European Transport \ Trasporti Europei (2021) Issue 84, Paper n° 3, ISSN 1825-3997

https://doi.org/10.48295/ET.2021.84.3

Threshold value estimation of journey-distance using generalized polynomial function

Subhojit Roy [1], Debasis Basu[2*]

[1]*Ph.D. Student in School of Infrastructure, Indian Institute of Technology Bhubaneswar*
[2]*Faculty in School of Infrastructure, Indian Institute of Technology Bhubaneswar*

Abstract
The present work demonstrates an experience in estimating the threshold value of journey distances travelled by transit passengers using generalized polynomial function. The threshold value of journey distances may be defined as that distance beyond which passengers might no more be interested to travel by their reported mode. A knowledge on this threshold value is realized to be useful to limit the upper-most slab of transit fare, while preparing of a length-based fare matrix table. Theoretically, the threshold value can be obtained at that point on the cumulative frequency distribution (CFD) curve of journey distances at which the maximum rate of change of the slope of curve occurs. In this work, the CFD curve of the journey distance values is empirically modelled using Newton's Polynomial Interpolation method, which helps to overcome various challenges usually encountered while an assumption of a theoretical probability distribution is considered a priori for the CFD.

Keywords: threshold value, journey length, transit-passengers, frequency distribution, generalized polynomial.

1. Introduction

Many studies on the pattern of trip lengths in general and journey distances in particular were carried out on public transport to obtain a knowledge on the variation of travel demand for various segments of journey length, and to prepare length-based fare matrix for transit service with an aim to optimize the revenue generation. As per Bandegani and Akbarzadeh (2016), the primary source of revenue generation for any transit agency is the fare, and it depends to a great extent on the pattern of transit demand. In urban areas of developing countries like India, transit fare is often considered as major socio-political concern (Lupo, 2013; Yook and Heaslip, 2015) and therefore, preparation of a fare structure often becomes a challenging task (Zhang et. al., 2019). In this aspect, it needs to be highlighted that fare-slab on various likely segments of journey length (Yook and Heaslip, 2015), and their corresponding passenger-kilometre demand (Narboneta and Teknomo, 2016) information could be an essential input. Usually, two types of model are applied for urban transit fare calculation such as the flat-fare model (Harmony, 2018) and the length-based graduated fare model (Moore, 2002). In a flat-fare model, the fare of a transit-line is predetermined and, it is in-sensitive to the passenger's trip-length pattern like journey length or tentative journey time. However, in a length based graduated model, the fare slab could be a function of the journey-length (Yook and

* Corresponding author: Debasis Basu, dbasu@iitbbs.ac.in

Heaslip, 2015; Narboneta and Teknomo, 2016). Among all predominant fare structure models, the length-based graduated fare model is usually considered by transit operators (Bullard et al., 2004). Many often the cause of poor revenue generation by transit operators is attributed to the lack of understanding (Cervero and Wachs, 2002) of the average passenger-distance demand, while fixing up of the length-based fare matrix. In such situation, knowledge on the journey length pattern (Yook and Heaslip, 2015) of transit users, and also on the upper limit of their journey length, called as threshold value, are usually found essential information for preparation of length-based fare matrix. The threshold value of journey length may be defined as that length beyond which transit passengers might no more be interested to travel. The demand beyond this value may be expected to drop significantly. The need for estimation of threshold value was also found in other contexts of transportation engineering such as limiting value of traffic queue (Fathy and Siyal, 1995), usual walk-extent in urban area (Daniels and Mulley, 2013.), limiting value of journey length estimation by bicycle (Lamont 2009), limiting value of waiting time at transit stops (Roy and Basu, 2021) etc. The knowledge on this threshold value of journey-length also becomes imperative to limit the upper-most slab of fare in a length-based fare matrix table. It may be highlighted that the upper-most fare slab in urban areas of developing countries are usually regulated by transit-regulatory authority. The transit operators need to plan to maximise their revenue generation by limiting the upper most fare slab considering the threshold value of journey length. Any new fare slab of journey lengths beyond this estimated threshold value may be suggested to keep at flat rate, which could be equal to the upper-most fare slab set out by the transit-regulatory authority.

In view to identify the threshold value, the current work proposes an approach for estimating the threshold value of journey length. As per Seneviratne (1985), the threshold value of journey length is that point on the cumulative frequency distribution (CFD) curve of journey length at which the maximum rate of change of the slope of curve occurs. At this point, even a small change in journey length is expected to affect large number of passengers opting of that transit service. Mathematically, the value of this point can be estimated by obtaining 3rd order derivative of the CFD function, and then finding roots to the solution of the 3rd order derivative. In view of the threshold value estimation of journey lengths, it necessitates knowledge on the functional form of CFD for the observed journey lengths. Modelling of journey length distribution was previously studied by many researchers (Gray and Sen, 1983; Kim et al., 2003), who assumed that the journey length distribution follows an exponential class of probabilistic process. Some other studies (Bandara and Wirasinghe, 1992; Arasan et al., 1994) showed normal and gamma distribution functions but for walk-length distribution. Many of the other works (Bandara and Wirasinghe, 1992; Jin et. al., 2009; Liang et. al., 2013; Ben-Edigbe et al., 2014) provided evidences of various distributional functions being employed for modelling of trip length distributions such as exponential or power function with negative exponent, Erlang distribution, log-normal distribution, gamma distribution etc. Though, all of these studies modeled the trip length or the journey length frequency distribution curves by a function of probabilistic processes, none of them demonstrated any procedure in terms of estimating threshold value of journey length. Arasan et al. (1994) once showed the threshold value estimation for walk length's distribution analytically by evaluating of the 3rd order derivative of a gamma distributed curve. Recently, Verma et al., (2018) showed the application of gamma and log-normal distributions in estimating the threshold value of walk-length using their 3rd order derivatives. In line with this, the work reported by

Roy and Basu (2021) can be appreciated, which showed the estimation of threshold value of waiting time for transit passengers at transit stops. Inspite of all these studies, there is a consensus that even though the pattern of journey length mostly remains similar in nature, they are often described (Jin and Fricker, 2008) by different distributional assumptions and within acceptable goodness-of-fit for all of these cases. It often leads to varying the interpretation by the analysts depending upon their distributional assumption. Apart from this, Jin and Fricker (2008) and Horbachov and Svichynskyi (2018) expressed their concern that there could exist no definite theoretical form for describing the general shape of an observed trip-length distribution curve. Besides, the distributional pattern may not always follow a known probabilistic process. Even then, it often becomes cumbersome to obtain the 3rd order derivative of CFD following of an exponential class of probabilistic process. Sometimes the 3rd order derivative doesn't even exist (Richards, 1984) due to non-differentiability of the function at every point in the function domain. The difficulty sometimes becomes even more, if an analyst does not have a priori knowledge on the afore-mentioned distributional pattern. In such situation, a priori assumption on the probability density function (PDF) may often lead to an erroneous conclusion (Kim et. al., 2003) on the assumption.

In order to overcome the above shortcomings, an alternative approach is explored obviating the knowledge on the assumption of theoretical probability distribution. The observed CFD curve could be modelled using a generalized polynomial function within the observed domain of the journey distance values. A generalized polynomial is an expression consisting of variables and coefficients, and it involves simple operations of addition, subtraction, multiplication, and contains non-negative integer exponentiation of variables. In this study, the CFD curve of the journey distance values is modelled using Newton's Polynomial Interpolation method (Krogh, 1970; Werner, 1984). Generally, Newton's polynomial is usually employed to model curves having limited number of observational data points, and its application encompasses various fields of engineering such as fluid dynamics (Rubin and Khoshla, 1977), solid structure (Hyman and Larrouturou, 1982), signal processing applications (Yabe and Aoki, 1991) etc. However, the application of such function is almost sparse in transportation engineering. In view of this, the study demonstrates a new experience in the application of Newton's polynomial function for estimation of the threshold value of journey length. The approach is demonstrated by estimating threshold values for two types of urban transit passengers such as city bus and shared-auto operational in Bhubaneswar, India.

2. Basic information of datasets used in the study

The Bhubaneswar is a mid-sized urban area spread over 135 square kilometres, and having population of around 885363 (eight hundred eighty five thousand and three hundred sixty three) as per the 2011 census data (MHA, 2011). The city is classified as B-2 type as per the HRA classification, Ministry of Finances (MoF), Govt. of India, and as Tier-II city by the Reserve Bank of India (RBI) classification (MoF, 2015). Out of the total area, about 32% is meant for residential purposes, and 15% is used for commercial and industrial activities (CUE, 2012). However, only 8% of the total area is so far meant for the road transport sector. Among various travel modes, city bus and shared-auto services (Fig. 1) are found to cater to the primary needs of public transport, and they often pose competitive to each other (Singh, 2012; Roy and Basu, 2021). Therefore, mode switching behaviour between city bus and shared-auto is often observed among transit

commuters. Table 1 shows the characteristics of operational service-line of city bus and shared-auto services in Bhubaneswar. In this work, a primary survey was carried out to obtain various travel related information for city bus and shared-auto passengers. In the survey, the actual travel information of passengers alighting either from city bus or from shared-auto services was collected at or nearby of urban local bus stops. The boarding-alighting activities of bus passengers were found to occur at designated urban bus stops. But, the same bus stops were usually found to be used by shared-auto services as well for boarding-alighting of their passengers.

Figure 1. Images of city bus and shared-auto services operational in Bhubaneswar, India

Table 1. Characteristics of operational service-lines of city bus and shared auto in Bhubaneswar (adapted from Bhide, 2015; CUE, 2012)

Service-line characteristics	City Bus	Shared Auto
No. of operational service lines	18	20
Max. length (kilometres) of a service line	30.5	25
Min. length (kilometres) of a service line	10.2	5.3
Avg. length (kilometres) of operational service lines	27.8	12.4
Std. deviation of length (kilometres) of operational service lines	4.64	3.65

During on field data collection, alighting passengers from city bus and shared auto were intercepted with a request to participate in the survey. The field study and data collection were conducted across major urban bus stop locations in the city encompassing various types of urban form such as residential, workplace, business, shopping area etc. The alighting-passengers were intercepted on random sampling basis, where probability of interception of any passenger was equally-likely. Apart from collecting of the boarding stop name and/or its location name, GPS (Global Positioning System) location information (i.e. Coordinates in the form of longitude and latitude) of the boarding point

was also taken using GOOGLE Map App on smart mobile phone devices (See Figure 2.0). The GPS location of the alighting point was also taken along with the boarding location for each of the survey respondent. While recording of GPS data points from a survey respondent, other travel related information such as types of transit service, service-line information (like route number or route name) of the transit service, transit fare, trip-purpose and other socio-demographic information of the respondent were also collected. The GPS information was recorded with a precision of 7 decimal places for each of the longitude and latitude coordinate.

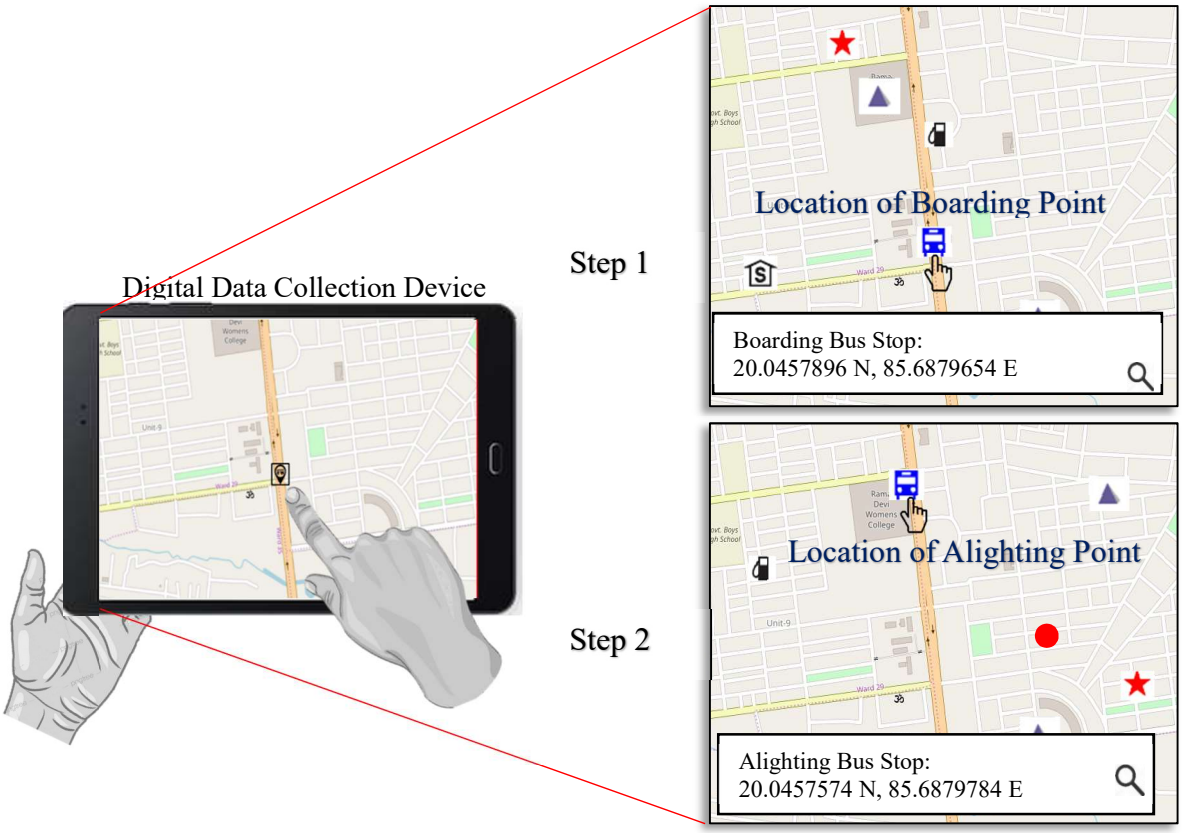

Figure 2. The image illustrates GPS based location collection on field using App in Smart Mobile phone device

It may be mentioned that such type of GPS data collected with 7 decimals of precision could have an error amount within of 5 metres (US DoD, 2020). The GPS data of survey respondents were then de-coded in GIS (geographical information system) environment to find out the journey distance travelled by a passenger along with other reported information such as types of transit mode, and its service line number or route name. While finding out of the journey distance being travelled, route numbers or names assigned for city-bus and shared-auto services were used as a guideline for mapping in GIS –based road network map. Although, several city bus passengers and shared-auto passengers were intercepted during the on-field survey, the final database was developed using responses from 993 city bus passengers and 596 shared-auto passengers. The following table (Table 2) shows the descriptive statistics of the observed journey-distances for two types of transit passenger.

Table 2. Characteristics of observed journey distance for city bus and shared auto passengers

	City bus passengers	Shared auto passengers
Sample size	993	596
Max. journ. dist. (kilometres)	26.0	22.0
Min. journ. dist. (kilometres)	1.00	0.50
Avg. journ. dist. (kilometres)	8.44	5.36
Std. Dev (kilometres)	4.21	3.17

3. Observed frequency distribution of journey distance

Figure 3 shows the frequency distribution (FD) of observed journey distances travelled by transit passengers; whereas Figure 4 shows the cumulative frequency distribution (CFD) of the same. It is observed that nature of the trend of distribution of journey distances for city-bus and shared-auto passengers are similar. The peak for the FD of journey distances of city bus passengers is found to occur around distance of 6.5 kilometres; whereas the same for shared-auto passengers is found to occur at 6 kilometres. This FD curves show that about 50% of transit passengers (more precisely 47% in case of city-bus and 50.3% in case of shared-auto passengers) travel about 6 kilometres by transit services. The proportion of passengers travelling beyond 6 kilometres by city-bus services follows a relatively flatter trend than that of shared-auto services. This fact is characterised by observing a longer tail of the distribution. On the other hand, the proportion of passengers travelling beyond 6 kilometres. by shared-autos follows a sharp decreasing trend. As per the observed datasets, the mean journey distance is observed to be 8.44 kilometres and 5.36 kilometres for city bus and shared-auto passengers respectively. The Standard deviation (Std. Dev.) is observed to be about 1 kilometres more for city-bus passengers than that of shared-auto passengers. The CFD curve (see Fig 4) implies that shared-auto passengers travel shorter journey distance than that of city-bus passengers. It is also noted from Fig. 4 that 85th percentile of city bus passengers travel up to a distance of 10.5 kilometres; whereas 85th percentile of shared-auto passengers travel up to a distance of 6.5 kilometres. As per Seneviratne (1985), the 85th percentile values could be used as threshold value of journey distance in the absence of values estimated from any demonstrated model.

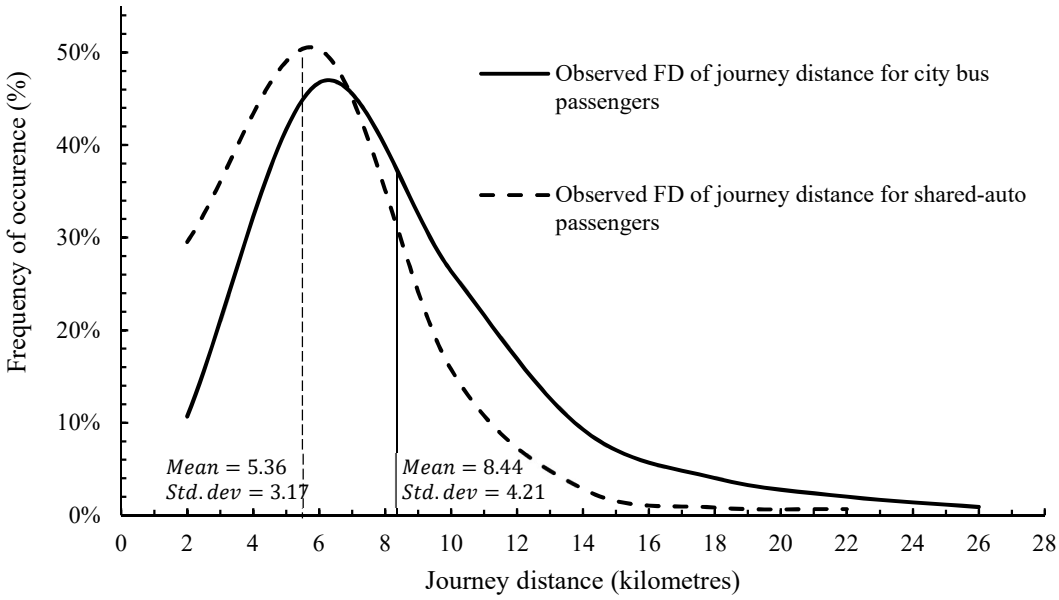

Figure 3. Observed FD of journey distance

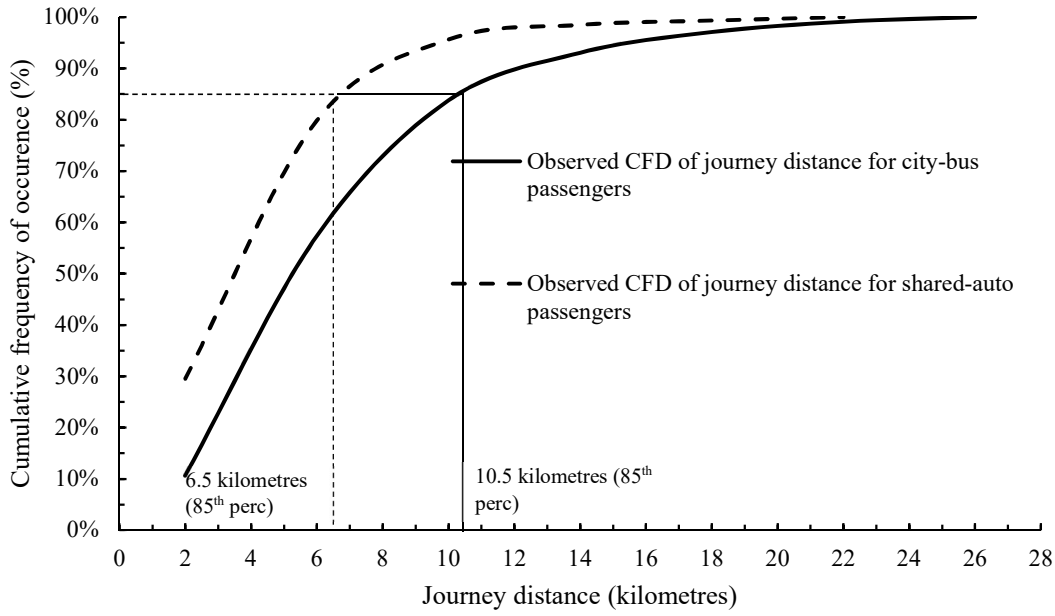

Figure 4. Observed CFD of journey distance

3.1. Estimation of threshold value of journey distance using generalized polynomial function

As mentioned, the study demonstrates an approach to estimate the threshold value of journey distance by modelling of the CFD curve using generalised polynomial function. According to the work of Seneviratne (1985), the threshold value of journey-distance may be estimated by finding of the maximum rate of change of slope of the modelled CFD curve. Beyond this threshold value of journey distance, the proportion of passenger-demand travelling by a particular transit mode is expected to fall significantly.

The CFD (say, $F(x_{ij})$) in Figure 4. represents a continuous function of journey distance in the domain of observed values for the journey distances (say, $x_{ij}, where\ i\ is\ boarding\ point\ and\ j\ is\ the\ alighting\ point$) The slope of CFD is given by the expression $\frac{dF(x_{ij})}{dx_{ij}}$, and then the rate of change of the slope can be expressed by $\frac{d^2F(x_{ij})}{dx_{ij}{}^2}$. Then, the threshold value of journey distance can be estimated by finding that point, where the maximum rate of the change of slope of the CFD occurs i.e.

$$\frac{d^3F(x_{ij})}{dx_{ij}{}^3} = 0 \text{ subjected to } 0 < x_{ij} \leq x_{ij_{max}} \quad (1)$$

where, $x_{ij_{max}}$ is the maximum journey distance observed in the dataset.

As per Seneviratne (1985), the threshold value should lie near about the observed 85th percentile value. Thus the estimated root to the solution of the above equation could be validated assuming the observed 85th percentile value as a guiding value. In other studies (Larsen et. al. 2010; Hou et. al. 2012), the value at 85th percentile was also found to be considered as limiting design value. The work assumes that the observed CFD conforms to a suitable generalized polynomial function. The Eqn. 1 signifies that the 3rd order derivative of $F(x_{ij})$ needs to be a polynomial function of at-least 1st degree in order to generate a single solution of the threshold journey distance. This implies that $F(x_{ij})$ needs to be of at-least 4th degree polynomial or even higher degree function can be considered. But, these higher degree of polynomial functions will eventually yield multiple roots of solution. Then, the selection of a suitable root as a solution may be carried out by finding feasible values and also by comparing the same with the 85th percentile value of the observed CFD. In this study, the generalized polynomial function is developed using Newton's Polynomial Interpolation method (Krogh 1970). As mentioned, Newton's polynomial is usually employed to model curves having limited number of observational data points This is the process of approximating any unknown function by a suitable n^{th} degree polynomial, whose values of the dependent variable say, y and the independent variable say x are known,. The procedure for estimating of the polynomial function by the above method is described in line with the works of Krogh (1970).

The Newton interpolation polynomial of the nth degree is generated by first estimating the consecutive ordinate and abscissa differences of n number of observed data points. In this regard, the observed domain space is first uniformly discretized into n number of segments. For each of the segments, mean ordinal and abscissa values are then estimated as y_n and x_n respectively. Then, the ordinal differences are estimated as $y_1 - y_0,\ y_2 - y_1, y_3 - y_2, \dots\dots,\ y_n - y_{n-1}$; where y_n is the observed ordinate value corresponding to the abscissa x_n. If the ordinal differences are estimated at small scale,

then the above differences are denoted by $\Delta y_0, \Delta y_1, \Delta y_2, \dots \dots, \Delta y_{n-1}$ respectively, and these are called the first forward differences. Thus the first forward differences are

$$\Delta y_i = y_i - y_{i-1}, where \;\; i = 0, 1, 2, \dots, n \quad (2)$$

The corresponding abscissa points are estimated at a suitable step size h, by the following equation

$$x_i = x_0 + ih, \; i = 0, 1, 2, \dots, n. \quad (3)$$

where h is the length of each of the n number of segments of the observed domain space. If $p_n(x)$ is a polynomial of the nth degree, then as per Newton's Interpolation method, the generalized polynomial function is represented as

$$p_n(x) = y_0 + \frac{\Delta y_0}{h}(x - x_0) + \frac{\Delta^2 y_0}{h^2 2!}(x - x_0)(x - x_1) + \cdot.. + \frac{\Delta^n y_0}{h^n n!}(x - x_0)(x - x_1)(x - x_2) \dots (x - x_{n-1}) \quad (4)$$

Using the above mentioned equation, it is possible to derive polynomial functions up to any degree n. However, the decision on up to what degree of the polynomial function to be estimated is a critical issue. In such case, the trend in change of approximation error associated with each degree of the polynomial could be considered as a guidance of selection. The approximation error is a function of the degree of polynomial, which is given by Krogh (1970).

$$\varepsilon(n) = \frac{\left((x - x_0)(x - x_0 - h) \dots (x - x_0 - nh) \; f^{n+1}(x)\right)}{(n+1)!} \quad (5)$$

Where, n is the degree of polynomial up to which the polynomial is being estimated.
It is mathematically evident that as the degree of the polynomial increases, the error decreases. As mentioned previously, the CFD of the observed journey distances needs to be of at-least 4th degree or higher degree of polynomial. Therefore, the values of error are computed for various polynomials of 4th degree and higher. Once, the errors of polynomials are computed, then they are plotted to identify the point of inflection. The point of inflection, on the error vs. degree of polynomial curve, indicates a point beyond which the rate of change of error becomes negligible with the increase in degree of polynomials. In addition, the two statistics such as Akaike Information Criterion (AIC) (Akaike, 1974) and Bayesian Information Criterion (BIC) (Schwarz, 1978) are also estimated to determine the goodness-of-fit of the estimated polynomials. The AIC and BIC are an estimator for relative measure of the goodness-of-fit of the calibration models for any given set of data. Any polynomial model developed relatively with lower AIC or BIC values (Burnham and Anderson, 2004) can be recognised as relatively better model. The AIC and BIC statistics are estimated using the following sets of equations

$$\text{AIC} = 2k - 2\ln\hat{L} \quad (6)$$

$$\text{BIC} = \ln nk - 2\ln\hat{L} \quad (7)$$

In the above set of equation, $\hat{L}$ is the maximized value of the likelihood function of the model M, i.e., $\hat{L} = p(x|\hat{\theta}, M)$, where $\hat{\theta}$ is the parameter estimate that maximize the likelihood function for a vector of observed variables x, number of observations n , and k is the number of parameters of the polynomial.

Figure 5 shows variation of the approximation error as the degree of polynomial increases. The error curve is observed to get flattened beyond 6th degree of polynomials for both the cases of city bus and shared-auto. This indicates that consideration of all polynomials from 4th degree (minimum requirement) upto 6th degree would be sufficient to find out the feasible value of threshold journey distances for city bus and shared-auto

passengers. The polynomials considered under this study are also tested for their goodness-of-fit using AIC, BIC statistics.

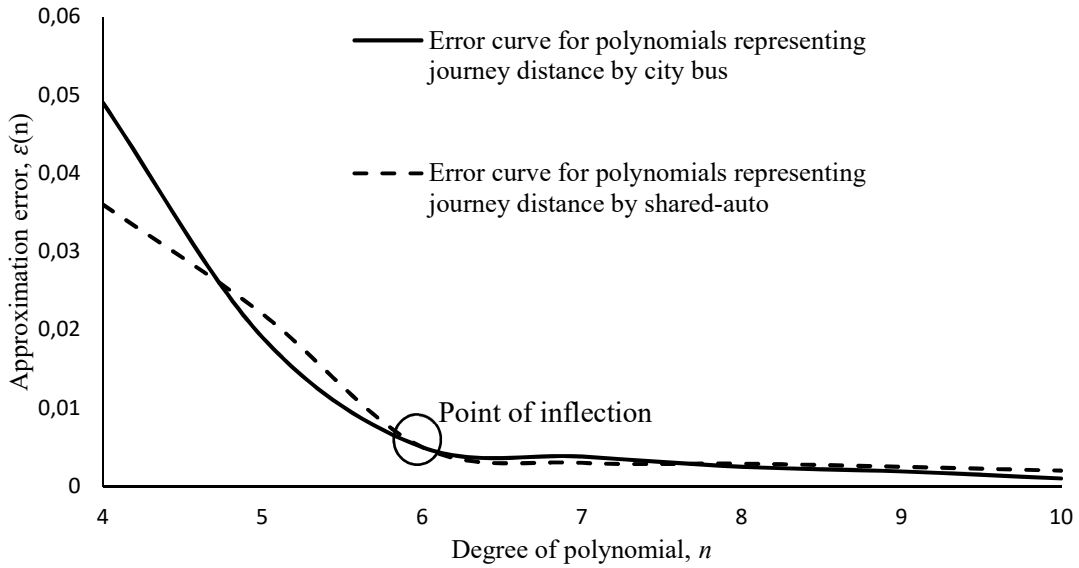

Figure 5. Variation of approximation error with degree of polynomial

The 4th , 5th , and 6th degree polynomial functions are estimated using Eq.4, and the roots of solutions for each polynomial function are estimated using Eq. 1 (refer Table 3). The Table 4 presents the approximation errors and the values of the goodness-of-fit measures for all polynomial functions under consideration for each type of transit passengers. The observed and the modelled CFD curves for journey distance data of city bus and shared-auto passengers are plotted and compared in Fig. 6 and 7 respectively. The AIC and BIC estimates for 6th degree polynomial are observed to be the least among estimates emanated from lower degree of polynomials. The 4th degree of polynomial has single root of solution for the threshold value of journey distance; whereas 5th and 6th degrees of polynomials have 2 and 3 roots respectively. If a comparison is made among all the roots of solution then, it is found that threshold values such as 12.79 kilometres. and 11.67 kilometres. for city bus passengers obtained from 5th and 6th degree of polynomials are found to be closer to the observed 85th percentile value. Similarly, threshold values such as 7.47 kilometres. and 6.79 kilometres. for shared-auto passengers obtained from 5th and 6th degree of polynomials are found to be closer to the observed 85th percentile value. But, as the statistics such as AIC and BIC estimates for 6th degree polynomial are found to be least for every case of transit passengers, so the current study considers the 6th degree of polynomial as the best fit to the observed CFD curve of journey distances. Therefore, the roots such as 11.67 kilometres and 6.79 kilometres. are finally selected as the threshold value of journey distances for city bus and shared auto passengers respectively.

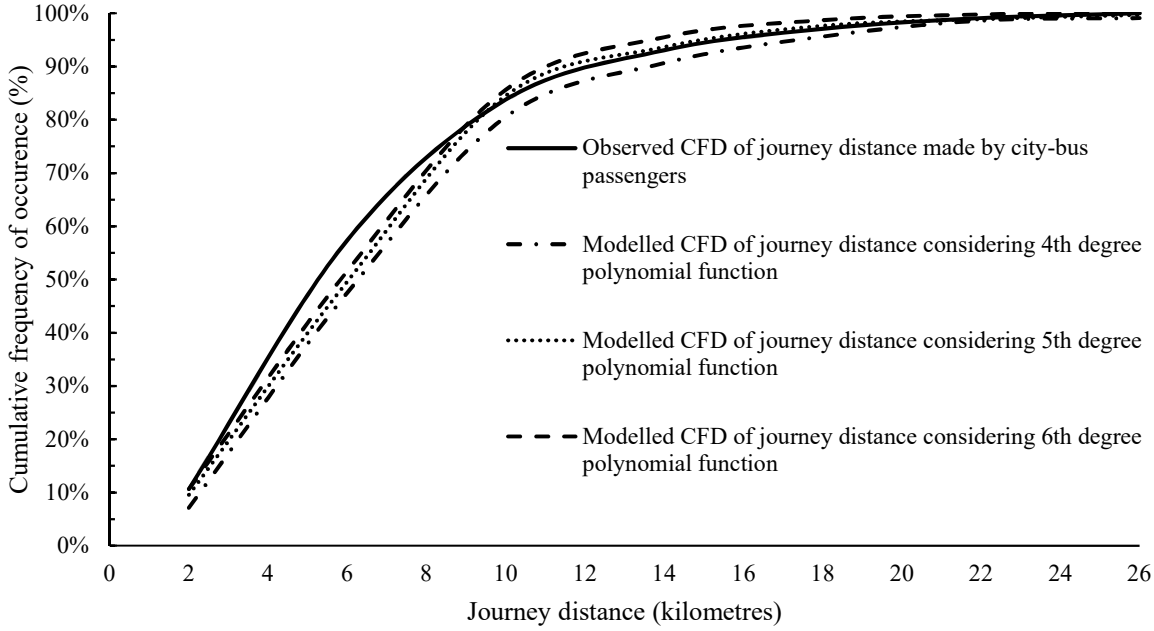

Figure 6. Fitting to the observed CFD of journey distances travelled by city-bus passengers by 4th , 5th , and 6th degree of polynomials

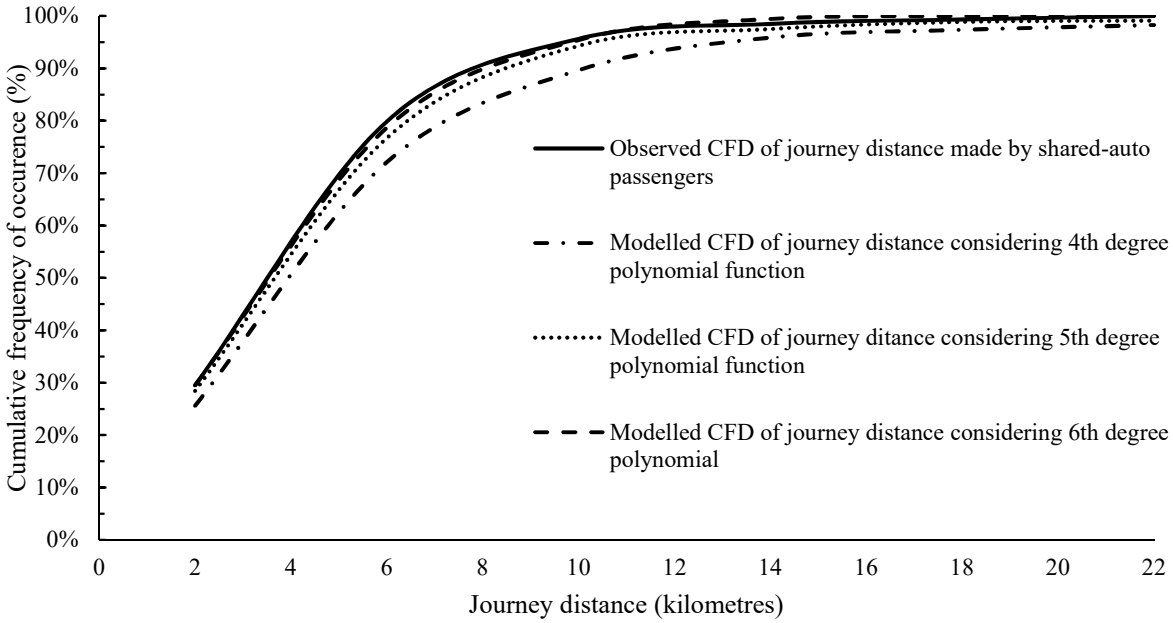

Figure 7. Fitting to the observed CFD of journey lengths travelled by shared-auto passengers by 4th , 5th , and 6th degree of polynomials

Table 3. Fitting to the CFD of observed journey lengths by various degrees of polynomials

Degree of Polynomial Function	Polynomial Function (F)	Third Order Derivative Function, $G = \frac{d^3F}{dx^3}$	1st root of $func., G$	2nd root of $func., G$	3rd root of $func., G$	85th perc as per obs. CDF
For City bus Passengers						
4th	$-2 \times 10^{-6}x^4 + 2 \times 10^{-4}x^3 - 0.01x^2 + 0.184x - 0.22 = 0$ (*Error: 0.049*)	$-48 \times 10^{-6}x + 12 \times 10^{-4} = 0$	25.55†			
5th	$1.67 \times 10^{-8}x^5 - 2.23 \times 10^{-6}x^4 + 8.23 \times 10^{-5}x^3 + 0.01x^2 + 0.036x - 0.26 = 0$ (*Error: 0.029*)	$1.002 \times 10^{-6}x^2 - 5.35 \times 10^{-5}x = 0$	12.79	40.67*		10.50
6th	$8.33 \times 10^{-9}x^6 - 1.6 \times 10^{-6}x^5 + 1.4 \times 10^{-4}x^4 - 4.5 \times 10^{-3}x^3 + 0.06x^2 + 0.39x - 0.07 = 0$ (*Error: 0.005*)	$9.99 \times 10^{-7}x^3 - 9.6 \times 10^{-5}x^2 + 3.36 \times 10^{-3}x - 0.027 = 0$	11.67	-56.47**	-58.42**	
For Shared auto Passengers						
4th	$-10^{-5}x^4 + 9 \times 10^{-4}x^3 - 0.0234x^2 + 0.267x - 0.156 = 0$ (*Error: 0.036*)	$-2.4 \times 10^{-4}x + 5.4 \times 10^{-3} = 0$	21.50†			
5th	$1.674 \times 10^{-8}x^5 - 2.26 \times 10^{-6}x^4 + 5.44 \times 10^{-5}x^3 - 0.0302x^2 + 0.3003x - 0.1962 = 0$ (*Error: 0.025*)	$1.004 \times 10^{-6}x^2 - 5.42 \times 10^{-5}x + 2.2 \times 10^{-4} = 0$	7.47	47.47*		6.50
6th	$8 \times 10^{-9}x^6 - 5.47 \times 10^{-7}x^5 - 1.2 \times 10^{-5}x^4 - 6 \times 10^{-4}x^3 - 0.028x^2 + 0.29x - 0.1864 = 0$ (*Error: 0.005*)	$9.6 \times 10^{-7}x^3 - 3.3 \times 10^{-5}x^2 - 2.9 \times 10^{-4}x - 0.0036 = 0$	6.79	37.47*	-12.56**	

* *Estimated values are infeasible as they lie beyond the domain of all observed values of journey distance*
** *Estimated values are implausible*
† *Value is significantly greater than that of 85th percentile*

Table 4. Approximation errors and goodness-of-fit measures for various degrees of polynomial

Journey-distance study	Degree of polynomial $fn.$	Approximation error, $\varepsilon(n)$	AIC	BIC
City bus passengers	4th degree	0.049	6402.87	7896.65
	5th degree	0.029	5578.87	6987.54
	6th degree	0.005	4689.54	5874.52
Shared-auto passengers	4th degree	0.036	3354.78	3567.89
	5th degree	0.025	2654.78	2687.41
	6th degree	0.005	1987.45	1897.12

The 6th degree polynomial is further used to estimate transit passenger demand in terms of passenger-kilometre travelled, and also to validate whether the passenger demand reduces sharply beyond the estimated threshold value of journey distance. In order to estimate, it is necessary to calculate the proportion of passengers travelling each segment of journey distance using the 6th degree polynomial function of the CFD. This proportion of passengers need to be multiplied with the sample observations as a part to estimate the total passenger demand. The proportion of travellers travelling for each segment (say s^{th} segment) of journey distance (say, X_s) is estimated as following.

If the proportion of transit passengers travelling by mode say k (i.e., either city bus or shared-auto) and for a distance segment, X_s is represented by $P_r(X_{sk})$, and $p_n(x)$ represents the polynomial of the CFD, then

$$P_r(X_{sk}) = p_n(X_{s+1,k}) - p_n(X_{sk}) \quad (8)$$

The sample demand for each segment of journey distance can be estimated as

$$D_{sk} = P_r(X_{sk}) \times No.of\ Sample\ Obsn.for\ mode\ k \quad (9)$$

Then, the total demand for each type of transit service is estimated by

$$TD_{sample,mode\ "k"} = \sum_{s,k} P(X_{sk}) \times No.of\ Sample\ Obsn.for\ mode\ k \quad (10)$$

Using the above Eqn. (10), the passenger-kilometre demand is estimated from the selected polynomial function. The estimated values are then compared with that of the observed data. The goodness-of-fit of the estimated demand (expressed in passenger-kilometre) is tested using χ^2 test, and is compared with the $\chi_{\text{critical}}{}^2$ values for the appropriate degrees of freedom and level of significance. The results are tabulated in Table 5.

Table 5. Comparison between the observed and the estimated passenger-demand

Journey distance study	Sample Demand (passenger-kilometres travelled)		
	Observed	6th deg. fn	$\chi_{\text{obs}}{}^2$ ($\chi_{\text{cr}}{}^2, sig.0.05, dof = (7-1) = 6$
City Bus passengers	8298	8267	14.56 (19.98)
Shared-Auto passengers	3508	3503	12.34 (13.87)

It is observed from Table 5 that the passenger-demand estimated by the 6th degree polynomial is indeed able to represent the observed passenger-kilometres demand more accurately within a level of significance of 0.05. Using this 6th degree polynomial, the change in passenger demand with each segment of journey distance is estimated and expressed by passenger-kilometres travelled. The change in passenger-demand for each successive increment of journey distance (say, between segmentX_{s+1}, and X_s) is then estimated by the following equation,

$$\Delta D_{sk} = D_{s+1,k} - D_{sk} \quad (11)$$

The variation of the change in passenger demand (i.e., ΔD_{sk}) for each successive increment of journey distance is plotted in Fig. 8.

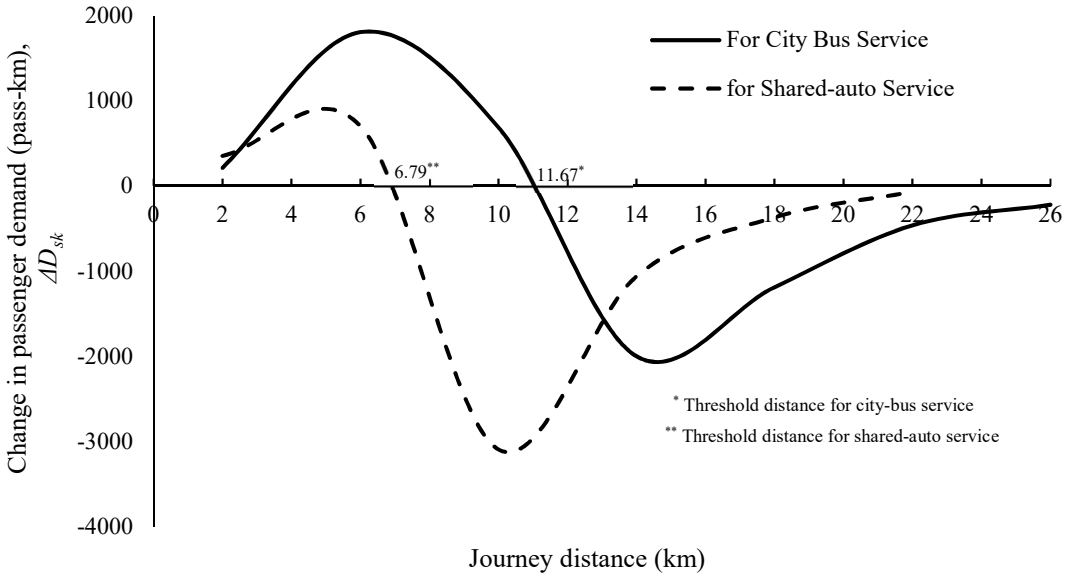

Figure 8. Variation of the change in passenger demand (ΔD_{sk}, in passenger kilometres) with each successive increment of journey distance

The above figure clearly shows that beyond the estimated threshold value of journey distance, (i.e. 11.67 kilometres. for city bus services and 6.79 kilometres. for shared-auto services), the values for change in passenger demand turns into negative. This implies that the total demand would start to drop significantly beyond the estimated value of journey distance. With regard to preparing of the length-based fare matrix, transit operators may be suggested not to create any new fare slab of journey lengths beyond such threshold value. Transit operators need to plan their revenue generation by limiting of the upper-most fare slab by considering the threshold value of journey distance. This upper-most fare slab in urban areas of developing countries is usually regulated by transit-regulatory authority. Any new slab of fare for journey lengths beyond this estimated threshold value may be kept at flat rate equalling the upper-most fare slab set out by the transit-regulatory authority. Overall, the study documents a new experience in the application of Newton's polynomial function for estimation of the threshold value of journey distance travelled by transit passengers.

4. Conclusion

The knowledge on the threshold value of journey-distance is realized to be useful to limit the upper-most slab of transit fare in a length-based fare matrix table. This value can be defined as that point on the cumulative frequency distribution (CFD) curve of journey distances at which the maximum rate of change of the slope of curve occurs. At this point, even a small change in journey distance is expected to affect large number of passenger demand opting of that transit service. In absence of the estimation of threshold value from any modelling approach, the 85th percentile value of the journey distances could have been assumed as threshold value. Seneviratne (1985) in his work mentioned that the true threshold value usually remains near about 85th percentile value of the journey distances.

The work demonstrates a novel experience to estimate the threshold value by modelling of the CFD (Cumulative frequency distribution) curve of journey distances using generalised polynomial function. A generalized polynomial function is an expression consisting of variables and coefficients, which involves simple operations of addition, subtraction, multiplication, and contains non-negative integer exponentiation of variables. In this work, the CFD curve of the journey distance values is modelled using Newton's Polynomial Interpolation method. Modelling of the CFD curve by polynomial function helps to overcome many challenges usually encountered, if any theoretical probability distribution is assumed as priori for the CFD. The current approach mentions that a polynomial of the CFD curve must be of 4 degree or higher in order to obtain the feasible root of solution. The demonstrated approach clarifies that as the degree of polynomial increases, the goodness-of-fit of the polynomial also betters.

This work finally considers the threshold value of journey distances for city bus and shared-auto passengers estimated from the 6th degree polynomial function. It is observed that the rate of change of approximation error beyond the 6 degree polynomial function becomes negligible with further increase in the degree of polynomials. The work estimates the threshold values at 11.67 kilometres. and 6.79 kilometres. for city bus and shared-auto passengers respectively. The study empirically confirms that the total passenger-demand starts to drop significantly beyond these estimated values of journey distance. The application of polynomial function in modelling of transportation engineering related problem is found to be almost sparse, and therefore the work demonstrates a new experience of the application of such function in estimating the threshold value considering journey distances data of transit passengers in Bhubaneswar.

References

Akaike, H. (1974). "A new look at the statistical model identification. IEEE Transactions on Automatic Control", 19(6), pp. 716-723.

Arasan, V. T., Rengaraju, V. R., Rao, K. K., (1994). "Characteristics of trips by foot and bicycle modes in Indian city", Journal of Transportation Engineering, 120(2), pp. 283-294.

Bandara, S., Wirasinghe, S.C., (1992). "Walking distance minimization for airport terminal configurations", Transportation Research Part A: Policy and Practice, 26 (2), pp. 59-74.

Bandegani, M. Akbarzadeh, M., (2016). "Evaluation of horizontal equity under a distance-based transit fare structure", Journal of Public Transportation, 19(3), pp.161-171.

Ben-Edigbe, J., Rahman, R. Mashros, N., (2014). "Headway distributions based on empirical Erlang and Pearson type III time methods compared". Research Journal of Applied Sciences, Engineering and Technology 7(21), pp.4410-4414.

Bhide, A., (2015). "Governance and Infrastructure Transformations in Bhubaneswar via the JNNURM". Centre for Urban Policy and Governance, School of Habitat Studies, Tata Institute of Social Sciences, Mumbai.

Bullard, R.D., Johnson, G. S., Torres, A. O., (2004). Highway Robbery: Transportation Racism & New Routes to Equity. Cambridge, MA: South End Press.

Centre for Urban Equity (CUE), (2012). "Walking and Cycling in Indian Cities: A Struggle for Reclaiming Road Edges". National Resource Centre for Ministry of Housing and Urban Poverty Alleviation, Govt. of India, CEPT University, Ahmedabad, India.

Cervero, R., Wachs, M. (2002). "An answer to the transit crisis: The case for length-based fares". Journal of Contemporary Studies, 5, pp. 59-70.

Daniels, R., Mulley, C., (2013) "Explaining walking distance to public transport: The dominance of public transport supply." Journal of Transport and Land Use, 6(2), pp. 5-20.

Fathy, M. Siyal, M.Y., (1995). "Real-time image processing approach to measure traffic queue parameters", IEE Proceedings-Vision, Image and Signal Processing, 142(5), pp.297-303.

Gray, R.H. and Sen, A.K., (1983). "Estimating gravity model parameters: A simplified approach based on the odds ratio". Transportation Research Part B: Methodological, 17(2), pp.117-131.

Harmony, X.J., (2018). "Fare Policy and Vertical Equity: The Trade-off between Affordability and Cost Recovery", Journal of Public Transportation, 21(2), pp.162-166.

Horbachov, P., Svichynskyi, S. (2018). "Theoretical substantiation of trip length distribution for home-based work trips in urban transit systems". Journal of Transport and Land Use, 11(1), pp. 593-632.

Hou, Y., Sun, C. Edara, P., (2012), "Statistical test for 85th and 15th percentile speeds with asymptotic distribution of sample quantiles". Transportation Research Record, 2279(1), pp.47-53.

Hyman, J. M., Larrouturou, B. (1982). "The numerical differentiation of discrete functions using polynomial interpolation methods". Applied Mathematics and Computation, 10, pp. 487-506.

Jin, L., Fricker, J.D., 2008. "Univariate Sensitivity and Uncertainty Analysis of Statewide Travel Demand and Land Use Models for Indiana". Transportation Research Board 87th Annual Meeting, Paper No. 08-1947, Accession Number: 01099559.

Jin, X., Zhang, Y., Wang, F., Li, L., Yao, D., Su, Y. Wei, Z., (2009). "Departure headways at signalized intersections: A log-normal distribution model approach". Transportation Research Part C: Emerging Technologies, 17(3), pp.318-327.

Kim, G., IM, H., Kim, W., Kang, K. (2003). "Analysis of trip length distribution between commodity-based model and truck trip-based model in Seoul metropolitan area". Proceedings of the Eastern Asia Society for Transportation Studies, 4, pp. 1242-1255.

Krogh, F.T., (1970). "Efficient algorithms for polynomial interpolation and numerical differentiation". Mathematics of Computation, 24(109), pp.185-190.

Lamont, M., (2009). "Reinventing the wheel: A definitional discussion of bicycle tourism". Journal of Sport and Tourism, 14(1), pp.5-23.

Larsen, J., El-Geneidy, A. Yasmin, F., (2010). "Beyond the quarter mile: Re-examining travel lengths by active transportation". Canadian Journal of Urban Research, 19(1), pp.70-88.

Liang, X., Zhao, J., Dong, L. and Xu, K., (2013). "Unraveling the origin of exponential law in intra-urban human mobility". Scientific Reports, 3(1), pp.1-7.

Lupo, T., (2013). "Strategic analysis of transit service quality using fuzzy AHP methodology", European Transport/Trasporti Europei, 53, pp.1-18.

Ministry of Finance (MOF), (2015). "Re-classification of Cities/Town on the basis of Census 2011 for the purpose of grant of House Rent Allowance to Central Govt. employees". Government of India, New Delhi, India. < https://doe.gov.in/sites/default/files/21-07-2015.pdf, accessed on 12/03/2020>

Ministry of Home Affairs (MHA), (2011). "Census of India 2011." Government of India, New Delhi, India

Moore, G.R., (2002). "Transit ridership efficiency as a function of fares". Journal of Public Transportation, 5(1), pp 105-131.

Narboneta, C., Teknomo, K. (2016). "A Study of Metro Manila's Public Transportation Sector: Implementing a Multimodal Public Transportation Route Planner". Asian Transport Studies, 4(2), pp. 460-477.

Richards, D., (1984). "Hyperspherical models, fractional derivatives and exponential distributions on matrix spaces". Sankhyā: The Indian Journal of Statistics, Series A, pp.155-165.

Roy, S. Basu, D., (2021). "An approach towards estimating critical value of waiting time at transit stops". Journal of Traffic and Transportation Engineering (English Edition), 8(2), pp. 257-266.

Rubin, S.G. Khosla, P.K., (1977). "Polynomial interpolation methods for viscous flow calculations". Journal of Computational Physics, 24(3), pp.217-244.

Schwarz, G. (1978). "Estimating the dimension of a model". The annals of Statistics, pp. 461-464.

Seneviratne, P.N., (1985). "Acceptable walking lengths in central areas". Journal of Transportation Engineering, 111(4), pp. 365-376.

Singh, S.K., (2012). "Urban Transport in India: Issues, Challenges, and the Way Forward", European Transport/Trasporti Europei, 52, pp.27-52.

US Department of Defence, (2020). Global Positioning System Standard Positioning Service Performance Standard, 5th ed., Washington D.C. < https://www.gps.gov/technical/ps/2020-SPS-performance-standard.pdf, accessed on 19/11/2020>

Verma, A., Verma, M., Rahul, T. M., Khurana, S., Rai, A., (2018). "Acceptable trip distance for walking in mass religious gatherings—A case study of world's largest human gathering Kumbh Mela in Ujjain, India", Sustainable Cities and Society, 41, pp. 505-512.

Werner, W., 1984. "Polynomial interpolation: Lagrange versus newton". Mathematics of Computation, pp.205-217.

Yabe, T., Aoki, T. (1991). "A universal solver for hyperbolic equations by cubic-polynomial interpolation I. One-dimensional solver". Computer Physics Communications, 66(2-3), pp. 219-232.

Yook, D. Heaslip, K., (2015). "Effective Modeling for a Length-Based Fare Structure with a Time-Expanded Network". Journal of Public Transportation, 18(1), pp.22-29.

Zhang, C., Cao, X., Nagpure, A. Agarwal, S., (2019). "Exploring rider satisfaction with transit service in Indore, India: An application of the three-factor theory". Transportation Letters, 11(8), pp.469-477.

Acknowledgment

The database of the work is developed using project grant under the **IMP**acting **R**esearch, **IN**novation and **T**echnology (**IMPRINT**) -India initiative (Project code # 7094). The grant is jointly supported by the Ministry of Human Resource Development (MHRD) curremtly known as Ministry of Education (MoE), GoI and the Ministry of Housing and Urban Affairs (MoHUA), GoI. The authors convey their sincere thanks to the Ministries for their kind support and encouragement.