concerning other classifications and provided in Table 4.1 for 2013 and 2019 with the annual growth rate of each building classification.

Table 4.1: Annual Growth Rate

Year	Households	General Purpose	Industrial
2013	53,628	6,448	21
2019	60,347	8,972	24
Annual Growth rate	2.0%	5.7%	2.3%

Figure 4.2 shows the relationship between the number of households and the total number of trips generated from each Traffic Analysis Zone (TAZ) considered for CoMTrans study in 2013. The number of households is calculated from the electricity consumption GPS points data. The number of trips generated from each zone is calculated from the home visit survey (HVS, 2013) conducted for the urban transport master plan for the Colombo metropolitan region and the suburbs (CoMTrans) in 2013. It is evident that trip generation is high in the suburban area where the number of households is also high.

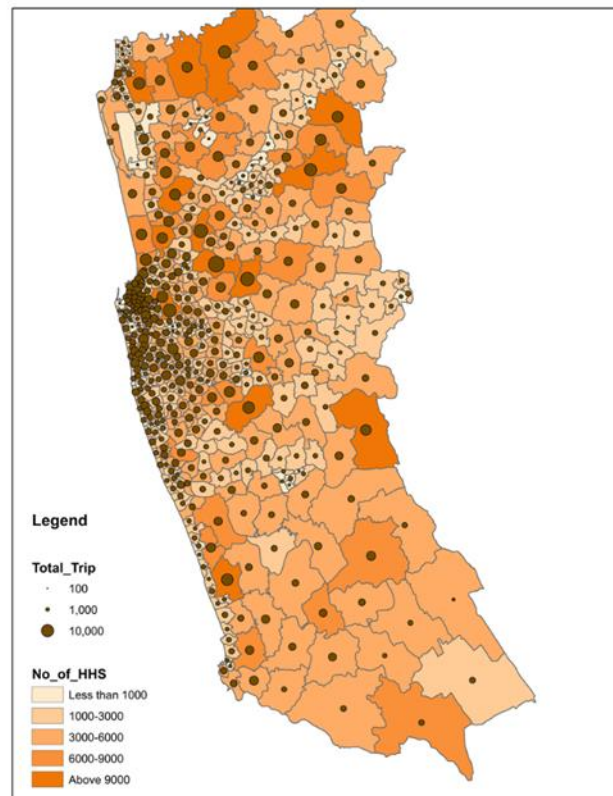


Figure 4.2: No of Households vs Total Trips Generated

4.3 Overview of Trip Generation Models

In general, trip generation can be analysed as two types of models. The first model involves a simple linear regression in which trip generation depends on socioeconomic characteristics. The regression models are structured with

dummy variables for discrete class analysis to treat the problem, as socio-economic attributes are not necessarily continuous, and their effects are not linear. For trip generation, households can be classified based on their cross-characteristics. In regression models, the explanatory variables are assumed to be continuous and vary widely.

In the second method, called the cross-classification method, the explanatory variables are taken to vary directly, so no assumptions are required on the variables' linearity range. The Federal Highway Administration developed the cross-classification technique in 1975 to determine the number of trips associated with residential land users. A cross-classification model groups households based on their socioeconomic characteristics, such as the number of vehicles owned, the size of the household, and income.

4.4 Regression Model for Trip Generation

Statistical regression analysis is the most widely used method for investigating and modelling the relationship between variables. Early-period regression analysis techniques were prevalent among developed countries, and numerous recent applications in travel estimations are found in developing countries. The concept of linear regression analysis is to find the best relationship between the independent variables (X) and the dependent variable (Y). To quantify the strength of the relationship and the use of methods that allow for prediction of the response values given the values of regressor X. The reasonable form of linear relationship between the two variables is shown in Equation 4.1, where β_0 is known as the intercept and β_1 as the slope.

$$Y = \beta_0 + \beta_1 X \quad (4.1)$$

The statistical error between the observed value of Y and the straight line ($\beta_0 + \beta_1 X$) is represented by ε . The more plausible regression model is represented in Equation 4.2.

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (4.2)$$

In most applications, as with this study, there will be more than one regressor that helps explain Y. The multiple regression equation is used in such situations. It is shown in Equation 4.3.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \varepsilon$$

(4.3)

This model contains Y as the natural dependent variable, and $X_1, X_2, \dots, X_n$ as the independent variable. As a result, multiple linear regression analysis is performed. The term linear indicates that the model is linear in the parameters $\beta_0, \beta_1, \dots, \beta_n$, not because Y is a linear function of X 's. Estimating unknown parameters in the regression model is one of the most critical objectives of regression analysis.

A second phase of regression analysis involves checking the model's adequacy, which determines the fit's quality and the model's appropriateness. Based on the results of the adequacy check, the model may be reasonable, or the original fit should be modified. As a result, regression analysis is an iterative procedure in which data are used to select a model, and the model is then fitted to the data. The model is either modified or adopted depending on the quality of the fit (Michael and Yasir 2005). For all steps of this travel forecasting model, regression analysis techniques were extensively used for both unknown parameter estimation and model adequacy checking.

4.4.1 Factors affecting Home based Work Trip Generation

Trip generation is discussed under three main categories based on trip purpose. They are home-based work (HBW), home-based non-work (HBNW) and non-home-based (NHB) trips. This research analyses the home-based work-related trip generation modelling. At the household level, home-based work trips are primarily influenced by two factors. The first and most important factor is the ownership of a vehicle. Second, it is affected by the household's average income. A direct correlation exists between the number of people and the trip rate per household. The Federal Highway Administration of the U.S. Department of Transportation (DOT) (1975) has shown consistent relationships between car ownership and income from trip generation. In many cities, this has been done, and the trend indicates that income and car ownership significantly affect travel demand. However, despite it being difficult to determine income or car ownership at an individual level in Sri Lanka, this concept can be applied using census data available at the aggregate zonal level (GND level).

4.5 Relationship between Income and Electricity Consumption

Several studies were conducted to validate the relationship between income and

electricity consumption (Athukorala and Wilson 2010, Kim 2020a, Rajmohan and Weerahewa 2007) (Table 4.2). The two studies, Athukorala and Wilson (2010) and Rajmohan and Weerahewa (2007), provide insights into the relationship between income (GDP) and electricity consumption. Athukorala and Wilson (2010) used unit root, integration, and error correction models based on data spanning from 1960 to 2007. They found that the long-run income elasticity of demand indicates a significant increase in electricity demand with an increase in household income (GDP). Specifically, their findings suggest that a 1% increase in GDP will lead to a 0.78% increase in electricity consumption. This highlights the positive relationship between income and electricity usage. On the other hand, Rajmohan and Weerahewa (2007) examined the "energy ladder" hypothesis and estimated Engle functions using data from Consumer Finances and Socio-Economic Surveys covering the period from 1978/79 to 2003/04. Their findings indicated that in 2003/04, the highest income decile consumed the most electricity, accounting for 64% of total consumption. This suggests that higher-income households tend to have higher electricity consumption levels, aligning with the concept of the energy ladder hypothesis. There is a consistent observation that higher income levels are associated with increased electricity consumption. Athukorala and Wilson (2010) demonstrated a positive income elasticity of electricity demand, indicating that a rise in income is likely to lead to a substantial increase in electricity usage. Rajmohan and Weerahewa (2007) further support this trend by showing that the highest-income households tend to consume a larger share of total electricity. These findings collectively indicate that income plays a significant role in shaping electricity consumption patterns, with higher-income individuals and households consuming more electricity compared to their lower-income counterparts.

Table 4.2: Literature on Income vs Electricity Consumption

References	Technique Used	Findings
(Athukorala and Wilson 2010)	Unit root, integration, and error correction models- Data from 1960-2007	The long-run income elasticity of demand shows that any future increase in household income (GDP) will likely significantly increase the electricity demand. A one per cent increase in GDP will lead to a 0.78 per cent increase in electricity consumption

(Rajmohan and Weerahewa 2007)	The 'energy ladder' hypothesis was tested, and Engle functions were estimated using Consumer Finances and Socio-Economic Survey data from 1978/79 to 2003/04.	The highest income decile consumed more electricity (64%) in 2003/04
-------------------------------	---	--

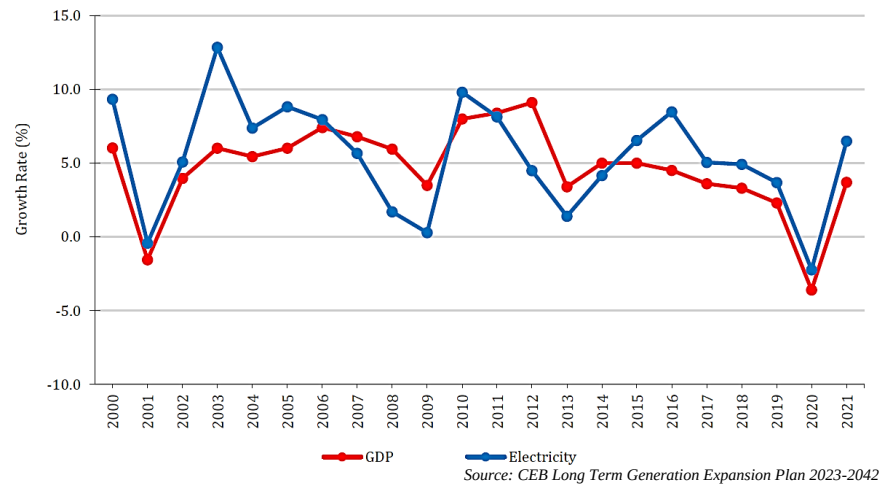


Figure 4.3: GDP Vs Electricity Growth Rate- Sri Lanka

Historical data in Sri Lanka shows a direct correlation between electricity demand and economic growth rates. Figure 4.3 compares the yearly growth rates of electricity demand and GDP, and it can be observed that there is a strong positive relationship between the two. Additionally, per capita electricity consumption in Sri Lanka has steadily increased from 2001 to 2020, as shown in Figure 4.4. This figure also depicts a strong positive relationship between per capita electricity consumption and per capita GDP ($r^2 = 91\%$ and $p\text{-value} < 0.05$), which can be used to predict future electricity demand based on the projected GDP for Sri Lanka and the provincial demand analysis.

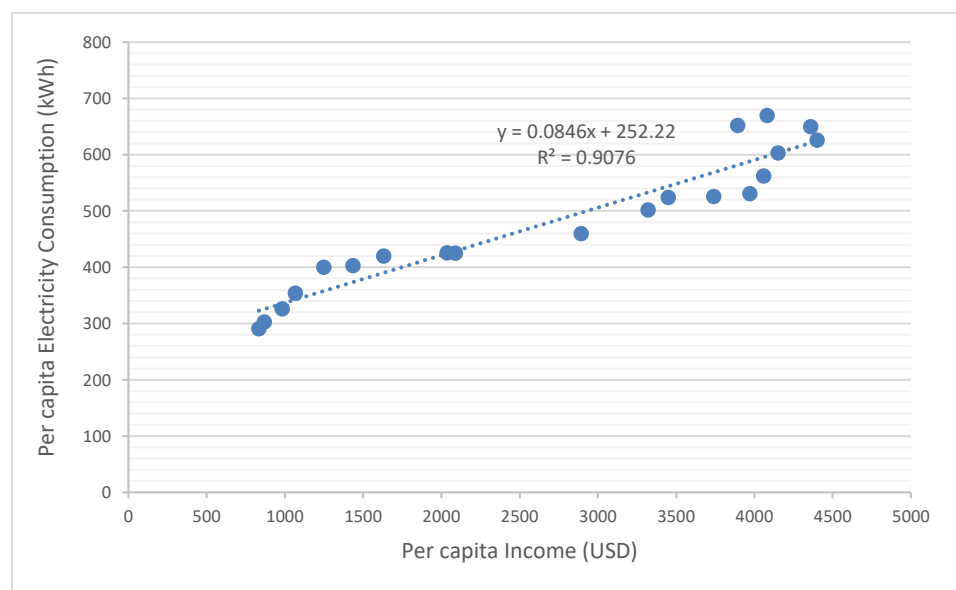


Figure 4.4: Relationship between Per Capita Income vs Per Capita Average Electricity Consumption

Similarly, Western Province exhibits the same trend between the GDP rate percentage and domestic sales consumption percentage from 2015 to 2020 (Fig. 4.5). The data on GDP is obtained from the CBSL website, and domestic consumption percentage is obtained from CEB sales and statistics reports.

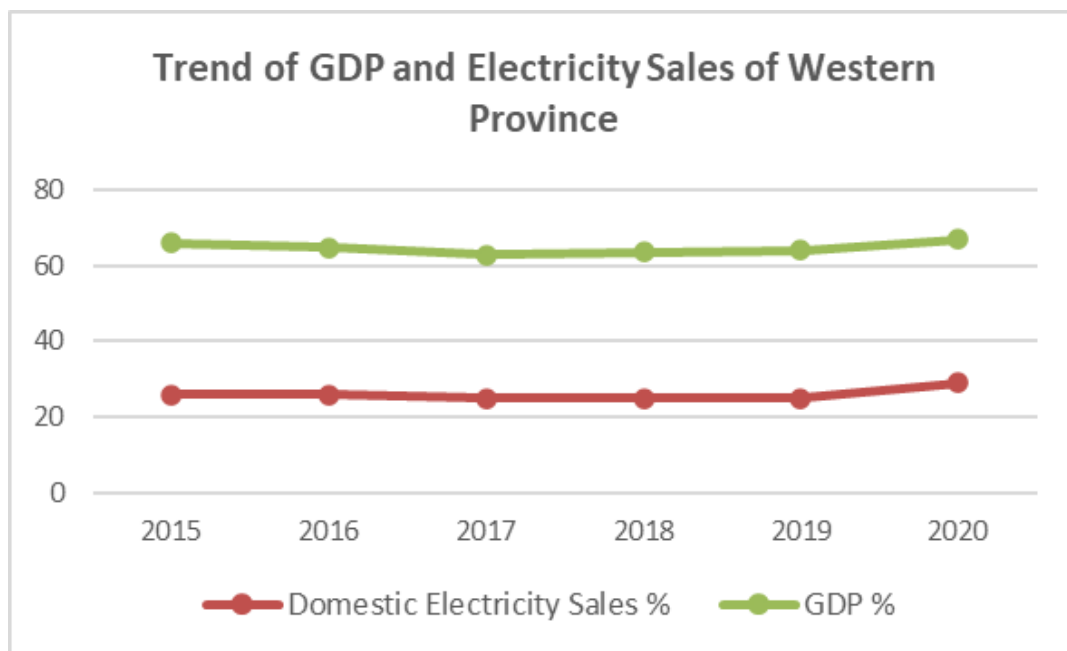


Figure 4.5: Trend of GDP and Domestic Electricity Sales of Western Province

Moreover, based on the 2013 electricity consumption data, it is evident that the average electricity consumption shows significant variation around Colombo city, where beyond a radius of approximately 20 km, the values remain constant and low (Figure 4.6). According to the income and expenditure survey in 2013, Colombo has the highest income receivers, with a mean income of 40,620 Sri Lankan Rupees (LKR) per month. Additionally, it is noteworthy that mean income is higher in urban areas compared to suburban areas in Sri Lanka. Furthermore, the average electricity consumption is higher in the suburban areas along the major corridors such as Negombo, Kandy, Galle, and Horana.

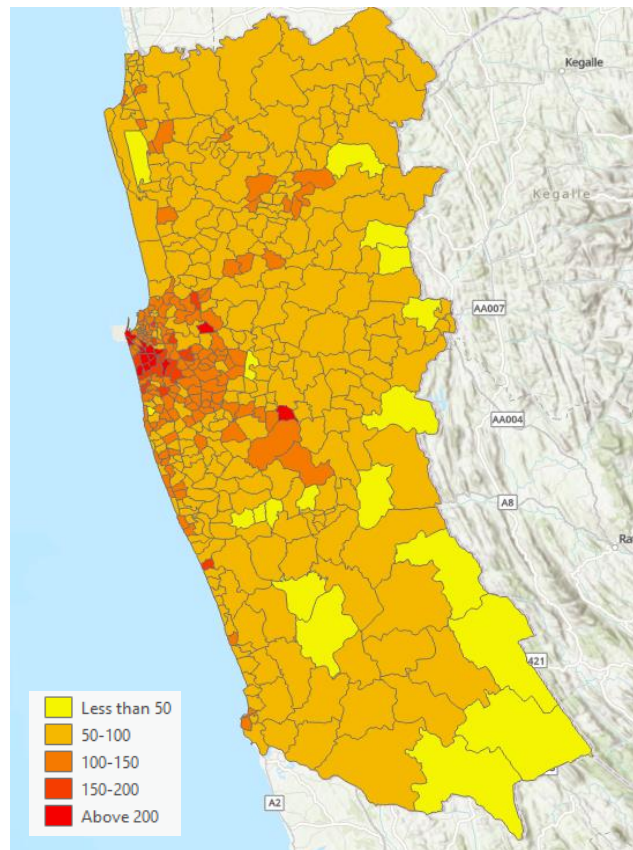


Figure 4.6: Average Electricity Consumption by TAZ (2013)

4.5.1 Relationship of Electricity Consumption based on HVS Survey 2013

The primary analysis was conducted on the data of HVS 2013 to show the relationship between electricity bills per month and income. The income group is categorised based on the CoMTrans study 2013 classification, and descriptions are as below,

- Low income: Household monthly income is less than 40,000 (LKR)
- Middle income: Household monthly income is between 40,000 to 60,000 LKR
- High income: Household monthly income is above 60,000 LKR

The below analysis is undertaken for the entire Western Province, Sri Lanka. The income and electricity expenses distribution are shown below (Figure 4.7).

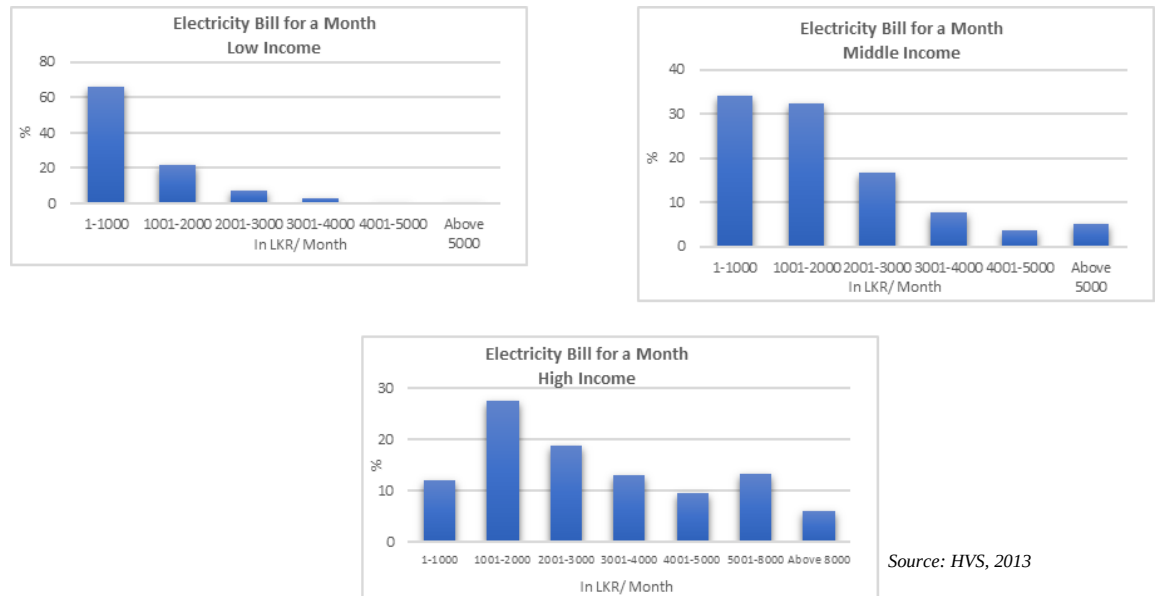


Figure 4.7: Households' Electricity Consumption vs Income (2013)

Based on figure 4.7, it is evident that the relationship between income and electricity expenses has strong relationship. Most low-income households' electricity consumption expenses fall under less than 1000 LKR, while middle-income expenses fall between 500 to 3000 LKR. Nearly 88 % of households under high income spend 1000 LKR on monthly electricity. The descriptive statistics of each income level are shown in table 4.3, and the conversion of electricity expense to the unit (kilowatt hour- kWh) is also shown in the same table based on the mean value of each group.

Table 4.3: Descriptive Statistics of Income vs Electricity Consumption

Source: HVS, 2013

Income Level	Mean	Stand. Error	Avg. Electricity unit (kWh)
Low	731.469	3.935	78
Middle	1347.852	8.025	97
High	2053.099	16.981	118

4.6 Developing Trip Generation Model and Validation

Trip generation models can be challenging to estimate and apply in many developing countries due to difficulties in conducting extensive household surveys to collect data. However, HF proxy data is becoming increasingly available due to advances in ICT and offers advantages over traditional data such as near real-time updates, high accuracy data collection at little or no cost, and the ability to collect data at a disaggregated level. The central aim of this section is to develop a trip generation model using HF proxy data, with a focus

on (1) investigating the input data required for the model development, (2) assessing the different sources of available HF proxy data and their usage as input to the model, and (3) calibration and validation of the developed model. The results of this section provide insight into how trip generation models can be developed under data constraints, which could be useful for developing countries.

The trip generation model was developed based on average monthly electricity consumption data and the number of households per zone, using the 462 TAZs considered in the CoMTrans 2013 study of Western Province, Sri Lanka. Fourteen zones were created in the external points of the study area, and the survey zones are the same as GN Divisions, except for two GN divisions (Thimbirigasyaya GN Division and Kurunduwatta GN Division) which were further divided into smaller survey zones due to their large size and location in CMC. The independent variables of the number of households and average monthly electricity unit were calculated from electricity data collected from CEB and LECO in 2013, while trip generation values were calculated from the HVS 2013. A multiple linear regression relationship (Eq 4.4) was developed between these variables, and the goodness of fitness and p-value between the variables are shown in Table 4.4.

$$\text{Home to work Trips} = 168.924 + 0.831488 (\text{No of households}) + 2.524818 (\text{Electricity Units/Month})$$

(4.4)

Table 4.4: Regression Analysis (Trip Generation)- 2013

Intercepts and Variables	Coefficients	Standard Error	t statistics	P-Value
Intercept	168.924	112.9558	1.495488	0.013546*
No of households	0.831488	0.01982	41.95125	5.5E-159***
Electricity Units/Month	2.524818	0.734969	3.435271	0.000646***
R²	0.796239			
F-Test	894.8684			
Observations	461			

$p < 0.05^*$, $p < 0.01^{**}$, $p < 0.001^{***}$

The value of the regression analysis is shown as the r^2 value of 79% between the independent and dependent variables, which shows the relationship between the variables is good with an acceptable p-value ($p < 0.05$). Based on the developed model, the number of generated trips from each DSD was calculated and compared with HVS 2013 trip generation (Figure 4.8).

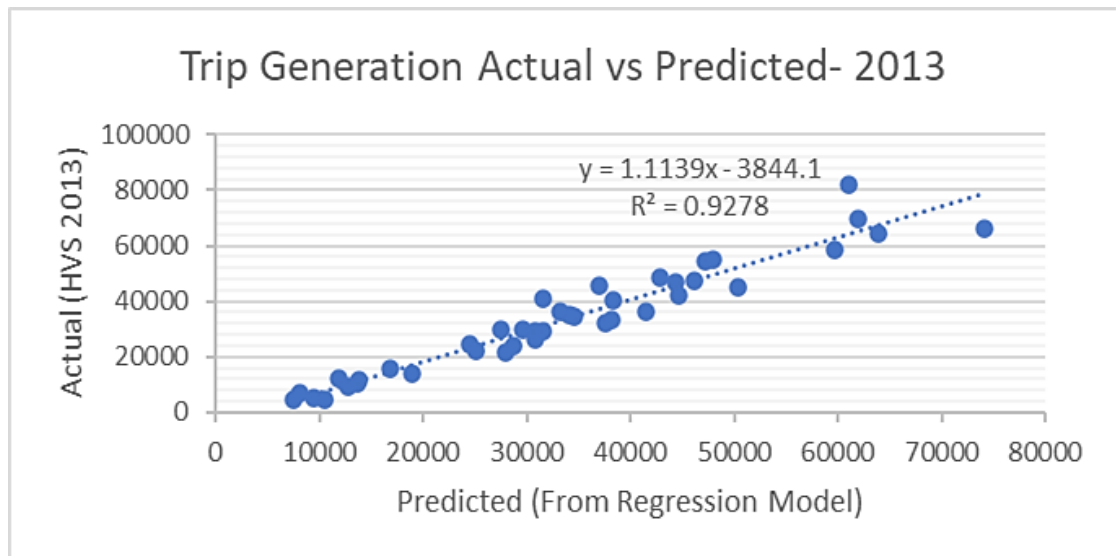
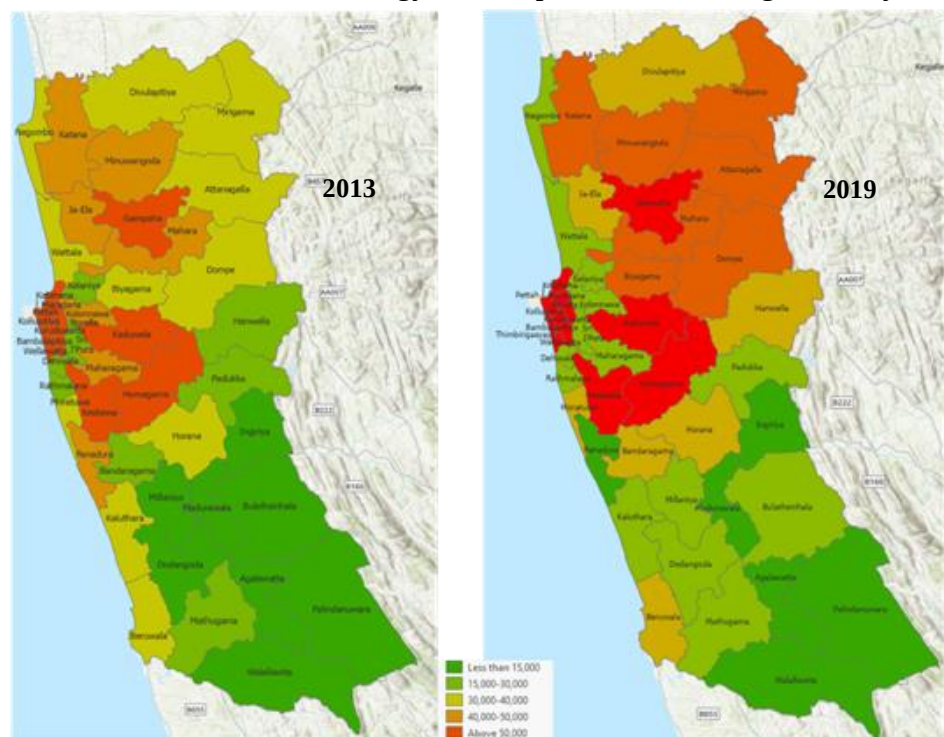


Figure 4.8: Home to Work Trip Generation (Actual vs Predicted)- Western Province (2013)

Figure 4.8 shows that the trip generation calculated from the regression model provides a perfect fit (92%) against the actual value calculated from the HVS 2013. The same regression model was used to calculate the trip generation for 2019 based on the electricity consumption data acquired for October 2019.

Figure 4.9 shows the comparison of trip generation between the years 2013 and 2019. The number of trip generation in the suburban areas has increased as the number of households, and their energy consumption increased significantly in



2019.

Figure 4.9: Home to Work Trip Generation in Western Province (2013 and 2019)

4.7 Developing Trip Attraction Model and Validation

The relationship between trip attraction and a zone depends on several factors, such as land use area, employment opportunities, and the number of employees working in those facilities. In this study, trip attraction models were developed for both general-purpose and industrial buildings within specific zones. The variables to develop a trip attraction model for these purposes were derived based on the area of the traffic zones and the area of the relevant land use. Chapter 3 explains how the area for different land use cover classifications, such as water, barren, and developed, was calculated using supervised classification from the Landsat satellite image of 2013 and validated with the existing land use survey conducted in 2013. The methodology (Fig 4.10) below shows how land use cover classifications for 2019 were calculated based on 2013 and electricity GPS data.

To derive the trip attraction for 2019, the model uses the GPS points of electricity customers, assuming that the growth rate of general-purpose and industrial areas is proportionate to the increase in GPS points of the customers between 2013 and 2019. Based on the satellite imagery and land use survey data for 2013, Figure 4.10 also explains how 2019 values were updated based on satellite imagery and GPS data collected from electricity customers. For example, in Thimbirigasyaya in 2013, general-purpose buildings shared 12% of the total area. The GPS points of electricity customers for general purpose increased from 6448 to 8972 with an overall growth rate of 39%.

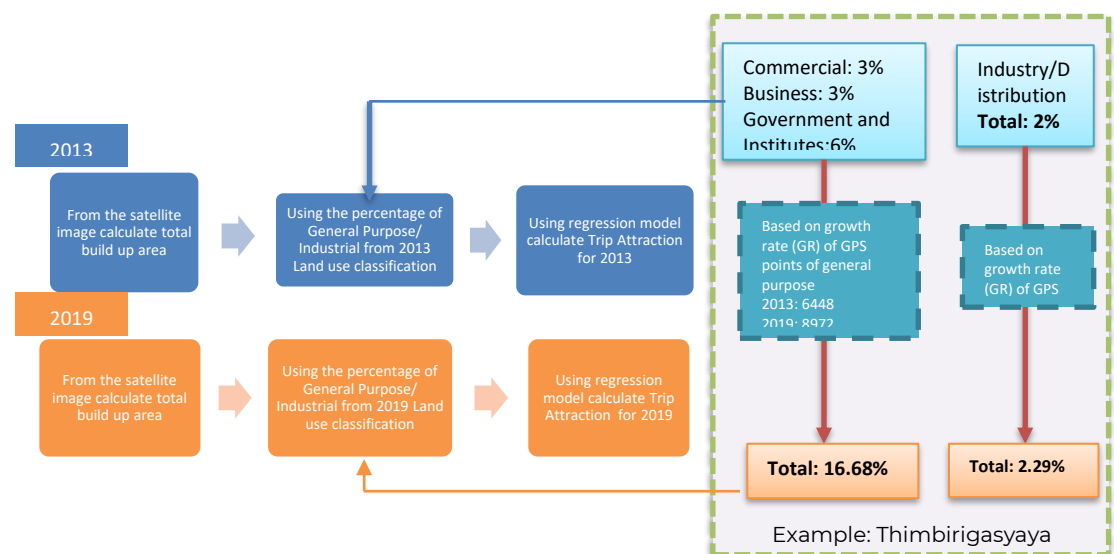


Figure 4.10: Methodology used for Calculating Area of General and Industrial

This percentage increase was applied to the 2013 share and calculated that in 2019, these customers shared 16.68% of the total area. Likewise, these percentages were calculated for all DSDs and are shown in Table 4.5.

Table 4.5: Area of General Purpose and Industrial Buildings in 2019

DSD	2013 Electricity Customers' GPS points		2019 Electricity Customers' GPS points		% Increase		% of area from Land use classification 2013		% of area for 2019	
	General Purpose	Industrial	General Purpose	Industrial	General Purpose	Industrial	General Purpose	Industrial	General Purpose	Industrial
Agalawatta	687	23	1008	18	47	-22	0.2	1.0	0.3	0.8
Attanagalla	4897	146	6544	180	34	23	0.7	1.7	0.9	2.1
Bandaragama	2587	276	4205	269	63	-3	1.0	1.0	1.6	1.0
Beruwala	3889	28	5640	34	45	21	2.0	1.0	2.9	1.2
Biyagama	6767	237	9482	233	40	-2	2.4	8.0	3.4	7.9
Bulathsinhala	1242	44	1899	44	53	0	0.1	0.1	0.1	0.1
Colombo	12321	64	15384	78	25	22	20.0	12.0	25.0	14.6
Dehiwala/Mt. Lavinia	2662	36	3427	34	29	-6	7.0	1.0	9.0	0.9
Divulapitiya	3600	202	5066	322	41	59	0.2	1.1	0.3	1.8
Dodangoda	1242	41	1947	40	57	-2	0.2	0.0	0.3	0.01
Dompe	4072	132	5631	171	38	30	0.4	0.7	0.6	0.9
Gampaha	7076	240	10273	302	45	26	2.0	0.6	2.9	0.8
Hanwella	3676	134	5034	129	37	-4	0.7	1.5	1.0	1.4
Homagama	6805	358	10861	337	60	-6	2.0	1.5	3.2	1.4
Horana	3311	237	5640	263	70	11	1.5	3.0	2.6	3.3
Ingiriya	1243	51	2112	61	70	20	0.2	1.0	0.3	1.2
JaEla	6107	144	8915	218	46	51	4.0	4.9	5.8	7.4
Kaduwela	8774	207	12993	207	48	0	3.0	1.5	4.4	1.5
Kalutara	4336	167	6331	181	46	8	3.5	0.9	5.1	1.0
Katana	7156	110	9790	165	37	50	3.0	4.9	4.1	7.4
Kelaniya	4419	155	6165	168	40	8	5.0	9.5	7.0	10.3
Kesbawa	7803	375	11093	415	42	11	2.0	1.6	2.8	1.8
Kolonnawa	4378	93	6463	98	48	5	8.0	5.4	11.8	5.7
Madurawala	596	38	1089	45	83	18	0.2	1.0	0.4	1.2
Mahara	4356	204	6497	235	49	15	1.0	1.0	1.5	1.2
Maharagama	7008	183	9662	217	38	19	3.2	1.6	4.4	1.9
Mathugama	2177	72	3634	55	67	-24	0.3	1.0	0.5	0.8
Millaniya	754	92	1350	107	79	16	0.1	0.1	0.2	0.1
Minuwangoda	4887	159	7017	206	44	30	1.0	0.4	1.4	0.5
Mirigama	3995	123	5326	155	33	26	0.6	0.1	0.9	0.1
Moratuwa	3426	874	4652	729	36	-17	5.2	5.7	7.1	4.8
Negombo	5358	28	7065	51	32	82	4.0	9.7	5.3	17.7
Padukka	1615	109	2486	107	54	-2	0.3	0.2	0.5	0.2
Palindanuwara	781	30	1212	17	55	-43	0.0	0.1	0.02	0.1
Panadura	5180	392	7267	380	40	-3	2.8	5.3	3.9	5.1
Rathmalana	2308	45	2978	58	29	29	7.0	15.6	9.0	20.1
Sri J'pura Kotte	4684	19	5796	40	24	111	7.0	1.2	8.7	2.5
Thimbirigasyaya	6448	21	8972	24	39	14	12.0	2.0	16.7	2.3
Walallawita	935	45	1432	31	53	-31	0.0	0.1	0.02	0.1
Wattala	4204	147	5959	207	42	41	5.0	5.0	7.1	7.0

4.7.1 Trip Attraction Model for the General Purpose and the Industrial Area

The trip attraction model has been developed based on the area of each land use occupied by different categories. The study considered two types of trip purposes based on the land use categorisation considered by the electricity providers: general purpose land use that includes shops, office banks etc. and industrial land uses. Due to the different characteristics of trip attraction levels, targeted analysis

and policy implications, as well as data availability, the data analysis aims to produce separate two-equation models for general-purpose buildings and industrial areas, examining the relationship between dependent and independent variables for each category individually. In the first model, the dependent variable considered is the number of trips attracted per square kilometre (km²) to the general-purpose building. The independent variable in this case is the area of the general-purpose building per square kilometre (km²). (Eq 4.5). The model was tested on 32 DSDs based on the available land use data for 2013. The goodness of fit of this developed model shows 70.7% with an accepted p-value ($p \leq 0.05$) (Table 4.6).

$$\text{General Purpose attracted trips/ km}^2 = 13.79302 + 8638.311 \text{ Area of general purpose building/ km}^2$$

-(4.5)

Table 4.6: Regression Model (General Purpose)- 2013

Intercepts and Variables	Coefficients	Standard Error	t statistics	P-Value
Intercept	13.79302	22.13066	0.623254	0.05000*
Area of general purpose building/ km ²	8683.311	1018.669	8.52417	0.00000***
R²	0.7077			
F-Test	72.66147			
Observations	32			

$p < 0.05^*$, $p < 0.01^{**}$, $p < 0.001^{***}$

Similarly, the model developed for the industrial area (Eq 4.6) uses industrial trip attracted per km² as a dependent variable, and the area of the industrial per km² is considered an independent variable (Table 4.7). The model gives an r² value of 79.57% with an acceptable level of p-values ($p < 0.05$).

$$\text{Trips attracted to the Industrial area/ km}^2 = 16.68874 + 4935.869 \text{ Area of industrial building/ km}^2$$

– (4.6)

Table 4.7: Regression Model (Industrial)- 2013

Intercepts and Variables	Coefficients	Standard Error	t statistics	P-Value
Intercept	16.68874	15.12292	1.103539	0.02783*
Area of industrial building/ km ²	4935.869	449.1363	10.98969	0.00000***
R²	0.795748			
F-Test	120.7733			
Observations	33			

$p < 0.05^*$, $p < 0.01^{**}$, $p < 0.001^{***}$

Table 4.8 shows the final attracted trips to the general purpose buildings for

2013 based on the developed model (Eq 4.5), and a total of 644,911 trips were attracted to the study area. As per the results, Colombo DSD is a highly trip-attracted zone for general purposes as it is the busiest commercial DSD in Western Province.

Table 4.8: Trip Attraction (General Purpose Buildings 2013)

DSD	Developed Area (Supervised Classification)	General Purpose Area%	Area of GP	Area TAZ	Total General Purpose area/ TAZ area	TA (General Purpose)
Agalawatta	38.03	0.20	0.18	89.78	0.00	2,789
Attanagalla	93.55	0.70	1.08	154.30	0.01	11,453
Bandaragama	37.75	1.00	0.57	57.40	0.01	5,747
Beruwala	32.62	2.00	1.44	72.03	0.02	13,431
Biyagama	58.24	2.40	1.45	60.27	0.02	13,318
Bulathsinhala	7.72	0.05	0.10	209.48	0.00	3,794
Colombo	21.40	20.0	4.16	20.82	0.20	116,036
Dehiwala/ Mt. Lavinia	7.70	7.00	0.59	8.40	0.07	14,526
Divulapitiya	110.41	0.20	0.41	205.23	0.00	6,374
Dodangoda	66.03	0.20	0.23	112.82	0.00	3,504
Dompe	118.96	0.43	0.78	182.16	0.00	9,275
Gampaha	76.16	2.00	1.81	90.69	0.02	16,910
Hanwella	95.40	0.70	1.02	145.88	0.01	10,828
Homagama	95.90	2.00	2.38	119.03	0.02	22,195
Horana	91.62	1.50	1.69	112.78	0.02	16,161
Ingiriya	59.41	0.20	0.19	94.05	0.00	2,921
JaEla	52.08	4.00	2.46	61.42	0.04	22,057
Kaduwell	77.50	3.00	2.63	87.75	0.03	23,937
Kalutara	40.75	3.50	2.72	77.68	0.04	24,544
Katana	83.74	3.00	3.27	109.00	0.03	29,734
Kelaniya	19.60	5.00	1.10	21.93	0.05	9,769
Kesbewa	41.00	2.00	1.03	51.45	0.02	9,593
Kolonnawa	22.90	8.00	2.08	26.04	0.08	18,344
Madurawala	42.34	0.20	0.13	62.92	0.00	1,954
Mahara	69.86	1.00	0.94	94.30	0.01	9,442
Maharagama	34.80	3.20	1.20	37.35	0.03	22,629
Mathugama	61.89	0.30	0.41	135.02	0.00	5,359
Millaniya	61.71	0.10	0.08	82.06	0.00	1,840
Minuwangoda	96.65	1.00	1.33	133.22	0.01	13,339
Mirigama	97.03	0.64	1.19	186.15	0.01	12,853
Moratuwa	14.30	5.21	1.00	19.21	0.05	23,055
Negombo	23.05	4.00	1.85	46.14	0.04	16,570
Padukka	67.40	0.30	0.31	104.96	0.00	4,166
Palindanuwara	99.53	0.00	0.00	283.23	0.00	3,907
Panadura	21.83	2.80	1.26	45.03	0.03	21,956
Rathmalana	10.00	7.00	0.92	13.15	0.07	22,742
Sri J'pura Kotte	14.00	7.00	1.16	16.52	0.07	28,571
Thimbirigasyaya	19.23	12.0	0.60	22.40	0.03	20,650
Walallawita	39.72	0.00	0.00	213.02	0.00	2,938
Wattala	38.93	5.00	2.88	57.69	0.05	25,698
Total						644,911

Figure 4.11 shows the validation of the above model with the total trip attraction values retrieved from HVS 2013 for the general-purpose buildings. The model yielded an overall actual vs prediction accuracy of 94.48%. The relationship between HVS 2013 trip attraction and predicted value based on the model developed is shown below (Eq 4.7).

$$TA (GP)_{HVS2013} = 1.3319 * TA (GP)_{model} - 4234.8 \quad (4.7)$$

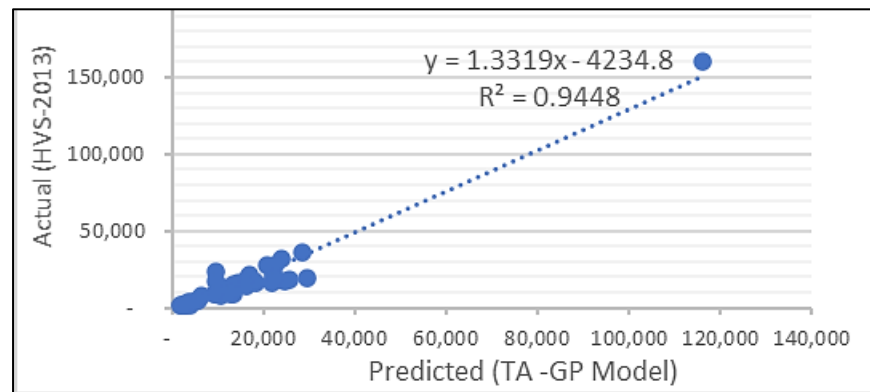


Figure 4.11: Trip Attraction (General Purpose) Actual vs Predicted - 2013

Similarly, the industrial area is calculated based on the land use survey data of 2013 from the supervised classification developed area. The area of industry and trip attracted to the industrial area using the developed model (Eq 4.6) is shown in table 4.9. A total number of 270,816 trips were attracted to the industrial area in 2013. As per the results, Biyagama DSD has the highest number of trip attractions for industrial purposes as Biyagama is one of the largest EPZ.

Table 4.9: Trip Attraction (Industrial)- 2013

DSD	Developed Area	Ind%	Area of Industrial	Area TAZ	Total Ind area/ TAZ area	TA (Industrial)
Agalawatta	38.029	1	0.38029	89.78	0.0042	3,375
Attanagalla	93.55	1.7	1.59035	154.3	0.0103	10,425
Bandaragama	37.75	1	0.3775	57.4	0.0066	2,821
Beruwala	32.62	1	0.3262	72.03	0.0045	2,812
Biyagama	58.24	8	4.6592	60.266	0.0773	24,003
Bulathsinhala	107.72	0.1	0.10772	209.48	0.0005	4,028
Colombo	21.4	12	2.568	20.82	0.1233	13,023
DehiwalaMt_Lavinia	17.7	1	0.177	8.399	0.0211	1,014
Divulapitiya	110.41	1.1	1.21451	205.23	0.0059	9,420
Dodangoda	66.03	0	0	112.82	0.0000	1,883
Dompe	118.96	0.7	0.83272	182.158	0.0046	7,150
Gampaha	76.16	0.6	0.45696	90.69	0.0050	3,769
Hanwellla	95.4	1.5	1.431	145.88	0.0098	9,498
Homagama	95.9	1.5	1.4385	119.033	0.0121	9,087
Horana	91.62	3	2.7486	112.78	0.0244	15,449
Ingiriya	59.41	1	0.5941	94.05	0.0063	4,502
JaEla	52.08	4.9	2.55192	61.42	0.0415	13,621
Kaduwela	77.5	1.5	1.1625	87.75	0.0132	7,202
Kalutara	40.75	0.9	0.36675	77.68	0.0047	3,107
Katana	83.74	4.9	4.10326	124.96	0.0328	22,339
Kelaniya	19.6	9.5	1.862	21.93	0.0849	9,557
Kesbawa	41	1.6	0.656	51.45	0.0128	4,097
Kolonnawa	22.9	5.4	1.2366	26.04	0.0475	6,538
Madurawala	42.34	1	0.4234	62.92	0.0067	3,140
Mahara	69.86	1	0.6986	94.3	0.0074	5,022
Maharagama	34.8	1.6	0.5568	37.35	0.0149	3,372
Mathugama	61.89	1	0.6189	135.02	0.0046	5,308
Millaniya	61.71	0.1	0.06171	82.06	0.0008	1,674
Minuwangoda	96.65	0.4	0.3866	133.22	0.0029	4,131
Mirigama	97.03	0.1	0.09703	186.15	0.0005	3,586
Moratuwa	14.3	5.7	0.8151	19.21	0.0424	4,344
Negombo	23.05	9.7	2.23585	46.14	0.0485	11,806
Padukka	67.4	0.2	0.1348	104.96	0.0013	2,417
Palindanuwara	99.53	0.1	0.09953	283.23	0.0004	5,218
Panadura	21.83	5.3	1.15699	45.03	0.0257	6,462
Rathmalana	10	15.6	1.56	13.15	0.1186	7,919

Sri J'Pura Kotte	14	1.2	0.168	16.52	0.0102	1,105
Thimbirigasyaya	19.2	2	0.0961517	22.4	0.0043	3,023
Walallawita	39.72	0.1	0.03972	213.02	0.0002	3,751
Wattala	38.93	5	1.9465	57.69	0.0337	10,570
Total						270,816

Figure 4.12 shows the validation of the above model with the total trip attraction retrieved from HVS 2013 for the industrial buildings and the relationship is shown in Equation 4.8.

$$TA (Ind)_{HVS2013} = 0.9626 * TA (Ind)_{model} + 1266.5$$

(4.8)

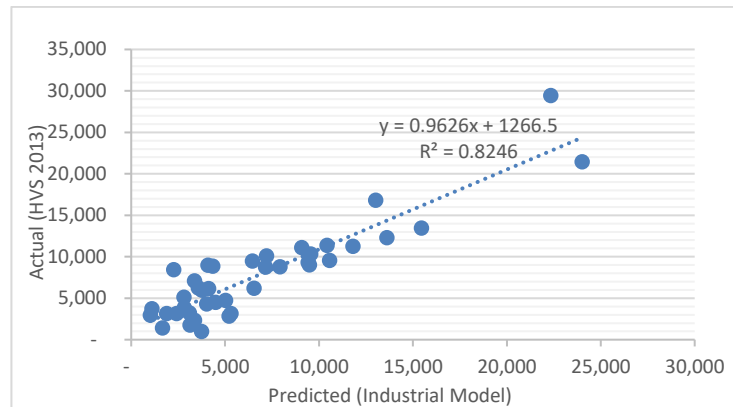


Figure 4.12: Trip Attraction (Industrial) Actual vs Predicted - 2013

All the points are within the diagonal line. So, the linearity assumption is satisfied for this model as well. The total number of trips attracted to the study in 2013 was 915,727 to the general purpose and industrial.

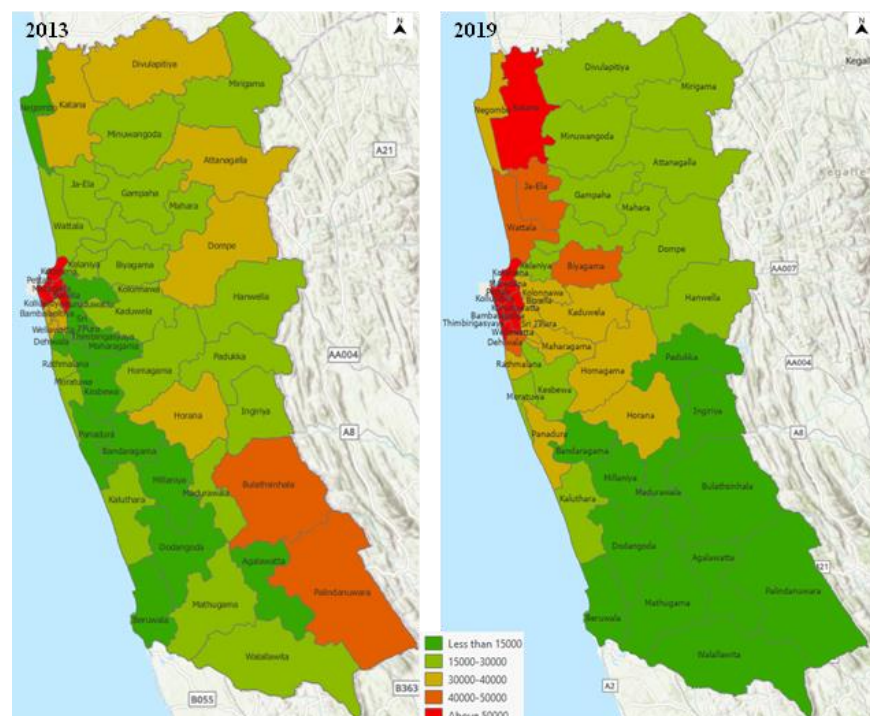


Figure 4.13: Summary of Trip Attraction between 2013 and 2019

Figure 4.13 shows that application of the developed model based on the 2013 data with the data of 2019. Based on that, nearly 1,177,741 trips were attracted to the Western Province in 2019.

4.8 Trip Distribution Model

There are several Trip Distribution models have been proposed based on different scenarios. One such famous model is the growth factor method. This model has been used to update the trip matrix or forecast the future trip matrix, using information related to future trip rates or growth factors. Another most used method is the gravity model.

4.8.1 Growth Factor Methods

This method has been applied where the fundamental trip matrix was obtained from a previous study or estimated from the survey data. There are several ways to update the growth factor based on the information. There are three types of growth factor models. The first one is the uniform growth factor model derived from the general growth rate r , for the whole study area. Equation 4.9 is, as always,

$$T_{ij} = r \cdot t_{ij} \text{ for each pair } i \text{ and } j$$

(4.9)

Where r = the ratio of expanded over the previous total no of trips.

The second one has singly constrained growth factor methods based on where the expected growth rate of the trip is available only for originating or attracted zones. This can be written as (Eq 4.10),

$$T_{ij} = r_i \cdot t_{ij} \quad \text{where } r_i \text{- origin specific growth factors}$$

(4.10)

Or

$$T_{ij} = r_j \cdot t_{ij} \quad \text{where } r_j \text{- Destination-specific growth factors}$$

The third method is doubly constrained growth factors. This is based on where the growth factor is available on the future number of trips originating and terminating each zone. In transport, this method is known as the Fratar method. This is the most widely used growth factor technique. This model can be represented mathematically as follows (Eq 4.11).

$$T_{ij (k+1)} = (T_{ij} \cdot F_{jk}) \cdot F_{ik}$$

(4.11)

Where,

$$F_{jk} = A_j / \sum_{i=1}^n T_{ijk} \quad F_{ik} = P_i / \sum_{j=1}^n (T_{ijk} \cdot F_{jk})$$

Where,

T_{ijk} = trips between i and j at iteration k

F_{jk} = Destination factor j

F_{ik} = Origin factor i

P_i = trip production of zone i

P_j = trip attraction of zone j

4.8.2 Synthetic or Gravity Models

The gravity model was the first trip distribution model approach. This theory was based on observable physical distribution phenomena. This approach aggregates destination choices rather than based on behavioural assumptions. Physical models can be derived using three approaches. The first one is based on Newton's law of gravity, called the gravity model. Entropy is the second approach. It is based on the analogy between particle behaviour in a system and the laws of statistical mechanics. The entropy concept is also related to information theory. The third is the network theoretic approach based on the analogy with electrical systems theory and optimal network flows.

4.8.2.1 Use of Gravity Model

Many external factors, such as total travel time or distance, can influence trip-making behaviour. Different trip distribution model has been developed to capture the important changes in the network to estimate future trips. The best-known such model is the gravity model, which is based on Newton's gravity model. This theory became popular in early 1955. The most straightforward formulation of the gravity model as applied in urban traffic distribution analysis is (Eq 4.12),

$$T_{ij} = \frac{\alpha P_i P_j}{d_{ij}^2}$$

(4.12)

P_i and P_j are the populations of origin (i) and destination (j), d_{ij} is the distance between i and j, and α is a proportionality factor. This model was soon considered too simple and later considered total trips instead of total populations. The model was further generalized by considering several factors that can affect the trip-making pattern instead of distance. This can be written as (Eq 4.13).

$$T_{ij} = \alpha O_i D_j f(c_{ij}) \quad (4.13)$$

Where $f(c_{ij})$ is a generalized function of the travel costs related to the trip.

4.8.3 Entropy Maximisation Model

In 1970, Wilson introduced the entropy model for trip distributions. In this concept, equilibrium refers to a physical system that is in equilibrium. Systems are described by their aggregate behaviour. As an alternative to the continuum approximation, continuous trip production and attractions functions may be used for densely populated urban areas. This section considers incremental opportunities at radial distances around the origin. As a result, the final model can also be modelled as a gravity-type model. In 1996, Jeppe Husted Rich developed a combined relationship between mode split and trip length. A simultaneous model set-up is presented to consider both mode-splitting and continuous trip length in a joint decision. In order to split modes and choose destinations, the well-known joint Logit model approach was used.

4.9 Application of Trip Distribution Model

The structure of the general trip distribution model is used gravity model with an explanatory variable of distance (d_{ij}) between OD pairs. The general format of that gravity model applied for this study is below (Eq 4.14).

$$T_{ij} = K \cdot \frac{P_i^\alpha A_j^\beta}{d_{ij}^\gamma} \quad (4.14)$$

Where:

T_{ij}	: Trip distribution from zone i to zone j
P_i	: Trip production of zone i
A_j	: Trip attraction of zone j
d_{ij}	: Impedance between zone i and zone j
α, β, γ	: Model parameters

The direct zone distance of OD pair is considered and calibrated using Arc GIS Pro (Appendix 1).

4.9.1 Calibration of Model Parameters based on CoMTrans Study 2013

The study considered a larger zone level to model trip distribution and selected 119 medium zones. However, the study has observed that the parameter for CMA is not improved and provided an incorrect value for trip distribution. The study considered the model development according to the selected zones based on the socioeconomic and population framework (Figure 4.14).

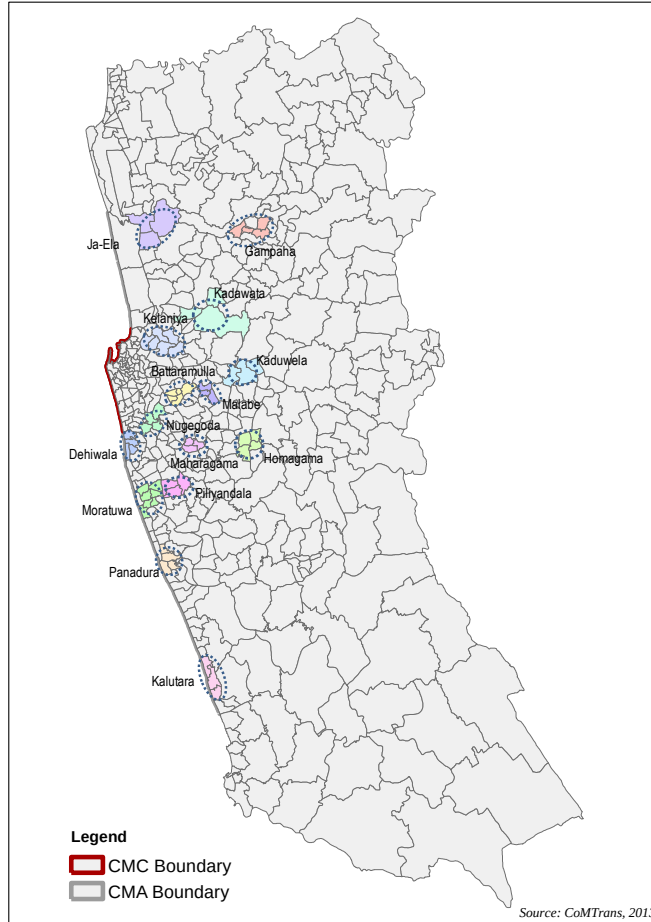


Figure 4.14: Selected TAZs for Gravity Model Application

The output of that model gives a promising result and significantly improves the performance of the models. Below values were calculated using the above methodology for home-to-work trips.

$$\begin{aligned}\alpha &= 0.461^* \\ \beta &= 0.461^* \\ Y &= 0.214^* \\ \text{And} \\ K &= 0.102^*\end{aligned}$$

Using the above values, the trip distribution for 2013 is calibrated and provided in appendix 2. The trip distribution model was validated with actual trips calibrated from the CoMTrans study HVS 2013. It is also recorded that the total number of trips recorded in 2013 based on the above value is 1,237,693 and based on the CoMTrans HVS study, it was recorded as 1,350,000 trips. The accuracy of this calculation shows nearly 91.6% of the actual value from the study.

4.9.2 Model Assessment

Figure 4.15 shows the distance vs cumulative % trips calculated based on the trip distribution of 2013. More than 85% of work-related trips end within the 25 km radius of the study area. It also can be seen that a sharp drop beyond the 25 km length of trips. Furthermore, the comparison of observed and synthesized work trip length frequency distribution is very close.

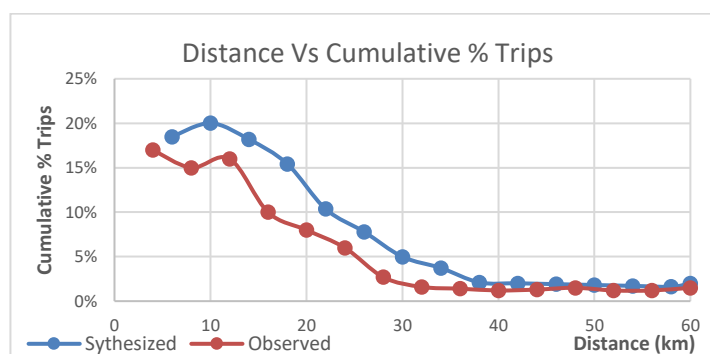


Figure 4.15: Distance Vs Cumulative % Trips

4.10 Sensitivity Analysis

Performing a sensitivity analysis reveals how much model outputs are affected by relatively small parameter changes. Sensitivity analysis also ensures the model response is reasonable and identifies the critical parameter affecting the trip generation and attraction most. A sensitivity analysis is used to determine what effect changes will be caused by if the independent variables increase or decrease in the future scenario.

A sensitivity test for the trip generation is performed based on the two different cases

1. 10% increment to the HHS
2. Based on the electricity consumption increase rate of Western Province

The annual electricity consumption (units in GWh) of households for Western Province is shown below (figure 4.16), and the annual growth rate is calculated based on that data. The annual growth rate of electricity consumption in Western Province is 5.3%.

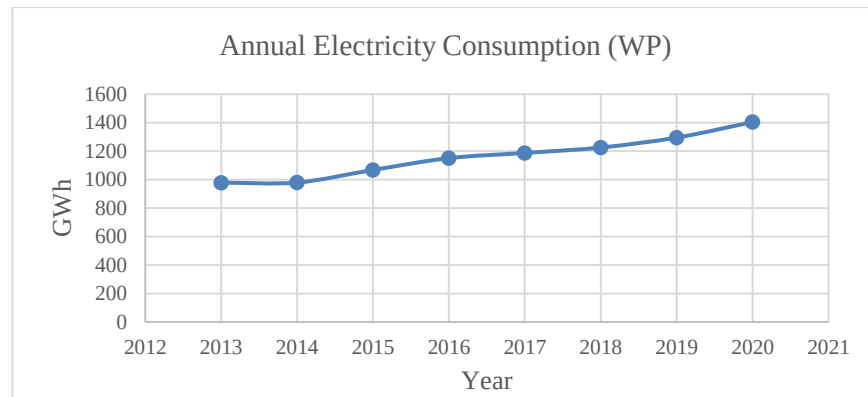


Figure 4.16: Annual Electricity Consumption (Western Province)

Based on the above two cases, the sensitivity analysis results are shown below (Table 4.10).

Table 4.10: Annual Electricity Consumption (Western Province)

Variation	TG Home-based work
Households (+10%)	+9.88%
Electricity Unit (+5.3% per year)	+0.51%

The increases or decreases of percentage in variable households significantly impact trip generation compared to electricity consumption. Suppose the variable of households increases by 10%. In that case, the trip generations of home-based work increased by 9.88%, which again validates the importance of the land use pattern of a zone concerning the residential has a substantial impact on generated trips of that zone.

Similarly, sensitivity for the trip attraction of general purpose and industrial land use has been tested based on the below case.

1. 10% increase in the general purpose area
2. 10% increase in the industrial area

Based on the above cases, the sensitivity results are shown in Table 4.11.

Table 4.11: Sensitivity Test for Trip Attraction

Variation	TA (Home based work trips)
General Purpose (+10%)	+9.57%
Industrial (+10%)	+7.90%

Results show that an increase or decrease in the land area for general purpose building and industrial significantly impacts trip attraction. However, general-purpose buildings play a significant role in attracting work-related trips compared to industrial purposes.

4.11 Summary and Conclusion

The importance of exploring household data as explanatory variables for the travel demand model is emphasized here, given the availability of long-range time series data that can be used to update the model frequently. However, household income can also provide a strong relationship between trip generation, and this study has shown a relationship between electricity consumer units and income using 2013 survey data. Several studies were conducted to validate the relationship between income and electricity consumption (Athukorala and Wilson 2010, Rajmohan and Weerahewa 2010, Kim 2020). This leads to a promising direction for ensuring the frequent update of trip generation with frequently available data on electricity consumption. Furthermore, consumer category information of electricity consumption data can also be used to develop trip generation models with the number of households in a particular zone. With these two independent variables, the developed trip generation model shows an R-squared of 79% with 92% accuracy from the CoMTrans survey 2013 between actual and predicted values from the model for 2013.

Electricity consumption data was used to develop trip attraction models for general and industrial areas separately. Due to the unavailability of land use survey data, satellite image data with supervised classification was used to obtain land use area information for 2013 and 2019. The developed models for general purpose and industrial trip attraction have an R-squared of 70% and 79.5% with 94% and 82% accuracy from the CoMTrans survey 2013 between actual and predicted values from the model for 2013, respectively. Finally, the gravity model was used to obtain the final trip distribution for 2013 and 2019 using the parameters considered in the CoMTrans study 2013.

The trip distribution output can be further developed into a four-step model by assigning trips to the transportation network, as described in Chapter 5. Additionally, Chapter 5 includes tests to assess the accuracy and reliability of the model using various scenarios based on land use and transportation.

CHAPTER 5

DEVELOPMENT OF THE LUTI MODEL WITH IDENTIFIED HFD AND SCENARIO TESTING

5.1 Introduction

The development of LUTI models has garnered significant attention over the years due to their ability to capture the complex interactions between land use and transport. In earlier chapters, the complexity of LUTI models and their data requirements were discussed along with the associated challenges and opportunities. This chapter details the development of the NRT-LUTI model through a three-step process.

The first step involved the development of the NRT-LUTI model using comprehensive data on travel behaviour, land-use patterns, and demographic characteristics. Key data affecting travel behaviours and land-use patterns were identified using proxy sources for the study area. A set of equations were developed to incorporate these data into the NRT-LUTI model and capture the relationship between travel behaviours and land-use patterns.

In the second step, the software used to capture the links between HF data and travel behaviour and land-use patterns was described in detail. The JICA STRADA software was employed to run each scenario, and key performance indicators such as passenger hours, passenger-km, travel time savings, overall system costs, and others were compared. This software also uses a four-step model, and trip assignment was the critical step to evaluate each scenario based on key performance indicators.

In the final step, the NRT-LUTI model was tested with various scenarios to evaluate the effects of policy interventions and changes in HFDs on travel behaviour and land-use patterns (Table 5.1). Overall, the chapter demonstrates the potential of the NRT-LUTI model to capture the complex interactions between land use and transport and provides a framework for evaluating policy interventions and changes in HFDs.

Table 5.1 Description of Scenario Cases

Scenario No	Year	Description	Origin Destination Matrix	Road Network
1	2013	Do Nothing	2013 OD developed based on electricity consumption data of 2013	2013 network developed for CoMTrans project

2	2019	Considering 2019 road network improvements and OD updated based on electricity consumption data	2019 OD developed based on electricity consumption data of 2019	2019 road network updated based on RDA data
3	2030	OD changed based on electricity consumption rate and 2019 road network	Update the OD for 2030 based on electricity consumption growth rate	2019 Road Network used in Scenario 2
4	2030	Land use improvement	2030 OD is modified based on the UDA development plans for 2030	2019 Road Network used in Scenario 2
5	2030	Land use and Transport changes	OD developed for scenario 4	Road network is updated from 2019 network based on development proposed for UOM study for 2030

Below are the details of various scenarios that were tested using the NRT-LUTI model presented in Table 5.1, which included changes in transportation infrastructure, land-use policies, and demographic factors.

- Scenario 1 involved a do-nothing approach for 2013, based on the 2013 road network and 2013 OD and parameters. This scenario was used as a validation year, and survey information from the CoMTrans study in 2013 was used to validate the accuracy of the model.
- Scenario 2 also involved a do-nothing approach, but for 2019. This scenario considered the 2019 updated network and the OD with updated parameters of 2019, such as the cost of travel time.
- Scenario 3 was designed to test a future scenario for 2030, where the OD was updated based on the electricity consumption rate for 2030.
- Scenario 4 was based on the future improvement plans proposed by the Urban Development Authority of Sri Lanka for 2030. The development percentages were proposed and tested based on the proposed development according to the residential, general purpose, and industrial plans.
- Scenario 5 was based on the future improvement plans proposed by the UOM study of the revised plan of CoMTrans 2013 for 2030 for the network with land use development OD.

Through the testing of these scenarios, the NRT-LUTI model demonstrated its effectiveness in capturing the complex interrelationships between transportation and land-use systems. The model can guide policy decisions on urban planning,

transportation, and environmental sustainability.

5.2 Methodology Used to Develop NRT-LUTI Model

Figure 5.1 illustrates the methodology framework for the NRT-LUTI model, which integrates the transport and land use models. The framework also identifies the key HF proxy data used in both models. The NRT-LUTI model, developed for this study, primarily uses electricity consumption data that includes GPS coordinates of customers and satellite data from Landsat. These data sources allow the model to capture the origin and destination of home-based work trips, as described in Chapter 4.

The NRT-LUTI model also considers land use changes, predicted by the model developed for trip generation and attraction. The integration of the transport and land use models in the the NRT-LUTI framework emphasizes the importance of considering the complex interaction between land use and transport.

The use of HF proxy data improves the accuracy and timeliness of the model outputs. The NRT-LUTI model framework demonstrates how the integration of these data sources can improve the understanding of urban transport and land use systems and inform policy decisions.

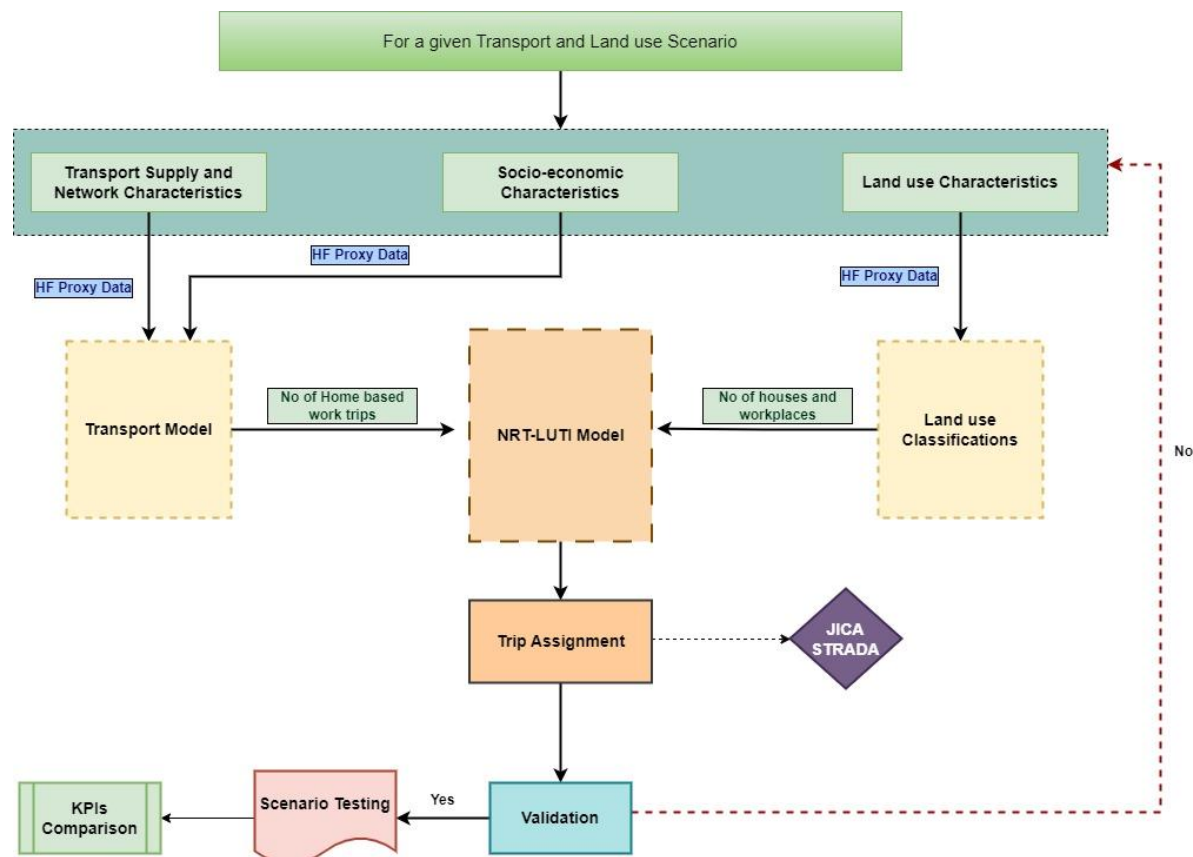


Figure 5.1: Overall Framework of NRT-LUTI model

5.2.1 Overview of Trip Assignment

The trip assignment is the final step in the four-step model that allocates trip interchanges between origin and destination based on the trip distribution matrix. It estimates the traffic volume at a specified route in the network, travel cost of inter-zones, and identifies congested road sections for future improvement. Each link in the network has a maximum capacity based on its road features. As trip flow increases and reaches the link's capacity, the average speed of that link reduces from the free flow and continues at the corresponding speed of maximum flow, leading to congestion. There are various models for assigning traffic, with the most commonly used being all-or-nothing, user equilibrium, and system optimum models.

The all-or-nothing assignment method assigns trips onto a single and minimum cost between any origin and destination zone. This model is suitable for uncongested networks where there are few alternative routes between specific origins and destinations, and the differences in the travel cost of those routes are high. However, a significant disadvantage of this model is that traffic on links is assigned without consideration of the link capacity, and the travel time is fixed and does not vary depending on congestion.

The user equilibrium (UE) assignment is a widely used trip assignment method based on Waldrop's first principle. It explains that UE does not allow drivers to switch routes to reduce their travel costs unilaterally. It presumes that each traveller seeks to minimize the time/cost associated with their chosen route. In this method, the travel time on each used path between the origin and destination is the same (Mamun, Siddique et al. 2015). Several assumptions are made for the UE assignment, and for a given O-D pair, the following equation can be written (Eq 5.1).

$$f_k(c_k - u) = 0: \forall k$$

(5.1)

$$c_k - u \geq 0: \forall k$$

where,

f_k = flow on the link k

c_k = travel cost of link k

u = minimum cost

Beckmann et al (1956) (Beckmann, McGuire et al. 1955) proposed a nonlinear

mathematical optimization program for the UE solutions (Eq 5.2).

$$\begin{aligned}
 \text{Minimize } Z &= \sum_a \int_0^{x_a} t_a(x_a) dx, \\
 \text{subject to } \sum_k f_k^{rs} &= q_{rs} : \forall r, s \\
 x_a &= \sum_r \sum_s \sum_k \delta_{a,k}^{rs} f_k^{rs} : \forall a \\
 f_k^{rs} &\geq 0 : \forall k, r, s \\
 x_a &\geq 0 : a \in A
 \end{aligned} \tag{5.2}$$

Where,

k =path

x_a = equilibrium in link a

t_a = travel time on link a

f_k^{rs} = flow on path k connection O-D pair r - s

q_{rs} = trip rate between r and s

$\delta_{a,k}^{rs}$ = definitional constraint where $\delta_{a,k}^{rs} = \begin{cases} 1 & \text{If link } a \text{ belongs to path } k, \\ 0 & \text{otherwise} \end{cases}$

System optimum assignment (SO) is a concept derived from Wardrop's second principle (Eq 5.3). The assignment can be viewed as a model in which drivers are instructed on routes to minimize congestion. Obviously, this is not a behaviourally realistic model, but it can be helpful to transport planners and engineers trying to optimize social equilibrium by minimizing travel costs.

$$\begin{aligned}
 \text{Minimize } Z &= \sum_a x_a t_a(x_a) \\
 \text{subject to } \sum_k f_k^{rs} &= q_{rs} : \forall r, s \\
 x_a &= \sum_r \sum_s \sum_k \delta_{a,k}^{rs} f_k^{rs} : \forall a \\
 f_k^{rs} &\geq 0 : \forall k, r, s \\
 x_a &\geq 0 : a \in A
 \end{aligned} \tag{5.3}$$

Where

x_a = equilibrium in link a

t_a = travel time on link a

f_k^{rs} = flow on path k connection O-D pair r - s

q_{rs} = trip rate between r and s

5.2.2 Overview of JICA STRADA

There are numerous computer-based simulation tools currently available for modelling, planning, and developing transport networks and systems. Due to the

difficulty of experimenting in a natural environment, it is more common to experiment with transport networks in a computer-based simulated model. Most models adopt the traditional four-step model, consisting of trip generation, trip distribution, mode choice model, and trip assignment. Recently, some models have been developed to adopt tour-based or activity-based modelling that can support microscopic simulation techniques at a disaggregated level.

For this traffic demand analysis, a four-step model was employed, which is a proven methodology with several applications in many countries. The System for Traffic Demand Analysis (STRADA) version 3.0 was used for trip assignment of present transport demand. This software is a cost-effective, comprehensive, and simple solution that has been used in several Master Plan projects in Sri Lanka. It is a window-based package developed by JICA in 1993 and consists of 17 individual modules, out of which 8 modules were used to build the private and public transport networks, trip assignments, and manipulate the assignment outputs through ArcGIS and STRADA. The software runs as two separate modules (Fig 5.2) called private or road assignment, which uses a user equilibrium module, and public passenger trips use transit assignment.

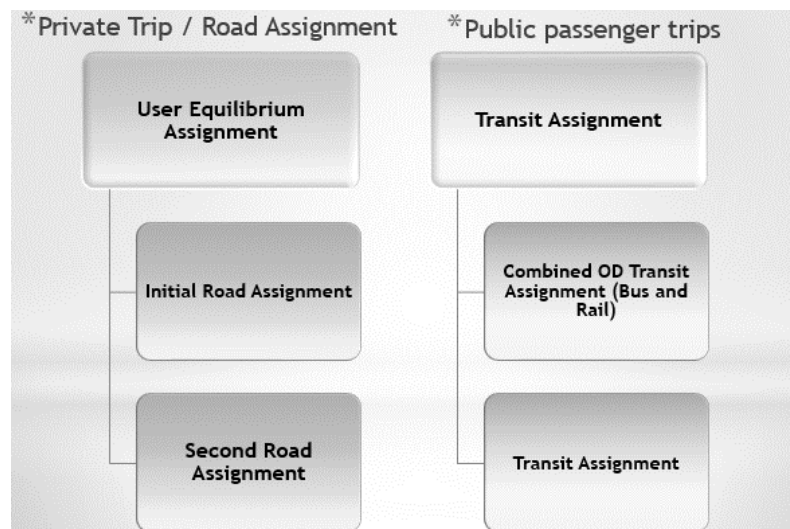


Figure 5.2: Flow Diagram for JICA STRADA Application

5.2.2.1 File Type of JICA STRADA- Input

5.2.2.1.1 Road Network Data (.INT)

The network data file contains various attributes of network links necessary to run the assignment programs. It contains,

- *Link and node names*
- *Link distance*

- *Link maximum capacity and velocity*

Link capacity is specified in PCUs. Maximum velocity is specified in kilometres per hour.

- *Link cost functions for velocity calculation (QV type)*
- *Colour code for each link*
- *Fare settings for each mode*
- *Directional control flag*

One-way, two-way and movement restrictions are controlled in directional control flag and defined separately for the above modes according to their numbers.

- *Road type flag*
- *Evaluation flag*

The evaluation flag excludes a given link from the calculation of evaluation indices.

- *User flags*

5.2.2.1.2 Assignment Parameters (.PAR)

This file contains the incremental assignment parameters and three user equilibrium assignment types. Users can specify various options for assignment.

- *Number of modes*

The user needs to specify the total number of modes in the network (INT).

- *Type of assignment*

Choose one of the four assignment types available, i.e., the incremental, the user equilibrium, the system optimum and the time-of-day user equilibrium.

- *Assignment option*

Users can select the following options for assignment. Click any number of accessible options in the tab.

- *Minimum route search by mode*

Select this option to do the minimum route search by mode. User must choose this option when you use the incremental assignment combined with the modal split method.

- *Impedances between zones*

Select this option to output the inter-zonal impedances into the

assignment results file.

- *Link traffic direction*

Select this option to separate the link traffic for each way in the assignment.

- *Minimum route information*

Select this option to output the minimum routes for all zone pairs.

- *Time Equivalence*

The time equivalence requires specifications per mode regarding the time value, the velocity factor, the passenger car unit, and the average passengers per vehicle in the Time Value tab. The modes displayed in the tab corresponding to the number specified in the General tab. User must always define the various parameters for each mode to run the assignment.

- *Base mode*

Specify the base mode by its mode number. If the minimum route search is not selected by mode in the General tab, the minimum route search will be performed by the maximum velocity of the base mode.

- *Time value*

Each mode's time value is specified and expressed in the minute/cost unit. Multiplied by this coefficient, the fare or toll is converted to a time value of fare resistance (i.e., delay) and added to the impedance in the minimum route search.

- *Velocity factor*

The velocity factor is used to adjust the maximum link velocity specified in the network data for each mode and expressed in positive value.

- *PCU equivalents and average passengers per vehicle*

The passenger car unit equivalents and the average passengers per vehicle convert OD trips by various modes relative to the road capacity. The following formula converts passenger trips recorded in each modal OD matrix to the traffic expressed in PCUs.

5.2.2.2 File Type of JICA STRADA- Outputs

5.2.2.2.1 Road Transport Assignment Results (.IRE)

This file type records various link results of road traffic assignment, including traffic volume, peak and average velocity, and traffic flow for both directions. The Volume and Capacity ratio (VCR) map has been developed for each

scenario based on the output information from JICA STRADA using Arc GIS Pro software.

5.2.3 Basic Assumptions and Inputs for the Model Development

This study exclusively uses the User Equilibrium Assignment to evaluate the scenarios. Figure 5.3 illustrates the methodology used for model development.

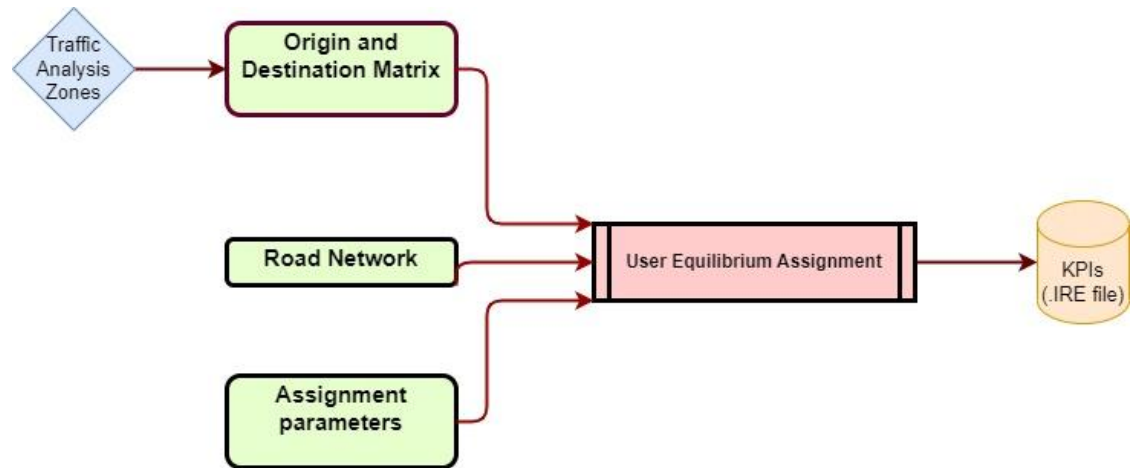


Figure 5.3: Overall Methodology of JICA STARDA Model Development

The origin and destination matrix includes home-based passenger trips that are assigned to the assignment based on TAZ. Road networks for the years 2013, 2019, and 2030 were developed using available information on relevant road parameters such as maximum capacity, velocity, and other restrictions like one-way and two-way. This information is provided in Annexures 7, 8 and 13 for 2013, 2019 and 2030 road networks respectively. The assignment parameters considered the time equivalence settings necessary to conduct the user equilibrium assignment. Furthermore, each input requirement is explained in detail below.

5.2.3.1 Travel Analysis Zones (TAZ)

The study area for this research covers the entire western province, which is divided into small zones called Traffic Analysis Zones (TAZs) based on the Divisional Secretary Divisions (DSDs) as the data related to population, land use etc can be easily retrieved from several sources. TAZs are created based on homogeneous characteristics such as population, land use, and accessibility, which can be easily acquired. In this study, a total of 40 TAZs were created in the study area (as shown in Figure 5.4). Each movement within the study area has a starting and ending point represented by a precise point known as the centroid.



Figure 5.4: Traffic Analysis Zones (TAZ) of Study Area (Western Province)

5.2.3.2 Time value (2013 and 2019)

The time value of 2013 is based on the values considered in the CoMTrans study of 2013 using below equation (5.4).

$$\text{Time value per workers} = (HHS_{\text{income}} + 0.3 HHS_{\text{income}}) / (\text{Avg Workers}/HHS * \text{Monthly working hours}) \quad (5.4)$$

The mean income of households was estimated using data from the Home Visit

Survey and the Household Income and Expenditure Surveys of 2012/2013. While averaged time values were used for economic analysis, worker's time values were used to forecast transport demand based on working hours and mean household income. Table 5.2 presents the values used for 2013.

Table 5.2: Time Value- 2013

Income Level	Mean Household Income (LKR.)	Avg. No. of Workers in Household	Monthly Working Hours	Social Security Cost	Worker's Time Value (LKR. /h)
Low	24,009	1.20	140	30%	186
Middle	56,810	1.72	140	30%	307
High	186,164	1.90	140	30%	909
All	70,366	1.36	140	30%	479

Source: CoMTrans, 2013

[Note Group C: Low income (less than 40,000 LKR), Group B: Middle Income (40,000-80,000 LKR), Group A: High income (Greater than 80,000 LKR)]

The mean household income was estimated based on the information from the Household Income and Expenditure Survey data of 2012/13. It was calculated for the Western Province by taking the percentage of the median income of Sri Lanka and the Western Province for the same year into account. In 2013, the median income of Sri Lanka and the Western Province was 30,400 LKR and 42,126 LKR, respectively, and an income distribution adjustment factor of 1.39 was applied. Based on this, the income distribution for 2013 is shown in Figure 5.5, with the same income range classified by the income and expenditure survey. The value of LKR 479 was applied for 2013 scenarios, considering the average value from all income groups.

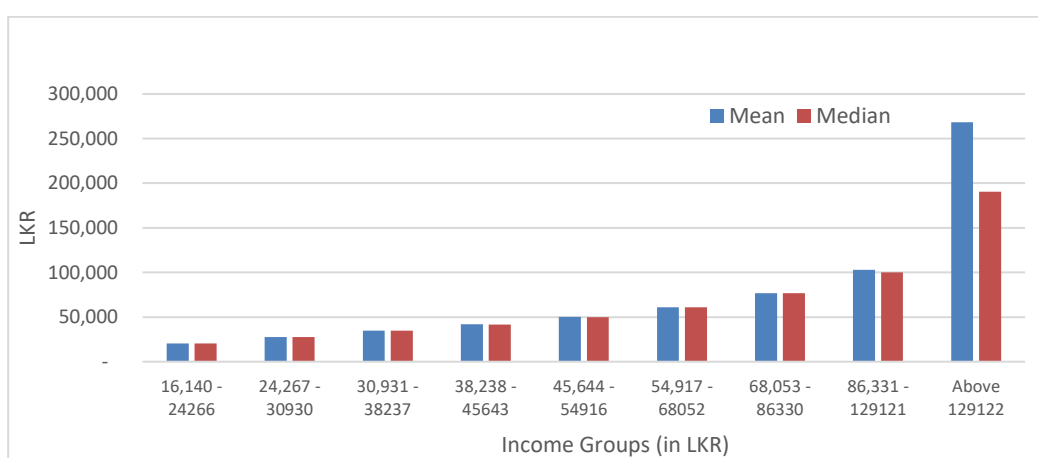


Figure 5.5: Household Income Distribution of Western Province (2013)

The 2019 time value was also calculated from the 2019 Income and Expenditure Survey, using the same methodology as for 2013. The median income of Sri

Lanka and Western Province in 2019 was 53,333 LKR and 75,000 LKR, respectively, with a Western Province income distribution adjustment factor of 1.41. Figure 5.6 shows the income distribution for 2019 with the income classification.

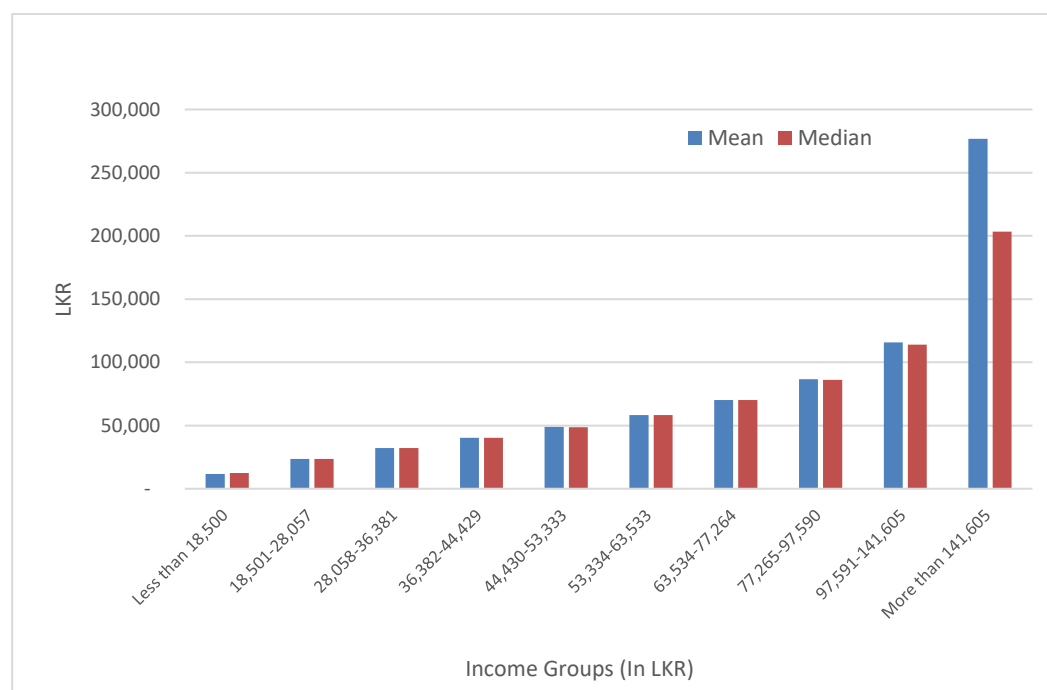


Figure 5.6: Household Income Distribution of Western Province (2019)

The time value for 2019 is presented in Table 5.3. As no household surveys were conducted after 2013, the average number of workers per household for 2019 was based on the 2013 value. The time value of 505 LKR per hour was applied for the 2019 scenario. The value calculated for 2019 was also used for the 2030 testing scenarios.

Table 5.3: Time Value - 2019

Income Level	Mean HHS Income (LKR.)	Avg. No. of Workers in HHS	Monthly Working Hour	Social Security Cost	Worker Time Value (LKR. /h)
Low	25,183	1.20	140	30%	195
Middle	65,635	1.72	140	30%	355
High	131,604	1.90	140	30%	643
All	74,141	1.36	140	30%	505

[Note Group C: Low income (less than 40,000 LKR), Group B: Middle Income (40,000-80,000 LKR), Group A: High income (Greater than 80,000 LKR)]

5.2.3.3 Road Network of 2013 and 2019

The CoMTrans study used the network for the base year (2013), which is included in Appendix 7 and contains detailed information. For the 2019 analysis, the road network was updated from the 2013 network, considering

improvement projects carried out by RDA. Appendix 3 provides a summary of the development completed by RDA between 2013 and 2019, and Figure 5.7 highlights the significant road improvements. Some road improvements were observed between 2013 and 2019, including the expansion of expressways.

In 2013, Western Province only had a southern expressway connecting its southern part to Kottawa, but by 2019, this expressway (Outer Circular Highway) had been completed up to Kadawatta, one of the main cities in Western Province (Kandy Corridor). Additionally, the Colombo-Katunayake Expressway was constructed to connect Katunayake airport to Colombo, benefiting the country's economy significantly. Furthermore, some road reconstruction projects were completed, such as the Battaramulla-Pannipitiya (B047), Ethul Kotte-Mirihana-Kohuwela (B120), and Colombo-Horana (B084) road sections. The Wellawatta area also implemented a one-way system.

Based on these improvements, the final road network used for 2019 is shown in Appendix 8. The values of other road parameters such as maximum velocity and capacity were calculated based on the CoMTrans study in 2013 (Appendix 8). However, due to the higher number of on-street parking and manoeuvring vehicles inside the Colombo Municipal Council (CMC) area, the CoMTrans study reduces 30% of the total maximum capacity.

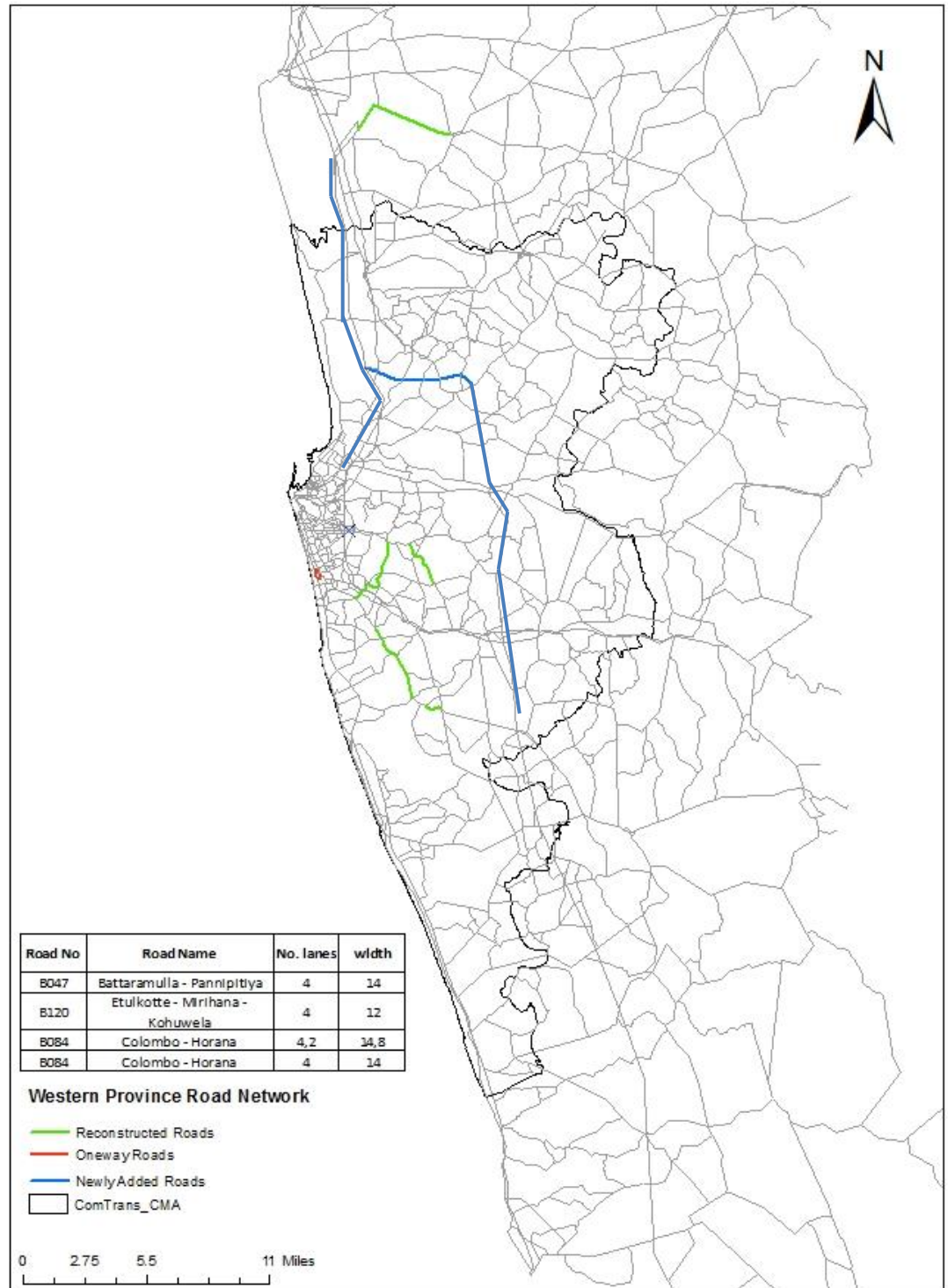


Figure 5.7: Road Projects Considered between 2013 and 2019 in Western Province

5.2.4 Origin and Destination Matrix of 2013 and 2019

The development of the 2013 and 2019 base year origin and destination matrix was discussed in Chapter 4. The finalized trip distribution matrix based on the trip generation and attraction for 2019 is provided in Appendix 4. The desire line maps for 2013 and 2019 are shown in figure 5.8. It was observed that many

work trips are distributed between Colombo and Gampaha DSDs, and most of the trip's end/start in Colombo DSD. It was also noted that the trips between Colombo to northern part of the province is significantly improved due to the improvement in the road network such as expressway development.

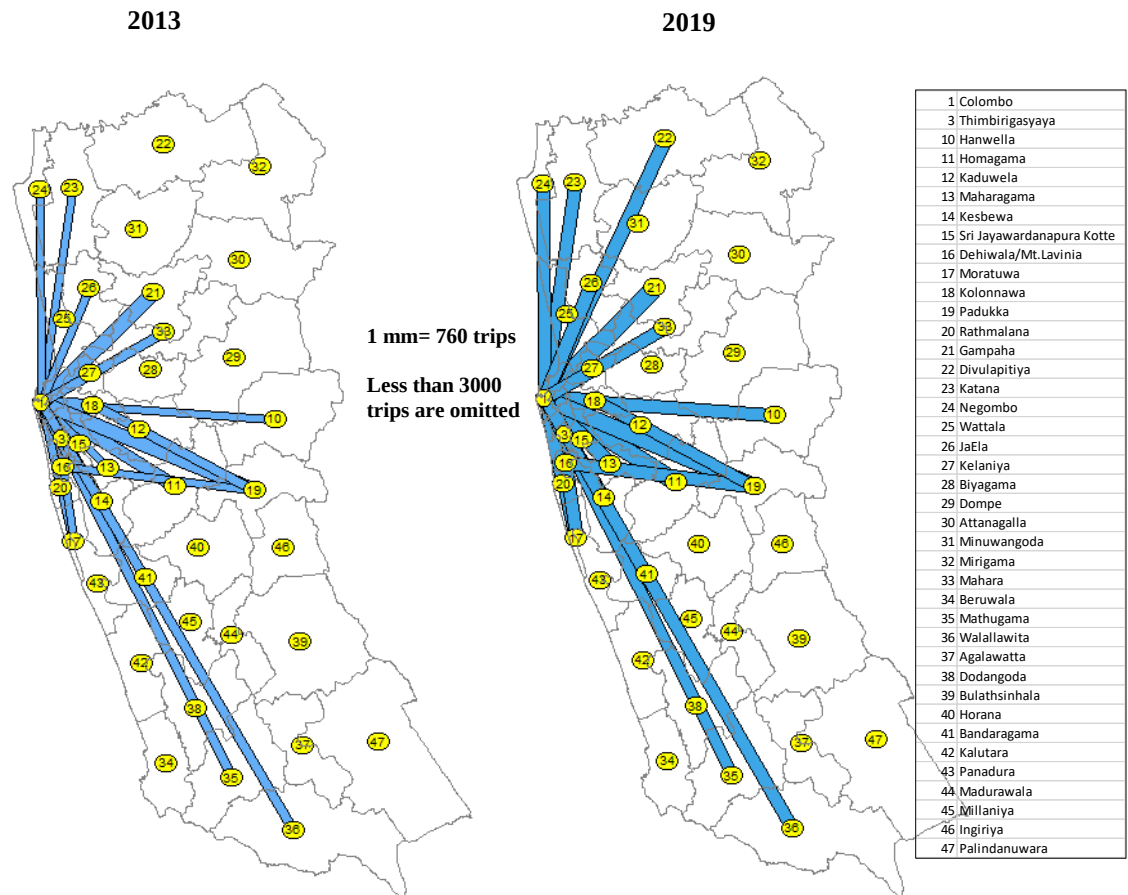


Figure 5.8: Desire Map of 2013 and 2019

5.3 Traffic Assignment Results and Validation of Scenario 1 (2013) and Scenario 2 (2019)

The assignment results of 2013 were verified using CVC survey count and screen line survey (SLS) counts for the CoMTrans study. The comparison was conducted at the CMC boundary central corridors point (Figure 5.9), and road segments for the selected location were identified. The line segment was selected from the JICA STRADA network using ArcGIS. The comparison was made for the total number of work-related passenger trips for the selected locations and line segments (Table 5.4).

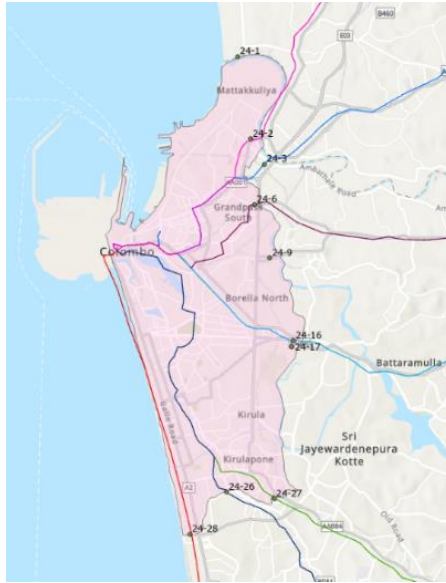


Figure 5.9: Survey Points Considered for the Comparison at CMC Boundary

According to Table 5.4, the total estimated number of passenger trips on all Sri Lankan roads in 2013 was 291,192, and the SLS count was 298,256. Furthermore, it was observed that road passenger trips per day from SLS and STRADA outputs were underestimated, such as A1 (Kandy) with 77%, and Road B84 (Horana) with 128% of the SLS passenger trips per day. On the other hand, Road AB10 (Low Level) and A4 (High Level) had only 129% and 128%, respectively, of the target passenger trips per day, which was higher than the target SLS counts. The validation results indicate that the estimated passenger trips per day for all Western Province major roads were quite accurate, with an average validation rate of 98%.

Table 5.4: Comparison of 2013 Assignment Output with Survey Data

Road No	Road Name	2013 Scenario 1 no of Pax trips/day	SLS 2013	Validation
A3	Negombo	58,263	50,200	116%
A1	Kandy	65,004	83,948	77%
AB10	Low Level Rd.	7,787	6,012	129%
A0	Malabe	28,621	28,289	101%
A4	High Level	36,312	28,330	128%
B84	Horana	30,394	38,977	78%
A2	Galle	64,811	62,499	104%
Total		291,192	298,256	98%

The SLS count conducted in 2013 was representative of 26% of the total counts, which is based on the home-based work trip share in Western Province. The distribution of overall home-based work trips at the major corridors of the CMC boundary has an accuracy rate of 99% when compared with the SLS data. The VCR map for 2013 is displayed in Figure 5.10.

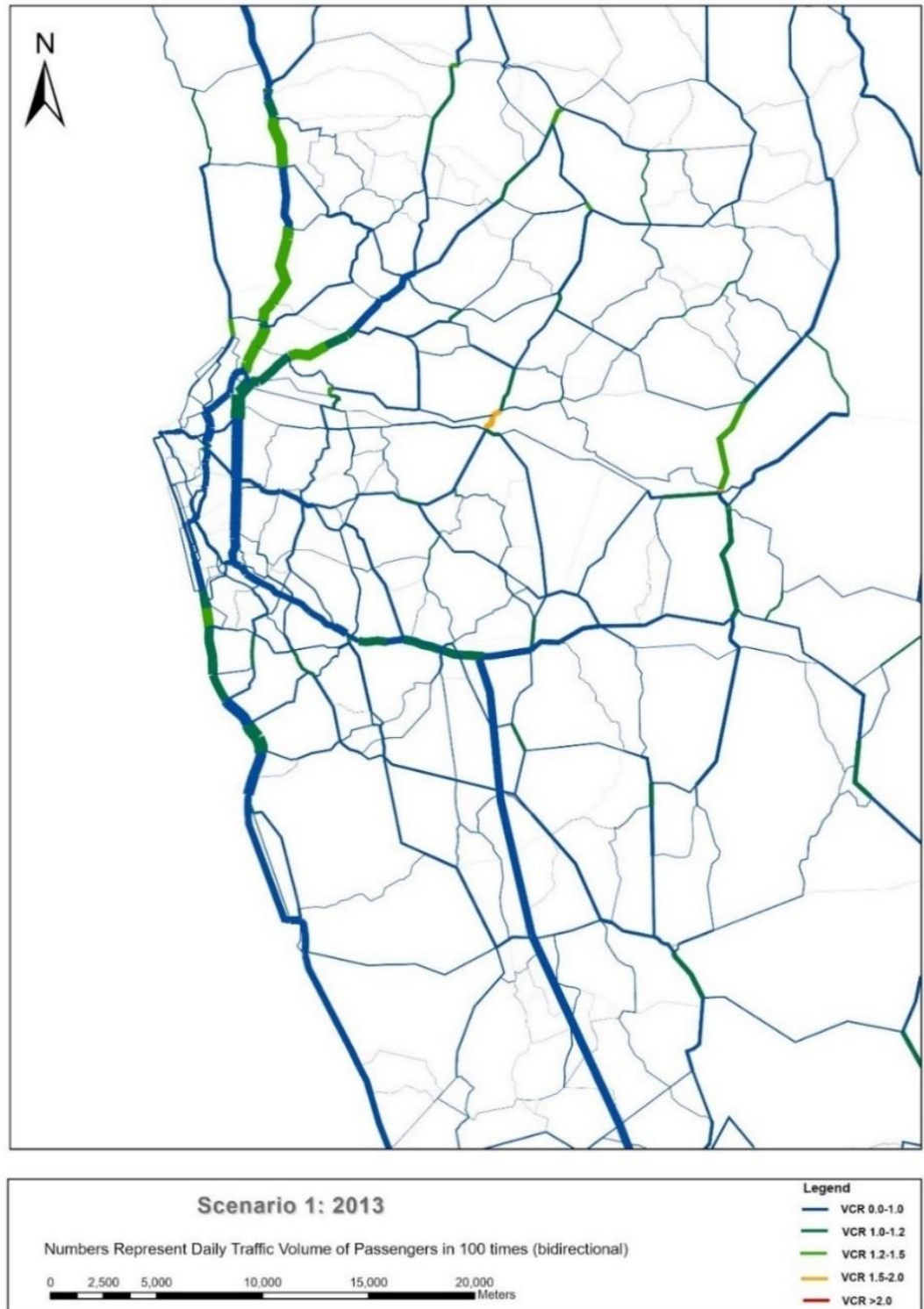


Figure 5.10: Volume/ Capacity Ratio (VCR) of Scenario 1 (2013)

The map in Figure 5.11 depicts the assignment results of the VCR map for 2019, which indicates that the majority of the major corridors that connect CMC and expressways are congested.

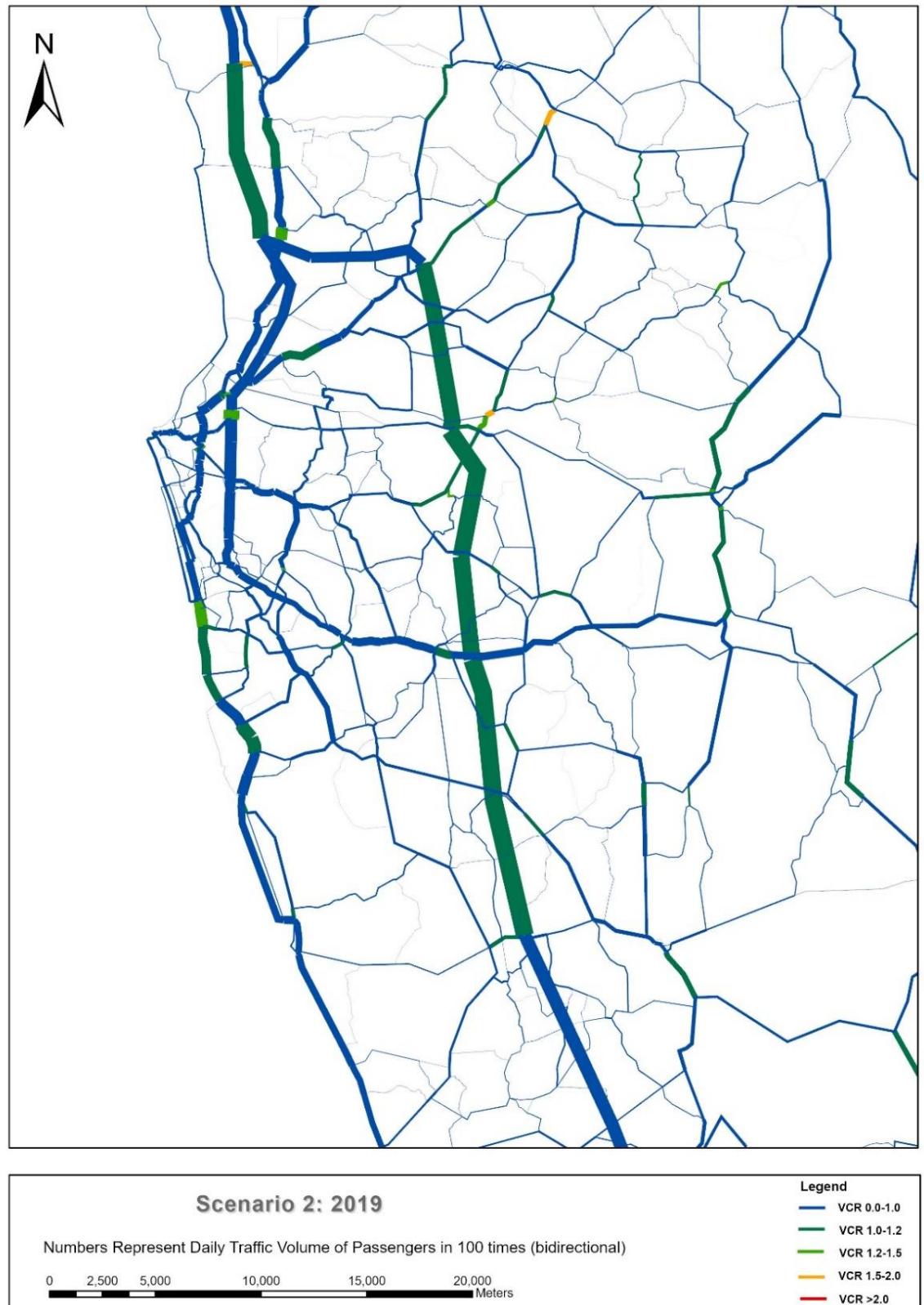


Figure 5.11: Volume/ Capacity Ratio of Scenario 2 (2019)

5.4 Scenario Testing for LUTI Integration

5.4.1 Overview

Three scenarios were formulated to assess the impact of land use and transport on the CMC area in 2030.

- I. Scenario 3: 2030 - do nothing
- II. Scenario 4: 2030 - with land use improvement
- III. Scenario 5: 2030 - with land use and transport development

Scenario 3 is a baseline scenario where the road network is based on the 2019 network, and the Origin-Destination (OD) matrix is estimated using the growth rate of GPS points for households, general purposes, and industrial purposes. The growth rate was determined using GPS customer points between 2013 and 2019.

Scenario 4 was tested with the land use improvement plan proposed by the Urban Development Authority (UDA) of Sri Lanka. The UDA has developed City Development Plans (2022-2030) with planning and development regulations, development zones, and zoning regulations as a strategic mechanism to achieve their objectives. Based on the development plan, proposed percentages of land use improvements for residential, commercial, and industrial areas were determined. Due to a lack of detailed information on each land use development, the percentages were determined from the proposed development map and the area considered for the development.

Scenario 5 was tested with the transport development plan proposed by the UOM study. The study team developed three cases (2020, 2025, and 2035) to assess the economic advantages and disadvantages of each development scenario. The transport impact was tested by considering the 2025 developments, and the Electronic Road Pricing (ERP) system was applied at each major corridor's point crossing the CMC boundary to manage traffic demand.

5.4.2 Scenario 3: Do Nothing (2030)

According to the GPS data, Scenario 3 OD has been updated according to the growth rate calculated from electricity GPS points data.

5.4.2.1 Development of Origin-Destination Matrix for 2030 (Scenario 3)

Trip generation for 2030 is calculated from the annual growth rate of households for each DSD and shown in Table 5.5. A total number of 1.76 million trips will be generated from Western Province in 2030.

Table 5.5: Trip Generation 2030

DSDs	GF/Year	HHS GPS (2030)	Average Electricity Unit (2019)	TG 2030
Agalawatta	3.34%	13,913	61.2	11,892
Attanagalla	2.27%	59,853	76.3	50,129
Bandaragama	2.87%	46,025	82.6	38,647
Beruwala	2.59%	48,049	73.6	40,307
Biyagama	2.69%	63,381	94.6	53,109
Bulathsinhala	3.27%	25,134	53.4	21,202
Colombo	3.67%	88,471	133.1	74,068
Dehiwala	2.09%	35,193	146.5	29,801
Divulapitiya	2.72%	54,865	66.2	45,956
Dodangoda	3.18%	25,307	68.1	21,383
Dompe	2.80%	58,600	70.1	49,071
Gampaha	2.53%	74,371	88.3	62,231
Hanwella	2.50%	42,993	65.0	36,081
Homagama	3.09%	103,117	87.9	86,132
Horana	3.25%	51,086	74.4	42,834
Ingiriya	3.03%	21,593	59.2	18,272
JaEla	2.30%	72,943	98.3	61,069
Kaduwela	2.39%	105,060	113.3	87,812
Kalutara	2.47%	58,485	72.3	48,981
Katana	2.97%	79,539	89.1	66,529
Kelaniya	2.50%	44,527	136.7	37,538
Kesbawa	2.46%	97,024	108.2	81,117
Kolonnawa	3.53%	79,004	117.6	66,156
Madurawala	2.94%	13,868	60.9	11,854
Mahara	2.54%	69,159	83.9	57,886
Maharagama	2.30%	68,273	105.8	57,204
Mathugama	3.60%	36,776	67.8	30,919
Millaniya	2.86%	21,710	63.0	18,380
Minuwangoda	2.74%	68,386	77.4	57,226
Mirigama	2.50%	58,619	68.5	49,083
Moratuwa	1.76%	39,882	85.9	33,547
Negombo	2.14%	44,804	94.8	37,662
Padukka	2.67%	26,348	66.1	22,244
Palindanuwara	4.23%	19,305	53.2	16,355
Panadura	2.22%	63,530	99.2	53,244
Rathmalana	1.78%	33,407	127.3	28,268
Sri Jayawardanapura	1.82%	35,090	124.5	29,660
Thimbirigasyaya	1.99%	73,534	223.3	61,875
Walallawita	3.50%	20,264	57.2	17,162
Wattala	2.87%	56,860	102.9	47,707
Total				1,760,594

Based on projections, the total number of trips in the Western Province is expected to increase by approximately 28% by the year 2030 compared to 2019. According to the results, the Western Province is projected to have 1,760,594 households in 2030, with a targeted growth rate of 2.72% and an average electricity unit consumption of 133.1. This information is illustrated on the map shown in Figure 5.12, which compares trip generation in 2013, 2019, and 2030. The analysis reveals that the growth rate of trip generation is higher in the CMC area and other urban areas.

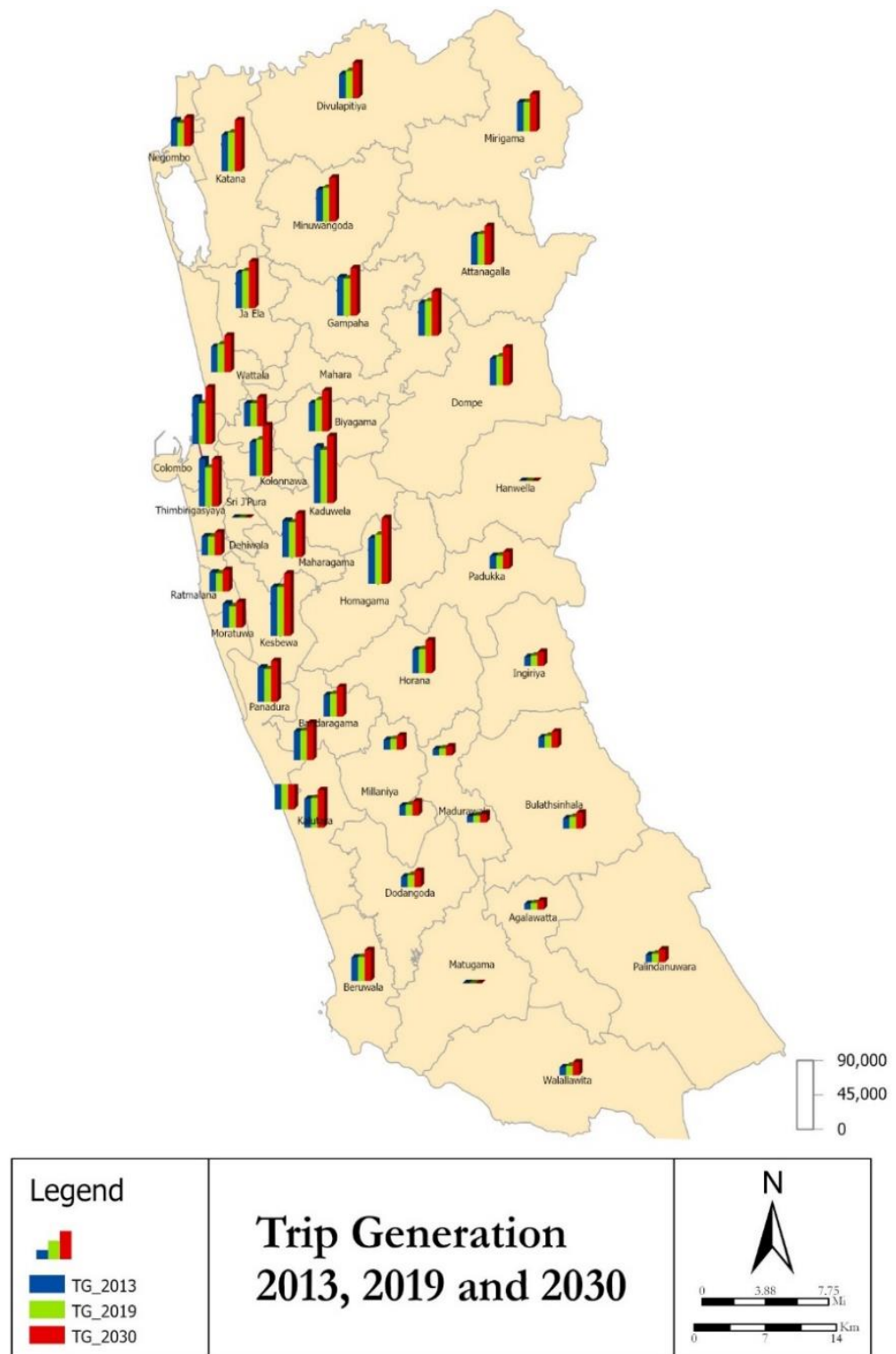


Figure 5.12: Trip Generation 2013, 2019 and 2030

Likewise, the trip attraction was calculated according to the annual growth rate calculated from the GPS point growth rate for general purpose and industry using 2013 and 2019 data (Table 5.6). A total number of 1.35 million trips will be attracted to the Western Province by 2030.

Table 5.6: Trip Attraction 2030

DSDs	Total (TAZ)	Annual GR (GP)	Area (GP) 2019	Area (GP) 2030	TA (GP) 2030	Developed Area	Annual GR (Ind)	Area (Ind) 2019	Area (Ind) 2030	TA (Ind) 2030	Total TA 2030
Agalawatta	89.73	6.60%	0.21	0.36	4,358	7.96	-4.00%	0.24	0.14	2,172	6,530
Attanagalla	154.21	4.95%	1.26	1.95	18,923	107.44	3.55%	1.87	2.60	15,398	34,320
Bandaragama	57.36	8.43%	0.67	1.29	11,930	35.98	-0.43%	0.37	0.35	2,689	14,619
Beruwala	51.62	6.39%	1.20	2.05	18,422	34.02	3.29%	0.35	0.47	3,200	21,622
Biyagama	60.23	5.78%	1.69	2.76	24,655	49.47	-0.28%	4.05	3.92	20,363	45,019
Bulathsinhala	209.30	7.33%	0.12	0.22	4,791	91.17	0.00%	2.80	2.80	17,302	22,093
Colombo	16.72	3.77%	3.90	5.52	187,479	15.00	3.35%	1.84	2.52	12,718	200,197
Dehiwala	8.39	4.30%	0.68	1.01	32,108	7.98	-0.95%	0.08	0.07	501	32,609
Divulpitiya	205.12	5.86%	0.48	0.79	9,625	111.15	8.08%	1.93	3.65	21,443	31,068
Dodangoda	112.74	7.78%	0.26	0.49	5,770	60.60	-0.41%	-	-	1,881	7,652
Dompe	182.08	5.55%	0.91	1.47	15,214	126.87	4.41%	0.91	1.35	9,697	24,912
Gampaha	90.63	6.41%	2.11	3.61	32,384	65.86	3.90%	0.40	0.58	4,364	36,748
Hanwella	145.79	5.38%	1.19	1.90	18,375	103.07	-0.63%	1.58	1.47	9,697	28,071
Homagama	118.96	8.10%	2.78	5.25	46,970	90.06	-1.00%	1.38	1.23	8,054	55,024
Horana	112.70	9.28%	1.97	3.99	35,971	75.13	1.75%	2.31	2.75	15,451	51,423
Ingiriya	93.99	9.24%	0.22	0.44	5,114	58.57	3.03%	1.80	2.40	13,397	18,511
JaEla	61.37	6.51%	2.86	4.92	43,280	49.07	7.16%	2.46	4.40	22,719	65,998
Kaduwela	87.68	6.76%	3.07	5.35	47,419	73.42	0.00%	1.13	1.13	7,024	54,443
Kalutara	77.64	6.51%	3.17	5.44	48,051	54.70	1.35%	0.50	0.58	4,150	52,202
Katana	124.86	5.36%	4.37	6.95	61,712	79.76	6.99%	2.77	4.91	26,305	88,017
Kelaniya	21.91	5.71%	1.28	2.08	18,265	19.68	1.35%	1.47	1.69	8,699	26,964
Kesbewa	61.41	6.04%	1.43	2.39	21,439	48.47	1.70%	0.79	0.94	5,675	27,113
Kolonnawa	26.02	6.71%	2.43	4.22	36,799	23.43	0.88%	1.29	1.42	7,438	44,237
Madurawala	62.84	10.57%	0.15	0.32	3,605	33.10	2.86%	1.02	1.34	7,639	11,244
Mahara	94.24	6.89%	1.10	1.93	17,988	70.17	2.39%	0.72	0.91	6,045	24,034
Maharagama	37.32	5.50%	1.39	2.24	19,818	33.08	2.88%	0.54	0.71	4,141	23,959
Mathugama	113.46	8.92%	0.40	0.79	8,356	48.26	-4.39%	1.48	0.77	5,674	14,030
Millaniya	82.01	10.19%	0.10	0.20	2,884	41.73	2.55%	0.04	0.05	1,638	4,522
Minuwangoda	133.13	6.21%	1.55	2.62	24,415	84.27	4.41%	0.34	0.51	4,750	29,164
Mirigama	186.03	4.91%	1.39	2.14	21,036	131.80	3.93%	0.13	0.19	4,058	25,093
Moratuwa	19.19	5.23%	1.17	1.84	56,917	15.56	-2.98%	0.91	0.61	3,331	60,248
Negombo	46.13	4.72%	2.15	3.27	28,866	24.45	10.51%	0.18	0.38	2,633	31,499
Padukka	104.89	7.45%	0.37	0.67	7,215	64.31	-0.31%	0.13	0.13	2,378	9,593
Palindanuwara	282.55	7.60%	-	-	3,897	89.03	-9.03%	2.73	0.02	4,803	8,700
Panadura	44.99	5.80%	1.47	2.41	66,515	35.48	-0.52%	0.47	0.44	2,947	69,462
Rathmalana	13.14	4.34%	1.07	1.59	50,475	11.54	4.32%	0.90	1.32	6,752	57,228
Sri J'Pura	16.51	3.61%	1.35	1.88	59,627	15.19	13.21%	0.19	0.46	2,532	62,160
Thimbirigasya	22.38	5.66%	3.70	6.00	52,140		2.25%	0.51	0.63	3,501	55,641
Walallawita	59.92	7.36%	-	-	827	16.64	-6.02%	0.51	0.17	1,851	2,677
Wattala	57.66	5.99%	3.36	5.58	48,963	40.22	5.87%	2.06	3.39	17,673	66,636
Total											1,695,409

Figure 5.13 shows trip attractions for 2013, 2019 and 2030. The trip attraction to the Western Province was noted to increase by 44% from 2019. Moreover, Colombo DSD is identified as a significant trip attractor, and CMC attracts 17% of total trips.

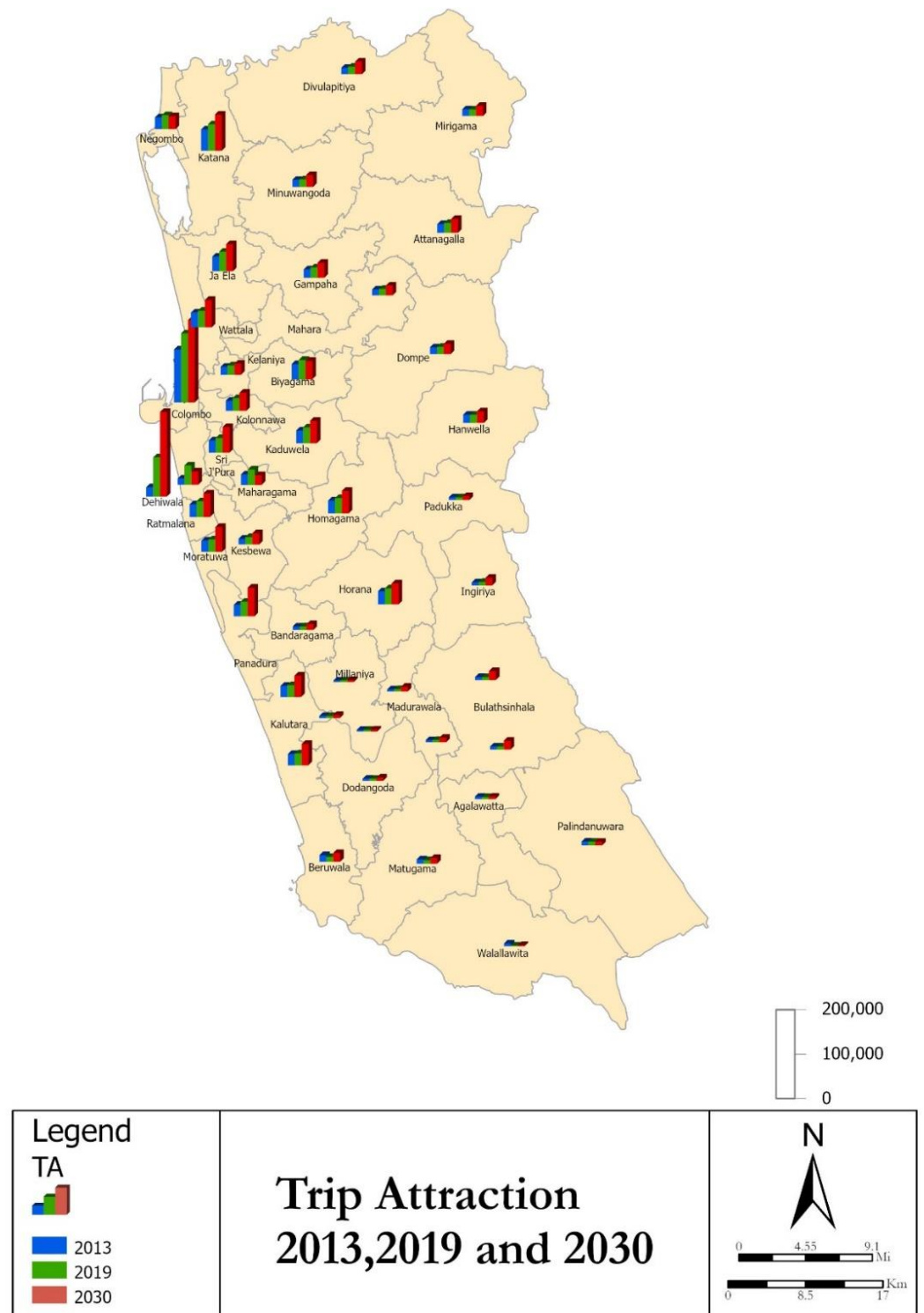


Figure 5.13: Trip Attraction 2013, 2019 and 2030

With trip generation and attraction development, trip distribution for 2030 was developed using the gravity model with the same parameters considered for 2013 and 2019. A total of 2,131,154 trips will be exchanged in Western Province, Sri Lanka (Appendix 5).

5.4.2.2 Estimated Results of Scenario 3

The volume/capacity map for the scenario 3 for 2030 is included below (Figure 5.14).

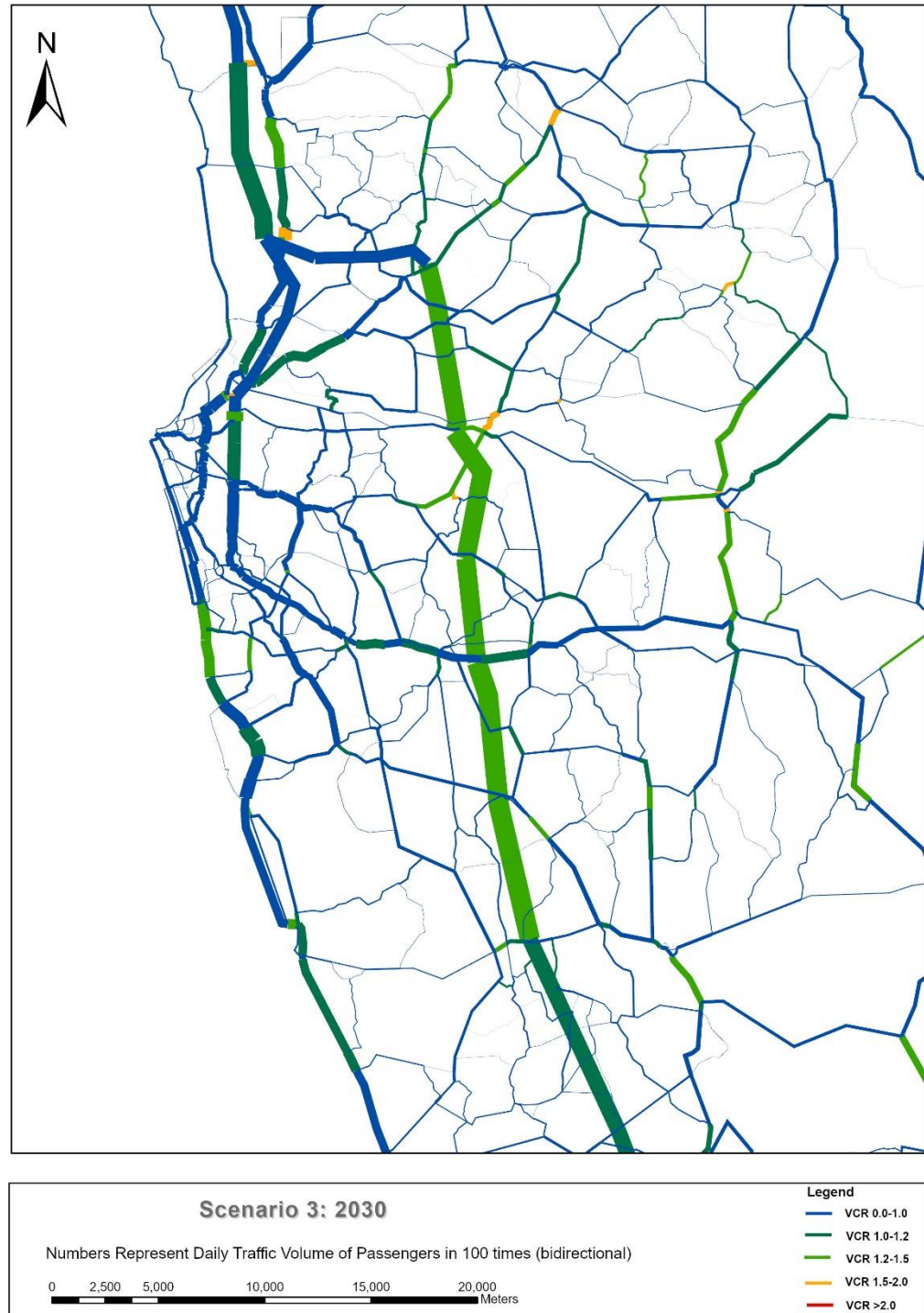


Figure 5.14: Volume/ Capacity Ratio of Scenario 3 (DNS 2030)

5.4.3 Scenario 4: Land use Improvement (2030)

This scenario was based on the land use development plan proposed by UDA for 2030. UDA has prepared an integrated plan considering the economic,

social, and environmental aspects to promote the development of those areas. In the Western Province, some DSD plans were proposed and available on their official website. According to the available information, the below DSDs development plan for 2030 is available (Figure 5.15).

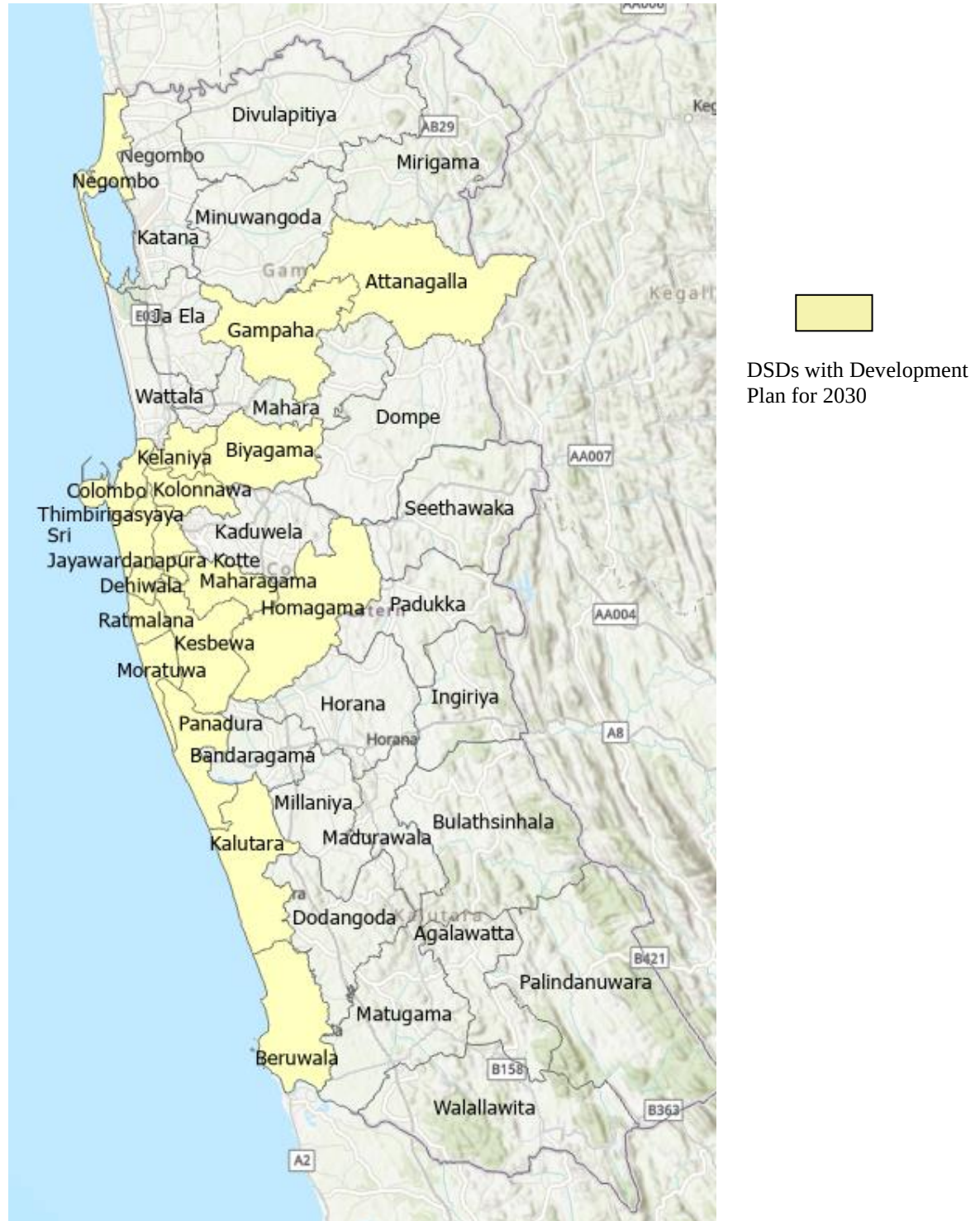


Figure 5.15: Land use Development Plan DSDs 2030

The residential, commercial, and industrial percentages were calculated based on the proposed development plan (Table 5.7) for the 2030 OD modification.

These percentages were developed according to the following categories that UDA has most developed a plan for:

- Concentrated Mixed Development Zone: Prominent use for residential, health, office and institutional, commercial, educational, and industrial
- Mixed Development Zone: Prominent use for residential, health, office and institutional, commercial, educational, industrial, and recreational
- Promoted Mixed Development Zone: Prominent use for residential, health, office and institutional, commercial, educational, industrial, and recreational
- Concentrated Logistics Zone: Development of industrial and offices
- Seafront Mixed Development Zone: Development along the coastal area including residential, health, office, institution, commercial, education, customer care services, tourism, industry, and recreation.
- Residential Prominent Mixed Development Zone: Priority given to residential developments and other basic facilities such as health and education
- Residential Development Zone: Priority given to residential developments
- Special Primary Residential Development Zone: Priority given for residential developments and some other basic facilities

Table 5.7: Land use Development Plan 2030

Land use development Plan-2030			
DSDs	Residential	Commercial	Industrial
Colombo	5%	10%	10%
Thimbirigasyaya	20%	5%	0
Dehiwela	10%	5%	0
Ratmalana	10%	5%	10%
Kalutara	3%	5%	0
Beruwela	3%	2%	0
Attanagalla	10%	5%	0
Biyagama	0%	10%	15%
Gampaha	12%	2%	0
Homagama	10%	5%	0
Kelaniya	10%	5%	3%
Sri J Pura	10%	5%	0
Maharagama	7%	3%	0
Moratuwa	12%	0%	5%
Negombo	3%	3%	0
Panadura	5%	4%	3%
Kesbewa	10%	3%	0
Kolonnawa	10%	3%	0%

5.4.3.1 Development of Origin-Destination Matrix 2030 (Scenario 4)

The trip generation and attraction matrix were developed based on the land use development plan proposed by UDA for 2030 (Table 5.7). By applying the above factor increase to the residential land use, the trip generation is increased by 4% for the Western Province from the do-nothing scenario of 2030, with the total generated trip value of 1.8 million. Figure 5.16 shows the comparison of trip generation of do-nothing scenario and the land use improvement for 2030. Thimbirigasyaya trip generation has significantly increased due to the increase in the residential plan for 2030.

According to the results for 2030, general-purpose and industrial trips will increase by 3% and 2%, respectively. The overall trip attraction increased to 1.7 million, showing an increment of 3% from the 2030 do-nothing scenario. The below Figure 5.17 shows the DSD-based increase of trip attraction.

Biyagama and Colombo DSDs show a significant improvement in trip attraction as the proposed development for the industrial and commercial is higher than other DSDs. Biyagama is focused as a model industrial city and is identified as an ideal for western region industrial development. Moreover, Biyagama DSD has been identified as an intermediate city with industrial economic value centered on the Port of Colombo (UDA 2021). Colombo DSD is also designed to create more competitive economic opportunities than other cities in the South Asian Region (UDA 2021). The gravity model was used to develop the trip distribution for scenario 4 (Appendix 6).

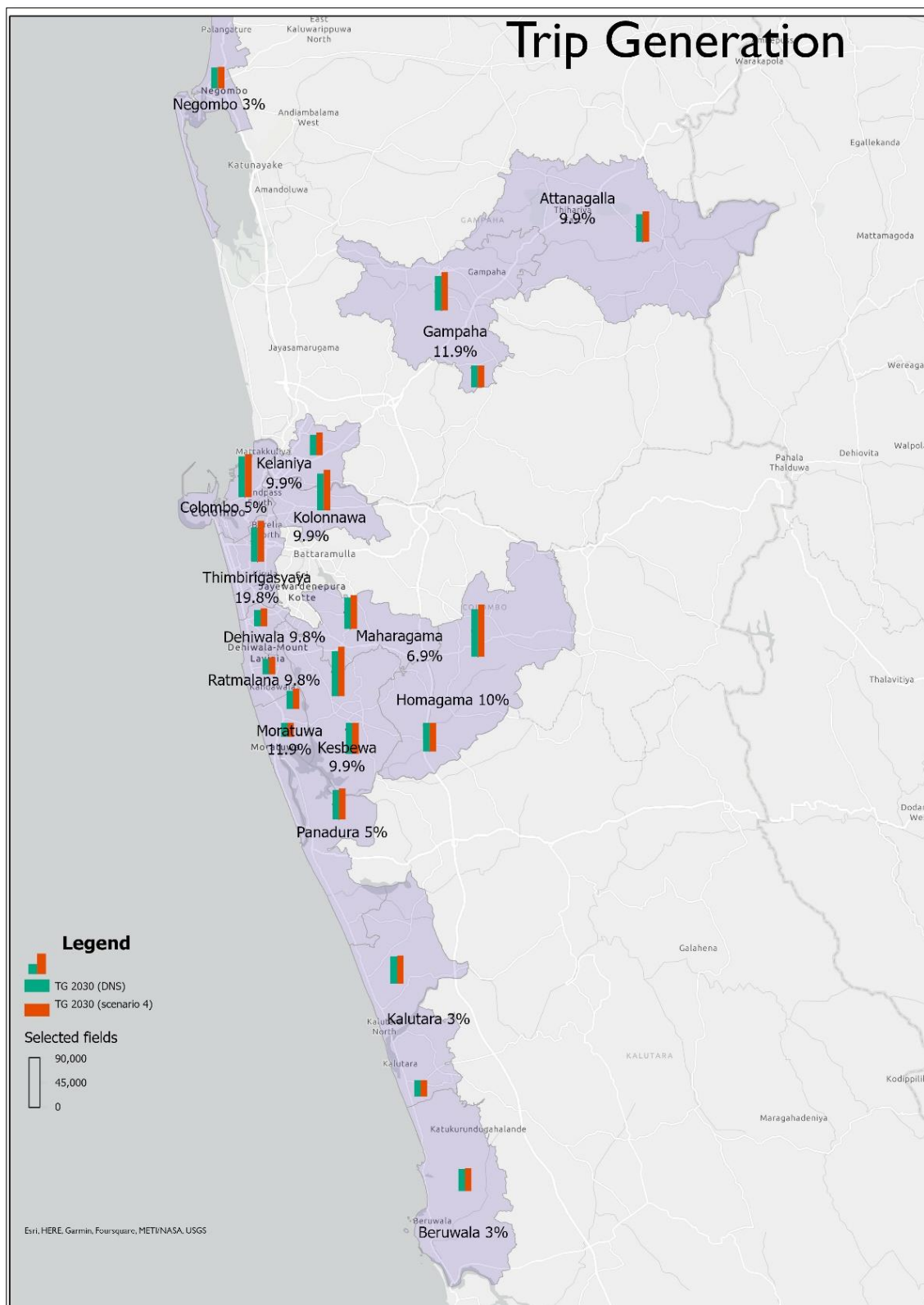


Figure 5.16: Trip Generation based on Land use Development Scenario 4

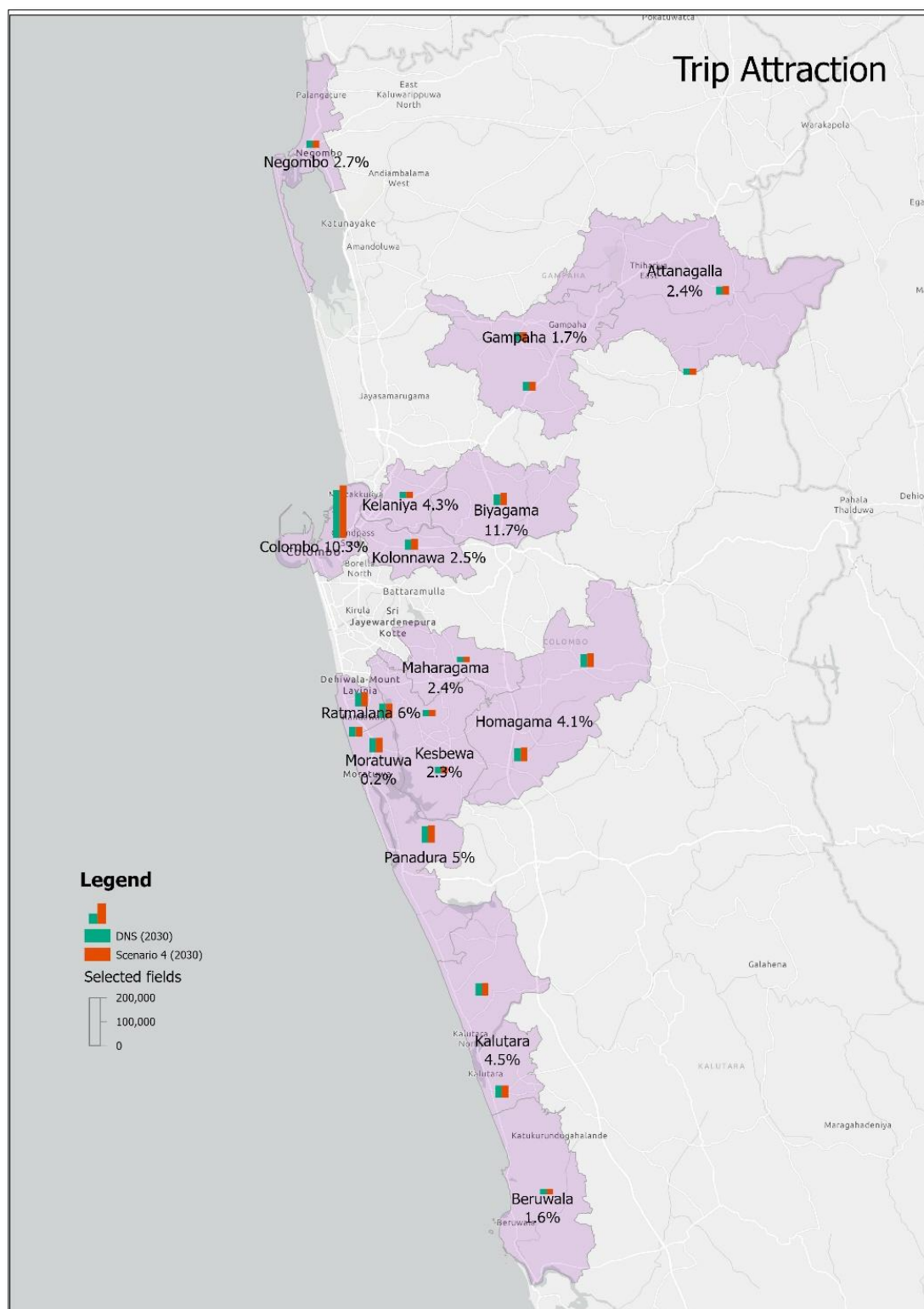


Figure 5.17: Trip Attraction based on Land use Development Scenario 4

5.4.3.2 Estimated Results of Scenario 4

The volume/capacity map (Figure 5.18) for the scenario 4 land use improvement 2030 case is attached below. Results show a significant trip increase with the land use development compared to the DNS of 2030.

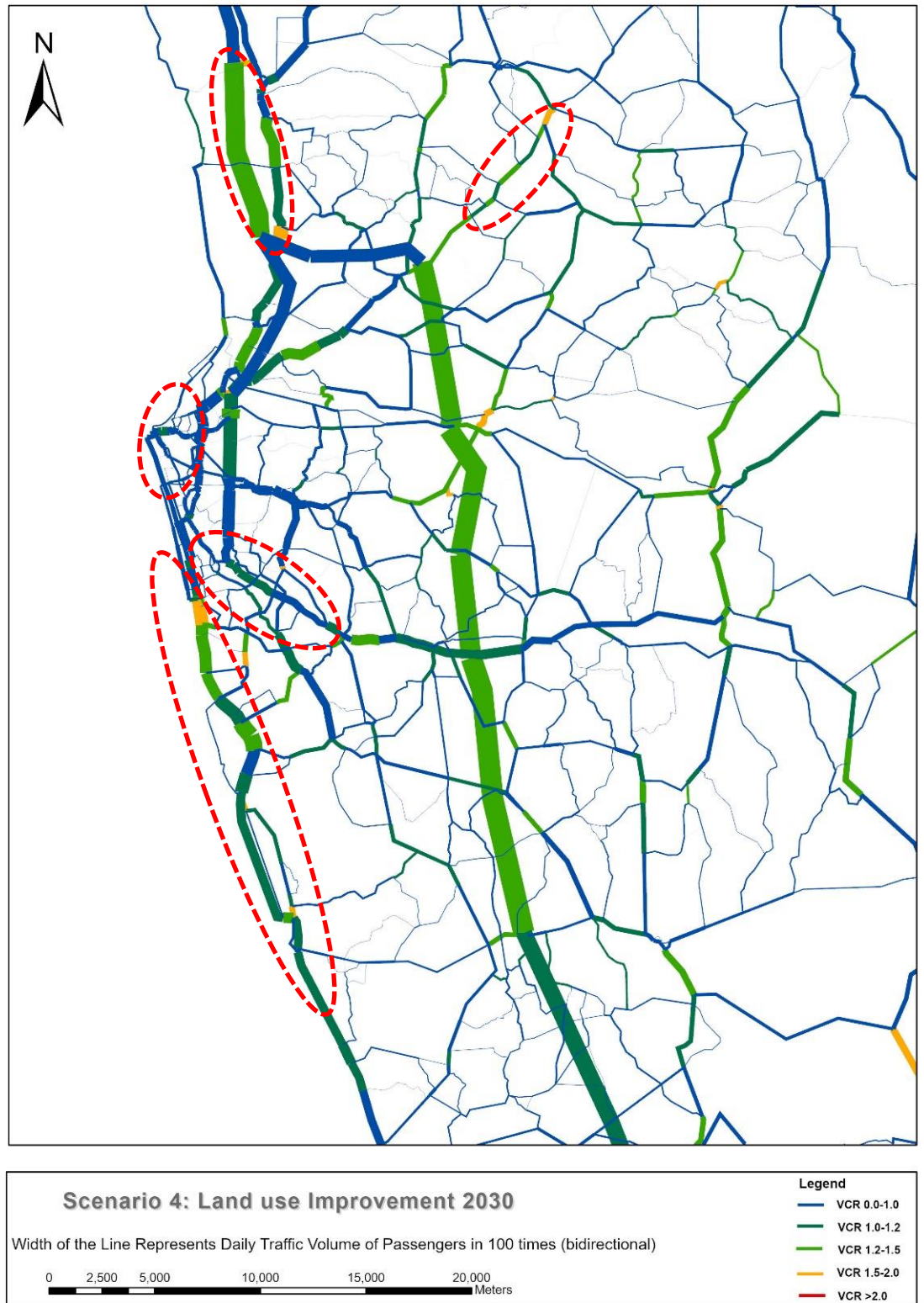


Figure 5.18: Volume/ Capacity Ratio Scenario 4 (2030)

5.4.4 Scenario 5: Land use and Transport Network Development

This case has been developed to test the impact of land use changes in trip distribution and how those effects can be eliminated with supportive transport developments. The OD developed based on land use improvements proposed by UDA in scenario 4 was used with the developments considered in the road

network for 2030.

5.4.4.1 Development of Road Network of 2030

As explained earlier, the intensive road improvement scenario considered in the UOM project has been tested for 2030. The below road projects are considered for the 2030 network (Table 5.8) and final network is provided in Appendix 13.

Table 5.8: Road Development Plan (Scenario 5- 2030)

Road	Description
Middle ring road	4 lanes (Vmax- 45kmph and Qmax: 40,000 PCU/Day)
Inner ring road	4 lanes (Vmax- 45kmph and Qmax: 40,000 PCU/Day)
Baseline extension	4 lanes (Vmax- 45kmph and Qmax: 40,000 PCU/Day)
Malabe road (Pettah to Kaduwela)	4 lanes (Vmax- 45kmph and Qmax: 40,000 PCU/Day)
Galle road (marine drive extension up to Ratmalana)	
Horana Road	Widening to 4 lanes (Vmax- 45kmph and Qmax: 40,000 PCU/Day)
Kandy road (Up to Kadawatta)	Widening to 6 lanes (Vmax- 60 kmph and Qmax: 61,000 PCU/Day)
Low level (Up to Malabe)	Widening to 4 lanes (Vmax- 45kmph and Qmax: 40,000 PCU/Day)
Port link expressway	4 lanes (Vmax- 45kmph and Qmax: 40,000 PCU/Day)
Central Expressway	Vmax- 100 kmph and Qmax: 72,000 PCU/Day

5.4.4.2 Estimated Results of Scenario 5

The volume/capacity map (Fig 5.19) for the scenario 5 land use and transport improvement 2030 case is attached below. Results show a significant decrease in trips along the development proposed for the transportation network compared to the land use scenario 4 case of 2030. A significant reduction can be observed along the middle and inner ring roads.

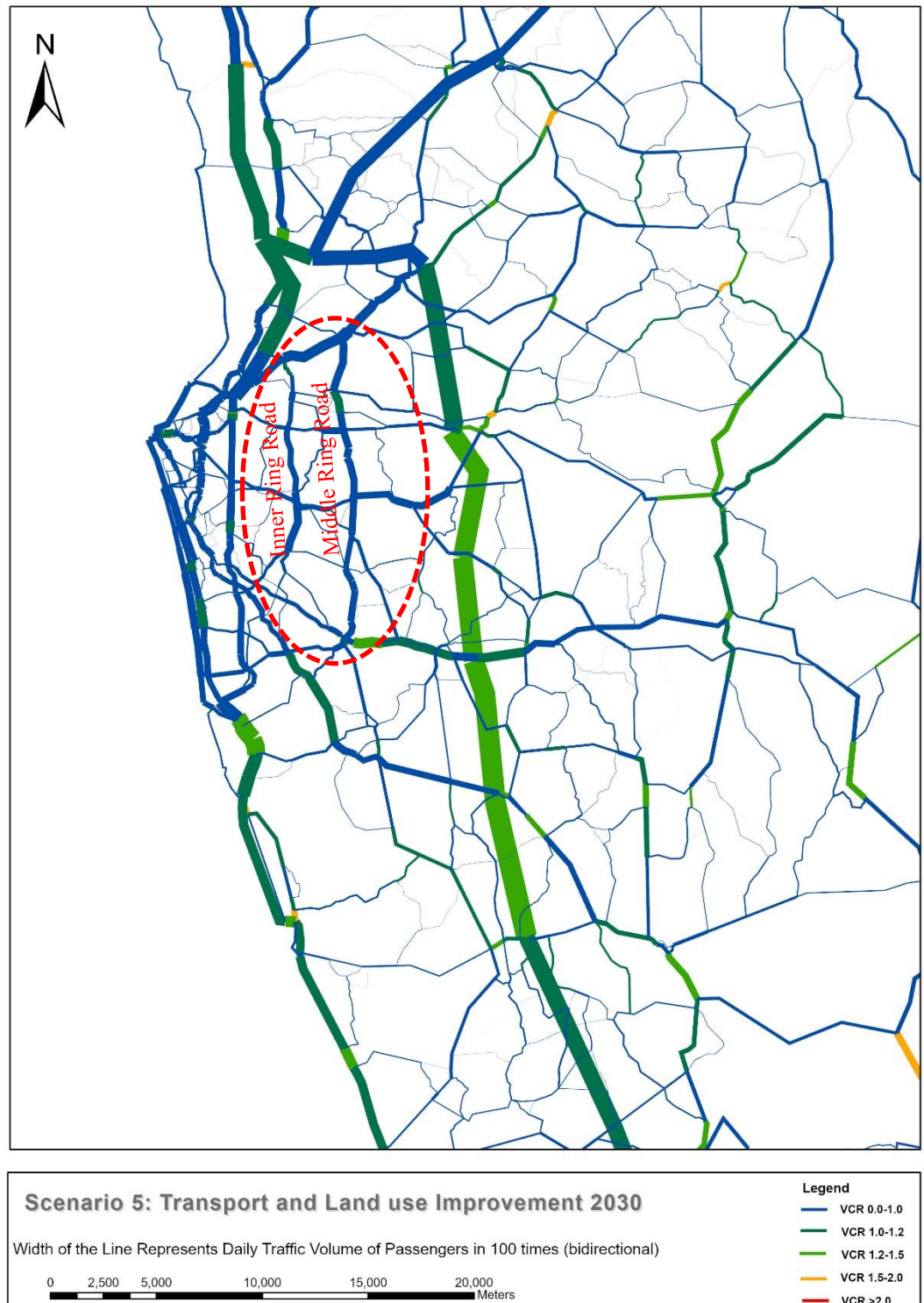


Figure 5.19: Volume/ Capacity Ratio of Scenario 5 (2030)

5.5 Cost Analysis of Each Scenario

In order to create a safe and efficient transportation system, various scenarios and options must be evaluated. To ensure that transportation systems meet the needs of users, policymakers and transportation planners must consider numerous factors, such as passenger kilometres, average trip length, and passenger hours. The cost of the system is, however, one of the most important

factors in transportation planning. There are a variety of costs involved, including the value of time (VOT), vehicle operating time (VOC), accident cost, and emission cost, which all play a role in determining whether a transportation project is feasible and sustainable. All scenarios have been compared with passenger kms, average trip length, passenger hours, average speed, system cost and project benefit. System cost is an aggregation of value of time (VOT), vehicle operating time (VOC), accident cost and emission cost.

5.5.1 Value of Time (VOT)

The methodology used to calculate the time value for 2013 and 2019 has been discussed above and summarised in Table 5.1 and 5.2.

5.5.2 Vehicle Operating Cost (VOC)

The unit VOC is assumed based on UOM revised study. The VOC is assumed based on “Assessing Public Investment in the Transport Sector 2001” by the Department of National Planning, Ministry of Finance and Planning. The price was converted to 2019 price (LKR) based on the Colombo Consumer Price Index (CCPI) (Table 5.9).

Table 5.9: Vehicle Operating Cost

Speed	Car
10	49.32
15	42.33
20	38.95
25	37.02
30	35.78
35	34.93
40	34.33
45	33.89
50	33.93
55	34.01
60	34.09
65	34.17
70	34.29
75	34.41
80	34.53
85	34.69
90	34.85
95	35.01
100	35.17

The VOC is calculated based on below Equation (5.5)

$$VOC = (VOC/km)_{speed} * Total\ Passengers\ kms$$

(5.5)

Where $(VOC/km)_{speed}$ is based on the average speed of the particular mode.

5.5.3 Accident Cost

Accident cost was estimated based on the value used for UOM study for 2013 and adjusted to 2019 CCPI value. Based on that accident cost for 2014 was 1.98

LKR/ vehicle km and it has increased to 2.44 LKR/ vehicle km in 2019.

Accident cost is calculated based on Equation 5.6.

$$\text{Accident cost} = (\text{Passenger kms/ Average vehicle occupancy rate}) * \text{Accident cost/km} \quad (5.6)$$

5.5.4 Emission Cost

Environmental improvements due to the decrease of GHG emission is also considered. While there are several greenhouse gases, carbon dioxide (CO₂) was considered for the analysis due to data availability. Unit emission factors of carbon dioxide for this study uses the value of 275g/km based on the value applied for UOM study and this was converted to the price based on the annual average European Union Allowance (EUA) Price (in 2012 it was 18.8 USD/ton). The emission cost is calculated based on Equation 5.7.

$$\text{Emission Cost} = (\text{Passenger kms/ Average vehicle occupancy rate}) * \text{Emission cost/km} \quad (5-7)$$

5.5.5 Estimation of Results

The total system cost for each scenario is calculated based on the cost considered above and shown in Equation 5.8.

$$\text{Total System Cost} = \text{VOT} + \text{VOC} + \text{Accident Cost} + \text{Emission Cost} \quad (5-8)$$

The economic benefits of each scenario are shown below in Table 5.10. The time value for 2013 was 233 billion per year, which increased by 46% in 2019, with the increment of daily home-based work trips by 42%. Similarly, VOC, accident and emission costs have been increased by 67%, 20% and 20%, respectively, which led to a system cost increase of 54%. During 2013-2019, the network's speed increased from 39 to 42 kmph as more expressways and roads were constructed. The DNS case of 2030 increases the VOT cost by 69% from the scenario in 2019.

Results show that home-based trips significantly impact the value of time. Land use development significantly impacted VOT as it increased the VOT by 25% from the DNS 2030 scenario. Similarly, VOC, accident and emission costs increased by 12%, 11% and 11%, respectively, leading to an increase in system cost by 16%. The impact can be noticed in the increase in passenger-km and hours due to the increase in trip distribution with the development of land use patterns proposed by UDA in a specific area. The development plan is mainly based in the Western Province's urban area, and those developments increased

the trip demand in those areas. Transport development should support the demand created by land use development to keep the balance between these variables. Due to the increase in demand, the average speed of the network dropped from 39 kmph to 37 kmph.

Based on the UOM study's 2030 road network improvement proposal and the development of land use patterns, scenario 5 was examined to assess the balance between land use and transport. It could be seen that there is a significant variation of VOT with the scenario 4 land use case. The VOT has reduced massively from 686 billion LKR to 615 billion LKR annually. It is nearly a 10% drop from scenario 4. Moreover, emission and accident costs were reduced by 1% and system costs by 5%. The proposed plan of land use and transport development of scenario 5 has reduced the overall system cost by 55,556 million LKR per year. Moreover, the average speed of the network is also improved from 37 kmph to 40 kmph. Scenario 5 explained that land use and transport development should be considered to balance the network equilibrium at the desired level.

Table 5.10: KPI of Scenario Cases

Scenario No.	Description	Year	Number of Work Trips (Daily)	Passenger kms (Daily)	Average Trips Length (kms)	Passenger Hours (Daily)	Average Speed (kms/hrs)	VOT in LKR Mn/year	VOC LKR Mn/year	Accident Cost LKR Mn/year	Emission Cost LKR Mn/year	System Cost LKR Mn/year
1	Do Nothing	2013	1,237,693	49,404,172	39.9	1,278,556	39	233,336	246,062	20,276	19,800	519,475
2	Considering 2019 network improvements and OD	2019	1,760,594	59,161,710	33.6	1,400,725	42	340,843	411,844	24,281	23,710	800,678
3	OD changed based on electricity consumption rate with 2019 network	2030	2,131,154	72,178,732	33.9	1,845,005	39	577,223	359,493	29,623	28,927	995,267
4	With Land use improvement	2030	2,131,172	80,219,833	37.6	2,186,894	37	684,185	402,155	32,923	32,150	1,151,414
5	LU Change OD and Transport Developed Network	2030	2,131,172	79,504,520	37.3	1,965,104	40	614,383	416,751	32,630	31,680	1,095,857

5.6 Summary and Conclusion

Using electricity consumption data as a tool for calibrating travel forecasting models has resulted in highly accurate predictions for the Western Province. By leveraging HF data, it is possible to develop models for trip generation and attraction, which can be used to forecast travel patterns and evaluate the impact of land use and transportation on planning. This, in turn, can assist decision-makers in making informed infrastructure development decisions. Additionally, this approach can help evaluate transport demand in both urban and suburban areas, taking into account local road network supply capacities, land use planning and relocation, and economic changes that affect travel demand and land use changes.

Evaluating various scenarios based on future development holds promise as an effective means of analysing the impact of land use and transport. The developed model has been tested based on VOT, VOC, accident costs, and emission costs, and the results indicate that land use and transport integration should be considered together for effective land use and transport planning. The developed model is simple and easy to use, making it suitable for relevant authorities, planners, and decision-makers. Furthermore, the model's variables can be updated in real-time, enhancing flexibility.

The model can also be further developed at a disaggregated level by gathering electricity consumption data based on individual buildings. This approach allows for the development of a disaggregated trip generation model that can provide insights into travel patterns at a smaller scale. The next chapter provides a detailed account of the development of such a model using the available HF proxy dataset for a small area within the study region. Overall, the use of electricity consumption data as a tool for travel forecasting and analysis shows great potential in enhancing land use and transport planning, with the developed model being a valuable addition to decision-making processes.

CHAPTER 6

DEVELOPING A FUZZY RULE-BASED TRIP GENERATION USING HIGH FREQUENCY DATA

[This chapter has been submitted to the Journal and details are below]

Amalan, T. P., Du, B., Kumarage, A. S (2023). A Fuzzy Logic based Trip Generation Approach using High-Frequency Data. Journal of Travel Behaviour and Society, Under Review]

6.1 Introduction

As the first and essential step of the classic four-step transport planning model, trip generation aims to predict the number of trips generated from each traffic zone. Cross classification and Multiple Linear Regression (MLR) are widely used to estimate trip generation (Kabakuş and Tortum 2019, Chang et al. 2014). Cross-classification can predict the trip rate based on socioeconomic characteristics (Stopher and McDonald 1983, Badoe and Mwakalonge 2011). The main disadvantage of cross classification is that the trip rate remains constant for the above stratification. Alternatively, MLR can predict the number of trips generated by each person, household, or zone by using a set of socioeconomic characteristics as independent variables. Hence, MLR is widely used to predict trip generation and attraction, considering specific influencing parameters as independent variables. Nevertheless, the erroneous in the independent variable for the prediction period can lead to imprecise values for forecasting. Furthermore, there is no guarantee that the relationship between the dependent and independent variables will remain same in the future years. When considering the traditional method to predict the vagueness and ambiguity of influencing variables for trip generation, new techniques such as fuzzy logic become more appropriate (Lafta and Ismael 2022). Fuzzy logic is a form of multi-valued logic derived from fuzzy set theory to deal with reasoning that is approximate rather than precise (Guelzim and Obaidat 2015).

Data availability is one of the most critical constraints when estimating independent variables for traditional trip generation models. In many cases, traditional trip generation models require survey data, which can be time-consuming and expensive. In addition, the data may not be available for all locations or may not be up-to-date, which could limit the accuracy of estimates. There has been an increase in interest in developing trip generation models using

alternative data sources and methods in recent years. As an alternative to traditional data collection methods such as the Household Travel Survey (HTS), High Frequency Data (HF data) offers several advantages. HF data provides up-to-date, highly accurate and disaggregated information. A disadvantage of HTS data collection is that it can be time-consuming, expensive, and has limited sample size. It is thus useful for researchers and practitioners to use HF data to develop models that require large amounts of accurate and timely data (Amalan et al. 2023).

This paper attempts to formulate a home-based work trip generation model that takes advantage of the fuzzy logic technique by using available HF data such as household monthly electricity consumption, GPS location of each household and satellite image data. In this study, electricity consumption data was used as a surrogate variable for household income, a very important independent variable for trip generation. With GPS location data and pretrained deep learning models, satellite data can be used to determine the area of households. Benefiting from fuzzy logic, the model can provide more accurate estimates of trip generation, improving transportation planning and reducing traffic congestion. Ultimately, the proposed home-based work trip generation model could be a valuable tool for transportation planners and policymakers. With fuzzy logic and frequently updated variables, the model provides more accurate trip generation estimates, which can reduce traffic congestion and improve transportation planning. In general, the main contribution is threefold: (1) proposing a home based work trip generation that utilises the HF data over traditional survey data such as HTS; (2) The use of fuzzy logic model that addresses the vagueness and ambiguity of influencing variables for trip generation, making it more appropriate than traditional methods; and (3) model can be frequently updated with HF Data, which has several benefits such as accurate at the disaggregated level and real time updates.

This paper is organised as follows. Section 2 provides a brief literature review of the background of the study. Section 3 describes the study area and methodology adopted based on the high-frequency data. The primary analysis of considered parameters is provided in Section 4. Section 5 conducts comprehensive numerical experiments to validate the proposed model. Based on the results from the model, the conclusion and future research recommendations

are discussed in Section 6.

6.2 Background Study

Analysing trip generation can be done using two types of models (Hirun 2015, Modi 2001). The first type of model involves linear regression using socio-economic characteristics to predict trips (Ortúzar and Willumsen 2011). Using this method, trip generation models are developed using regression analysis and a zonal trips prediction equation is created. Many studies indicate that number of home-based work trips were highly dependent on socio economic variables, such as household size, household income and car ownership (Kitamura 2009, Takyi 1990). A second type of model, developed by the Federal Highway Administration in 1975, is called cross-classification (Huntsinger, Roupail, and Bloomfield 2013). In this method, the explanatory variables are directly varied without any assumptions about linearity. According to cross-classification models, households are classified based on socioeconomic characteristics such as number of vehicles owned, size of family, and income.

The concept of fuzzy logic was first introduced in the 1960s (Salini et al. 2017b) and computer-based approach, computing based on degrees of truth rather than usual Boolean logic (1 or 0). Compared to alternative clustering algorithms, such as K-means, Fuzzy C-Means (FCM) clustering is computationally efficient and improves the convergence of the squared-error criterion local minima (Millward, Hafezi, and Daisy 2019). A fuzzy logic system simulates human reasoning and cognition in artificial intelligence systems. The advantages of fuzzy logic for modelling nonlinear and complex systems are that it manages uncertainty and imprecision quite robustly, and it integrates expert knowledge quite easily. Fuzzy logic based models are typically developed in three steps, namely (1) Fuzzification (fuzzy inputs), (2) inference engines or if-then rules and (3) defuzzification. Several inferencing models use fuzzy sets to explain the outputs of a system based on its inputs. In the beginning, Mamdani fuzzy inference was introduced as a way of synthesising human-assisted linguistic control rules for creating control systems and was introduced by Professor Ibrahim Mamdani in 1977. This interference has several advantages, such as,

- using the model when the data is uncertain and imprecision
- utilising the expert's valuable judgements

- reducing the high requirements for precise data in modelling and predicting stages
- providing a cost-effective alternative to ordinary expensive survey-based methods

Fuzzy logic is one of the most popular techniques used in transportation modelling, especially in estimating trip generation, trip distribution, and Origin-Destination (OD) matrices. Despite using different variables and techniques, all the papers show an improvement in accuracy when compared to linear regression and gravity models. "Estimating origin-destination matrices with fuzzy weights" (Xu and Chan 1993) and "Estimation of origin-destination matrices using fuzzy Logic" (Sbai and Ghadi 2018) use fuzzy set theory to estimate origin-destination matrices with superior accuracy over traditional methods. Sbai and Ghadi found that the fuzzy logic model had a lower Root Mean Square Error (RMSE) (24.45) than the gravity model (26.23). Xu and Chan (1993) developed an approach for estimating origin-destination matrices using fuzzy weights. As a result of their approach, they found that OD matrices obtained by fuzzy sets were more accurate than those obtained from conventional OD estimation methods. A maximum entropy optimization model incorporating fuzzy entropic parameters is proposed by Lopez-Ospina et al (López-Ospina et al. 2022) to estimate the OD trip matrix. For representing the membership functions of OD trips, the model uses fuzzy sets that consider the uncertainty and imprecision of the data. A more robust and reliable model is achieved by incorporating fuzzy entropic parameters, which are introduced to handle uncertainty in the data. The model is compared with several other OD estimation models in a real-world case study. Finally, "Spatial distribution of urban trips in recently expanded Surat city through Fuzzy logic with various clustering Techniques: A case study of the typical metropolitan city in India" (Salini et al. 2017a) uses fuzzy logic-based trip distribution models with FCM and subtractive clustering techniques and finds that the subtractive clustering model performs better than the FCM model and the gravity model. It is evident that the proposed model provides better accuracy when sparse or noisy data is present.

Various contexts are illustrated by the reviewed papers where fuzzy logic-based

approaches are demonstrated to improve trip generation modelling accuracy. Several factors are considered in these papers, including land use, population density, accessibility, household size, income, and parking availability, to generate predictions for trip generation. Both "Prediction of future trips using fuzzy logic based trip generation model" (Rassafi, Rezaei, and Hajizamani 2012) and "Application of fuzzy logic to model trip generation phase of sequential travel demand analysis" (Agarwal 2012) utilize fuzzy logic to generate trips, with significant improvements in accuracy compared to linear regression models. Rassafi, Rezaei, and Hajizamani (2012) compared traditional linear regression models with fuzzy logic models to predict future trips. In order to evaluate the performance of the models, they used statistical parameters such as R^2 and RMSE. As a result of incorporating fuzzy logic into trip generation models, the R^2 values improved from 0.84 to 0.96 for productions and from 0.801 to 0.987 for trip attractions. According to Agarwal (2012), the trip generation phase of sequential travel demand analysis is modelled using multiple linear regression and fuzzy logic. Vehicle ownership, household size, and income were the variables considered. A comparison of the model's results with field data confirmed its validity, and the author believes it will be useful in the future. In general, fuzzy logic-based approaches have shown promise for improving the accuracy of trip generation estimates and can be applied to several different contexts in transport planning. Overall, fuzzy logic has shown promise for improving accuracy and providing a deeper understanding of travel behaviour in transportation modelling.

In addition, fuzzy logic approach has been used in optimisation of transport system, such as mode identification, traffic signal control and route planning. "Hierarchical process of travel mode imputation from GPS data in a motorcycle-dependent area" (Nguyen and Armoogum 2020) paper explains the identification of modes using fuzzy based approach using set of rules based on acceleration and speed related variables. Traffic light controllers based on fuzzy logic can respond ideally to fluctuating traffic volumes, like oversaturated or unusual loads (Kulkarni and Waingankar 2007, Tunc, Yesilyurt, and Soylemez 2021). The use of fuzzy logic in routes planning literature has been widely acknowledged as a means of handling uncertainty and imprecision (Nasibov, Diker, and Nasibov 2016).

Deep learning is commonly used to extract building footprints from high-resolution aerial or satellite imagery. Several pre-trained models have been developed for this purpose, such as building footprint extraction developed for the USA, Africa, Australia, and China are famous among developed models, which are available in the ArcGIS living atlas for the world. These models have the precisions of 71.8%, 78.6%, 79.4% and 58%, respectively (Esri, 2020). All these models used the Mask RCNN model architecture implemented using ArcGIS API for Python. Most of the research uses high resolution satellite images to develop deep learning algorithm to detect building footprint. The paper "Continental-Scale Building Detection from High Resolution Satellite Imagery" presents a methodology of using 50 cm satellite imagery with widely used U-Net model for building detection at a continental scale (Sirko et al. 2021). The results achieved high precision and recall rates. Using a Convolutional Neural Network (CNN), the paper "CNN-based building footprint detection from Sentinel-1 SAR imagery" discusses an automated method for extracting building footprints from SAR imagery collected by Sentinel-1. Using a CNN-based classifier, the authors proposed a two-stage method of pre-processing SAR images and detecting buildings and achieved F1 score 0.75, which indicates that the proposed model is able to detect building footprint accurately (Rapuzzi et al. 2020).

There were some studies that used other images such as airborne light detection and ranging (LiDAR) measurements to identify building footprint models accurately compare to the aerial and satellite images. (Zhang, Yan, and Chen 2006, Alexander et al. 2009, Wang et al. 2012). There have been several studies that have pointed out that using aerial and satellite imagery to analyse urban structures can be challenging because of their complexity. According to them, inaccurately extracting building footprints from imagery is difficult due to irregular building shapes, obstructions caused by trees and other structures, and variances in building heights and orientations. (Shirowzhan et al. 2020, Rezaei et al. 2018).

6.3 Methodology

6.3.1 Overall Methodology of the Study

The methodology illustrated in Figure 6.1 of this study encompasses the analysis

of three distinct high-frequency data types: monthly electricity consumption records for households, GPS coordinates of each dwelling, and building footprint data extracted from 2013 satellite images. The process begins with the application of a pre-trained deep learning algorithm to derive building footprint details from satellite imagery. Subsequently, the relevant monthly electricity consumption data is pinpointed based on the GPS locations of each household, drawing from information provided by electricity providers. The household's trip generation rate for home-based work purposes is then computed using data from the Household Travel Survey (HTS) of 2013. Finally, a fuzzy logic inference system is employed to establish a nuanced relationship between the trip generation rate, area, and monthly electricity consumption data for the year 2013. The study further delves into the in-depth approach taken for the preparation of each input, elucidating the comprehensive methodology adopted for

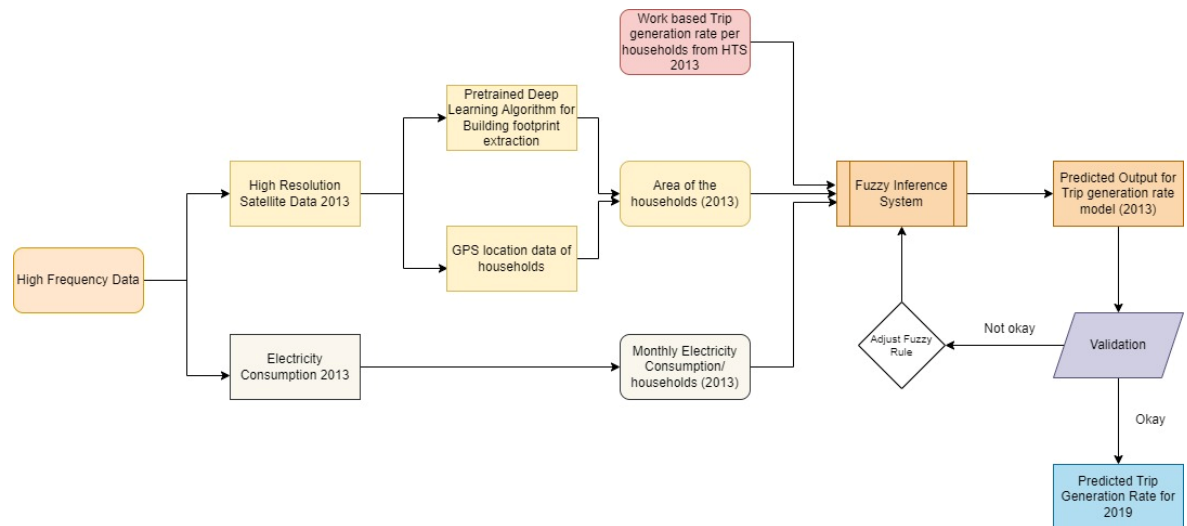


Figure 6.1: Overall Methodology

6.3.1 Preparation of Satellite Image of the Study Area

Due to the limitation and high cost of acquiring high-resolution satellite imagery for Sri Lanka, the study uses freely available GE images (high resolution). Since this study analysed 2013 and 2019 data, both years' images were retrieved manually from Google Earth Pro 7.3.4. Data from 2013 were gathered between Dec 2012 to Jan 2014, and 2019 were acquired between Dec 2018 and Jan 2020 due to the cloud coverage and availability of historical satellite images. Eighty-eight images at 760 m eye elevation levels were collected for Thimbirigasyaya DSD each year. These images are geo-referenced

and mosaicked using Arc GIS Pro software. As these images are collected in different timelines, the colour difference can be observed in the final mosaicked images (Figure 6.2).

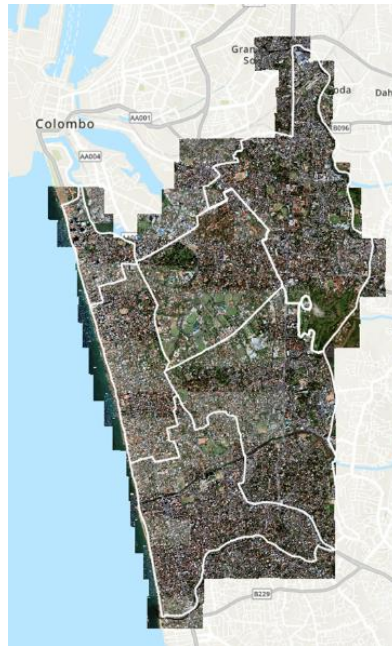


Figure 6.2: Mosaicked Image of Thimbirigasyaya

6.3.2 Detecting Building Footprint Using Deep Learning

6.3.2.1 Overview of the Deep Learning Model

Mask Region- based Convolutional Neural Network (Mask-RCNN) is commonly used technique for the object detection. The accuracy of the Mask-RCNN model can be measured using the Average Precision (AP) metric. AP, a common metric in object recognition and segmentation tasks, measures model accuracy at different levels of recall. It is calculated using the following equation 6.1:

$$\text{Precision} = \text{True Positives} / (\text{True Positives} + \text{False Positives}) \quad (6.1)$$

Here, true positives represent the number of correct positive detections of the model, while false positives represent the number of false positive detections (He et al. 2017). In other words, precision represents the model's ability to correctly identify positive detections without including too many false positives. A high precision score indicates the model is making accurate detections with a low false positive rate, while a low precision score indicates the model is making

many false positives. Precision is often used in conjunction with recall, which measures the proportion of true positive detections out of all actual positive instances in the data set. Precision and recall combine to give a more complete picture of the model's performance. The AP is then calculated as the area under the precision recall curve. Higher AP values indicate better performance, with a perfect value of 1.0 indicating the model has achieved 100% accuracy at all memory levels. In general, the accuracy of a Mask R-CNN model can be affected by several factors, including the quality and size of the training data set, the complexity of the objects to be detected, and the model's hyperparameters such as B. the learning rate and the number of epochs. Therefore, it is important to carefully tune these parameters and train the model on a diverse and representative data set to achieve the best possible precision.

In deep learning, it is common to monitor training and validation loss during the training process to assess a model's performance. Training loss is a measure of how well the model fits the training data. During training, the model attempts to minimize the value of the loss function by adjusting the neural network's weights and biases. In general, since the model iteratively updates its parameters, the training loss should decrease over time. However, if the training loss remains high or increases over time, this may indicate that the model is not learning well and may require adjustments to the architecture or hyperparameters. The validation loss, on the other hand, is the measure of how well the model performs on invisible data. During training, part of the dataset is usually reserved for validation purposes. After each training epoch, the model is evaluated on the validation set to assess its performance. The validation loss should generally decrease over time, indicating that the model gets better as it generalises to new data. However, if the validation loss starts to increase or stays high, it may indicate that the model is overfitting the training data and is not generalizing well. By monitoring training and validation loss during training, it is possible to detect problems with the model's performance and adjust as needed to improve its accuracy and generalization abilities. Generally, the training loss is calculated by averaging the loss function across all samples. If N is the size of a batch, the training loss can be calculated as follows (Equation 6.2):

$$\text{Training Loss} = 1/N * \text{sum}(\text{loss}_i) \quad (6.2)$$

Loss_i represents the loss function associated with the *i*th training sample. As with the validation loss, the validation loss is calculated as the average of all validation samples. The validation loss, for a validation set of size *M*, can be calculated as follows (Equation 6.3):

$$\text{Validation Loss} = 1/M * \text{sum}(\text{loss}_j) \quad (6.3)$$

A validation sample is represented by a loss function, loss_j. It is important to note that the loss function used in these equations will vary depending on the specific deep-learning task and architecture used. Cross-entropy loss functions are often used in classification tasks, while mean-squared error loss functions are commonly used in regression tasks.

6.3.2.2 Methodology Used for Building Footprint Extraction Using Deep Learning

The area of households is derived from Google Earth satellite images and Global Positioning System (GPS) data from electricity providers in Sri Lanka. To apply different tariff calculations, the electricity providers in Sri Lanka classify their customers into different categories, such as domestic (residential), commercial, industrial, and others. Several studies used open-source datasets to detect building blocks and land use classification with satellite image data such as foursquare data, Flickr photos etc.(Spyratos et al. 2016a, Jayanetti, Meedeniya, et al. 2017, Estima, Fonte, and Painho 2014). In addition, GPS devices have become a standard method of collecting data for transportation analysis since the 1990s (Sadeghian et al. 2022). The overall methodology of detecting the building footprint area of households with GPS data of electricity consumers is shown in Figure 6.3.

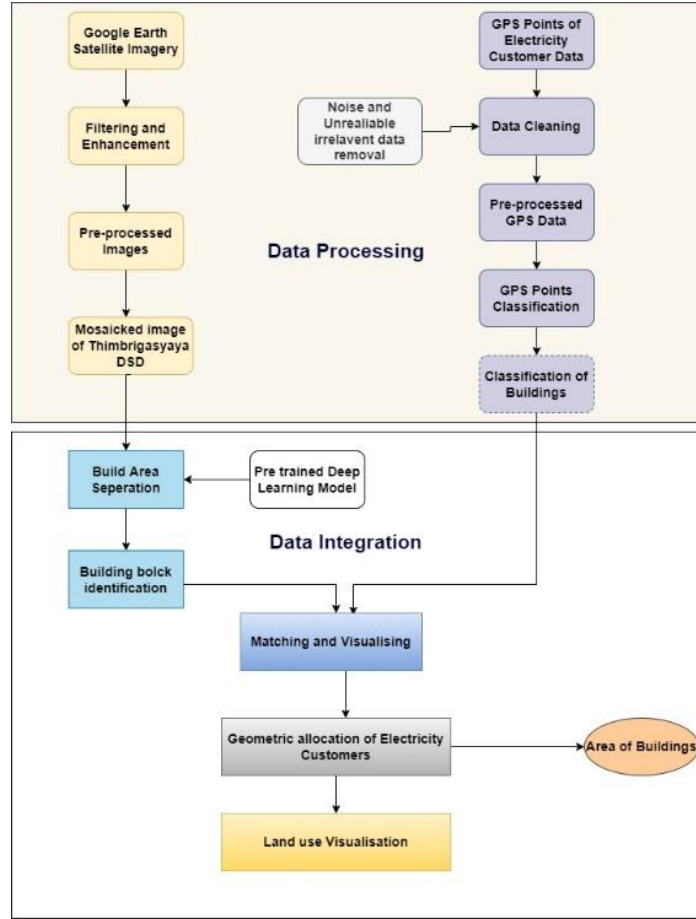


Figure 6.3: Methodology of Building Footprint Extraction and Map Matching of GPS

6.3.3 Overview of the Fuzzy Logic Model

This research adopted Mamdani systems used as an inference system to develop the fuzzy model, which is well suited for human inputs and rule-based control. The fuzzy set differs from a classical set and belongs to any other set simultaneously or partially. Fuzzy set A defined as follow (Equation 6.4),

$$A = \{x, \mu_A(x)\} \quad (6.4)$$

Where $\mu_A(x)$ is a membership function of element x in set A , where $0 \leq \mu_A(x) \leq 1$. The membership function is a graphical representation of each input's participation magnitude. A linguistic variable is an essential concept in fuzzy logic. In this research, all the variables are based on Gaussian membership functions, and the equation is provided below (Equation. 6.5),

$$f(x, \sigma, c) = e^{\frac{-(x-c)^2}{2\sigma^2}} \quad (6.5)$$

Where σ is the standard deviation and c is mean.

The fuzzy rules are controlled by AND operation and OR operation.

Defuzzification uses the centroid method for computing a single crisp output value for a fuzzy set. This method returns the centre of gravity of the fuzzy set along the x-axis. The centroid is computed using the following formula (Equation 6.6),

$$x_{centroid} = \frac{\sum_i \mu(x_i) x_i}{\sum_i \mu(x_i)} \quad (6.6)$$

Where $\mu(x_i)$ is a membership function value for point x_i .

The result from the fuzzy logic for 2013 shows that the predicted value for work from home from the fuzzy model is closer to the relevant value calculated from the HTS data. Based on the output from the fuzzy model trip rate and the total households detected from the deep learning model, the trips were calculated for the entire households using the below formula (Equation 6.7),

$$TG_{Zone\ i} = \left(\frac{TG_{Fuzzy\ i}}{\sum_i HHS_{DL}} \right) \times \sum_i HHS_{Total} \quad (6.7)$$

wherein

$TG_{Zone\ i}$ = total trip generation for a zone i

$TG_{Fuzzy\ i}$ = total trip generation calculated fuzzy model for zone i

$\sum_i HHS_{DL}$ = total number of households detected by deep learning modal for the zone i

$\sum_i HHS_{Total}$ = total number of GPS points of HHS from electricity data for zone i

6.4 Data Analysis

6.4.1 Description of the study area

Thimbirigasyaya DSD area (Figure 6.4), located in Western Province, Sri Lanka, is selected as a study area for this research. It is a very complex and challenging DSD in the Colombo district concerning land-use patterns and traffic congestion. It had a population of 236,903 in 2012 (Department of Census and Statistics, 2012). The total area of this DSD is 22 km², and nearly 50% of the land was utilised for residential purposes in 2012 (CoMTrans 2014).

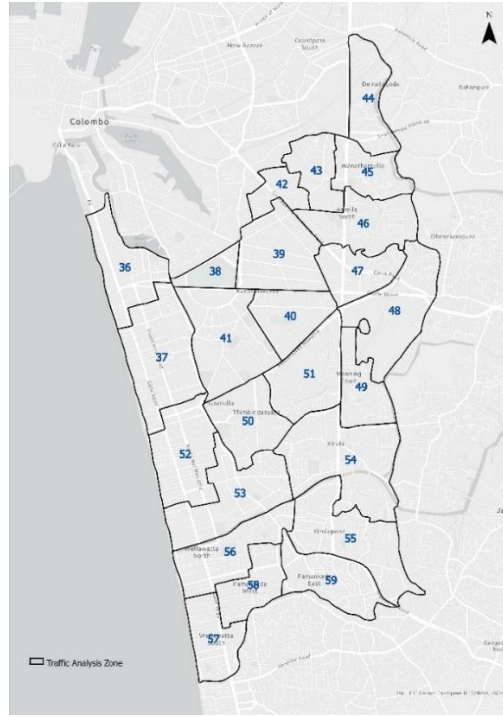


Figure 6.4: Study Area and TAZs of Thimbirigasyaya DSD

6.4.2 Using the Deep Learning Model

Deep learning is commonly used to extract building footprints from high-resolution aerial or satellite imagery. Several pre-trained models have been developed for this purpose, such as building footprint extraction developed for the USA, Africa, Australia, and China are famous among developed models. These models have the precisions of 71.8%, 78.6%, 79.4% and 58%, respectively. All these models used the Mask RCNN model architecture implemented using ArcGIS API for Python. These models perform well in their relevant countries and need to retrain the model based on the extra number of samples to apply to other countries. This kind of learning is called transfer learning, where the existing model is retrained based on the new additional samples.

This research attempts to fine-tune the pre-trained model of building footprint extraction deep learning technique developed for USA and Africa (descriptions are available in appendixes 9 and 10) using the following methodology (Figure 6.5).

Requirements	• Arc GIS Pro • Nvidia GPU • CUDA toolkit
Pre trained model	• Building Footprint Extraction-USA
Create Training Data	• Same as pre trained model parameters
Export Training Data	• About 200 samples
Final model	• Building Footprint Extraction-Sri Lanka

Figure 6.5: Methodology used Transfer Learning

Arc GIS Pro software has been used to retrain the model. After satisfying the basic requirements to develop a deep learning environment, the existing model has been used to detect the building footprint first. Initially, 30 samples were collected based on the non-detected or error-detected buildings (Figure 6.6). The sample of the precision result of the first attempt with the train and validation graph (Appendix 11). The graph illustrates that the validation loss is greater than the training loss, which indicates that model performs well on training data but poorly on the new data. The description of input and output commands is given in Appendix 11.



Figure 6.6: Building Detected and Samples

Unfortunately, the precision value for the first case shows 18% with 30 samples. It may be because of this that fewer samples were collected. However, then again model gives 0% precision with 200 samples. Several cases have been tried (Table 6.1) by increasing the sample size to 200, but the maximum average precision recorded with African Model with 239 samples is 69.2%. The train and validation loss graph and ground truth/predictions images are shown in Appendix 12.

Table 6.1: Cases Considered for Transfer Learning

No of Cases	Description	Precision
Case 1	200 Samples with USA model	0
Case 2	New Model for SL (1723 Samples)	1.33%
Case 3	Model for SL using USA model (1723 Samples)	0
Case 4	New samples avoiding grey colour buildings (USA Model)	7.7%
Case 5	Small area of image with USA model (239 samples)	63.5%
Case 6	Small area of image with USA model (497 Samples)	54.78%
Case 7	Small area of image with African Model (239 Samples)	69.2%

Using USA model, the average precision score is reduced when sample sizes increase. There was also confusion with the colour of the road section in some of the samples with grey roofs. So, these grey buildings were avoided in most of the samples for cases 5, 6 and 7. Even though several models were tested based on the different scenarios, the precision scores are shallow from the original model, and using the retrained model in the study area is unsuitable. The problems identified as the mosaicked image of the study area used to create from different timeline images, and each timeline image has a different wavelength to capture the colours accurately. It was concluded that the transfer learning for building extraction could be performed only with higher-resolution satellite imagery.

Due to the above problem, this research used a pre-trained deep learning model developed for the USA as it accurately detected the building footprints for the study area compared to other models. Developing a new building footprint extraction model using deep learning may be time-consuming and requires high-performing computer technology. Moreover, this model requires 8-bit, 3-band high-resolution (10-40 cm) imagery and has an average precision score of 71.8 % (Esri 2020).

6.4.3 Extraction of Building Footprint

ArcGIS Pro software extracted the building footprint from the mosaicked satellite image of Thimbirigasyaya DSD. The task was performed using ArcGIS Pro's "Detect objects using deep learning" feature. The critical parameter in that tool is setting the threshold value, which can affect the number of buildings detected and the categorisation of the detected footprint. By decreasing the threshold value, more footprints will be detected, but the footprint of the building may move away from the basic form (Pepe et al. 2021). For this purpose, the model adopted the threshold value of 50%. There were significant issues with detecting several buildings together, recognising complex building shapes, and enhancing image clarity (Figure 6.7).



Figure 6.7: Sample of Detected Buildings using DL

While detecting building footprints, some polygons of building blocks were removed due to the above issue, and the building footprints of accurately detected buildings were reviewed. The ArcGIS pro tool "Regularise Building Footprint" was used to regularise the shape. Finally, the near GPS points were

detected for each building block (150 meters). Table 6.2 shows the percentage of detected households with the actual counts of households detected from GPS.

Table 6.2: Detected HHS vs GPS counts (2013 and 2019)

Category	GPS Points	Detected using DL	%
Households- 2013	53,629	27,641	51.54
Households-2019	60,347	29,468	48.83

6.5 Primary Analysis of Data Collection

6.5.1 Relationship between Income and Electricity Consumption

Several studies were conducted to validate the relationship between income and electricity consumption (Athukorala and Wilson 2010, Rajmohan and Weerahewa 2010, Kim 2020b). The primary analysis was conducted on the data of HTS 2013 to show the relationship between electricity bills per month and income. The income group is categorised based on the CoMTrans project classification, and descriptions are as below,

- Low income: Household monthly income is less than 40,000 Sri Lankan Rupees (LKR)
- Middle income: Household monthly income is between 40,000 to 60,000 LKR
- High income: Household monthly income is above 60,000 LKR

The below analysis is undertaken for the entire Western Province, Sri Lanka. The income and electricity expenses distribution are shown below (Figure 6.8). Based on Figure 6.8, it is evident that the relationship between income and electricity expenses has strong relationship. Most low-income households' electricity consumption expenses fall under less than 2000 LKR (Figure 6.8a), while middle-income expenses fall between 100 to 3000 LKR (Figure 6.8b). The majority of high-income households spend above 2000 LKR for electricity (Figure 6.8c).

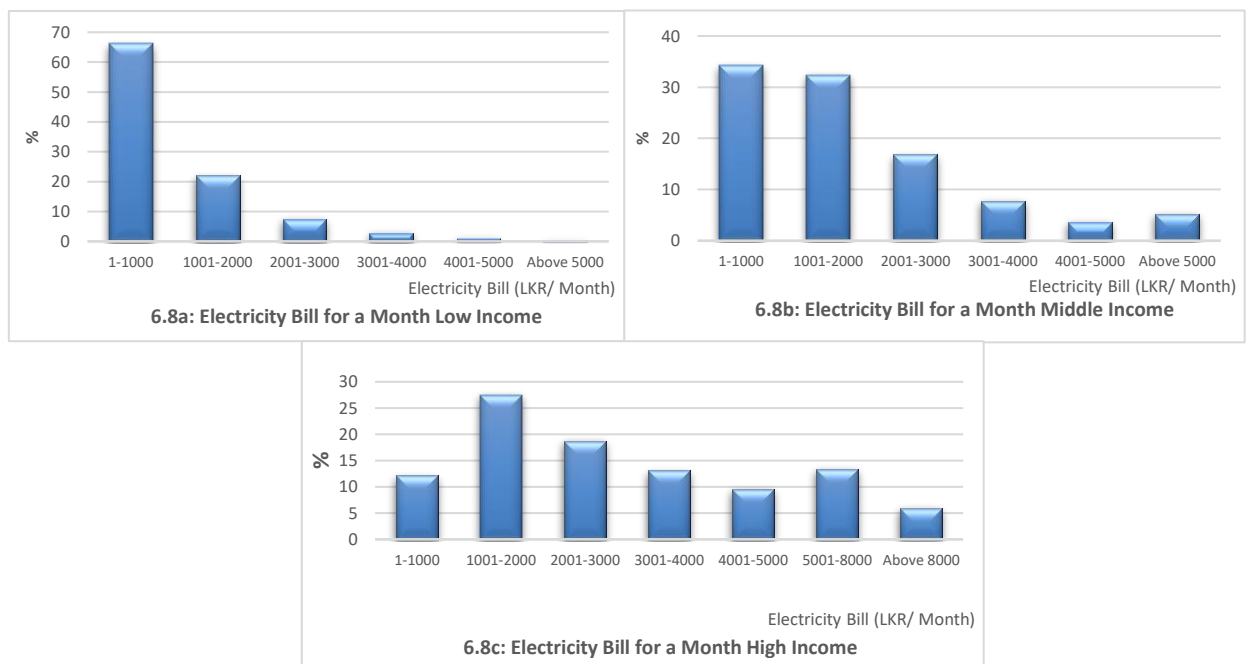


Figure 6.8: Electricity Consumption according to the Income (Thimbirigasyaya DSD)

The descriptive statistics of each income level for the entire Western Province are shown in Table 6.3, and the conversion of electricity expense to the unit (kWh) is also shown in the same table based on the mean value of each group. According to this dataset, different income groups have different levels of electricity consumption. The average income of higher income groups and the average amount of electricity used by higher income groups correlate. Moreover, standard errors provide information about the uncertainty of inferences about a larger population based on the sample means. Based on the table provided, it is evident that higher income levels are associated with higher average electricity consumption. The table indicates that electricity consumption is positively related to income level.

Table 6.3: Descriptive Statistics of Income vs Electricity Consumption

Income Level	Mean	Stand. Error	Average Electricity unit (kWh)
Low	731.469	3.935	78
Middle	1347.852	8.025	97
High	2053.099	16.981	118

6.6 Validation of Electricity Domestic Customers' GPS Location Data

The electricity domestic customer data for 2013 and 2019 were collected from

two major electricity providers in Sri Lanka. In addition to the customer GPS point data, monthly electricity consumption for each customer was also collected, and the complete consumption data were available for October for both years. GPS point data for 2013 was analysed and validated with the latest census data in 2013. Table 6.4 shows the total number of households recorded from GPS data for the Colombo district and Thimbirigasyaya DSD and compares the accuracy with the census 2013 data. The accuracy of GPS counts of households against the census data shows an acceptable range of 102%.

Table 6.4: Households Census vs GPS data 2013

Location	Total HHS from Census 2013	Total HHS from GPS points 2013	%
Colombo District	572,475	553,713	96.7
Thimbirigasyaya DSD	52,285	53,628	102.6

The validation of the GPS point data for 2013 against the latest census data in 2013 indicates that the dataset is reliable and can provide an accurate representation of the number of households in the Colombo district and Thimbirigasyaya DSD. The GPS point data for 2013 and 2019 was further analyzed to calculate the annual growth rate of domestic customers, which was found to be 2% over the six-year period.

6.7 Results of Fuzzy Logic Model for Trip Generation

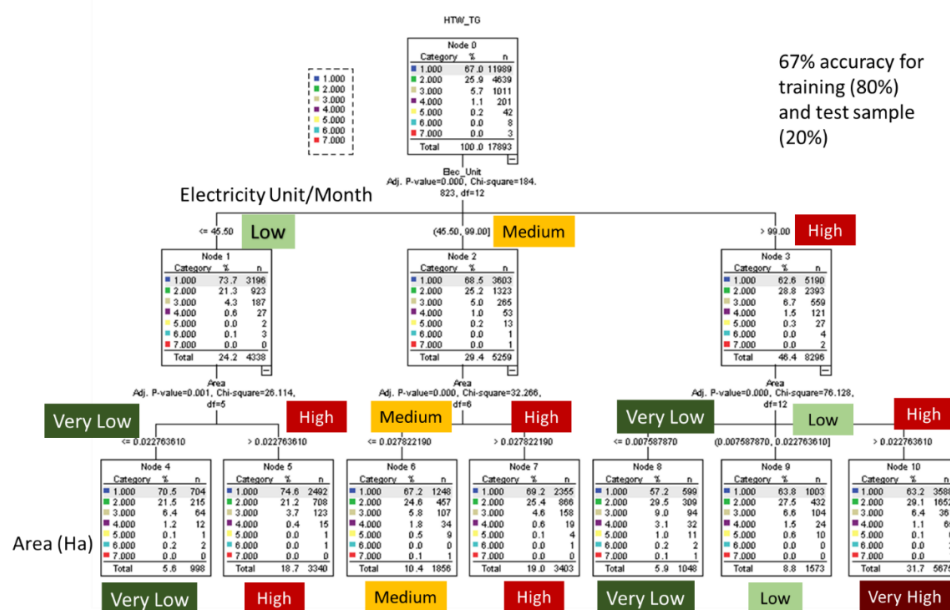


Figure 6.9: Decision Tree (Trip Generation vs Electricity Consumption and Area of Households)

The initial fuzzy model rules related to work-related trip generation, electricity consumption and area of households were derived from the HTS survey in 2013 and developed using MATLAB. A decision tree was used to determine the relationship between these variables, the parameters range, and fuzzy rules. (Figure 6.9). The total number of households, information on work trip generation (Home based work only), electricity consumption (electricity unit (kWh)/month) and floor area of the households were extracted from HTS 2013 survey. The data was split between 80% (14,314 HHS) for training data and 20% (3,579 HHS) for testing data, and the model shows 67% accuracy. According to the results, electricity consumption data is divided into three categories [Less than or equal to 45.5: Low (L), between 45.5 and 99: Medium (M), above 99 units is High(H)]. Moreover, based on the further classification of TG, the area of HHS is divided into four categories; [Less than or equal to 75.8 m²: Very Low (VL), between 75.8 to 227.6 m²: Low (L), between 227.6 to 278.22 m²: Medium (M) and above 278.22 m²: High(H)].

Table 6.5 briefly describes the mean and standard deviation of variables considered at each classification. The means of the Electricity Unit variable increase as the classification goes from Low to Medium to High, with means of 40, 91.94, and 258.1, respectively. For the High classification, the standard deviation is 33, indicating greater variability in electricity usage among households. As the classifications rise from Very Low to Low to Medium to High, the mean area of households increases from 70 to 227 to 274.7 to 2120. There is however a significant amount of variation in household size among households in the High classification, as indicated by the 4805 standard deviation. As the classification of Trip Generation increases from Very Low to Low to Medium to High to Very High, the means increase by 0.8354, 1.085, 1.344, 1.64, and 1.931. There is a similar level of variation in trip generation across all classifications, with a standard deviation of 0.124. A useful insight into household behaviour and energy use patterns can be gained from the table's distribution of these three variables.

Table 6.5: Brief Description of Mean and Standard Deviation (SD)

Variables	Classification	Mean	Standard Deviation
Electricity Unit (kWh)	Low	40	27.6
	Medium	91.94	34.7
	High	258.1	33

Area of Households (m ²)	Very Low	70	205
	Low	227	243
	Medium	274.7	257
	High	2120	4805
Trip Generation (Trip/HHS)	Very Low	0.8354	0.124
	Low	1.085	0.124
	Medium	1.344	0.124
	High	1.64	0.124
	Very High	1.931	0.124

Based on the relationship of the decision tree, 13 fuzzy rules were derived, and some of the rules were assumed* based on the adjacent rules.

Rule 1: if the area of HHS is VL AND the electricity unit is L, THEN TG is VL

Rule 2: if the area of HHS is L AND the electricity unit is L, THEN TG is VL*

Rule 3: if the area of HHS is M AND the electricity unit is L, THEN TG is L*

Rule 4: if the area of HHS is H AND the electricity unit is L, THEN TG is H

Rule 5: if the area of HHS is VL AND the electricity unit is M, THEN TG is VL*

Rule 6: if the area of HHS is L AND the electricity unit is M, THEN TG is L*

Rule 7: if the area of HHS is M AND the electricity unit is M, THEN TG is M

Rule 8: if the area of HHS is H AND the electricity unit is M, THEN TG is H

Rule 9: if the area of HHS is VL AND the electricity unit is H, THEN TG is VL

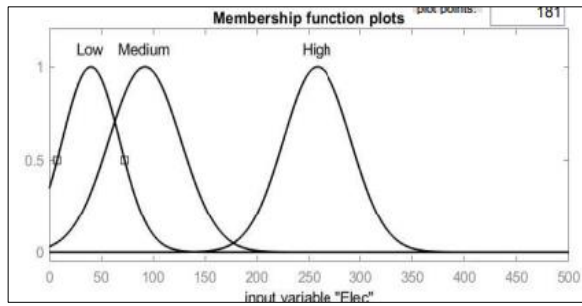
Rule 10: if the area of HHS is L AND the electricity unit is H, THEN TG is L

Rule 11: if the area of HHS is M AND the electricity unit is H, THEN TG is M*

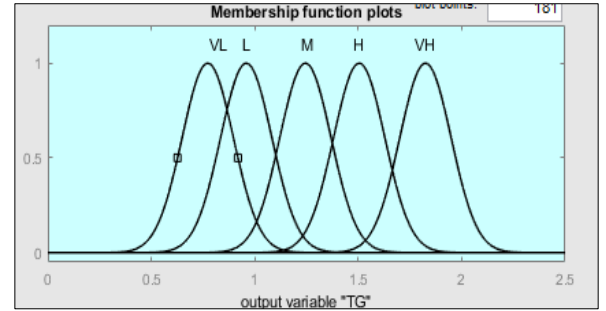
Rule 12: if the area of HHS is H AND the electricity unit is H, THEN TG is VH

Rule 13: if the area of HHS is zero OR the electricity unit is zero, THEN TG is zero

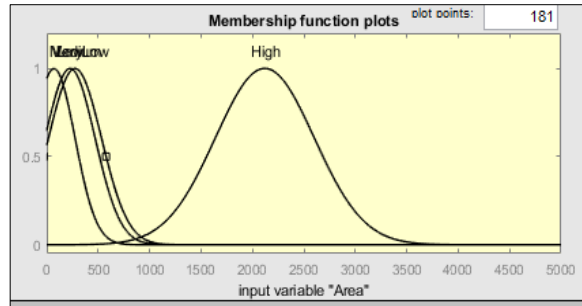
The membership function of input and output variables and the surface view of the rules and values assigned to the fuzzy model is shown in Figure 6.10.



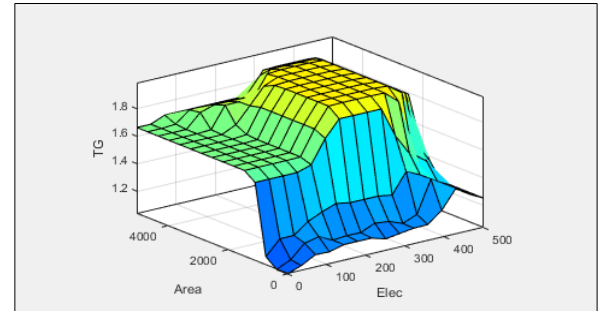
6.10a: Electricity Membership Function



6.10c: Output Variable (TG)



6.10b: Area Membership Function



6.10d: Surface View of Fuzzy Model

Figure 6.10: Fuzzy Logis Model Input and Output

As illustrated in Figure 6.10b, the fuzzy variable "Area of Households" belongs to the linguistic terms "Very Low", "Low", "Medium", and "High". Degrees of membership are represented by the y-axis, ranging from 0 to 1, while Area of Households is represented by the x-axis. A membership function is gaussian-shaped and indicates how much each value of a variable belongs to a linguistic term. Taking 100 square meters as an example, it would have a membership value of 0.5 for "Medium", 0.5 for "Very Low", and 0 for "High". According to Figure 6.10a, "Low", "Medium", and "High" are members of the fuzzy variable "Electricity Unit". The x-axis represents the values of the variable "Electricity Unit" and a y-axis represents the degree of membership. Members of the variable are gaussian-shaped functions that indicate how many values belong to which linguistic term. An example would be a value of 200 kWh having a membership value of 0 for the term "Medium", 0 for the term "Low" and 0.3 for the term "High". Figure 6.10c illustrates the fuzzy membership functions for "Trip Generation" based on the linguistic terms "Very Low", "Low", "Medium", "High", and "Very High". Membership degree is represented on the y-axis, which ranges from 0 to 1, and the variable "Trip Generation" is represented on the x-axis. Each value of the variable is assigned to a linguistic term based on how much it belongs to the membership functions. With 1.2 trips/household,

membership values would be 0 for "Medium", "Low" and "Very High", 1 for "Very Low" and 0.3 for "High". A 3D surface view of the fuzzy logic model based on rules and membership functions is shown in Figure 6.10d. Household density is shown on the x-axis, electricity units are shown on the y-axis, and trip generation is shown on the z-axis. Depending on the degree of membership, the surface is coloured red for high membership and blue for low membership. Based on the defined rules and membership functions, the fuzzy logic model maps input values to output values. As an example, an input parameter of 150 square meters for "Area of Households" and 150 kWh for "Electricity Unit" result in an output parameter of approximately 1.1 trips/household.

After preparing the fuzzy model based on the HVS 2013 values, the electricity consumption data and area of households detected using the deep learning building footprint model of USA for 2013 are entered into the model for each household to calculate the work related trip generation rate/household of Thimbrigasyaya DSD.

Table 6.6: Fuzzy Model Output for 2013 and 2019

Source	HTW	No of HHS	HTW_TG/HHS
HTS-2013	69,790	63,746	1.0947
Fuzzy Model-2013	26,401	24,397	1.0821
Fuzzy Model-2019 (Predicted)	31,376	29,468	1.0647

Table 6.6 shows the predicted value of the work trip generation rate/household versus the observed ones calculated from the HTS 2013. The result from the fuzzy logic for 2013 shows that the predicted value for work from home from the fuzzy model is closer to the relevant value calculated from the HTS data. Based on the output from the fuzzy model trip rate, it was observed that the entire home-based work trip for the number of households from HTS is 69,177 trips which gives the precision of 88.7% of the actual value (69,790 trips). Based on the developed model, the data values for 2019 and the trip rate calculated for 2019 are 1.0647. Table 6.7 shows the calculation for each zone for the study area.

Table 6.7: Total Trip Generation for Households (HHS) Based on Fuzzy Model

TAZ	TG (Fuzzy)	Total HHS Detected	Trip Rate/H HS	Total HHS in CEB	% Detected HHS	TG for the whole HHS	Trip Rate/HHS (Population)
36	522	554	0.9429	1232	45%	1162	0.9429
37	1273	1188	1.0718	2589	46%	2775	1.0718
38	108	87	1.2394	156	56%	193	1.2394

39	996	839	1.1870	1447	58%	1718	1.1870
40	389	344	1.1305	552	62%	624	1.1305
41	345	197	1.7496	477	41%	835	1.7496
42	367	748	0.4911	1375	54%	675	0.4911
43	670	854	0.7843	2255	38%	1769	0.7843
44	906	809	1.1201	2873	28%	3218	1.1201
45	1185	1102	1.0756	3298	33%	3547	1.0756
46	1104	980	1.1261	2562	38%	2885	1.1261
47	682	620	1.1004	1220	51%	1342	1.1004
48	742	678	1.0943	1432	47%	1567	1.0943
49	910	838	1.0857	2314	36%	2512	1.0857
50	818	740	1.1050	1318	56%	1456	1.1050
51	597	522	1.1441	1256	42%	1437	1.1441
52	1485	1307	1.1362	2749	48%	3123	1.1362
53	1467	1302	1.1269	3144	41%	3543	1.1269
54	2394	2115	1.1319	4841	44%	5480	1.1319
55	2368	2207	1.0728	4961	44%	5322	1.0728
56	1877	1629	1.1523	3859	42%	4447	1.1523
57	1681	1520	1.1058	3820	40%	4224	1.1058
58	1855	1695	1.0943	3695	46%	4043	1.0943
59	1660	1522	1.0907	3716	41%	4053	1.0907
Total	26,401	24,397	1.0821	57,141	45%	61,951	1.0842

It was observed that the entire home-to-work trip for the number of households from HVS is 61,951 trips which gives a precision of 88.7% of the actual value (69,790 trips). Based on the developed model, the data values for 2019 and the trip rate calculated for 2019 is 1.0647. The map illustration of the calculated

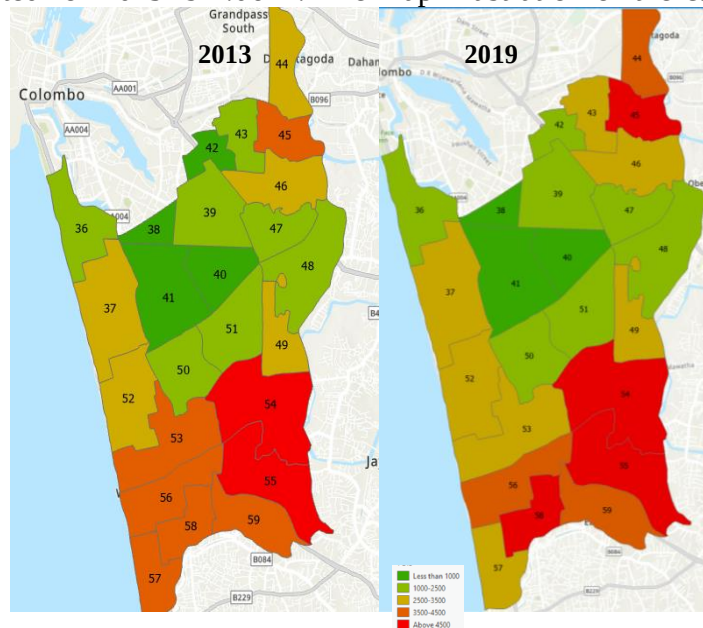


Figure 6.11: Work based Trip Generation for 2013 and 2019

2013 and 2019 is shown in Figure 6.11.

6.8 Summary and Conclusion

The study explores the alternative way to predict trip generation based on the available data. Although many studies use MLR for this purpose, as this method

is still an effective tool in prediction, it has some inherent limitations. The main issue with MLR is the reliance on relatively exact data and the availability of such data for future prediction. Therefore, a detailed survey such as HVS is needed for accurate prediction to reformulate the independent variables that have impacted the dependent variable. Most of the time, the prediction of independent variables is based on several assumptions which can affect the accuracy of the result.

Applying fuzzy logic based on available data that can be frequently updated can mitigate such disadvantages of the MLR model and its data dependency in this study. The study area is selected in Western Province, Sri Lanka, an excellent example of a developing country with limited data availability. The difficulty of collecting survey data frequently is a challenge due to financial and technological constraints. Based on the data availability of the study area, the data of electricity consumption with GPS point data of customers based on the tariff category and satellite data of the study area have been considered a valid data set as an input. Moreover, these data can be updated frequently and available at little or no cost. The electricity consumption data can be used as a surrogate variable for income. The satellite images can be used for building footprint extraction to calculate building area using a pre-trained deep learning algorithm and map match with the GPS category of residence. The relationship of these selected data was first tested with the HVS 2013, showing 67% accuracy on the training dataset (80%) and test dataset (20%) using the decision tree algorithm. Based on the relationship derived from the decision tree, fuzzy rules and variable parameters were derived and applied to the model.

The output of the trip rate calculated for home-to-work purposes from HVS 2013 and the updated model with collected data on electricity consumption and area of households shows nearly 88.7% accuracy. It is also evident that the fuzzy model can handle insufficient data with some assumptions and expert experience. In developing countries such as Sri Lanka, where limited resources are available for data collection and analysis, such high frequency, freely available data can reduce uncertainty in model application and assumptions.

Besides, the model can be further improved with the correct classification of households such as apartments, single units, and adjoined households. The current study detected footprints are applied the same in all those categories, and

the number of units inside the building (especially for apartments) and household types are not considered.

6.8.1 Detecting Building Height from the Satellite Image

Chapter 3 explains categorising domestic customers (single, adjoining, and apartments) based on GPS share points. This information has been used to detect each household unit's building height, especially in separating the apartment household units. Most of the apartment types of residence can be identified in Thimbirigasyaya DSD in Western Province, especially in the Wellawatte area. Figure 6.12 shows the number of condominium buildings approved per annum inside Colombo city.

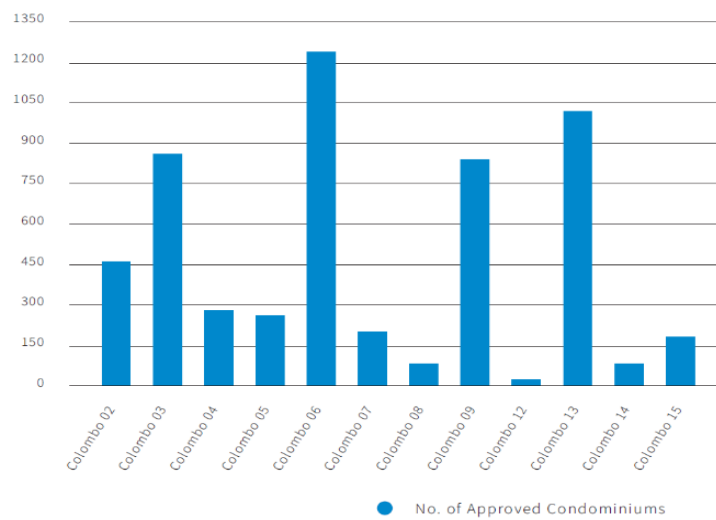


Figure 6.12: No of Condominium Units in Colombo

The shape area is the same for the single and adjoining households, and the model for calculating the shape area for each apartment unit. The following equation was developed to calculate the area of each household in the apartment building.

$$\text{Total volume (V)} = \text{Area (A)} * \text{Height (H)}$$

$$\text{The volume of per household unit (v)} = \text{Area of household unit (a)} * \text{Standard Height (h)}$$

$$\text{Total volume (V)} = n * \text{volume of per household unit (v)}$$

$$n * v = V$$

$$n * a * h = A * H$$

$$a = (A * H) / (n * h)$$

(6.8)

The equation 6.8 requires all variables to be known except the standard height of households, the area of each unit, and the height of each household unit. The following methodology has been adopted to extract each unit's building height and area (Figure 6.13).

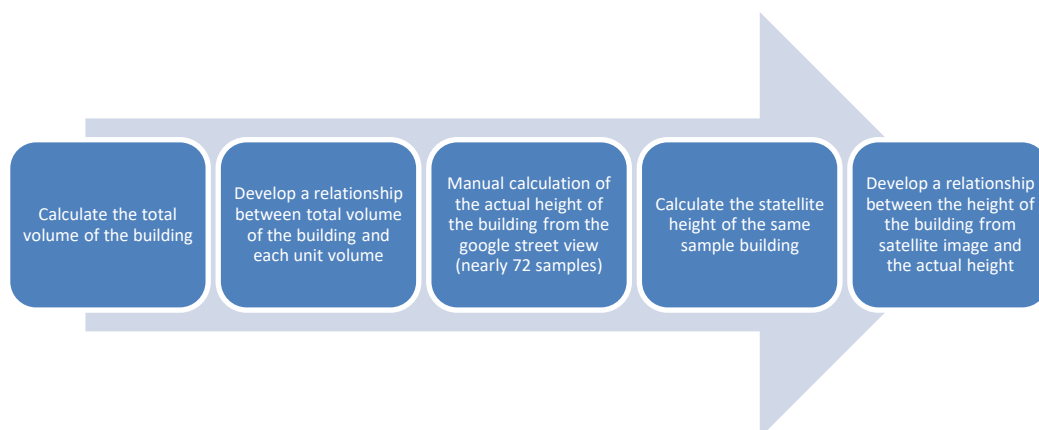
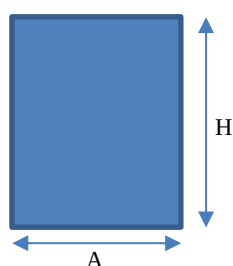


Figure 6.13: Methodology for Extracting Building Height from Google Satellite Images

The total volume of the building can be calculated from the area detected from deep learning (A), and the actual height can be calculated based on the manually counted floor area using google street map (H) (Equation 6.9).



$$\text{Total Volume (V)} = A * H \quad (6.9)$$

Table 6.8: Standard Height of the Building Applied in Sri Lanka

Floor Level	Height Level (m)
G	2.7
G+1	5.4
G+2	5.4
G+3	8.1
G+4	10.8
G+5	13.5
G+6	16.2
G+7	18.9
G+8	21.6
G+9	24.3
G+10	27
G+11	29.7
G+12	32.4
G+13	35.1
G+14	37.8
G+15	40.5
G+16	43.2
G+17	45.9
G+18	48.6

G+19	51.3
G+20	54
G+21	56.7
G+22	59.4

The Table 6.8 shows the standard height of the building applied in Sri Lanka based on the floor level of the building. The ground floor level is allocated for parking for most of the apartments in Western Province. Due to that, the G+1 and G+2 is considered the same height, assuming that the building with the above two levels is considered an apartment and the ground floor is allocated for parking. Also, the balcony area is not considered when counting the no of floors using a google street image. The sample apartment building from the street view and the counting of floors is shown in Figure 6.14. Nearly 72 buildings' actual height was calculated with the area information available based on the deep learning method.



Figure 6.14: Methodology for Sample Apartment Building from the Street View and the Counting of Floors

Furthermore, when the actual height was determined based on the counting of floors and the standard height level, the research also tried developing a model to detect the building height based on the image shadow height measured from the satellite images (Figure 6.15). The shadow of the building was measured for the same sample considered for the above has been used.



Figure 6.15: Building Height Based on the Image Shadow

The information on shape area detected by deep learning, satellite image shadow length, no of floors and actual height calculated based on the number of floors for all the samples is summarised in Table 6.9.

Table 6.9: Samples of Buildings Information

Sample No	Shape Area (m ²)	Length from Satellite Image (m)	Floor No	Actual Height (m)
Sample 1	622.25	11.05	G+7	18.9
Sample 2	1358.75	6.5	G+3	8.1
Sample 3	1422.375	6.54	G+3	8.1
Sample 4	1425	6.02	G+3	8.1
Sample 5	1492.25	5.32	G+3	8.1
Sample 6	1358.75	6.47	G+3	8.1
Sample 7	1418.5	6.7	G+3	8.1
Sample 8	1324.125	7.04	G+3	8.1
Sample 9	552.5	5.47	G+3	8.1
Sample 10	598.5	5.09	G+3	8.1
Sample 11	940.18	17.53	G+8	21.6
Sample 12	648.5	14.04	G+7	18.9
Sample 13	1127.875	10.02	G+3	8.1
Sample 14	1049.96	6.46	G+3	8.1
Sample 15	837.25	3.87	G+2	5.4
Sample 16	880.125	4.05	G+2	5.4
Sample 17	807.81	14.18	G+7	18.9
Sample 18	730.16	19	G+8	21.6
Sample 19	370.75	15.55	G+7	18.9
Sample 20	1745	15.71	G+9	24.3
Sample 21	99.75	3.5	G+1	5.4
Sample 22	462.875	8.55	G+2	5.4
Sample 23	559.75	10.68	G+5	13.5
Sample 24	1306.02	18.63	G+8	21.6
Sample 25	494.625	24.79	G+8	21.6
Sample 26	385.625	21.92	G+6	16.2
Sample 27	1041.34	27.63	G+11	29.7
Sample 28	642.375	16.06	G+6	16.2
Sample 29	405	10.52	G+2	5.4
Sample 30	403.125	11.03	G+6	16.2
Sample 31	878	29.2	G+11	29.7
Sample 32	456.625	9.39	G+6	16.2
Sample 33	317.5	8.76	G+2	5.4
Sample 34	184	7.73	G+2	5.4
Sample 35	719.53	13.02	G+7	18.9
Sample 36	717.5	12.11	G+7	18.9
Sample 37	295.125	2.49	G	2.7
Sample 38	903	19.92	G+8	21.6
Sample 39	928.5	17.52	G+7	18.9
Sample 40	427.875	12.99	G+5	13.5

Sample 41	696.75	9.61	G+6	16.2
Sample 42	796.125	15.85	G+9	24.3
Sample 43	216.375	3.97	G+2	5.4
Sample 44	666.75	12.12	G+6	16.2
Sample 45	330.75	10.6	G+6	16.2
Sample 46	133	10.04	G+3	8.1
Sample 47	562.5	8.46	G+4	10.8
Sample 48	177	7.25	G+3	8.1
Sample 49	448.13	2.54	G	2.7
Sample 50	529.5	17.78	G+7	18.9
Sample 51	745.67	13.57	G+5	13.5
Sample 52	418.62	13.57	G+5	13.5
Sample 53	418.62	13.57	G+5	13.5
Sample 54	418.62	13.57	G+5	13.5
Sample 55	418.62	13.57	G+5	13.5
Sample 56	418.62	13.57	G+5	13.5
Sample 57	2203	28.55	G+10	27
Sample 58	1169.75	23.01	G+10	27
Sample 59	448	5.82	G+2	5.4
Sample 60	106.125	6.52	G+2	5.4
Sample 61	2159.375	15.14	G+14	23.9
Sample 62	977.32	17.07	G+10	27
Sample 63	288.5	5.26	G+4	10.8
Sample 64	2981.5	32.24	G+12	32.4
Sample 65	1421.5	10.74	G+8	21.6
Sample 66	99.75	13.35	G+3	8.1
Sample 67	738.21	21.37	G+9	24.3
Sample 68	2081.25	21.28	G+9	24.3
Sample 69	718.875	21.67	G+8	21.6
Sample 70	331.25	10.835	G+4	10.8
Sample 71	285.37	5.324	G+1	5.4
Sample 72	258.125	13.31	G+5	13.5

A relationship has been developed between the actual and shadow heights from the satellite image (Table 6.10 and Equation 6.10).

$$\text{The actual height of the building} = 2.031259 + 0.99058 * (\text{height from satellite image}) \quad (6.10)$$

Table 6.10: Relationship between Actual Height and Shadow Height

SUMMARY OUTPUT

Regression Statistics								
Multiple R	0.889726							
R Square	0.791613							
Adjusted R Square	0.788636							
Standard Error	3.478896							
Observations	72							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	3218.277	3218.277	265.9136	1.54E-25			
Residual	70	847.19	12.10271					
Total	71	4065.467						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	2.031259	0.865999	2.345567	0.021837	0.304078	3.75844	0.304078	3.75844
Length (m)	0.990508	0.060742	16.30686	1.54E-25	0.869362	1.111654	0.869362	1.111654

A strong relationship with r^2 of 79% has been observed between these two variables with an acceptable p-value (Fig 6.16).

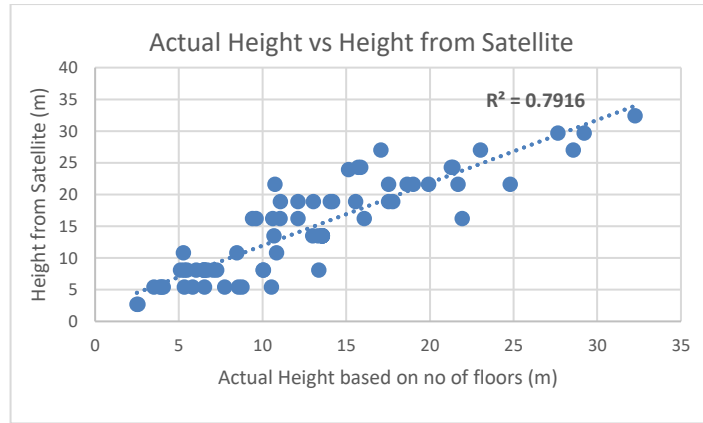


Figure 6.16: Actual Height vs Height from Satellite

The model developed in the equation can be used to calculate the actual height of the building using the satellite building height. With the information on the maximum household height of 6m (UDA, 2018) and the actual height of the building, the equation for calculating the area can be used.

Extracting satellite height information from the mosaicked image is time-consuming due to the problem of the resolution of GE images and the time difference between each image. A significant problem with the GE image is that the time taken differences between the images can cause the building slope to differ depending on the elevation angle, which makes deep learning shadow extraction ineffective. This model is practical where the high-resolution satellite image can be available for the same period.

CHAPTER 7

CONCLUSION AND RECOMMENDATION

7.1 General

Developing accurate travel demand models that effectively account for changes in land use is crucial for making informed decisions regarding transportation infrastructure investments and resource allocation in developing countries. The dynamic nature of land use changes in these countries often presents challenges in accurately capturing the demand for transport.

The study has employed a methodology for the transport model by updating the variables using HF proxy data of the recent year with the model developed using transport and travel survey data in 2013 due to the unavailability of HF data such as CDR or smart card. Among the different variables related to the trip generation model, the monthly electricity consumption and the floor area of households are considered valuable variables since they can be derived from several sources and updated frequently. Additionally, the unavailability of high-quality data, particularly in countries like Sri Lanka, further complicates the situation. To overcome these challenges, alternative approaches have been explored. Additionally, the unavailability of high-quality data, particularly in countries like Sri Lanka, further complicates the situation. To overcome these challenges, alternative approaches have been explored. Surrogate variables such as electricity consumption and satellite images have been utilized to develop travel demand models that can better account for changes in land use. These models have shown promising results and can provide valuable insights for policymakers and planners.

The use of high-frequency (HF) data, particularly electricity consumption, has enabled the calibration and validation of trip generation and attraction models, leading to more accurate estimations. However, obtaining accurate data on essential urban travel forecasting attributes like income and land use classification remains a challenge, especially in developing countries. Nevertheless, the use of alternative data sources like HF data can partially mitigate this issue.

While HF data has shown potential, incorporating other independent variables like vehicle ownership can further enhance the accuracy of these models. However, such data may not always be readily available. Future research could

focus on addressing this limitation and exploring ways to improve the models. Furthermore, the development of more complex land-use-transport interaction (LUTI) models that consider minor impacts in the system, such as pedestrian route choice and emissions, is worth pursuing. Medium-resolution satellite images obtained from Landsat have been used to calibrate and validate trip generation and attraction models. However, their limitations in providing a broader view of land use have been identified. To address this, additional data sources, including GPS data of electricity customers, have been incorporated. This approach has proven successful in accurately estimating trip generation and attraction, particularly in areas with consistent electricity consumption patterns. In conclusion, accurate travel demand models that account for changes in land use are essential for informed decision-making in transportation infrastructure investments and resource allocation. The use of HF data offers a promising approach to overcome data availability challenges. Future research should aim to develop more complex models that integrate additional independent variables and explore ways to enhance their accuracy.

7.2 Summary of Key Findings

Efficient urban planning involves designing and managing urban areas to maximise economic, social, and environmental viability while minimizing adverse effects of negative impacts. The main objective of such efficient urban planning is to create liveable and sustainable cities that meet the needs of all residents while providing connected urban to reduce urban sprawl, encouraging mixed-use developments, providing sustainable transportation, etc. Transport patterns and infrastructure needs are heavily influenced by land use and development. Consider, for example, a city with a dispersed pattern of land use, with residential areas far from employment centres and other essential services. When this occurs, private vehicle use may increase, resulting in traffic congestion, air pollution, and other harmful environmental and human health effects.

Land use and transport development concern issues in developing countries, where rapid urbanization is increasing and LUTI models require extensive data and resources to develop, implement, and maintain. Applying these models in developing countries can be challenging due to limited resources and funding.

Thus, the broader aim of this dissertation is to explore freely, or low-cost HF data related to land use and transport and develop an integrated model with available data in Sri Lanka as an example for developing countries. A secondary aim was to develop a disaggregated model with the data explored from the primary aim.

7.2.1 Objective1: Identification of suitable HF data inputs, which can support the development of LUTI models based on previous studies.

The second and third chapters are dedicated to answering the first research objective, “Identification of suitable HF data inputs, which can support the development of LUTI models based on the previous studies”. Most of the studies that have been reviewed focus on a passive collection of data sources and lack the necessary attention to establishing dynamic LUTI models that can capture short-term impacts. Additionally, applying LUTI models is primarily concentrated in developed countries and limited in developing countries. Developing countries like Sri Lanka, face limitations due to the significant data requirements and associated costs for data acquisition.

The availability of HF datasets at aggregated and disaggregated levels can help address these data availability issues and the impacts of rapid land use changes. Though there are several sources of HF data related to land use and transport modelling, only a few studies have focused on these aspects. For example, data such as CDR and the smart card contain an enormous set of individual mobility, which can be a valuable source of information for the microsimulation and complex transport model. Still, it requires additional effort in data provision, and the availability of such data is very limited in some developing countries like Sri Lanka and other low-income countries. In light of the limitations in exploring HF data, the study also incorporates the use of HF proxy data. The term HF proxy data describes alternative datasets which are available at high frequencies and can be used as surrogate variables for actual HF data. A proxy dataset allows similar patterns and trends to be examined that are seen in HF data, but from alternative sources.

This study focuses specifically on the Western Province of Sri Lanka, where data is unavailable and collecting survey data is frequently challenging due to financial and technological constraints. The study leverages electricity

consumption data with GPS point data of customers, based on their tariff category, and satellite data of the study area as input data. The electricity consumption data can be used as a surrogate variable for income, while satellite images can be used for mapping with GPS coordinates to calculate the building area. The preliminary analysis and background information about the electricity and satellite imagery data are valuable in initiating the model development process. These inputs are particularly helpful in setting the tone for the calibration of the model, thereby providing a solid foundation for subsequent research. The GPS information of each electricity customer with their tariff categories gave a promising direction for the model development. The accuracy of the GPS points of domestic customers in Western Province, Sri Lanka, and the total number of households from the 2013 census data has been validated to 94%.

7.2.2 Objective 2: Developing a methodological framework for the Land Use and Transport Integrated Model based on the identified vital HF data inputs.

Chapter 4 addressed objective 2, which was to develop a methodological framework for a LUTI model based on key household electricity consumption data inputs. The study found that household electricity consumption data is a useful explanatory variable for travel demand modelling, particularly when long-range time series data are available for frequent model updates. While household income has traditionally been used to establish the relationship between trip generation and travel behaviour, this study found that electricity consumption units also have a strong relationship with income, based on previous literature and survey data from 2013. This suggests that trip generation models could be updated more frequently using electricity consumption data.

The study also found that the consumer category information of electricity consumption data can be used to develop trip generation models, using the number of households in a particular zone as an additional independent variable. The resulting model achieved an R-squared value of 79%, and 92% accuracy between actual and predicted values based on the CoMTrans survey in 2013.

In addition, electricity consumption data was used to develop separate trip attraction models for general and industrial areas. Since land use survey data was unavailable, satellite image data with supervised classification was used to

obtain land use area information for 2013 and 2019. The developed models for general and industrial trip attraction achieved R-squared values of 70% and 79.5% for general purpose and industrial, with 94% and 82% accuracy between actual and predicted values based on the CoMTrans survey in 2013, respectively. Finally, a gravity model was used to obtain the final trip distribution for 2013 and 2019 based on the parameters considered in the CoMTrans study in 2013. Overall, the study demonstrated the value of incorporating household electricity consumption data in travel demand modelling and highlighted the potential for satellite imagery to complement traditional survey data in obtaining land use information.

7.2.3 Objective 3: Assessing the land use and transport impact on each based on a developed model by testing with different scenarios.

Chapter 5 addresses research objective 3, which is to assess the impact of land use and transport on each other based on a developed model by testing different scenarios. The study emphasizes the importance of high-frequency data in developing accurate trip generation and attraction models that can inform infrastructure development decisions, aid in evaluating transport demand, and address changes affecting land use and travel demand. The model's viability is tested by evaluating various scenarios based on future development and assessing its simplicity, ease of use, and real-time update capabilities, making it suitable for planners and decision-makers.

7.2.4 The comparison of different scenarios reveals key insights regarding work trips, passenger kilometres, average trip length, passenger hours, average speed, and costs. Scenario 5, which incorporates land use changes, modified origin-destination patterns, and a developed transport network in 2030, shows the same number of work trips as Scenarios 3 and 4. However, Scenario 4, with land use improvements, has the highest passenger kilometres, while Scenario 1 (Do Nothing) has the longest average trip length. Scenario 2 (2019 Network Improvements and OD) exhibits the highest average speed. In terms of costs, Scenario 4 incurs the highest value of time and emission costs, while Scenario 2 has the highest vehicle operating and accident costs. Overall, the comparison underscores the potential benefits of land use improvements and a developed transport network in optimising travel patterns, increasing passenger kilometres, and enhancing system efficiency, highlighting the importance of careful analysis in transportation and land use decision-making. Objective 4: Developing a disaggregated Trip Generation Model based on the available HF data and assessing the accuracy with the aggregated model.

Chapter 6 addresses research objective 4, which is to develop a disaggregated trip generation model based on available high-frequency data and assess its accuracy compared to the aggregated model. Traditional models have limitations, such as reliance on accurate and frequently available data, which can be costly and challenging to obtain in developing countries. The study explores an alternative approach by developing a fuzzy model based on electricity consumption data and building footprint information from satellite imagery. The model was developed for Thimbirigasyaya DSD, and its accuracy was assessed compared to the aggregated fuzzy model. The relationship between the variables was tested using the HVS 2013 and the decision tree algorithm, resulting in 67% accuracy for both the training and test datasets. Fuzzy rules and variable parameters were derived and applied to the model, resulting in nearly 88.7% accuracy in calculating the trip rate for home-based work purposes compared to the HVS 2013.

7.3 Recommendation and Future Research

The land use and transport integrated (LUTI) model developed in this study can be employed for analysing the impact of each other. This study aims to create a methodology that can be applied in similar demographic, social, and economic contexts in Sri Lanka and other countries. This was achieved through an integrated model that combines the expertise of urban planners, transport planners, economists, data scientists, and environmentalists. By accounting for land-use changes and their impact on travel demand, these models can provide valuable insights for policymakers and planners. These models could help identify potential transportation infrastructure projects and better allocate resources to improve mobility. The model has used freely available and low-cost novel data sources, such as HF data, which can provide real-time modelling outputs. The resulting methodology may have high publication and practical implementation value for planning institutions and consultants.

The outcome of this research is an extension of the LUTI model by integrating various socioeconomic attributes. The developed planning tools can be utilized in the proposed urban development programs in Sri Lanka in the upcoming years, with available HF data related to land use and transport. Moreover, in some developing countries, including Sri Lanka, there is a lack of access to such

data, which requires additional effort to obtain it. The research also aims to provide a promising direction for developing countries to achieve sustainable land use and transport integration by exploring more data that can be updated frequently. For example, appropriate remote sensing technology can be used in satellite imagery data for detailed land use classification. However, free or low-cost satellite data availability has been limited in developing countries. This leads to a promising direction to establish a more dynamic planning LUTI model by discovering more HF data applications and sources if the aforementioned barriers are addressed.

Nowadays, there is a noticeable shift towards incorporating activity-based transport analysis in LUTI models, as well as disaggregated simulations at individual levels. LUTI model calibration is improved by micro-level simulations, but finding relevant HF data to understand land use and transport dynamics remains challenging. By relying on certain assumptions and expert knowledge, fuzzy models are effective in handling incomplete data. Using readily available and high-frequency data, the approach seeks to reduce uncertainties associated with model applications and assumptions, which is particularly crucial in developing countries with limited data collection and analysis resources. By integrating fuzzy models and accessible data, LUTI modelling is not only made more robust but also addresses the practical limitations often present in regions with limited resources. However, the current model can be improved by correctly classifying households, such as apartments, single units, and adjoined households, and considering the number of units inside the building (especially for apartments) and household types. The study focuses primarily on HTW trip generations stems from the observed positive relationship between these variables and the effectiveness of fuzzy logic in their prediction. This limitation prompts avenues for future research to broaden the application of fuzzy models to other trip categories, such as HTE and HTO trips. By extending the analysis to encompass a more comprehensive range of HF data and trip types, researchers can offer a more holistic understanding of the potential applications and effectiveness of fuzzy models in the field of transportation modelling.

This research introduces an innovative land use and transport model approach by utilising HF data. The method is unique and has not been previously proposed

in the literature. Moreover, HF data has the potential to be applied in any part of the world due to the availability and ease of implementation with the appropriate methodology. The self-correcting power of those existing HF data sets in developing countries, such as satellite imagery, available mobility data, and land registry data, may not provide an accurate ‘proxy result’ at first. However, eventually, it will improve performance as new data batches and applications are ingested. A continuous effort must be made to improve data collection methods, enhance the quality of data, and expand the coverage of the data. It is also possible to enhance the utilisation and interpretation of the information provided by HF data by developing robust data analysis techniques and tools. Investment in data refinement and updating makes it possible for urban planners to harness the full potential of HF data to support sustainable development, inform evidence-based decision-making, and address urban and transportation challenges.

7.4 Significance and Contribution of the study

This research significantly contributes to advancing LUTI models in the context of developing countries, with a primary focus on Sri Lanka. By innovatively integrating HF data, including electricity consumption and satellite imagery, the study addresses the critical issue of data limitations prevalent in such regions. The methodology not only explores the potential of HF proxy data but also empirically tests and validates its effectiveness in creating dynamic LUTI models. The findings showcase the model's accuracy in various scenarios, emphasizing its practical applicability to urban planning. Moreover, the study proposes a novel disaggregated trip generation model based on available HF data, introducing a fuzzy model approach. This research holds significance in its interdisciplinary nature, combining expertise in urban planning, transport planning, economics, data science, and environmental studies. Beyond its contributions to academic research, the study offers valuable insights for policymakers and planners, providing a methodology that can be applied globally in similar demographic, social, and economic contexts, thereby promoting sustainable development in urban areas.

This research makes a substantial contribution to the field of urban planning and transportation modelling, particularly in developing countries like Sri Lanka.

The primary contribution lies in the innovative integration and application of HF data, such as electricity consumption and satellite imagery, to enhance the precision and dynamism of LUTI models. By addressing the inherent data limitations in developing nations, the study provides a practical solution for resource-constrained environments. The developed LUTI model, validated through empirical testing, not only offers insights into current urban dynamics but also provides a robust framework for predicting and accommodating short-term impacts. Additionally, the study introduces a novel disaggregated trip generation model, leveraging HF data, and demonstrates its accuracy, paving the way for more nuanced and localized modelling approaches. The interdisciplinary nature of the research, combining insights from urban planning, economics, data science, and environmental studies, further enriches the methodology's applicability and potential impact. Overall, this research significantly advances the understanding and application of HF data in urban and transportation planning, contributing valuable tools and insights for sustainable development in similar global contexts.

BIBLIOGRAPHY OR LIST OF REFERENCES

1. Acheampong, Ransford A., and Elisabete A. Silva. 2015. "Land use–transport interaction modeling
2. A review of the literature and future research directions." *Journal of Transport and Land Use* 8 (3):11-38.
3. Agarwal, Abhishek. 2012. "Application of fuzzy logic to model trip generation phase of sequential travel demand analysis." Bachelor of Technology in Civil Engineering, Department of Civil Engineering, National Institute of Technology Rourkela.
4. Agarwal, Chetan, G. M. Green, Morgan Grove, T. P. Evans, and Charles Schweik. 2002. *A Review and Assessment of Land-Use Change Models. Dynamics of Space, Time, and Human Choice*.
5. Ahmed, Tanjeeb, Suman Kumar Mitra, Rezwana Rafiq, and Sanjana Islam. 2020. "Trip Generation Rates of Land Uses in a Developing Country City." *Transportation Research Record* 2674 (9):412-425. doi: 10.1177/0361198120929327.
6. Alam, Akhtar, M. Sultan Bhat, and M. Maheen. 2020. "Using Landsat satellite data for assessing the land use and land cover change in Kashmir valley." *GeoJournal* 85 (6):1529-1543. doi: 10.1007/s10708-019-10037-x.
7. Alexander, C., S. Smith-Voysey, C. Jarvis, and K. Tansey. 2009. "Integrating building footprints and LiDAR elevation data to classify roof structures and visualise buildings." *Computers Environment and Urban Systems* 33 (4):285-292. doi: 10.1016/j.compenvurbsys.2009.01.009.
8. Amalan, Thivya P., Bo Du, Amal S. Kumarage, Rohan Wickramasuriya, and Pascal Perez. 2023. "High frequency data in Land use and transport integrated model: A review of sources and application." *Asian Transport Studies* 9:100092. doi: <https://doi.org/10.1016/j.eastsj.2022.100092>.
9. Anderstig, Christer, and Lars-Göran Mattsson. 1998. "Modelling Land-Use and Transport Interaction: Policy Analyses Using the IMREL Model." In, 308-328.
10. Athukorala, P. P. A. Wasantha, and Clevo Wilson. 2010. "Estimating short and long-term residential demand for electricity: New evidence from Sri Lanka." *Energy Economics* 32:S34-S40. doi: <https://doi.org/10.1016/j.eneco.2009.08.005>.
11. Badoe, Daniel A., and Judith L. Mwakalonge. 2011. "Estimating Household Trip Rates for Cross-Classification Cells with No Data: Alternative Methods and Their Performance in Prediction of Travel." *Journal of Urban Planning and Development* 137 (3):262-271. doi: doi:10.1061/(ASCE)UP.1943-5444.0000073.
12. Bain, Stuart, David A. Hensher, and Zheng Li. 2011. "R-Tresis: developing a transport model system for regional New South Wales." *Journal of Transport Geography* 19 (4):615-622. doi: <https://doi.org/10.1016/j.jtrangeo.2010.06.018>.
13. Bandara, Lasantha, and Amila Jayasinghe. 2013. *An Application of Metronamica Tool to Simulate Urban Land Use Changes: A Case Study of Galle City, Sri Lanka*.
14. Bellemans, Tom, Bruno Kochan, Davy Janssens, Geert Wets, and Harry Timmermans. 2008. "Field Evaluation of Personal Digital Assistant Enabled by Global Positioning System: Impact on Quality of Activity and Diary Data." *Transportation Research Record* 2049 (1):136-143. doi: 10.3141/2049-16.
15. Borgogno-Mondino, E., A. Lessio, L. Tarricone, V. Novello, and L. de Palma. 2018. "A comparison between multispectral aerial and satellite imagery in precision viticulture." *Precision Agriculture* 19 (2):195-217. doi: 10.1007/s11119-017-9510-0.
16. Bouanan, Y., Zerguini, S., & Gaussier, N. 2017. "MODELING AND SIMULATING HOUSEHOLDS AND FIRMS LOCATION CHOICE USING
17. AGENT-BASED MODELS: APPLICATION TO THE URBAN AREA OF BORDEAUX." *Proceedings of the 2017 Winter Simulation Conference*.

18. Brandi, Adrienne, Stefano Gori, Marialisa Nigro, and Marco Petrelli. 2014. "Development of an Integrated Transport-land Use Model for the Activities Relocation in Urban Areas." *Transportation Research Procedia* 3:374-383. doi: <https://doi.org/10.1016/j.trpro.2014.10.018>.
19. Bwambale, Andrew, Charisma F. Choudhury, and Stephane Hess. 2019. "Modelling trip generation using mobile phone data: A latent demographics approach." *Journal of Transport Geography* 76:276-286. doi: <https://doi.org/10.1016/j.jtrangeo.2017.08.020>.
20. CBSL. 2018. Economic and Social Statistica of Sri Lanka.
21. Cervero, Robert. 2013. "Linking urban transport and land use in developing countries." *Journal of Transport and Land Use* 6 (1):7-24.
22. Chang, Justin S., Dongjae Jung, Jaekyung Kim, and Taeseok Kang. 2014. "Comparative analysis of trip generation models: results using home-based work trips in the Seoul metropolitan area." *Transportation Letters* 6 (2):78-88. doi: 10.1179/1942787514Y.0000000011.
23. Chen, H., D. Wang, and C. Liu. 2020. "Towards Semantic Travel Behavior Prediction for Private Car Users." Proceedings - 2020 IEEE 22nd International Conference on High Performance Computing and Communications, IEEE 18th International Conference on Smart City and IEEE 6th International Conference on Data Science and Systems, HPCC-SmartCity-DSS 2020.
24. Chen, Z., and W. Fan. 2020. "Extracting bus transit boarding and alighting information using smart card transaction data." *Journal of Public Transportation* 22 (1):1-23. doi: 10.5038/2375-0901.22.1.3.
25. Chester, Mikhail V., and Arpad Horvath. 2009. "Environmental assessment of passenger transportation should include infrastructure and supply chains." *Environmental Research Letters* 4 (2):024008. doi: 10.1088/1748-9326/4/2/024008.
26. Dong, Honghui, Mingchao Wu, Xiaoqing Ding, Lianyu Chu, Limin Jia, Yong Qin, and Xuesong Zhou. 2015. "Traffic zone division based on big data from mobile phone base stations." *Transportation Research Part C: Emerging Technologies* 58:278-291. doi: <https://doi.org/10.1016/j.trc.2015.06.007>.
27. Donnelly, Rick. 2018. "Oregon's transportation and land use model integration program: A retrospective." *Journal of Transport and Land Use* 11 (1). doi: 10.5198/jtlu.2018.1210.
28. Du, B., and P. A. Dublanche. 2018. "Bus Bunching Identification Using Smart Card Data." 2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS), 11-13 Dec. 2018.
29. Eboli, Laura, Carmen Forciniti, and Gabriella Mazzulla. 2012. "Exploring Land Use and Transport Interaction through Structural Equation Modelling." *Procedia - Social and Behavioral Sciences* 54:107-116. doi: <https://doi.org/10.1016/j.sbspro.2012.09.730>.
30. Estima, Jacinto, Cidalia Fonte, and M. Painho. 2014. *Comparative study of Land Use/Cover classification using Flickr photos, satellite imagery and Corine Land Cover database*.
31. Figliozzi, Miguel A., and Nicholas B. Stoll. 2018. "A Study of Bus High-Resolution GPS Speed Data Accuracy." *Transportation Research Record* 2672 (8):187-198. doi: 10.1177/0361198118793273.
32. Frumkin, Howard. 2002. "Urban Sprawl and Public Health." *Public health reports (Washington, D.C. : 1974)* 117:201-17. doi: 10.1093/phr/117.3.201.
33. Ge, Qian, and Daisuke Fukuda. 2016. "Updating origin-destination matrices with aggregated data of GPS traces." *Transportation Research Part C: Emerging Technologies* 69:291-312. doi: <https://doi.org/10.1016/j.trc.2016.06.002>.
34. Giles-Corti, Billie, Fiona Bull, Matthew Knuiman, Gavin McCormack, Kimberly Van Niel, Anna Timperio, Hayley Christian, Sarah Foster, Mark Divitini, Nick Middleton, and Bryan Boruff. 2013. "The influence of urban design on neighbourhood walking following residential relocation: Longitudinal results from the RESIDE study." *Social Science & Medicine* 77:20-30. doi: <https://doi.org/10.1016/j.socscimed.2012.10.016>.
35. Gong, Z., B. Du, Z. Liu, W. Zeng, P. Perez, and K. Wu. 2020. "SD-seq2seq : A Deep Learning Model for Bus Bunching Prediction Based on Smart Card Data." 2020 29th International Conference on Computer Communications and Networks (ICCCN), 3-6 Aug. 2020.

36. Griffith, Daniel A., and Yongwan Chun. 2018. "3.01 - GIS and Spatial Statistics/Econometrics: An Overview." In *Comprehensive Geographic Information Systems*, edited by Bo Huang, 1-26. Oxford: Elsevier.
37. Guelzim, Tarik, and Mohammad S. Obaidat. 2015. "Chapter 29 - Formal methods of attack modeling and detection." In *Modeling and Simulation of Computer Networks and Systems*, edited by Mohammad S. Obaidat, Petros Nicopolitidis and Faouzi Zarai, 841-860. Boston: Morgan Kaufmann.
38. Gulhan, Gorkem, and Huseyin Ceylan. 2016. "Relation Between Land Use and Transportation Planning in the Scope of Smart Growth Strategies: Case Study of Denizli, Turkey." In.
39. Guo, D., J. Wang, Y. An, Y. Zhang, and D. Liu. 2020. "Research on Pre-processing Method of Travel Data Based on Mobile GPS: Three-step Data Pre-processing Method." *ACM International Conference Proceeding Series*.
40. Hansen, Walter G. 1959. "How Accessibility Shapes Land Use." *Journal of the American Institute of Planners* 25 (2):73-76. doi: 10.1080/01944365908978307.
41. He, K., G. Gkioxari, P. Dollár, and R. Girshick. 2017. "Mask R-CNN." 2017 IEEE International Conference on Computer Vision (ICCV), 22-29 Oct. 2017.
42. Hirun, Wirach. 2015. "TRIP RATES FOR CONDOMINIUM CONSTRUCTION PROJECT." *Journal of Urban and Environmental Engineering* 9:73-81. doi: 10.4090/juee.2015.v9n1.73-81.
43. Hu, Qiong, Wu Wenbin, Tian Xia, Qiangyi Yu, Peng Yang, Zhengguo li, and Qian Song. 2013. "Exploring the Use of Google Earth Imagery and Object-Based Methods in Land Use/Cover Mapping." *Remote Sensing* 5:6026-6042. doi: 10.3390/rs5116026.
44. Hunt, John, and John Abraham. 2003. *Design and Application of the PECAS Land Use Modeling System*.
45. Huntsinger, Leta F., Nagui M. Roupail, and Peter Bloomfield. 2013. "Trip Generation Models Using Cumulative Logistic Regression." *Journal of Urban Planning and Development* 139 (3):176-184. doi: doi:10.1061/(ASCE)UP.1943-5444.0000151.
46. Hwang, Jeong Hwan, Hyunmyung Kim, Sungjin Cho, Tom Bellemans, Won Do Lee, Keechoo Choi, Seung Hoon Cheon, and Chang-Hyeon Joh. 2017. "An examination of the accuracy of an activity-based travel simulation against smartcard and navigation device data." *Travel Behaviour and Society* 7:34-42. doi: <https://doi.org/10.1016/j.tbs.2017.01.001>.
47. Iqbal, M. S., C. F. Choudhury, P. Wang, and M. C. Gonzalez. 2014. "Development of origin-destination matrices using mobile phone call data." *TRANSPORTATION RESEARCH PART C-EMERGING TECHNOLOGIES* 40:63-74. doi: 10.1016/j.trc.2014.01.002.
48. Jayanetti, J. A. A. M, D. A. Meedeniya, M. D. N. Dilini, M. H. Wickramapala, and J. H. Madushanka. 2017. "Enhanced land cover and land use information generation from satellite imagery and foursquare data." *Proceedings of the 6th International Conference on Software and Computer Applications, Bangkok, Thailand*.
49. Jayanetti, J., Madhuka Wickramapala, Dulani Meedeniya, and M. Dilini. 2017. *Enhanced Land Cover and Land Use Information Generation from Satellite Imagery and Foursquare Data*.
50. Jayasinghe, Amila, Kazushi Sano, and Kasemsri Rattanaporn. 2017. "Application for developing countries: Estimating trip attraction in urban zones based on centrality." *Journal of Traffic and Transportation Engineering (English Edition)* 4 (5):464-476. doi: <https://doi.org/10.1016/j.jtte.2017.05.011>.
51. Johnston, Robert A., David R. Shabazian, and Shengyi Gao. 2003. "UPlan: A Versatile Urban Growth Model for Transportation Planning." *Transportation Research Record* 1831 (1):202-209. doi: 10.3141/1831-23.
52. Jokar Arsanjani, Jamal, Marco Helbich, and Mohamed Bakillah. 2013. "Exploiting volunteered geographic information to ease land use mapping of an urban landscape." *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-4/W1:51-55. doi: 10.5194/isprsarchives-XL-4-W1-51-2013.
53. Joubert, Anke, Matthias Murawski, and Markus Bick. 2021. "Measuring the Big Data Readiness of Developing Countries – Index Development and its Application to Africa." *Information Systems Frontiers*. doi: 10.1007/s10796-021-10109-9.

54. Kabakuş, Nuriye, and Ahmet Tortum. 2019. "Comparative analysis of trip generation models according to household characteristics for developed, developing and non-developed provinces in Turkey." *Sādhanā* 44. doi: 10.1007/s12046-019-1104-2.
55. Kii, Masanobu, Rolf Moeckel, and Jean-Claude Thill. 2019. "Land use, transport, and environment interactions: WCTR 2016 contributions and future research directions." *Computers, Environment and Urban Systems* 77:101335. doi: <https://doi.org/10.1016/j.compenvurbsys.2019.04.002>.
56. Kim, Min-Jeong. 2020a. "Determining the Relationship between Residential Electricity Consumption and Factors: Case of Seoul." *Sustainability* 12 (20):8590.
57. Kim, Min-Jeong. 2020b. "Determining the Relationship between Residential Electricity Consumption and Factors: Case of Seoul." *Sustainability* 12 (20). doi: 10.3390/su12208590.
58. Kitamura, Ryuichi. 2009. "A dynamic model system of household car ownership, trip generation, and modal split: model development and simulation experiment." *Transportation* 36 (6):711-732. doi: 10.1007/s11116-009-9241-9.
59. Kulkarni, G. H., and P. G. Waingankar. 2007. "Fuzzy logic based traffic light controller." 2007 International Conference on Industrial and Information Systems, 9-11 Aug. 2007.
60. Lafta, Safa Ali, and Mohammed Qadir Ismael. 2022. "Trip generation modeling for a selected sector in Baghdad city using the artificial neural network." *Journal of Intelligent Systems* 31 (1):356-369. doi: 10.1515/jisys-2022-0023.
61. Lee, Jacky, Jeffrey A. Cardille, and Michael T. Coe. 2018. "BULC-U: Sharpening Resolution and Improving Accuracy of Land-Use/Land-Cover Classifications in Google Earth Engine." *Remote Sensing* 10 (9). doi: 10.3390/rs10091455.
62. Lefebvre, Antoine, Christophe Sannier, and Thomas Corpetti. 2016. "Monitoring Urban Areas with Sentinel-2A Data: Application to the Update of the Copernicus High Resolution Layer Imperviousness Degree." *Remote Sensing* 8 (7). doi: 10.3390/rs8070606.
63. Lei, D., X. Chen, L. Cheng, L. Zhang, S. V. Ukkusuri, and F. Witlox. 2020. "Inferring temporal motifs for travel pattern analysis using large scale smart card data." *Transportation Research Part C: Emerging Technologies* 120. doi: 10.1016/j.trc.2020.102810.
64. Li, D., J. Cao, R. Li, and L. Wu. 2020. "A Spatio-Temporal Structured LSTM Model for Short-Term Prediction of Origin-Destination Matrix in Rail Transit with Multisource Data." *IEEE Access* 8:84000-84019. doi: 10.1109/ACCESS.2020.2991982.
65. Lierop, Dea, Geneviève Boisjoly, Emily Grise, and Ahmed El-Geneidy. 2017. "Evolution in Land Use and Transportation Research." In, 110-131.
66. Lopes, André Soares, Camila Bandeira Cavalcante, David Sousa Vale, and Carlos Felipe Grangeiro Loureiro. 2020. "Convergence of planning practices towards LUT integration: Seeking evidences in a developing country." *Land Use Policy* 99:104842. doi: <https://doi.org/10.1016/j.landusepol.2020.104842>.
67. López-Ospina, Héctor, Cristián E. Cortés, Juan Pérez, Romario Peña, Juan Carlos Figueroa-García, and Jorge Urrutia-Mosquera. 2022. "A maximum entropy optimization model for origin-destination trip matrix estimation with fuzzy entropic parameters." *Transportmetrica A Transport Science*. doi: <https://doi.org/10.1080/23249935.2021.1913257>.
68. Malinowski, Radek, Stanisław Lewiński, Marcin Rybicki, Ewa Gromny, Małgorzata Jenerowicz, Michał Krupiński, Artur Nowakowski, Cezary Wojtkowski, Marcin Krupiński, Elke Krätzschar, and Peter Schauer. 2020. "Automated Production of a Land Cover/Use Map of Europe Based on Sentinel-2 Imagery." *Remote Sensing* 12 (21). doi: 10.3390/rs12213523.
69. Mamei, M., N. Bicocchi, M. Lippi, S. Mariani, and F. Zambonelli. 2019. "Evaluating Origin-Destination Matrices Obtained from CDR Data." *SENSORS* 19 (20). doi: 10.3390/s19204470.
70. Michael, Anderson, and Abdullah Yasir. 2005. "A Small Community Through Trip Rate Methodology."

71. Miller, Eric. 2018. "Integrated Urban Modelling: Past, Present and Future." *The Journal of Transport and Land Use* 11 (1):387-399.
72. Miller, Eric, and Paul Salvini. 1998. *The Integrated Land Use, Transportation, Environment (ILUTE) Modeling System: A Framework*.
73. Millward, Hugh, Mohammad Hesam Hafezi, and Naznin Sultana Daisy. 2019. "Activity travel of population segments grouped by daily time-use: GPS tracking in Halifax, Canada." *Travel Behaviour and Society* 16:161-170. doi: <https://doi.org/10.1016/j.tbs.2019.05.005>.
74. Miyamoto, Kazuaki, Varameth Vichiensan, Nao Engineer, and Keiichi Kitazume. 2007. "Applications of RURBAN Integrated with a Transport Model in Detailed Zone System."
75. Modi, Kevin B. 2001. "Transportation Planning Models : A Review."
76. Moeckel, Rolf, Carlos Llorca Garcia, Ana T. Moreno Chou, and Matthew Bediako Okrah. 2018. "Trends in integrated land-use/transport modeling
77. An evaluation of the state of the art." *Journal of Transport and Land Use* 11 (1):463-476.
78. Mukherjee, Jaideep, and B. Raghuram Kadali. 2022. "A comprehensive review of trip generation models based on land use characteristics." *Transportation Research Part D: Transport and Environment* 109:103340. doi: <https://doi.org/10.1016/j.trd.2022.103340>.
79. Nasibov, Efendi, Ahmet Can Diker, and Elvin Nasibov. 2016. "A multi-criteria route planning model based on fuzzy preference degrees of stops." *Applied Soft Computing* 49:13-26. doi: <https://doi.org/10.1016/j.asoc.2016.07.052>.
80. Nassir, Neema, Alireza Khani, Sang Gu Lee, Hyunsoo Noh, and Mark Hickman. 2011. "Transit Stop-Level Origin–Destination Estimation through Use of Transit Schedule and Automated Data Collection System." *Transportation Research Record* 2263 (1):140-150. doi: 10.3141/2263-16.
81. Navarro Cerrillo, Rafael M., Guillermo Palacios Rodríguez, Inmaculada Clavero Rumbao, Miguel Á Lara, Francisco J. Bonet, and Francisco-Javier Mesas-Carrascosa. 2020. "Modeling Major Rural Land-Use Changes Using the GIS-Based Cellular Automata Metronamica Model: The Case of Andalusia (Southern Spain)." *ISPRS International Journal of Geo-Information* 9 (7). doi: 10.3390/ijgi9070458.
82. Nguyen, Minh Hieu, and Jimmy Armoogum. 2020. "Hierarchical process of travel mode imputation from GPS data in a motorcycle-dependent area." *Travel Behaviour and Society* 21:109-120. doi: <https://doi.org/10.1016/j.tbs.2020.06.006>.
83. Nitze, Ingmar, Brian Barrett, and Fiona Cawkwell. 2015. "Temporal optimisation of image acquisition for land cover classification with Random Forest and MODIS time-series." *International Journal of Applied Earth Observation and Geoinformation* 34:136-146. doi: <https://doi.org/10.1016/j.jag.2014.08.001>.
84. Niu, Fangqu, Fang Wang, and Mingxing Chen. 2019. "Modelling urban spatial impacts of land-use/ transport policies." *Journal of Geographical Sciences* 29 (2):197-212. doi: 10.1007/s11442-019-1592-3.
85. Ortúzar, J.d.D , and L.G Willumsen. 2011. "Trip Generation Modelling." In *Modelling Transport*, 139-173.
86. Pepe, Massimiliano, Domenica Costantino, Vincenzo Saverio Alfio, Gabriele Voza, and Elena Cartellino. 2021. "A Novel Method Based on Deep Learning, GIS and Geomatics Software for Building a 3D City Model from VHR Satellite Stereo Imagery." *ISPRS International Journal of Geo-Information* 10 (10):697.
87. Pourmoradnasseri, M., K. Khoshkhal, A. Lind, A. Hadachi, and Ieee. 2019. "OD-Matrix Extraction based on Trajectory Reconstruction from Mobile Data." 2019 INTERNATIONAL CONFERENCE ON WIRELESS AND MOBILE COMPUTING, NETWORKING AND COMMUNICATIONS (WIMOB).
88. Rajmohan, K., and J. Weerahewa. 2007. "Household Energy Consumption Patterns in Sri Lanka." *Sri Lankan Journal of Agricultural Economics* 09:205967.
89. Rajmohan, K., and Jeevika Weerahewa. 2010. "Household Energy Consumption Patterns in Sri Lanka." *Sri Lankan Journal of Agricultural Economics* 9. doi: 10.4038/sjae.v9i0.1833.

90. Rapuzzi, A., C. Nattero, R. Pelich, M. Chini, P. Campanella, and Ieee. 2020. "CNN-BASED BUILDING FOOTPRINT DETECTION FROM SENTINEL-1 SAR IMAGERY." *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Electr Network, Sep 26-Oct 02.
91. Rasouli, Soora, and Harry Timmermans. 2014. "Activity-based models of travel demand: promises, progress and prospects." *International Journal of Urban Sciences* 18 (1):31-60. doi: 10.1080/12265934.2013.835118.
92. Rassafi, Amir Abbas, Roohollah Rezaei, and Mehdi Hajizamani. 2012. "PREDICTING URBAN TRIP GENERATION USING A FUZZY EXPERT SYSTEM." *Iranian Journal of Fuzzy Systems* 9 (3):127-146. doi: 10.22111/ijfs.2012.151.
93. Rezaei, M., H. Arefi, H. Rastiveis, and M. Sajadian. 2018. "Building Extraction and Modeling Using LiDAR Point Clouds Imaging on Two-Dimensional Surface." *ISSGE* 7 (3):139-150.
94. Sadeghian, Paria, Xiaoyun Zhao, Arman Golshan, and Johan Håkansson. 2022. "A stepwise methodology for transport mode detection in GPS tracking data." *Travel Behaviour and Society* 26:159-167. doi: <https://doi.org/10.1016/j.tbs.2021.10.004>.
95. Salini, P. S., A. Kedia, S. Dhulipala, K. Saw, and B. K. Katti. 2017a. "Spatial distribution of urban trips in recently expanded Surat city through Fuzzy Logic with various clustering Techniques: A case study of typical metropolitan city in India." *Transportation Research Procedia*.
96. Salini, P. S., A. Kedia, S. Dhulipala, Krishna Saw, and B. K. Katti. 2017b. "Spatial distribution of urban trips in recently expanded Surat city through Fuzzy Logic with various clustering Techniques: A case study of typical metropolitan city in India." *Transportation Research Procedia* 25:2396-2407. doi: <https://doi.org/10.1016/j.trpro.2017.05.245>.
97. Salvini, Paul, and Eric J. Miller. 2005. "ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems." *Networks and Spatial Economics* 5 (2):217-234. doi: 10.1007/s11067-005-2630-5.
98. Sbair, A., and F. Ghadi. 2018. "Estimation of origin destination matrix using fuzzy logic." 2018 4th International Conference on Optimization and Applications (ICOA), 26-27 April 2018.
99. Shield, C. M., and P. T. Blythe. 1997. "The Use of Smart Cards in Transportation Systems: A European Perspective." *IFAC Proceedings Volumes* 30 (8):257-262. doi: [https://doi.org/10.1016/S1474-6670\(17\)43833-X](https://doi.org/10.1016/S1474-6670(17)43833-X).
100. Shirowzhan, S., S. Lim, J. Trinder, H. Li, and S. M. E. Sepasgozar. 2020. "Data mining for recognition of spatial distribution patterns of building heights using airborne lidar data." *Advanced Engineering Informatics* 43:101033. doi: <https://doi.org/10.1016/j.aei.2020.101033>.
101. Silva, Cecília, and Paulo Pinho. 2006. "A METHODOLOGY TO ASSES THE CONTRIBUTION OF THE LAND USE AND TRANSPORT SYSTEMS TO SUSTAINABLE URBAN MOBILITY."
102. Silva, Cecília, José Pedro Reis, and Paulo Pinho. 2014. "How Urban Structure Constrains Sustainable Mobility Choices: Comparison of Copenhagen and Oporto." *Environment and Planning B: Planning and Design* 41 (2):211-228. doi: 10.1068/b37138.
103. Sirko, Wojciech, Sergii Kashubin, Marvin Ritter, Abigail Annkah, Yasser Salah Edine Bouchareb, Yann Dauphin, Daniel Keyzers, Maxim Neumann, Moustapha Cissé, and John Quinn. 2021. "Continental-Scale Building Detection from High Resolution Satellite Imagery." *ArXiv abs/2107.12283*.
104. Small, Kenneth, and Erik Verhoef. 2007. *Urban Transportation Economics*. Vol. 70.
105. Soltani, Ali, Michael Tanko, Matthew I. Burke, and Reza Farid. 2015. "Travel Patterns of Urban Linear Ferry Passengers: Analysis of Smart card Fare Data for Brisbane, Queensland, Australia." *Transportation Research Record* 2535 (1):79-87. doi: 10.3141/2535-09.
106. Spyridon, Spyridon, Demetris Stathakis, Michael Lutz, and Chrisa Tsinaraki. 2016a. "Using Foursquare place data for estimating building block use." *Environment and Planning B: Urban Analytics and City Science* 44 (4):693-717. doi: 10.1177/0265813516637607.

107. Spyratos, Spyridon, Dimitris Stathakis, Michael Lutz, and Chrisa Tsinaraki. 2016b. "Using Foursquare place data for estimating building block use." *Environment and Planning B: Planning and Design* 44. doi: 10.1177/0265813516637607.
108. Still, Ben. 1996. "The Importance of Transport Impacts on Land Use in Strategic Planning." *Traffic Engineering and Control*:564-71.
109. Still, Ben, Tony May, and A Bristo. 1999. "TRANSPORT IMPACTS ON LAND USE: POTENTIAL METHODS AND THEIR RELEVANCE OF STRATEGIC PLANNING." *World Transport Research: Selected Proceedings of the 8th World Conference on Transport Research* World Conference on Transport Research Society.
110. Stopher, Peter R, and Kathie G McDonald. 1983. "Trip generation by cross-classification: an alternative methodology." *Transportation Research Record* 944:84-91.
111. Takyi, ISAAC K. 1990. "Trip generation analysis in a developing country context." *Transportation Research Record* 1285:9-21.
112. Talukdar, Swapan, Pankaj Singha, Susanta Mahato, Shahfahad, Swades Pal, Yuei-An Liou, and Atiqur Rahman. 2020. "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review." *Remote Sensing* 12 (7). doi: 10.3390/rs12071135.
113. Todd, Litman. 2019. "Autonomous vehicle implementation predictions: implications for transport planning." doi: <https://doi.org/APO-261151>.
114. Tsumura, Yuma, Yoshihisa Asada, Hiroshi Kanasugi, Ayumi Arai, Ryosuke Shibasaki, Hirohisa Kawaguchi, and Kaoru Yamada. 2022. "Examining potentials and practical constraints of mobile phone data for improving transport planning in developing countries." *Asian Transport Studies* 8:100043. doi: <https://doi.org/10.1016/j.eastsj.2021.100043>.
115. Tunc, Ilhan, Atakan Yasin Yesilyurt, and Mehmet Turan Soylemez. 2021. "Different Fuzzy Logic Control Strategies for Traffic Signal Timing Control with State Inputs." *IFAC-PapersOnLine* 54 (2):265-270. doi: <https://doi.org/10.1016/j.ifacol.2021.06.032>.
116. Veldhuisen, Jan, Harry Timmermans, and Loek Kapoen. 2000. "RAMBLAS: A Regional Planning Model Based on the Microsimulation of Daily Activity Travel Patterns." *Environment and Planning A: Economy and Space* 32 (3):427-443. doi: 10.1068/a325.
117. Vorraa, T. 2004. "Cube Land: Integration Between Land Use And Transportation." *Urban Transport X* 75:7.
118. Waddell, Paul. 2002. "UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning." *Journal of the American Planning Association* 68:297-314. doi: 10.1080/01944360208976274.
119. Waddell, Paul, Alan Borning, Hana Sevcikova, and David Socha. 2006. *Opus (the Open Platform for Urban Simulation) and UrbanSim 4*.
120. Wagner, Peter, and Michael Wegener. 2007. "Urban Land Use, Transport and Environment Models." *disP - The Planning Review* 43 (170):45-56. doi: 10.1080/02513625.2007.10556988.
121. Wang, J. F., C. Q. Zeng, B. Lehrbass, and Ieee. 2012. "BUILDING EXTRACTION FROM LIDAR AND AERIAL IMAGES AND ITS ACCURACY EVALUATION." *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Munich, GERMANY, Jul 22-27.
122. Wang, P., R. Zhang, S. Sun, M. Gao, B. Zheng, D. Zhang, Y. Zhang, G. R. Carmichael, and H. Zhang. 2023. "Aggravated air pollution and health burden due to traffic congestion in urban China." *Atmos. Chem. Phys.* 23 (5):2983-2996. doi: 10.5194/acp-23-2983-2023.
123. Wegener, Michael. 2004. "Overview of land use transport models." *Journal of Discrete Algorithms - JDA* 5.
124. Wilson, Alan. 2010. "Entropy in Urban and Regional Modelling: Retrospect and Prospect." *Geographical Analysis* 42:364-394. doi: 10.1111/j.1538-4632.2010.00799.x.

125. Xu, Weici, and Yupo Chan. 1993. "Estimating an origin-destination matrix with fuzzy weights." *Transportation Planning and Technology* 17 (2):127-144. doi: 10.1080/03081069308717506.
126. Yan, Yingwei, Michael Schultz, and Alexander Zipf. 2019. "An exploratory analysis of usability of Flickr tags for land use/land cover attribution." *Geo-spatial Information Science*. doi: 10.1080/10095020.2018.1560044.
127. Zaragozaí, Benito, Sergio Trilles, Aaron Gutiérrez, and Daniel Miravet. 2021. Development of a Common Framework for Analysing Public Transport Smart Card Data. *Energies* 14 (19). doi:10.3390/en14196083.
128. Zhang, K., J. Yan, and S. C. Chen. 2006. "Automatic Construction of Building Footprints From Airborne LIDAR Data." *IEEE Transactions on Geoscience and Remote Sensing* 44 (9):2523-2533. doi: 10.1109/TGRS.2006.874137.
129. Zhong, Ming, John Douglas Hunt, and John Edward Abraham. 2007. "Design and Development of a Statewide Land Use Transport Model for Alberta." *Journal of Transportation Systems Engineering and Information Technology* 7 (1):79-89. doi: [https://doi.org/10.1016/S1570-6672\(07\)60010-5](https://doi.org/10.1016/S1570-6672(07)60010-5).
130. Zhu, Yi, Mi Diao, Joseph Ferreira, and Chris Zegras. 2018. "An integrated microsimulation approach to land-use and mobility modeling." *Journal of Transport and Land Use* 11. doi: 10.5198/jtlu.2018.1186.
131. Zondag, Barry, Michiel de Bok, Karst T. Geurs, and Eric Molenwijk. 2015. "Accessibility modeling and evaluation: The TIGRIS XL land-use and transport interaction model for the Netherlands." *Computers, Environment and Urban Systems* 49:115-125. doi: <https://doi.org/10.1016/j.compenvurbsys.2014.06.001>.

Appendix 1

Distance Matrix

2013 OD Matrix

174

Appendix 3 RDA Projects

Road No	Road Name	Section-From	From-km	Section-To	To-km	Total Length (km)	Province	Date of Commencement	Date of Completion	Revised Date of Completion	Year of Completion
B216	Kesbewa - Kindelpitiya - Bandaragama	Kesbewa	0	Bandaragama	11.78	11.78	WP	15/10/2012	14/10/2014		2014
B458	Waskaduwa - Bandaragama	Waskaduwa	0	Bandaragama	12.07	12.07	WP	15/10/2012	14/10/2014		2014
B123	Galagedera - Horana	Galagedera	0	Horana	19.05	19.05	WP	15/11/2012	14/11/2014		2014
B285	Meepe - Ingiriya	Meepe	0	Ingiriya	18.13	18.13	WP	25/06/2013	21/06/2015	30/09/2016	2016
B207	Katukurunda - Neboda	Nagoda	2.7	Neboda	15.9	13.2	WP	25/07/2013	24/07/2015	31/07/2017	2017
Non-RDA	Roads Under Akuregoda Defence Headquarters Development		0		10.98	10.98	WP	27/11/2013	27/11/2015	1/9/2017	2017
B400	Seeduwa - Udugampola	Seeduwa	0	Udugampola	11.7	11.7	WP	17/10/2013	16/04/2015		2015
B598	Seeduwa - Katunayake	Seeduwa	0	Katunayake	3.9	3.9	WP	17/10/2013	16/04/2015		2015
Non-RDA	Kelanimulla - Mulleriya - Koswatta	Viaduct bridge	0		0.6	0.6	WP	26/11/2013	26/05/2015	23/08/2015	2015
B408	Talduwa - Meewitigammana	Talduwa	0	Meewitigammana	17.3	17.3	Sab/WP	17/02/2014	16/02/2016		2016
B460	Wattala - Mahara	Wattala	0	Mahara	7.24	7.24	WP	4/4/2014	3/7/2015	30/12/2015	2015
A002	Colombo - Galle - Hambantota - Wellawaya	Cross Jn	17.2	Walana Jn	25.2	8	WP	4/7/2014	4/1/2016	30/09/2017	2017
B240	Kotte - Bope	Polduwa Jn	0	Thalahena	3.7	3.7	WP	4/7/2014	3/10/2015	15/06/2016	2016
B047	Battaramulla - Pannipitiya	Battaramulla	0	Thalawathugoda	3.5	3.5	WP	4/7/2014	3/10/2015	15/06/2016	2016
B151	Hendala - Hunupitiya	Hendala	0	Hunupitiya	4.02	4.02	WP	10/7/2014	9/7/2015	11/12/2015	2015
E02	Outer Circular Highway	Kadawatha	8.65	Kaduvela	17.5	8.85	WP	9/1/2012	8/1/2015	18/07/2015	2015
E02	Outer Circular Highway	Kottawa	17.5	Kaduvela	28.5	11	WP	1/12/2009	30/04/2014	13/03/2015	2015
B120	Etulkotte - Mirihana - Kohuwela	Etulkotte	0	Kohuwela	5.95	5.95	WP	10/12/2013	9/9/2014	31/01/2015	2015
AB005	Approach Road to Sri Jayawardanapura Hospital	Thalapathpitiya Jn	1.2	Sri Jayawardhanapura Hospital	2.6	1.4	WP	23/12/2014	22/03/2016	15/07/2016	2016
B096	Dematagoda - Wellampitiya	Dematagoda Jn	0	Wellampitiya Jn	1.93	1.93	WP	23/12/2014	22/03/2016	15/07/2016	2016
B232	Kolonnawa - Yakbedde	Kolonnawa Jn	0	Yakbedde	2.41	2.41	WP	23/12/2014	22/03/2016	15/07/2016	2016
B469	Welikada - Kohilawatta	Welikada Jn	0	Kohilawatta Jn	6.03	6.03	WP	23/12/2014	22/03/2016	15/07/2016	2016
Non-RDA	Aluth Mawatha and Centre Road	0.54th km	0.54	1.54th km	1.54	1	WP	23/12/2014	22/03/2016	15/07/2016	2016
A001	Colombo - Kandy	Peliyagoda	5.86	Kiribathgoda	13.1	7.24	WP	13/12/2013	12/12/2014		2014
B207	Katukurunda - Neboda	Katukurunda	0	Nagoda	2.72	2.72	WP	25/05/2012	17/08/2013	31/01/2014	2014
A004	Colombo - Ratnapura - Wellawaya - Batticaloa	Nugegoda	7.92	Homagama	22.28	14.36	WP	9/12/2013	9/6/2015	30/11/2015	2015
B084	Colombo - Horana	Pepiliyana Jn	4.1	10.60th km	10.6	6.5	WP	28/03/2013	12/7/2015	28/02/2016	2016
B084	Colombo - Horana	13.40th km	13.4	13.80th km	13.8	0.4	WP	28/03/2013	12/7/2015	28/02/2016	2016

B639	Piliyandala Bypass	0	0	2.86th km	2.86	2.86	WP	28/03/2013	12/7/2015	18/11/2014	2014
B208	Katunayake - Veyangoda	Airport Jn	0	4th km	4	4	WP	7/1/2015	31/03/2015	30/06/2015	2015
B208	Katunayake - Veyangoda	4th km	4	9.50th km	9.5	5.5	WP	7/1/2015	31/03/2015	30/06/2015	2015
B363	Pelawatta - Kankotayawatta - Tinniyawela - Morawaka	Pelawatta	0	20th km	20	20	WP	1/7/2011	30/06/2013	15/03/2014	2014
B363	Pelawatta - Kankotayawatta - Tinniyawela - Morawaka	20th km	20	Neluwa	29	9	WP/SP	1/7/2011	30/06/2013	15/03/2014	2014
B322	Negombo - Giriulla	3.70th km	3.7	12th km	12	8.3	WP	2/5/2013	30/09/2013	31/01/2014	2014
B421	Tiruwanaketiya - Agalawatte	Kalawana	33.45	Agalawatte	67.8	34.35	Sab/WP	12/12/2011	11/12/2013	15/07/2014	2014
B445	Veyangoda - Ruwanwella	Urapola	12.4	Ruwanwella	32.2	19.8	WP/Sab	24/10/2014	23/04/2017	30/05/2017	2017
B295	Moratuwa - Piliyandala	Kospelena Jn	2.62	Piliyandala	5.12	2.5	WP	21/04/2015	15/10/2016	19/06/2017	2017
B389	Ratmalana - Mirihana	Pepiliyana Jn	5.7	Mirihana Jn	7.84	2.14	WP	21/04/2015	15/10/2016	19/06/2017	2017

Appendix 4 2019 OD Matrix

[illegible]

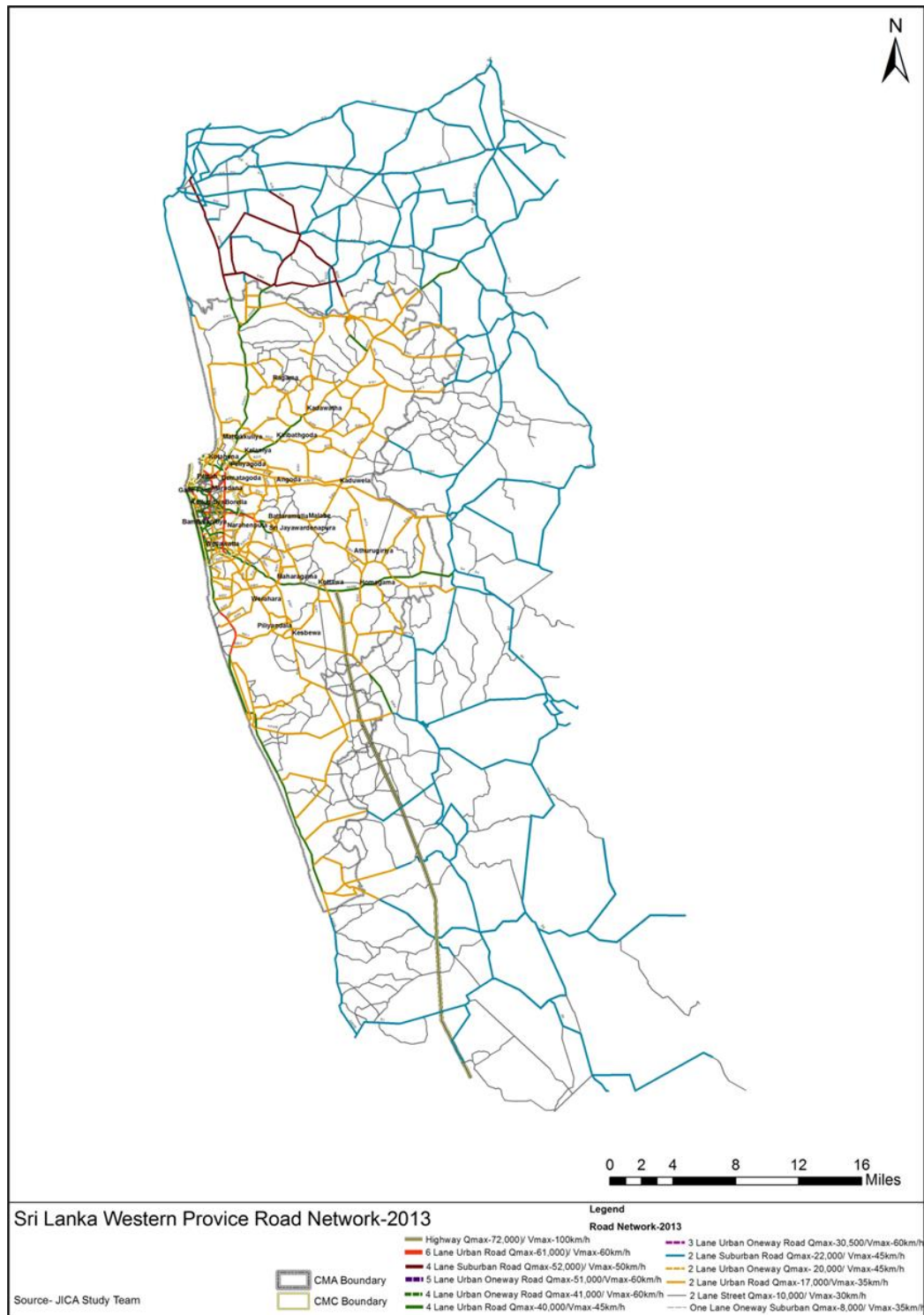
[illegible]

Matrix (Scenario 4) Land use Development

[illegible]

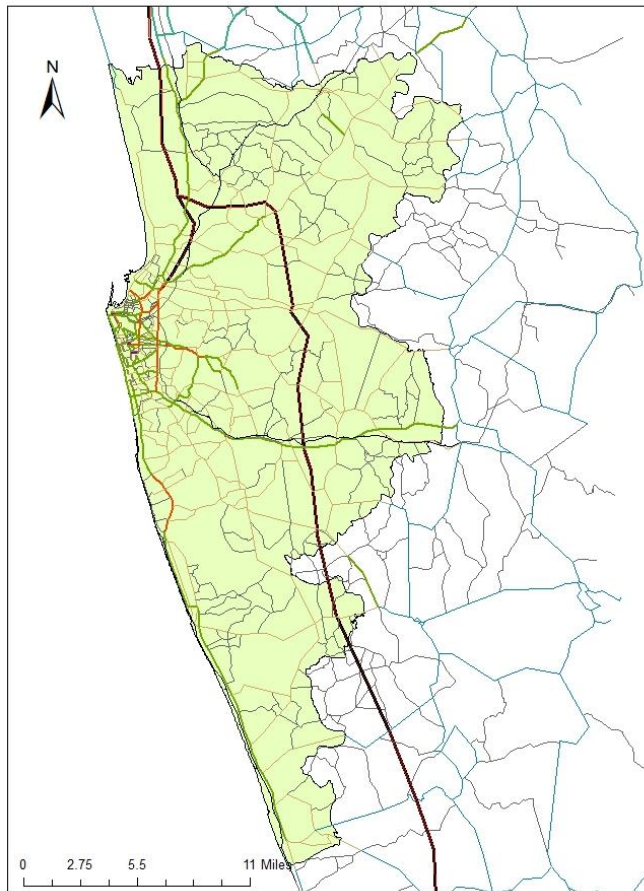
Appendix 7

Road Network 2013

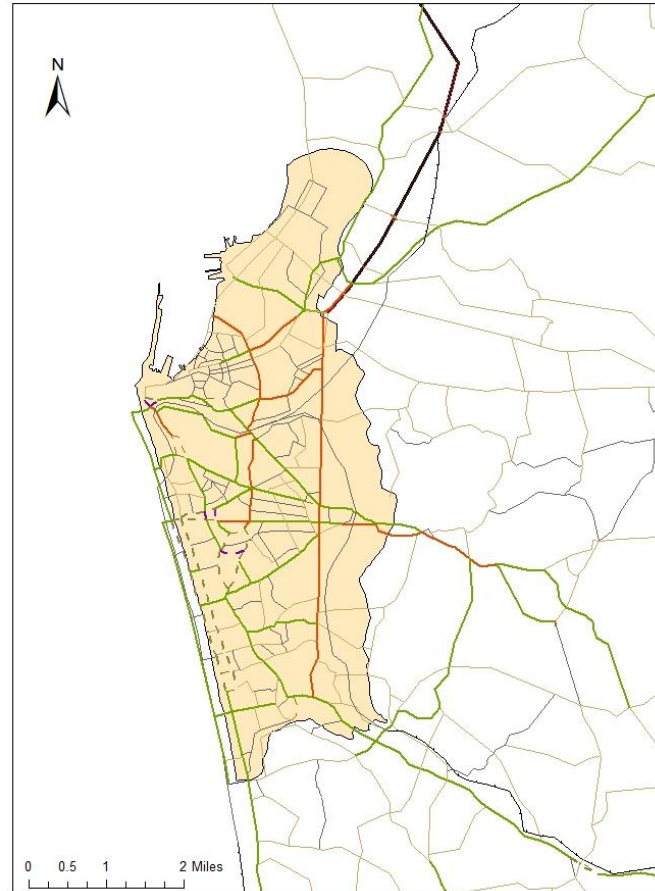


Appendix 8 Road Network 2019 and Capacity

WP Road Capacity Network_2019



WP Road Capacity Network_2019



WP Road Capacity Network_2019

- One Lane Suburban Oneway (Qmax-8000/Vmax-35km/h)
- Railway Line (Qmax-9000/Vmax-65km/h)
- Two Lane Street (Qmax-10000/Vmax-30km/h)
- Two Lane Urban (Qmax-17000/Vmax-35km/h)
- - - Two Lane Urban Oneway (Qmax-20000/Vmax-45km/h)
- Two Lane Suburban (Qmax-22000/Vmax-45km/h)
- - - Three Lane Urban Oneway (Qmax-30500/Vmax-60km/h)
- Four Lane Urban (Qmax-40000/Vmax-45km/h)
- - - Four Lane Urban Oneway (Qmax-41000/Vmax-60km/h)
- Four Lane Suburban (Qmax-52000/Vmax-50km/h)
- - - Five Lane Urban Oneway (Qmax-51000/Vmax-60km/h)
- Horana Expressway (Qmax-60000/Vmax-60km/h)
- Six Lane Urban (Qmax-61000/Vmax-60km/h)
- Four Lane Expressway (Qmax-72000/Vmax-100km/h)
- CMC Boundary

	Observation	CMC Capacity (30%)	Observation
	Q day	Q day	Velocity
	PCU/2dir/day	PCU/2dir/day	km/h
4 Lane Expressway	72,000	50400	100
6 Lane Urban	61,000	42700	60
6 Lane (Suburban)	79,000	55300	60
4 Lane (Urban)	40,000	28000	45
4 Lane (Suburban)	52,000	36400	50
2 Lane (Urban)	17,000	11900	35
2 Lane (Suburban)	22,000	15400	45
2 Lane (Street)	10,000	7000	30
5 Lane (Urban, One way)	51,000	35700	60
4 Lane (Urban, One way)	41,000	28700	60
3 Lane (Urban, One way)	30,500	21350	60
2 Lane (Urban, One way)	20,000	14000	45
1 Lane (Urban, One way)	8,000	5600	35

Appendix 9

Deep Learning Model- USA

Building Footprint-USA

```
{
  "ModelParameters": {
    "pointrend": false,
    "backbone": "resnet50",
    "backend": "pytorch"
  },
  "Framework": "arctgis.learn.models._inferencing",
  "ModelConfiguration": "_maskrcnn_inferencing",
  "InferenceFunction": "ArcGISInstanceDetector.py",
  "ModelType": "InstanceDetection",
  "MaskRCNNkwarg": (Jayasinghe, Sano et al. 2017),
  "ExtractBands": [
    0,
    1,
    2
  ],
  "SupportsVariableTileSize": true,
  "Classes": [
    {
      "Value": 1,
      "Name": "1",
      "Color": [
        187,
        194,
        137
      ]
    }
  ],
  "ArcGISLearnVersion": "1.9.0",
  "ModelFile": "usa_building_footprints.pth",
  "ImageHeight": 512,
  "ImageWidth": 512,
  "ImageSpaceUsed": "MAP_SPACE",
  "LearningRate": "1.0000e-04",
  "ModelName": "MaskRCNN",
  "backend": "pytorch",
  "average_precision_score": {
    "1": 0.7187471585753209
  },
  "resize_to": null,
  "IsMultispectral": false
}
```

Appendix 10

Deep Learning Model- Africa

Building Footprint-Africa

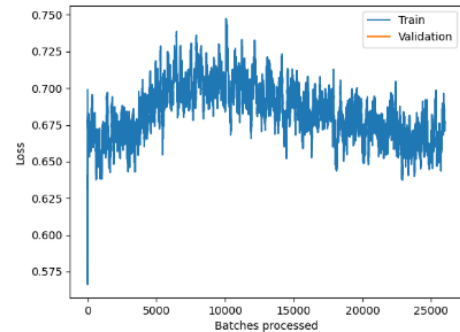
```
{
  "ModelParameters": {
    "pointrend": false,
    "backbone": "resnet50",
    "backend": "pytorch"
  },
  "Framework":
"arcgis.learn.models._inferencing",
  "ModelConfiguration":
"_maskrcnn_inferencing",
  "InferenceFunction":
"ArcGISInstanceDetector.py",
  "ModelType": "InstanceDetection",
  "MaskRCNNkwargs": (Jayasinghe,
Sano et al. 2017),
  "ExtractBands": [
    0,
    1,
    2
  ],
  "SupportsVariableTileSize": true,
  "Classes": [
    {
      "Value": 1,
      "Name": "1",
      "Color": [
        137,
        102,
        55
      ]
    }
  ],
  "MinCellSize": null,
  "MaxCellSize": null,
  "ArcGISLearnVersion": "1.9.0",
  "ModelFile": "BuildingFootprintExtraction_Africa.pth",
  "ImageHeight": 400,
  "ImageWidth": 400,
  "ImageSpaceUsed": "MAP_SPACE",
  "LearningRate": "1.0000e-04",
  "ModelName": "MaskRCNN",
  "backend": "pytorch",
  "average_precision_score": {
    "1": 0.7866484022095899
  },
  "resize_to": 400,
  "IsMultispectral": false
}
```

MaskRCNN

Backbone: resnet50

Learning Rate: 1.0000e-04

Training and Validation loss



Analysis of the model

Average Precision Score: {'1': 0.7866484022095899}

Sample Results

Appendix 11

Deep Learning Model- 30 Samples

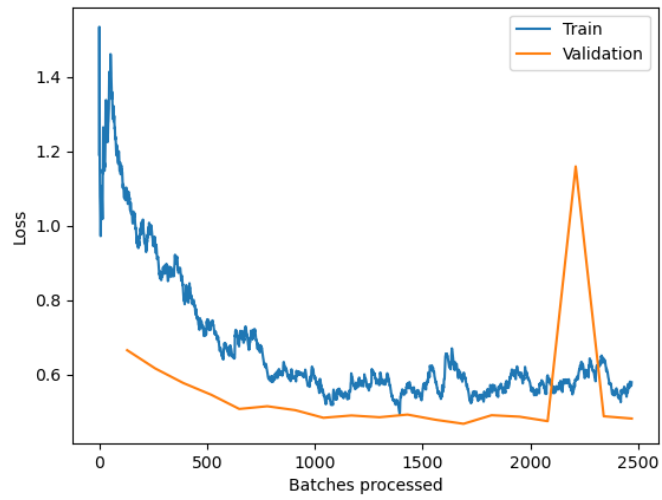
Building Footprint- Fine-tuned model with 30 Samples

```
{
  "ModelParameters": {
    "pointrend": false,
    "backbone": "resnet50",
    "backend": "pytorch"
  },
  "Framework": "arcgis.learn.models._inferencing",
  "ModelConfiguration": "_maskrcnn_inferencing",
  "InferenceFunction": "[Functions]System\\DeepLearning\\ArcGISLearn\\ArcGISInstanceDetector.py",
  "ModelType": "InstanceDetection",
  "MaskRCNNkwargs": {},
  "ExtractBands": [
    0,
    1,
    2
  ],
  "SupportsVariableTileSize": true,
  "Classes": [
    {
      "Value": 1,
      "Name": "1",
      "Color": [
        230,
        0,
        0
      ]
    }
  ],
  "ModelFormat": "NCHW",
  "MinCellSize": {
    "x": 0.15093142649469277,
    "y": 0.15706269495873806,
    "spatialReference": {
      "wkid": 102100,
      "latestWkid": 3857
    }
  },
  "MaxCellSize": {
    "x": 0.15093142649469277,
    "y": 0.15706269495873806,
    "spatialReference": {
      "wkid": 102100,
      "latestWkid": 3857
    }
  },
  "ArcGISLearnVersion": "2.0.0",
  "monitored_valid_loss": 0.5951891541481018,
  "ModelFile": "Output Model_30 Samples.pth",
  "ImageHeight": 224,
  "ImageWidth": 224,
  "ImageSpaceUsed": "MAP_SPACE",
  "LearningRate": "slice('1.9055e-05', '1.9055e-04', None)",
  "ModelName": "MaskRCNN",
  "backend": "pytorch",
  "average_precision_score": {
    "1": 0.18333333482344946
  }
},
```

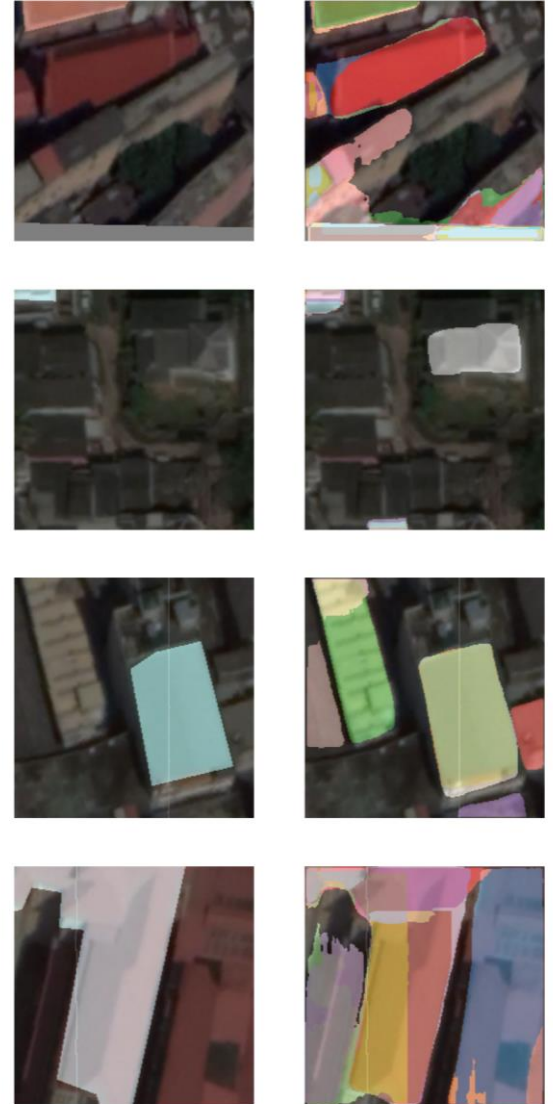
Appendix 12

Case 5: Retraining USA model- Pre trained

Accuracy: 63.5% and Accuracy of the original model is 71%

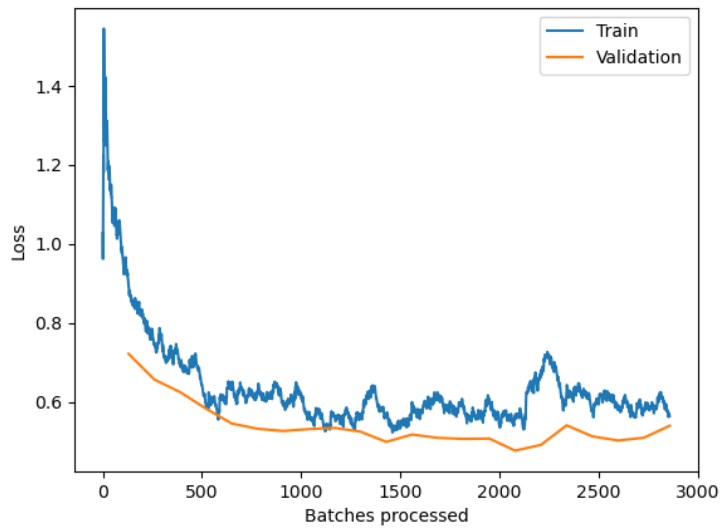


Ground Truth / Predictions

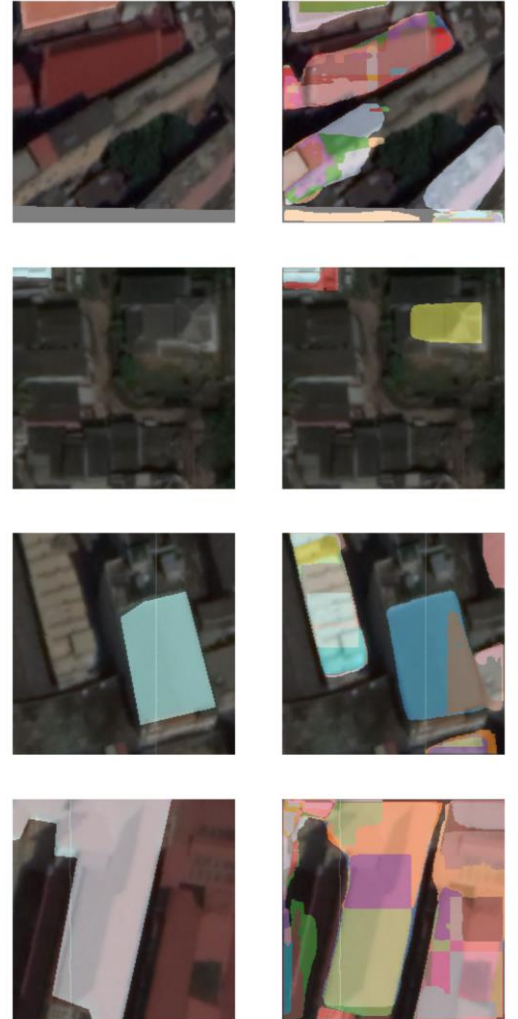


Case 5: Retraining Africa- Pre trained

Accuracy: 69.2% and Accuracy of the original model is 78%



Ground Truth / Predictions



Appendix 13

Road Network 2030

