

LB/TH/01/2026
TH6100

**TRAVEL DEMAND MODELING WITH MOBILE
NETWORK BIG DATA.**

(STUDY AREA: WESTERN PROVINCE OF SRI LANKA)

N. K. Bhagya Jeewanthi

(188062D)

Doctor of Philosophy

Department of Transport Management and Logistics Management

Faculty of Engineering

University of Moratuwa

Sri Lanka

August 2025

**TRAVEL DEMAND MODELING WITH MOBILE
NETWORK BIG DATA.
(STUDY AREA: WESTERN PROVINCE OF SRI LANKA)**

N. K. Bhagya Jeewanthi
(188062D)

Thesis submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy

Department of Transport Management and Logistics Management

Faculty of Engineering

University of Moratuwa

Sri Lanka

August 2025

DECLARATION OF ORIGINALITY

I declare that this is my work. This thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or institute of higher learning and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Signature:

Date: 05/08/2025.

COPY RIGHT STATEMENT

I hereby grant the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part, in print, electronic, or other medium. I retain the right to use this content in whole or in part in future works (such as articles or books).

Signature:

Date: 05/08/2025

STATEMENT OF THE SUPERVISORS

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

Name of the supervisor: Prof. Amal S Kumarage

Signature of the supervisor:

Date:07/08/2025

Co- Supervisor: Dr. Shehan Perera

Signature of the supervisor:

Date:08/08/2025

ACKNOWLEDGEMENT

Prima Facea, I sincerely thank my research supervisor, Senior Prof. Amal S. Kumarage. I have been fortunate to have a supervisor like him who guided me from start to end, continuously followed up on my progress, and gave me advice at each step of the research. I am very grateful for your encouragement and continuous attention throughout the research. Further, I extend my gratitude to Dr. Shehan Perera for the continuous guidance he gave me throughout the period.

Special thanks should go to the postgraduate research coordinator, Prof. A.I.T. Gamage, and former coordinators, Dr. T. Sivakumar and Dr. Varuna Adhikariwattage, for conducting research progress reviews and supporting the process. I would like to express my profound gratitude to Prof. Asoka Perera, Head of the Department of Transport and Logistics Management of the University of Moratuwa.

A very special thank you goes to Sriganesh Lokanathan (Former Team leader, LirneAsia) for lending me the data. I also thank Mr. Buddhi Weerasinghe and Mr. Anuradha Munasinghe, who contributed immensely to the primary data processing and analysis stages. I thank you all for trusting me in providing data and for the guidance given from start to finish. I gratefully acknowledge the funding received for my PhD from the AHEAD project.

I am sincerely thankful to all the staff members and non-academic staff of the Department of Transport and Logistics Management. I am also thankful to my family members, colleagues, senior colleagues, and friends who supported me in many ways throughout the research. Your encouragement, advice, and opinions are really appreciated.

ABSTRACT

The increasing travel demand necessitates dynamic traffic control mechanisms, which rely heavily on human mobility data for effective implementation. Traditional gathering of mobility information, such as household surveys, is costly and time-consuming. However, the advent of big data offers an unprecedented opportunity to analyze human mobility on a large scale in a cost-effective and real-time manner. Specifically, Call Detail Record (CDR) data, a subset of Mobile Network Big Data (MNBD), provides valuable insights into human mobility patterns. Yet, this data requires extensive processing to convert into structured travel information suitable for integration into travel demand models.

In response to this challenge, this study introduces a novel methodology that leverages CDR data to inform the four-step modeling process, encompassing Trip Generation/Attraction, Origin-Destination (OD) information, and Network Assignment. Utilizing a stay-based approach with pseudo-anonymized CDR data, the methodology identifies trip nodes by tracking the presence of mobile phone users within specific temporal and spatial cells. This approach facilitates extracting individual mobility characteristics, including activity sequences and trip generations/attractions. Moreover, it allows for determining trip purposes based on the origins and destinations and generating OD matrices from CDR data using mapped spatio-temporal data. Individual commuting routes are then defined based on the frequency of appearances and weighted caller activities along each potential route.

The methodology was applied to a CDR dataset comprising randomly selected 400,000 users over one month within the Western Province of Sri Lanka. The findings were validated against data from traditionally collected household visit surveys. The validation results demonstrated a significant correlation at each stage of the four-step process, underscoring the potential of repurposed, privacy-preserved mobile phone data to significantly enhance dynamic travel demand analysis.

Keywords – Mobile Network Big Data, Call Detail Record, OD Matrix, Network Assignment

TABLE OF CONTENTS

DECLARATION OF ORIGINALITY	i
COPY RIGHT STATEMENT	ii
STATEMENT OF THE SUPERVISOR.....	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
LIST OF FIGURES.....	x
LIST OF TABLES	xiii
LIST OF ACRONYMS.....	xiv
LIST OF APPENDICES	xv
1 INTRODUCTION.....	1
1.1 Background.....	1
1.2 Mobile Network Big Data	3
1.3 Purpose of the Research	7
1.4 Research Gap	9
1.5 Research Objectives	12
1.6 Scope of the Research.....	15
1.7 Overview of the Thesis.....	16
2 LITERATURE REVIEW.....	17
2.1 Traditional Travel Demand Modelling.....	17
2.1.1 Trip Generation and Attraction.....	18
2.1.2 Trip Distribution	19
2.1.3 Network Assignment	19
2.2 Use of Digital Data in Transport Analysis	20
2.3 CDR and Human Mobility	22

2.3.1	Origin Destination Matrices.....	22
2.3.2	Mobility Pattern Recognition and Significant Location Identification.....	24
2.3.3	Route Estimation from CDR data.....	25
2.4	Use of CDR in the Sri Lankan Context.....	25
2.5	Drawbacks in Call Detail Record Data.....	27
2.6	Techniques Used in Overcoming the Limitations of CDR.....	28
2.7	Research Gap Identification	29
3	METHODOLOGY.....	31
3.1	Study Area	32
3.2	Data Description – CDR Data	33
3.2.1	Data Privacy and Security Measures	33
3.2.2	Attributes in CDR Data.....	33
3.3	Household Visit Survey Data	34
3.4	Sampling Technique	36
3.4.1	Identifying the Sample Required from the Lowest Populated DSD.....	39
3.5	Overview of the Research Methodology	42
4	OD ESTIMATION FROM CDR DATA.....	46
4.1	Initial Processing	46
4.1.1	Removing NA	46
4.1.2	Splitting Time	47
4.1.3	Load Share Effect	47
4.1.4	Spatial Mapping.....	51
4.2	Descriptive Statistics	53
4.2.1	Temporal Measures.....	53
4.2.2	Spatial Measurements	61
4.3	Identifying Movements from CDR Data	69
4.4	Behavioral Representation of the Subscribers.....	72

4.5	Home-based Trip Identification and validation.....	80
4.5.1	Stage 1: Constructing a trip from CDR Data	80
4.5.2	Stage 2: Identifying Significant Locations	83
4.5.3	Stage 3: Trip Extraction and Classification	88
4.5.4	Stage 4: Estimation of Origin-Destination Matrices.....	90
4.5.5	Results and Validation	94
4.6	Conclusion	103
5	NETWORK ASSIGNMENT	104
5.1	Methodology	105
5.2	Stage 1 – Initial Data Processing	106
5.2.1	Defining the Sample	106
5.2.2	Enroute Cell Identification.....	107
5.2.3	Route-Cell Matching.....	109
5.2.4	User-Case Example.....	110
5.3	Stage 2 – Defining Trip Statistics.....	111
5.4	Stage 3- Determining the Route Choice	115
5.5	Stage 4 – Validation using STRADA.....	116
5.5.1	Introduction to STRADA.....	116
5.5.2	Results and Validation	121
5.6	Summary and conclusion.....	126
6	SUMMARY AND CONCLUSION.....	127
6.1	Utilizing CDR for the Four-Step Demand Modeling Process.....	127
6.2	Validation with HVS Data.....	128
6.3	Application of CDR-based Findings	129
6.4	Summary of Key Contributions of the Study	130
6.5	Limitations of the Study	132
6.6	Future Research	133

7	REFERENCES	135
8	APPENDIX A - SAMPLE REPRESENTATION FROM EACH DSD.....	143
9	APPENDIX B - HOUSEHOLD SURVEY QUESTIONNAIRE	144
10	APPENDIX C - SUMMARY OF THE LITERATURE REVIEW	145

LIST OF FIGURES

Figure 1-1 Data Transmission Flow in Mobile Telecommunication Systems	4
Figure 1-2 Standard Voronoi Cell Format.....	5
Figure 1-3 Attributes within a CDR Entry.....	7
Figure 1-4 Relationship among urban computing and big data.....	9
Figure 2-1 Illustration of the Traditional Travel Demand Model.....	18
Figure 3-1 District and DSD segregation in the Western Province of Sri Lanka,.....	32
Figure 3-2 Overview of the sampling population -Comparison with HVS population.....	37
Figure 3-3 Graphical Representation of Multi-Staged Random Sampling technique	39
Figure 3-4 Conceptual Framework of the Study	42
Figure 3-5 Methodology of the study at Macro level	44
Figure 3-6 Summary of expected outputs from the study	45
Figure 4-1 Removing NA from the CDR data.....	46
Figure 4-2 Splitting Time component within CDR Data.....	47
Figure 4-3 Graphical Representation of the Load Sharing Effect	48
Figure 4-4 Methodolog used within the study for Load Share Identification	49
Figure 4-5 Discrepancy Analysis Between CDR and HVS Distributions Across Speed Thresholds	50
Figure 4-6 User Connectivity among Cell Towers.....	52
Figure 4-7 Summary on Spatio-temporal measures derived from CDR Data.....	53
Figure 4-8 Caller Activity Distribution of CDR Sample.....	55
Figure 4-9 CDF of Caller Activities	55
Figure 4-10 Hourly Distribution of Caller Activities of the CDR Sample.....	56
Figure 4-11 Average Inter-Event Time of the CDR Sample	57
Figure 4-12 Cumulative Distribution of AIT.....	57
Figure 4-13 Maximum Inter Event Time Distribution of the CDR Sample.....	58
Figure 4-14 CDF of MIT	59
Figure 4-15 Day-wise Caller Activity Distribution of the CDR Sample.....	60
Figure 4-16 CDF of Day wise caller activity distribution	60
Figure 4-17 Cell Appearance Distribution of the CDR Sample	62
Figure 4-18 Cumulative Distribution of the cells visited	62
Figure 4-19 Saptial Activity Space of user 1 as derived from CDR data.....	64
Figure 4-20 Spatial Activity space of user 2 as derived from CDR data.....	64

Figure 4-21	Activity Space Distribution derived from CDR data	65
Figure 4-22	CFD of Activity Space	66
Figure 4-23	Entropy Values distribution derived from CDR data.....	68
Figure 4-24	CDF of Entropy Value	68
Figure 4-25	Generation of Time-Stamped Matrices from CDR data	71
Figure 4-26	User Behavior Model 1 Based on CDR Data	73
Figure 4-27	User Behavior Model 2 Based on CDR Data	74
Figure 4-28	User Behavior Model 3 Based on CDR Data	75
Figure 4-29	User Behavior Model 4 Based on CDR Data	75
Figure 4-30	User Behavior Model 5 Based on CDR Data	76
Figure 4-31	User Behavior Model 6 Based on CDR Data	77
Figure 4-32	Socio-Economic Classification Using Spatio-Temporal Features of CDR Data.....	79
Figure 4-33	Approach for OD Estimation Based on CDR Data.....	81
Figure 4-34	Theoretical Trip Structure.....	81
Figure 4-35	Personalized Activity Trajectories	82
Figure 4-36	User Mobility Mapping Across Time Periods Using CDR Data.....	83
Figure 4-37	Significant location categorization based on User Behaviour	84
Figure 4-38	Integration of Appearance and Activity Layers.....	84
Figure 4-39	CDR-Based Core Time Window Identification.....	85
Figure 4-40	Trip Classification Based on Origin-Destination Combinations	89
Figure 4-41	Spatial Mapping of Segmented User Locations to Administrative Zones.....	91
Figure 4-42	Cell Tower Allocation to District and DSD Zones via GIS.....	91
Figure 4-43	Trip Type Classification Across DSD and District Levels	94
Figure 4-44	Validation of Trip Distribution derived from CDR data at Intra District Level	96
Figure 4-45	Validation of HBW Trip Distribution derived from CDR data at at DSD level.....	96
Figure 4-46	Validation of HBO Trip Distribution derived from CDR data at at DSD level.....	97
Figure 4-47	Validation of NHB Trip Distribution derived from CDR data at at DSD level.....	97
Figure 4-48	Intensity map of Intra District Trips by Trip Type	98
Figure 4-49	Trip Distance Distribution Across All Trip Types.....	101
Figure 4-50	Hourly Distribution of HBW trips	102
Figure 4-51	Hourly Distribution of HBO trips	102
Figure 4-52	Hourly distribution of NHB trips.....	103
Figure 5-1	Methodology for Inferring Travel Routes Using CDR Data	105
Figure 5-2	Identification of Enroute Locations from Call Detail Records	107

Figure 5-3 Filtered User Set for Route Estimation Using CDR Data.....	108
Figure 5-4 Location-Based Mapping of Daily Commute Patterns	110
Figure 5-5 Mapping CDR derived locations onto the Most Probable Travel Path.....	111
Figure 5-6 Steps in deriving the trip statistics(a) and (b)	113
Figure 5-7 Route Choice Representation of a Single User.....	114
Figure 5-8 Steps in deriving the trip statistic (c)	114
Figure 5-9 Route Choice generation function via CDR Data.....	116
Figure 5-10 Initial Interface of STRADA.....	119
Figure 5-11 Sample OD format in STRADA	119
Figure 5-12 Parameter interface of STRADA	120
Figure 5-13 Network Assignment Methodology with CDR Data and STRADA,	121
Figure 5-14 Distribution of the Route Choices.....	122
Figure 5-15 Sample Route Assignment comparison	124
Figure 5-16 Route Assignment comparison with CDR Vs STRADA.....	125
Figure 5-17 Route Assignment comparison of CDR Vs STRADA after outlier detection.....	125
Figure 5-18 Route assignment from CDR Data (A) and Traditional Data (B).....	126

LIST OF TABLES

Table 1-1 Scope and Direction of the Study	12
Table 3-1 Attributes in a CDR Entry	34
Table 3-2 Estimation of the target sample size from the HVS	35
Table 3-3 Total Sample Calculation according to the Multistage Random Sampling Method ...	41
Table 3-4 Sample Comparison-HVS and CDR Sample	41
Table 4-1 Sample Representation of Load Shared Records	50
Table 4-2 Cell tower coverage at district level of Sri Lanka	52
Table 4-3 Sample Entropy calculation.....	67
Table 4-4 Socio-Economic Segmentation of Users Using CDR data.....	77
Table 4-5 Time-of-Day Categorization for Regularity Analysis	87
Table 4-6 Regularity Measurement of Cell Tower Visits by Time Category.....	87
Table 4-7 Significant Location Extraction via Regularity and Ideal-Typical Matching	88
Table 4-8 Trip Type Classification Using Sample CDR Entries	90
Table 4-9 Classification of Intra- and Inter-Zone Trips at District and DSD Levels	93
Table 4-10 Comparison of Trip Type Proportions: CDR vs. Household Travel Survey Data	95
Table 4-11 Expansion of HBW Trips to Actual Trip Counts Using Population Scaling	98
Table 4-12 Comparison of Expanded HBW Trip Counts (CDR vs. HVS) at District Level	99
Table 5-1 Example Calculation of Determining Route Choice for a Single User.....	115
Table 5-2 Steps in Traditional Network Assignment.....	118

LIST OF ACRONYMS

MNBD	Mobile Network Big Data
CDR	Call Detail Records
VLR	Visitor Location Register
CA	Caller Activities
HVS	Household Visit Survey
WP	Western Province
DSD	Divisional Secretariat Divisions
GND	Grama Niladhari Divisions
BTS	Base Transceiver Station
O- D MATRIX	Origin – Destination Matrix

LIST OF APPENDICES

Appendix A	Sample Representation from each DSD	134
Appendix B	Household Survey Questionnaire	144
Appendix C	Summary of the Literature Review	145