

# DEVELOPMENT OF MODE CHOICE BEHAVIOR MODEL FOR INTER-REGIONAL PUBLIC TRANSPORT– A CASE STUDY OF INDIA

**Pinakkumar Ramanuj<sup>1,\*</sup>, Harishkumar Varia<sup>2</sup>, Ami Shah<sup>3</sup> and Arvindkumar M Jain<sup>4</sup>**

*Received: February 03, 2022; Revised: November 19, 2022; Accepted: December 13, 2022*

## Abstract

Mode choice behavior of the inter-regional public transport passengers is important for proficient planning and operation of transport systems in developing countries. It is the responsibility of the authority to satisfy the demand for long regional trips having a significant movement of passengers by providing efficient public transport services. Most of the studies for the mode choice behavior of the passengers have been done for the urban mass transport within the urban conglomerate. The study aims to improve knowledge of the factor affecting passengers' decisions on the mode of travel for inter-regional public transportation in India. The Multinomial Logit (MNL) model and Artificial Neural Network (ANN) models among Gujarat State Road Transport Corporation (GSRTC) buses, Railways, and privately operated buses were developed for the trips between Surat city and Bhavnagar region of Gujarat, India. The final ANN model reflects a difference of attributes has 90.52% mode prediction capability against 70.11% of the MNL model. The developed model gives the proper insight for improving the transportation facilities in the region. It is revealed that the improvement in the service level and egress parameters are more important to attract travelers. The sleeper seat availability, lesser travel cost, lesser egress distance, and availability of night journey have been found important during travelers' thinking process to select the traveling mode. There are 13.1% and 20.1% rises in the probability of choosing GSRTC bus and train mode, respectively, with the improvement in the service level. Moreover, the GSRTC bus should improve its connectivity route to attract travelers.

**Keywords:** Egress and ingress modes, interregional passenger trips, modal split analysis, mode choice model, multinomial logit model

---

<sup>1</sup> Research Scholar, Gujarat Technological University, Ahmedabad, India, E mail: pinakramanuj@gmail.com

<sup>2</sup> Civil & Infrastructure Engineering, Adani Institute of Infrastructure Engineering, Ahmedabad, India, E mail: hrvaria.7@gmail.com

<sup>3</sup> Smt. S. R. Patel Engineering College, Unjha, India, E mail: principal@srpec.org.in

<sup>4</sup> Jivraj Mehta Institute of Technology, Mogar, India, E mail: amjain\_asu@yahoo.com

\* Corresponding author

## Introduction

Travel demand amongst people is increased due to rapid urbanization, higher population growth, and improvement in an individual's living standards. Hence, the primary goal of transportation planners is to balance the supply of transportation with the rising demand for travel. A Mode choice analysis is the third step of the classical travel demand model used to predict a mode of transportation (Ortuzar and Willumsen, 2011; Ashalatha *et al.*, 2013). In order to handle challenges like forecasting demand for new transportation modes, reducing traffic congestion, mode shift owing to improvements in current transportation facilities, allocating funds, etc., it is essential to analyze mode choice decisions properly. The modal split analysis plays a vital role because it directly impacts certain policy decisions (Ortuzar and Willumsen, 2011; Minal and Sekhar, 2014). Therefore, it is necessary to study the relationship between the traveler's mode choice behavior and the various explanatory variables. Most of the studies regarding mode choice modeling have been carried out in urban area for planning and improvement of urban transportation systems. In comparison, insignificant research is available for developing the mode choice model for the inter-regional transportation system (Miller, 2004). The available travel modes for low to middle-income groups are either bus or railway systems in developing countries. Generally, such public transportation facilities are provided by either the government or privately operated buses for travel distances between 300 km to 600 km. Personal car or air travel service mode is not generally affordable for low to middle-income group travelers. Some of the urban/regional centers have typical attractions due to higher employment or business activities. This leads to considerable to and fro person trips daily, weekly, monthly or seasonal basis between developed centers to the undeveloped/semi-developed regions. Such type of travel demand of the lower and middle-income group travelers must be taken care of by the authorities.

Due to the distinction in culture, supply available, and economic condition, the model formulated for the developed country cannot be applied directly to developing countries (Haire and Machemehl, 2010). Because of a developing country, a large part of India's population depends on the extensive rail and road network (Jana and Varghese, 2017). The proportion of travelers' dependence on bus travel is high though the road network is used by both public and private modes (Bardhan *et al.*, 2015). Therefore, mode choice analysis for the inter-regional passenger trips

between highly acceptable and most competitive public transportation modes, i.e., bus and railway trips, is important to analyze for Indian conditions. The research aims to enhance the understanding of the factors affecting mode choice decisions for inter-regional public transportation passenger trips in Indian conditions. The study's main objectives are to understand the influence of the various attributes in the mode choice model developed for inter-regional trips and suggest alternative strategies to improve the existing public transport facilities in the study area. The Bhavnagar region consists of a group of the Bhavnagar, Amreli, and Botad districts of Gujarat state, India is selected for the study based on higher trip rate and common joining route with the Surat as well as their geographic situation. Surat city (21.1702° N, 72.8311° E) is situated in the southern part of Gujarat state of India, with a 55.3% decennial population growth rate between 2001 to 2011 (SMC, 2020). It is one of the uppermost developing cities in India. People from the other regions of Gujarat, including the Saurashtra region, migrate to the Surat city area for business opportunities and better livelihoods. Travelers mainly prefer private bus mode, having very high and demand-oriented fares though government-operated GSRTC bus and railway services are available in the study area. The study will be helpful in improving existing travel services for increasing the ridership in the study area by analyzing trip makers' mode choice decisions. The mode choice model is developed for native home-bound trips of the migrated population made by public transport services. For the study, trip data was collected by questionnaire for the trips originating from the Surat city region to the Bhavnagar region.

This study addresses a need for more research on traveler's mode choice behavior between large metropolitan centers and non-metropolitan areas in developing countries. The urban and regional transport systems can be considered as a macro systems because they are defined by the interactions among micro transportation networks (Tamin, 2000). Due to cultural differences, supply constraints, and economic conditions, the model designed for developed countries cannot be immediately applied to developing countries (Almasri and Alraee, 2013). From the theoretical perspective, parameters considered in the research will be valuable to inter-regional mode choice modeling in similar circumstances of a developing country. The current study is beneficial for understanding travelers' behavior while choosing between existing passenger

travel modes such as private buses operated by travel agencies, GSRTC buses, and railways. Furthermore, an arc elasticity obtained for some policy parameters and sensitivity analysis is valuable for recommending ways to improve the transportation service in the region to meet travelers' needs.

The study uses the novel idea of adding differences of attributes as an additional input in the ANN model. The significance of the difference of attributes is in line with the actual thought process of the mode choice decision makers. Moreover, the study focuses on the traveler's thought process to investigate the impact of the parameters that travelers perceive as governing parameters for their mode choice decisions.

The article is organized as follows. The next section provides a background of the study, followed by a study of methodology. Details of the study area with data collection and data analysis are presented in sequence prior to the model development section. The subsequent section is a conclusion based on the model interpretation and analysis. The last section shows the references. Several important works of literature related to the proposed study are presented onwards.

## Literature Review

### A Brief Review of the Modeling Approach

Modeling of mode choice can be done by means of a discrete choice model (Ben-Akiva and Lerman, 1985). Broadly, the modeling of mode choice has two fundamental approaches, namely the aggregate discrete choice model and the disaggregate discrete choice model. Aggregate travel demand models are estimated, with data on travel behavior at some level of aggregation within a geographic zone that serves as an observational unit for the statistical analysis. Most transportation studies collect micro-level data to estimate the models directly from this information without first aggregating it to a zonal level (Ben-Akiva and Lerman, 1985). Such models have come to be called disaggregate models. The former was popularly used up to the late 1970s; in comparison, the latter became more popular during the 1980s because the elasticity of demand can be more precisely estimated for the modal attributes by the disaggregate discrete choice model (Ortuzar and Willumsen, 2011). Moreover, disaggregated models can provide a rich prediction of behavioral attributes (Koppelman and Bhat, 2006). The disaggregate models have a higher degree of transferability of the model in context with location and are less likely to suffer from biases (Koppelman and Bhat, 2006).

### Review of Utility-Based Discrete Choice Models

According to Ben-Akiva and Morikawa (1990), utility-based models are considered as one of the most accurate tools for making predictions about human decisions and behavior. Based on the error distribution, there are three popular families of models, namely Logit models, Probit Models, and General Extreme Value (GEV) models. The logit model is the most popular and widely used model for calibrating mode choice because it has a simple model formulation (Domencich and McFadden, 1975). Ghareib (1996) has suggested that the logit model shows more accurate predictions when analyzing the situation of a binary mode choice. Sayed and Razavi (2000) have developed a binary logit model for U. S. freight transport market using the information on individual shippers and individual shipments. The Multinomial logit model (MNL) has been used by many researchers within the urban context. Ashalatha *et al.* (2013) have modeled the mode choice preference of commuters of Thiruvananthapuram city of India by using the MNL model for predicting bus, car, and two-wheeler modes. Subbarao and Krishnarao (2013) have developed a multinomial logit regression model for the Mumbai Metropolitan Region of India. Similarly, there were several studies on mode choice models like binary logit (Manssour *et al.*, 2013), multinomial logit (Miskeen and Rahmat, 2011; Philip *et al.*, 2013; Islam *et al.*, 2016), nested logit (Yao and Morikawa, 2005) and mixed logit models (Walker, 2002; Hensher and Greene, 2003; Barus *et al.*, 2016) carried out within intercity/ inter-regional context.

### Review of Factors Affecting Mode Choice

The significant characteristics in determining the mode choice of the travelers are classified into three groups: characteristics of trip makers, characteristics of the journey, and transport facility available (Ortuzar and Willumsen, 2011). Mode choice of travelers is influenced by factors like travel time, travel cost, waiting time, number and ease of transfers, comfort, income, car ownership, egress and ingress time, etc., as analyzed by many researchers. Rajalakshmi (2013) studied mode choice for a suburban area in the Kottayam district of Kerala, India. In the study, work trips were modeled using Artificial Neural Network (ANN) and observed the importance of travel time and travel cost in mode choice decisions. In the study made by Ashalatha *et al.* (2013), it was observed that age group, gender, monthly income, vehicle ownership, distance, time per distance, and cost per distance were the main parameters responsible for a modal shift. Almasri and Alraee (2013) have modeled work trips in developing cities and have

observed that total travel time, total travel cost divided by personal income, vehicle ownership, and age are the most influencing factors. However, the authors have pointed out that the model formulated for developed countries can't be directly suitable for developing countries due to different conditions and circumstances. The influencing attributes of the model developed in an urban context might differ from an intercity context (Haire and Machemehl, 2010).

Al-Ahmadi (2006) modeled the mode choice model for Saudi Arabia. The factors affecting the mode choice were in-vehicle travel time, out-of-pocket cost, group trips, income, travel distance, nationality, and car ownership. Manssour *et al.* (2013) have depicted the importance of Out-of-vehicle travel time, total travel cost, gender, age, nationality, income level, car ownership, convenience, comfort, and weather conditions for the mode choice decisions. Bhat (1995) as well as Yao and Morikawa (2005), have found that service frequency also affects the mode choice decision of individuals for intercity travel. Further, Zhao *et al.* (2018) have found that better feeder accessibility is important in enhancing ridership. Zhou *et al.* (2017) analyzed the mode choice of rural residents in China. The result indicates that income, vehicle ownership, convenience, and reliability of service are noticeable affecting factors. Jana and Varghese have analyzed mode choices for inter-regional travel in India for long-distance travel (Jana and Varghese, 2017). The Binary logit model was developed by dividing trips with reference to journey duration and rural/urban trips. Authors have observed the importance of trip purpose, time of year, type of stay, and rural/urban trip in travel decisions about the mode choice for long travel. Philip *et al.* (2013) have developed the MNL model using an activity-based approach for the rural area in Kerala, India. The main influencing factors observed were vehicle ownership, license ownership, trip walking time, waiting time, trip cost, and duration.

Likelihood Ratio Test Statistics (LRTS) has been used by various researchers to check the MNL model's validity (Koppelman and Bhat, 2006; Ashalatha *et al.*, 2013). However, the non-nested hypothesis test proposed by Horowitz (1982) is useful for the case where both models have the same number of parameters (Koppelman and Bhat, 2006). For better results, researchers have looked forward to soft computing techniques like an artificial neural network, fuzzy logic, and hybrid techniques. An ANN can be effectively used to predict the mode choice of travelers (Cantarella and de Luca, 2005; Sikdar and Sekhar, 2005; Rajalakshmi, 2013). The result obtained from the neural network model has higher prediction accuracy compared with the

conventional logit model (Sayed and Razavi, 2000; Zhao *et al.*, 2020).

### Literature Gap

A large part of the population in India is extensively relying on the rail network and road network by public bus transportation mode for long-distance travel (Jana and Varghese, 2017). Therefore, it is helpful to study the mode choice between popular public transportation in Indian conditions. Industrialization resulted in a considerable population shift from rural to urban in India. There are regular social as well as concentrated festival trips between urban and rural areas. Inter-regional modal analysis is much more useful for studying such trips. Zhu *et al.* (2010) studied the mode choice behavior of a large population shift during the spring festival in China. ANN has been used to model traveler's behavior and observed that the resident's travel choice is complex during concentrated demand. However, trip makers' behavior regarding mode choice may differ from region to region; based on their social, cultural, and economic conditions (Haire and Machemehl, 2010).

However, intercity passenger travel mode choice modeling has not developed to the same extent as urban mode choice modeling (Landinois, 1987). The authors have emphasized the usefulness of the work for the government, public transport agency, and private carriers to make a marginal decision.

Looking at the above reviewed literature, it can be deduced that more research on traveler's mode choice behavior among public transportation between large urban centers and non-urban areas is needed, particularly in developing countries like India.

## Materials and Methods

### Study Methodology

The proposed study intends to analyze trip makers' mode choice behavior within the study space for passenger trips.

Defining the study problem is the first task in the methodology which is followed by the literature review concerned with the study problem. The next phase is the definition of the study area and preliminary data collection. A pilot survey is useful to determine the most relevant attributes for the travelers in the study area is done. The subsequent section includes data collection with a well-designed questionnaire and data analysis. In the next section, MNL and ANN models have been developed. An elasticity analysis and scenario development are conducted helpful for policy

decisions. The ANN models were compared with the MNL model. The last section discusses a few important findings obtained from the data analysis, model result, and sensitivity testing for the important variables by elasticity or scenarios. The chart in Figure 1 presents the methodology of the proposed study.

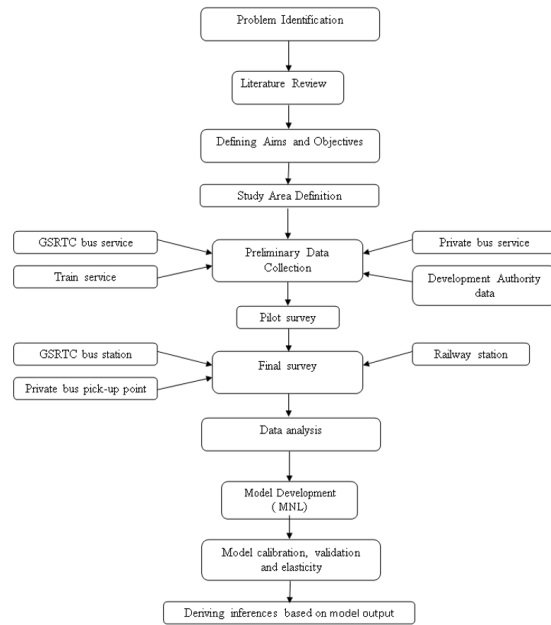


Figure 1. Proposed methodology

### MNL Model

The discrete choice model is a mathematical function based on utility maximization theory. The decision-maker selects the mode which can offer the greatest utility, divided into two components: (i) measurable systematic component ( $V_{in}$ ), and (ii) random part ( $\varepsilon_{in}$ ). The utility associated with individual  $n$  to an alternative  $i$ , denoted by  $U_{in}$  can be obtained by Equation (1).

$$U_{in} = V_{in} + \varepsilon_{in} \quad (1)$$

The logit model is easier than the probit model from an analytical point of view (Minal and Sekhar, 2014). The most underlying assumption of the logit model is that the random component is distributed independently, identically extreme values. The expression for selecting alternative  $j$  among  $m$  alternatives is as shown in Equation (2).

$$P_{(j)} = \frac{\exp(V_j)}{\sum_{m=1}^m \exp(V_m)} \quad (2)$$

where,

$P_{(j)}$  = probability of choosing alternative  $j$  by the decision-maker

$V_j$  = systemic component of the alternative  $j$

Furthermore, The MNL model structure selected for the proposed study has been widely used for intercity mode choice models mainly due to its simple mathematical form, ease of estimation and interpretation, and the ability to add or remove choice alternatives. The software NLOGIT was used to build the MNL model. The best model was chosen based on the coefficient of the variable with other statistics parameters.

**Model performance:** Mcfadden's Pseudo  $R^2$ ;  $\rho_c^2$  and  $\bar{\rho}_c^2$ , are considered as goodness to fit measures for the model. The McFadden pseudo  $R^2$ , lies between 0 and 1 measures how well-observed outcomes are replicated by the model. Its greater value suggests that the model is more accurate. The  $\rho_c^2$  and  $\bar{\rho}_c^2$  represent the rho-squared with respect to the constant-only model and adjusted rho-squared for the constant-only model, respectively. During the model calibration, the significance of variables is judged based on the logical sign, p-value, and likelihood ratio test. Likelihood Ratio Test Statistics are used to check the model's validity. The NULL hypothesis rejection is decided based on the LRTS value compared with critical, at a 90/95% confidence level with a particular degree of freedom (Koppelman and Bhat, 2006; Almasri and Alraee 2013; Ashalatha *et al.*, 2013; Miskeen *et al.*, 2013). However, the non-nested hypothesis test proposed by Horowitz (1982) is used for the case where both models have the same number of parameters (Koppelman and Bhat, 2006).

### ANN Model

The mode selection decisions of travelers are behavioral and complex in nature. Therefore, Artificial Neural Network (ANN) is also helpful in predicting the mode choice as ANN has good potential to understand the complex relationship (Sayed and Razavi, 2000). The feed-forward back propagation technique was used for the model developed in MATLAB software. The relative importance of input variables is found by proportioning of weights (Sikdar and Sekhar, 2005).

### Study Area

#### Case Study Description

The study area selected for the present study is Surat city and Bhavnagar Region. Bhavnagar, Amreli, and Botad districts of Gujarat state, India, are grouped as a Bhavnagar region, which is shown in Figure 2. Surat, a textile hub, a diamond processing and trading center in Gujarat, is one of the fastest-growing cities in India. Bhavnagar region

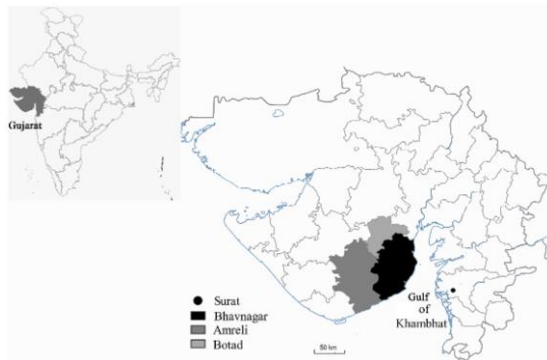


Figure 2. Location of the study area

has lesser industrial growth compared to Surat, and it is a semi-arid area. People from the Bhavnagar region migrate to Surat for employment, business, and improvement in the standard of living. There is a higher trip rate between the Surat city area and the native region of the migrated people of the Bhavnagar region.

The distance by roadway between Surat city and Bhavnagar region is comparatively longer due to the bay of Khambhat, as shown in Figure 3. Also, the distance by railway between the origin and the destination is much longer than the distance by roadway. The travel distance between Surat to destination district headquarter are presented in Table 1.

#### Data Collection

The present study is concentrated on the native home-bound trips made by travelers using public transportation systems available between Surat city and the Bhavnagar region. The data collected for the trips originated from the Surat city region to the Bhavnagar region. The main survey (followed by the pilot survey) was carried out at Surat with a carefully designed questionnaire to suit the objective of the study. Passenger trips made by privately operated buses, Gujarat State Road Transport Corporation (GSRTC) buses and Indian railway mode were considered due to their major share.

Data pertaining to trips made during one year were collected from randomly selected travelers who had migrated from the Bhavnagar region but living in Surat city by intercept survey done at Surat. The data were collected by face-to-face interview technique at a railway station, GSRTC bus station/pick-up point, and private bus service office/pick-up point in Surat city during June of 2017 by conducting a Revealed Preference (RP) survey. The questionnaire includes the socio-economic parameters of travelers, trip data, and data

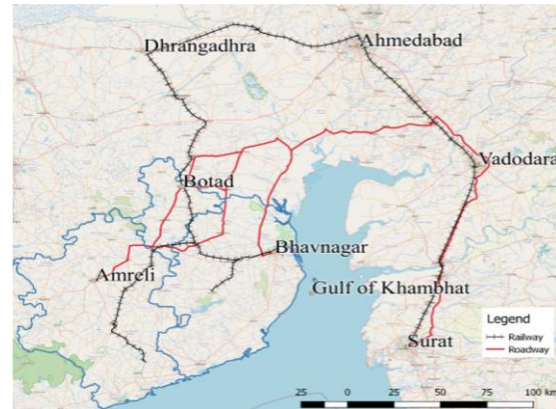


Figure 3. Connecting routes in the study area

Table 1. Mode-wise travel distance and travel time

| Origin-Destination | Travel distance (KM)  |         |
|--------------------|-----------------------|---------|
|                    | Private bus/GSRTC bus | Railway |
| Surat - Bhavnagar  | 335                   | 526     |
| Surat - Amreli     | 427                   | 537     |
| Surat - Botad      | 338                   | 434     |

representing existing transportation facilities. Based on the supply data collected in the preliminary survey, 2,216 cleared trip data were collected. Respondents were also asked for the governing criteria behind particular mode selection. Such selected dummy variables were availability of reserved seats, convenient time, lesser travel time, lesser travel cost, better service, ingress distance, egress distance, sleeper seat availability, and a night journey. Inter-regional travel is not a daily practice due to higher travel time in the study area, and it has not been performed by every individual in the city. Choice-based sampling, considered as an endogenous sampling process, was used for data collection. Respondents were asked to rank the level of service/service level (LS) from 1 to 5 offered by the mode based on regularity, fare structure, travel time, vehicle condition, etc., perceived by travelers. Travelers were also asked to tick the parameters responsible for their mode selection. Mode-wise distribution of the data is presented in Figure 4, which reflects a higher share of bus mode agrees with the Indian scenario (Jana and Varghese, 2017).

#### Sample Size

Sufficient sample size mainly depends on population size, desired precision, and degree of variability. Assuming normally distributed population, the empirical formula given by Levy and Lemeshow (2008) and also by Hogg *et al.* (2015) is presented in Equation (3) is used for the study (Ashalatha *et al.*, 2013).

$$n_0 = \frac{Z^2 pq}{E^2} \quad (3)$$

where,

$n_0$  = Sample size

$Z = 1.96$  for a 95% confidence level

$q = 1-p$

$E$  = Desired error margin

$p$  = Hypothesized true proportion for population (0.5 for the worst case).

Based on Equation (3), the minimum sample sizes were obtained as 385 and 1,083 for the desired error margin of 0.05 and 0.03, respectively. For the present study, 2,216 cleared trip sample data were finally obtained from the interview of 1,256 trip makers. It was ensured that data contains trips toward all the districts of the study area by GSRTC bus (gbus), Private bus (pbus), and Railway mode (train). For the proposed study, mode-wise sample data were collected in view of matching their existing seating capacity available.

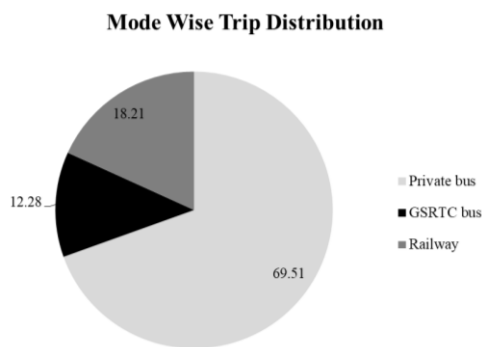


Figure 4. Mode wise trip distribution

### Data Analysis

After removing incomplete data, cleared data were analyzed so as to understand the travel attributes of the travelers within the study area. The available modes for passenger transportation in the study area include the GSRTC bus, privately operated bus, and train. The mode-wise distribution of collected data is presented percentagewise in Figure 4. The collected data from the respondents were fed into Microsoft Excel for sorting, coding, and appropriate statistical analysis. Socio-economic characteristics of travelers, like income, gender, age, vehicle ownership, etc., were explored. The primary data analyses of the dataset are summarized in Table 2. More than 70% of respondents have unavailability of a car indicates the level of dependence on public transport in the study area. Roughly 70 percent of the households are middle-income people who fall in the monthly income range of Rs. 20000 to Rs. 50000. A traveler's income

determines the amount of money he is willing to spend on a trip.

Travel cost for the main mode has been influenced by factors such as sleeping or seating seat, Air Conditioned (AC) or Non AC bus, types of GSRTC bus, route, class, type of train, trip season, etc.

In the case of railway mode, only second-class seating and sleeping class were considered. Travelers using First Air Condition (AC), Second AC, and Third AC on the railway are assumed a choice transit riders due to their comparatively higher travel costs. The ranges of travel costs from Surat to destination district headquarter are presented in Table 3, along with travel time. There is demand-oriented travel cost fluctuation during peak festival season and off-peak season. From Figure 4 and Table 3, it is clear that the private bus has the maximum share, though it has a comparatively higher fare structure.

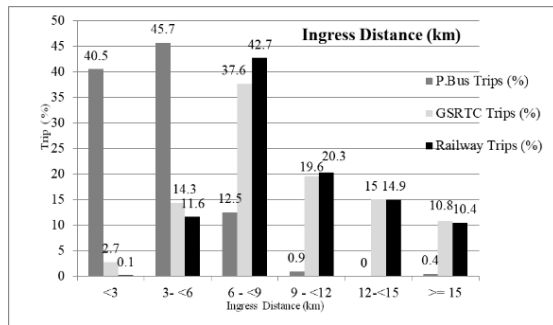
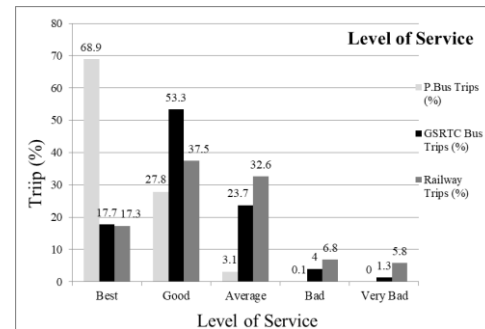
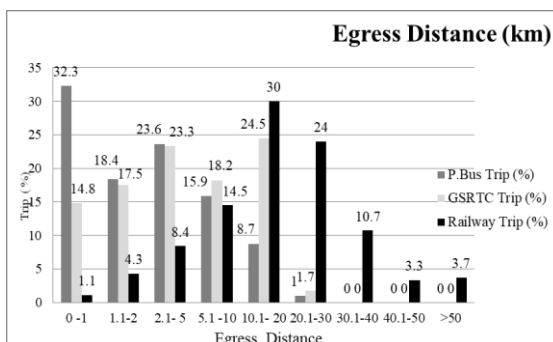
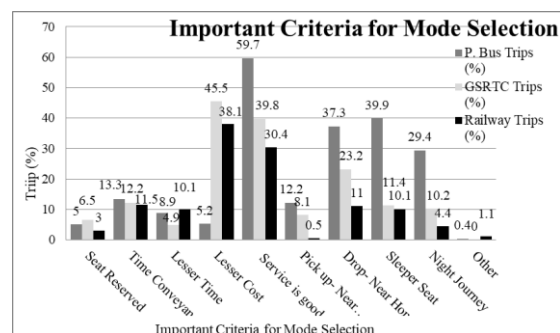
Zhao *et al.* (2018) have suggested that feeder accessibility is indeed crucial in enhancing the mode share of trunk service, sometimes even more important than the attributes of trunk service. Travelers are more concerned with out-of-vehicle time and out-of-vehicle cost attributes compared to in-vehicle travel attributes. Out-of-vehicle distances were analyzed to study the effect of ingress and egress attributes in the present study. The distribution of ingress and egress distances in

Table 2. Data frequency table

| Variable                     | Category          | Count | Percentage |
|------------------------------|-------------------|-------|------------|
| Gender                       | Male              | 1,509 | 75.3       |
|                              | Female            | 495   | 24.7       |
| Age                          | <18               | 26    | 1.3        |
|                              | 18-40             | 1,375 | 68.6       |
|                              | 41-60             | 532   | 26.6       |
|                              | >60               | 71    | 3.5        |
|                              | 0                 | 10    | 0.5        |
| No. of Two Wheeler           | 1                 | 406   | 20.3       |
|                              | 2                 | 1,373 | 68.5       |
|                              | 3                 | 194   | 9.7        |
|                              | $\geq 4$          | 21    | 1          |
|                              | 0                 | 1,471 | 73.4       |
| Car Ownership (No. of a car) | 1                 | 516   | 25.75      |
|                              | $\geq 2$          | 17    | 0.85       |
|                              | 0                 | 6     | 0.3        |
| Income in Rs/month           | <10,000           |       |            |
|                              | 10,001 to 20,000  | 124   | 6.2        |
|                              | 20,001 to 30,000  | 452   | 22.6       |
|                              | 30,001 to 50,000  | 943   | 47.1       |
|                              | 50,001 to 100,000 | 457   | 22.8       |
|                              | >100,000          | 22    | 1.1        |
|                              | 1                 | 1,257 | 62.7       |
| Earning Members in HH        | 2                 | 626   | 31.2       |
|                              | 3                 | 111   | 5.5        |
|                              | 4                 | 10    | 0.5        |
|                              |                   |       |            |

**Table 3. Mode-wise travel cost and travel time**

| Origin-Destination | Travel Cost range (Rs) |                                         | Travel time (Hours-h Minutes-m) |                       |          |
|--------------------|------------------------|-----------------------------------------|---------------------------------|-----------------------|----------|
|                    | GSRTC Bus              | P Bus: Off-peak Season<br>(Peak Season) | Railway                         | Private bus/GSRTC bus | Railway  |
| Surat - Bhavnagar  | 190-304                | 350-450 (600 – 1000)                    | 185-335                         | 7 h 45 m to 8 h 45m   | 9 h 33 m |
| Surat - Amreli     | 227-337                | 400-500(600– 800)                       | 150-305                         | 8 h 30 m to 11 h 0m   | 9 h 30 m |
| Surat- Botad       | 170-307                | 400-500 (700 – 900)                     | 140-290                         | 7 h 30 m to 9 h 0 m   | 7 h 40 m |

**Figure 5. Distribution of ingress distance (km)****Figure 7. Level of service perceived by travelers****Figure 6. Distribution of egress distance (km)****Figure 8. Important criteria for mode selection**

Figures 5 and 6 represent study area characteristics on either end. Lesser ingress distance for pbus reveals due to a number of pick-up facilities available in the Surat city area. There is a very high egress distance for railway mode because of its poor track distribution in the Bhavnagar region. Whereas pbus mode has a lesser egress distance due to its well-distributed network throughout the Bhavnagar region. The higher value of feeder attributes decreases the utility of the mode.

Latent variables indicating comfort and convenience for the travel mode perceived by the travelers were defined through five points Likert scale. The level of service perceived by travelers is represented in Figure 7. The higher values of the perceived level of service leads to improve the utility of the mode for travelers.

The study focuses on the traveler's thought process to investigate the impact of the parameters that travelers regard as governing parameters for the mode selection. Mode-wise classification of governing variables perceived by the respondent for the mode selection is presented in Figure 8. Better service level is an important criterion for the private bus mode selection. Because of higher travel costs in pbus mode, lower travel cost is an important criterion for the gbus and train travelers. The travelers for whom less ingress and egress distance were more important; they preferred pbus due to its well-distributed network in the study area. A variable representing the availability of night journeys was important for pbus users as mostly all private buses run during the night in the study area. Based on the number of buses and trains that run

**Table 4. Variable code and definition**

| Sr. | Parameter                  | Code | Definition                                                                                                                                 |
|-----|----------------------------|------|--------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | Age                        | AG   | 1, if $\leq 20$ ; 2, if 21 - 39; 3, if 40 - 59; 4, if $\geq 60$                                                                            |
| 2   | Gender                     | GND  | 1, if male; 2, if female                                                                                                                   |
| 3   | Occupation                 | OCC  | 1= Self-employed; 2 = Diamond industry; 3 = Private Service; 4 = Housewife; 5 = Other; 6 = Retired; 7 = Student                            |
| 4   | Income Level               | IN   | 1: < 10000; 2: 10001-20000; 3: 20001 - 30000; 4: 30001 – 50000; 5: 50001 - 100000; 6: > 100000. Unit Indian Rupees.                        |
| 5   | Average Income             | INC  | Average income of income level                                                                                                             |
| 6   | No. of TW                  | TW   | Number of two wheelers in household. Discrete value                                                                                        |
| 7   | Member in Household        | HH   | Number of members in household. Discrete value                                                                                             |
| 8   | Earning Member             | EM   | Number of earning members in household. Discrete value                                                                                     |
| 9   | Car Ownership              | CW   | Number of cars in household. Discrete value                                                                                                |
| 10  | Out-of-Vehicle Distance    | DST  | 0: <10; 1: 10 -19.9; 2: 20- 29.9; 3: 30- 39.9; 4: 40- 49.9; 5: 50-59.9; 6: 60- 69.9; 7: 70- 79.9; 8: 80- 89.9; 9: 90- 99.9; 10: $\geq 100$ |
| 11  | Egress Distance            | DD   | Continuous. Unit –kilometer                                                                                                                |
| 12  | Total Travel Time          | T    | Continuous. Unit – Hours                                                                                                                   |
| 13  | Total Travel Cost          | C    | Continuous. Unit Indian Rupees.                                                                                                            |
| 14  | Main Mode Travel Time      | IVT  | Continuous. Unit – Hours                                                                                                                   |
| 15  | Total egress time          | WTD  | Continuous. Unit – Minutes.                                                                                                                |
| 16  | Main mode Cost             | IVC  | Continuous. Unit Indian Rupees.                                                                                                            |
| 17  | Egress Cost                | OVCD | Continuous. Unit Indian Rupees.                                                                                                            |
| 18  | Waiting time at the egress | WD   | Continuous. Unit – Minutes.                                                                                                                |
| 19  | Perceived Level of Service | LS   | Discrete: 1: Very Bad; 2: Bad; 3: Average; 4: Good; 5: Best                                                                                |
| 20  | Lesser egress Distance     | RD   | Governing criteria for mode selection - Lesser egress Distance (1 if yes, 0 otherwise)                                                     |
| 21  | Lesser Travel Cost         | RC   | Governing criteria for mode selection - Lesser travel cost (1 if yes, 0 otherwise)                                                         |
| 22  | Service Availability       | S    | Index based on frequency of mode availability between origin and destination taluka.                                                       |
| 23  | Night Journey              | RN   | Governing criteria for mode selection - Night journey (1 if yes, 0 otherwise)                                                              |
| 24  | Better Service/ Facility   | RS   | Governing criteria for mode selection - Better service availability in journey (1 if yes, 0 otherwise)                                     |
| 25  | Sleeper Seat Availability  | RSL  | Governing criteria for Mode Selection - Sleeper seat availability in journey (1 if yes, 0 otherwise)                                       |
| 26  | Urban Destination          | UR   | Destination is in urban (1 if yes, 0 otherwise)                                                                                            |
| 27  | Peak Month                 | PM   | Trip in the peak period (1 if yes, 0 otherwise)                                                                                            |
| 28  | Mode Selected              | MOD  | 1: Private bus (pb/pbus); 2: GSRTC bus (gb/gbus); 3: Railway (tr)                                                                          |

between Surat and destination taluka, an accessibility index has been derived.

## Results and Discussion

### Model Development

In this study, two competing models, viz. an MNL model, and the ANN models are developed and analyzed.

### MNL Model

Based on the analyses of prior work concerned with mode choice modeling, choice variables for the model were selected. The variables' code and definition are tabulated in Table 4.

There are three popular inter-regional public transportation modes considered, i.e., privately operated bus (pb), GSRTC bus (gb), and railway/train (tr) with reference to descriptive data analysis for modeling.

A simple multinomial logit model was estimated with socio-economic variables. The analysis includes calibration and estimation of the utility functions for the different models and then choosing the best model. The model with the logical sign, maximum log-likelihood value, and maximum goodness of fit index ( $\rho^2$ ) value was considered the most suitable model. The first trial model includes total travel time (T) and total travel cost (C) as generic variables. The private bus mode was considered as a base mode when adding constants for the mode utilities. The utilities for different modes are presented in Equations (4a) to (4c) for the first trial model.

$$U_{pb} = \beta_1(T) + \beta_2(C) \quad (4a)$$

$$U_{gb} = \beta_G + \beta_1(T) + \beta_2(C) \quad (4b)$$

$$U_{tr} = \beta_T + \beta_1(T) + \beta_2(C) \quad (4c)$$

The significance of variables was checked, and the non-significant variables were eliminated based on the logical sign, p-value, and likelihood ratio test. Expected influencing variables were added in their various form, and their statistical parameters were checked through software. Mcfadden's Pseudo  $R^2$ ;  $\rho_c^2$  and  $\bar{\rho}_c^2$ , are considered as goodness to fit measures for the model. The estimation result of the final model among all trial models developed during refinement is presented in Table 5. The correlation matrix has been used to check the correlation between independent variables used in the model. The influence of various variables in the model with its statistics is described in this section.

**Income:** The ability of a traveler to spend the amount for travel costs is influenced by his income. Income group, income per household member (HH), log (income), and log (income) per HH were added to the model during the model development procedure. Though these variables were statistically significant, they were eliminated due to unjustified logical signs in the model.

**Car Ownership:** Car ownership correlates with the traveler's income level. Car ownership (CW) is defined as the number of cars occupied in a household, CW/HH and CW per earning members in household (EM) were added to the model in sequence. The sign of the parameters was as expected, and the variables were statistically significant. Later due to interaction with other variables during the model calibration process, it became insignificant even at a 90% confidence level, therefore, eliminated from the model.

**Two-Wheeler Ownership:** Considering the importance of vehicle ownership, the number of two-wheelers in the household (TW) variable was added as an alternative specific. The variable TW was replaced with TW/HH and TW/EM for model refinement. Considering proper negative sign and significance levels, TW and TW/HH variables were replaced with TW/EM. These have a correct intuitive negative sign with more than a 90% confidence level.

**Perceived Level of Service:** Attitudes and perceptions variables are also significant in the traveler's mode choice decision. The overall service level (LS) of the mode perceived by the travelers was included in the model. The importance of the level of service for public transportation has also been realized by Arasan and Vedagiri (2011). LS has a correct positive sign as per intuition with more than a 99% confidence level.

**Travel Time:** Total travel time (T) and T/distance were alternatively added to the model, which contains positive signs that could not be justified. Travelers are more sensitive to out-vehicle

**Table 5. Model estimation**

| Variables                          | Estimated Coefficient | Significance | Exp ( $\beta$ ) |
|------------------------------------|-----------------------|--------------|-----------------|
| Constant (tr)                      | 0.3459                | 0.1913       | 1.41            |
| Constant (gb)                      | -1.1633               | 0.000        | 0.31            |
| Waiting time at egress (WD)        | -0.0373               | 0.000        | 0.96            |
| Egress distance (DD)               | -0.0347               | 0.000        | 0.97            |
| Perceived level of service (LS)    | 0.7966                | 0.000        | 2.22            |
| Egress cost/ income (OVCD/Inc)     | -0.0340               | 0.0037       | 0.97            |
| TW/EM (tr)                         | -0.2387               | 0.0788       | 0.79            |
| TW/EM (gb)                         | -0.3079               | 0.0452       | 0.74            |
| Lesser egress distance RD (tr)     | -0.7516               | 0.0002       | 0.47            |
| Lesser travel cost RC (tr)         | 2.639                 | 0.000        | 14              |
| Lesser travel cost RC (gb)         | 2.8247                | 0.000        | 16.86           |
| Night journey RN (pb)              | 1.1217                | 0.0001       | 3.07            |
| Sleeper seat availability RSL (pb) | 1.5665                | 0.000        | 4.79            |
| Service accessibility S(tr)        | 0.0408                | 0.0022       | 1.04            |
| Urban destination UR (gb)          | 2.3364                | 0.000        | 10.34           |
| LL(0)                              | -2201.619             |              |                 |
| LL(C)                              | -1644.212             |              |                 |
| LL( $\theta$ )                     | -1068.041             |              |                 |
| $\rho_0^2$                         | 0.515                 |              |                 |

travel time compared to in-vehicle travel time (Almasri, 2011). Therefore, total travel time was split into in-vehicle travel time of main mode (IVT) and total egress time (WTD) as out-vehicle travel time. IVT and IVT/distance variables have a positive sign which could not be justified; therefore, these are eliminated from the model. In contrast, WTD has a correct negative sign with more than a 0.01 significance level. The total egress time (WTD) was further split into egress travel time (TTD) and waiting time at egress (WD) for model refinement. Variables TTD, TTD/Travel distance, and TTD/egress distance (DD) were eliminated from the model due to their unjustified sign. They were removed from the model with a chi-square value justified at a 99% confidence level. The influence of TTD may be taken care of by the egress distance (DD) variable in the model. WD has been retained in the model and has a correct negative sign with more than a 0.05 significance level in the final model.

**Travel Cost:** Total travel cost (C) and C per unit distance were included in the model successively, which have a positive sign against intuition. Total travel cost was separated into the main mode cost (IVC) and egress cost (OVCD). Due to the positive sign of C, C/distance, IVC, and IVC/income, these variables were removed from the model. In contrast, egress travel cost per average income of income level (OVCD/Inc) was retained due to its proper negative sign with a confidence level of more than 99%.

**Egress Distance:** Egress distance (DD) has a higher variation because the destination region also includes a rural region. The egress distance has

a correct negative sign, as expected at a more than 99% confidence level. Therefore, it has been retained in the final model.

**Supply Available:** The accessibility index (S) between origins to destination region for the mode has been derived based on mode-wise supply availability. S (tr) and S (gb) were included in the model, but S (gb) was removed due to its insignificance level. The S (tr) has a correct positive sign with more than a 99.8% confidence level.

**Urban Destination:** The present study is a case of a multi-destination. The destination region consists of urban and rural regions. This factor has been taken into consideration by including UR as a dummy variable. Variable UR (tr) was eliminated due to its lower significance level because its influence might be considered by S (tr) variable. Wherein UR (gb) has a positive sign that supports intuition and has more than a 0.001 significance level. Travelers were asked to state the governing factors for selecting a particular mode. The governing factors, such as lesser egress distance (RD), lesser travel cost (RC), sleeping seat availability (RSL), availability of night journey (RN), and better service (RS) were analyzed to improve the model.

**Lesser Egress Distance:** To identify whether lesser egress distance is one of the important governing criteria for mode selection RD (tr) and RD (gb) were added to the model. However, RD (gb) was eliminated because of its poor significance level. Wherein RD (tr) has been retained in the model which has a correct negative sign with more than a 0.01 significance level. The negative sign of the RD (tr) reflects that travelers for whom less egress distance was important do not prefer train mode due to its poor distribution network in the destination region.

**Lower Travel Cost:** To evaluate the significance of the lower travel cost for the travelers, RC (tr) and RC (gb) were included in the model. Both the variables have a correct positive sign with an acceptable significance level. Usually, the travel costs of the GSRTC bus and train were lesser compared to the private bus mode. Therefore, the positive signs for RC (tr) and RC (gb) were justified.

**Peak Month:** To check the changes in mode preference during peak duration, a dummy variable representing peak month (PM) was added to the model. However, it was not significant even at a 90% confidence level. Therefore, it was eliminated from the model.

The model estimation results are shown in Table 5, and the final utilities for different modes are presented in Equations (5a) to (5c). All the parameters have a logical sign and desired significance level in the MNL model. The developed model was tested with 212 trip data. The log-

likelihood value for the fitted model was -98 against the log-likelihood value of the intercept-only model as -200. The McFadden  $\rho^2$  value for the testing data was 0.54, satisfactory for the logistic model.

$$U_{pb} = -0.0347 DD - 0.0373 WD - 0.0340 OVCD/Inc + 0.7966 LS + 1.5665 RSL(pb) + 1.1217 RN(pb) \quad (5a)$$

$$U_{tr} = 0.3459 - 0.0347 DD - 0.0373 WD - 0.0340 OVCD / Inc + 0.7966 LS - 0.2387 TW / EM(tr) - 0.7516 RD(tr) + 2.6390 RC(tr) + 0.0408 S(tr) \quad (5b)$$

$$U_{gb} = -1.1633 - 0.0347 DD - 0.0373 WD - 0.0340 OVCD/Inc + 0.7966 LS - 0.3079 TW/EM(gb) + 2.8247 RC(gb) + 2.3364 UR(gb) \quad (5c)$$

All variables in the final MNL model are statistically significant at a significant level larger than 0.05, except for TW/EM<sub>tr</sub>(0.07) with logically justified signs. The importance of the egress attributes is revealed in the model. The egress attributes such as egress distance, egress waiting time, and egress cost per average income of income level are significant in the mode choice decision. The value of Exp ( $\beta$ ) of these attributes is less than one, which implies that for each unit increase in the attribute, the odds are decreased in choosing the mode. The perceived level of service variable has a positive sign, as expected. The rise in the perceived level of service leads to increased chances of selecting the particular mode. The value of Exp ( $\beta$ ) for lesser travel cost as governing parameters for GSRTC mode is the highest. From the MNL model, GSRTC bus and railway mode with lesser travel cost as a governing attribute have 16.8 and 14 times the odds of selection over private bus mode. A positive sign of the night journey and sleeper seat availability as a governing criterion for private bus mode reveals the popularity of private buses. The probability of choosing GSRTC for the urban destination type is 10.3 times compared to the reference mode.

### Analysis of the MNL Model

It is worthwhile to know the extent to which the mode selection probability changes in response to changes in the value of important attributes. Elasticity measures the percentage change in the probability of choosing a particular alternative in the choice set with respect to a given percentage change in an attribute of the alternative. The elasticity for important attributes in the final model was obtained.

An arc elasticity was obtained with the probability weighted sample enumeration method which was recommended by Hensher *et al.* (2005). Direct and cross elasticities calculated for egress distance, waiting time, egress travel cost per average income of income level, and the number of two-wheelers per number of earning members in the household are presented in Table 6.

If the egress parameter DD was increased by 1% for gbus and train mode, the probability of choosing the gbus and train mode would be reduced by 0.16% and 0.33%, respectively. The direct elasticity of WD and OVCD/Inc for the train mode was higher than the alternative mode. This reflects

that there might be a more reduction in the probability of choosing the mode train with an increase in egress waiting time and egress cost.

Sensitivity testing for various scenarios was carried out because the elasticity for the categorically coded data is not meaningful. To study the mode sharing change with reference to improvement in the Perceived Level of Service (LS) variable for gbus and train were obtained as presented in Tables 7 and 8. The rise in the probability of choosing gbus and train mode were 13.1% and 20.1%, respectively, with the improvement in the level of service perceived by the travelers.

The binary variable UR for gbus was simulated to study the influence of destination type. It reveals that there is a 27.5% rise in the probability of choosing gbus if the change in the destination from rural to urban which is presented in Table 9.

**Table 6. The elasticity of important variables in the model**

| Attribute<br>(Alternative) | Direct Elasticity (DE) and Cross Elasticity (CE) |             |             |
|----------------------------|--------------------------------------------------|-------------|-------------|
|                            | pbus                                             | Gbus        | train       |
| DD (gbus)                  | 0.018 (CE)                                       | -0.157 (DE) | 0.037 (CE)  |
| DD (train)                 | 0.064 (CE)                                       | 0.122 (CE)  | -0.334 (DE) |
| WD (pbus)                  | -0.043 (DE)                                      | 0.092 (CE)  | 0.105 (CE)  |
| WD (gbus)                  | 0.019 (CE)                                       | -0.017 (DE) | 0.043 (CE)  |
| WD (train)                 | 0.047 (CE)                                       | 0.094 (CE)  | -0.239 (DE) |
| OVCD/Inc (pbus)            | -0.019 (DE)                                      | 0.039 (CE)  | 0.046 (CE)  |
| OVCD/Inc (gbus)            | 0.009 (CE)                                       | -0.075 (DE) | 0.017 (CE)  |
| OVCD/Inc (train)           | 0.023 (CE)                                       | 0.045 (CE)  | -0.116 (DE) |
| TW/EM (gbus)               | 0.023 (CE)                                       | -0.171 (DE) | 0.047 (CE)  |
| TW/EM (train)              | 0.025 (CE)                                       | 0.052 (CE)  | -0.117 (DE) |

### ANN Model

Models developed by conventional approaches are mathematically manageable but compromise on their performance due to their inherent constraints. While ANN has shown good potential to successfully model complex relationships where nonlinearity and inconsistent data adversely affect other approaches (Sayed and Razavi, 2000). In the study, a multilayer feed-forward neural network with a single hidden layer was developed with the

**Table 7. Sensitivity testing of attribute LS on mode gbus**

| Change in<br>Attribute<br>LS <sub>gb</sub> | Affected<br>Mode | Base    |                    | Scenario |                    | Scenario- Base       |                                 |
|--------------------------------------------|------------------|---------|--------------------|----------|--------------------|----------------------|---------------------------------|
|                                            |                  | % Share | Number of<br>Trips | % Share  | Number of<br>Trips | Change in<br>% Share | Change in<br>Number of<br>Trips |
| 3 to 4                                     | pbus             | 73.66   | 1,476              | 68.11    | 1,365              | -5.55                | -111                            |
|                                            | gbus             | 5.88    | 118                | 14.68    | 294                | 8.79                 | 176                             |
|                                            | train            | 20.456  | 410                | 17.22    | 345                | -3.24                | -65                             |
| 4 to 5                                     | pbus             | 68.11   | 1,365              | 65.24    | 1,307              | -2.87                | -58                             |
|                                            | gbus             | 14.68   | 294                | 18.96    | 380                | 4.28                 | 86                              |
|                                            | train            | 17.22   | 345                | 15.80    | 317                | -1.41                | -28                             |
| 3 to 5                                     | pbus             | 73.66   | 1,476              | 65.24    | 1,307              | -8.42                | -169                            |
|                                            | gbus             | 5.88    | 118                | 18.96    | 380                | 13.08                | 262                             |
|                                            | train            | 20.456  | 410                | 15.80    | 317                | -4.66                | -93                             |

**Table 8. Sensitivity testing of attribute LS on mode train**

| Change in<br>Attribute<br>LS <sub>tr</sub> | Affected<br>Mode | Base    |                    | Scenario |                    | Scenario- Base       |                                 |
|--------------------------------------------|------------------|---------|--------------------|----------|--------------------|----------------------|---------------------------------|
|                                            |                  | % Share | Number of<br>Trips | % Share  | Number of<br>Trips | Change in<br>% Share | Change in<br>Number of<br>Trips |
| 3 to 4                                     | pbus             | 73.72   | 1,477              | 67.33    | 1,349              | -6.39                | -128                            |
|                                            | gbus             | 13.87   | 278                | 11.47    | 230                | -2.40                | -48                             |
|                                            | train            | 12.42   | 249                | 21.20    | 425                | 8.78                 | 176                             |
| 4 to 5                                     | pbus             | 67.33   | 1,349              | 58.56    | 1,174              | -8.77                | -175                            |
|                                            | gbus             | 11.47   | 230                | 8.86     | 178                | -2.60                | -52                             |
|                                            | train            | 21.20   | 425                | 32.57    | 653                | 11.37                | 228                             |
| 3 to 5                                     | pbus             | 73.72   | 1,477              | 58.56    | 1,174              | -15.16               | -303                            |
|                                            | gbus             | 13.87   | 278                | 8.86     | 178                | -5.00                | -100                            |
|                                            | train            | 12.42   | 249                | 32.57    | 653                | 20.16                | 404                             |

Table 9. Sensitivity testing of attribute UR on mode gbus

| Change in Attribute UR <sub>gb</sub> | Affected Mode | Base    |                 | Scenario |                 | Scenario-Base     |                           |
|--------------------------------------|---------------|---------|-----------------|----------|-----------------|-------------------|---------------------------|
|                                      |               | % Share | Number of Trips | % Share  | Number of Trips | Change in % Share | Change in Number of Trips |
| 0 to 1                               | pbus          | 71.88   | 1,440           | 52.64    | 1,055           | -19.24            | -385                      |
|                                      | gbus          | 8.30    | 166             | 35.776   | 717             | 27.47             | 551                       |
|                                      | train         | 19.82   | 397             | 11.59    | 232             | -8.23             | -165                      |

Table 10. Correctly Classified mode (%)

| Sr | Mode      | Correctly classified (%) |
|----|-----------|--------------------------|
| 1  | Train     | 51.51                    |
| 2  | Gbus      | 82.93                    |
| 3  | Pbus      | 87.44                    |
|    | Aggregate | 80.34                    |

Table 11. Network sensitivity analysis

| No. of Hidden nodes | Coefficient of Correlation (R) |            |         |        |
|---------------------|--------------------------------|------------|---------|--------|
|                     | Training                       | Validation | Testing | Total  |
| 5                   | 0.8328                         | 0.8066     | 0.7717  | 0.8175 |
| 10                  | 0.8646                         | 0.7816     | 0.8103  | 0.8424 |
| 15                  | 0.9079                         | 0.8282     | 0.754   | 0.873  |
| 20                  | 0.8786                         | 0.8233     | 0.8527  | 0.8659 |
| 25                  | 0.9398                         | 0.8653     | 0.8254  | 0.9105 |
| 30                  | 0.8858                         | 0.8601     | 0.8199  | 0.8725 |
| 35                  | 0.9099                         | 0.7861     | 0.8556  | 0.8845 |
| 40                  | 0.9048                         | 0.8246     | 0.7756  | 0.871  |
| 45                  | 0.9514                         | 0.8413     | 0.7671  | 0.9058 |

backpropagation Levenberg-Marquardt algorithm. In ANN, it is not always possible to find the significance of the input variables in advance (Sayed and Razavi, 2000). Even the statistical tests used to find input significance in logistic regression cannot be used due to the nonlinear nature of ANN (Dreiseitl and Ohno-Machado, 2002). Therefore, the input variables that had been excluded in the MNL model due to their insignificant contribution or logical signs in the MNL model were excluded in the ANN model to make it more efficient.

So, input variables of the MNL model were used in the ANN-1 model after normalization between 0 and 1. The nodes in the input layer were 11. After the number of trials, 35 hidden nodes were observed as an optimum number to avoid over fitting. The coefficient of correlation (r) value, used to measure the correlation between actual and predicted output in training, validation, and testing, are 0.85, 0.76, and 0.75, respectively. However, the correctly classified percentage shown in Table 10 reveals that the model needs improvement to serve the aim. The mode selection process of decision-makers is highly concerned with the difference in attributes (Train, 2002).

Therefore, the difference in the generic variables, such as egress distance, waiting time at

Table 12. Prediction success rate

| No. of Hidden nodes | Prediction success (%) |              |             |              |
|---------------------|------------------------|--------------|-------------|--------------|
|                     | Train                  | Gbus         | Pbus        | Total        |
| 5                   | 52.05                  | 69.92        | 88.73       | 79.74        |
| 10                  | 52.05                  | 73.98        | 89.66       | 80.89        |
| 15                  | 74.79                  | 81.71        | 91.1        | 86.98        |
| 20                  | 60.27                  | 85.77        | 87.72       | 82.49        |
| 25                  | <b>84.93</b>           | <b>82.52</b> | <b>93.4</b> | <b>90.52</b> |
| 30                  | 64.11                  | 84.55        | 88.73       | 83.73        |
| 35                  | 66.03                  | 83.74        | 90.6        | 85.28        |
| 40                  | 59.45                  | 85.77        | 89.38       | 83.48        |
| 45                  | 83.01                  | 81.3         | 93.68       | 90.22        |
| 50                  | 62.19                  | 85.37        | 89.52       | 84.03        |

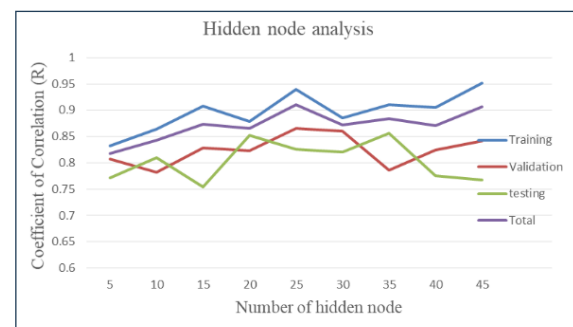


Figure 9. Hidden node analysis

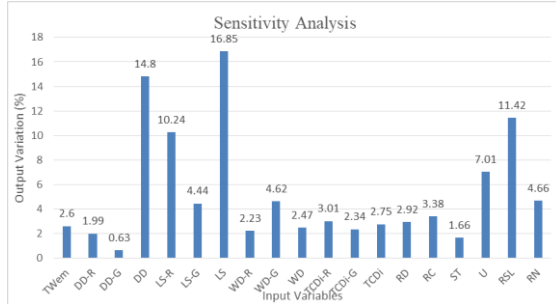
egress, egress cost/income were added to the model to make it more realistic. The revised ANN model (ANN-2) has been developed with 19 input variables as input nodes. The most efficient model has been selected based on the results of the various trials with the varying number of hidden nodes, as shown in Figure 9 and Table 11. The percentage of the correctly classified modes is tabulated in Table 12 for all the trials.

The model with 35 hidden nodes has a maximum r value in testing. However, considering the prediction accuracy of modes, as shown in Table 12, the model with 25 hidden nodes is considered the best model. It has r value in training, validation, and testing as 0.9398, 0.8653, and 0.8254, respectively.

The cross-classification data presented in Table 13 indicates that the ANN-2 model has a higher prediction capacity than the previous ANN model.

**Table 13. Prediction success table**

| Actual output | Predicted output |      |      |
|---------------|------------------|------|------|
|               | Train            | Gbus | Pbus |
| Train         | 84.9             | 14.5 | 0.5  |
| Gbus          | 10.6             | 82.5 | 6.9  |
| Pbus          | 0.3              | 6.3  | 93.4 |

**Figure 10. Sensitivity analysis of input variables**

The performance of ANN is good in predicting the observed choice. However, it suffers from the lack of methods for the interpretation of the significance of input variables. This black box image of the neural network is revealed by finding the effect of the variables on the output. For this purpose, a relative variation in the output with respect to varying the input value by its standard deviation from the mean was determined by keeping all other variables at their mean value. The procedure was repeated for all variables. It is found that perceived level of service, egress distance, type of destination, egress waiting time as well as sleeping seat availability and night journey as governing parameters, have more relative importance (RI), which is shown in Figure 10. The importance of these variables has been observed in the MNL model also.

### Models' Performance Comparison

The ANN-2 model's minimum prediction success rate for the GSRTC bus mode is 82.5%, and its overall prediction rate is 90.52%, which is higher than the MNL model's 70.11% success rate. Additionally, the private bus mode has a 93.4% consistently high prediction success rate.

However, the MNL model is useful for analyzing the influencing attributes. The ANN-2 model has more prediction power. Therefore, the ANN model is more useful whenever the mode prediction is only concerned. However, due to the black box image of the ANN model, the MNL model is advantageous for the analysis of the attributes which are useful for policy decisions.

## Conclusions

The results obtained by developing the inter-regional mode choice model for the passenger travel between Surat city and the Bhavnagar region of Gujarat state may help the government and the public transport agencies to make marginal decisions. Several conclusions derived from this study are as follows:

- In the final MNL model, all the parameters have expected signs and are found to be significant at a 95 percent confidence level. The best-fitted ANN-2 model is found with 25 hidden nodes.
- A comparison of the prediction result of all models shows that the performance of the ANN-1 model with revised input is better than other models. The prediction success rate of the ANN-2 model is increased to 90.52% compared with 70.11% of the MNL model. Thus the performance of the conventional ANN model can be increased by adding input variables representing the difference in attributes.
- The MNL model indicates that vehicle ownership, egress distance, waiting time at egress, transport service availability, type of destination (Urban/rural), and facility/service level offered by the mode have impacted the mode choice decision. All these variables, except vehicle ownership and transport service availability, have been found important in the ANN-2 model analysis also.
- Additionally, travelers were asked about their governing parameters for the mode choice. It is observed that lesser egress distance, lesser egress cost, availability of night journey, and availability of sleeping seats are important parameters for the mode selection
- Governing dummy parameters such as availability for night journeys and reserved sleeping seats for the private bus has a strong positive effect with  $\beta = 1.12$  and  $1.567$  reflecting the significant reasons for private bus popularity. Sensitivity analysis of the ANN-2 model also reveals the very high importance of these parameters with RI of 4.66% and 11.42%, respectively. Railway and GSRTC modes should change their timing to attract more travelers.
- Vehicle ownership (TW/EM) in the household has a negative effect on choosing GSRTC and Railway mode compared to the private bus, though it has a comparatively higher fare structure. This reflects that travelers with

higher vehicle ownership (indirectly higher income) prefer costly travel modes.

- Travel time parameters have no proper effect on the mode choice decision might be due to the overnight journey. Minor travel time saving is not governing criterion when a sleeping facility is available during an overnight journey. The effect of egress travel time is indirectly considered in the egress distance parameter, which has a proper negative sign with a 95 percent confidence level.
- The private bus popularity reveals that the effect of travel cost is dominated by better service level offered and favorable egress parameters.
- Levels of service perceived to have a strong positive effect ( $\beta = 0.7965$ ) on choosing the mode. The same has been agreed by ANN-2 model analysis, also with 16.85% RI. The level of service of the modes is ranked based on regularity, fare structure, travel time, vehicle condition, etc., perceived by travelers. The probability of choosing the less ranked modes GSRTC bus and train can increase up to 13.1% and 20.2%, respectively, with the improvement in the level of service.
- Destination in the urban area has a strong positive effect ( $\beta=2.3371$ ) on choosing the GSRTC bus, reflecting its uneven distribution network in the rural region. There is a 27.5% rise in the probability of choosing GSRTC bus if a change in destination variable UR changes from 0 to 1. It may be the scope for the GSRTC to attract passengers having a destination in the rural region by providing the proper connectivity in the rural area.

#### Application of the Study

- Data collection technic and methodology used in the study might be helpful for similar types of the study.
- The ANN-2 model developed with differences of several attributes as input is more precise. Therefore, it can be incorporated in future studies as mode choice decision is influenced by differences in attributes.
- The output of the model developed is useful for the improvement of the existing public transport facilities.
- Based on the model's recommendations, researchers might hypothetically enhance future intercity travel modeling and offer further ideas to transport organizations. It might also improve the decision-making process of government and private services.

#### Limitations and Recommendations for Future

- The present research is limited to the Bhavnagar region due to restrictions on funds. Future studies may include similar regions. Such supporting studies can effectively validate the models for geographical transferability and give generalized concluding statements.
- Due to the unavailability of regular air services and lesser travel distance, the air service is excluded from the study. Once regular air mode service data is available, it can be included.
- The perceived level of service attribute is observed as an important attribute in the choice of the mode. The respondents were asked to judge the level of service based on a few parameters. The importance of these individual parameters is not analyzed in the study. However, the influence of each attribute on travelers' perceived level of service can be evaluated individually for a better understanding of an individual's perception of those parameters.
- The present study highlights the importance of governing criteria responsible for travelers' mode choices perceived by them. These criteria should be analyzed in-depth as they concern the decision maker's thinking process. Moreover, the relationship between the governing criteria and the socio-economic characteristics of the travelers might be studied, which might be useful for future modal split analysis.

#### References

- Al-Ahmadi, H. (2006). Development of intercity mode choice models for Saudi Arabia. *Journal of King Saudulaziz University*, 17(1):3-20. <https://doi.org/10.4197/Eng.17-1.1>.
- Almasri, E. (2011). Factors affecting travel choice of shared taxi versus bus for palestinian university student trips. *International Review of Civil Engineering*, 2(1):35-45.
- Almasri, E. and Alraee, S. (2013). Factors affecting mode choice of work trips in developing cities- gaza as a case study. *Journal of Transportation Technologies, Scientific Research*, 3(4):247-259. <https://doi.org/10.4236/jtts.2013.34026>.
- Arasan, T. and Vedagiri, P. (2011). Modeling Model Shift from Personal Vehicles to Bus on Introduction of Bus Priority Measure. *Asian Transport Stud.* 1(3): 288-302.
- Ashalatha, R., Manju, V.S., and Zacharia, A.B. (2013). Mode Choice Behavior of Commuters in Thiruvananthapuram City. *Journal of Transportation Engineering*, 139(5):494-502. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000533](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000533).
- Bardhan, R., Varghese, V., and Jana, A. (2015). Analyzing Regional Travel Patterns in India: Disaggregated Analysis of Social, Health and Pilgrimage Trips. *Journal of the*

- Eastern Asia Society for Transportation Studies, 11:362-378.
- Barus, L.S., Martell-Flores, H., Hadiwardoyo, S.P., and Batoz, J.L. (2016). Intercity mode choice modelling: Considering the intracity transport systems with application to the jakarta-bandung corridor. *International Journal of Technology*, 7(4):581-591. <https://doi.org/10.14716/ijtech.v7i4.2798>.
- Ben-Akiva, M. and Lerman, S.R. (1985). *Discrete Choice Analysis: Theory and application to Travel Demand*. The MIT Press, Cambridge, London. 384p.
- Ben-Akiva, M. and Morikawa, T. (1990). Estimation of switching models from revealed preferences and stated intentions *Transportation Research Part A: General* 24(6):485-490. [https://doi.org/10.1016/0191-2607\(90\)90037-7](https://doi.org/10.1016/0191-2607(90)90037-7).
- Bhat, C.R. (1995). A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research Part B: Methodological*, 29(6):471-483. [https://doi.org/10.1016/0191-2615\(95\)00015-6](https://doi.org/10.1016/0191-2615(95)00015-6).
- Cantarella, G.E. and de Luca, S. (2005). Multilayer feed forward networks for transportation mode choice analysis: An analysis and a comparison with random utility models. *Transportation Research Part C: Emerging Technologies*, 13(2):121-155. <https://doi.org/10.1016/j.trc.2005.04.002>.
- Domencich, T.A. and McFadden, D. (1975). *Urban Travel Demand: A Behavioral Analysis*. North-Holland, Amsterdam. 215p.
- Dreiseitl, S. and Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*, 35:352-359. [https://doi.org/10.1016/S1532-0464\(03\)00034-0](https://doi.org/10.1016/S1532-0464(03)00034-0).
- Ghareib, A. (1996). Evaluation of logit and probit models in mode-choice situation. *Journal of Transportation Engineering*, 122(4):282-290. [https://doi.org/10.1061/\(ASCE\)0733-947X\(1996\)122:4\(282\)](https://doi.org/10.1061/(ASCE)0733-947X(1996)122:4(282)).
- Haire, A.R. and Machemehl, R.B. (2010). Regional and modal variability in effects of gasoline prices on U.S. transit ridership. *Transportation Research Record: Journal of the Transportation Research Board*, 2144(1):20-27. <https://doi.org/10.3141/2144-03>.
- Hensher, D.A. and Greene, W.H. (2003). The Mixed Logit model: The state of practice. *Transportation (Amst)*, 30:133-176. <https://doi.org/10.1023/A:1022558715350>.
- Hensher, D.A., Rose, J.M., and Greene, W.H. (2005). *Applied Choice Analysis A Primer*. Cambridge University Press, Cambridge. 717p. <https://doi.org/10.1017/CBO9780511610356>.
- Hogg, R.V., Tanis, E.A., and Zimmerman, D.L. (2015). *Probability and Statistical Inference*. 9th ed. Pearson, NY. 1,091p.
- Horowitz, J.L. (1982). Evaluation of Usefulness of Two Standard Goodness-of-Fit Indicators for Comparing Non-Nested Random Utility Models. *Transportation Research Record: Journal of the Transportation Research Board*. 874:19-25.
- Islam, M., Akter, S., Imran, M., Hossain, I., and Hassan, N. (2016). Festival Time Transportation Demand Modeling in Bangladesh. *International Journal Science Engineering Investigations*, 5(48):40-44.
- Jana, A. and Varghese, V. (2017). Analyzing mode choice for inter-regional travel in India. In: *Proceedings of the World Conference on Transport Research-WCTR 2016*; July 10-15, 2016; Shanghai, *Transportation Research Procedia*, Elsevier, 25:5,220-5,234. <https://doi.org/10.1016/j.trpro.2018.02.049>.
- Koppelman, F.S. and Bhat, C. (2006). *A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models*. U.S. Department of Transportation Federal Transit Administration, U.S., p. 241.
- Landinois, C. (1987). Supply models for intercity passenger transport: a review. *Transport Reviews*. 7(2):119-143. <https://doi.org/10.1080/01441648708716649>.
- Levy, P.S. and Lemeshow, S. (2008). *Sampling of Populations: Methods and Applications*. 4<sup>th</sup> ed. Wiley, 616p. <https://doi.org/10.1002/9780470374597>.
- Manssour, A., Ahmed, M., and Riza, A. (2013). Modeling of intercity transport mode choice behavior in libya: a binary logit model for business trips by private car and intercity bus. *Australian Journal of Basic and Applied Sciences*, 7(1):302-311.
- Miller, E.J. (2004). The trouble with intercity travel demand models. *Transportation Research Record: Journal of the Transportation Research Board*, 1895(1):94-101. <https://doi.org/10.3141/1895-13>.
- Minal, C. and Sekhar, R. (2014). Mode choice analysis: the data, the models and future ahead. *International Journal for Traffic and Transport Engineering (IJtte)*, 4(3):269-285. doi: 10.7708/ijtte.2014.4(3).03. [https://doi.org/10.7708/ijtte.2014.4\(3\).03](https://doi.org/10.7708/ijtte.2014.4(3).03).
- Miskeen, M.A. and Rahmat, R.A. (2011). Development of disaggregate mode choice models of intercity travel in libya. In: *Proceedings of the 6<sup>th</sup> IEEE Joint International Technology and Artificial Intelligence Conference*; August 20-22, 2011, Chongqing, China, IEEE, p. 197-201.
- Miskeen, M.A., Alhodairi, A.M., and Rahmat, R.A. (2013). Modeling a Multinomial Logit Model of Intercity Travel Mode Choice Behavior for All Trips in Libya. *International Journal of Civil Architectural, Structure and Construction Engineering*, 7(9):308-317.
- Ortuzar, J.D. and Willumsen, L.G. (2011). *Modelling Transport*. 4<sup>th</sup> Ed. Wiley, Chichester, UK, 586p. <https://doi.org/10.1002/9781119993308>.
- Philip, M., Shreelatha, T., and Soosan, G.T. (2013). Modelling on mode choice behavior of rural middle class residents - an activity based approach. *International Journal of Emerging Technology & Advanced Engineering*, 3(7):150-155.
- Rajalakshmi, P. (2013). Mode choice modelling based on work trips- artificial neural network model. *International Journal of Innovative Research in Science, Engineering and Technology*, 2(1):360-369.
- Sayed, T. and Razavi, A. (2000). Comparison of neural and conventional approaches to mode choice analysis. *Journal of Computing in Civil Engineering*, 14(1):23-30. [https://doi.org/10.1061/\(ASCE\)0887-3801\(2000\)14:1\(23\)](https://doi.org/10.1061/(ASCE)0887-3801(2000)14:1(23)).
- Sikdar, P.K. and Sekhar, R. (2005). Analysis of artificial neural networks for mode choice modeling. In: *Proceeding of International Conference on Structural and Road Transportation Engineering (START 2005)*; January 3-5, 2005; Kharagpur, India.
- Subbarao, S.V. and Krishnarao, K. (2013). Analysis of Household Activity and Travel Behaviour: A case of Mumbai Metropolitan Region. *International Journal of Emerging Technology & Advanced Engineering*, 3(1):98-109.
- Surat Municipal Corporation [SMC]. (2020). Surat Municipal Corporation, Surat: SMC. Available from: <https://www.suratmunicipal.gov.in/TheCity/City/Stml13>. Accessed date: January 25, 2020.
- Tamin, O.Z. (2000). *Perencanaan dan Pemodelan Transportasi*. Penerbit ITB, Indonesia. 1281p.
- Train, K. (2002). *Discrete Choice Methods and Simulation*. Cambridge University Press, Cambridge. 378p. <https://doi.org/10.1017/CBO9780511753930>.
- Walker, J. (2002). Mixed logit (or Logit Kernel) model: dispelling misconceptions of identification. *Transportation Research Record: Journal of the Transportation Research Board*,

- 1805(1):86-98. <https://doi.org/10.3141/1805-11>.
- Yao, E. and Morikawa, T. (2005). A study of an integrated intercity travel demand model. *Transportation Research Part A: Policy and Practice*, 39(4):367-381. <https://doi.org/10.1016/j.tra.2004.12.003>.
- Zhao, S., Wu, N., and Wang, X. (2018). Impact of Feeder Accessibility on High-Speed Rail Share: Wuhan-Guangzhou Corridor, China. *Journal of Urban Planning and Development*, 144(4):04018029-1-8. [https://doi.org/10.1061/\(ASCE\)UP.1943-5444.0000475](https://doi.org/10.1061/(ASCE)UP.1943-5444.0000475).
- Zhao, X., Yan, X., Yu, A., and Van Hentenryck, P. (2020). Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behaviour and Society*. 20:22-35. <https://doi.org/10.1016/j.tbs.2020.02.003>.
- Zhou, X, Du, H., Liu, Y., Huang, H., and Ran, B. (2017). Investigating the intention of rural residents to use transit in cixi, China. *Journal of Urban Planning and Development*, 143(2):05016017-1-8. [https://doi.org/10.1061/\(ASCE\)UP.1943-5444.0000367](https://doi.org/10.1061/(ASCE)UP.1943-5444.0000367).
- Zhu, M., Li, X., and Li, J. (2010). Travel mode choice of residents in spring festival based on neural network. In: *Proceedings of the International Conference on Computational and Information Sciences*; Dec 17-19, 2010; Chengdu, China, IEEE, p. 1,118-1,121. <https://doi.org/10.1109/ICCIS.2010.277>.