Available online at www.CivileJournal.org

# Civil Engineering Journal

(E-ISSN: 2476-3055; ISSN: 2676-6957)

Vol. 8, No. 09, September, 2022


# Modeling Trip-generation and Distribution using Census, Partially Correct Household Data, and GIS

Akash Anand [1*], Varghese George [1]

[1] Department of Civil Engineering, National Institute of Technology Karnataka, Mangalore-575025, Karnataka, India.

Received 26 April 2022; Revised 10 August 2022; Accepted 19 August 2022; Published 01 September 2022

**Abstract**

The efficiencies of urban transport systems in several cities are drastically affected due to difficulties imposed by rapid urbanization and the proliferation of private modes of transport. The conventional four-stage travel demand modeling approach provides an ideal platform to formulate strategies to rectify problems in urban transport. Trip generation is the first stage in this exercise (where trip production and trip attractions are modelled), followed by trip distribution in the second stage. The present work related to the development of models for trip generation and trip distribution necessitated the use of census data related to the number of households in each zone since the available revealed preference (RP) data compiled based on household interview surveys was partially incorrect. A review of the literature indicated that studies on the use of sparsely available and partially inaccurate data such as revealed preference and zone-specific secondary data on trip generation and trip distribution were limited. In the present study, the use of the initial trip generation regression models developed based on existing household survey data resulted in prediction errors ranging between 26% and 32%. Modeling efforts after applying corrections to zone-specific characteristics based on secondary data and the use of trip rate per household later resulted in prediction errors of less than ±5%. In the latter phase of work related to trip distribution modeling, a log-linear regression model was developed based on a smaller refined set of the revealed preference data obtained by eliminating erroneous data in a stage-wise manner. The use of the calibrated and validated model ensured that the errors in predicted trip frequencies were less than 0.6%. Here, the information on the inter-zonal aerial distances that formed part of the trip distribution model was obtained using GIS approaches that employed the moment area method, which considered the intensity of land use at the sub-zone level. The combined strategy incorporates the use of GIS-based approaches to determine inter-zonal aerial distances, and the use of the refined relationship between trip interchanges and the inter-zonal aerial distances in the development of a reliable log-linear regression model for trip distribution contributed towards attaining higher accuracies in travel demand estimation. The modeling approaches described herein do not rely on the use of sophisticated technology, and time-consuming data processing. The study will provide the basic framework for transport planners to formulate better strategies for travel demand modeling where available data is noisy and less reliable.

*Keywords:* Trip Distribution; Moment Area Method; Travel Demand Modeling; Log-Linear; Gravity Model; Regression Modeling.

## 1. Introduction

The urban transport scenario has undergone drastic changes over the past few decades. Transport planners are often challenged to provide efficient transport systems that can address the evolving mobility needs of trip-makers. To provide effective travel opportunities in the face of restrictions imposed by unchecked urbanization, it is required to exploit the potential of better approaches that can assist in decision-making in the field of urban transport management.

The conventional *four-stage travel demand modeling* approach is a powerful tool that can effectively assist in evolving strategies to rectify problems in urban transport [1]. *Trip generation* is the first stage in this exercise where *trip*

* Corresponding author: akashcv14f12@nitk.edu.in; akashnitr1987@gmail.com

http://dx.doi.org/10.28991/CEJ-2022-08-09-013

*productions* and *trip attractions* to various zones of the city are modelled, followed by *trip distribution* in the second stage, *modal split* in the third stage, and *trip assignment* in the fourth stage [2, 3].

The conventional *four-stage travel demand modeling* (TDM) process, a *systems-based approach* that gained importance worldwide in planning urban transportation systems, evolved with systematic top-down sequential procedures for travel demand forecasting [4]. However, the capability of models to replicate field observations will depend to a large extent on the accuracy of the supporting data.

One of the most important challenges in this study was that since details on the origin-destination travel matrix that describes the travel desires of the trip-makers were not available, it was imperative to synthesize the same using secondary data sources as demonstrated by Brijs et al. [5], Priyanto and Friandi [6], and Bwambale et al. [7], and by extracting relevant information from *revealed preference* (*RP*) data on inter-zonal trip movements based on household surveys as reported by Hu [8], and Bourgeat [9]. In the present study related to Amritsar city, the available *revealed preference* (RP) data on household interview surveys for modeling *trip generation* and *trip distribution* was found partially incorrect.

The primary scope of the study was focused on the development of regression-based *trip generation* models for *trip attractions* and *trip productions* by incorporating the use of secondary data such as Census data with details related to households in each zone as in investigations performed by Yue et al. [10], Chmelik [11], and Hussein et al. [12]. The present study also incorporates the development of a log-linear regression model, and its calibration and validation based on a smaller refined set of *revealed preference* (RP) obtained by eliminating erroneous data so that inter-zonal trips could be correctly synthesized.

As part of the development of a regression model for *trip distribution*, it was required to obtain details on the inter-zonal aerial distances between centroids of various zones in the city. A similar approach was adopted by several researchers including Mozolin et al. [13], Jin et al. [14], and Chmelik [11]. This was accomplished by downloading digitized high-resolution images through *google-earth* and by overlaying them with supplementary information on land use and ward boundaries. The layers of high-resolution images were processed using *Gimp*, and the *geo-referencing* of these images was performed using *QGIS.*

The present study will assist transport planners in devising strategies to refine primary household survey data that could be partially incorrect, inconsistent, and noisy to perform modeling exercises for *trip generation* and *trip distribution* as part of *travel demand modeling* (TDM). The approach adopted in this study using noisy *RP* data and secondary data such as Census data for modeling *trip generation* and *trip distribution* is unique in the absence of a readily-available OD (origin-destination) matrix. The present work will provide the basic framework for extracting details on inter-zonal aerial distances between centroids using google-earth maps and *QGIS*, a free public domain *GIS* software.

## 2. Literature Survey

In the early stages of travel demand modeling, travel demand forecast was mainly based on the extrapolation of trends and the use of growth factor methods. But the late 1940s proved that the travel desires of trip-makers traveling from one zone to another zone in the city can be represented by an origin and destination (OD) matrix table and that the land use characteristics play a major role in trip-making [15].

*Trip productions* and *trip attractions* in each zone of the city depend to a large extent on the zone-specific characteristics or variables such as population, number of households, number of employees, number of students, vehicle ownership, employment potential, commercial floor spaces, and the educational opportunities. The use of *multiple linear regression* (MLR) models for the study of trip generation based on similar zone-specific variables was demonstrated by Hill and Dodd [16] for Toronto city, and Modesto. Domencich et al. [17] suggested a regression model that considered the influence of cost of travel by various modes, the service characteristics, and other socio-economic factors on *trip generation*, *trip distribution*, and *modal choice* simultaneously. Wilson [18] pioneered the development of spatial interaction models and explained *trip generation* and *trip distribution* based on the concept of *entropy.*

Thamizharasan and Rengaraju [19] and Chari et al. [20] performed studies using regression models for predicting trip movements in Chennai and Hyderabad respectively. HATS [21] conducted an extensive study on *trip attractions* for Hyderabad city. The main objective of the study was to investigate the influence of land use characteristics, and socio-economic factors on travel demand. As part of this study, the trip production and attraction rates were computed using the *category-analysis* method, and subsequently, *multi-linear regression* models were developed to model *trip generation.* Sarna et al. [22] adopted the traditional four-stage planning process for the projection of transport demand for New Delhi for the base year 1981, where *trip generation* modeling was performed using *multi-linear regression*, while trip distribution was performed using the BPR calibration procedure. The study indicated that the road system would be insufficient to handle future traffic, and the six most preferred transport networks were identified to improve road-based transport systems.

Björketun [23] performed studies on *trip assignment* using TransCAD, a *GIS-based* software. The database on the road network and the origin-destination (OD) matrix were made available by the local authority for Linköping city in Sweden based on the centroid-to-centroid trips. The *GIS* software was found to provide a user interface that was highly capable of modeling *disaggregated trip assignment.* Varagouli et al. [24] performed studies on the application of a regression model to identify factors affecting inter-city travel demand for Xanthi in Northern Greece. Information on travel time and travel costs for various modes on the road links was obtained based on roadside interviews at junctions collected over one week during summer. The origin-destination travel matrix was computed and later verified against field data.

Kitamura [25] performed studies on the application of a dynamic simultaneous equation system for the study of car ownership, *trip generation,* and *modal split* based on household data for The Netherlands. In this study, *trip generation* was analyzed using the linear regression model. The investigations indicated that *trip generation* was influenced by the number of workers in a household, vehicle ownership, and household income. Sofia et al. [26] developed *multiple linear regression* (MLR) models using the step-wise approach for modeling household *trip generation* based on socio-economic factors for Al-Diwaniyah city in Iraq which comprised 60 zones spread over an area of 52 sq. km. The results indicated that trip productions were influenced by the household size, gender of the trip-maker, the number of working members, and the number of students in the household.

Mustafa and Zhong [27] proposed a *GIS*-based high-fidelity travel demand model (HFTDM) for the estimation of traffic volume counts on important road sections of the area of Beresford, Bathurst in New Brunswick. The GIS-based approach was adopted in the estimation of *trip generation*, *trip distribution*, and *traffic assignment* for private modes of travel. The study revealed that the *GIS* approach could perform better in the estimation of traffic counts when compared to traditional *regression-based* approaches, and *ANN.* Semeida [28] adopted the use of the *multi-linear regression* method to identify and select the main variables that affected travel demand in North East Egypt, where subsequently, the *generalized logit model* (GLM) was used to model *mode-choice preferences*. Chang et al. [29] developed and compared numerous *trip generation* models. Out of all the approaches available, it was found that the regression modeling approach and the *category analysis* method provided reliable and accurate estimates.

Urade et al. [30] developed models for estimating *trip generation* for four major categories of trips such as work trips, education trips, business trips, and shopping trips based on studies performed in 4 major zones located in the IT hub of Nagpur City. Mamei et al. [31] performed studies on modeling mobility in the regions Piemonte, Lombardia, and Emilia Romagna in Italy. Details on the trips generated and the OD matrix were obtained based on *call detail records* (CDR) and mobile position data. Individual trips were aggregated to construct the OD data for various origin-destination pairs. The studies performed by Cordera et al. [32] indicated that the use of the *gravity models* which considered the effect of spatial separation was capable of providing realistic estimates of the travel demand as demonstrated in the case of regional railway lines of Cantabria (Spain).

Calvo et al. [33] adopted a geographically weighted regression (GWR) approach to analyze the influence of socio-economic and land use factors in Metro-trips for Madrid in Spain. The study indicated that the number of Metro trips could be increased significantly by improving sustainable travel facilities in the activity zones. Doustmohammadi et al. [34] utilized the stepwise linear regression and the Bayesian linear regression approaches for developing *trip generation* equations related to freight traffic in Birmingham using GPS Census data on truck movements, and details on employees. The approach was considered appropriate for analysis related to cities with a population ranging between 200,000 and 1,000,000.

Machado and Quintanilha [35] adopted the use of a GIS-based approach to determine activity centers and zones that generate a large number of trips in Brazil incorporating changes in land use. This approach was considered less tedious when compared to the time and effort spent on traffic data collection. Yang et al. [36] demonstrated the use of a *random forest regression* approach to model *trip generation* for resident and non-resident trip-makers of 36 zones in Nanjing city in China using cell phone signals collected from 76,061 signal towers. The study indicated that the trip-making characteristics of resident and non-resident trip-makers of the city varied significantly and that mixed land use influenced *trip generation.*

Aboelenen et al. [37] adopted the traditional approach to determining *trip generation* developed for residential apartments and villas in Qatar based on household surveys. The regression models relating trips generated to household income, household size, vehicle ownership, and driving license had *R-square* values ranging between 0.54 and 0.65. The study indicated that the number of employees, students, and the number of members holding driving licenses in households influenced the *trip generation.*

Gebre and Quezon [38] too adopted the use of linear regression to develop *trip generation* models for work trips dependent on public modes for Hawassa city in Ethiopia. The independent variables used in regression models included household characteristics, socio-economic factors, and trip-making characteristics. The study indicated that trip production was significantly influenced by monthly household income.

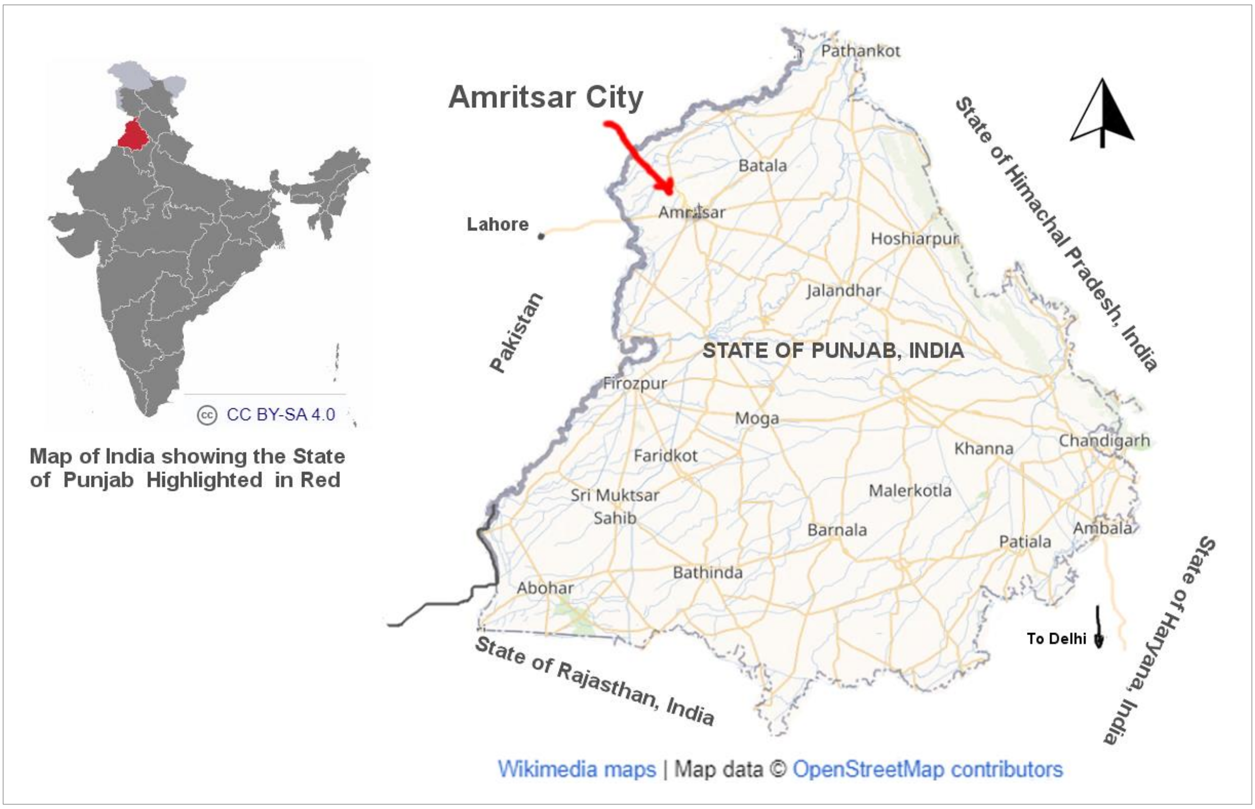
Hussein et al. [12] demonstrated the use of the category analysis approach for trip generation in addition to the use of ordered logit modeling as part of discrete choice modeling for the London Census Metropolitan Area (CMA) in Ontario, Canada. The study indicated that although the predictions made based on the discrete choice modeling approach were reliable, the category analysis method was more versatile, and could be easily adapted in trip predictions at the disaggregate and aggregate levels.

Mohd Shafie et al. [39] developed a regression-based *trip generation* model incorporating information on land use related to the location of petrol-pump stations that provided fast-food services and convenience stores in Penang State and Klang Valley in Malaysia. The study revealed that the use of the independent variable related to the *number of restaurant seats* provided better predictions when compared to the total floor area and the number of petrol-pump stations. Moreover, it was observed that incorporating information related to population and socio-economic characteristics further improved the accuracy of trip predictions.

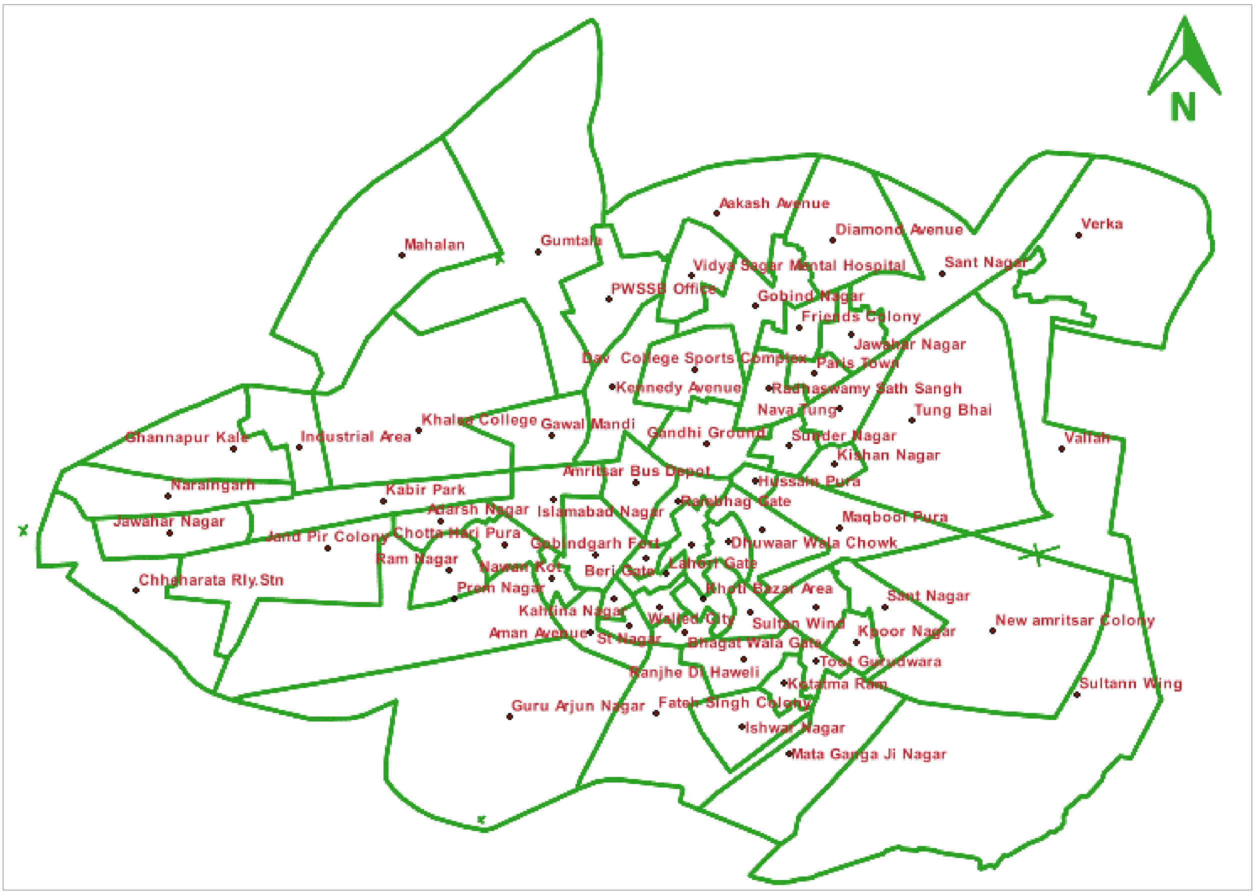
Based on a review of literature, it was observed that studies on the use of sparsely available data such as *revealed preference* data (where the available data on OD-matrix is only partially correct), supported by secondary data such as Census data (with information on zone-specific characteristics) in modeling *trip generation* and *trip distribution* were very limited.

## 3. Details on the Study Area

Situated at 31.63°N and 74.87°E with an average elevation of 234 meters, Amritsar is one of the most ancient cities in India. The city lies on the main Grand Trunk Road (GT Road) connecting Delhi and Lahore in Pakistan. Amritsar is considered the gateway for travellers coming to India by the land route from central Asia from ancient times. Among all the famous places, the Golden Temple of Amritsar is a major tourist destination and pilgrimage center that attracts thousands of people from around the world. It has an important place in the history and culture of the Sikh religion. Figure 1-a, provides details on the location of Amritsar city in the State of Punjab, India, along with the location of nearby cities (Website: Creative Commons-1, Website: Creative Commons-2) [40, 41]. Out of a total of 111 Traffic Analysis Zones (TAZs) of Amritsar city, trip movements among 65 internal zones of the municipal area were considered important in the study conducted by CES and Systra-MVA [42] as shown in Figure 1-b. A *base map* of the city with details on place names was downloaded from *google.maps* [43] at a resolution of 1:5000. This base map was developed by stitching together 515 screenshots downloaded from a *google-base map* [43] using a powerful image processing software, *Gimp* at an optimum resolution of 1:5000. The boundaries of the 65 wards were then redrawn in *Gimp* using details available from Website: MCA [44]. The wards defined by the city authorities were assumed to represent the traffic analysis zones (TAZs) often mentioned as ‘zones’ in this study.




(a)




(b)

**Figure 1. a) Map of India showing the State of Punjab highlighted on the left along with the #Map of the State of Punjab on the right showing the location of Amritsar City; b) Details on 65 zones of Amritsar city considered in this study**

## 4. Basic Theoretical Aspects

A *trip* refers to a one-way person-movement from an origin zone to a destination zone. *Trip production* refers to the number of trips originating from a zone, while *trip attraction* refers to the trips attracted by a zone. *Trip generation* is the first stage in the travel demand estimation process where *trip productions* and *trip attractions* of various activity zones of the study area are estimated. In the present study, a regression-based approach was adopted for modeling *trip generation.*

The *origin-destination* (OD) trip matrix that represents the inter-zonal trips from one zone to another is the basic data that can be used in *trip distribution*. According to the *gravity model* based on Newton's concept of gravity, the spatial distribution of trip interchanges ($t_{ij}$) between an OD pair is proportional to the *trip productions* ($O_i$) at *i*, and the *trip attractions* ($D_j$) at *j*, and is inversely proportional to a function of the spatial separation ($C_{ij}$) between the zones. The gravity model proposed by Voorhees [45] was one of the earliest models based on the above concept. The *gravity model* can be derived as follows [45]:

$$t_{ij} \propto O_i \tag{1}$$

$$t_{ij} \propto D_j \tag{2}$$

$$t_{ij} \propto 1/C_{ij}\text{, or} \tag{3}$$

$$t_{ij} \propto f\left(C_{ij}\right) \tag{4}$$

Hence,

$$t_{ij} \propto O_i . D_j . f\left(C_{ij}\right) \tag{5}$$

$$t_{ij} = K . O_i . D_j . f\left(C_{ij}\right) \tag{6}$$

where, *K* is a constant of proportionality.

# 5. Methodology Adopted

To perform a study on *trip generation* for a given study area, it is required to extract information on the *trip productions* by summing up the trips in each row of the OD matrix, and the information on *trip attractions* by summing up the trips in each column of the OD matrix. In this study, an alternative approach was adopted where information gleaned from noisy RP data, in addition to secondary data was used to develop the OD matrix.

In the present study, the *trip generation* model was developed based on reliable information on *zone-specific characteristics* such as *population*, *number of households*, *household size*, *number of students*, and the *number of employees* in each zone for the 65 *traffic analysis zones* (TAZs) of the city based on Census 2011 [46, 47] since available data on the same recorded based on *household interview surveys* (*HIS*) were not accurate.

Also, the development of a reliable model for *trip distribution* requires dependable information on *origin-destination* (OD) travel matrix. It was initially planned to use the *revealed preference* (RP) data obtained based on *household interview surveys* (*HIS*) conducted by CES and Systra-MVA compiled by Kanthi [48] to determine the OD travel matrix. Since the *RP* data collected was partly erroneous and noisy due to inconsistent sample sizes, and errors in data entry and reporting, it was required to devise a strategy to eliminate the errors from the existing data to maintain consistency and accuracy.

To obtain a reliable subset of data points as part of the OD matrix for Amritsar city using the *RP* data, it was first required to search for a trend between trip interchanges ($T_{ij}$) and the inter-zonal aerial distances ($C_{ij}$). This was achieved by eliminating intra-zonal trips, and by eliminating data points lying very close to the $x$ and $y$ axes. The emergence of a trend was checked, and further levels of refining were performed to obtain a reliable trend between $T_{ij}$ and $C_{ij}$. This part of the study focused on the development of a linear regression model based on the *gravity model*.

As part of *trip distribution* modeling, it was also required to obtain reliable information on the inter-zonal aerial distances ($C_{ij}$). Thus, it was required to obtain the coordinates of the zone-centroids using a digitized high-resolution *google-earth map* using *Gimp* and *QGIS* software. In this exercise, since most of the larger zones had mixed land uses, it was required to use the information on the intensity of activities at the sub-zone level based on visual observation where weights were assigned at the subzone level. The *moment area method* was then used to compute the coordinates of the zone-centroids for each of the zones.

The sections below provide details on the step-by-step approach adopted in the development of reliable *trip generation* and *trip distribution* models. Figure 2 provides an overview of the overall methodology adopted in the present study.

# 6. Modeling Trip Generation

## 6.1. Correction of Errors on Zone-Specific Characteristics Reported based on Household Survey

Modeling approaches related to urban transport planning and operation require reliable information on zone-specific characteristics and household-based travel demand. Such information can only be attained based on a *household interview survey* (*HIS*). It is also possible to extract information on zone-specific characteristics such as population, number of households, number of students and employees, and so on using *HIS*. However, in the present study, the *desired sample-size criterion* of 1 in 25 households (or 4% of the households) suggested by BPR [49] was not satisfied for any of the zones. Also, 37 zones of the city had weighted average sample sizes of much less than the recommended minimum sample size of 1 in 100 (or 1%) for the *revealed preference* data [49].

Also, in the case of the origin-destination (OD) matrix compiled by Kanthi [48] as shown in Table 1 using *revealed preference* (RP) data collected based on the above-mentioned *household interview surveys*, it was observed that the inter-zonal trips recorded were inconsistent with the expected trip interchanges. Closer scrutiny of the data forms indicated errors in reporting of data by the respondents, in addition to errors in data entry. Additionally, as cited above, the recommended desired sample-size criterion was not maintained in many of the zones. Table 2 provides details on *zone-specific characteristics* such as population, number of households, and so on as reported by Kanthi [48] based on *household interview surveys* undertaken by CES and Systra-MVA [42] in 2011, in addition to corrected zone-specific characteristics based on Census [46, 47]. As part of applying corrections to data related to zone-specific characteristics, it was proposed to supplement with details based on information on various wards/ zones obtained from Census [46, 47] which was more reliable.

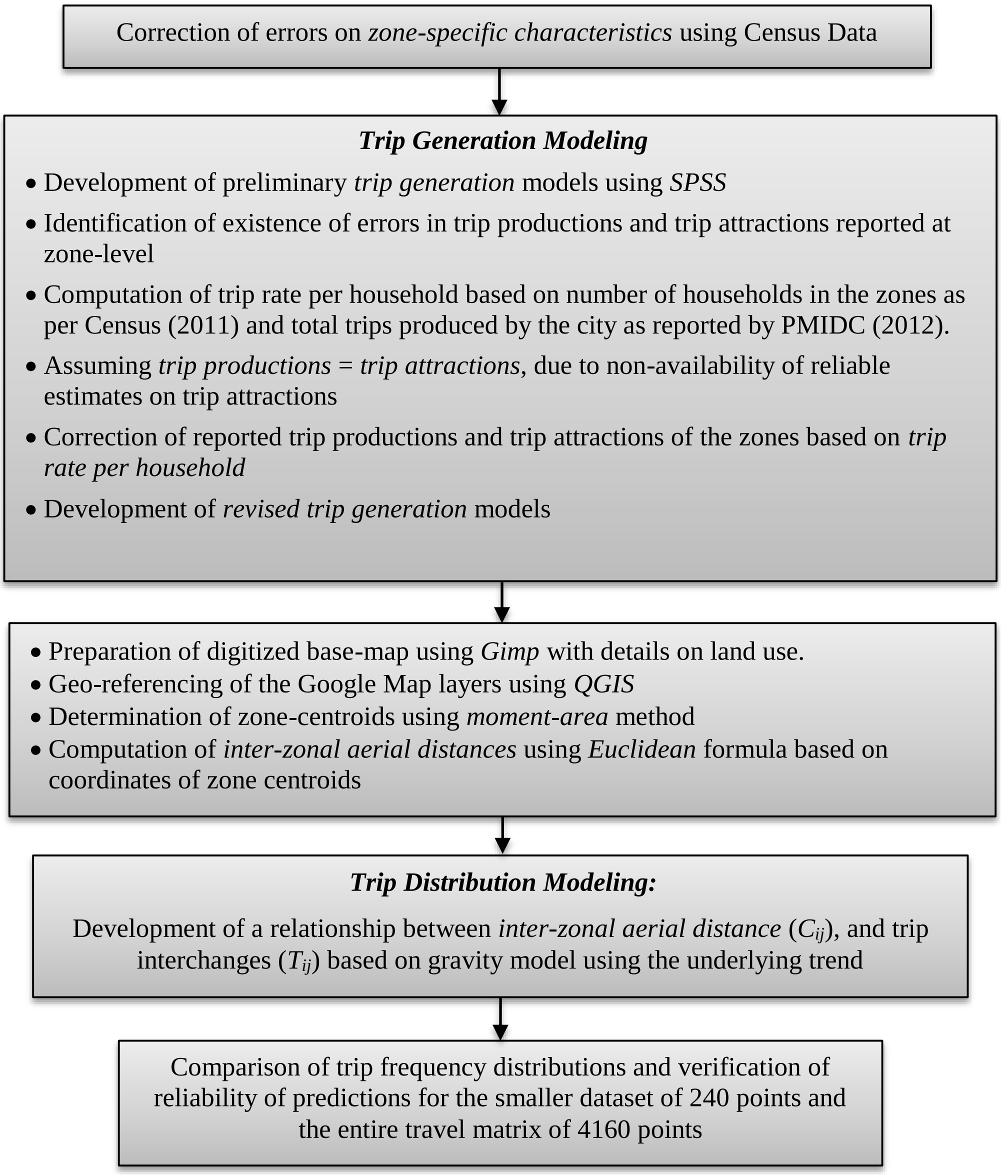



**Figure 2. An outline of the overall methodology adopted in the present study**

**Table 1. Origin-Destination trips compiled by Kanthi [48]: Based on partially erroneous RP data**

| O/D | 1 | 2 | 3 | 4 | 5 | 6 | - | 64 | 65 | Oi |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0 | 432 | 322 | 0 | 0 | 0 | - | 0 | 0 | **4873** |
| **2** | 124 | 0 | 87 | 120 | 52 | 210 | - | 387 | 219 | **9785** |
| **3** | 0 | 40 | 0 | 120 | 58 | 202 | - | 136 | 120 | **10199** |
| **4** | 0 | 73 | 105 | 0 | 1043 | 568 | - | 50 | 158 | **11490** |
| **5** | 0 | 47 | 128 | 365 | 0 | 1291 | - | 136 | 194 | **11623** |
| **6** | 0 | 160 | 115 | 599 | 1020 | 0 | - | 151 | 205 | **14556** |
| **7** | 15 | 0 | 209 | 160 | 205 | 415 | - | 62 | 73 | **15863** |
| - | - | - | - | - | - | - | - | - | - | |
| **64** | 0 | 177 | 153 | 64 | 59 | 165 | - | 0 | 137 | **12186** |
| **65** | 0 | 108 | 194 | 57 | 104 | 145 | - | 80 | 0 | **14022** |
| **Dj** | **5186** | **11076** | **11748** | **11797** | **11493** | **15094** | **-** | **11356** | **15203** | **1045672** |

**Table 2. Details on zone-specific characteristics reported by Kanthi [48], and Census [46, 47]: Partial data**

| Zone No. | Population | | No. of Households | | Household-Size | | No. of Employees | | No. of Students | | Sample Size* (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (Kanthi, 2012) | (Census, 2011) | (Kanthi, 2012) | (Census, 2011) | (Kanthi, 2012) | (Census, 2011) | (Kanthi, 2012) | (Census, 2011) | (Kanthi, 2012) | (Census, 2011) | (Kanthi, 2012) | (Census, 2011) |
| 1 | 20828 | 19747 | 3433 | 3915 | 6.07 | 5.04 | 303 | 6474 | 5264 | 6547 | 0.379 | 0.400 |
| 2 | 17765 | 16684 | 3740 | 3490 | 4.75 | 4.78 | 1990 | 6011 | 5321 | 7299 | 0.788 | 0.839 |
| 3 | 22985 | 16119 | 4732 | 3443 | 4.86 | 4.68 | 1131 | 5481 | 6084 | 7841 | 0.365 | 0.521 |
| 4 | 19347 | 17587 | 3583 | 3681 | 5.4 | 4.78 | 1666 | 6128 | 5181 | 7016 | 1.396 | 1.535 |
| 5 | 21208 | 24344 | 3900 | 4862 | 5.44 | 5.01 | 4144 | 9674 | 5063 | 9091 | 1.084 | 0.945 |
| 6 | 18235 | 17209 | 3802 | 3622 | 4.8 | 4.75 | 5541 | 6843 | 4754 | 5776 | 1.283 | 1.360 |
| 7 | 19081 | 20530 | 3388 | 4422 | 5.63 | 4.64 | 1438 | 8008 | 4527 | 7959 | 1.153 | 1.072 |
| 8 | 17673 | 14038 | 3027 | 2816 | 5.84 | 4.99 | 1487 | 5816 | 4685 | 5027 | 1.047 | 1.318 |
| - | - | - | - | - | - | - | - | - | - | - | - | - |
| 64 | 27627 | 19250 | 5133 | 4031 | 5.38 | 4.78 | 1994 | 7102 | 6375 | 8387 | 0.572 | 0.821 |
| 65 | 23317 | 16155 | 4115 | 3275 | 5.67 | 4.93 | 0 | 5450 | 6368 | 6414 | 0.678 | 0.978 |

* The sample size was computed based on data forms compiled for each zone.

* For a city with a population of 1.1 million, the minimum sample size is 1:100 and the desired sample size is 1:25 [50, 51].

## 6.2. Identification of Existence of Errors on Zone-Specific Trip Productions and Trip Attractions Reported

The traditional approach to *trip generation* modeling is based on the development of regression models to predict *trip productions* and *trip attractions* as adopted in this study. *Trip productions* are computed based on the row-totals, and *trip attractions* are obtained using the column-totals of the origin-destination matrix. The modeling of *trip productions* and *trip attractions* is performed based on independent variables such as zone-specific characteristics.

In the present study, the zone-specific characteristics recorded as part of the *household interview survey* were not accurate. Due to this reason, the corrected values of zone-specific characteristics such as population, households, employees, students, and so on recorded in Table 2 based on Census [46, 47] were used in the development of regression models.

However, while developing the regression models for *trip generation* using *trip productions* and *trip attractions* based on Table 1, and the zone-specific characteristics as given in Table 2 using the statistical software *SPSS*, it was observed that the *adjusted R-square* ($R^2_{adj}$) values for *trip productions* and *trip attractions* lay between 0.68 and 0.74. This indicated that the predictions made at the zone level using these models could result in significant errors ranging between 32% and 26%.

This also indicated that there could be inaccuracies in the reporting of values related to dependent variables such as *trip productions* and *trip attractions* which were extracted from *revealed preference* data collected as part of a *household information survey* compiled by Kanthi [48]. Due to this reason, it was required to determine the values of dependent variables such as zone-specific *trip productions* and *trip attractions* based on alternative approaches.

The following are the important regression models developed as part of this exercise along with details on the values of the adjusted $R$ square ($R_{adj}$), $F$-test value, the significance of the $F$-test ($F$ Sig), and the significance of the Student's $t$-test ($t$-sig) for the independent variables tested:

$$Prod = 0.847\ Pop \quad (R^2_{adj} = 0.691;\ F\text{-}test = 146.02;\ F \text{ Sig} = 0.00;\ t\text{-sig} = 0.00) \tag{7}$$

$$Prod = 2.263\ Nstud \quad (R^2_{adj} = 0.735;\ F\text{-}test = 181.07;\ F \text{ Sig} = 0.00;\ t\text{-sig} = 0.00) \tag{8}$$

$$Prod = 2.273\ Nemp \quad (R^2_{adj} = 0.686;\ F\text{-}test = 142.70;\ F \text{ Sig} = 0.00;\ t\text{-sig} = 0.00) \tag{9}$$

$$Prod = 4.138\ No\_HH \quad (R^2_{adj} = 0.700;\ F\text{-}test = 153.01;\ F \text{ Sig} = 0.00;\ t\text{-sig} = 0.00) \tag{10}$$

$$Attr = 0.846\ Pop \quad (R^2_{adj} = 0.684;\ F\text{-}test = 141.82;\ F \text{ Sig} = 0.00;\ t\text{-sig} = 0.00) \tag{11}$$

$$Attr = 2.258\ Nstud \quad (R^2_{adj} = 0.727;\ F\text{-}test = 174.39;\ F \text{ Sig} = 0.00;\ t\text{-sig} = 0.00) \tag{12}$$

$$Attr = 2.268\ Nemp \quad (R^2_{adj} = 0.678;\ F\text{-}test = 138.01;\ F \text{ Sig} = 0.00;\ t\text{-sig} = 0.00) \tag{13}$$

$$Attr = 4.130\ No\_HH \quad (R^2_{adj} = 0.694;\ F\text{-}test = 148.17;\ F \text{ Sig} = 0.00;\ t\text{-sig} = 0.00) \tag{14}$$

where, *Prod* = trip productions of the zones; *Attr* = trip attractions of the zones; *Pop* = population in the zones; *Nstud* = number of students in the zones; *Nemp* = number of employees in the zones; and *No_HH* = number of households in the zones.

In this exercise, the cross-correlations obtained using *SPSS* software were examined for correlations between the dependent variables *trip productions* and *trip attractions* against the independent variables consisting of corrected zone-specific characteristics such as *population*, *number of employees*, *number of students*, and *number of households*. It was found that the *cross-correlation* (*r*) values between the dependent and independent variables ranged from 0.8 to 0.9. Also, the *cross-correlation* (*r*) values between the independent variables were higher than 0.9 indicating that it would be sufficient to use any one of the above-mentioned independent variables to predict the dependent variables. Hence, it was not planned to develop multi-linear regression (MLR) models using various combinations of the independent variables. A few simple examples are illustrated below.

For example, on using a regression model to predict the trip productions (*Prod*) with the number of students in the zones (represented as *Nstud*) as an independent variable, the value of the regression coefficient for *Nstud* was obtained as 2.263 as in Equation 8 mentioned above. However, by using a combination of independent variables such as the population in the zones (denoted as *Pop*), and the number of students in the zones (represented as *Nstud*), the following regression model with the related statistics can be obtained:

$$Prod = -0.083\ Pop + 2.474\ Nstud \quad (R^2_{adj} = 0.731;\ F\text{-}test = 89.278;\ t\text{-sig Pop} = 0.77;\ t\text{-sig } Nstud = 0.002) \tag{15}$$

Here, the coefficient for the independent variable *Pop* is found to have a negative value, which is not logically correct. Also, the variable *Pop* failed in the *t-test* for significance with a value of 0.77. Additionally, the *F-test* value for the model is found to have reduced to 89.278 when compared to the same for the regressions with independent variables *Pop* and *Nstud* when used separately as in Equations 7 and 8.

Also, on using a regression model to predict the trip productions (*Prod*) with the number of employees in the zones (represented as *Nemp*) as an independent variable, the value of the regression coefficient for *Nemp* was obtained as 2.273 as in Equation 9. But by using a combination of independent variables such as the population in the zones (denoted as *Pop*), and the number of employees in the zones (represented as *Nemp*), the following regression model with the related statistics can be obtained:

$$Prod = 1.161\ Pop - 0.778\ Nemp \quad (R^2_{adj} = 0.686;\ F\text{-}test = 72.00;\ t\text{-sig Pop} = 0.299;\ t\text{-sig } Nemp = 0.778) \tag{16}$$

Here, the coefficient for the independent variable *Nemp* is found to have a negative value, which is not logically correct. Also, the variables *Pop* and *Nemp* have failed in the *t-test* for significance with values of 0.299 and 0.778 respectively. Moreover, the *F-test* value for the model is found to have reduced to 72.00 when compared to the same for the regressions with independent variables *Pop* and *Nemp* when used separately as in Equations 7 and 9.

Similar observations can be made while testing various other combinations of variables in the predictions for trip productions, and trip attractions. Due to this reason, MLR equations were not developed as part of *trip generation* modeling in this study.

### 6.3. Correction of Reported Zone-Specific Trip Productions and Trip Attractions based on Trip Rate per Household, and Development of Revised Trip Generation Models

Because of the high errors in prediction as mentioned above, it was required to adopt an alternative approach in the development of refined *trip generation* models. According to studies made by PMIDC [42, 52] in Amritsar city, the total *trips produced* by 65×65 zones of the urbanized area of the city was reported as 1,045,672. Also, according to Census [46, 47], the total households for the 65 zones of the city stood at 233,866. This indicated that the *trip rate per household* was around 4.4712. The corrected *zone-specific trip productions* can then be computed by multiplying the *trip rate per household* by the values of *zone-specific households* obtained from the Census [46, 47] as reported in Table 2. The *trip productions in the zone* were thus computed as,

$$\begin{aligned} &Trip\ productions = Trip\ rate\ per\ household \times Number\ of\ zone - specific\ households = \\ &4.4712 \times Number\ of\ zone - specific\ households \end{aligned} \tag{17}$$

Since reliable estimates of *trips-attracted* were not available as in the case of *trip productions*, it was required to assume that the *trip attractions in the zones* were equal to the *trip productions in the zones*. This is true in the case of Amritsar city as 98.36% of the total trips performed constitute work trips, educational trips, and shopping trips which are *home-based* as indicated in surveys undertaken in 2011 [42, 52].

Using the details on corrected characteristics of the zones reported in Table 2, and the newly assumed *trip productions* or, *trip attractions* of the zones (computed based on *trip rate per household*) as in Table 3, several regression models for *trip productions* and *trip attractions* were again developed using *SPSS* software. The details of these models and the important statistical indicators are provided in Table 4**.** Here, it may be observed that regression models 1, 2, and 3 that have *adjusted R-square* values greater than 0.95 can be employed to model *trip generation*. In these models, the independent variables used include the zone-specific population (*Pop*), the *number of students in the zones* (*Nstud*), and the *number of employees in the zones* (*Nemp*).

**Table 3. Dependent and independent variables for developing regressions for revised trip productions and trip attractions based on household trip rate (Partial)**

| **Zone No.** | **Production** (based on trip rate per household) | **Attraction** (based on trip rate per household) | **Population** (Census, 2011) | **No of Households** (Census, (2011) | **Household Size** (Census, 2011) | **No of Employees** (Census, 2011) | **No of Students** (Census, 2011) |
|---|---|---|---|---|---|---|---|
| **1** | 17505 | 17505 | 19747 | 3915 | 5.04 | 6474 | 6547 |
| **2** | 15605 | 15605 | 16684 | 3490 | 4.78 | 6011 | 7299 |
| **3** | 15394 | 15394 | 16119 | 3443 | 4.68 | 5481 | 7841 |
| **4** | 16459 | 16459 | 17587 | 3681 | 4.78 | 6128 | 7016 |
| **5** | 21739 | 21739 | 24344 | 4862 | 5.01 | 9674 | 9091 |
| **6** | 16195 | 16195 | 17209 | 3622 | 4.75 | 6843 | 5776 |
| **7** | 19772 | 19772 | 20530 | 4422 | 4.64 | 8008 | 7959 |
| **8** | 12591 | 12591 | 14038 | 2816 | 4.99 | 5816 | 5027 |
| **9** | 12564 | 12564 | 13511 | 2810 | 4.81 | 5196 | 6374 |
| - | - | - | - | - | - | - | - |
| **64** | 18024 | 18024 | 19250 | 4031 | 4.78 | 7102 | 8387 |
| **65** | 14643 | 14643 | 16155 | 3275 | 4.93 | 5450 | 6414 |
| **Total** | **1045672** | **1045672** | | | | | |

**Table 4. Summary of statistics for regressions to estimate trip productions and assumed trip attractions based on zone-specific characteristics**

| **Model. No.** | **Dependent Variable** | **Independent Variables** | $R^2$ | **Adj.** $R^2$ | **Std. Error** | ***F*-Test Value** | ***F* Sig** | **Coefficient** | ***t*-test Value** | ***t*-sig** |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Pro_Census Or Att_assumed | Pop | 0.999 | 0.999 | 548.302 | 58784.453 | 0.00 | 0.921 | 242.455 | 0.00 |
| 2 | Pro_Census Or Att_assumed | Nstud | 0.959 | 0.958 | 3364.314 | 1499.083 | 0.00 | 2.339 | 38.718 | 0.00 |
| 3 | Pro_Census Or Att_assumed | Nemp | 0.995 | 0.995 | 1221.33 | 11796.66 | 0.00 | 2.476 | 108.612 | 0.00 |

*The above multiple linear regression models were developed using SPSS [46, 47]; Pro_Census = Trip productions computed based on Census; Att_assumed = Trip attractions assumed equal to the computed trip productions based on Census; Pop= Population in the zone as per Census; Nemp = Number of employees as per Census; and Nstud = Number of students as per Census.*

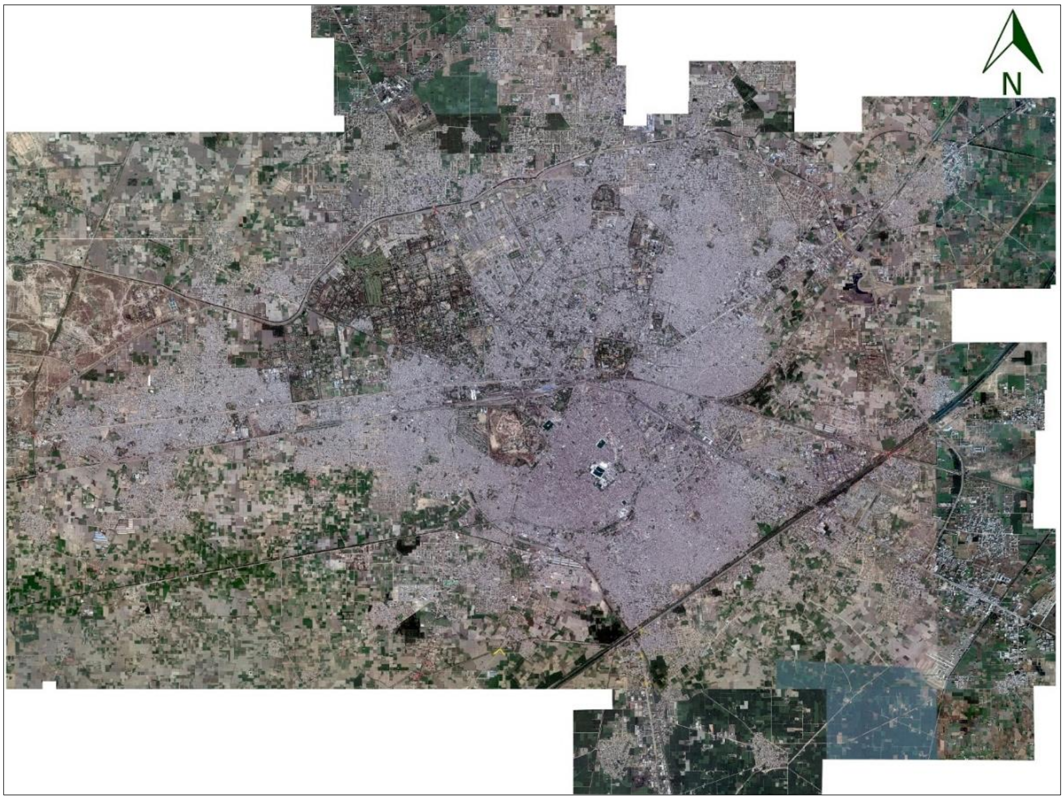
The total *trip productions* and *trip attractions* predicted based on zone-specific population (*Pop*) using *model 1* were equal to 1,042,925 which tallied closely with the actual total trips of 1,045,672 reported by PMIDC [42, 52] within an *average* or *mean absolute error* (MAE) of 2.6%. This indicated that the alternative approach based on *trip rate per household* adopted in determining zone-specific *trip productions* and *trip attractions* was acceptable.

## 7. Preparation of Digitized Images Using Land Use Data in QGIS, and Deriving Inter-Zonal Aerial Distances

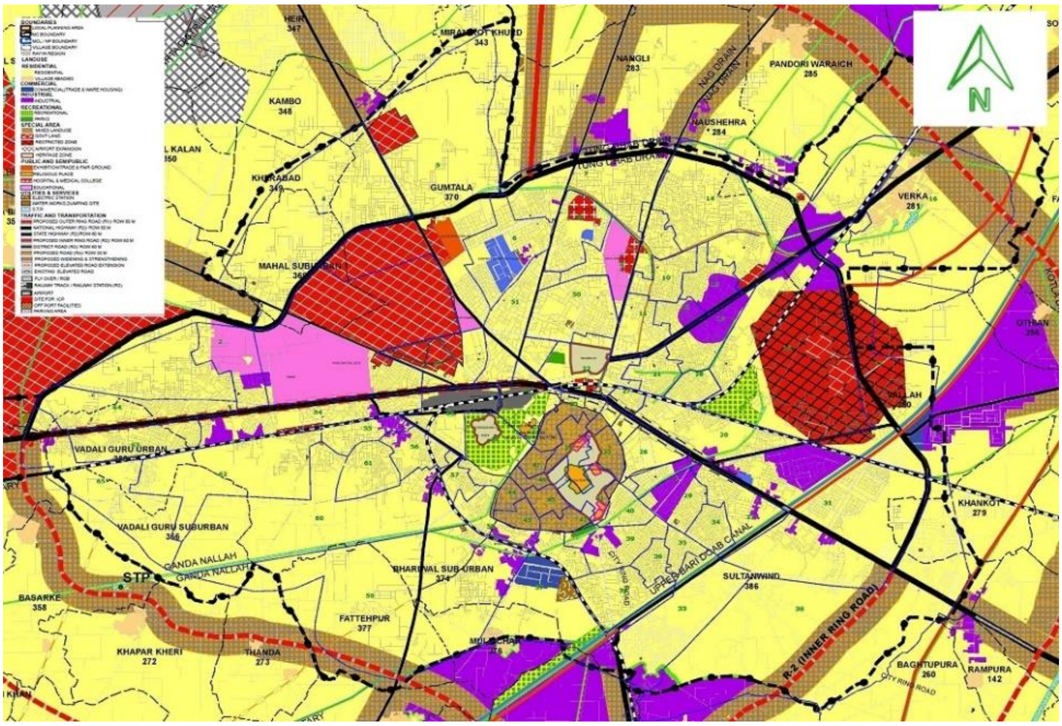
This section provides details on the preparation of a digitized *google-earth* base-map for the 65 zones of the city at a scale of 1:5000, along with details on superimposing zone boundaries using *Gimp* software. It was also required to superimpose details of the boundaries of the subzones using a land use map on the digitized *google-earth* base-map to identify the density of activities. This exercise was followed by the computation of the area covered by the sub-zones, the determination of the coordinates of the zone-centroids using the *moment area* method, and the computation of the *inter-zonal aerial distances* between the zone-centroids using the *Euclidean formula*. The resulting details on *inter-zonal aerial distances* computed will constitute an important input to the *trip distribution* model developed in the next stage.

### 7.1. Creation of Digitized Base-Map Using Gimp, and Superimposing of Land Use Maps

To perform a comprehensive study on the various activity centers in the 65 zones assigned as the TAZs (Traffic analysis Zones) of Amritsar city, a high-resolution *google-earth* base-map of scale 1:5000 encompassing 65 zones, covering an approximate area of 15×9 km was downloaded from Website: Google.maps [43]. 515 screenshots downloaded from Website: Google.maps [43] were stitched together using an image processing software such as *Gimp (GNU Image Manipulation Program)*. This *google-earth* base-map also comprised information on the density of the build-up areas of the city (See Figure 3-a).




(a)




(b)

**Figure 3. a) Base map of Amritsar city with 65 wards/ zones, b) Amritsar city urban development land use map**

Additionally, a map comprising details on the boundaries of these 65 zones/wards was downloaded from Website: MCA [44] and superimposed on the base map to create a ward-boundary map in *Gimp*. Subsequently, a land use map downloaded from Website: ADA [53] was used for reference to identify the type of land use (such as residential, commercial, industrial, recreational, religious, and educational land uses in addition to government-owned, and mixed land use areas) in each sub-zone of the zones, and also to identify the boundaries of the sub-zones (See Figure 3-b).

The centroids of the sub-zones were then located on a separate layer in *Gimp* based on a visual inspection of the density of activities and the location of major sub-activity centers in the sub-zones. Each sub-zone was assigned a particular weight based on the density of activities and land use. Each of the image layers including the base map was then saved in a *.tif* format.

### 7.2. Geo-Referencing of the Image Layers Using *QGIS*

To obtain the coordinates of the sub-zone centroids, and to obtain the land areas under each sub-zone, it was required to geo-reference various layers of image files using *GIS*-based software namely *QGIS*. The *.tif* image layers in *Gimp* were exported to *QGIS* and geo-referenced to create the *modified geo-referenced .tif* files.

These geo-referenced files were reloaded as *raster* files in *QGIS* and polygonized, and the *.shp* files were generated for each of the layers.

The above *geo-referenced .tif* files and the *.shp* files generated by *QGIS* were again re-opened in *QGIS*, and the area of the sub-zones was determined. The coordinates of the zone-centroids were then found using the "Moment Area Method" in which the coordinates of the sub-zone centroids, the areas of the sub-zones, and the weightage given to each sub-zone were used. The formula for the computation of *X* and *Y* coordinates of the zone-centroids is provided below:

$$X = (w_1 a_1 x_1 + w_2 a_2 x_2 + w_3 a_3 x_3) / (w_1 a_1 + w_2 a_2 + w_3 a_3) \tag{18}$$

$$Y = (w_1 a_1 y_1 + w_2 a_2 y_2 + w_3 a_3 y_3) / (w_1 a_1 + w_2 a_2 + w_3 a_3) \tag{19}$$

where, $x_1$, $x_2$, $x_3$, $y_1$, $y_2$, $y_3$ = $x$ and $y$ coordinates of the sub-zone centroids of a zone; $w_1$, $w_2$, $w_3$ = weightage given to each sub-zones 1, 2 and 3 of the corresponding zone respectively; $a_1$, $a_2$, $a_3$= areas of the sub-zones 1, 2 and 3 respectively.

Table 5 gives the details on sub-zone activity centers and computations for the zone centroids. The inter-zonal aerial distances between the zone-centroids were computed using a spread-sheet application using the *Euclidean* formula,

$$\text{Distance} = [(x_1 - x_2)^2 + (y_1 - y_2)^2]^{1/2} \tag{20}$$

where ($x_1$, $y_1$) represents the centroid of the origin zone, and ($x_2$, $y_2$) represents the centroid of the destination zone.

**Table 5. Details on sub-zone activity centers and zone-centroid computations: Partial data**

| Ward No. | Activity Center | | | | | Sub- Zone Co-ordinates obtained from *QGIS* | | Zone-Centroid Co-ordinates | |
|---|---|---|---|---|---|---|---|---|---|
| | Sub-Zone | Type of Development | Area (sq. Km) | Weightage | Weighted Area | $x_i$ | $y_i$ | *X* | *Y* |
| 1 | a | Residential | 1.381 | 0.25 | 0.345 | 2539 | 5515 | 3027 | 5688 |
| | b | Residential | 0.370 | 1.00 | 0.370 | 2987 | 6136 | | |
| | c | Residential | 0.509 | 1.00 | 0.509 | 3387 | 5480 | | |
| 2 | a | Residential | 0.181 | 1.00 | 0.181 | 4131 | 5227 | 3968 | 5711 |
| | b | Industrial | 0.139 | 0.75 | 0.104 | 4074 | 5527 | | |
| | c | Educational | 0.152 | 0.75 | 0.114 | 4157 | 5796 | | |
| | d | Educational | 0.698 | 0.25 | 0.175 | 3611 | 6266 | | |
| … | … | … | … | … | … | … | … | … | … |
| 64 | a | Residential | 0.778 | 1.00 | 0.778 | 2565 | 4989 | 2091 | 5015 |
| | b | Residential | 0.912 | 0.50 | 0.456 | 1284 | 5060 | | |
| 65 | a | Residential | 1.037 | 1.00 | 1.037 | 2314 | 3118 | 1630 | 3641 |
| | b | Residential | 1.021 | 0.50 | 0.511 | 1676 | 3586 | | |
| | c | Residential | 0.717 | 1.00 | 0.717 | 608 | 4437 | | |

Table 6 gives details on the inter-zonal aerial distances computed using the *Euclidean* formula.

**Table 6. Inter-zonal aerial distances based on centroid co-ordinates obtained from QGIS and the Euclidean formula: Partial data**

| O/D | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | … | 64 | 65 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0.000 | 0.941 | 2.698 | 3.705 | 5.243 | 5.825 | 7.756 | 7.059 | 7.806 | … | 1.152 | 2.478 |
| **2** | 0.941 | 0.000 | 1.761 | 3.148 | 4.469 | 4.954 | 6.911 | 6.179 | 6.896 | … | 2.001 | 3.122 |
| **3** | 2.698 | 1.761 | 0.000 | 2.540 | 3.102 | 3.320 | 5.308 | 4.507 | 5.165 | … | 3.740 | 4.692 |
| **4** | 3.705 | 3.148 | 2.540 | 0.000 | 1.968 | 3.040 | 4.582 | 4.182 | 5.152 | … | 4.834 | 6.173 |
| **5** | 5.243 | 4.469 | 3.102 | 1.968 | 0.000 | 1.218 | 2.632 | 2.232 | 3.232 | … | 6.394 | 7.589 |
| **6** | 5.825 | 4.954 | 3.320 | 3.040 | 1.218 | 0.000 | 1.990 | 1.240 | 2.127 | … | 6.949 | 8.003 |
| **7** | 7.756 | 6.911 | 5.308 | 4.582 | 2.632 | 1.990 | 0.000 | 0.970 | 1.448 | … | 8.894 | 9.983 |
| **8** | 7.059 | 6.179 | 4.507 | 4.182 | 2.232 | 1.240 | 0.970 | 0.000 | 1.028 | … | 8.177 | 9.199 |
| **9** | 7.806 | 6.896 | 5.165 | 5.152 | 3.232 | 2.127 | 1.448 | 1.028 | 0.000 | … | 8.895 | 9.824 |
| **10** | 8.355 | 7.432 | 5.682 | 5.828 | 3.923 | 2.790 | 2.020 | 1.728 | 0.701 | … | 9.422 | 10.293 |
| **11** | 7.772 | 6.835 | 5.074 | 5.625 | 3.863 | 2.649 | 2.620 | 1.965 | 1.190 | … | 8.795 | 9.576 |
| **12** | 8.454 | 7.518 | 5.757 | 6.192 | 4.359 | 3.167 | 2.702 | 2.272 | 1.291 | … | 9.484 | 10.269 |
| **13** | 9.071 | 8.143 | 6.387 | 6.583 | 4.671 | 3.546 | 2.598 | 2.459 | 1.440 | … | 10.126 | 10.961 |
| **14** | 9.171 | 8.280 | 6.577 | 6.233 | 4.265 | 3.363 | 1.727 | 2.122 | 1.483 | … | 10.280 | 11.259 |
| **…** | … | … | … | … | … | … | … | … | … | … | … | … |
| **64** | 1.152 | 2.001 | 3.740 | 4.834 | 6.394 | 6.949 | 8.894 | 8.177 | 8.895 | … | 0.000 | 1.449 |
| **65** | 2.478 | 3.122 | 4.692 | 6.173 | 7.589 | 8.003 | 9.983 | 9.199 | 9.824 | … | 1.449 | 0.000 |

## 8. Modeling Trip Distribution

The next stage in the *travel demand modeling* (TDM) exercise is related to the estimation of the inter-zonal *trip distribution* matrix or *the trip-interchange* matrix. Modeling of *trip interchanges* assumes great importance as it can provide the platform for deriving answers to several questions in the field of urban transport planning and management. The *trip distribution* matrix forms the basis for performing analyses related to *modal-split* modeling (where mode-choices of trip-makers are modelled), followed by *trip-assignment* (where trips by various modes for each origin-destination pair are assigned along selected routes). But the biggest challenge in the present study was that details on the *trip-interchange* matrix were not available. An alternative approach was hence adopted where information extracted from partially correct and noisy *RP* data, in addition to secondary data was used to build up the OD matrix. This section provides details on various steps involved in the development of a reliable *trip-interchange* matrix using the *regression* approach based on data related to *traffic analysis zones* (TAZ).

### 8.1. Development of a Suitable Relationship between Inter-Zonal Aerial Distance ($C_{ij}$), and Trip Interchanges ($T_{ij}$) based on the Gravity Model using the Underlying Trend

Based on the fundamentals of the *gravity model* proposed by Newton, and based on the contribution of researchers in the field of transport modeling including Voorhees [45], the gravity model as derived in Equation 6 can be reproduced as: $t_{ij} = K.O_i.D_j.f(C_{ij})$. where, $t_{ij}$ = spatial distribution of trip interchanges from the origin zone *i* to the destination zone *j*; $O_i$ = trip productions at zone *i*, $D_j$ = trip attractions at zone *j*; $C_{ij}$ = spatial separation between the two zones *i* and *j*; and *K* = constant of proportionality. In the present study, it was proposed to adopt the use of a log-linear model based on the following assumption to model the trip interchanges ($T_{ij}$):

$$T_{ij} = K.Pro^p.Att^q.f(C_{ij}) \tag{21}$$

where, $T_{ij}$ = spatial distribution of trip interchanges from the origin zone *i* to the destination zone *j*; *Pro* = trip productions at zone *i* computed based on trip rate per household as shown in Table 3; *Att* = trip attractions at zone *j* as in Table 3; $f(C_{ij})$ = a function of the spatial separation between the two zones *i* and *j*; and *K* = constant of proportionality.

To develop a regression model for trip distribution, it is necessary to have reliable information on *origin-destination* (OD) data. However, the data on inter-zonal trip interchanges compiled by Kanthi [48] as in Table 1 in Section 6.1 based on *household interview surveys* (*HIS*), was not reliable due to inconsistent sample sizes and errors due to data-entry and reporting as mentioned earlier. Thus, it was required to extract a smaller set of real-life data that could effectively describe the underlying trend in trip-making by eliminating extreme data points in the scatter plot.

In the search for an underlying trend between $T_{ij}$ and $C_{ij}$, the scatter plots and the resulting trend equations were examined using *SPSS* at various stages of elimination of extreme data points. In the initial stage, the scatter plot for $T_{ij}$ Vs $C_{ij}$ for 4225 data points representing 65x65 cells of the OD matrix (as in Table 1) did not show a clear trend. The *adjusted R square* value for the *cubic* trend was quite low at around 0.066, while the *quadratic* and *linear* trends had lower *adjusted R square* values of 0.063 and 0.048 respectively. Figure 4 provides details on the *cubic* trend at the initial stage of filtering the dataset.

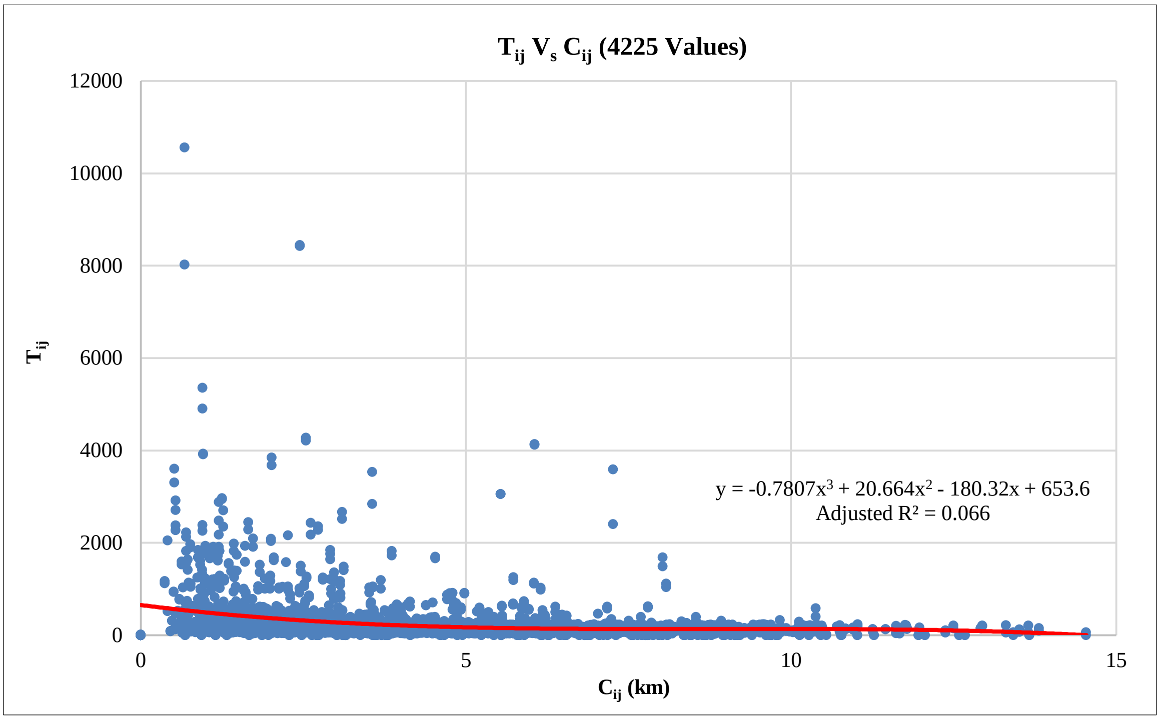



**Figure 4. Scatter plot obtained using all 65×65 data points of the original OD matrix**

Various levels of corrections were performed where the scatter points lying very close to the *x* and *y* axes were eliminated since these points would not contribute to the development of a meaningful trend. While deleting these extreme data points in various stages, it was ensured that the scatter plot remained closer to the *cubic* function.

of 240 data points were identified that followed the underlying *cubic* trend with an *adjusted R-square* value of 0.983, a high *F*-test value of 4679.306, and an *F* significance value of 0.00. See Figure 5. The details of the cubic trend followed by the refined scatter plot along with the important statistical details are given below:

$$T_{ij} = \mathrm{b}_0 + \left(\mathrm{b}_1 \times C_{ij}\right) + \left(\mathrm{b}_2 \times C_{ij}^2\right) + \left(\mathrm{b}_3 \times C_{ij}^3\right) \tag{22}$$

($R_{adj}$ = 0.983; *F* Sig = 0.00; *t*-sig for $\mathrm{b}_0$, $\mathrm{b}_1$, $\mathrm{b}_2$, and $\mathrm{b}_3$ = 0.00)

where, the values of regression coefficients are given as, $\mathrm{b}_0$ = 723.91; $\mathrm{b}_1$ = -160.14; $\mathrm{b}_2$ = 12.81; and $\mathrm{b}_3$ = -0.338; $T_{ij}$ = trip interchanges between origin zone *i* and destination zone *j*; and $C_{ij}$= inter-zonal distance between centroids of origin zone *i* and destination zone *j*.

Out of the 240 data points, 180 data points were set apart for calibration of the gravity model developed, while 60 data points were set apart for validation. The information on the 180 datasets related to *inter-zonal trip interchanges* ($T_{ij}$) obtained from Table 1 was compiled based on *household interview surveys* by Kanthi [48], the data on *trip productions in the zones* (*Pro*), and *trip attractions in the zones* (*Att*) computed as in Table 3 based on *trip rate per household*, and the *inter-zonal aerial distances between zone-centroids* ($C_{ij}$) computed using GIS approaches as summarized in Table 6, were then used to develop the following log-linear expression using *SPSS* software.

$$\mathrm{Ln}\left(T_{ij}\right) = -1.285 \times \mathrm{Ln}\left(C_{ij}+1\right) + 0.344 \times \mathrm{Ln}\,(Pro) + 0.428 \times \mathrm{Ln}\,(Att) \tag{23}$$

$$T_{ij} = Pro^{0.344}.Att^{0.428}.f\left(C_{ij}\right) \tag{24}$$

where, $T_{ij}$= Trip inter-changes between origin *i* and destination *j*; $C_{ij}$= Inter-zonal distance between origin *i* and destination *j*; *Pro*= Production of zone *i*; *Att*= Attraction of zone *j*; and $f(C_{ij}) = 1/\,(C_{ij}+1)^{1.285}$. Here, *K* = constant of proportionality = 1.

The above expressions developed using 180 datasets for calibration show that $T_{ij}$ is proportional to the *trip production in the zones* (*Pro*) and the *trip attraction in the zones* (*Att*) and that it is inversely proportional to a function of the *inter-zonal distance* ($C_{ij}$). Thus, the formulation of the regression models developed above is logically correct and follows the form of the generalized expression for the *gravity model* as in Equation 6.

The *adjusted R-square* value for the log-linear model was 0.99 with a high *F*-test value of 48354.475 and a very low *F* Sig value of 0.00. The *t-Sig* values for the coefficients of all the variables $C_{ij}$, *Pro*, and *Att* were much lower than 0.05. The low values of *t*-significance indicated that the coefficients were indeed meaningful. The above regression model as shown in Equation 24 was then used to predict the *inter-zonal trip interchanges* ($T_{ij}$) for the 180 datasets kept apart for calibration, and the 60 datasets set apart for validation. The results indicated that the *average* or *mean absolute errors* (*MAE*) in the prediction of $T_{ij}$ using the log-linear model were 15.18% and 16.06% respectively for the calibration and the validation exercises. See Table 7.

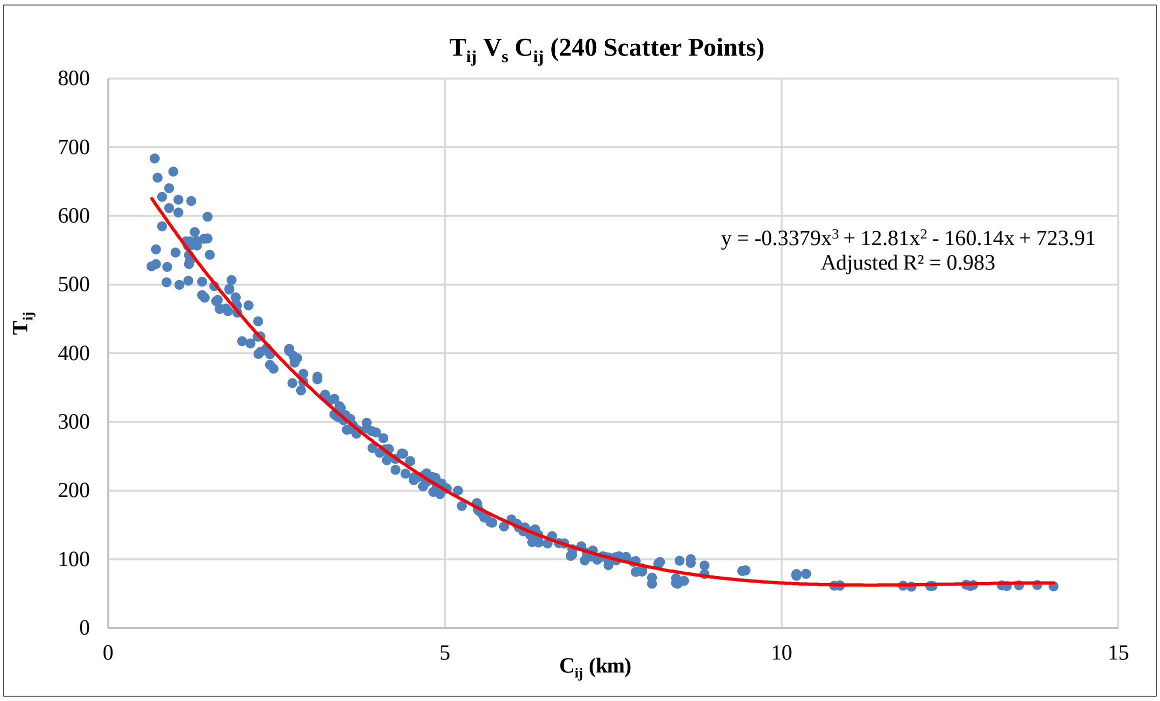



**Figure 5. Scatter plot obtained using 240 data points of the OD matrix**

**Table 7. Details on error computations for calibration and validation datasets for regression models developed using *SPSS*: $T_{ij}$ Vs $C_{ij}$, *Pro*, and *Att*: *Partial data***

| Calibration (180 Datasets) | | | | | | | Validation (60 Datasets) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Cell No. (Row ×Col)** | $T_{ij}$ | $C_{ij}$ | ***Pro*** | ***Att*** | ***Pre_$T_{ij}$*** | **Abs. Error** | **Cell No. (Row x Col)** | $T_{ij}$ | $C_{ij}$ | ***Pro*** | ***Att*** | ***Pre_$T_{ij}$*** | **Abs. Error** |
| 23×49 | 526 | 0.649 | 14116 | 14630 | 845 | 60.51 | 19×21 | 611 | 0.911 | 17393 | 14643 | 771 | 26.07 |
| 40×41 | 683 | 0.697 | 15220 | 14433 | 831 | 21.59 | 21×19 | 640 | 0.911 | 14643 | 17393 | 782 | 22.19 |
| 45×47 | 529 | 0.714 | 15363 | 14357 | 821 | 55.12 | 42×40 | 664 | 0.974 | 15386 | 15220 | 722 | 8.60 |
| 47×45 | 551 | 0.714 | 14357 | 15363 | 826 | 49.82 | 61×62 | 461 | 1.783 | 23139 | 19316 | 599 | 29.87 |
| 48×47 | 655 | 0.741 | 14330 | 14357 | 786 | 19.87 | 62×61 | 461 | 1.783 | 19316 | 23139 | 608 | 31.88 |
| 42×43 | 627 | 0.804 | 15386 | 13526 | 750 | 19.50 | 55×57 | 492 | 1.804 | 16195 | 14782 | 468 | 4.89 |
| 43×42 | 584 | 0.804 | 13526 | 15386 | 758 | 29.71 | 57×55 | 494 | 1.804 | 14782 | 16195 | 472 | 4.41 |
| 54×55 | 503 | 0.871 | 19262 | 16195 | 835 | 65.92 | 53×51 | 506 | 1.837 | 12654 | 13955 | 414 | 18.26 |
| 42×46 | 525 | 0.883 | 15386 | 11183 | 654 | 24.54 | 23×42 | 481 | 1.898 | 14116 | 15386 | 436 | 9.39 |
| - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 31×65 | 61 | 12.214 | 20545 | 14643 | 66 | 8.41 | 65×50 | 100 | 8.658 | 14643 | 13454 | 93 | 6.96 |
| 15×65 | 61 | 12.247 | 20062 | 14643 | 66 | 7.48 | 2×18 | 78 | 8.86 | 15605 | 24637 | 120 | 53.54 |
| 1×16 | 63 | 12.744 | 17505 | 16280 | 62 | 0.86 | 18×2 | 91 | 8.86 | 24637 | 15605 | 116 | 27.10 |
| 1×30 | 61 | 12.808 | 17505 | 16892 | 63 | 3.81 | 29×64 | 84 | 9.471 | 12976 | 18024 | 91 | 8.98 |
| 30×63 | 62 | 13.275 | 16892 | 16248 | 59 | 5.30 | 65×13 | 61 | 10.87 | 14643 | 25482 | 95 | 54.28 |
| 64×30 | 61 | 13.350 | 18024 | 16892 | 61 | 0.79 | 63×18 | 62 | 10.873 | 16248 | 24637 | 97 | 55.61 |
| 17×65 | 62 | 13.528 | 13056 | 14643 | 50 | 18.53 | 30×1 | 62 | 12.851 | 16892 | 17505 | 70 | 12.26 |
| 16×65 | 60 | 14.045 | 16280 | 14643 | 52 | 14.10 | 63×16 | 62 | 13.8 | 16248 | 16280 | 61 | 1.08 |
| **Average Absolute Error / MAE** | | | | | | **15.18** | **Average Absolute Error /MAE** | | | | | | **16.06** |

*$T_{ij}$= Trip inter-changes between Origin 'i' and Destination 'j';* *Pro= Production of zone 'i';*
*$C_{ij}$= Inter-zonal distance between Origin 'i' and Destination 'j';* *Att= Attraction of zone 'j';*
*Pre_$T_{ij}$= Predicted Trip inter-changes between i and j*
*Abs. Error = Absolute error between actual and predicted trip interchanges*
***Regression Model:*** Log Linear Model $T_{ij} = Pro^{0.344} \times Att^{0.428} / C_{ij}^{1.285}$

Using the above-developed *log-linear model*, the *inter-zonal trip interchanges* for all the 4160 cells (total of 65 x 65 cells - 65 intra-zonal cells on the diagonal of the OD matrix) of the OD matrix were predicted. The total number of *predicted total trip productions* for all the zones put together was 1,082,986 which tallied closely with the *actual total trip productions* of 1,045,672 as reported by PMIDC [52] within an error of 3.57%. These errors were within a limit of ±5%. Hence, it can be said that the regression model developed is reliable and provides predictions close to actual observations.

## 8.2. Comparing Trip Frequency Distributions and Verification of the Reliability of Predictions for the Smaller Dataset of 240 Data-Points, and Also for the 4160 Data-Points of the Entire Travel Matrix

A comparison of the observed and predicted trip frequencies for various distance ranges for the city as shown in Table 8 was performed to test the reliability of the model developed. Table 8 provides information on the percentage of *actual trip frequencies* for various distance ranges for the smaller dataset comprising 240 data points along with comparisons to the percentage of *predicted trip frequencies*. The *average* or *mean absolute error* between the observed and predicted *percentages of trip frequencies* for various distance ranges is 0.42% as shown in Table 8. Figure 6 provides a graphical representation of the same. Here, it can be observed that the differences between the observed and predicted trip frequencies for various distance ranges for the smaller set of data points are much less than ±3% as stipulated by BPR [49].

**Table 8. Trip frequency distribution and prediction errors for various distance ranges for the smaller set of data: *Log-linear Model* (*Partial*)**

| Inter-zonal Distance Range | | Mid-Range Distance (Km) | Trip Frequency Actual | Actual Trip Frequency (%) | Trip Frequency Predicted | Predicted Trip Frequency (%) | Error in Prediction (%) | Absolute Error in Prediction (%) |
|---|---|---|---|---|---|---|---|---|
| From | To | | | | | | | |
| 0 | 0.5 | 0.25 | 0 | 0 | 0 | 0 | 0.00 | 0.00 |
| 0.5 | 1 | 0.75 | 6438 | 10.12 | 8620 | 13.59 | -3.47 | 3.47 |
| 1 | 1.5 | 1.25 | 12241 | 19.23 | 13101 | 20.65 | -1.42 | 1.42 |
| 1.5 | 2 | 1.75 | 7286 | 11.45 | 7141 | 11.26 | 0.19 | 0.19 |
| 2 | 2.5 | 2.25 | 6739 | 10.59 | 6093 | 9.60 | 0.98 | 0.98 |
| 2.5 | 3 | 2.75 | 1887 | 2.97 | 1549 | 2.44 | 0.52 | 0.52 |
| - | - | - | - | - | - | - | - | - |
| 13 | 13.5 | 13.25 | 0 | 0.00 | 0 | 0.00 | 0.00 | 0.00 |
| 13.5 | 14 | 13.75 | 247 | 0.39 | 224 | 0.35 | 0.03 | 0.03 |
| 14 | 14.5 | 14.25 | 0 | 0.00 | 0 | 0.00 | 0.00 | 0.00 |
| 14.5 | 15 | 14.75 | 60 | 0.09 | 50 | 0.08 | 0.02 | 0.02 |
| | | **Total** | **63645** | **100** | **63438** | **100** | **Average** | **0.42** |

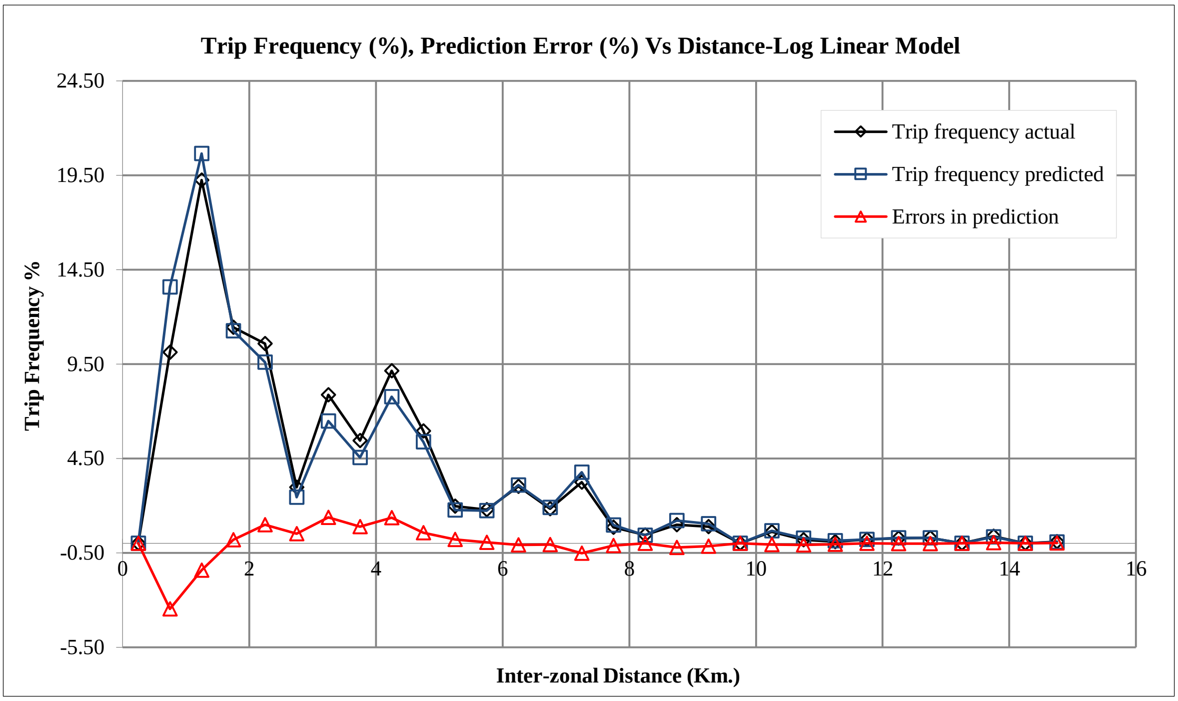



**Figure 6. Trip frequency distribution and prediction errors for various distance ranges for the smaller dataset**

In the next step, it was proposed to compare the observed trip-frequency distributions for various distance ranges for the 4160 inter-zonal cells of the OD matrix compiled based on *household interview surveys* to the trip frequencies predicted using the log-linear model. Table 9 provides a summary of computations related to the trip-frequency distributions. Here, it can also be observed that the *average* or *mean absolute error* between the observed and predicted *percentages of trip frequencies* for various distance ranges for the 4160 data points of the entire OD matrix is 0.52%. Additionally, the differences between the observed and predicted trip frequencies for various distance ranges for the smaller set of data points are much less than ±3% as stipulated by BPR [49]. Figure 7 provides a graphical representation of the actual and predicted trip frequencies in addition to the errors in prediction.

**Table 9. Trip frequency distribution and prediction errors for various distance ranges for the entire 4160 datasets (*Partial*)**

| Inter-zonal Distance Range | | Mid-Range Distance (Km) | Trip Frequency Actual | Actual Trip Frequency (%) | Trip Frequency Predicted | Predicted Trip Frequency (%) | Error in Prediction (%) | Absolute Error in Prediction (%) |
|---|---|---|---|---|---|---|---|---|
| From | To | | | | | | | |
| 0.0 | 0.5 | 0.3 | 5659 | 0.54 | 8075 | 0.75 | -0.20 | 0.20 |
| 0.5 | 1.0 | 0.8 | 140045 | 13.39 | 116252 | 10.73 | 2.66 | 2.66 |
| 1.0 | 1.5 | 1.3 | 123324 | 11.79 | 135086 | 12.47 | -0.68 | 0.68 |
| 1.5 | 2.0 | 1.8 | 82497 | 7.89 | 113678 | 10.50 | -2.61 | 2.61 |
| 2.0 | 2.5 | 2.3 | 108921 | 10.42 | 117744 | 10.87 | -0.46 | 0.46 |
| - | - | - | - | - | - | - | - | - |
| 13.0 | 13.5 | 13.3 | 267 | 0.03 | 238 | 0.02 | 0.00 | 0.00 |
| 13.5 | 14.0 | 13.8 | 781 | 0.07 | 436 | 0.04 | 0.03 | 0.03 |
| 14.0 | 14.5 | 14.3 | 0 | 0.00 | 0 | 0.00 | 0.00 | 0.00 |
| 14.5 | 15.0 | 14.8 | 0 | 0.00 | 100 | 0.01 | -0.01 | 0.01 |
| | | **Total** | **1045672** | **100** | **1082986** | **100** | **Average** | **0.52** |

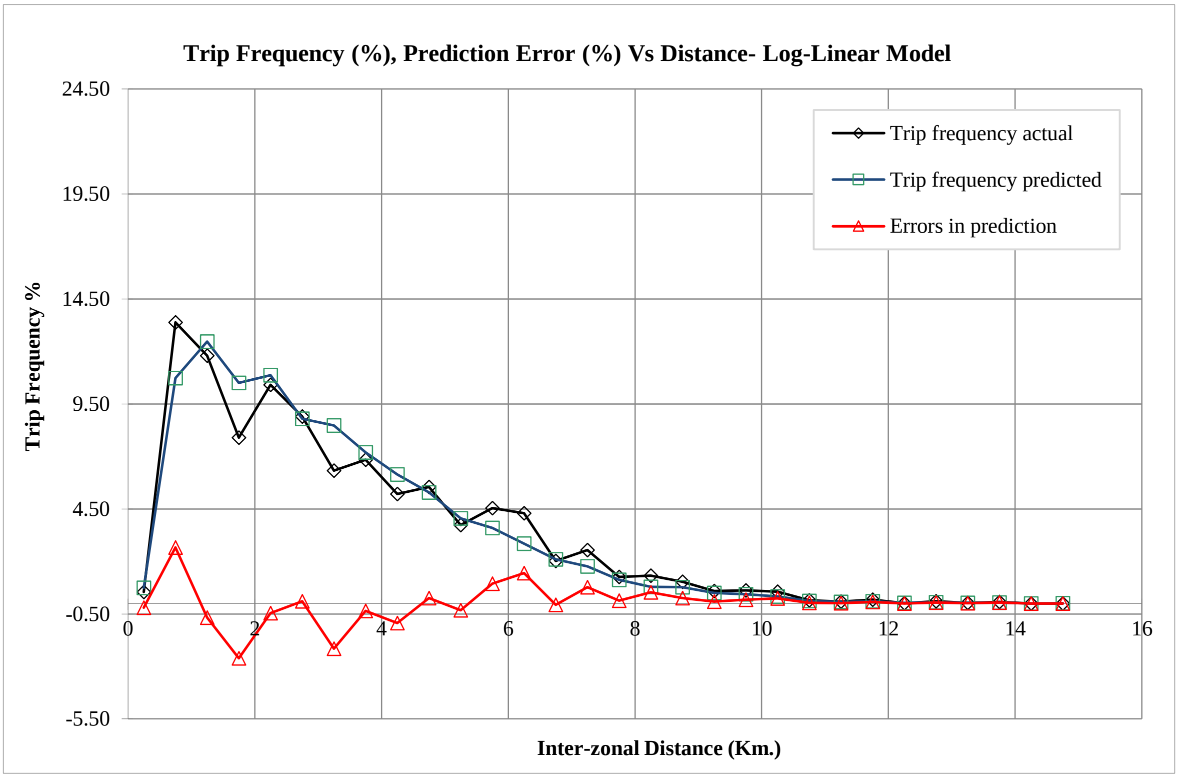



**Figure 7. Trip frequency distribution and prediction errors for various distance ranges for the entire 4160 inter-zonal data**

It may also be observed in Table 9 that the predicted *total trip production* for all the zones put together for all distance ranges is 1,082,986 which tallies closely with the actual *total trip production* of 1,045,672 as reported by PMIDC [52] and that the prediction error is 3.57%. Thus, the predictions of trip interchanges using the log-linear regression as part of the *trip distribution* exercise are considered acceptable, and within tolerable ranges.

## 9. Conclusions

Because of the non-availability of reliable data on the *origin-destination* (OD) travel matrix, the present study focused on the development of an alternative approach for modeling *trip generation* (involving *trip production* and *trip attractions*) and *trip distribution*. The *revealed-preference* (RP) data for the 65 zones of the city compiled based on a *household interview survey* (*HIS*) used in this study was also not completely dependable as the *desired* and the *minimum sample-size criteria* were not fulfilled for a majority of the zones. Hence, it was necessary to extract necessary information from the noisy *revealed-preference* (RP) data after correction of errors due to data entry and reporting. Information based on secondary data such as the Census data was used in applying corrections to zone-specific characteristics as part of the overall approach.

The initial phase of this study focused on *trip generation* modeling where regression models for *trip productions* and *trip attractions* were developed as in Equations 8-14 with *adjusted R-square* ($R^2_{adj}$) values between 0.68 and 0.74 using corrected values of *zone-specific characteristics* such as the population, number of households, number of students, number of employees, and so on. The trip production and trip attractions estimated using these models were not accurate as they resulted in prediction errors ranging from 26% to 32%. This indicated that there could be inaccuracies in the reporting of values related to dependent variables such as *trip productions* and *trip attractions,* which were extracted from the *revealed preference* data collected.

The errors in the values of *trip production* and *trip attractions* used as dependent variables were then corrected using a modified approach involving the computation of trip production and trip attractions based on *trip rate per household*. The resulting regression models developed are given in Table 4. These models indicate that the independent variables such as the population (*Pop*), the number of employees (*Nemp*), and the number of students (*Nstud*) in each zone have a significant influence on the prediction of trip production and trip attractions when used separately in the regression models. The high *adjusted R-square* values of more than 0.95 indicate that errors in the prediction of *trip generation* could be contained within +/- 5%. This approach resulted in a significant improvement in predictions related to *trip generation*. These regression models also indicate that an increase in the population of the zone by 1000 is likely to increase the number of total trips produced by 921. A similar increase in the number of students is expected to generate 2339 trips, and an increase in the number of employees by 1000 is predicted to generate 2476 trips.

Similar studies using the multiple linear regression (MLR) approach for *trip generation* were performed by Sofia et al. [26] for 70 zones in Al Diwaniyah city in Iraq. Here, the independent variables such as the number of students per household, car ownership, area of the dwelling, and so on were found to have a significant influence on the overall trips produced. The approach using MLR involving several independent variables was found necessary as the inter-correlation among independent variables was quite low. The *adjusted R-square* values of the MLR models were found to range between 0.81 and 0.88.

Aboelenen et al. [37] also performed similar investigations on the use of regression models for *trip generation* for residential areas in Qatar. The regression models with *R-square* values between 0.54 and 0.65 indicated that the number of employees, students, and family members having driving licenses significantly influenced trip generation. Lafta and Ismael [54] developed a trip generation model for 11 zones of the Al-Karada region in Baghdad city using an approach involving the application of artificial neural networks (ANN) and MLR. The approach provided predictions with accuracies ranging between 83.72% and 72.46%.

The present paper also deals with the implementation of the log-linear regression approach to developing a *trip distribution* model for predicting the inter-zonal trip interchanges based on partially correct *revealed-preference* (RP) data as a reliable origin-destination (OD) data was not available. To model *trip distribution,* it was required to extract a smaller set of the actual data that could reveal the underlying trend in trip-making as explained in Section 8.1. As part of this exercise, regression models relating the inter-zonal trip interchanges ($T_{ij}$) and the inter-zonal aerial distances ($C_{ij}$) between the centroids of the zones of the city were tested using *SPSS* software, which revealed that a *cubic function* could probably fit the trend. After the elimination of data points close to the *x* and *y* axes and the extreme points in various stages, a clear trend emerged for a set of 240 data points where the cubic function fitted the scatter plot with a high *adjusted R-square* value of 0.983. The information on these 240 points concerning the corresponding *trip productions* and *trip attractions* (obtained based on the *trip generation* models developed as in Table 4), in addition to the information on the inter-zonal aerial distances ($C_{ij}$), were then used to develop the log-linear model formulated based on the *gravity model* to predict the inter-zonal trip interchanges ($T_{ij}$ ). The calibration and validation exercises also indicated that the *average* or *mean absolute errors* (*MAE*) in the prediction of $T_{ij}$ were 15.18% and 16.06%, respectively.

The prediction error measured by comparing the percentage of trip frequencies for various distance ranges for the smaller dataset of 240 points was less than 0.5%, as shown in Table 8. Similarly, in the predictions made for the entire travel matrix of 4160 points (for the inter-zonal trips), the prediction error related to the percentages of trip-frequency distribution was less than 0.6%, as in Table 9. This is much less than the allowable error of 3% as recommended by BPR [49]. Also, the predicted *total trip production was* 1,082,986, which was very close to the actual value of 1,045,672 tallying within a prediction error of around 4%. The results thus indicated that the approach adopted in identifying the relationship between $T_{ij}$, and $C_{ij}$ from the scatter-plot and the subsequent modeling using the log-linear regression based on the gravity model was able to provide reliable predictions.

Similar studies were performed by Novačko et al. [55] on modeling *trip distribution* for Croatia where traffic counts on road links were obtained using the entropy maximization approach, along with noisy data on the OD matrix partially corrected using the fuzzy-logic approach, were used in obtaining the trip interchanges. The exponential form of the gravity model was used in modeling *trip distribution*, and the accuracy of prediction was verified based on the *GEH statistic* [56, 58]. Here, it can be seen that the trend followed by $f(Cij)$ Vs $Cij$ in the present study where $f(C_{ij})=1/(C_{ij}+1)^{1.285}$ as mentioned in Section 8.1 also follows the exponential form like the expression developed by Novačko et al. [55].

Zhou et al. [59] investigated the application of geographically and temporally weighted regression (GTWR) modeling for *trip distribution* where factors such as land use related to the built environment were considered for studies conducted in Hangzhou, China. The coordinates of 126 traffic analysis zones (TAZ), the density of land use activities, the travel time, and the distance were considered in the computations. The data was further supplemented by information on vehicular movements using license plate detectors. The estimated trip distributions using the GTWR model were found to have *R-square* values close to 0.823 when compared to actual observations, while the gravity model formulation was found to have *R-square* values close to 0.422. This indicated that the use of land use data as part of the GTWR model provided better predictions, resulting in errors of close to 17.7%. In the present study, the coordinates of the centroids of the TAZs were computed considering the density of land use as part of the *moment-area method* as described in Sections 7.1 and 7.2. The gravity model based on the log-linear regression model provided predictions where errors related to trip-frequency distributions for various distance ranges were within +/- 1%, while the errors related to the cell-wise prediction of trip interchanges in calibration and validation were 15.18% and 16.06%. This indicates that the present study provides reasonably accurate predictions using noisy RP data and secondary data, in addition to the use of land use density data in identifying the centroids of TAZs.

In the modeling exercise related to *trip distribution* described in this study, a GIS-based approach was adopted to prepare a high-resolution *Google-Earth* base-map of 1:5000. To locate the centroid of the activity centers/zones, it was necessary to obtain details on land use and the intensity of activity of sub-zones of most of the larger zones. The use of the *moment area* method in the determination of the coordinates of the zone-centroids assisted in the precise computation of *inter-zonal aerial distances* between the zone-centroids using the *Euclidean formula* for the 65 wards of Amritsar city. The *inter-zonal aerial distances* thus obtained were then used in developing the model for *trip distribution* based on the Gravity model formulation. Merrett and Khalife [60] observe that the use of the above-mentioned classical approach is indeed an effective strategy in handling data related to mixed land uses.

Thus, the combination of various strategies, including the use of GIS-based approaches to determine *inter-zonal aerial distances* and the use of the refined relationship between *trip interchanges* and the *inter-zonal aerial distances* in the development of a reliable log-linear regression model for *trip distribution*, contributed towards attaining higher accuracies in travel demand estimation.

The present study will provide the basic framework for transport planners to formulate better strategies for travel demand modeling where available data is noisy and less reliable. The approach to trip generation and trip distribution modeling described in this study is considered more cost-effective since it does not rely on the use of sophisticated GPS technology and the use of electronic sensors. The techniques adopted in the present study in the compilation of the database are less time-consuming when compared to the efforts required to process cell-phone-related call record data and video-graphic data.

One of the limitations of the present study is that intra-zonal trips were not considered in the modeling approach for *trip distribution* modeling. Also, the *inter-zonal aerial distances* ($C_{ij}$) used in modeling *trip distribution* were based on the shortest distances between the centroids of the TAZs, instead of the actual travel distances. It is also felt that the additional use of *stated preference* (*SP*) data could assist in modeling future travel demand as well.

## 10. Declarations

### 10.1. Author Contributions

Conceptualization, A.A. and V.G.; methodology, A.A. and V.G.; software, A.A. and V.G.; validation, A.A. and V.G.; formal analysis, A.A. and V.G.; investigation, A.A. and V.G.; resources, V.G.; data curation, A.A. and V.G.; writing—original draft preparation, A.A. and V.G.; writing—review and editing, A.A. and V.G.; visualization, A.A. and V.G.; supervision, V.G.; project administration, V.G. All authors have read and agreed to the published version of the manuscript.

### 10.2. Data Availability Statement

The data presented in this study are available in the article.

### 10.3. Funding

The authors received no financial support for this research work, authorship, and/or publication of this article.

### 10.4. Conflicts of Interest

The authors declare no conflict of interest.

## 11. References

[1] McNally, M. G. (2007). The Four-Step Model. Handbooks in Transport, 35–53, Emerald Group Publishing Limited, Bingley, United Kingdom. doi:10.1108/9780857245670-003.

[2] Hutchinson, B. G. (1974). Principles of urban transport systems planning. Scripta Book Company, Washington, United States.

[3] Papacostas, C. S. (1987). Fundamentals of transportation engineering. Prentice-Hall, Hoboken, United States.

[4] Janssens, D., Wets, G., Brijs, T., & Vanhoof, K. (2004, August). The simulation of activity diary data using sequential probability distributions. Proceedings of the ISCTSC Triennial International Conference on Transport Survey Methods, 1-6 August, 2004, Playa Herradura, Costa Rica.

[5] Janssens, D., Wets, G., Brijs, T., & Vanhoof, K. the Simulation of Activity Diary Data Using Sequential. ISCTC Triniel International Conference on Transport Survey Methods, 32(0), 1–21.

[6] Priyanto, S., & Friandi, E. P. (2010). Analyzing of public transport trip generation in developing countries; a case study in Yogyakarta, Indonesia. International Journal of Civil and Environmental Engineering, 4(6), 167-171.

[7] Bwambale, A., Choudhury, C. F., & Sanko, N. (2019). Car Trip Generation Models in the Developing World: Data Issues and Spatial Transferability. Transportation in Developing Economies, 5(2), 1–15. doi:10.1007/s40890-019-0075-7.

[8] Hu, S. (2010). Modelling trip generation/trip accessibility using logit models. PhD Thesis, Edinburgh Napier University, Edinburg, Scotland.

[9] Bourgeat, P. (2015). A Revealed/stated Preference Approach to Bus Service Configuration. Transportation Research Procedia, 6, 411–423. doi:10.1016/j.trpro.2015.03.031.

[10] Yue, X., Yang, X., Song, D., & Yuan, Y. (2022). The interaction effect of severe weather and non-weather factors on freeway travel volume. Science of the total environment, 808, 152057. doi:10.1016/j.scitotenv.2021.152057.

[11] Chmelík, J. (2015). Assessments of modal split in long-distance passenger transport. Review of Economic Perspectives, 15(1), 49–69. doi:10.1515/revecp-2015-0011.

[12] Hussein, R., Madar-Vani, G., & Maoh, H. (2021). A Comparison of Micro-and Zone-based Trip Generation Models with an Application to London, Ontario (No. TRBAM-21-01811). Transportation Research Board 100th Annual Meeting, January, 2021, Washington, United States.

[13] Mozolin, M., Thill, J. C., & Lynn Usery, E. (2000). Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation. Transportation Research Part B: Methodological, 34(1), 53–73. doi:10.1016/S0191-2615(99)00014-4.

[14] Jin, T. G., Saito, M., & Schultz, G. G. (2012). Development of a Statewide Commodity Flow Distribution Model Using Composite Friction Factors. Procedia - Social and Behavioral Sciences, 43, 406–417. doi:10.1016/j.sbspro.2012.04.114.

[15] Bruton, M. J. (2021). Introduction to Transportation Planning. Routledge, London, United Kingdom. doi:10.4324/9781003155690.

[16] Hill, D. M., & Dodd, N. (1966). Studies of trends of travel between 1954 and 1964 in a large metropolitan area. Highway Research Record, (141), Washington, United States.

[17] Domencich, T. A., Kraft, G., & Valette, J. P. (1968). Estimation of urban passenger travel behavior: An economic demand model. Charles River Associates, 64-78, Cambridge, United States.

[18] Wilson, A. (2013). Entropy in Urban and Regional Modelling. Routledge, London, United Kingdom. doi:10.4324/9780203142608.

[19] Thamizharasan, V., & Rengaraju, V. R. (1986). Methodology of Approach for Intercity Travel Demand Modelling. Indian Highways, 14(12), 23–33.

[20] Chari, S.R., Chandrasekhar, B.P., & Anjaneyalu, M.V.L.R. (1987). Experiences in the calibration of trip distribution model for HUDA area. Proceedings of 1987 Workshop on Transportation Research, Indian Institute of Management Ahmedabad, India.

[21] Ramamoorthy, N.V., (1986). Hyderabad Area Transportation Study (HATS). A Report on Zonal Trip Attraction Models for HATS Project. Hyderabad, India.

[22] Sarna, A., Prakash, A., Mukhopadhaya, D., Bhatia, N., & Suthar, H. H. (1992). Planning of Mass Rapid Transit System- a case study of Delhi. Journal of the Indian Roads Congress, 53(3), 553-588.

[23] Bjorketun, U. (1970). Traffic assignment in a GIS environment. WIT Transactions on the Built Environment, 26. doi:10.2495/UT960061.

[24] Varagouli, E. G., Simos, T. E., & Xeidakis, G. S. (2005). Fitting a multiple regression line to travel demand forecasting: The case of the prefecture of Xanthi, Northern Greece. Mathematical and Computer Modelling, 42(7–8), 817–836. doi:10.1016/j.mcm.2005.09.010.

[25] Kitamura, R. (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.

[26] Sofia, G. G., Ali, A. H. A., & Al-Zubaidy, H. A. (2012). Trip generation modeling for selected zone in Al-Diwaniyah city. Journal of Engineering and Sustainable Development, 16(4), 167-180.

[27] Mustafa, R., & Zhong, M. (2011). Applying GIS-Based Traditional Travel Demand Model for Improved Network-Wide Traffic Estimation: New Brunswick Case Study. ICTIS 2011. doi:10.1061/41177(415)103.

[28] Semeida, A. M. (2014). Derivation of travel demand forecasting models for low population areas: the case of Port Said Governorate, North East Egypt. Journal of Traffic and Transportation Engineering (English Edition), 1(3), 196–208. doi:10.1016/S2095-7564(15)30103-3.

[29] Chang, J. S., Jung, D., Kim, J., & Kang, T. (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.

[30] Urade, S., Khode, B., & Sujesh Ghodmare, P. (2017). Analysis of Trip Generation Using Land Use Pattern. International Journal of Scientific Engineering and Applied Science (IJSEAS), 3(4), 120–131.

[31] Mamei, M., Bicocchi, N., Lippi, M., Mariani, S., & Zambonelli, F. (2019). Evaluating origin–destination matrices obtained from CDR data. Sensors (Switzerland), 19(20), 4470. doi:10.3390/s19204470.

[32] Cordera, R., Sañudo, R., dell'Olio, L., & Ibeas, Á. (2018). Trip distribution model for regional railway services considering spatial effects between stations. Transport Policy, 67, 77–84. doi:10.1016/j.tranpol.2018.01.016.

[33] Calvo, F., Eboli, L., Forciniti, C., & Mazzulla, G. (2019). Factors influencing trip generation on metro system in Madrid (Spain). Transportation Research Part D: Transport and Environment, 67, 156–172. doi:10.1016/j.trd.2018.11.021.

[34] Doustmohammadi, M., Anderson, M., & Doustmohammadi, E. (2019). Regression Analysis to Create New Truck Trip Generation Equations for Medium Sized Communities. Current Urban Studies, 07(03), 480–491. doi:10.4236/cus.2019.73024.

[35] Soares Machado, C. A., & Quintanilha, J. A. (2019). Identification of trip generators using remote sensing and geographic information system. Transportation Research Interdisciplinary Perspectives, 3, 100069. doi:10.1016/j.trip.2019.100069.

[36] Yang, F., Li, L., Ding, F., Tan, H., & Ran, B. (2020). A data-driven approach to trip generation modeling for urban residents and non-local travelers. Sustainability (Switzerland), 12(18), 7688. doi:10.3390/su12187688.

[37] Aboelenen, K. E., Mohammad, A. N., Elgaar, M. I., & Choe, P. (2021). Trip Generation Rates Using Household Surveys in the State of Qatar. Journal of Traffic and Logistics Engineering, 9(1), 10–19. doi:10.18178/jtle.9.1.10-19.

[38] Gebre, G., & Quezon, E. T. (2021). Modeling Public Transport Users' Trip Production in Hawassa City, Ethiopia. Journal of Civil Engineering, Science and Technology, 12(2), 75–90. doi:10.33736/jcest.3972.2021.

[39] Mohd Shafie, S. A., Leong, L. V., & Mohd Sadullah, A. F. (2021). A trip generation model for a petrol station with a convenience store and a fast-food restaurant. Sustainability (Switzerland), 13(22), 12815. doi:10.3390/su132212815.

[40] Wikipedia (2022). Location of Punjab in India. Available online: https://en.wikipedia.org/wiki/Punjab,_India#/media/File:IN-PB.svg (Accessed on August 2022).

[41] Wikipedia (2022). Map. Available online: https://en.wikipedia.org/wiki/Amritsar#/map/0 (Accessed on August 2022).

[42] Consulting Engineering Services (India) Private Limited in Association with SYSTRA MVA Consulting (2012). Comprehensive Mobility Plan for Various Cities of Punjab State (Package-1): Amritsar City. Draft Final Report Document No. 2010112, March 2012, Punjab Municipal Infrastructure Development Company, Department of Local Government of Punjab, Punjab, India.

[43] Google.maps (2022). Amritsar. Available online: https://www.google.co.in/maps/place/Amritsar,+Punjab/@31.6336712,74.8000796,12z/data=!3m1!4b1!4m5!3m4!1s0x391964aa569e7355:0xeea2605bee84ef7d!8m2!3d31.6339793!4d74.8722642 (Accessed on August 2022).

[44] Municipal Corporation Amritsar (2022). Amritsar Municipal ward Boundary. Available online: http://www.amritsarcorp.com/WardWise.jpg (Accessed on August 2022).

[45] Voorhees, A. M. (2013). A general theory of traffic movement. Institute of Traffic Engineers, New Haven, Connecticut. doi:10.1007/s11116-013-9487-0.

[46] Census (2014). Census of India 2011 - Punjab - Series 04 - Part XII A - District Census Handbook, Amritsar. Ministry of home Affairs, New Delhi, India. Available online: https://censusindia.gov.in/nada/index.php/catalog/982 (Accessed on August 2022).

[47] Census (2014). Census of India 2011 - Punjab - Series 04 - Part XIIB - District Census Handbook, Amritsar. Ministry of home Affairs, New Delhi, India. Available online: https://censusindia.gov.in/nada/index.php/catalog/1017 (Accessed on May 2022).

[48] Kanthi, R. V. (2012). Multinomial Logit and Nested Logit Models for Integrated Studies on Vehicle ownership and Mode choice for Amritsar City using BIOGEME and NLOGIT based RP data. Master Thesis, Department of Civil Engineering, National Institute of Technology Karnataka, Karnataka, India.

[49] Bureau of Public Roads. (1963). Calibrating and Testing the Gravity Model for Any Sized Urban Area. U.S. Department of Commerce, Bureau of Public Roads, Office of Planning, Washington, United States.

[50] Bureau of Public Roads (1954). Conducting a Home Interview Origin-Destination Survey. Procedure Manual 2B, Public Administration Service, Chicago, United States.

[51] Kadiyali, L. R. (2013). Traffic Engineering and Transport Planning. Khanna publishers, New Delhi, India.

[52] Mir, I. S., Cheema, P. P. S., & Singh, S. P. (2021). Implementation analysis of solid waste management in Ludhiana city of Punjab. Environmental Challenges, 2, 100023. doi:10.1016/j.envc.2021.100023.

[53] ADA. (2022). Amritsar Development Authority. Available online: https://www.adaamritsar.gov.in/en (accessed on June 2022).

[54] Lafta, S. A., & Ismael, M. Q. (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.

[55] Novačko, L., Šimunović, L., & Krasić, D. (2014). Estimation of Origin-Destination Trip Matrices for Small Cities. Promet-Traffico, 26(5), 419–428. doi:10.7307/ptt.v26i5.1501.

[56] DMRB (1996). Traffic Appraisal of Road Schemes. Design Manual for Roads and Bridges, 12(2), UK Highways Agency, May, London, United Kingdom.

[57] Wisconsin Department of Transportation. (2022). Model Calibration, The GEH formula, Wisconsin DOT Micro-simulation Guidelines. Wisconsin, United States. Available online: http://www.wisdot.info/microsimulation/index.php?title=Model_Calibration#The_GEH_Formula (Accessed on August 2022).

[58] Smith, J., & Blewitt, R. (2010). Traffic modelling guidelines: TFL traffic manager and network performance best practice, version 3.0. Transport for London, London, United Kingdom.

[59] Zhou, Y., Shen, X., & Jin, S. (2019). A Geographically and Temporally Weighted Regression Model to Explore the Spatiotemporal Influence of Built Environment on Trip Distribution: A Case Study in Hangzhou. CICTP 2019. doi:10.1061/9780784482292.407.

[60] Merrett, T.H., & Khalife, B. (1993). Geographical Information Systems Development Using a General-Purpose Database Programming Language. GIS: Technology and Applications, Proc. of the Far East Workshop on Geographic Information Systems, 21-22 June 1993. 121-138.