IJTE, Vol. 11, No. 2, Autumn 2023, pp. 1419-1441
Research Paper

Random Effect Models to Predict Operating Speed on Horizontal Curves for Intercity Buses Based on Naturalistic Driving Data

Amir Hossein Sheikh Azami[1,*], Omid Titidezh[2], Ayyoob Jafari[3], Mohammad Soleimani Varaki[4]

Received: 2022/09/09 *Accepted: 2023/07/10*

Abstract

The most common cause of death and injury worldwide is traffic accidents. Several studies have shown that these accidents occur primarily on rural roads, especially those with horizontal curves. Studies have indicated that inconsistency in geometric design is one of the main reasons for this problem. Road geometric consistency refers to the alignment of the road geometry with the expectations of the driver. In evaluating road design consistency, operating speed is the most common criterion, which can be estimated by using operating speed models. Buses account for the majority of passenger movement, so it is important to study their operating speed. In this work, the random effect model is used for 150 buses passing through 690 horizontal curves of the Tehran-Qaemshahr highway to estimate the performance speed in horizontal curves. Among all the highways connecting the northern area of the country to the capital, that expressway was chosen for the study because of many reasons such as the minimum changes in the geometry, high volume of daily bus traffic, various topographic conditions, and the availability of data needed for the study. Besides, the variables of driving time, the geometric design of the curve, the driver's data, and the characteristics of the bus are examined in this study. Finally, the developed model revealed that the average radius of the curve, the volume of traffic, the reverse curves, the percentage of heavy vehicles, the location of fixed speed cameras, the posted speed limit, and the darkness were all significant factors.

Keywords: Operating speed models; Random Effect Models; Intercity Buses; Four-lane Divided Rural Highway; Horizontal curves

---

* Corresponding author, Email: amirhosseinsheikhazami@yahoo.com

[1] PhD Candidate, Faculty of Engineering and Technology, Shomal University, Mazandaran, Iran

[2] Faculty member, Faculty of Engineering and Technology, Shomal University, Mazandaran, Iran

[3] Faculty Member, Faculty of Electricity and Computer, Islamic Azad University, Qazvin, Iran

[4] Faculty Member, Department of Basic Science, Islamic Azad University, Mazandaran, Iran

1. Introduction

Studies indicated that increasing the speed of vehicles is associated with increasing the severity or probability of accidents [Gargoum and El-Basyouny, 2016; Mirbaha et al., 2013]. The majority of fatalities occur on rural roads and in horizontal curves [Fitzsimmons et al., 2013b and Campbell et al., 2008]. The operating speed of vehicles has been used as an alternative criterion for evaluating the safety of horizontal curves by researchers [Jacob and Anjaneyulu, 2013]. Operating speed is an indicator of how drivers behave and their level of comfort and safety on different roads. The operating speed represents three major components of driving: the road, the driver, and the vehicle. Inconsistency among these components adversely affects road safety. According to previous studies, operating speed has become an interesting topic for road designers. According to AASHTO design guidelines, the operating speed is the speed at which drivers choose to travel in free-flow traffic [Aashto, 2001].In free traffic conditions, 85% of drivers move at this speed or below. The operating speed reflects the speed behavior of drivers affected by horizontal and vertical alignments as well as cross-sectional levels. As a result, the researchers collect data on the operating speeds of isolated vehicles traveling in free flow. Free-flow speed has a normal distribution, as shown in many studies. To assess whether speed design consistency has been achieved, engineers calculate the 85th percentile of the distribution (V85), separating the population of cautious drivers from a small group of more aggressive drivers. To evaluate road safety, many studies have focused on determining the operation speed and considering road geometric adaptation methods. So naturally, the operating speed was necessary to that evaluation. In those studies, operating speed was calculated under free flow conditions, which dictates calculations under ideal weather conditions, land use, and traffic conditions. This means that only geometric characteristics were effective in estimating operating speed. The use of Naturalistic Deriving (ND) data allows us to build a model for estimating operating speeds under a variety of conditions, especially in the case of free traffic flow.

[Sil et al., 2019] developed a model for vehicle speed prediction through vehicle speed and road geometrical data obtained from isolated four-lane highways to plan and improve the safety of the road network and convert two-lane roads into four-lane highways. However, because of the absence of opposite direction flow, greater road width incurves, and more space for overtaking slower vehicles, the speed-influencing components in isolated four-lane rural highways are expectedly different from the rural two-lane roads [Nama et al., 2016].

Most of the existing operating speed prediction models are limited to passenger cars [Misaghi and Hassan, 2005; Fitzpatrick et al., 2000]. In spite of the considerable importance of bus safety as a means of public transportation with a high capacity for passengers, relatively few studies have been conducted on the geometric consistency of this vehicle class. To the best of the authors' knowledge, studies on performance speed determination have been developed for only two classes: passenger cars and trucks. Also, according to our understanding, only Jacob, in her study, calculated the performance speed for buses besides passenger cars and trucks.

To investigate the effect of various factors on the operating speed of buses in a road network, particularly on four-lane arterial roads separated by a median, extensive studies are necessary. There are several models of operating speed developed in previous studies. Most of those models apply to two-lane rural roads, however, the model format, independent variables, and regression coefficients are, in most cases, substantially different from one model to the other. There may have been differences in driver behavior between localities, and this highlights that no single model is universally accepted [Misaghi and Hassan, 2005].

In this study, we try to identify and evaluate all the variables affecting the operating speed. Then, among the identified effective variables, significant variables are included in the model so that the developed model would be complete and comprehensive for predicting the operating speed. Thus, to forecast the operating speed on this type of road, a random-effect linear regression model is proposed for curves, which takes into account road geometry data, drivers' natural behavior data, traffic data, and other influential factors along the route, such as the locations of fixed speed cameras and speed limit signs. As part of the formulation structure of random effect models, multiple observations are taken into account in randomly selected areas of four-lane rural roads.

The rest of this paper is structured as follows. Section 2 reviews the relevant literature. Sections 3 and 4 present the data collection process and the theoretical foundations of the regression model, respectively. Section 5 presents the modeling results.

2. Literature Review

Studies have been conducted on operating speed estimation because operating speed is an important factor in determining the quality and efficiency of a route, designing geometric components of the road, and ensuring road safety. Most of these researchers have widely studied the effect of geometric road features on vehicle speed for two-lane rural roads. Operating speed estimation studies can be categorized according to factors such as the types of subjects, the types of information, the models used, the types of vehicles, roadside usage, and others. Table 1 summarizes the method in some research on operating speed.

Table 1. A summary of some methods in some research on operating speed

The writer Country/Year	Research Objective	Type of Road	Type of Vehicles	Type of information used	Method of data collection	Sites conditions	Statistical analysis method
Lamm USA/1988	Measuring the consistency of horizontal design as defined by operating speed and accidents expected	Two-Lane Rural Roads	Passenger cars and trucks	Geometric and traffic volume data	Geometric Design: - Traffic volume: -	AADT range: 400 to 5,000 vehicles per day	Linear regression
Passeti USA/1999	Investigating the effect of spiral transition curves on the speed at which vehicles traverse a horizontal curve	Two-Lane Rural Roads	Passenger cars	Geometric and Speed data	Geometric Design: - Speed data: equipment radar or laser gun	Free-flowing traffic conditions	Linear regression
Fitzpatrick USA/2000	Models for 85th percentile speed of horizontal and vertical alinement	Two-Lane Rural Roads	Passenger cars	Geometric and Speed data	Geometric Design: - Speed data: equipment radar or laser gun	Free-flowing traffic conditions	Linear regression
Gong USA/2008	Prediction of operating speed in horizontal curves	Four-Lane Rural Roads	Passenger cars	Geometric and Speed data	Geometric design: - measuring in ArcMap Speed data: equipment radar or laser gun	Free-flowing traffic conditions	Linear regression and multiple linear regression
Wang China/2010	Relationship between Operating Speed and Posted Speed Limit	Mountainous Freeway	Passenger cars and trucks	Speed and Posted speed limit data	Speed data: using a MetroCount 5600 Posted Speed Limit: -	Free-flowing traffic conditions	Linear regression
Khairi Malaysia/2011	Develop the 85th percentile operating speed model at mid-curve	Two-Lane Rural Roads	Passenger cars	Geometric and Speed data	Geometric design: the respective Public Works Department of Malaysia (PWD) officials Speed data: equipment radar or laser gun	Free-flowing traffic conditions	Multiple linear regression
Semeida Egypt/2012	Impact of highway geometry and posted speed on operating speed	Multi-lane highways	Passenger cars	Geometric and Speed data	Geometric design: is collected directly from the site Speed data: equipment radar or laser gun	Free-flowing traffic conditions	Linear regression And Artificial

The writer Country/Year	Research Objective	Type of Road	Type of Vehicles	Type of information used	Method of data collection	Sites conditions	Statistical analysis method
							Neural Network
Jacob India/2013	Develop operating speed and speed reduction models for vehicles at tangent and mid-curve	Two-Lane Rural Roads	Different classes of vehicles	Geometric and Speed data	Geometric Design: - Speed data: vehicles equipped with GPS	Free-flowing traffic conditions	Multiple linear regression
Bella Italy/2014	Effect of light on operating speed using a driving simulator	Two-Lane Rural Roads	Passenger cars	Geometric and Speed data	Geometric Design: - Speed data: vehicles equipped with GPS	Free-flowing traffic conditions	Multiple linear regression
Borujerdian Iran/2016	Analysis of geometric design impacts on vehicles operating speed	Two-Lane Rural Roads	Passenger cars and trucks	Geometric and Speed data	Geometric design: is collected directly from the site Speed data: video recording	Free-flowing traffic	Linear
Maji India /2018	study the effect of rural four-lane divided highway geometry on operating and design speed of passenger car, SUV, LCV, and HCV	Four-Lane Divided Rural Highways	Different classes of vehicles	Geometric, Speed and Traffic data	Geometric design: obtained from the as-built plan and profile drawings provided by the National Highways Authority of India. Speed data: Video cameras installed at vantage points	Free-flowing traffic conditions	Multiple linear regression
Bamdad Mehrabani Iran/2019	Develop passenger car V85 prediction model for four-lane divided horizontal curves	Four-lane rural highways	Passenger cars	Geometric and Speed data	Geometric design: using Google Earth photos and video images and a GPS device. Speed data: speed gun	Free-flowing traffic conditions	Multiple linear regression
Sil India /2019	Develop passenger car V85 prediction model for four-lane divided horizontal curves	Four-lane divided highways	Passenger cars	Geometric and Speed data	Geometric design: from the National Highways Authority of India (NHAI) as plan and profile drawings.	Free-flowing traffic conditions	Multiple linear regression

The writer Country/Year	Research Objective	Type of Road	Type of Vehicles	Type of information used	Method of data collection	Sites conditions	Statistical analysis method
					Speed data: synchronized video cameras		
Gao China/2020	Verify the impact of large vehicles on the traffic flow by using field tests	Suburban highway with two lanes	Different classes of vehicles	Traffic flow, Speed and Vehicle type data	Traffic flow: laser roadside traffic survey instrument Speed: are collected directly from the site Vehicle type: are collected directly from the site	Free-flowing traffic conditions	Linear regression
JHS India/2021	Operating speed prediction model of the combined curve (combination of horizontal and sag vertical curve)	Two-Lane Rural Roads	Different classes of vehicles	Traffic, Geometric and Crash data	Traffic: site using TIRTL (Transportable Infra-Red Traffic Logger) Geometric: using Total station Vehicle type: police stations	Free-flowing traffic conditions	multiple linear regression
Malaghan Indiana/2022	Develop operating speed and speed reduction models for heavy passenger vehicles.	Two-Lane Rural Roads	Heavy passenger vehicles	Geometric and Speed	Geometric design: is collected directly from the site Speed data: GPS device which was placed near the dashboard of the vehicle	Enabled to maintain free-flow conditions	OLS (ordinary least square) regression

According to Table 1, most studies used traditional and on-the-field methods to collect data. Geometrical and vehicle speed data have been collected using a variety of methods and tools. Speed measurement is mainly done by laser gun, radar gun, LIDAR gun, and stopwatch. The use of a radar gun is usually subject to human error since it requires at least two people to aim and read the speed number. The use of speed detection tools that may attract the driver's attention may also affect the driver's behavior in some cases. Even though the detected speed was recorded at a specific and predetermined location, it is usually used for modeling the operating speed for the entire road segment. Therefore, because of the two variables, road conditions and driver behavior, the speed of a point is not representative of the speed of a specific segment [Nie & Hassan, 2007]. Previous studies have gathered the geometric components of the route through maps or field visits and measurements. Due to the low scale of the maps, one of the most significant weaknesses of these measurements is their low accuracy. A field measurement error also makes it difficult to determine the exact location of curve elements. Among the variables investigated in each of the studies, Table 2 shows the influential variables briefly.

Table 2. Summary of influential variables in each of the mentioned studies

The writer Country/Year	Number of Lane	Lane width	Shoulder width	Median width	Curve length	Curve radius	Degree of curvature	Vertical curves	Clothoide	Deflection angle	Superelevation	Slope	Type of surface	Type of median	Type of shoulder	Operating speed in the previous part	Previous geometric component	previous curve radius	Next one curve radius	Tangent length	Speed limit	Percentage of heavy equipment	Traffic volume	Land use
Lamm USA/1988	-	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-
Passeti USA/1999	-	-	-	-	✓	✓	✓	-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
Fitzpatrick USA/2000	-	-	-	-	-	-	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Gong USA/2008	-	-	-	-	✓	-	-	-	-	-	-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-
Wang China/2010	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-
Khairi Malaysia/2011	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
Semeida Egypt/2012	-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-
Jacob India/2013	-	-	-	-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-
Bella Italy/2014	-	-	-	-	-	-	✓	-	-	✓	-	-	-	-	-	-	-	✓	-	✓	-	-	-	-
Borujerdian	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	✓	-	-	-	-	-	-	-

The writer Country/Year	Number of Lane	Lane width	Shoulder width	Median width	Curve length	Curve radius	Degree of curvature	Vertical curves	Clothoide	Deflection angle	Superelevation	Slope	Type of surface	Type of median	Type of shoulder	Operating speed in the previous part	Previous geometric component	previous curve radius	Next one curve radius	Tangent length	Speed limit	Percentage of heavy equipment	Traffic volume	Land use
Iran/2016																								
Maji India /2018	-	-	-	-	✓	✓	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-	-
Mehrabani Iran/2019	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	✓
Sil India /2019	✓	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-
Gao China/2020	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	✓	-
JHS India/2021	-	-	-	-	✓	✓	✓	✓	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Malaghan Indiana/2022	-	-	-	-	-	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-

As you can see in Table 2, the researchers investigated the effect of newer variables over time. Most of the studies introduced curve radius, curve length, and curvature of the arch as variables influencing operating speed. Therefore, more variables have been collected and analyzed in this research to closely examine the factors affecting the operating speed. In addition, new methods have been used to collect the data in order to increase the accuracy of the databases, as well as more sophisticated and accurate equipment.

3. Methodology

Four rural transportation routes between Tehran and Mazandaran provinces in Iran were selected for this case study. Tehran-Qaemshahr highway, known as Firuzkuh road, is a two-directional four-lane road (two lanes in each direction) with a topographical variety of plains, hills, and mountains. The total length of the highway is 504 km (in both directions), and the width of each lane is fixed and equal to 3.65 meters alongside the entire road. Moreover, there were different types of usages (residential, commercial, etc.) along the entire road length, which is the reason there were different posted speed limits at different points. In the following sub-sections, we will explain the data and the modeling approach.

3.1. Data Description

There are two valuable tools for improving roadway safety and efficiency: Naturalistic Driving (ND) and the Roadway Information Database (RID). Using ND in traffic safety research can be defined as studies undertaken to provide insight into driver behavior during everyday trips by recording details of the driver, the vehicle, and the surroundings through unobtrusive data-gathering equipment and without experimental control [Van Schagen et al., 2011]. On the other hand, the RID data includes information about the physical characteristics of the roadways whereas ND provides information on vehicle trips performed on the roadways. Therefore, when ND is coupled with the RID, the ND data can explain drivers' behavior within a roadway environment. In this way, linking the ND and RID data permits us to observe and analyze the interrelationships between the driver, the vehicle, the road, and other vehicles in normal conditions, conflict situations, and real accidents [Cruzado, 2009].

3.2. Data Collection

By using ArcGIS software, the ND and RID data were linked, resulting in a map with information from both databases overlaid. This new shapefile, containing both sets of data, can be utilized for several potential safety studies [Ishak et al., 2016]. Also, traffic data, meteorological data, technical specifications of buses, locations of posted speed limit signs, and fixed speed control cameras with RID and ND data along the Firuzkuh road for 2020 have been collected and analyzed. As a road feature such as slope has different effects on operating speed in opposite directions, the gathered data were then separated according to driving direction, as shown in figure 1. Consequently, all the influencing elements of the two directions of the movement were considered.

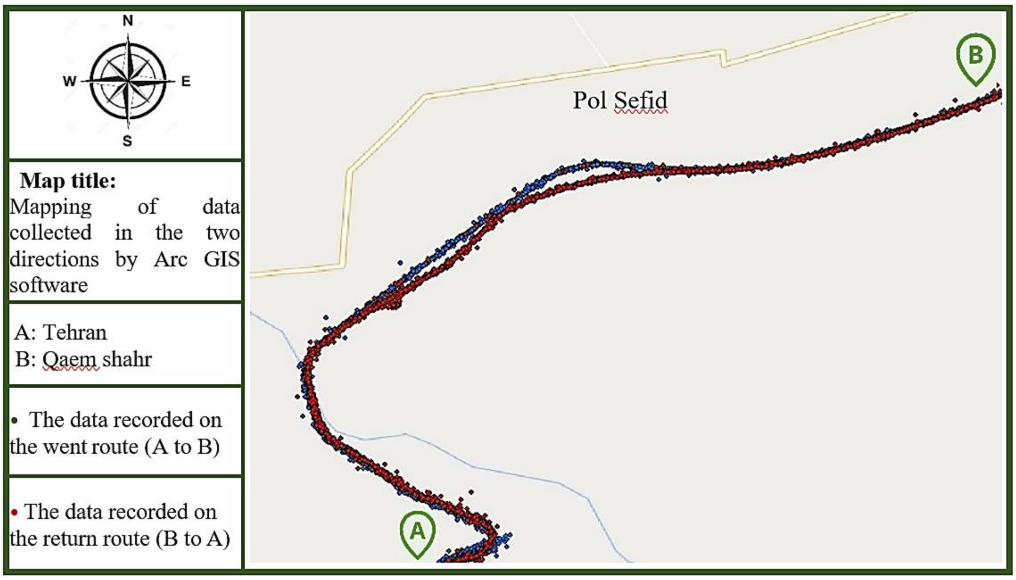

Fig 1. Mapping of data collected in the two directions by Arc GIS software

The different parts of the database are explained below:

1. The RID data was collected by a Road Surface Profiler (RSP) machine along the route. The RSP machines have road scanning equipment and imaging cameras that accurately collect road geometrical data. Also, the road geometrical data that were chosen to enter the analysis process include curve radius, length, direction, longitudinal slope, transverse slope, and design speed. The width of each lane along the route was about 3.65 meters. Driving behavior is influenced by the curves selected up to a radius of 1000 meters. Therefore, 223 curves were collected in one direction and 240 curves in the other direction. Figure 2 shows the location of the selected curves.

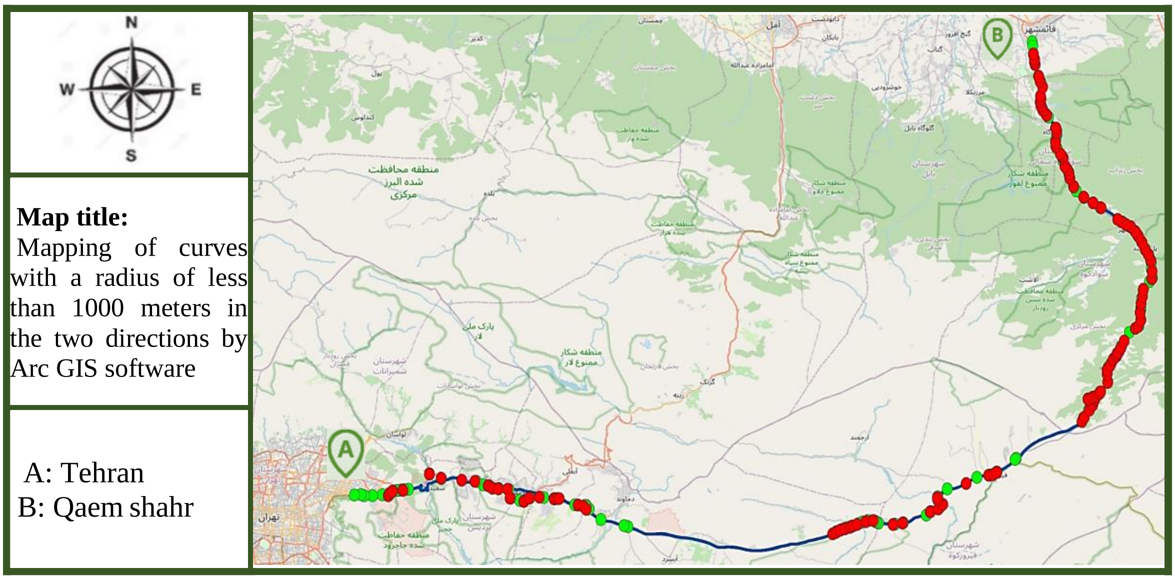

Fig 2. Mapping of curves with a radius of less than 1000 meters in the two directions by Arc GIS software

2. In this study, buses are equipped with Fleet Management Systems (FMS). The FMS had a GPS device that was placed near the dashboard of the vehicle, a camera fixed to the windshield, and an antenna fixed to the roof for high positioning accuracy by combining signals from different satellites (see Figure 3).

ND Data includes (speed, acceleration/deceleration, direction, mileage) were continuously and imperceptibly recorded by these devices for several months to several years. There were no experimental controls on the driving condition.

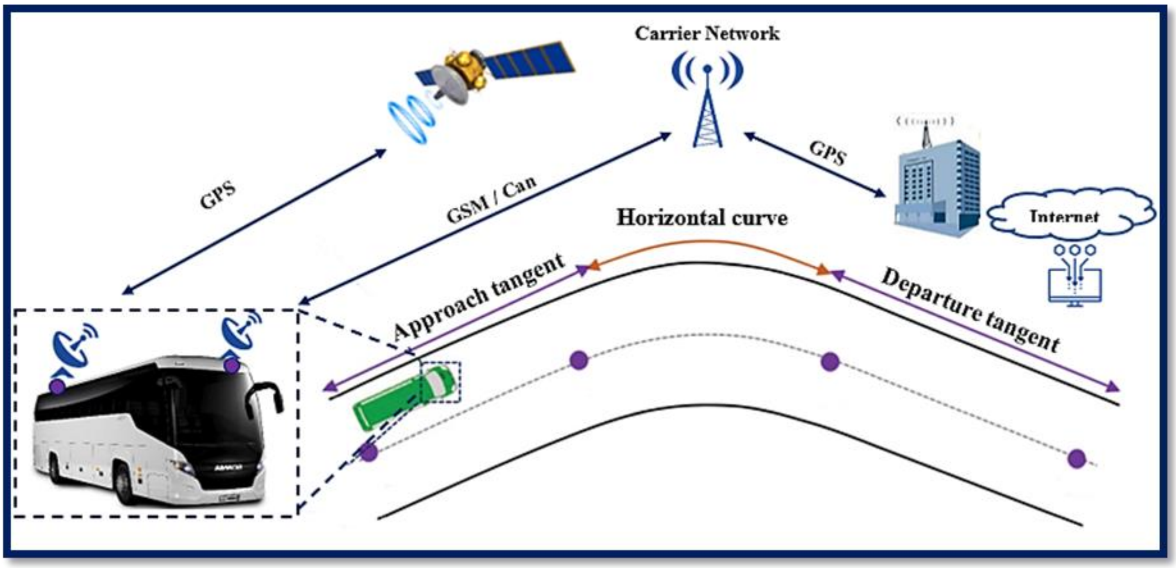

Fig 3. Diagrammatic representation of FMS device arrangement and road geometric elements in the direction of travel

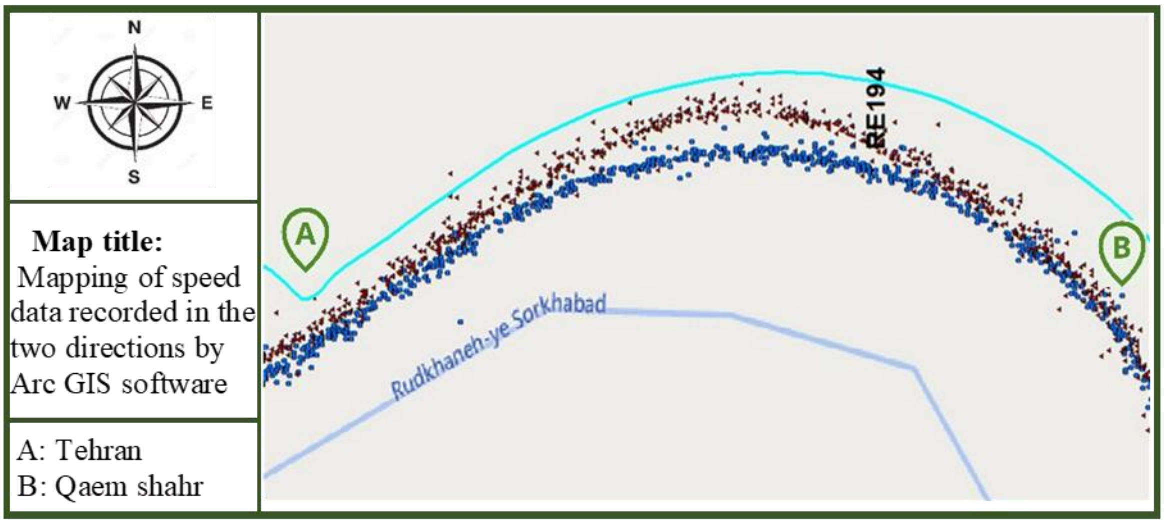

Fig 4. Mapping of speed data recorded in the two directions by Arc GIS software

One of the most important data in this section is the speed of buses during the study, in which the lower and upper limits of speed observations were 11 to 115 km/h, respectively.

3. As the parameters affecting drivers' choices of speed and behavior, traffic data including traffic volume, the percentage of heavy vehicles, the average speed of automobiles, etc., are of great importance. Therefore, that dataset was extracted from the SANET system database of the Iran Ministry of Roads and Urban Development for 2019.

4. Meteorological information was prepared from the National Meteorological Organization for four road weather stations of Imamzadeh Hashem, Amin Abad, Shurab, and Gaduk along the route for the study's duration.

5. To check the technical specifications, all the buses were classified into three groups: Maral, Dorsa, and 4212. They differed only in

terms of the type of cabin design and other comforting facilities and were not significantly different in terms of parameters affecting speed, such as engine power, cabin weight, and safety systems. The only specific parameter was the mileage of buses used in the model to determine the wearing out.

6. The Fix Speed camera location was gathered in a field visit in both directions of the route.

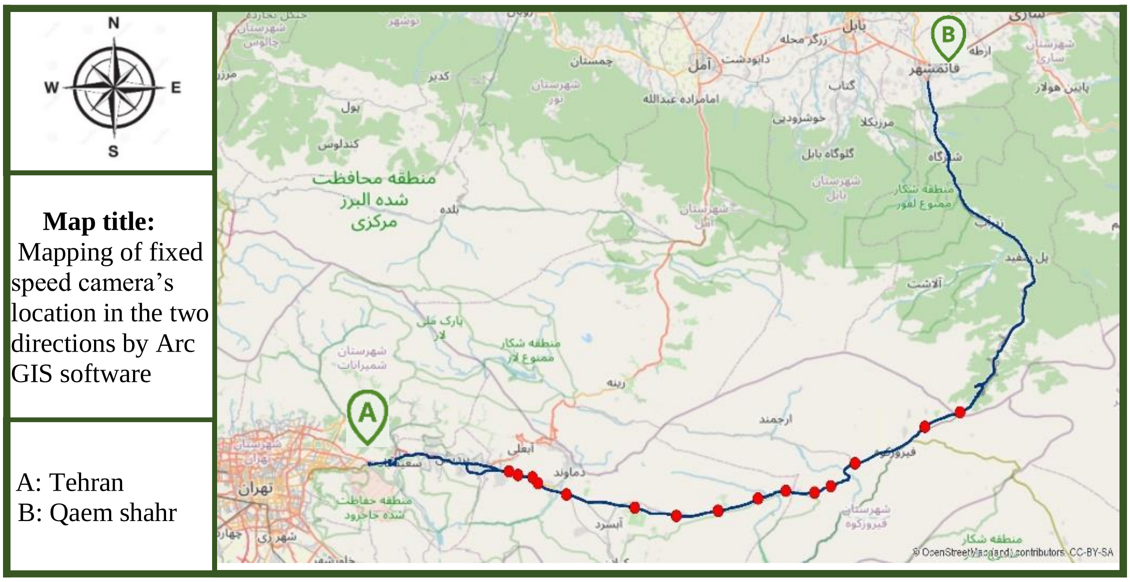

Fig 5. Mapping of fixed speed camera's location in the two directions by Arc GIS software

7. The locations of speed limit signs posted in both directions of the road have also been recorded. For the purpose of investigating the effect of these signs on drivers' operating speeds, a range of 500 meters ahead of the signs was determined as the point where the driver saw the sign and read its message.

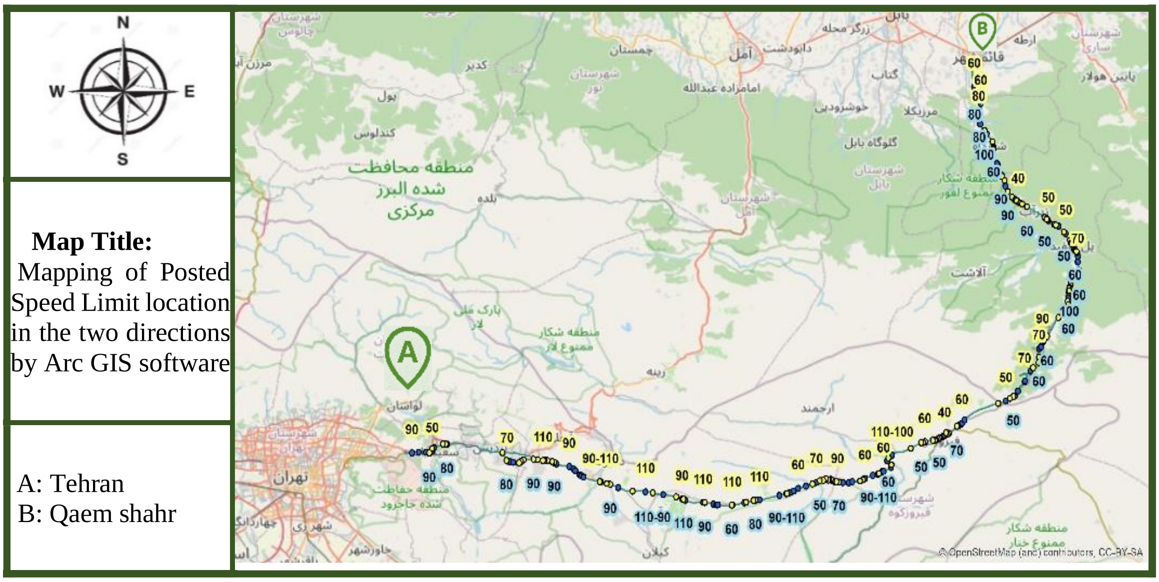

Fig 6. Mapping of Posted Speed Limit location in the two directions by Arc GIS software

In the next step, removing incorrect data and cleaning the data was done to link databases to each other and enter the model. By examining the databases, the outlier data were removed using Quantile Layers test. Since a large part of the meteorological data has not been fully recorded because of the technical defects of the stations' equipment and only the data of one month was completely and continuously available, the decision fell not to enter this type of data at all into the modeling. Also, in many cases of examining the driver’s information data, different drivers drive the same bus on each trip. Due to this, it was impossible to assess the impact of the individual characteristics of a specific driver on the selected speed at various stages along the route. It was also avoided entering this part of the data into the modeling.

3.3. Modeling Approach

The authors of this research adopted the structure of the model proposed by [Bassani et al., 2016] used for rural two-lane roads to evaluate the operating speed. This structure separates the central tendency index from the index of dispersion for individual speeds, which is a function of driving skills and drivers' decisions.

3.4. Random Effect Model

Data collected include multiple measurements along the route according to the previously described method. Multiple observations were, in fact, available for the entire length of the route in both directions. Hence random effects (RE), which properly account for the sampling design, were included in the model to remove any dependence between estimation errors of individual observations. Based on the background, previous researchers mainly used linear regression models to predict and select operating speed. Only in one case [Bassani et al., 2016] random effects model were used. The random effects model is a hierarchical linear statistical model in which the model parameters are random variables. It assumes that the analyzed data is drawn from a hierarchy of different populations, and the differences are related to the hierarchy.

Random effects are not directly estimated (although they are obtainable after estimation) but are summarized concerning their estimated variances and co-variances. The random effects may be random coefficients, and the grouping structure of the data may comprise multiple levels of nested groups. Random effect models evaluate any difference between the speed predictions for curves and drivers and the corresponding predictions for a specific curve and driver. As assumed, this observed heterogeneity follows a normal distribution function:

$$\alpha \sim N(0.\sigma^2) \qquad (1)$$

The dependent variable ($V_{mn,i}$), which represents the observed speed (i) for a driver (m) in a curve (n), is then derived from an RE model.

$$V_{mn,i} = \beta o + \sum_{k=1}^{K} \beta_K^C \,.X_{ki} \; + \sum_{n} \alpha \; + \epsilon_{mn,i} \qquad (2)$$

Where βo is the constant value of the model, β_K^C is the calibration parameter for the variables affecting the estimated mean (X_k). Subscript K indicates the number of significant variables affecting the dependent variable (operating speed of buses) (X_k). X_{ki} is the independent variable of the model. $\sum_n \alpha$ is the sum of normally distributed random effects (α) and $\epsilon_{mn,i}$ is the error of each measurement, where (m) is the driver index and (n) is the curve index.

If the random effects are removed from Equation (2), this model becomes a fixed effect (FE) model, widely used in the literature on modelling operating speeds. Unlike the RE model, the FE model does not properly account for the sampling design of this research when multiple measurements are included in the database. Such observations have unknown effects, so it is impossible to suppose the independence of error for individual observations.

3.5. Variable Selection

The Bayesian Information Criterion (BIC) hypothesized by [Schwarz, 1978] was usable for selecting variables that significantly affect driver's speed behavior. The researchers prioritized the model with the lowest function BIC (f_BIC) value, calculated according to equation (3).

$$f_{BIC} = -2.L^{\cap} + k.ln(n) \qquad (3)$$

Where Lˆ is equal to the maximum value of the log-likelihood function, n is equal to the number of observations, and k is equal to the number of parameters in the model. According to the structure of equation (2), the number of parameters is equal to the sum of the coefficients β0 and βC. Only the variables that contributed to minimizing the BIC function were selected and included in the model.

3.6. Model Comparison

Model comparison is a very common problem, we often have multiple competing hypotheses about how our data were generated and we want to see which model is best supported by the evidence. If we can describe our data generating process explicitly as a set of deterministic and stochastic components, then we can use likelihood-based methods (e.g., LRT, AIC, BIC, Bayesian model selection) to infer which data generating.

The Akaike Information Criterion (AIC) measures how much information is lost by a model in order to determine which model is most appropriate. In this way, AIC provides a balance between the number of model parameters and the fit of a model to the data. Considering this issue, it can be said that a model that is recognized as a suitable model by AIC does not have overfitting, nor does it suffer from underfitting, and it can be considered a model with a suitable fit. The AIC evaluation criterion can be considered as an estimator of the relative quality of the statistical model according to the data collected in the random sample. In this way, with the assistance of the AIC evaluation criterion, an ordered relation between the models can be obtained to compare and measure the superiority between them. Also, according to the Akaike information criterion, the most suitable model has the lowest AIC value.

The Deviance Information Criterion (DIC) is a hierarchical modeling generalization of the Akaike information criterion (AIC). DIC is an asymptotic approximation as the sample size becomes large, like AIC. It is only valid when the posterior distribution is approximately multivariate normal.

4. Results and Discussion

4.1. Description of Data

Different databases were subjected to the chi-squared test (χ2 test) and the Kolmogorov–Smirnov test (K-S or KS test) to examine the normal distribution of the data. In all cases, the assumption of non-normality of the data was rejected with a significance level of 95%. In other words, goodness-of-fit tests indicated that the speed data was sufficient in each direction to form a normal distribution.

The following graphs indicate the distribution of recorded speeds, the percentage of heavy vehicles, regional speed limits, traffic volume, and the radius of curves:

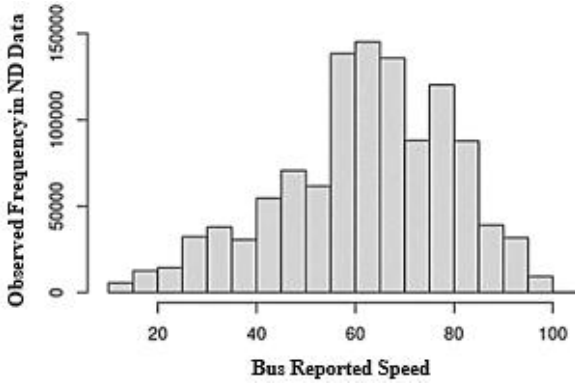

Fig 7. Distribution diagram of recorded speeds

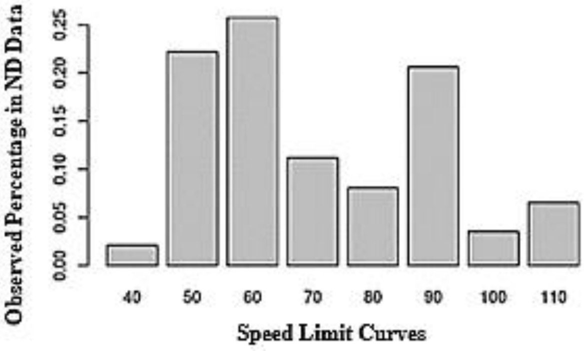

Fig 8. Distribution diagram of speed limit on curves

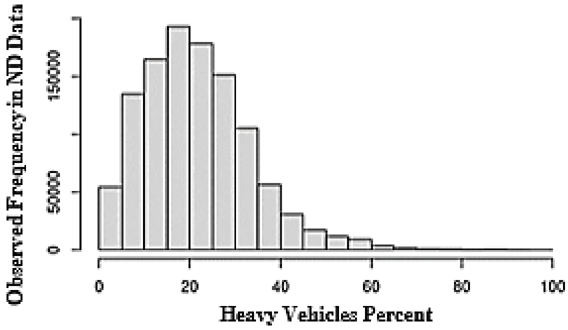

Fig 9. Distribution diagram of heavy vehicle percent

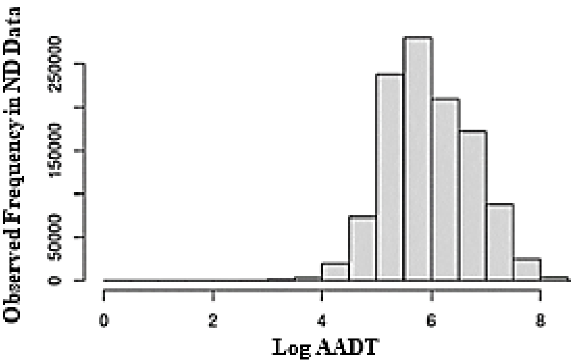

Fig 10. Distribution diagram of traffic volume

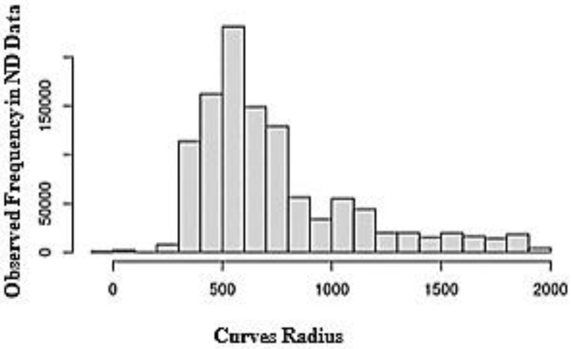

Fig 11. Distribution diagram curves radius

Tables 3, 4, 5, and 6 contain a set of research variables, including geometric variables, ND, traffic, and posted speed limit, respectively. These tables describe the data of the final model; we address each separately.

As explained, Table 3 contains geometric road variables. The set of these variables was obtained by the RSP machine of the National Company for Soil Mechanics. The machine was equipped with hardware and software simultaneously and had a set of sensors and many cameras through which it collected and recorded the geometrical variables of the longitudinal profile of the route. Part of those collected variables used in the modeling are:

Meanrad (average radius of the curve in meters): Since each curve consists of several curves, the average radius of the curves is the criterion for calculation. Length (curve length in meters): curve length is the distance between the beginning and end of a curve. Slope 1 (slope of the curve): It results from dividing the elevation difference of the points at the beginning and at the end of the curve by the length of it. Slope 2 (the slope of the tangent leading to the curve): It is calculated by dividing the elevation difference of the beginning and end points of the straight path connecting two consecutive curves, by the length of that path. The elevation digits used for calculating slopes 1 and 2 can be computed through the use of Digital Elevation Models (DEMs). Height 1 (average height in the curve): It is a ND data set and obtainable through the GPS installed on the buses. Height 2 (average height in the curve): It is obtainable through DEM. Posted Speed Limit: the permitted speed announced to the road users through the signs installed by the road; has been collected and recorded in the field.

Table 3. Geometric variables in modeling

Variable	n	mean		median	min	max	range
Meanrad	416	990.69	454.31	921.61	-1	1996.62	1997.62
Length	416	457.52	496.12	263	24	2918	2894
Slope1	416	0.14	3.77	0.30	-23.33	20.45	43.78
Slpoe2	416	0.18	4.30	0.28	-25.52	19.86	45.38
Height 1	390	1355.46	739.83	1655.22	74.54	2276	2201.46
Height 2	416	1361.61	730.80	1677.74	73.54	2261.08	2187.54

Variable	n	mean		median	min	max	range
Posted Speed Limit	416	72.12	17.61	70	40	110	70

In cartography, a digital elevation model is a series of cells that represent a continuous topographic elevation surface. Each cell representing the elevation (Z) is a feature of a certain position (X and Y).

Digital models represent a "bare earth" because they only contain information about the elevation of geological (terrain) features, such as valleys, mountains, and landslides. They do not include elevation data on non-terrestrial features, such as vegetation or buildings. Topographic maps of land and bathymetric maps can be created using digital elevation models.

Table 4 shows a part of the ND variables used in the modeling. As explained earlier, these variables are obtainable by the FMS installed on the buses. These variables are:

- Speed: The speed is calculated based on the GPS of the buses.
- In-speed: The speed calculated through the buses ECU.
- V-speed: The speed calculated by the system based on the distance and travel time of each bus.
- Mileages: traversed bus distance in kilometers.

Table 4. ND variables in modeling

Variable	n	mean	Standard deviation	min	max	range
Speed	1302113	60.39	20.40	0	137	137
Inspeed	1302113	58.52	24.00	0	120	120
Vspeed	1302113	58.18	24.36	0	200	200
Mileages	1302113	1466676511	4676360116	11565922	72398136875	72386570953

Table 5 also indicates a part of the traffic variables. Again, these variables are obtainable from SANET.

SANET is a traffic flow counting system used for better road management. The system provides the user with traffic information such as the number of passing vehicles, the average, maximum, and minimum speed of the vehicles, distances between vehicles, and, in some instances, the classification of the vehicles.

Table 5. Summary of traffic variables in modeling

Variable	n	mean	Standard deviation	median	min	max	range
Device Operation (minutes)	248973	57.66	8.36	60	5	110	105
Traffic volume (Counted)	248973	762.44	738.08	505	0	5553	5553
Percentage of Heavy Vehicles	248973	16.55	13.79	13	0	100	100
Average Speed	248973	76.57	14.08	78	0	146	146
Unauthorized Speed	248973	71.84	124.18	23	0	3040	3040
Traffic Volume (Selected value)	261337	810.07	790.84	524	0	12973	12973
Route Capacity	261744	3372.36	826.27	3300	1300	5000	3700

The traffic flow counting system consists of magnetic loops embedded in different sections of each road, each connected by a cable to a counting device beside the road. The output data of the loops is saved into the memory of the device and will then be sent to the main server of the system through GSM/GPRS network.

Besides the above variables, the researchers paid special attention to the elements affecting the operating speed on the road, such as the fixed speed control cameras and the permitted speeds announced by the signs. These features are usually missed in studying operating speed because of the assumption of free flow speed conditions. For example, 4.2 percent of the curves are within a five hundred meter range of a fixed speed control camera. The speeds within the curves were calculated according to the last speed limit sign before entering the curve. Table 6 shows the contribution of the different posted speed limits in curves.

Table 6. Speed observations within 500 meters distance of speed limit signs

Posted Speed Limit (km/h)	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**
Frequency (%)	2	22	26	11	8	21	3.5	6.5

As could be seen, the largest share of observations is within 500 meters distance of speed limit signs of 60, 50, and 90 (km/h), respectively.

4.2. Results of the Random Effects Model

The developed random effects model was calibrated. Table 7 shows the combination of model calibration results. Specifically, the analysis was performed through R-software version 3.0.2. REML algorithm estimated the model by implementing the package (lme4). All the variables are based on the BIC criterion calibrated by the model.

Table 7. Model Coefficients and Significant Variables for Random Effect (RE) Models

Variable	**Estimate**	**Standard Error**	**t-value**	**Pr(>|t|)**
Intercept	7.46E+01	3.91E+00	19.094	< 2e-16 ***
Mean Radius	5.85E-03	1.88E-03	3.112	0.00203 **
Traffic volume	-5.77E-01	2.15E-02	-26.884	< 2e-16 ***
Fix Speed Camera	-2.74E+00	9.80E-02	-27.918	< 2e-16 ***
Percentage of Heavy vehicles	-3.06E-02	1.20E-03	-25.563	< 2e-16 ***
Reverse Curve	-9.52E+00	1.79E+00	-5.313	2.03e-07 ***
Posted Speed Limit	8.93E-03	1.41E-03	6.335	2.37e-10 ***
Nightfall	-8.66E-01	1.54E-01	-5.61	4.27E-08***

**: p-value<0.01, ***: p-value<0.001

The obtained results in Table 7, compared with the similar studies, are reported as follows:

With each meter increase in curve radius, the speed increases by one kilometer per hour. For example, [Nasiri et al., 2012; Donnell et al., 2001], obtained similar results in their research. Their model showed that the operating speed also increases with an increase in the curve radius.

Like the findings of [Gressai et al., 2021], the modeling results showed that the speed of cars decreases with the increase in traffic volume.

According to Table 7, buses usually reduce their speed by two kilometers per hour when they are within the range of a speed control camera. The findings of [Rogerson et al., 1994] also confirm that. They found that a speed reduction of over 15 km/h occurs within 1 km of a speed camera.

As the percentage of heavy vehicles increases, the speed of buses decreases by thirty kilometers per hour. Researchers have presented results similar to the current research in separate studies. They showed that as the proportion of large vehicles increases in the traffic flow, the average speed of traffic flow and the average speed of cars decreases.

In reverse curves, the speed of buses is nine kilometers per hour less than their speed in simple curves. For example, Sabhanayagam's research results prove that significant speed reductions are achievable with minimal reverse curves [Sabhanayagam, 2018].

On the other hand, with an increase in the posted speed limit by one kilometer per hour, the speed of cars increases by nearly nine meters per hour. [Gao et al., 2020] also presented a similar result, while [Semeida, 2013] achieved different results in their study in Egypt. They state that because of the bad behavior of drivers in Egypt, posted speed limits have very little effect on operating speed.

During the night, the speed of drivers decreases by about 0.8 km per hour. Some researchers presented similar results. They found that operating speeds at night and day did not differ significantly from each other in short tangents, while in long tangents, the operating speed at night was lower than during the day. Besides analyzing the effect of the variables listed in Table 7, other variables and factors on speed were also considered in this study:

- None of the longitudinal slope, the transverse slope of the curve, and the slope of the tangent leading to the curve were significant.
- The effect of the distance traveled by buses (milage) was insignificant, with a p-value of 0.37.

4.3. Model Evaluation Indicators

Table 8 shows the statistical analysis of the construed RE model. It is confirmed that the data fit the random effects model well, although it must be noted that random effects in the data are completely dependent on the sampling design.

Table 8. Composition of statistical analysis of the model

Groups	Name	Variance	Standard .Dev	Corr
Curve id	(Intercept)	179.10	13.38	
Driver id	(Intercept)	29.18	5.40	
	Nightfall	23.62	2.81	-0.15
Residual		108.9	10.43	

Based on Table 8, it can be inferred that a significant proportion of the variance is attributed to changes in the arch level, while the characteristics of the car and the driver contribute to a smaller extent. Furthermore, although drivers may exhibit varying behaviors during the day and night, the effect on speed is relatively minor. A portion of the variance is also associated with the items not included in the model.

4.4. Validation of the Model

To check the normal distribution of data (Normality Test), a Q-Q plot is used. The diagram shows the population distribution compared to the normal distribution. In an ideal Q-Q diagram, all the points are on a straight line with an angle of forty-five degrees. Going out of this situation indicates problems in the data, resulting in abnormal errors in the regression model.

The following diagrams show the multiple dispersion of the errors, which seems appropriate. The following table presents the parameters that influence the speed of drivers in this model, all of which are constant.

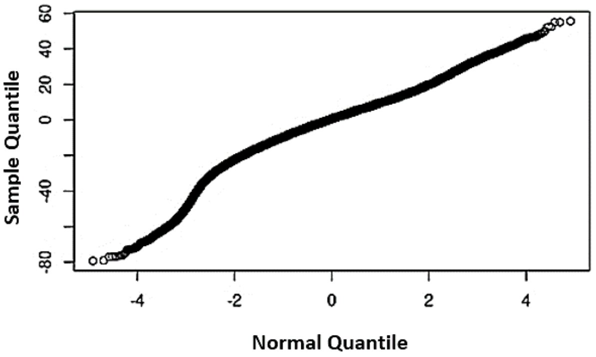

Fig 12. Q-Q plot chart

In a regression model, the ideal errors should have a normal distribution. It means that the largest number of estimates should be within a small distance from the observation value, and the number of estimates with a large distance from the observation value should be small. The diagram below shows a Bell Curve distribution around the zero value.

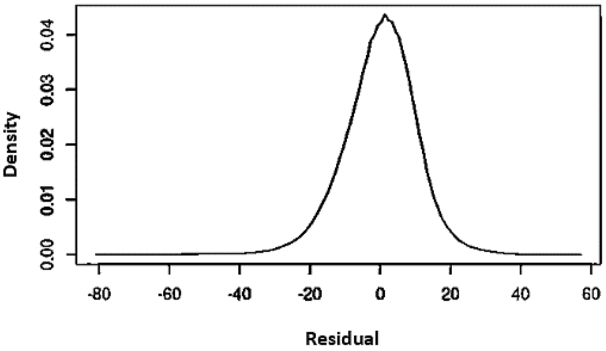

Fig 13. Standard normal distribution curve

Based on the above two diagrams, the errors have a normal distribution, and the correctness of the model is confirmable. During the development of the random effect model, some simple linear models were also developed, and the following are the diagrams and results of those linear models:

Table 9. Linear models created for the final model

Variable	Estimate	Standard. Error	t value Pr	(>|t|)
(Intercept)	6.17E+01	2.08E-01	296.7	<2e-16
Mean Radius	5.23E-03	5.41E-05	96.64	<2e-16
Traffic volume	-1.23E+00	2.38E-02	-51.67	<2e-16
Fix Speed Camera	8.26E+00	7.76E-02	106.51	<2e-16
Percentage of Heavy vehicles	-8.33E-02	1.54E-03	-53.98	<2e-16
Reverse Curve	-5.67E+00	4.65E-02	-121.96	<2e-16
Posted Speed Limit	2.21E-01	8.41E-04	262.36	<2e-16

To compare these two models, the researchers benefited from Akaike, Bayesian, similarity, deviation, and second power of the object provided in the Nonlinear Least Squares (NLS) package. You can see the results in the table below.

Table 10. Results of comparative tests for two linear and random effect models

Model	[5]AIC	BIC	logLik	DIC[6]	Pr(>Chisq)
Linear	89367416	9367511	-4683700	9367400	
Random Effect	8406068	8406211	-4203022	8406044	< 2.2e-16 ***

As you can see, the AIC, BIC, and deviance indices in the random effect model are lower than in the linear model, and the log-likelihood shows more similarity. The presence of all these characteristics confirms the superiority of a model that has been tailored to fit the candidate.

5. Conclusion

A random effects model was used to study the effect of different factors on the prediction of the operating speed of buses on a highway in the north of Iran. The model classified the parameters affecting the speed of the drivers in two levels, the vehicle, and the curve. In this research, the effect of variables was considered as fix-effects. As a result of the speed data analysis, the speed of the buses along the curves followed a normal distribution. The analysis of the variables at the random effect level showed that a large share of the variance results from changes in the level of the curves, and far fewer changes are for the characteristics of the car and the driver; however, different drivers present different behaviors during the day and night. Therefore, some variance concerns the cases not included in the model.

According to the model analysis, the presence of heavy vehicles in the traffic flow and reverse curves has a significant impact on the operating speed of buses. As the percentage of heavy vehicles increases, the speed of buses decreases by thirty kilometers per hour. Additionally, buses on reverse curves traveled at a lower speed than on other curves by nine kilometers per hour.

Based on all the factors that affect operational speed, this article developed a speed prediction model. By calculating the speed using the model, it is possible to identify inconsistent sections along a road. To mitigate the risk of bus accidents, geometric improvements can be implemented on the inconsistent sections or

---

[5] . Akaike Information Criterion

[6] . Deviance information criterion

warning or speed signs can be installed that are tailored to the characteristics of those sections. The number of sites could be increased for future studies to build more accurate prediction models. The influence of meteorological data and individual characteristics of drivers on speed is also examinable. Through the use of the random effect model, the models were constructed in this study. It is possible to compare the results using an alternative estimation method.

6. References

– Gargoum, S. A. and El-Basyouny, K. (2016) "Exploring the association between speed and safety: A path analysis approach", Accident Analysis & Prevention, Vol. 93, pp.32-40.

– Mirbaha, B., Saffarzadeh, M. and Noruzoliaee, M. H. (2013) "A model for predicting schoolchildren accidents in the vicinity of rural roads based on geometric design and traffic conditions", International Journal of Transportation Engineering, Vol. 1, No. 1, pp.25-33.

– Fitzsimmons, E. J., Souleyrette, R. R., and Nambisan, S. S. (2013b). "Measuring horizontal curve vehicle trajectories and speed profiles: Pneumatic road tube and video methods." J. Transp. Eng., 10.1061/(ASCE)TE.1943-5436.0000501, 255–265.

– Campbell, J. L., Richard, C. M., and Graham, J. (2008). Human factors guidelines for road systems, Transportation Research Board, Washington, DC.

– Jacob, A., and Anjaneyulu, M. V. L. R. (2013). "Operating speed of different classes of vehicles at horizontal curves on two-lane rural highways." J. Transp. Eng., 10.1061/(ASCE)TE.1943-5436.0000503, 287–294.

– Aashto, A. (2001). Policy on geometric design ofhighways and streets. American Association of State Highway and Transportation Officials, Washington, DC, 1(990), 158.

– Sil, G., Maji, A., Nama, S., & Maurya, A. K. (2019). Operating speed prediction model as a tool for consistency based geometric design of four-lane divided highways. Transport, 34(4), 425-436.

– Nama, S., Maurya, A. K., Maji, A., Edara, P., and Sahu, P. K. (2016).“Vehicle speed characteristics and alignment design consistency for mountainous roads." Transp. Dev. Econ., 2(2), 1–11.

– Misaghi, P.; Hassan, Y. 2005. Modeling operating speed and speed differential on two-lane rural roads, Journal of Transportation Engineering 131(6): 408–418.

– Fitzpatrick, K., et al. (2000a). Alternative design consistency rating methods for two-lane rural highways, Federal Highway Administration, Washington, DC.

– Nie, B., & Hassan, Y. (2007). Modeling driver speed behavior on horizontal curves of different road classifications (No. 07-0782).

– Lamm, R., Choueiri, E. M., Hayward, J. C., & Paluri, A. (1988). POSSIBLE DESIGN PROCEDURE TO PROMOTE DESIGN CONSISTENCY IN HIGHWAY GEOMETRIC DESIGN ON TWO-LANE RURAL ROADS. Transportation Research Record, (1195).

– Passetti, K. A., & Fambro, D. B. (1999). Operating speeds on curves with and without spiral transitions. Transportation research record, 1658(1), 9-16.

– S. K. S. Abbas, M. A. Adnan, and I. R. Endut, "Exploration of 85 th Percentile Operating Speed Model on Horizontal Curve: A Case Study for Two-Lane Rural Highways," Procedia-Social and behavioral sciences, vol. 16, pp. 352-363, 2011.

– Bella, F., Calvi, A., & D'Amico, F. (2014). Analysis of driver speeds under night driving conditions using a driving simulator. Journal of safety research, 49, 45-e1.

– Gong, H., & Stamatiadis, N. (2008). Operating speed prediction models for horizontal curves on rural four-lane highways. Transportation Research Record, 2075(1), 1-7.

– Wang, L., He, Y., & Wang, S. (2010). Research on Relationship between Operating Speed and Posted Speed Limit for Mountainous Freeway. In ICCTP 2010: Integrated Transportation Systems: Green, Intelligent, Reliable (pp. 816-823).

– Semeida, A. M. (2013). Impact of highway geometry and posted speed on operating speed at multi-lane highways in Egypt. Journal of Advanced Research, 4(6), 515-523.

– Boroujerdian, A. M., Seyedabrishami, E., & Akbarpour, H. (2016). Analysis of geometric design impacts on vehicle operating speed on two-lane rural roads. Procedia engineering, 161, 1144-1151.

– Maji, A., Sil, G., & Tyagi, A. (2018). 85th and 98th percentile speed prediction models of car, light, and heavy commercial vehicles for four-lane divided rural highways. Journal of Transportation Engineering, Part A: Systems, 144(5), 04018009.Bamdad 2019.

– Bamdad Mehrabani, B., & Mirbaha, B. (2019). Modeling the operating speed in tangents and curves of four-lane highways based on geometric and roadside factors. International Journal of Transportation Engineering, 6(4), 355-366.

– Gao, C., Xu, J., Jia, X., Dong, Y., & Ru, H. (2020). Influence of large vehicles on the speed of expressway traffic flow. Advances in Civil Engineering, 2020.

– JHS, M. (2021). Operating Speed Prediction Model for Combined Curve on Two-Lane Two Way Non-Urban Roads.

– Malaghan, V., Pawar, D. S., & Dia, H. (2022). Speed prediction models for heavy passenger vehicles on rural highways based on an instrumented vehicle study. Transportation letters, 14(1), 39-48.

– Van Schagen, I., Welsh, R., Backer-Grøndahl, A., Hoedemaeker, M., Lotan, T., Morris, A., Sagberg, F., & Winkelbauer, M. (2011). Towards a large-scale European Naturalistic Driving study: main findings of PROLOGUE. PROLOGUE Deliverable D4.2. SWOV Institute for Road Safety Research, Leidschendam, The Netherlands.

– Cruzado, I. (2009). Factors affecting driver speed choice along two-lane rural highway transition zones. The Pennsylvania State University.

– Bassani, M., Cirillo, C., Molinari, S., & Tremblay, J. M. (2016). Random effect models to predict operating speed distribution on rural two-lane highways. Journal of transportation engineering, 142(6), 04016019.

– Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 461-464.

– Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. Biometrika, 94(2), 443-458.

– Nasiri, H. A., Mohammadi, R. A., Nazari, H. (2012). "Geometric design consistency in two-lane rural road (case study in mazandran two-lane rural road)." The 12th international conference on traffic and transportation engineering, Iran.

– Donnell, E. T., Ni, Y., Adolini, M., and Elefteriadou, L. (2001). "Speed prediction models for trucks on two-lane rural highways." Transportation Research Record 1751, Transportation Research Board, National Research Council, Washington, DC, 44–55.

– Gressai, M., Varga, B., Tettamanti, T., & Varga, I. (2021). Investigating the impacts of urban speed limit reduction through microscopic traffic simulation. Communications in transportation research, 1, 100018.

– Rogerson, P. A., Newstead, S. V., & Cameron, M. H. (1994). Evaluation of the Speed Camera Program in Victoria 1990-1991, Phase 3: Localised Effects On Casualty Crashes and Crash Severity. Phase 4: General Effects On Speed (No. 54).

– Sabhanayagam, A. (2018). Impact of S-Curve on Speed in a Modern Roundabout.

– Ishak, S., Osman, O. A., Ashly, G., Theriot, M., Ye, M., & Bakhit, P. (2016). Exploring Naturalistic Driving Study (NDS) Data and Roadway Information Database (RID) for Emerging Applications in Traffic Safety.