WILEY | Hindawi

Research Article

A New Big Data Approach to Understanding General Traffic Impacts on Bus Passenger Delays

Yaiza Montero-Lamas [ID],[1] **Margarita Novales** [ID],[1] **Alfonso Orro** [ID],[1] **and Graham Currie** [ID][2]

[1]*Universidade da Coruña, Group of Railways and Transportation Engineering, ETS Ingenieros de Caminos, Canales y Puertos, Elviña, 15071 A Coruña, Spain*
[2]*Public Transport Research Group, Monash University, Melbourne, Australia*

Correspondence should be addressed to Yaiza Montero-Lamas; y.montero@udc.es

This paper presents a new method to quantify the potential user time savings if the urban bus is given preferential treatment, changing from mixed traffic to an exclusive bus lane, using a big data approach. The main advantage of the proposal is the use of the high amount of information that is automatically collected by sensors and management systems in many different situations with a high degree of spatial and temporal detail. These data allow ready adjustment of calculations to the specific reality measured in each case. In this way, we propose a novel methodology of general application to estimate the potential passenger savings instead of using simulation or analytical methods already present in the literature. For that purpose, in the first place, a travel time prediction model per vehicle trip has been developed. It has been calibrated and validated with a historical series of observations in real-world situations. This model is based on multiple linear regression. The estimated bus delay is obtained by comparing the estimated bus travel time with the bus travel time under free-flow conditions. Finally, estimated bus passenger time savings would be obtained if an exclusive bus lane had been implemented. An estimation of the passenger's route in each vehicle trip is considered to avoid average value simplifications in this calculation. A case study is conducted in A Coruña, Spain, to prove the methodology's applicability. The results showed that 18.7% of the analyzed bus trips underwent a delay exceeding 3 min in a 2,448 m long corridor, and more than 33,000 h per year could have been saved with an exclusive bus lane. Understanding the impact of different factors on transit and the benefits of a priority bus system on passengers can help city councils and transit agencies to know which investments to prioritize given their limited budget.

1. Introduction

Improving bus systems to attract new users is essential to achieving more sustainable mobility. Traffic delay is a critical factor affecting bus travel time performance [1, 2]. As traffic in cities grows, traffic congestion will cause a rise in the number of transit vehicles required to maintain headway and, therefore, an increase in the operation costs [3]. The increase in travel time for transit users will also result in ridership loss [2].

Advances in big data availability provide much potential to improve our understanding of traffic impacts on bus travel time [4]. This paper proposes a new methodology for calculating the delay in bus travel time due to general traffic, bus ridership, and accumulated rainfall. Furthermore, the methodology determines the bus user time savings that the implementation of a dedicated or exclusive bus lane (DBL) can generate. An accurate evaluation of these savings can only be made in before-and-after studies, but an estimation of its value is necessary for planning purposes. Analytical methods based on traffic theory or simulation studies can be used, but the proposed methodology, which has general applicability and is based on automatically collected big data sources, allows estimation for each corridor or street section employing real performance information in local conditions and configurations.

This methodology provides data on the effect of mixed traffic on the transit travel time and, consequently, on bus user time. Similarly, it quantifies the estimated savings and

benefits to buses and bus users if an exclusive bus lane exists; therefore, this methodology is useful for transit planning, providing data for decision-making to justify its possible implementation. This type of operating environment results in decreased bus travel time and increased transit reliability, two of the most important attributes for users [5], making it more attractive to noncaptive public transport travelers.

The remainder of this paper is organized as follows: Section 2 discusses the previous studies related to this research. Section 3 presents a case study and a descriptive analysis of the variables considered. Section 4 explains the methodology: a multiple linear regression (MLR) model is developed in the first part, and its applications for delay calculations are explained in the second one. Section 5 presents the results and discussion. Finally, the conclusions and future work are presented in Section 6.

2. Literature Review

The factors that affect the choice of public transport have been widely analyzed in the literature, focusing mainly on fares, quality of service and income, and car ownership [6]. Among quality of service attributes, public transport providers can control some operational service attributes that affect user satisfaction, like frequency, speed, crowdedness, and reliability. Improving public transport user satisfaction is a way to maintain and grow demand [7]. Several authors highlighted the importance of reliability, waiting time, or service time on buses and other transit systems [5, 8, 9]. All these variables are affected by mixed-traffic congestion.

The travel time of on-road transit, as well as the factors influencing it, has been the topic of many studies for decades. Factors that influence bus travel time include corridor configurations, bus stops and ridership, and interaction with traffic. The Transit Capacity and Quality of Service Manual (https://www.trb.org/Main/Blurbs/169437.aspx) [3] distinguishes four main operating environments depending on the degree of protection of the bus against traffic vehicles (grade separated, exclusive, semi-exclusive, and mixed traffic). In mixed traffic corridors, where transit shares lanes with other vehicles, buses are exposed to a wide variety of possible traffic-related delays.

To avoid the aforementioned delays and free transit from traffic interference, two main bus priority approaches have been deeply studied so far dedicated bus lane (DBL), and intermittent bus lane (IBL) [10, 11]. In a dedicated or exclusive bus lane, only transit use is permitted (in case carpools, taxis, or bicycles are allowed, it would be a semi-exclusive line). However, the intermittent bus lane changes its status from a bus lane to a mixed-traffic lane if the bus is not using it or traffic conditions do not entail a delay for the bus.

The effects of implementing a DBL in Rome were studied by Russo et al. [12]. Their results showed a bus travel time reduction of 18% due to the exclusive bus lane provision. Since turning a mixed-traffic lane into an DBL normally entails a loss of road capacity, Gan et al. [13] developed a tool to evaluate whether the implementation of a proposed DBL (or improvements on an existing one) is justified. They proposed a corridor simulation (CORSIM) and empirical models to estimate the average person's travel time with and without a bus lane. As long as the average person's travel time in mixed traffic is longer, a DBL implementation is recommended. Yang et al. [14] analyzed the effects of converting a general traffic lane into a DBL by employing a microsimulation model, considering various demand levels and bus share ratios. Significant reductions in bus delays were found for all demand levels, especially for high values. Under all analyzed scenarios, average passenger delays decrease when the bus share ratio increases. Eichler and Daganzo [11] evaluated by traffic flow theory analysis the implementation of a variant of the IBL proposed by Viegas and Lu [10], forcing traffic to leave the lane by variable message signs. They found that there was a definite niche between transit signal priority and DBL as an adequate solution. Kampouri et al. [15] also employed microsimulation to evaluate the effects of a theoretical IBL application. They proposed that a DBL lane could change to a shared lane activated by the traffic volumes observed. They focused on the critical level of traffic volumes and bus service frequencies to activate the shared use of the line and obtain traffic flow and environmental benefits.

Surprenant-Legault and El-Geneidy [16] calculated the impact of introducing a reserved bus lane on bus travel time, transit reliability, and on-time performance on two parallel bus routes in Montreal. Their data were obtained from automatic vehicle location (AVL) and automatic passenger counting (APC) before and after the implementation. The results showed that the exclusive bus lane produced travel time savings from 1.3 to 2.2% and an increment of the odds of being on schedule of 65% in both cases. Arias et al. [17] calculated the potential travel time savings of bus lanes for all segments with bus networks. They used General Transit Feed Specification (GTFS) trip-stop data and ridership data from the Metropolitan Atlanta Rapid Transit Authority (MARTA) bus network to estimate the potential user-weighted travel time savings for each segment between stops based on schedule padding. Therefore, the potential delay is calculated as the differences between the minimum travel time registered and the actual travel time.

Regarding the data source, in addition to AVL, APC, and automatic fare collection (AFC), mobile phones have provided a relevant amount of low-cost information that has been used to estimate bus demand and travel time. Mobile phones and other devices with active Bluetooth can be located by sensors that calculate and record the speeds and travel times of the vehicle they are in [18, 19], although recent techniques of anonymization can prevent it. Besides, real-time bus tracking information can be collected from bus rider's smartphones connected to a Wi-Fi network [20]. Transit operation studies using big data also get information from GPS points, smartcard data, web data, and social media data [21]. Wang et al. [22] proposed a broadly adaptable alternative to AVL-APC systems by extracting fined-grained information from raw big data (GPS signal and AFC records). They analyzed bus delays at the route and segment levels. They used average bus speeds to detect road congestion and performed a regression analysis on the difference between scheduled and actual travel times to identify reasons for delays.

A big data real-time processing and analysis was performed in Sydney to study the impact of COVID-19 on bus delays [23]. They considered bus position, bus network, and bus timetable data to quantify the transit delays. Their findings revealed a significant decrease in bus delays in March 2020 around the eastern and central suburbs and a drop in traffic congestion in the central urban areas. The results also showed a relevant impact on people's travel behavior due to the COVID-19 pandemic. COVID-19 has led to a very active line of transit research due to its relevant impacts on mobility, including its emotional and psychological impact on transit users [24].

Most of the research performed the analysis of the factors influencing bus travel time as a first step to calculating bus delays [16, 25]. Levinson [1] calculated traffic speeds and bus delays based on surveys in a representative sample of several cities in the United States. The study concluded that, during the peak hour, 26% of the transit travel time in the central business district (CBD), or 3 minutes per mile (min/mi), is caused by the traffic delay, 15% in the city (0.9 min/mi), and 16% in the suburbs (0.7 min/mi). Among the author's recommendations is to provide a priority bus lane to reduce traffic-induced congestion, but time savings are not calculated in the paper.

The impact of traffic congestion on bus travel time in New Jersey was studied by McKnight et al. [26] using multiple linear regression with a 690 records data. The main aim was to measure the extent to which general traffic congestion increases bus travel time. The travel time data for buses and cars are mainly from field measures. The variables considered were bus travel time rate, car traffic time rate, values per mile of the number of bus stops, actual bus stops, left turns, signals, boardings, alightings, and the sum of boardings and alightings in the route segment. Variations in car travel time were observed to affect bus travel time more than twice as much as other variables such as boardings per mile and bus stops per mile. McKnight et al. [26] concluded that bus travel time increases proportionally with car travel time and suggested that bus operations can benefit from an improvement in car traffic flow.

The influence of the weather on bus operations has also been the object of research [27]. Meteorological conditions affect the choice of transport mode [28]. Arana et al. [29] observed that the number of bus boardings in Gipuzkoa (Spain) increases when the temperature rises. In contrast, the results indicated that wind and rain caused a decrease in the number of bus boardings. Wang et al. [30] studied the impact of winter weather (snowfall, temperature, and visibility) on bus travel time in Harbin, China, based on historical bus GPS data for two years. They conclude that the travel time series of consecutive bus trips displays autocorrelation, meaning that the travel time of a bus is influenced by the travel times of the two preceding buses. They also found that Bus Line 18 increased by 0.483 min when the cumulative snowfall level increased by 1 unit. Novales et al. [31] observed that rainy conditions increase the bus's lost time at stops.

After reviewing the literature, there is a relevant field of research in the use of the automatically collected big data to evaluate potential DBL implementations. This study aims to obtain separately the influence on bus travel time of factors specific to the transit operation from the bus delays produced by general traffic, using information from thousands of recorded trips along with the concurrent situation. As a result, potential time savings can be obtained by comparing the travel time in mixed traffic with the time that could have been spent under DBL conditions for each trip. Therefore, this study fills the existing gap by obtaining the delay for each actual bus trip. Our method for estimating delays improves on previous approaches by being more accurate than comparing actual bus travel time with the minimum bus travel time registered (as Arias et al. [17] conducted), which could correspond not only to a situation with low traffic but also with low ridership, reduced dwell times, and skipped stops. It is also more accurate than Wang et al. [22] proposal, which infers road congestion using bus speeds below 10 km/h instead of using the real traffic data. Moreover, our approach is based on the analysis of the real data of the corridor rather than traffic theory or simulation models that often require a simplification of the actual circumstances. Thereby, it is possible to calculate savings specifically for each trip and link them, through big data and an alighting stop prediction algorithm, to the bus users who actually performed each trip, avoiding the simplification of considering average travel times.

3. Case Study and Descriptive Analysis

This paper investigates, in the first part, the influence of several variables on bus travel time on the mixed traffic including general traffic time, traffic flow rate, occupancy, ridership, and accumulated rainfall. The traffic flow rate and occupancy were obtained from inductive loops, and general traffic time was acquired from sensors located throughout the city to detect active Bluetooth devices. More than 20.5 million data are processed in this new approach. In the second part, this paper proposes a new methodology that estimates the potential time savings for buses and passengers based on transit, traffic, and weather data from the corridor under analysis. The variables to be used in our approach were selected for obtaining a methodology of general applicability, as the utilized data are usually registered in many cities. The methodology is contrasted in a case study developed for the city of A Coruña, in the northwest of Spain. This section presents the corridor of the case study, the variables considered, the sources of information, and the data processing.

A Coruña is in the autonomous region of Galicia, and its population was 245,711 inhabitants in 2019 [32]. In 2020 and 2021, the COVID-19 pandemic had a significant impact on A Coruña bus services [33]. This study considered data from 2019 along a 2,448 m long corridor called Ronda de Outeiro in the southeast—northwest direction. This corridor has been chosen due to its relevance in the city bus network and the simultaneous availability of a bus line, inductive loops, and Bluetooth sensors. In addition, this corridor is served by bus line 14, which is the bus line with the highest annual demand in the city (more than 2.7 million users in 2019) and is the one to be studied in this research. This corridor is

composed of sections with two or three mixed-traffic lanes in the considered direction and some areas with turn-exclusive lanes: two stretches of turn right and four of turn left in the selected stream. There are 19 signalized intersections and five crosswalks with traffic lights located between intersections. The traffic lights cycle length varies between 80 and 105 s. The stream analyzed is a corridor with 10 bus stops and is travelled by different regular bus lines operated by Compañía de Tranvías de La Coruña (CTC) throughout A Coruña city. The locations of inductive loop detectors, Bluetooth sensors, and bus stops are shown in Figure 1.

Bus stop arrival times are analyzed to obtain bus travel times of the complete corridor. These data were obtained from the CTC's Transit Management System, which records boardings per bus stop, bus stop arrival time, payment method, and type of ticket. Bus travel time was the variable to be estimated. A regression model presented in the next section was developed to predict bus travel time using as inputs the explanatory variables listed in Table 1. A summary of these variables and their statistics is shown in Table 1, according to the naming that will be used in the analysis.

3.1. Bus Travel Time.
The dependent variable, bus travel time, was obtained as the difference between bus arrival time at the final stop (559) and bus arrival time at the initial stop (119) of the studied stream for every bus trip on line 14 in 2019. This travel time includes bus time serving stops but not passenger waiting and walking times. The variations in bus travel time depending on the hour are shown in Figure 2. The significant variability of bus travel times can be observed in almost all time slots, especially on weekdays. During regular weekdays, average bus travel time was higher than on Saturdays and Sundays or holidays from 6 a.m. to 8 p.m. During the peak hours, during regular weekdays, higher average bus travel times were at 8 a.m., 1 p.m., and from 5 p.m. to 7 p.m. The higher average values on Saturday were from 11 a.m. to 12 a.m. and from 7 p.m. to 8 p.m. On Sundays and holidays, the results indicated that the highest average travel time was from 7 p.m. to 9 p.m. The difference between average bus travel time for peak and off-peak hours was 355.05 s during regular weekdays, 174.60 s on Saturdays, and 201.59 s on Sundays and holidays.

3.2. General Traffic Travel Time.
For the explanatory variables shown in Table 1, general traffic travel times were acquired from sensors located in different parts of the city (Figure 1). These sensors detect active Bluetooth devices, encode and anonymize them, and record the information every three min to provide information on real-time general traffic speeds and travel times. According to city council measures for 2019, the average percentage of vehicles that have a Bluetooth device in Ronda de Outeiro is 33.88%, providing an adequate sample of general traffic in the corridor. The detection of the same device by two consecutive sensors allows us to determine the travel time of the vehicle with that device between the intersections where those sensors are installed. The four Bluetooth detection

sensors divide, between each pair, the corridor into three vectors shown in Figure 1.

Once the corridor is subdivided into three vectors, general traffic travel time for each vector is calculated by averaging the three-min traffic travel time records only for the periods when each bus is travelling along that vector. This means that only general traffic conditions which could really affect bus speeds are considered. The total general traffic time affecting each bus trip along the corridor is obtained as the sum of the general traffic times of the three vectors considered. The bus is considered to be in vector 44 from stop 129 to stop 122 (the first one located after the Bluetooth sensor 2). Similarly, it will be on vector 62 during the period when it travels the distance between stops 122 and 124. Finally, from stops 124 to 559, the bus is considered to be in vector 23. Data from a total of 525,600 general traffic travel time records were considered for this research.

3.3. Accumulated Rainfall.
The accumulated rainfall in A Coruña in 2019 was obtained from the Torre de Hércules meteorological station [34], which registers data for each ten-minute period. The accumulated rainfall corresponding to each bus vehicle trip was calculated by adding the values for the 10-min periods when that bus was travelling through the considered corridor. A total of 52,560 rainfall data pieces were processed for this research.

3.4. Bus Ridership.
On the urban buses of A Coruña, boardings (and fare-payment) are only allowed through the front door, and alightings through the rear door(s) without smartcard check-out. All boardings per stop and the corresponding payment method are recorded, but alightings are not registered, lacking therefore direct information about the origin-destination matrixes of the network and about the bus load between each pair of stops. For this reason, in the bus travel time model phase of this research, bus ridership was divided into two different independent variables: previous and stream ridership. Stream ridership (SR) corresponds to the sum of the boardings in the studied corridor, from bus stops 119 to 127 (Figure 1). Since there is no information about the alightings, the previous ridership (PR) was considered a proxy for crowding when the bus enters the stream. The more users inside the bus as it enters the corridor, the more likely it is to serve a stop when no passengers are waiting to board, only for alighting. Furthermore, if the bus occupancy is high and there are standees, boardings will last longer and, consequently, the bus travel time will also be longer. Previous ridership (PR) was obtained as the sum of the passengers who boarded the bus at any of the 14 previous stops of bus line 14 in the considered direction, from bus stops 15 to 269 (Figure 1).

3.5. Traffic Flow Rate and Occupancy.
Both traffic flow rate and occupancy data were obtained from inductive loops installed throughout the city and provided by the A Coruña council. The loops considered in this research are shown in Figure 1. Loops record the information once a minute, and
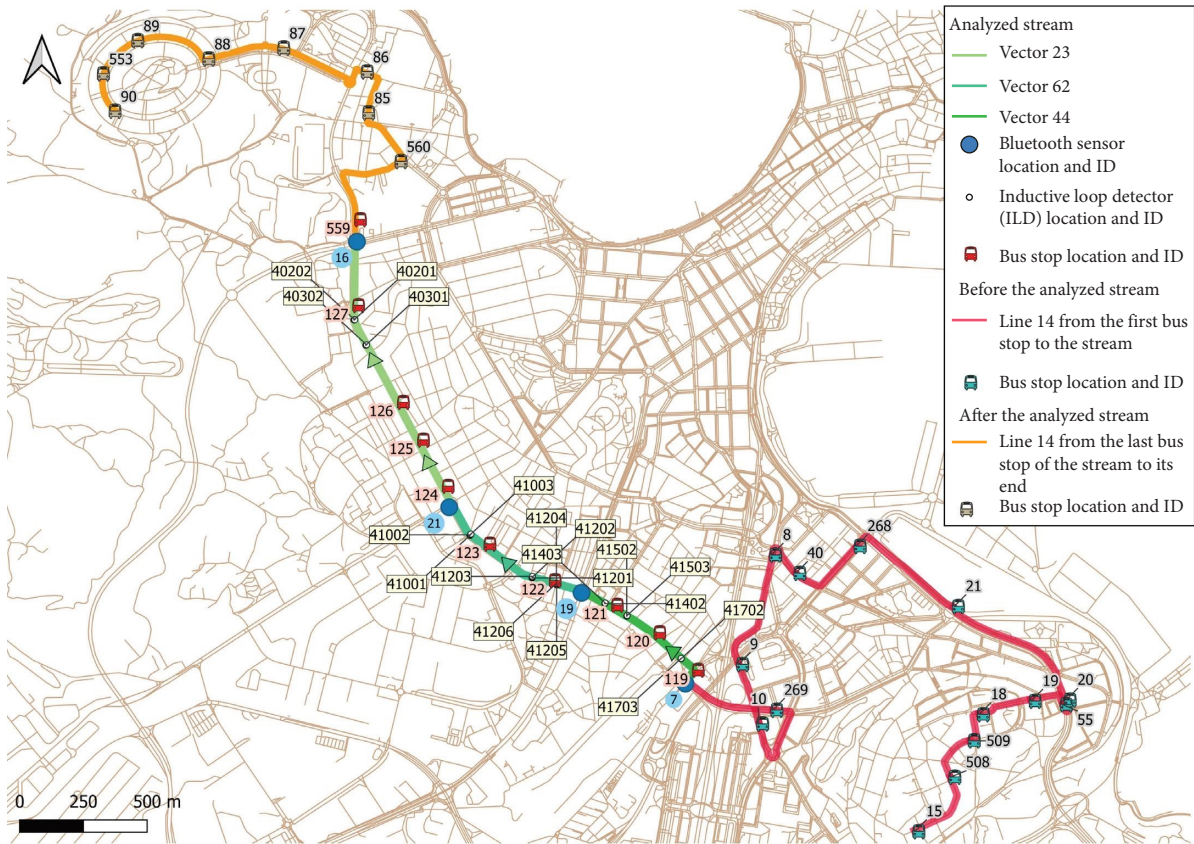
Figure 1: Corridor, bus stops, Bluetooth sensors, and inductive loop detectors location. Source: Background image © OpenStreetMap contributors.

a total of 19,972,800 values were processed to describe traffic conditions. Following the criteria established to calculate general traffic travel time, traffic flow rate and occupancy were considered in each vector only for the periods when each bus was travelling along that vector. The traffic flow rate was calculated as follows: the traffic flow rates of all the loops located at the same cross section of the street were added and averaged along the stream. Road occupancy was calculated as the mean value along the stream of the averaged values of all the loops located at the same cross section. The values of the inductive loops were only considered if they were located in lanes that are not turn-exclusive.

Data cleansing was performed to remove erroneous values related to all the aforementioned variables. The initial number of bus trips considered was 26,752. These data were examined, and inconsistent data were removed. The inconsistencies can be related to shift changes during a service, a lack of bus arrival time at a stop, erroneous values, or a lack of data for any of the variables (due to the lack of data quality or malfunctioning of the loops or Bluetooth sensors). Finally, the number of bus trips selected was 21,591.

The scatter plot matrix of all the variables considered in the model is represented in Figure 3, which shows that there is a relationship between all the independent variables and the bus travel time. The diagonal of the graph shows the univariate distribution of each variable. The last row displays the relationship between the dependent (BTT) and

independent variables. Furthermore, Figure 3 exposes a relationship between the independent variables. For this reason, the Pearson's correlation coefficient test, which measures the strength of the linear relationship between two variables, is performed. The results of the Pearson's correlation coefficient test are shown in Table 2. It is relevant for the subsequent discussion to highlight the fact that high values of SR and PR correspond to high TFR values.

4. Methodology

The goal of establishing a bus-exclusive lane is to reduce bus travel time and its variability. To quantify and justify the benefits of implementing an exclusive bus lane in a corridor that is a part of a bus line, a novel methodology is developed to calculate bus delays due to mixed traffic and convert them to passenger time savings. A summary of this procedure is represented in Figure 4, and the present section explains it in detail for a general case.

4.1. Estimated Bus Travel Time. After data processing (purple in Figure 4), the second step of the methodology consists of performing an MLR analysis to predict bus travel times along the length of the corridor where the implementation of a DBL is under study (dark green in Figure 4). Other statistical approaches can also be considered if they fit better

TABLE 1: Definitions and summary statistics of the considered variables.

| Variables	Abbrev	Definitions	Mean	Std.	Min	Max
(a) Independent variables used in multiple linear regression (MLR)						
General traffic time (s)	GTT	Travel time of the general traffic	373.521	62.251	227.557	1,174.473
Accumulated rainfall (l/m²)	AR	Accumulated rainfall	0.028	0.179	0.000	6.500
Previous ridership	PR	Number of boardings from the first bus stop (15) to bus stop 269 (included) in line 14 (Figure 1)	35.608	17.061	0.000	125.000
Stream ridership	SR	Number of boardings from bus stop 119 to 127 (included) in line 14 (Figure 1)	26.404	14.883	0.000	112.000
Traffic flow rate (veh/h)	TFR	Traffic flow rate	668.512	215.058	0.000	1,194.444
Inductive loop occupancy (%)	ILOP	Roadway occupancy: Percentage of time that the loop has a vehicle standing over it	5.158	3.060	0.207	41.847
(b) Dependent variable used in MLR						
Bus travel time (s)	BTT	Travel time of the bus in the stream	784.919	143.279	329.000	1,892.000
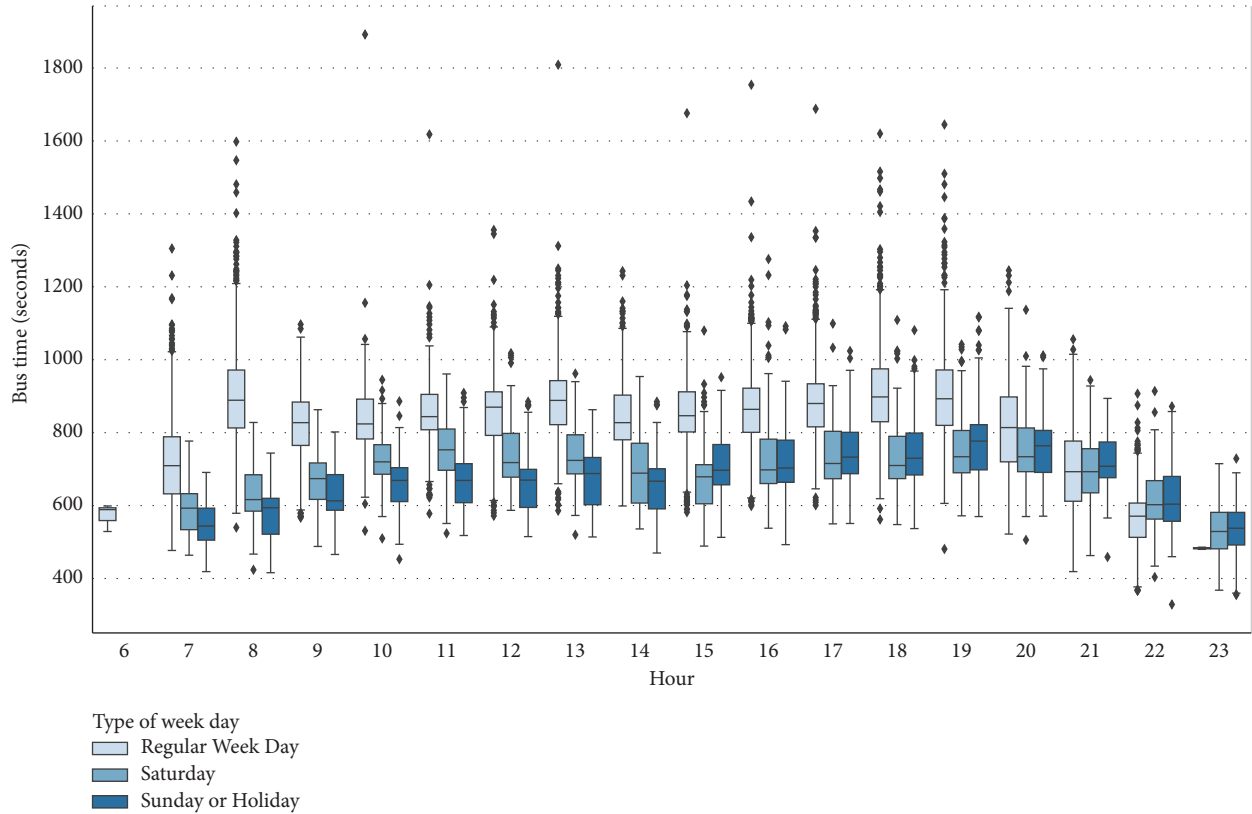
Figure 2: Bus travel time variations by hour considering the type of day.

with the data of the case studied. A k-fold cross-validation method should be applied to verify the consistency of the model. Furthermore, it is necessary to ensure that there are no multicollinearity problems by calculating the variance inflation factor (VIF) and discarding other issues that can affect the robustness of the models. The results for the case study presented are shown and discussed in Section 5. For the variables considered in this research, the MLR model leads to the following equation to determine the estimated bus travel time (E_BTT):

$$\text{E\_BTT} = \beta_0 + \beta_1 \, \text{GTT} + \beta_2 \, \text{SR} + \beta_3 \text{PR} + \beta_4 \, \text{TFR} \\ + \beta_5 \, \text{ILOP} + \beta_6 \, \text{AR}. \tag{1}$$

4.2. Estimated Bus Delay.

The MLR model allows in determining the internal bus travel time (IBTT), that is, the bus travel time derived directly from bus operation, and, consequently, the bus delay due to general traffic in the current mixed traffic operating environment (light green in Figure 4).

To calculate this bus delay produced by the mixed-traffic environment, it is necessary to obtain the time employed in the internal bus operations that will continue to be present in a DBL environment. Equation (1) is used to calculate the internal bus travel time for each trip (IBTT, under free-flow conditions) by equaling the independent variables related to general traffic (TFR and ILOP) to zero and considering the first decile value of GTT ($\text{GTT}_{FF} = 307.23$ s). This value of GTT reflects situations of free-flow for general traffic without

interference from other vehicles in bus travel time. The IBTT under these assumptions is formulated in equation (2). This approach is reflected in the first and second boxes of the light green zone of Figure 4.

$$\text{IBTT} = \beta_0 + \beta_1 \, \text{GTT}_{FF} + \beta_2 \, \text{SR} + \beta_3 \text{PR} + \beta_6 \, \text{AR}. \tag{2}$$

The estimated bus delay per bus trip is obtained by subtracting the IBTT from the E_BTT (equations (1) minus (2)), shown in the third and fourth boxes of the light green area in Figure 4. These data indicate how factors not related to the bus operation are detrimental to BTT according to their influence obtained in the model.

4.3. Estimated Bus Alightings.

The total user delay caused by bus delays (analyzed trip by trip) in the corridor under study should be estimated considering the actual boarding and alighting stops of each passenger on the line. There will be passengers who do not use the corridor at all, others who use it partially, and others who complete the whole corridor in their trip. Arias et al. [17] considered that the data of load or, at least, boardings and alightings at each stop are available. However, this is not common in most cities.

It is usual that the transit management system in mixed-traffic transit systems only registers boardings at each stop, while the information on alighting stops and the bus load between stops is not available. To solve this lack of information and determine the number of passengers affected by bus delays, alighting points can be estimated from
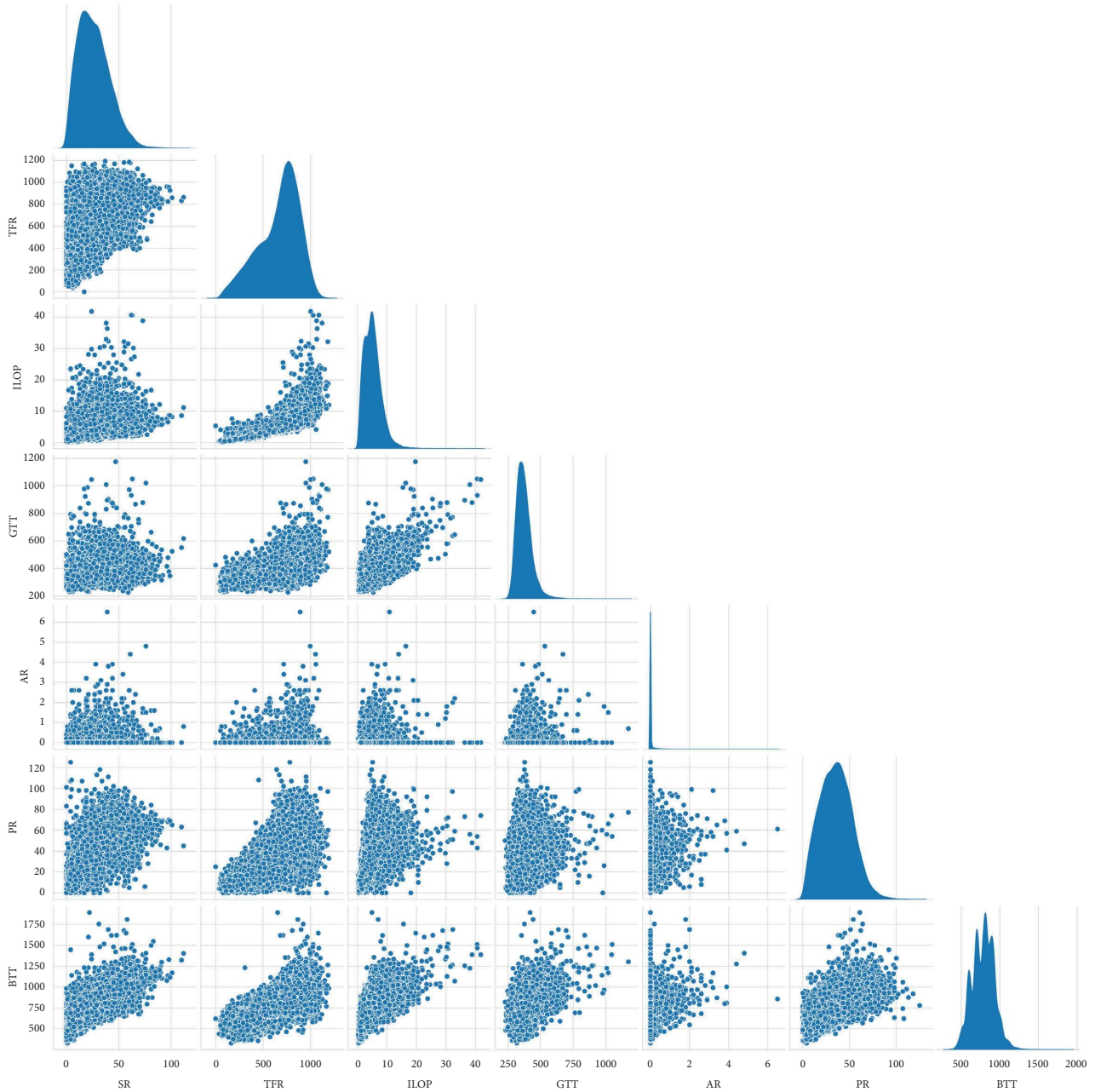
FIGURE 3: Scatter plot matrix.

TABLE 2: Pearson's correlation coefficients.

	SR	TFR	ILOP	GTT	AR	PR	BTT
SR	1						
TFR	0.550	1					
ILOP	0.444	0.796	1				
GTT	0.291	0.549	0.662	1			
AR	0.072	0.052	0.107	0.131	1		
PR	0.627	0.671	0.543	0.392	0.062	1	
BTT	0.703	0.673	0.626	0.477	0.096	0.623	1

smartcard uses data. For the case study of the present research, we developed a specific algorithm (yellow in Figure 4). The process was based on smartcard information for trips with exactly two uses in the network on the same day. If a round trip is made, the alighting stop is located near the boarding stop of the reverse trip, establishing certain conditions based on bus travel times and walking times. The algorithm considers the network as a whole, permitting return trips (the second trip of the day) on a viable different line. The consistency of each trip was verified by considering the entire network. The complete description of the algorithm is outside the scope of this paper. The application of the algorithm to the historical data of the analysis period allows obtaining an origin-destination (OD) matrix of the line. If the transit management system of a company provides origin-destination (OD) information, this step is not necessary to the application of our methodology.
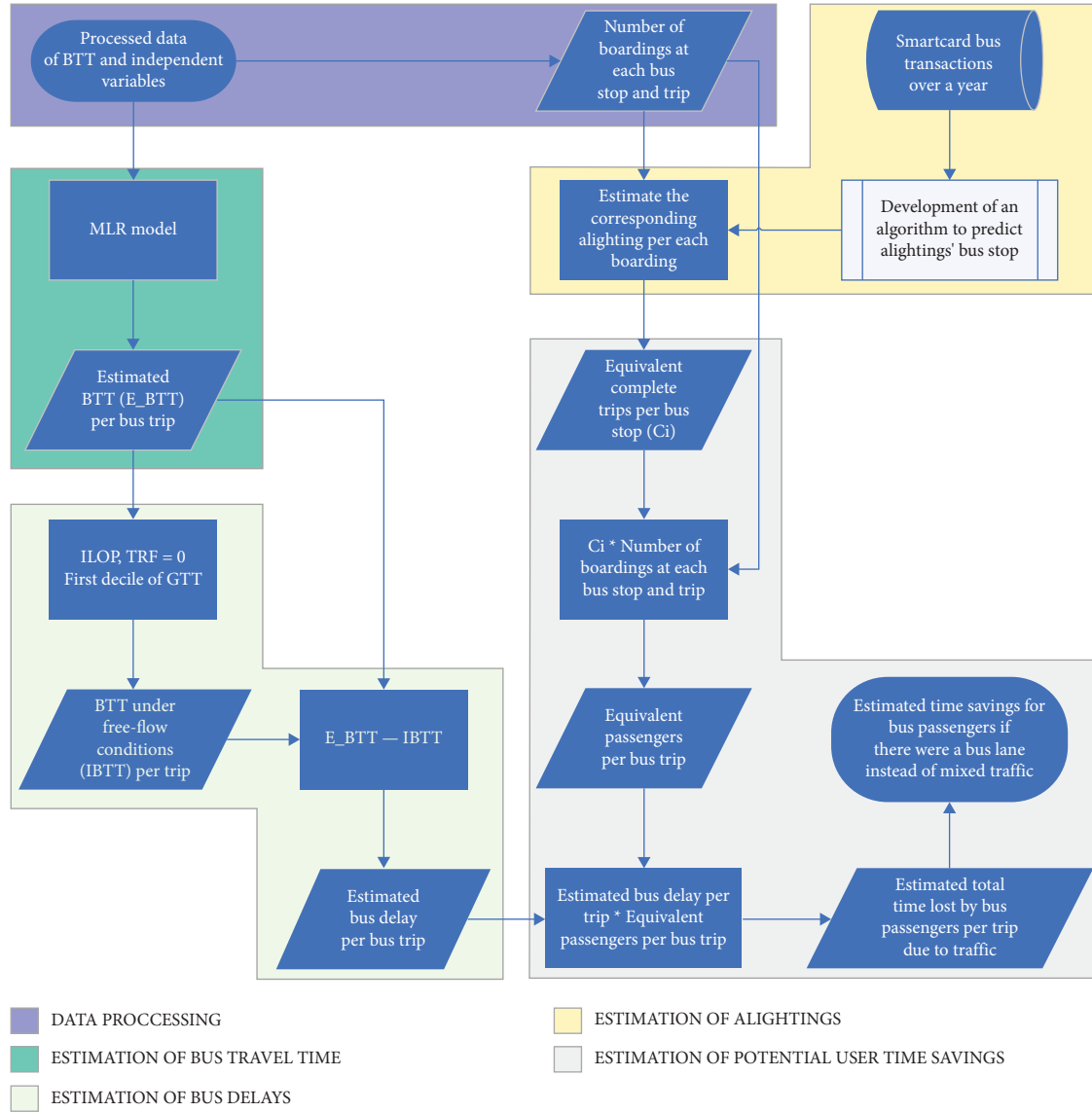
FIGURE 4: Flowchart of the methodology to estimate bus users time savings.

4.4. Estimated Time Savings for Bus Users. The OD matrix is used to estimate the share of the boardings at each stop of the line that alight at each of the other stops of the entire line. The proportion of users who board at stop i that alight at stop j (AS_{ij}) is obtained using equation (3), where B_i is the total number of boardings at stop i and T_{ij} are the trips from stop i to stop j in the OD matrix obtained in the previous step.

$$AS_{ij} = \frac{T_{ij}}{B_i}; B_i = \sum_j T_{ij}. \tag{3}$$

With this information, it is possible to estimate the average percentage of the corridor that a passenger who boards at stop i will travel. In the first place, with the simplified assumption that stops were evenly spaced along the considered corridor, the proportion of use of the corridor (CP_{ij}) for each i–j pair was established using equation

(4). In this equation, k_1 is the order number of the stop at which the stream begins, and k_2 is the order number of the stop where the corridor ends.

$$CP_{ij} = \begin{cases} 0, & \text{if } j \le k_1 \text{ or } i \ge k_2, \\ \dfrac{\min(j, k_2) - \max(i, k_1)}{k_2 - k_1}, & \text{otherwise.} \end{cases} \tag{4}$$

In second place, an equivalent coefficient per stop is calculated (C_i (first box of the grey area in Figure 4)) (equation (5)). This coefficient indicates the average percentage of the length of the corridor travelled by passengers boarding at each stop, considering the data of the estimated share of alightings obtained from the algorithm. For example, if half of the users that board at a stop complete the corridor (i.e., alight at the last stop of the corridor or later)

and the rest travel only through 60%, the equivalent coefficient is 0.8.

$$C_i = \sum_j \text{AS}_{ij} \cdot \text{CP}_{ij}. \tag{5}$$

The total number of equivalent travelers in the selected stream is obtained, for each vehicle trip n by multiplying the measured number of boardings at each bus stop i (B_{in}) times the coefficient (C_i) and adding the results for all the stops (second and third boxes of the grey area in Figure 4). This can be interpreted as the number of users of bus n that would traverse the entire corridor in which the DBL is implemented without necessarily having to be an integer.

When the estimated bus delay and equivalent travelers per vehicle in the stream are determined, the total time lost by passengers per trip could be obtained by multiplying both data (fourth box of the grey area in Figure 4). Adding the values for all the trips provides the final result of estimated time savings for bus passengers in the analysis period if there were a DBL instead of mixed traffic (last box of the grey area in Figure 4).

5. Results and Discussion

The methodology developed in this study (Figure 4) and described in Section 4 has been applied to the case study presented in Section 3. The aims were, on the one hand, to determine the delay caused by the influence of general traffic on the bus travel time on this specific corridor and, on the other hand, to calculate time savings for line 14 bus users if a bus lane is implemented in the entire studied stream between bus stops 119 and 559 (Figure 1).

After processing more than 20.5 million data related to traffic, an MLR model was performed considering 21,591 bus trips, and the results are shown in Table 3. All the independent variables were significant at the 99% level of confidence, and the model explained 64.40% of the variation in BTT. To guarantee that correlation between variables does not compromise the robustness of the model, the VIF was calculated, and multicollinearity problems do not exist (all VIF <3.70). The absence of problems of heteroscedasticity of residuals has been checked by the analysis of the scatterplot of residuals vs. E_BTT and verifying that "robust standard errors" led to the same conclusions about significant variables. The Durbin-Watson test did not indicate a lack of independence of the residuals. Due to the large sample size, the normality assumption of the residuals is not necessary to perform the usual test on the coefficients [35–37].

The results indicated that, holding the rest of the variables constant at their means, a single boarding in the stream would increase bus travel time by 4.23 s. Following this reasoning, one more point in the ILOP percentage would increase bus travel time by 7.96 s. In McKnight et al. [26], the impact of car travel time on bus travel time was more than twice larger than that of boardings per mile and bus stops per mile. As Table 3 shows, the influence of SR is more than four times larger than GTT, considering the standardized coefficient (SC), and more than twice as much as any other variable. In our model, the influence of traffic on BTT is reflected by three variables: TFR, ILOP, and GTT. The combined influence of the three had a similar weight to that of SR. McKnight et al. [26] concluded that bus travel time increases proportionally with car travel time and suggested that bus operations can benefit from an improvement in car traffic flow. However, in comparison with the present methodology, they did not employ big data or average speed obtained from sensors the car travel time did not match the specific bus trip (and therefore they did not calculate potential users' time savings) and they did not consider traffic variables (flow rate and occupancy).

AR is the variable with the least impact on BTT. Note that the influence of AR on GTT is already present in the GTT value provided by Bluetooth sensors; therefore, the AR coefficient of the model only reflects its influence on the internal bus travel time. Mazloumi et al. [38] studied weather conditions, but, unlike in this research, they were not significant and were not included in the subsequent analysis.

To verify the consistency of the model, a k-fold cross-validation (CV) method was applied. For each iteration, one-fold was used to test the model and the remainder to train it. The results of each iteration of this 10-fold CV are shown in Table 4. The coherence among the MLR coefficients for all the iterations, as well as the similarity between R^2 for test and training samples for each of them, guaranteed the reliability of the model and rejected the existence of overfitting or selection bias. Table 4 shows the MAPEs (mean absolute percentage error) on the test and training samples, with values under 8.90% in every training sample and under 9.07% in every test sample. The MAPE values are considered a sign of the goodness of fit of the model.

After the validation of the linear regression, the estimated bus travel time (E_BTT) is calculated by applying equation (1). The next step consists of estimating bus delays per trip obtained as the difference between the E_BTT and the internal bus travel time (IBTT) (equation (2)). The IBTT result considers the travel time spent by the bus to go through the corridor in very low traffic conditions (which in Arias et al. [17] was approximated as the minimum recorded time). Nevertheless, our approach is more accurate as it takes into account the operating time at stops, which can be high at peak hours when many passengers benefit from the exclusive bus lane (avoiding the overestimation of the suppressed delay which could derive from not considering this aspect). It also considers the additional influence of weather conditions, which is not usually considered in this kind of models. An increase in ridership may be expected if preferential treatment is given to the bus. Equation (2) can be used to determine its effect on the internal bus travel time, should an estimation of new values of SR and PR exist.

After subtracting the IBTT from the E_BTT, bus delay is obtained. The results reveal that, for the 21,591 bus vehicle trips considered for this case study, 133.75 s per bus trip were lost on average in the analyzed stream due to the effect of traffic (17% of the average bus travel time). This result is in line with the ones from Levinson [1], who concluded that between 12% and 26% of bus travel time was caused by traffic delays, depending on the location of the corridor (26% for CBD, which is not the case of the analyzed corridor, and 15%

Table 3: Regression results using the ordinary least squares model.

		Coef	Std. error	SC	R^2	N
					0.644	21,591
β_0	Constant	438.478*	4.253			
β_1	Stream ridership (SR)	4.238*	0.052	0.440		
β_2	Traffic flow rate (TFR)	0.116*	0.005	0.174		
β_3	Inductive loop occupancy (ILOP)	7.964*	0.352	0.170		
β_4	General traffic time (GTT)	0.231*	0.013	0.100		
β_5	Accumulated rainfall (AR)	14.112*	3.301	0.018		
β_6	Previous ridership (PR)	0.822*	0.051	0.098		

*Coefficients are significant at the 99% level.

Table 4: Results of the 10-fold cross-validation.

Cases	SR	TFR	ILOP	GTT	AR	PR	Intercept	R^2 on test samples	R^2 on training samples	MAPE on test samples (%)	MAPE on training samples (%)
0	4.229	0.121	7.599	0.227	10.826	0.817	438.956	0.659	0.642	8.826	8.877
1	4.235	0.117	7.926	0.233	14.013	0.815	437.768	0.638	0.644	9.065	8.854
2	4.226	0.112	8.153	0.238	16.096	0.825	437.915	0.669	0.641	8.503	8.913
3	4.207	0.118	7.832	0.238	12.616	0.828	435.458	0.644	0.644	8.895	8.862
4	4.252	0.113	8.203	0.229	14.142	0.833	439.345	0.638	0.644	9.031	8.857
5	4.243	0.112	8.261	0.223	16.329	0.834	441.234	0.645	0.643	8.771	8.879
6	4.223	0.117	7.733	0.224	15.395	0.833	441.317	0.673	0.640	8.673	8.897
7	4.250	0.116	8.097	0.236	15.227	0.801	436.627	0.610	0.647	8.956	8.862
8	4.265	0.117	8.002	0.225	8.408	0.802	440.042	0.621	0.646	8.990	8.859
9	4.244	0.116	7.828	0.237	16.115	0.830	436.307	0.632	0.645	9.047	8.853

and 16.7% for city and suburbs). However, our model allows us to determine the specific delay of each bus trip, showing high variability (Figure 5).

The use of mean values of bus delays and ridership to estimate time savings can bias the results. It is common for the largest number of transit users to coincide with the highest traffic values (see Figure 3), who experience the greatest delays, as it was observed in this study. These peak hour users are a target for sustainable mobility policies and would benefit the most from the implementation of a bus lane. Nevertheless, not all travelers will obtain the same savings, as it depends on the actual stretch of the line they travel on. The proposed methodology improves on the previous approaches by considering the estimated delay of each bus journey and the boarding and estimated alighting stops of each passenger using the large amount of information collected by the transit operator as well as a specific algorithm already described in the previous section. The calculation considers, on the one hand, the specific conditions of general traffic, weather, and transit demand in each section at the time that the passenger makes the trip and, on the other hand, an estimation of the part of the corridor that he/she has travelled. This could be of high relevance for an accurate evaluation in the decision-making process.

An OD matrix for line 14 was obtained and included 619,691 actual trips measured that fulfilled algorithm conditions. This information was used to estimate the share of
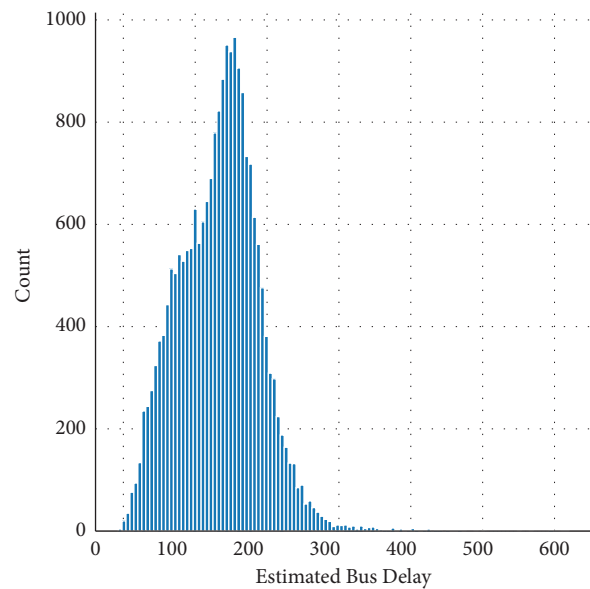

Figure 5: Estimated bus delay (s).

the boardings at each stop that alighted at each of the other stops in the corridor for the 21,591 bus trips analyzed. Since the corresponding alighting per each boarding were estimated for the studied corridor, the coefficient related to the equivalent complete trips per bus stop is calculated. The

TABLE 5: Coefficient related to the number of equivalent complete trips per bus stop i before and along the corridor.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Bus stop	**15**	**508**	**509**	**18**	**19**	**55**	**20**	**21**	**268**	**40**	**8**	**9**	**10**	**269**	*119*	*120*	*121*	*122*	*123*	*124*	*125*	*126*	*127*	*559*
C_i	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}	C_{17}	C_{18}	C_{19}	C_{20}	C_{21}	C_{22}	C_{23}	C_{24}
	0.439	0.497	0.404	0.436	0.398	0.426	0.483	0.521	0.631	0.722	0.719	0.750	0.734	0.686	0.811	0.728	0.656	0.601	0.530	0.431	0.327	0.220	0.110	0.000

*Bold for the stops before the studied corridor and italic for the ones along it.
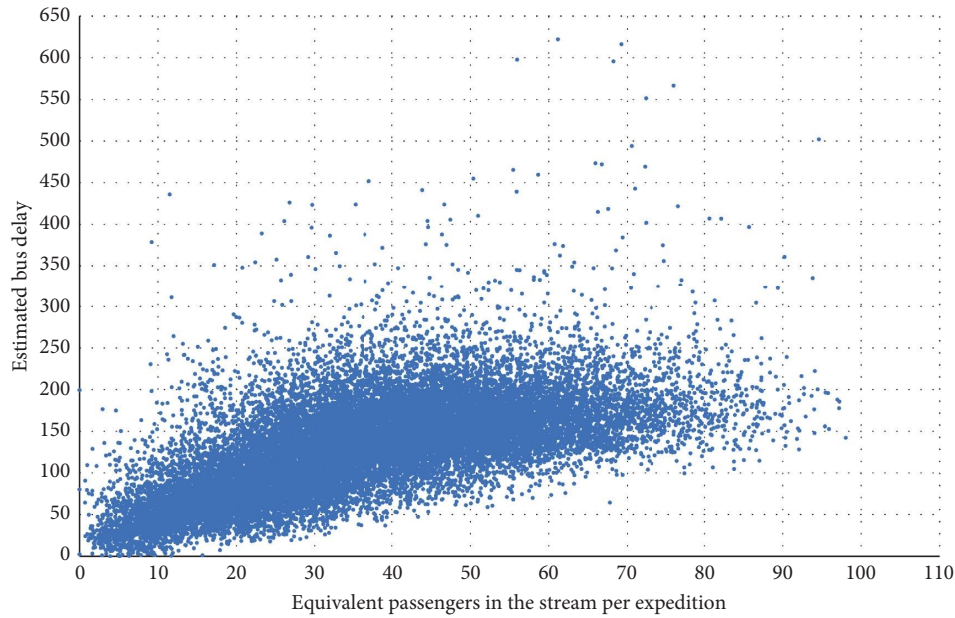
FIGURE 6: Relationship between estimated bus delays and equivalent travelers per expedition in the stream.

results are displayed in Table 5 (bold for the stops before the studied corridor and italic for the ones along it). For example, for the passengers who boards at stop number 15 (first of the line), the matrix showed that 34.17% do not use the corridor, 18.75% travel the entire corridor and the remaining 47.08% partly use it (e.g., 16.48% alight at stop 123); while for the passengers who board at stop number 119 (15th of the line and first of the corridor), 44.43% travel along the entire corridor, and the rest alight before (e.g., 18.91% alight at stop 123). According to the detailed shares obtained and considering a DBL implemented in the entire studied corridor, a passenger who boards at stop number 15 is expected to travel 43.9% (C_1) of the DBL, while a passenger who boards at stop number 119 is expected to travel 81.1% (C_{15}) of the DBL.

As stated in the previous section, the total number of equivalent passengers per bus trip is obtained by multiplying these coefficients C_i by the corresponding number of boardings at each bus stop i per trip. The relationship between the number of equivalent passengers in the stream and the estimated bus delay per expedition is shown in Figure 6. As expected, the results indicated that the largest estimated bus delays due to mixed traffic generally occurred on bus vehicle trips with the largest ridership (higher number of equivalent bus users in the stream per bus trip). This means that higher bus delays per trip would affect more bus passengers per trip.

Finally, by multiplying the bus delay and equivalent users per vehicle in the stream, the estimated total time lost by bus passengers is obtained, and therefore, the time savings for line 14 bus passengers if there were a bus exclusive lane instead of mixed traffic. The results show that, adding the estimated time wasted by bus users per trip and considering the 21,591 vehicle trips studied in this research, a total of 33,756.62 h were lost by bus passengers in 2019 in the stream analyzed due to the bus delays caused by the lack of a DBL.

In the 2,448 m long corridor of the case study, an exclusive bus lane would reduce bus travel time by more than 2 min on average per trip. 4,046 bus vehicle trips (18.74%) were observed to undergo a delay exceeding 3 min, which represents 22.93% of the average bus travel time (784.91 s) and an increase of 27.64% of the average bus travel time under free-flow conditions (651.16 s). 33 bus vehicle trips had a bus delay longer than 400 s.

6. Conclusions and Future Work

Exclusive bus lanes are a relevant measure to improve transit performance and to attract new passengers. The assessment of time savings is crucial for councils or transit agencies to decide on whether and where an exclusive or dedicated bus lane (DBL) should be implemented based on the benefits it would provide. Nevertheless, the cost-benefit analysis of DBLs often lacks information about the actual number of affected bus users and time savings, which weakens the arguments in favor of bus lanes. The methodology presented in this paper, of the general applicability, allows the accurate calculation of these time savings and is thus a tool to promote more sustainable mobility.

There are many methods, both by analytical and simulation procedures, that allow bus travel times to be calculated in different situations. However, it is difficult for these methods to gather the wide variety of geometric configurations, signalization, driving habits, stop operations, and traffic and weather situations that influence the travel times of general traffic and buses. Therefore, these methods do not allow the estimation of the delay derived from sharing lanes for each of the buses that circulate during the year, in their specific circumstances. It is not accurate to consider that the bus trip time in the bus lane is that of the fastest trips or to assume that a free flow will be performed without any type of delay (for example, at intersections).

Nowadays, there is a large amount of information that is automatically registered in cities. New technologies that allow traffic speeds to be collected, such as Bluetooth sensors or license plate recognition, are being added to the usual data from loops, AVL, APC, or weather stations. All these systems provide huge amounts of information on the corridor operation, considering the local trait and the wide variety of specific characteristics mentioned in the previous paragraph. These types of data have already been used, for example, to estimate arrival times at stops.

Using this wealth of information, this paper has presented a new methodology to quantify the delay in bus travel time caused by the influence of general traffic and to estimate the potential time savings for bus users if an exclusive bus lane is implemented. The proposal is based on analyzing the influence of different factors on the travel time obtaining a model specifically adapted to the local circumstances of the analyzed corridor (considering the spatial and temporal variations). In this way, it is possible to estimate the internal operating times of the public transport system for each bus trip (highly influenced by its actual ridership), separately from those caused by the coexistence with other vehicles on the street. It also allows us to consider the influence of rainy weather with the frequency and intensity actually present throughout the year. In this paper, a linear regression model is proposed for this analysis, but other modelling strategies adapted to the circumstances of each study can be developed with the same approach.

As it has been shown in the case study, bus delays present a great variability, from buses in which the bus lane will not lead to improvements in travel time to others that will avoid long delays. Each of those buses will have a different number of passengers. The potential time savings of the specific passengers using each bus can be calculated once the specific delay of each trip is determined. In this way, our methodology avoids the underestimation of users' time savings that would derive from the simplified approach of multiplying the average number of passengers in the corridor by the average delay of a bus trip. In the case of public transport systems that collect data on the specific stops where each user boards and alights, this can be done directly. In many cases, the alighting information is not available, but data from smartcard cancellations is, which allows the estimation of the alighting stops through the application of suitable algorithms. In this paper, this last approach has been proposed, allowing for consideration of the actual length of the affected corridor that each passenger goes through.

The findings of this study have to be seen in light of some limitations. Although it is not very common, if there is direct information on alightings, several improvements in model construction can be performed. The alightings information will allow for a more accurate estimation of the time spent serving stops and the influence of crowding consideration of previous ridership will not be necessary in that case. It will also allow the direct estimation of trip lengths travelled. If the transit management system provides information about dwell time, a more accurate estimation of internal times can be performed. In future work, traffic light priority for the bus could also be considered. Some assumptions have been made

as well. We have assumed that there is no fare evasion due to the exhaustive fare control exerted by the driver, as boarding is only allowed through the front door. In case where there is a significant percentage of fare evasion, it should be considered to avoid underestimating of passengers. In addition, to estimate the potential time savings in 2019, it has been assumed that the number of passengers on the line would be the same, as well as the number of bus trips and stops. The implementation of the DBL leads to travel time savings and it may produce an increase in ridership, so we have remained on the conservative side, and higher time savings may be expected. In future work, if there are measurements before and after the installation of a dedicated bus lane, validation of equation (2) should be performed.

Despite these limitations, our methodology is an improvement of the existing academic approaches for the determination of the benefits of an exclusive or dedicated bus lane, which is highly valuable for decision-making and justification of this kind of infrastructure.

Data Availability

Due to confidentiality issues, the raw data of this study cannot be shared.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

References

[1] H. S. Levinson, "Analyzing transit travel time performance," *Transportation Research Record*, pp. 1–6, 1983.

[2] J. de Dios Ortúzar and L. G. Willumsen, *Discrete Choice Models, Modelling Transport*, John Wiley & Sons, 2011.

[3] National Academies of Sciences, *Engineering, and Medicine, Transit Capacity and Quality of Service Manual*, National Academy Press, Washington, DC, USA, 3 edition, 2013.

[4] A. Comi, A. Nuzzolo, S. Brinchi, and R. Verghini, "Bus travel time variability: some experimental evidences," *Transportation Research Procedia*, vol. 27, pp. 101–108, 2017.

[5] E. Echaniz, R. Cordera, A. Rodriguez, S. Nogués, P. Coppola, and L. dell'Olio, "Spatial and temporal variation of user satisfaction in public transport systems," *Transport Policy*, vol. 117, pp. 88–97, 2022.

[6] N. Paulley, R. Balcombe, R. Mackett et al., "The demand for public transport: the effects of fares, quality of service, income and car ownership," *Transport Policy*, vol. 13, no. 4, pp. 295–306, 2006.

[7] J. Allen, J. C. Muñoz, and J. de Dios Ortúzar, "On the effect of operational service attributes on transit satisfaction," *Transportation*, vol. 47, no. 5, pp. 2307–2336, 2020.

[8] L. dell'Olio, A. Ibeas, and P. Cecín, "Modelling user perception of bus transit quality," *Transport Policy*, vol. 17, no. 6, pp. 388–397, 2010.

[9] D. A. Hensher, J. M. Rose, W. Leong, A. Tirachini, and Z. Li, "Choosing public transport—incorporating richer behavioural elements in modal choice models," *Transport Reviews*, vol. 33, no. 1, pp. 92–106, 2013.

[10] J. Viegas and B. Lu, "Widening the scope for bus priority with intermittent bus lanes," *Transportation Planning and Technology*, vol. 24, no. 2, pp. 87–110, 2001.

[11] M. Eichler and C. F. Daganzo, "Bus lanes with intermittent priority: strategy formulae and an evaluation," *Transportation Research Part B: Methodological*, vol. 40, no. 9, pp. 731–744, 2006.

[12] A. Russo, M. W. Adler, and J. N. van Ommeren, "Dedicated bus lanes, bus speed and traffic congestion in Rome," *Transportation Research Part A: Policy and Practice*, vol. 160, pp. 298–310, 2022.

[13] A. Gan, H. Yue, I. Ubaka, and F. Zhao, "Development of operational performance and decision models for arterial bus lanes," *Transportation Research Record*, vol. 1858, no. 1, pp. 18–30, 2003.

[14] G. Yang, D. Wang, and X. Mao, "Modelling the modal shift effects of converting a general traffic lane into a dedicated bus lane," *Promet - Traffic & Transportation*, vol. 32, no. 5, pp. 625–637, 2020.

[15] A. Kampouri, I. Politis, and G. Georgiadis, "A system-optimum approach for bus lanes dynamically activated by road traffic," *Research in Transportation Economics*, vol. 92, Article ID 101075, 2022.

[16] J. Surprenant-Legault and A. M. El-Geneidy, "Introduction of reserved bus lane: impact on bus running time and on-time performance," *Transportation Research Record*, vol. 2218, no. 1, pp. 10–18, 2011.

[17] D. Arias, K. Todd, J. Krieger et al., "Using gtfs to calculate travel time savings potential of bus preferential treatments," *Transportation Research Record*, vol. 2675, no. 9, pp. 1643–1654, 2021.

[18] T. Tsubota, A. Bhaskar, E. Chung, and R. Billot, "Arterial traffic congestion analysis using bluetooth duration data," *Proceedings of the 34th Australasian Transport Research Forum (ATRF)*, pp. 1–14, 2011.

[19] F. S. Imam, A. Bhaskar, and E. Chung, "Integrating bluetooth and smart card data for better estimation and prediction of bus speed on arterial corridors with low frequency buses," *Proceedings of the 39th Australasian Transport Research Forum (ATRF)*, pp. 1–15, 2017.

[20] R. Zhang, W. Liu, Y. Jia et al., "WiFi sensing-based real-time bus tracking and arrival time prediction in urban environments," *IEEE Sensors Journal*, vol. 18, no. 11, pp. 4746–4760, 2018.

[21] T. F. Welch and A. Widita, "Big data in public transportation: a review of sources and methods," *Transport Reviews*, vol. 39, no. 6, pp. 795–818, 2019.

[22] Y. Wang, S. Ram, F. Currim, E. Dantas, and L. A. Sabóia, "A big data approach for smart transportation management on bus network," in *Proceedings of the 2016 IEEE international smart cities conference (ISC2)*, pp. 1–6, Trento, Italy, September 2016.

[23] Y. Ou, A.-S. Mihăiță, and F. Chen, "Big data processing and analysis on the impact of COVID-19 on public transport delay," *Data Science for COVID-19*, vol. 2, pp. 257–278, 2022.

[24] T. Campisi, S. Basbas, M. A. Al-Rashid, G. Tesoriere, and G. Georgiadis, "A region-wide survey on emotional and psychological impacts of COVID-19 on public transport choices in Sicily, Italy," *Transactions on Transport Sciences*, vol. 12, no. 3, pp. 34–43, 2022.

[25] A. Tirachini, "Estimation of travel time and the benefits of upgrading the fare payment technology in urban bus services," *Transportation Research Part C: Emerging Technologies*, vol. 30, pp. 239–256, 2013.

[26] C. E. McKnight, H. S. Levinson, K. Ozbay, C. Kamga, and R. E. Paaswell, "Impact of traffic congestion on bus travel time in northern New Jersey," *Transportation Research Record*, vol. 1884, no. 1, pp. 27–35, 2004.

[27] M. Hofmann and M. O'Mahony, "The impact of adverse weather conditions on urban bus performance measures," *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005*, pp. 84–89, 2005.

[28] J. Anta, J. B. Pérez-López, A. Martínez-Pardo, M. Novales, and A. Orro, "Influence of the weather on mode choice in corridors with time-varying congestion: a mixed data study," *Transportation*, vol. 43, no. 2, pp. 337–355, 2016.

[29] P. Arana, S. Cabezudo, and M. Peñalba, "Influence of weather conditions on transit ridership: a statistical study using data from Smartcards," *Transportation Research Part A: Policy and Practice*, vol. 59, pp. 1–12, 2014.

[30] Y. Wang, Y. Bie, and Q. An, "Impacts of winter weather on bus travel time in cold regions: case study of Harbin, China," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 11, 2018.

[31] M. Novales, A. Orro, J. B. Pérez-López, J. Feal, and M. R. Bugarín, "Increasing boarding lost time at regular bus stops during rainy conditions: a case study," *Journal of Public Transportation*, vol. 23, no. 1, pp. 63–80, 2021.

[32] Instituto Nacional de Estadística (INE), 2022, https://www.ine.es/index.htm.

[33] A. Orro, M. Novales, Á. Monteagudo, J. B. Pérez-López, and M. R. Bugarín, "Impact on city bus transit services of the COVID-19 lockdown and return to the new normal: the case of A Coruña (Spain)," *Sustainability*, vol. 12, no. 17, p. 7206, 2020.

[34] MeteoGalicia, 2022, https://www.meteogalicia.gal/observacion/estacionshistorico/historico.action?idEst=10157.

[35] A. F. Schmidt and C. Finan, "Linear regression and the normality assumption," *Journal of Clinical Epidemiology*, vol. 98, pp. 146–151, 2018.

[36] W. D. Berry and S. Feldman, *Multiple Regression in Practice, No. 50*, Sage, Thousand Oaks, CA, USA, 1985.

[37] U. Knief and W. Forstmeier, "Violating the normality assumption may be the lesser of two evils," *Behavior Research Methods*, vol. 53, no. 6, pp. 2576–2590, 2021.

[38] E. Mazloumi, G. Rose, G. Currie, and M. Sarvi, "An integrated framework to predict bus travel time and its variability using traffic flow data," *Journal of Intelligent Transportation Systems*, vol. 15, no. 2, pp. 75–90, 2011.