Journal of South Asian Logistics and Transport
Volume 4 Issue 1 March 2024
JSALT 4.1.3: pp. 45-68
DOI: https://doi.org/10.4038/jsalt.v4i1.77

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

Thivya P Amalan [a*] ***and R C Wickramasuriya*** [b]

[a] *SMART Infrastructure Facilities, University of Wollongong, Australia*
[b] *Endeavour Energy, Australia.*

[*] Correspondence should be addressed to *tamalan@uow.edu.au*

ABSTRACT

The attraction and generation of a trip play a crucial role in transport planning and forecasting of trips. Multiple linear regression (MLR) is the most popular method of calculating trip attractions (TA) and trip generations (TG) to produce a distribution that can be used to forecast trips with updated values of independent variables such as electricity consumption, no of households, area of the land use etc. Literature shows that the surveys require independent variables to be updated; making them expensive and time consuming. This study aims to develop an MLR model for TA and TG based on available survey data from 2013 for Western Province, Sri Lanka and to update those independent variables using High Frequency (HF) proxy data for the predicted year (2019). Data from HF proxy sources such as electricity consumption, GPS customer points, and Landsat satellite imagery data have been used to update the independent variables of TG and TA for 2019. In the initial stage of this research, data from a home visit survey conducted in 2013 and land use data from the Western Province of Sri Lanka were used to develop the MLR model for TG and TA. A correlation and a regression analysis were performed using these surveyed data. According to the study, the MLR model for TG for home-based work trips has an r2 value of 0.79 and TA for general purposes (including shops and businesses) and industrial has an r2 value of 0.74 and 0.79, respectively, indicates the strong relationship between the considered variables to predict TG and TA. The model is validated with 2013 survey data and would be helpful for real-time estimation of the TG and TA of each zone.

Keywords: *Trip Generation and attraction, Multiple Linear Regression, High-Frequency Proxy Data, GPS data, Landsat Satellite Imagery*

1. INTRODUCTION

Transport planners and engineers have long been developing different methods and models to forecast travel demand in urban areas. Estimating travel demand is a core element of designing transport facilities and developing transport and land use planning. A four-step demand model is the typical approach that planners and engineers use. One of the main steps of the four-step travel demand is to estimate trip distribution considering trip generation (TG) and trip attraction (TA). Accordingly, trip generation aims at estimating the total number of trips produced by households, and trip attraction identifies the number of trips attracted to the zones based on land use activities.

However, many studies argued about 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 other data related to travel demand [2, 3]. Amalan et al. argued in detail about the data requirement of building a transport model, especially in developing countries and brought a new concept of data called "High-Frequency Data" to solve the current issues [4]. Though there are several other HF data available to predict the trip distribution such as mobile big data [5],public transport smart card data and Global Positioning Data (GPS), there are significant inefficiencies such as absence of interpersonal attributes and limitation in analysing the data especially in developing countries [4].

In this background, employing different methods to estimate TG and TA models can solve the above-mentioned data constraints in developing countries. Accordingly, the study is focused on developing home based work trip generation and attraction model based on freely available high frequency data such as electricity consumption and satellite images. The study area is based on Western Province, Sri Lanka, the smallest province but highly urbanised in Sri Lanka. Colombo, Gampaha, and Kalutara are the three administrative districts that make up the Western Province in Sri Lanka, which has high population density, with approximately 5.8 million people in 2013 in a land area of 3,684 sq. km. The importance of developing an effective transport 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 transport model can help policymakers and planners make informed decisions about transportation policies and investments.

2. INCOME VS ELECTRICITY CONSUMPTION

Several studies were conducted to validate the relationship between household income and electricity consumption. In the study conducted by Athukorala and Wilson (2010), the researchers employed unit root, integration, and error correction models on data spanning from 1960 to 2007. Their findings revealed a significant relationship between household income (GDP) and electricity demand, indicating that a one percent increase in GDP corresponds to a 0.78 percent increase in electricity consumption, highlighting the long-run income elasticity of demand. In a separate investigation by Rajmohan and Weerahewa (2007), the 'energy ladder' hypothesis was examined, and Engle functions were estimated using Consumer Finances and Socio-Economic Survey data from 1978/79 to 2003/04. The results showed that in the fiscal year 2003/04, the highest income decile accounted for a substantial 64 percent of electricity consumption, suggesting disparities in electricity usage across income groups. These studies contribute valuable insights into the dynamics of electricity demand in relation to income and underscore the significance of income distribution in shaping consumption patterns.

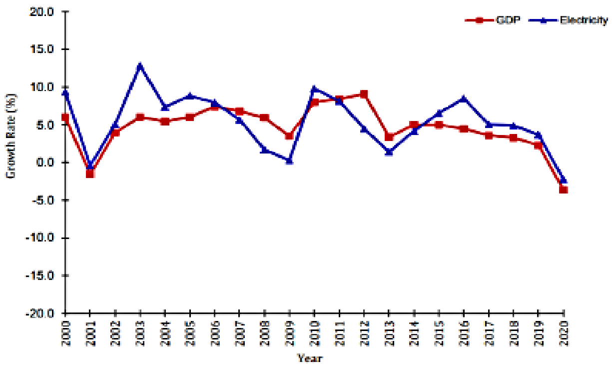

Figure 1: GDP Vs Electricity Growth Rate - Sri Lanka

Historical data in Sri Lanka shows a direct correlation between electricity demand and economic growth rates [11]. Figure 1 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.

3. OVERVIEW OF TRIP GENERATION/ATTRACTION MODELS

In general, trip generation/attraction can be analysed in 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 [3,6]. For trip generation, households can be classified based on their cross-characteristics such income level, number of members per households etc. 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[6] . 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.

3.1. Regression Model for Trip Generation/Attraction

Statistical regression analysis is the most widely used method for investigating and modelling the relationship between variables. Early-period regression analysis techniques were trendy 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 use methods that allow for prediction of the response values given the values of regressor X. The appropriate form of the linear relationship between the two variables is shown in Equation 1, where β_0 is known as the intercept and β_1 as the slope.

$$Y = \beta_0 + \beta_1 X \quad (1)$$

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 as in Equation 2.

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (2)$$

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 3.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n + \varepsilon \quad (3)$$

This model contains Y as the natural dependent variable, and X_1, X_2,... 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 β_0, β_1, ..., β_n, not because Y is a linear function of X's. Estimating unknown parameters in the regression model is one of the essential 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[7]). For all steps of this travel forecasting model, regression analysis techniques were extensively used for both unknown parameter estimation and model adequacy checking.

4. DATA SET AND METHODOLOGY

This study is based on the Western Province of Sri Lanka, the smallest province in Sri Lanka but highly populated, with about 5.4 million people in 2013. This study uses electricity consumption data acquired from the Ceylon Electricity Board (CEB) and Lanka Electricity Company (Private) LTD (LECO), which are the primary providers of electricity in Western Province, Sri Lanka (Fig 2).

Figure 2: Area Classification of CEB and LECO

CEB	Area Name
Colombo City	Colombo North Colombo East Colombo West
WS1	Sri J'Pura
WS2	Avissawela
WS3	Colombo South (Dehiwela)
WS4	Ratmalana
WS5	Homagama
WS6	Horana
WS7	Bandaragama
WS8	Kalutara
WN1	Negombo
WN2	Veyangoda
WN3	Ja-ela
WN4	Kelaniya
WN5	Gampaha
WN6	Divulapitiya

LECO
Kotte
Nugegoda
Moratuwa
Kaluthara
Kelaniya
Negombo

Source: CEB Website

This data source is not available publicly but can be available upon request from CEB and LECO 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 3. Since the study focuses only on home-based work trips, the classification of domestic, general purpose and industrial customers details were used for modelling purposes.

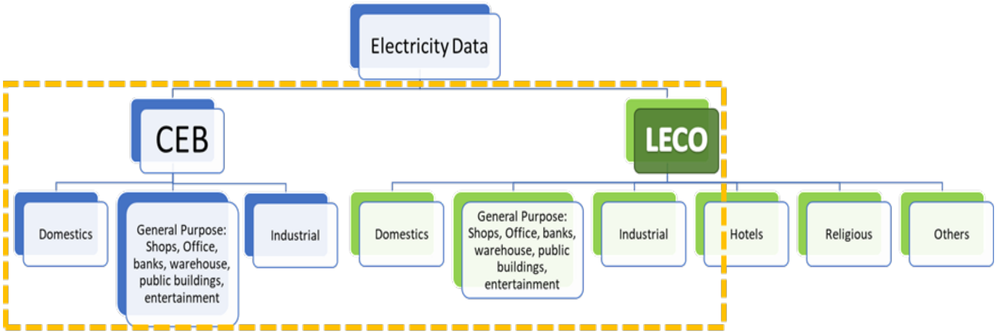

Figure 3: Electricity Classifications

5. DATA ANALYSIS

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.

5.1. Validation of Electricity GPS Data

The comparison between the number of households calculated from the GPS points of electricity consumption data and the census and statistics data of 2012 shows the validation of 94%. Furthermore, the GPS data was analysed concerning other classifications and provided in table 1 for 2013 and 2019 with an annual growth rate of each building classification based on the electricity consumption data from CEB and LECO.

Table 1: Annual Growth Rate

Year	Households	General Purpose	Industrial
2013	53,628	6448	21
2019	60,347	8972	24
Annual Growth rate	2.0%	5.7%	2.3%

Figure 4 shows the relationship between the number of households and the total number of trips generated from each zone for 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. Trip generation is high in a suburban area with a high number of households.

5.2. Relationship between Income and Electricity Consumption

Several studies were conducted to validate the relationship between income and electricity consumption [8-10]. The primary analysis was conducted on the data of HVS 2013 to show the relationship between electricity bills per month and income.

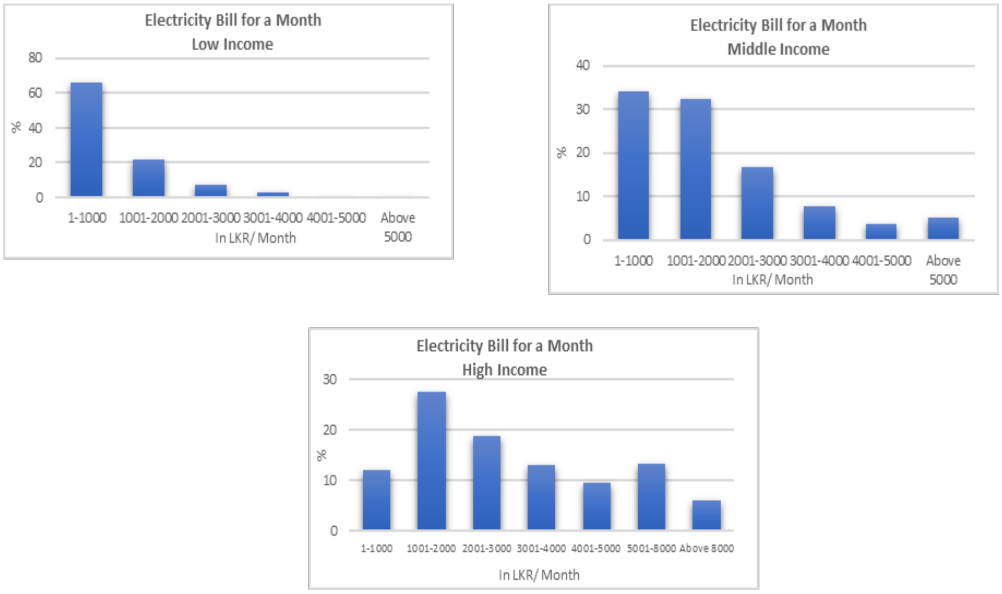

Figure 5: Income vs Electricity Consumption

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

The following analysis is undertaken for the entire Western Province, Sri Lanka. The income and electricity expenses distribution based on HVS 2013 are shown in Figure 5.

It is clear from the information presented in Figure 5 that there is a strong correlation between household income and electricity expenses. Most low-income households' electricity consumption expenses are less than LKR 1,000, while middle-income household expenses fall between LKR 500 to LKR 3,000. Nearly 88% of households under high income spent LKR 1,000 monthly on electricity. The descriptive statistics of each income level are shown in Table 2, 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.

Table 2: Descriptive Statistics of Income vs Electricity Consumption

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

5.3. Calculating area using Supervised Classification

Landsat satellite images have been used to obtain the area of specific land use to develop trip attraction model. 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. 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 digitises the classification as an image. The Study used the supervised classification of satellite data using Arc GIS Pro and 2013 land use classification data to obtain the area of general purpose and industrial buildings. This study uses Landsat satellite images that are available for 2013 and 2019, and these satellite images can be retrieved for free with a revisiting time of 16 to 18 days.

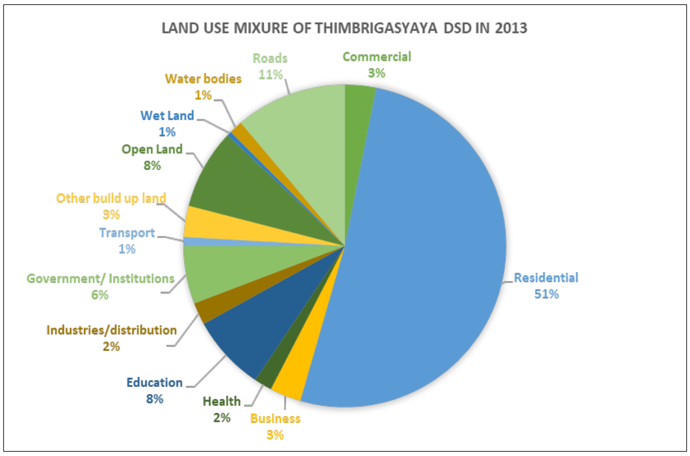

Figure 6: Land use classification of Thimbirigasyaya DSD (2013)

Source: CoMTrans,2013

The study tested the model in a small area to validate the supervised classification. Thimbirigasyaya Divisional Secretariat Division (DSD) was selected for testing as this area is famous for the mixed land use DSD in Western Province (Fig 6). DSD level is the small administrative units in Sri Lanka and responsible for local governance and public administration at the divisional level.

A fair 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. The results show 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: 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 4.

Table 4: 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 5 is from the land use survey conducted for the CoMTrans study in 2013. The values for all DSDs are unavailable, and the comparison of the supervised classification vs actual classification from land use survey of 2013 has been conducted based on the available DSDs results. Validation of land use classifications is performed according to each district's basis, and total developed area comparisons are based only on DSD built up area data. According to that the accuracy of each built up area in Kalutara, Gampaha and Colombo are 88%, 99% and 104% respectively, with the overall accuracy of 96% for the entire study area. For the DSDs that do not have the actual values, the supervised classification values are applied directly. According to the derived developed area from the supervised classification, the area of general purpose is calculated based on the detail land use classification available with CoMTrans study in 2013.

Moreover, to derive the trip attraction for 2019, the model uses the electricity customers GPS points assuming that the growth rate of general purpose and industrial area is proportionate to the increase in GPS points of the customers between 2013 and 2019.

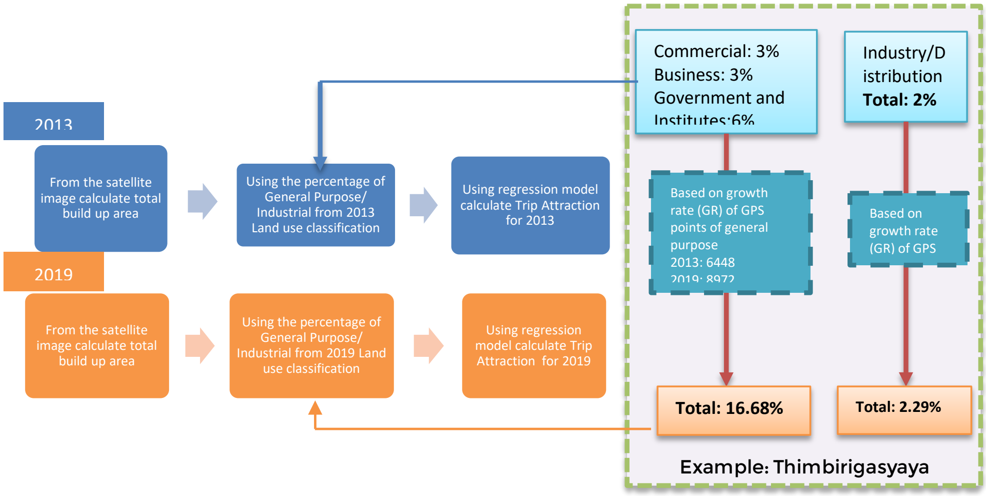

Figure 7: Methodology used for Calculating Area of General and Industrial

Figure 7 illustrates the methodology for computing the area of industrial and general-purpose buildings. The calculation is based on the integration of satellite imagery and land use survey data from 2013. Furthermore, the figure illustrates the process for updating values using GPS data from electricity customers and satellite imagery from 2019. For example, in Thimbirigasyaya in 2013 general purpose building shares 12% of the total area and the GPS points of electricity customers for the general purpose is increased from 6,448 to 8,972 with overall growth rate is 39%. This percentage increase is applied to the 2013 share and calculated that in 2019 these customers shared 16.68% of the total area. Likewise, this percentages have been calculated to all DSDs and shown in table 5.

Table 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
Kesbewa	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

6. DEVELOPING TRIP GENERATION MODEL AND VALIDATION

The trip generation model has been developed based on the average monthly electricity consumption data and the number of households per zone. The zones are based on the 461 traffic analysis zones considered in the study of CoMTrans 2013 of Western Province, Sri Lanka. This is the smallest traffic analysis zone considered by this study based on the homogeneous characteristics in terms of land use, population, accessibility etc. The independent variables of number of households and average monthly electricity unit are calculated using from the electricity consumption data of

2013 collected from the CEB and LECO and trip generation values were calculated from the HVS 2013. The multiple linear regression relationship (Equation 4) was developed between these variables and the goodness of fitness and p value between the variables are shown in table 6.

*Home to Work Trips = 168.924 + 0.831488 (No of households)
+ 2.524818 (Electricity Units/Month) -------------(4)*

Table 6: Regression Analysis (Trip Generation)- 2013

Regression Statistics	
Multiple R	0.892323
R Square	**0.796239**
Adjusted R Square	0.79535
Standard Error	913.5382
Observations	461

ANOVA

	df	SS	MS	F	Significance F
Regression	2	1.49E+09	7.47E+08	894.8684	6.1E-159
Residual	458	3.82E+08	834552.1		
Total	460	1.88E+09			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	168.924	112.9558	1.495488	**0.013546**	-53.0518	390.8998
No of HHS	0.831488	0.01982	41.95125	**5.5E-159**	0.792538	0.870438
Avg Elec unit/Month	2.524818	0.734969	3.435271	**0.000646**	1.080488	3.969148

The research conducted investigates the determinants of the dependent variable in a given context using MLR analysis. The model demonstrates a high level of significance ($p < 0.05$) and a strong explanatory power, with an R-square of 0.796239. Both the Number of Households (No of HHS) and Average Electricity Units per month emerge as highly significant predictors (p = 5.5E-159 and p = 0.000646, respectively). The intercept, representing the expected value when independent variables are zero, is 168.924 (p = 0.013546). These results collectively indicate that variations in the dependent variable are substantially explained by changes in the specified independent variables. The findings offer valuable insights for stakeholders and policymakers in the context studied, suggesting the need for attention to household numbers and electricity consumption patterns. (Fig 8).

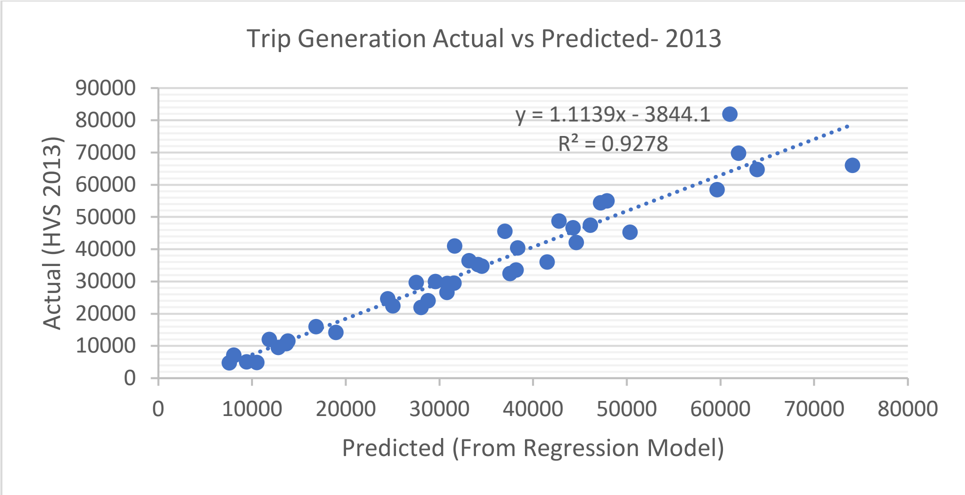

Figure 8: Trip Generation (Actual vs Predicted)

Figure 7 shows that the trip generation calculated from the regression model provides a very good fit (92%) against the actual value calculated from the HVS 2013. The same regression model was used to calculate the trip generation for the year of 2019 based on the electricity consumption data acquired for the October 2019.

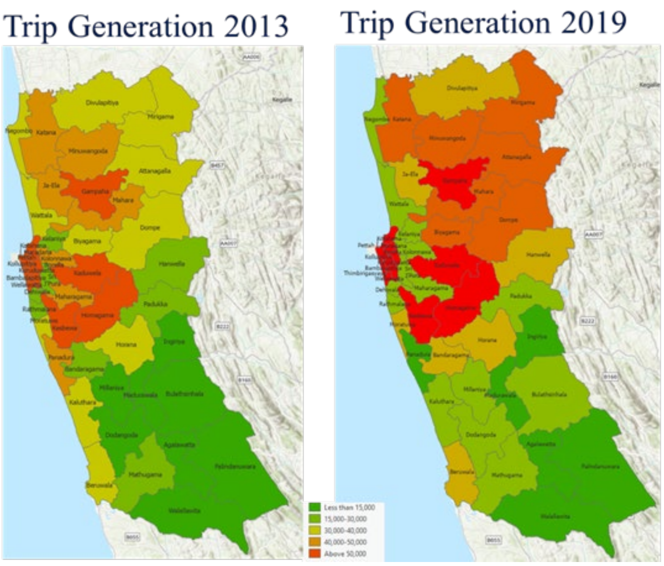

Figure 9: Trip Generation (2013 and 2019)

Figure 9 shows the comparison of trip generation between the year of 2013 and 2019. It can be observed that number of trip generation in the suburban areas has been increased as the number of households and their energy consumptions are increased significantly in 2019.

7. DEVELOPING TRIP ATTRACTION MODEL AND VALIDATION

Trip attraction model has been developed based on the area of each land use occupied by different categories. The study considered two different types of trip purposes based on the land use categorisation considered by the electricity providers namely, general purpose land use that includes shops, office banks etc and industrial land uses (Fig 2). Data analysis aims to produce two equation models for general purpose buildings and industrials between dependent variables and independent variables separately. The first model considered the trips attracted per km^2 to the general-purpose building as a dependant variable and the area of the general-purpose building per km^2 is considered as an independent variable (Equation 5). The model was tested to 32 DSDs based on the available land use data for 2013. The goodness of fit of this developed model shows 70.7% with accepted p-value ($p \leq 0.05$) (table 7).

General Purpose attracted trips/ km^2 = 13.79302 + 8638.311 Area of general purpose building/ km^2 -----------(5)

Table 7: Regression Model (General Purpose) - 2013

Regression Statistics	
Multiple R	0.841295
R Square	**0.707777**
Adjusted R Square	0.698037
Standard Error	84.67702
Observations	32

ANOVA

	df	*SS*	*MS*	*F*	*Sig. F*
Regression	1	520997.1	520997.1	72.66147	1.64E-09
Residual	30	215105.9	7170.198		
Total	31	736103.1			

	Coef.	*Standard Error*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	13.79302	22.13066	0.623254	**0.05**	-31.4038	58.98985	-31.4038	58.98985
GP/km2	8683.311	1018.669	8.52417	**1.64E-09**	6602.91	10763.71	6602.91	10763.71

The regression model is statistically significant, as evidenced by the low p-value (1.64E-09) in the ANOVA table 7. The model explains a substantial portion of the variance in the dependent variable, as indicated by the R-square of 0.707777. The coefficient for the intercept is 13.79302 ($p < 0.05$), representing the estimated value of the dependent variable when the independent variable is zero. The coefficient for GP/km^2 is 8683.311 ($p < 0.05$), suggesting that for every one-unit increase in GP/km^2, the dependent variable is estimated to increase by 8683.311 units. The 95% confidence interval for the GP/km^2 coefficient is (6602.91, 10763.71). These findings suggest a significant positive relationship between the independent variable (GP/km^2) and the dependent variable. However, careful consideration is needed due to the relatively small sample size (32 observations) and potential outliers, as reflected in the wide confidence intervals for the coefficients.

$$\textit{Trips attracted to the Industrial area/ km}^2 = 16.68874 + 4935.869 \textit{ Area of industrial building/ km}^2 \text{ --------------------(6)}$$

Table 8: Regression Model (Industrial) - 2013

Regression Statistics	
Multiple R	0.892047
R Square	**0.795748**
Adjusted R Square	0.789159
Standard Error	65.98064
Observations	33

ANOVA

	df	*SS*	*MS*	*F*	*Sig. F*
Regression	1	525779.8	525779.8	120.7733	3.21E-12
Residual	31	134956.8	4353.444		
Total	32	660736.6			

	Coeff.	*Standard Error*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	16.68874	15.12292	1.103539	**0.027829**	-14.1547	47.53214	-14.1547	47.53214
Ind (Area)/km2	4935.869	449.1363	10.98969	**3.21E-12**	4019.849	5851.889	4019.849	5851.889

The regression model is highly significant (p-value of 3.21E-12) according to the ANOVA Table 8, indicating a strong relationship between the dependent and independent variables. The model explains a substantial proportion of the variance in the dependent variable, with an R-square of 0.795748. The coefficient for the intercept is 16.68874, representing the estimated value of the dependent variable when the independent variable is zero. The coefficient for Ind (Area)/km² is 4935.8, suggesting that for each additional unit increase in Ind (Area)/km², the dependent variable is estimated to increase by 4935.869 units. The 95% confidence interval for the Ind (Area)/km² coefficient is (4019.849, 5851.889). These results suggest a significant positive relationship between the independent variable (Ind (Area)/km²) and the dependent variable.

Table 9: Trip Attraction (General Purpose) - 2013

DSD	Developed Area (Supervised Classification)	GP%	Area of GP	Area TAZ	Total GP area/ TAZ area	TA (GP)
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	107.72	0.05	0.10	209.48	0.00	3,794
Colombo	21.40	20.00	4.16	20.82	0.20	116,036
DehiwalaMt_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
Kaduwela	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	12.00	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 10 shows the validation of above model with the total trip attraction values retrieved from HVS 2013 for general-purpose buildings. The model yielded an overall actual vs prediction accuracy of 93.23%.

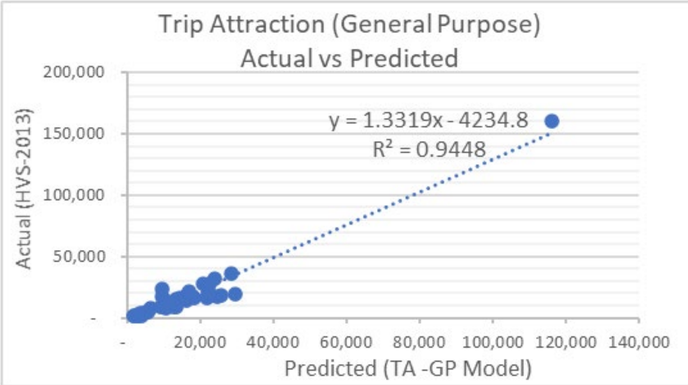

Figure 10: Trip Attraction (General Purpose) Actual vs Predicted - 2013

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 (Equation 8) is shown in the table 10. The 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 attraction for the industrial purpose as this Biyagama is a one of the largest export processing zones (EPZ).

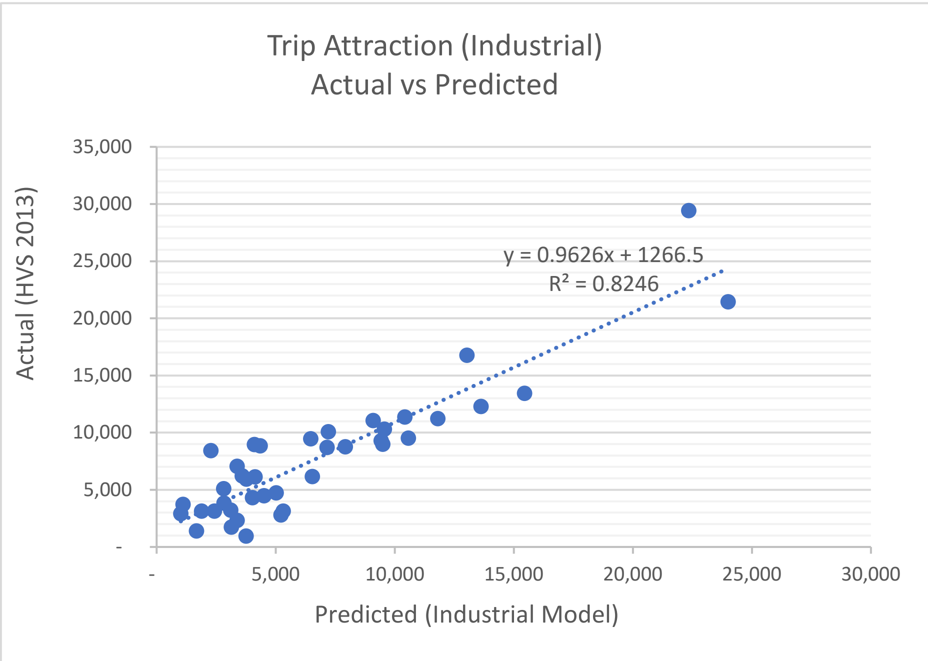

Figure 11: Trip Attraction (Industrial) Actual vs Predicted - 2013

Table 10: 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
Hanwella	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
Kesbewa	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 11 shows the validation of above model with the total trip attraction retrieved from HVS 2013 for the industrial buildings. 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 12 shows that application of developed model based on the 2013 data with the data of 2019. Based on that nearly 1,177,741 trips attracted to the Western Province in 2019.

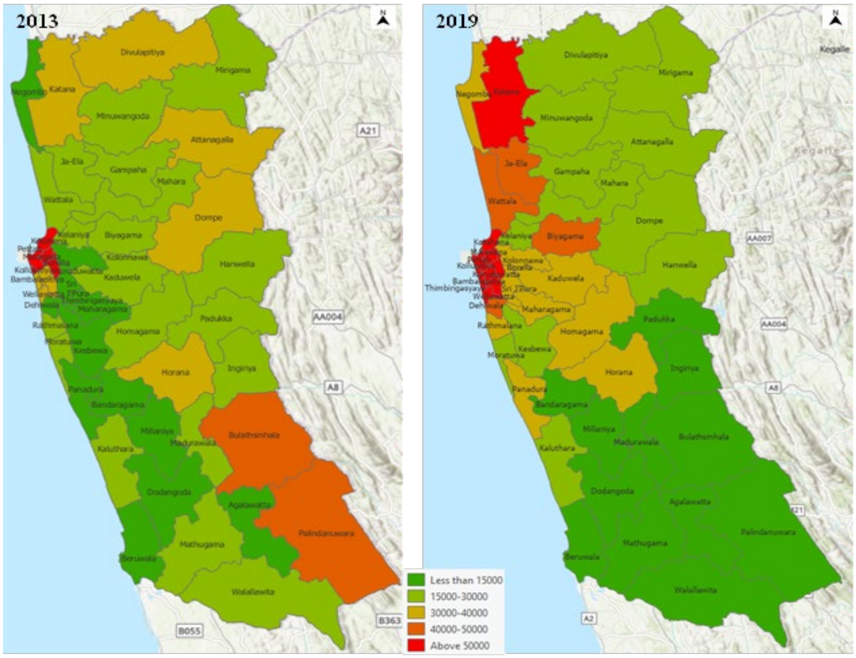

Figure 12: Summary of Trip Attraction between 2013 and 2019

8. 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

- 10% increment to the HHS
- 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 13), and the annual growth rate is calculated based on that data. The annual growth rate of electricity consumption in Western Province is 5.3%.

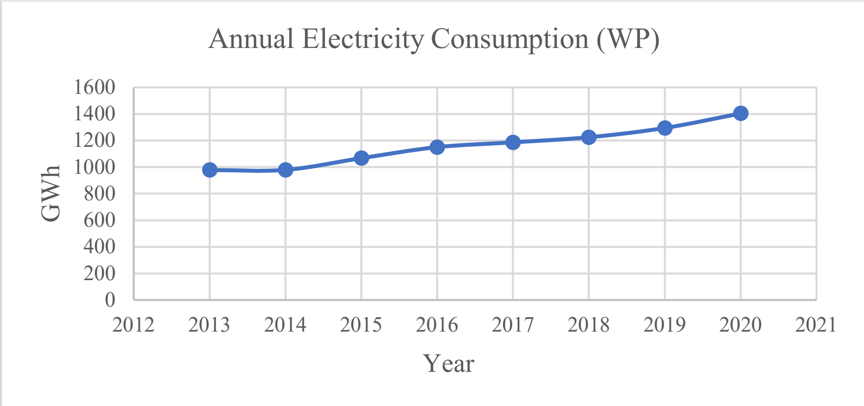

Figure 13: Annual Electricity Consumption (Western Province)

Based on the above two cases, the sensitivity analysis results are shown below (Table 11).

Table 11: Sensitivity Test for Trip Generation

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.

- 10% increase in the general-purpose area
- 10% increase in the industrial area

Based on the above cases, the sensitivity results are shown in Table 12.

Table 12: 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.

9. CONCLUSION AND RECOMMENDATIONS

The importance of exploring HF data as the explanatory variables for the travel demand model is again emphasized here, having the availability of long-range time series data, which can be used to update the model frequently. However, household income can show e a strong relationship between trip generation, and this study has shown a relationship between the electricity consumer unit and the household income using 2013 survey data. This leads to a promising direction to ensure 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 model with the number of households in a particular zone. With these two independent variables, the developed trip generation model shows the r^2 of 79% with 92% accuracy from the CoMTrans survey 2013 between actual and predicted trip generation from the model for 2013. This represents the proportion of the variance in the dependent variable of trip generation that is predictable from the independent variables of number of household and electricity consumption in a regression model.

The electricity consumption data was used to develop trip attraction models for general and industrial areas separately. Due to the non-availability of land use survey data, the satellite image data with supervised classification was used to get the land use area information for 2013 and 2019. The developed models for general purpose and industrial trip attraction have r^2 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. This explains the proportion of variance in between the total trip attraction of each household can be predictable based on the area of each trip attractions.

In summary, the study highlights the potential of using high-frequency data, household income, and electricity consumption information to develop accurate and regularly updated travel demand and attraction models. The results contribute valuable insights into the relationship between socio-economic factors, land use, and travel behaviour, offering a foundation for informed urban planning and transportation management. Despite the valuable insights provided by the study into travel demand and attraction models, several limitations should be considered. The reliance on available data, including household income and land use information derived from satellite imagery, introduces potential uncertainties, and relies on the quality and completeness of the datasets. The temporal constraints of the study may limit its ability to capture evolving travel behaviour over time, and the generalizability of the findings to different regions may be constrained by the specificity of the geographic area studied. Additionally, the simplification of variables, such as using household income and electricity consumption as proxies for trip generation, may oversimplify the complex nature of travel behaviour. The study's use of regression models with a limited number of variables might oversimplify the underlying relationships, and the reliance on survey data from a specific year (2013) may not fully capture dynamic changes in travel behaviour over time. Furthermore, the predictive models based on historical data may have limited applicability to future scenarios, necessitating caution when extrapolating results to changing urban landscapes and transportation systems. Acknowledging these limitations is crucial for a nuanced interpretation of the study's findings and informs avenues for future research refinement.

REFERENCES

[1] Mukherjee, J. and B. Raghuram Kadali, *A comprehensive review of trip generation models based on land use characteristics.* Transportation Research Part D: Transport and Environment, 2022. 109: p. 103340.

[2] Jayasinghe, A., K. Sano, and K. Rattanaporn, *Application for developing countries: Estimating trip attraction in urban zones based on centrality.* Journal of Traffic and Transportation Engineering (English Edition), 2017. 4(5): p. 464-476.

[3] Ahmed, T., et al., *Trip Generation Rates of Land Uses in a Developing Country City.* Transportation Research Record, 2020. 2674(9): p. 412-425.

[4] Amalan, T.P., et al., *High frequency data in Land use and transport integrated model: A review of sources and application.* Asian Transport Studies, 2023. 9: p. 100092.

[5] Kumarage, A., D. Maldeniya, and S. Lokanathan, *ORIGIN-DESTINATION MATRIX ESTIMATION FOR SRI LANKA USING MOBILE NETWORK BIG DATA*. 2015.

[6] Huntsinger, L.F., N.M. Rouphail, and P. Bloomfield, *Trip Generation Models Using Cumulative Logistic Regression.* Journal of Urban Planning and Development, 2013. 139(3): p. 176-184.

[7] Anderstig, C. and L.-G. Mattsson, *Modelling Land-Use and Transport Interaction: Policy Analyses Using the IMREL Model*. 1998. p. 308-328.

[8] Athukorala, P.P.A.W. and C. Wilson, *Estimating short and long-term residential demand for electricity: New evidence from Sri Lanka.* Energy Economics, 2010. 32: p. S34-S40.

[9] Rajmohan, K. and J. Weerahewa, *Household Energy Consumption Patterns in Sri Lanka.* Sri Lankan Journal of Agricultural Economics, 2010. 9.

[10] Kim, M.-J., *Determining the Relationship between Residential Electricity Consumption and Factors: Case of Seoul.* Sustainability, 2020. 12(20).

[11] CBSL. 2019. Economic and Social Statistics of Sri Lanka.