





# Bus Journey Time Prediction: A Comparison of Whole Route and Segment Journey Time Predictions Using Machine Learning

Laura Dunne<sup>(✉)</sup>  and Gavin McArdle 

School of Computer Science, University College Dublin, Belfield Dublin 4, Ireland  
laura.dunne2@ucdconnect.ie, gavin.mcardle@ucd.ie

**Abstract.** Accurately predicted bus journey times are essential for bus network reliability and making bus transport attractive. The most common approach when predicting bus journey times with machine learning (ML) is to predict journey times for each stop pair segment. Segment data can be very noisy, leading to inaccuracies. To investigate this, this paper compares the classic stop pair segment approach to three other methods. Firstly, a naive method of calculated historical averages is introduced as a baseline. We then explore two methods based on predicting the whole bus route journey time from origin to terminus. To estimate a passenger's journey, where the whole route is not travelled, we estimate the proportion of the whole journey time the passenger's journey will take. The first of these methods calculates this proportion from similar historical journeys, and the second proposed method trains an ML model to predict this proportion for each segment of the passenger's journey. The results show that this novel proposed approach results in less error across most metrics, when compared to the segment prediction method. An interesting insight from the analysis shows the proposed approach has enhanced benefits during peak travel time and during the working week. Gains in prediction accuracy at these times would benefit the most commuters. This research can be applied to make robust scheduling decisions that will increase bus network reliability, improve bus network satisfaction and uptake, and lead to more sustainable cities.

**Keywords:** bus journey time prediction · machine learning · random forest

## 1 Introduction

Scheduled bus services are an important component of the transportation network in an increasingly urban world. Many urban centres are rapidly expanding and have exceeded the road and parking infrastructure for every inhabitant to have a private car [11] even if air quality issues could be overcome by increased

or complete electrification of personal vehicles. As stated by UN Sustainable Goal 11.2 [26], more sustainable forms of transport are necessary for continued development. Walking and cycling are ideal for shorter journeys in mild weather, as they require minimal infrastructure and have health benefits associated with active transportation [17]. For many reasons, much of the population will continue to depend on public bus and rail services. Buses have several advantages over rail transport; buses are cheaper for the same quality of service [2], require far less infrastructure, and are more flexible as buses can easily be rerouted as urban centres evolve. However, buses are not without limitations and often lack reliability compared to rail services. They operate in more complex and less controlled environments than trains, stopping more regularly and interacting with other traffic and cyclists. Many governments have policies encouraging switches to more sustainable transport since the Paris Agreement in 2015 [4]. Many of these policies focus on improving the bus network to promote its use. Since resources are not infinite, the bus services must also be optimised within the existing infrastructure. The most significant factor for passengers is a low waiting time [19] which is directly related to the reliability of the bus network [23], and accurately predicting journey times for bus scheduling and real-time passenger information (RTPI) is key to a reliable bus network [3].

The prediction of bus journey times is the subject of much research. The main techniques used are simple historical averages [15], statistical methods including regression models, Kalman filters and Machine Learning (ML) [21]. The literature in this area has several weaknesses; there is a lack of standard syntax, and the studies tend to be small due to the complexity of bus data [18,21,27]. There is also no standard benchmark dataset to allow comparisons between studies. Comparing one study with another is usually impossible as bus routes have different characteristics, affecting the error metrics. The longer the bus route, the higher the absolute error will be [10]. Bus routes on networks with low reliability will have a higher level of irreducible error regardless of the prediction methods employed [8,21].

A common approach in many studies that predict bus journey times with ML is training multiple models for each consecutive stop pair segment. Generally, a model is built to predict the journey time for the segment between every two consecutive stops on the bus network [7,10,16,20,21,27]. This approach will result in a number of models that is one less than the number of stops on the route. There are many reasons for this segment prediction approach: stops are where bus arrival times are often monitored with Automatic Vehicle Location (AVL) systems and are typically the only place passengers can embark and disembark. Stop arrival time is most relevant to the service user, and it is a natural and intuitive way to conceptualise bus routes. However, real-world data is messy, and the measurement of bus location can be inexact. The GPS readings themselves are not exact and have reported errors up to 30m depending on the age of the GPS unit, local conditions and the speed of the bus. [28]. Many GPS units deployed on buses are old, and buses often operate in densely urban areas. The presence of tall buildings, or the so-called 'urban canyons', is known to

impact the accuracy of GPS readings [14]. A recent study [25] showed a 13-second discrepancy between the GPS recorded time of arrival of a bus at bus stops and the actual time of arrival. There is also the compounding factor of the frequency of recording of position. Often this is around 30 s but can be longer, so the timing of arrival at stops is often interpolated [22]. These factors will likely create significant noise in the data regarding journey times between individual stops, increasing errors in the predictions from models trained on segment data. We observed that studies that predicted both longer and shorter sections of the same route tended to have lower error metrics for longer sections [5, 6, 12, 20, 21]. We conducted a provisional experiment to test the theory that whole route prediction methods were more accurate than segment prediction methods. Two methods predicted the whole route journey time. The first method had a single Random Forest (RF) model trained on the whole journey times between the origin and terminus stop. The second method trained an RF model for each consecutive stop pair segment on the route and returned the sum of these models' predictions to estimate the whole journey time. The evaluation revealed the first method was superior across multiple error metrics. Predicting by segment resulted in a mean absolute error (MAE) of 286 s versus 266 s for whole journey prediction. Segment prediction had a mean absolute percentage error (MAPE) of 0.099 and a coefficient of determination ( $R^2$ ) of 0.877, and the corresponding values for whole journey prediction were 0.094 and 0.895, respectively.

These results motivated further exploration, as regardless of the accuracy of whole journey time predictions, they are only useful for predicting journey times for bus schedules. They are not useful for predicting individual passenger trip times (i.e. partial journeys). Partial journey predictions are needed for journey planning and RTPI. We sought to harness the accuracy of whole journey time predictions to improve partial journey time predictions. An experiment was designed to compare four methods of journey time prediction for several bus routes in Dublin, Ireland. The methods were a naive historical averages (HA) method, the most common method: segment prediction (SP) and two methods based on whole journey prediction. Both of these methods predict the whole route journey time and estimate the proportion of the whole route journey time the partial journey is likely to take. As described in [8], the first of these methods calculates the historical average proportion similarly to how HA calculates journey time and is called Whole Journey Prediction with Calculated Proportion (WJP-C). The novel second method uses an RF model to predict the proportion and this method is called Whole Journey Prediction with Predicted Proportion (WJP-P).

This paper makes the following contributions:

- Challenges the status quo regarding how bus routes are treated conceptually for ML modelling by assessing the SP method, an approach often used without much discussion of the rationale.
- Presents an approach for predicting partial journey times that significantly improves upon the SP method on most metrics with the consumption of similar computing resources.

- Performs deep analysis of the results of the four methods and examines the results by segment length, bus route, time of day and day of the week.

The remainder of this paper is organised as follows: Sect. 2 describes the data processing, Sect. 3 outlines the methodology used, Sect. 4 demonstrates the results and discusses the analysis, and Sect. 5 presents the conclusions of this study and discusses the planned further work.

## 2 Data

The National Transport Authority (NTA) in Ireland provided the historical bus data used in this study. As is typical with AVL data, there were some data quality issues, such as the bus stop arrival events not being recorded or duplicate arrival events at the same stop. As a result, not every unique trip had the same stop sequence; some trips were invalid and could not be included in the analysis. When selecting a subset of these routes for analysis, the routes with the highest quality data were desired. The inclusion criteria for this study were that the bus routes had at least 80% valid unique trips and at least 3000 unique trips in total. Based on these criteria, 16 bus routes were selected from the Dublin Bus network. These were eight head sign pairs for route numbers: 4, 27A, 32, 42, 56A, 79A, 120, and 184. The final dataset was all valid data for these routes for a year from January 1st to December 31st 2018. The Dublin Bus network in 2018 contained 253 routes and served over 1.3 million people in the Dublin area.

The most common stop sequence present on each route in the data was found, and this stop sequence was confirmed to be correct by comparison to the GTFS (General Transit Feed Specification) data published by Dublin Bus. The raw data was structured as bus arrival events at bus stops. Trips that contained extreme outliers, such as whole journey times or segments with journey time outliers greater than twelve standard deviations (SD) from the mean were removed, as were trips with impossible values such as negative journey times. Less than one and a half per cent of data was lost at this step. Additional features for the time, day of the week, and month were extracted from the timestamps. The time group feature had variable granularity. Peak travel periods are 30 min, and off-peak travel periods are 60 min long and are encoded from 0 to 29 starting at midnight. This was to avoid too coarse a granularity during rapidly changing peak travel periods yet allow for enough data in each time group during off-peak periods when there are fewer buses on the network. This approach was benchmarked against homogenous 30-minute and 60-minute granularities and was found to improve error metrics. Further details on the data cleansing procedure can be found in [8].

The cleaned and preprocessed data format is shown in Table 1. Following processing, the data was split into a training set and a testing set in the ratio of 85:15. This could not be done using a standard test/train split as many rows in the dataset refer to the same unique trip. To maintain data integrity, 15% of the trip IDs were randomly selected, and all rows with those trip IDs became the test set. The training set was then further processed in three ways. Firstly,

**Table 1.** Sample of the data after preprocessing. The stop arrival times are defined in seconds after midnight and journey time is in seconds. Day 0 is Monday and Month 1 is January. Time groups 9 and 10 are 30-minute periods during the morning peak travel period from 08:00 to 08:30 and 08:30 to 09:00

| TripID | First Stop | Second Stop | First Stop Arrival Time | Second Stop Arrival Time | Month | Day | Time Group | Journey Time |
|--------|------------|-------------|-------------------------|--------------------------|-------|-----|------------|--------------|
| 20858  | 324        | 327         | 30579                   | 30669                    | 1     | 0   | 9          | 90           |
| 20858  | 327        | 7113        | 30669                   | 30745                    | 1     | 0   | 10         | 76           |
| 20858  | 7113       | 127         | 30745                   | 30812                    | 1     | 0   | 10         | 67           |
| 20858  | 127        | 112         | 30812                   | 30877                    | 1     | 0   | 10         | 65           |
| 20858  | 112        | 113         | 30877                   | 30958                    | 1     | 0   | 10         | 81           |
| 20858  | 113        | 114         | 30958                   | 30964                    | 1     | 0   | 10         | 6            |

the training set is used to create a Reference Dataset. This contains the average journey time and the average proportion of the full journey for each unique segment on each route for each combination of time/day in our dataset. An example of the resulting Reference Dataset is shown in Table 2.

**Table 2.** Sample of the Reference Dataset. All of these samples are from Route 4 in direction 1, and take place on a Monday (Day = 0) between 13:00 and 17:30 in the afternoon (time groups 15 through 20) on the segment between stops 273 and 405. The Reference Dataset contains the mean journey time and mean proportion of the whole journey time for each day/time combination.

| Day of Week | Time Group | First Stop | Second Stop | Mean Journey Time | Mean Proportion |
|-------------|------------|------------|-------------|-------------------|-----------------|
| 0           | 15         | 273        | 405         | 249.79            | 0.058591        |
| 0           | 16         | 273        | 405         | 253.63            | 0.059462        |
| 0           | 17         | 273        | 405         | 251.87            | 0.059878        |
| 0           | 18         | 273        | 405         | 262.81            | 0.058941        |
| 0           | 19         | 273        | 405         | 257.38            | 0.059040        |
| 0           | 20         | 273        | 405         | 263.17            | 0.060619        |

Secondly, the training set undergoes further processing to train RF models to predict whole journey times. It is restructured to represent unique journeys instead of arrival events. All arrival events except for the first and last were dropped for predicting whole journeys. The target feature, the historical journey time, is calculated for each journey in the dataset. The resulting data structure is shown in Table 3 and contains three temporal features: month, day and time group.

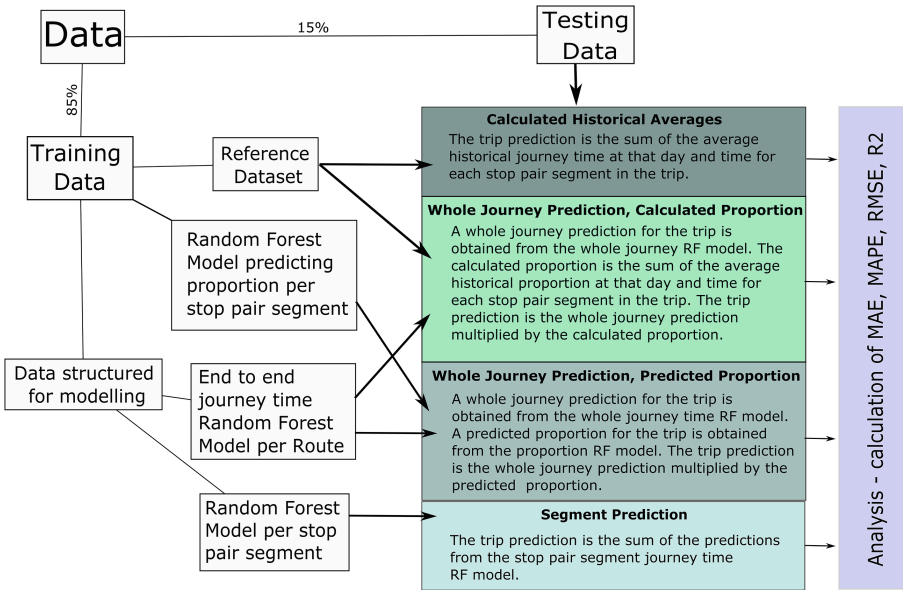
Traffic volume and passenger load have the most significant impacts on bus journey times [24]. These features are difficult to measure directly, but as they offer a cyclical pattern, they are encoded in temporal features [1]. The training data is also used to structure the data for segment prediction, similar to for whole journey predictions, but the details of the intermediate bus stops are not removed.

**Table 3.** Sample of the data prior to modelling

| Month | Day | Time Group | Total Journey Time |
|-------|-----|------------|--------------------|
| 3     | 3   | 10         | 5442               |
| 1     | 2   | 7          | 5097               |
| 1     | 2   | 15         | 4799               |
| 1     | 2   | 26         | 2807               |
| 4     | 4   | 19         | 5402               |

### 3 Methodology

Once the dataset was processed, four methods were implemented to predict journey time as shown in Fig. 1. These four methods are described in this section.



**Fig. 1.** Methodology Flow Diagram

#### 3.1 Historical Averages (HA)

The Reference Dataset described in Sect. 2 was used in two ways. Firstly, it was used in the naive baseline method, HA. To produce an estimate for a passenger’s partial trip time, this method references the Reference Dataset for each segment at the time and day that the partial trip occurs and sums these historical average

times to get a total. HA is shown in Eq. 1 where  $n$  is the number of segments in the trip and  $\bar{T}$  is the average historical journey time for the day of the week,  $d$  and time of day,  $t$  that the trip takes place.

$$\sum_{i=1}^n \bar{T}_{i,dt} \quad (1)$$

### 3.2 Whole Journey Prediction with Calculated Proportion (WJP-C)

Secondly, the Reference Dataset is used to calculate the proportion of the whole journey, the passenger's partial journey historically represented. The dataset is referenced for each segment on the partial journey at the time and day the trip takes place and the sum of these historical average proportions for the segments is returned. This value will always be a ratio between 0 and 1 depending on how much of the whole journey the passenger travels. It is multiplied by the prediction returned from the whole journey RF model to get an estimate for the passenger's journey time. That model has been trained on the whole journey dataset restructured for modelling as described in Fig. 3. RF was used throughout as it needs minimal hyperparameter tuning, is scalable and has previously been shown to be the best of the traditional ML algorithms for this dataset [8]. WJP-C is shown in Eq. 2 where  $\hat{W}$  is the whole journey time prediction from the random forest model,  $n$  is the number of segments in the trip and  $\bar{P}$  is the average historical proportion for the day of the week,  $d$  and time of day,  $t$  that the trip takes place.

$$\hat{W} \cdot \left( \sum_{i=1}^n \bar{P}_{i,dt} \right) \quad (2)$$

### 3.3 Whole Journey Prediction with Predicted Proportion (WJP-P)

An RF model is trained for each segment on the route that will predict the proportion of the whole journey that the segment will take. The training data is sequentially filtered to just the data corresponding to the relevant stop pair, and an RF model is trained for each pair. The number of models trained depends on the length of the route and is always one less than the number of stops. WJP-P is shown in Eq. 3 where  $\hat{W}$  is the whole journey time prediction from the RF model,  $n$  is the number of segments in the trip and  $\hat{P}$  is the predicted proportion.

$$\hat{W} \cdot \left( \sum_{i=1}^n \hat{P}_i \right) \quad (3)$$

### 3.4 Segment Prediction (SP)

In a similar way to how WJP-P builds a model to predict the proportion of each segment, the SP method sequentially filters the data to each consecutive stop pair segment and builds an RF to predict journey time, and like before, the number of models trained will be one less than the number of stops on the route. SP is shown in Eq. 4 where  $\hat{S}$  is the segment prediction from the RF model and  $n$  is the number of segments in the trip.

$$\sum_{i=1}^n \hat{S}_i \quad (4)$$

### 3.5 Testing

It was not possible to access real passenger journeys for this experiment, so a simulated passenger journey was extracted from each of the 21706 test journeys. A random sequence of stops was generated from each unique journey in our test set. This was achieved by choosing two random indices from a sequential list of stops for the route the unique journey is from. We check that the same index has not been chosen twice, and the lower index becomes the boarding stop and the higher index becomes the disembarking stop of the pseudo passenger. This is similar to the approach used in [8, 13]. Quotas were used to ensure the distribution of the length of the partial sample journeys was uniform from each route, with a similar number of journeys for each possible number of segments.

For each test journey, four predictions were made - one for each of the four methods. Analysis was then performed, including the calculation of MAE, MAPE, root mean squared error (RMSE), mean percentage error (MPE) and  $R^2$ . The resulting predictions were also assessed for skew and were found to be right skewed with skew values of between 0.97 and 0.999. Since the dataset was not normally distributed, one-way ANOVA and Kruskal-Wallis tests were performed on the predicted values of the methods to evaluate statistical significance. The results were analysed to show the results of the method by segment length, route, day of the week and time of day. The results are presented and discussed in the next Section.

## 4 Results and Discussion

As can be seen in the results presented in Table 4, the proposed method WJP-P outperforms the other methods using MAE, MAPE, RMSE and  $R^2$ , the results of the commonly used SP are comparable to WJP-P, and both outperform the other two methods on all metrics. The MAE for WJP-P is a 5% improvement over the commonly used SP method. The MAPE is 2.5% better for WJP-P compared to SP. WJP-P surpasses SP by 6.4% on RMSE.  $R^2$  is high for all methods due to the size and quality of the data, and WJP-P surpasses the other methods with an  $R^2$  of 0.954.

The magnitude of the error is of primary importance, but the bias or direction of the error is also important and is rarely reported in the literature. MPE is not a good metric for assessing the accuracy of the results, as even large positive and negative error values could negate each other. Still, it was included in our analysis as it is a good indicator of the bias in the results. We can see from the MPE results in Table 4 that all methods return a small positive MPE. HA and WJP-C return smaller MPE values than SP and WJP-P. A negative MPE means the method tends to overpredict journey time instead of underpredicting it. If a bus journey is underpredicted, the bus will arrive later than expected, and if a bus journey is overpredicted, the bus will arrive earlier than expected. When bus scheduling is taking place, methods of prediction that overestimate should be used as this reduces the likelihood of late departure on the return journey, which is one of the causes of unreliability on bus networks [9]. For journey planning without arrival-time bound transfers (e.g. arriving in the office by 9 am), methods that overpredict are superior.

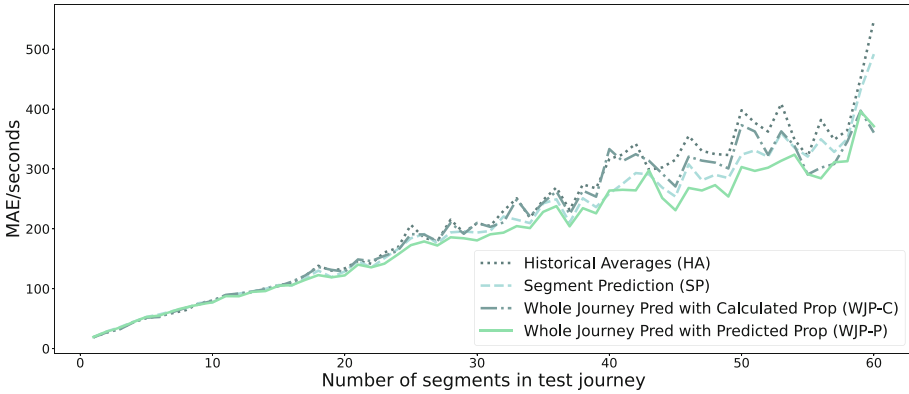
The results were statistically significant with ANOVA and Kruskal-Wallis tests with p-values of 0.013 and 0.019, respectively. These results were stable over multiple runs with different test/train splits and various test/train sizes. The remainder of this Section will discuss a deeper analysis of the performance of the methods by the number of segments travelled, by the route and by temporal features.

**Table 4.** Full Results

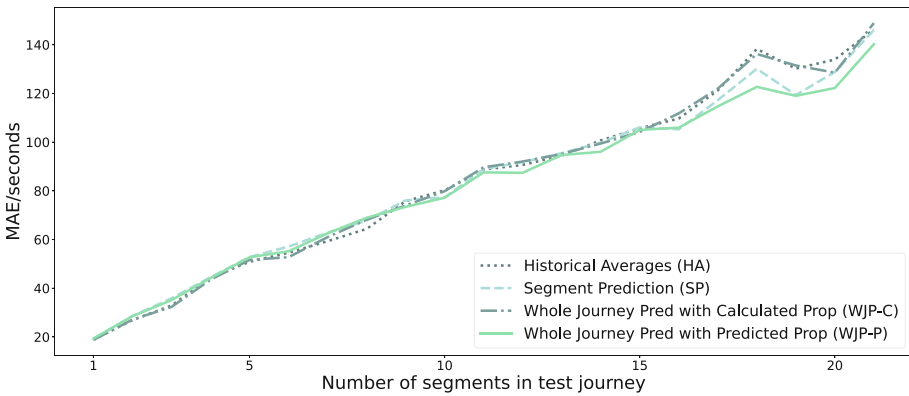
| Metric         | HA       | SP       | WJP-C           | WJP-P         |
|----------------|----------|----------|-----------------|---------------|
| MAE/s          | 164.88   | 154.15   | 159.92          | <b>146.36</b> |
| MAPE           | 0.1309   | 0.1276   | 0.1298          | <b>0.1245</b> |
| RMSE/s         | 289.46   | 245.81   | 275.52          | <b>229.98</b> |
| MPE            | -0.01099 | -0.02309 | <b>-0.01069</b> | -0.02298      |
| R <sup>2</sup> | 0.927    | 0.948    | 0.934           | <b>0.954</b>  |

#### 4.1 Impact of Number of Segments

As shown in Figs. 2 and 3, WJP-P outperforms the other methods on all lengths of journey that exceed 9 segments in length. Up to 9 segments in length, WJP-C and HA methods are superior. This threshold is likely due to the cumulative noise in the data. It can be considered that methods involving calculated averages, especially HA, perform well on short trips and especially WJP-P but also SP, perform well on medium and long trips.



**Fig. 2.** The number of stop pair segments in the partial journey vs MAE. WJP-P outperforms all other methods for trips with a length of greater than 9 segments. The number of segments in the test journey in this chart has no relationship to where in the whole bus journey the test journey is. There could be a test journey of 1 segment at the beginning or the end of the whole bus journey.



**Fig. 3.** The number of stop pair segments in the partial journey vs MAE: trips with a segment length of 1 to 22. This enlarged part of Fig. 2 more clearly shows HA and WJP-C outperforming the other methods until the number of segments in the trip exceeds 9 and WJP-P outperforming other methods for trips longer than 9 segments.

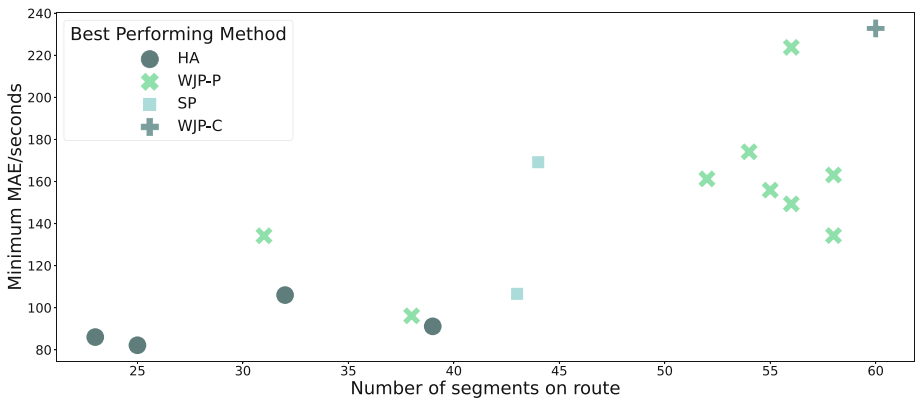
### 4.2 Impact of Route

Generalising from the findings in Fig. 2 and 3, it was theorised that the methods using calculated averages, HA and WJP-C, would perform best on short routes and WJP-P would perform best on medium and longer routes. This was largely found to be the case, as can be seen in Table 5 and Fig. 4. Even though HA was included as a naive baseline method, and performed poorest overall, it was the best performing method on four routes. These routes were all short routes, with an average length of 30.5 segments, and were four of the six shortest routes.

WJP-P was the top scoring method on nine routes which have an average length of 50.89 segments. Despite being the dominant method in the literature, SP was the superior method on only two routes, the 184 in both directions. WJP-C was the top-performing method on one route, the longest one. It was also observed that some of the routes had very minimal differences in the MAE between the methods, and others had wide variation. The column *Max minus Min MAE* in Table 5 shows the remainder when the MAE of the best performing method is subtracted from the MAE of the worst performing method. Several factors were examined to try to elucidate the cause of this instability between methods including the percentage of data retained after outlier removal, SD, variance, skew, kurtosis and the number of outliers at various thresholds. None of the factors studied showed a strong correlation or had a linear relationship with the *Max minus Min MAE*. However, it can be seen in Table 5, which is arranged in order of increasing SD, that the routes with a large difference in the performance of methods all have an SD larger than 60s. The results across these sixteen routes echo what is seen across the literature, with different methods performing better on different datasets. It is clear that bus routes shouldn't be treated as a homogenous group when assessing methods for the prediction of journey times.

### 4.3 Impact of Temporal Variables

An analysis of MAE by time of day and the day of the week was conducted for the two best-performing methods overall, WJP-P and SP, as shown in Figs. 5 and 6. A pattern emerged that seemed to correspond to peak and off-peak travel times/days on the bus network, so this was quantified with a reliability index, defined as one divided by the SD of the whole journey times on the network, as described by Sterman and Schofer [24]. Figure 5 shows WJP-P outperforming SP, especially during the morning and evening peak travel periods when the network reliability is low. The left section of the graph shows WJP-P outperforming SP



**Fig. 4.** The best method by the number of segments on the bus route and the MAE of the best performing method.

from time group 5 to 12 (5 am to 11 am) and again from time group 20 to 25 (5 pm to 9 pm). There is minimal difference between the two methods during off-peak times. Similarly, Fig. 6 shows WJP-P outperforming SP from Monday to Friday. During the conventional working week and especially during peak travel times is the time with the greatest number of passengers are on the network, and improving journey times at these times will benefit the most people. This is an important finding and is a strong argument for WJP-P over SP.

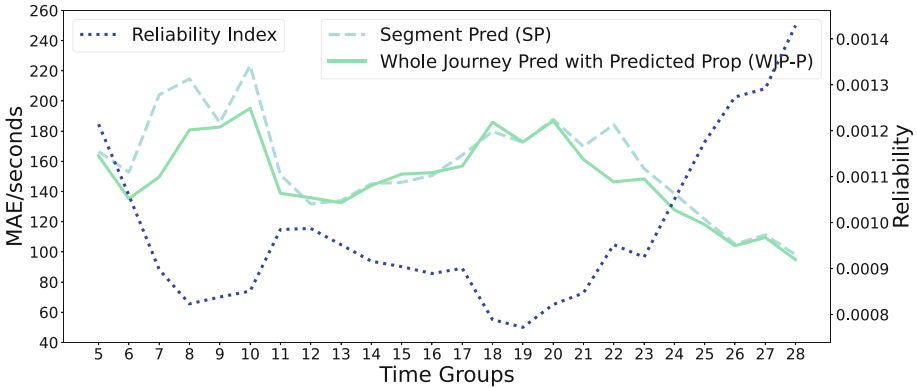


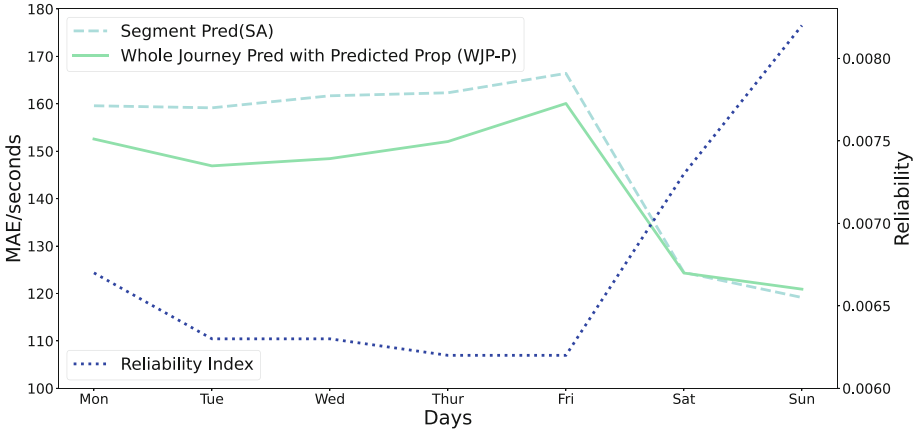
Fig. 5. The time groups in the study vs MAE and reliability

Table 5. Results by Route

| Route | Num Trips | Percent Data | Num Segs | SD    | Min MAE | Max minus Min MAE | Best Method |
|-------|-----------|--------------|----------|-------|---------|-------------------|-------------|
| 27A.1 | 8775      | 99.51        | 39       | 31.73 | 91.15   | 7.93              | HA          |
| 32.2  | 6423      | 98.75        | 58       | 32.89 | 134.30  | 6.46              | WJP-P       |
| 32.1  | 6512      | 98.20        | 56       | 39.66 | 149.37  | 17.21             | WJP-P       |
| 184.1 | 6848      | 98.85        | 43       | 44.88 | 106.60  | 8.83              | SP          |
| 79A.1 | 7566      | 99.01        | 32       | 45.43 | 106.00  | 2.42              | HA          |
| 120.2 | 13618     | 99.52        | 23       | 52.47 | 86.04   | 2.79              | HA          |
| 42.1  | 10200     | 98.25        | 55       | 53.38 | 155.89  | 16.77             | WJP-P       |
| 42.2  | 11164     | 98.67        | 58       | 56.32 | 163.08  | 8.06              | WJP-P       |
| 27A.2 | 8436      | 99.45        | 38       | 56.74 | 96.15   | 7.19              | WJP-P       |
| 120.1 | 13944     | 99.18        | 25       | 58.62 | 82.15   | 1.90              | HA          |
| 56A.1 | 4238      | 97.85        | 54       | 62.51 | 174.12  | 147.37            | WJP-P       |
| 4.1   | 15016     | 98.69        | 56       | 69.59 | 223.77  | 30.49             | WJP-P       |
| 56A.2 | 4294      | 99.02        | 52       | 73.85 | 161.25  | 201.51            | WJP-P       |
| 4.2   | 15629     | 99.08        | 60       | 75.94 | 232.77  | 23.54             | WJP-C       |
| 79A.2 | 7240      | 99.38        | 31       | 81.71 | 134.20  | 22.77             | WJP-P       |
| 184.2 | 6361      | 98.63        | 44       | 88.43 | 169.19  | 21.34             | SP          |

#### 4.4 Computational and Storage Resources

The computational and storage resources consumed in this experiment are presented in Table 6. All methods were tested on the same partial journeys on a 2017 MacBook Pro with a 3.3 GHz Intel Core i5 processor and 16 GB of memory. Data processing time has been provided separately from training time because after the initial year of data is processed, the time taken for processing additional daily data would be minimal. HA does not do any ML model training, so its average training time is zero. WJP-C trains a single RF model per route, and a short training time of 0.51 s reflects this. WJP-P and SP show similar prediction times with 42.91 s and 45.40 s, respectively. Both of these methods train a model per segment, but the training time for WJP-P is shorter than SP. The average prediction time is likely the most significant of these measurements for journey planning applications and RTPI, as this will determine the speed at which information is returned to passengers. HA and WJP-C are very similar and over 40 times faster than WJP-P and SP, which are also very similar to each other. The same pattern is seen for storage with the models and data required for HA and WJP-C 18 times smaller than the other two methods. These results are specific to RF models, which have a larger storage size and a shorter training time than neural network models.



**Fig. 6.** The days of the week in the study vs MAE and reliability

**Table 6.** Computational and storage resources consumed by the different methods

| Measurement                            | HA     | SP       | WJP-C  | WJP-P    |
|--|--------|----------|--------|----------|
| Data Processing/Route (in seconds)     | 54.83  | 20.32    | 102.7  | 132.29   |
| Training Time/Route (in seconds)       | 0.00   | 45.40    | 0.51   | 42.91    |
| Prediction Time/Test case (in seconds) | 0.0157 | 0.6719   | 0.0156 | 0.6735   |
| Storage /Route (in megabytes)          | 846.30 | 16017.62 | 852.80 | 16739.08 |

## 5 Conclusion

From the results of our experiments and analysis, we conclude that the commonly used SP method is not the best approach. It is not the best approach overall, nor on the majority of the bus routes, across multiple metrics. We also conclude bus routes are not a homogenous population and that attempting to define a single best algorithm for predicting bus journey times will result in sub-optimisation. The optimum method for journey time predictions should be determined based on many factors, including those identified in this study: the application, the trip's length, the temporal features of the trip, and also possible factors related to the bus route characteristics and data profile.

The novel method we present in this paper represents a significant contribution as WJP-P outperforms SP on most metrics, and on the majority of bus routes at a similar computational cost. The other method based on whole journey time prediction, WJP-C, has an overall reduction of 3.7% MAE compared to SP but at a significantly reduced computational cost. We suggest this as a good value option, balancing the accuracy of predictions with computational and storage costs.

Analysis of the results has provided important insights into the nature of bus journey time prediction regarding the bias in and variability between methods predicting for trips of different lengths, at different times of day and week and for bus routes with different characteristics. An important finding is the enhanced benefit of WJP-P at peak travel times and during the traditional working week. These findings can be applied to predict bus journey times for timetabling, and result in more achievable timetables, improving the reliability of the bus network.

Planned further work involves the application of these methods to more bus routes to validate the results. We are especially interested in applying the method to bus routes with different characteristics in different cities, and to bus routes with a lower frequency and a lower quality of data.

**Acknowledgements.** This publication has emanated from research supported in part by a grant from Science Foundation Ireland under Grant number 18/CRT/6183.

## References

1. Almeida, F., Lobo, A., Couto, A., Ferreira, J.P., Ferreira, S.: Urban factors influencing the vehicle speed of public transport. *Transp. Res. Procedia* **62**, 318–324 (2022). 24th Euro Working Group on Transportation Meeting
2. Avenali, A., Catalano, G., Gregori, M., Matteucci, G.: Rail versus bus local public transport services: a social cost comparison methodology. *Transp. Res. Interdiscip. Perspect.* **7**, 100200 (2020)
3. Cats, O., Loutos, G.: Evaluating the added-value of online bus arrival prediction schemes. *Transp. Res. Part A Policy Pract.* **86**, 35–55 (2016)
4. Chakhtoura, C., Pojani, D.: Indicator-based evaluation of sustainable transport plans: a framework for Paris and other large cities. *Transp. Policy* **50** (2016)
5. Chen, G., Yang, X., An, J., Zhang, D.: Bus-arrival-time prediction models: link-based and section-based. *J. Transp. Eng.* **138**(1), 60–66 (2012)

6. Chien, S.I.J., Ding, Y., Wei, C.: Dynamic bus arrival time prediction with artificial neural networks. *J. Transp. Eng.* **128**(5), 429–438 (2002)
7. Cristóbal, T., Padrón, G., Quesada-Arencibia, A., Alayón, F., de Blasio, G., García, C.R.: Bus travel time prediction model based on profile similarity. *Sensors* **19**(13), 2869 (2019)
8. Dunne, L., McArdle, G.: A novel post prediction segmentation technique for urban bus travel time estimation. In: *Proceedings of the Tenth International Workshop on Urban Computing* (2021)
9. El-Geneidy, A.M., Horning, J., Krizek, K.J.: Analyzing transit service reliability using detailed data from automatic vehicular locator systems. *J. Adv. Transp.* **45**(1), 66–79 (2011)
10. Gal, A., Mandelbaum, A., Schnitzler, F., Senderovich, A., Weidlich, M.: Traveling time prediction in scheduled transportation with journey segments. *Inf. Syst.* **64**, 266–280 (2017)
11. Gössling, S.: Why cities need to take road space from cars - and how this could be done. *J. Urban Des.* **25**(4), 443–448 (2020)
12. He, P., Jiang, G., Lam, S.K., Sun, Y.: Learning heterogeneous traffic patterns for travel time prediction of bus journeys. *Inf. Sci.* **512**, 1394–1406 (2020)
13. He, P., Jiang, G., Lam, S.K., Tang, D.: Travel-time prediction of bus journey with multiple bus trips. *IEEE Trans. Intell. Transp. Syst.* **20**(11), 4192–4205 (2019)
14. Ma, L., et al.: Estimating urban road GPS environment friendliness with bus trajectories: a city-scale approach. *Sensors* **20**(6), 1580 (2020)
15. Maiti, S., Pal, A., Pal, A., Chattopadhyay, T., Mukherjee, A.: Historical data based real time prediction of vehicle arrival time. In: *Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems*, pp. 1837–1842 (2014)
16. Mazloumi, E., Rose, G., Currie, G., Sarvi, M.: An integrated framework to predict bus travel time and its variability using traffic flow data. *J. Intell. Transp. Syst.* **15**(2), 75–90 (2011)
17. Morabia, A., et al.: Potential health impact of switching from car to public transportation when commuting to work. *Am. J. Public Health* **100**(12), 2388–2391 (2010)
18. Nimpanomprasert, T., Xie, L., Kliewer, N.: Comparing two hybrid neural network models to predict real-world bus travel time. *Transp. Res. Procedia* **62**, 393–400 (2022). 24th Euro Working Group on Transportation Meeting
19. dell’Olio, L., Ibeas, A., Cecin, P.: The quality of service desired by public transport users. *Transp. Policy* **18**(1), 217–227 (2011)
20. Pandurangi, A., Byrne, C., Anderson, C., Cui, E., McArdle, G.: Design and development of an application for predicting bus travel times using a segmentation approach. In: *Proceedings of the 6th International Conference on Geographical Information Systems Theory, Applications and Management 2020*, pp. 72–80 (2020)
21. Palys, L., Ganzha, M., Paprzycki, M.: Machine learning for bus travel prediction. In: *Computational Science - ICCS 2022*, pp. 703–710 (2022)
22. Singh, N., Kumar, K.: A review of bus arrival time prediction using artificial intelligence. *WIREs Data Min. Knowl. Discov.* e1457 (2022)
23. Soza-Parra, J., Muñoz, J.C., Raveau, S.: Factors that affect the evolution of headway variability along an urban bus service. *Transportmetrica B Transp. Dyn.* **9**(1), 479–490 (2021)
24. Serman, B.P., Schofer, J.L.: Factors affecting reliability of urban bus services. *Transp. Eng. J. ASCE* **102**(1), 147–159 (1976)

25. Sutandi, A.C., Dermawan, A.: Accuracy of bus timetable using information communication and technology GPS to inform trans metro bandung bus passenger. In: Proceedings of the 6th International Conference on Civil, Offshore and Environmental Engineering, pp. 725–734 (2021)
26. U.N.: Goal 11 — Department of Economic and Social Affairs. <https://sdgs.un.org/goals/goal11>. Accessed 15 July 2022
27. Yu, B., Wang, H., Shan, W., Yao, B.: Prediction of bus travel time using random forests based on near neighbors. *Comput. Aided Civ. Infrastruct. Eng.* **33**(4), 333–350 (2018)
28. Zhao, X., Carling, K., Håkansson, J.: An evaluation of the reliability of GPS-based transportation data. In: Proceedings of the Computer Science Conference, Vienna (2017)