Department for Transport Framework for Transport-Related Technical and Engineering Advice and Research

Ref PPRO 4/45/12

LOT 2 - ROAD

Bus Journey Time Variability in Urban Areas

PILOT STUDY AND ANALYSIS REPORT

Tuesday, 15 December 2015

Ref: PPRO 04/91/20 Task 119

Work Package Manager: Catherine Goodwin

PREPARED BY WSP ON BEHALF OF THE PBWSP SUPPLIER GROUP

13/03/2015

Table of Contents

1 Introduction

1.1 Pilot Study Context

1.1.1 DfT awarded the PB/WSP supplier consortium a contract to carry out research into bus journey time variability in urban areas in England (outside of London). The first stage of this project was the submission of a Technical Report to outline the factors which were understood to affect variability from existing research and literature and to propose a methodology for the 'real-world' collection of data necessary to derive robust estimates of the impact of the different factors which affect journey time variability.

1.1.2 Following identification of a suitable location and service, engagement of local stakeholders and agreement between DfT and PB/WSP on the scope of the data to be gathered, a Pilot Study was conducted on buses operating in service between Bath and Bristol in November 2013.

1.1.3 From the Technical Report's literature review, it was possible to determine the implied importance of various factors influencing bus journey variability. These are provided in Table 1 below.

Element	Implied Importance	Complexity of existing research with respect to journey time variability
General traffic	High	Simple relationships
Priority at junctions	High	Simple examinations for whole route or service; complex at specific junctions
Route length	High	Complex relationships for specific routes
Number of bus stops in route	High	Some investigation
Priority between junctions (segregated lanes)	Medium	Simple examination
Vehicle spacing policies	Medium	No substantial studies
Ticket type	Medium	Investigated within dwell time
Physical characteristics of the stops	Low	No substantial studies
Experience of bus drivers	Low	No substantial studies
Number of doors	Low	No substantial studies
Frequency of public transport use	Low	No substantial studies
Weather	Low	No substantial studies

Table 1 Implied Importance of Factors Affecting Bus Journey Time Variability

1.1.4 The factors with the highest implied importance are also those factors which are analysed and presented in this report, with the addition of consideration of passenger numbers and whether passengers travel with cash or non-cash fares. These latter factors were identified as potential key issues from the project team's work with the local stakeholders.

1.1.5 Chapter 3 of the Technical Report identified that the measure of variability which would be adopted in this study would be the coefficient of variation (CV), as it is the recommended measure of urban travel time variability (by DfT, through WebTAG, and others). Testing of each variable would be done against the calculated CV and repeated by combining variables using multi-variate regression.

1.2 Overview of Data Collected

1.2.1 The data collection process of the Pilot Study, described fully in Appendix A, generated one observation record for each bus journey on which the enumerators travelled. As the survey period covered 15 consecutive weekdays (between 04/11/2013 and 22/11/2013) and 5 enumerators were deployed to collect data on each day, undertaking 4 separate bus journeys during each shift, a total of 150 bus journey records were due to be collected for each direction of travel. Initial analysis has focused on the direction from Bristol to Bath and, primarily due to the disruptive effects of late running on previous journeys, 6 records of journeys are not available. Therefore in total, there are 144 records collected for Bristol to Bath and which have been used to calculate the bus travel time in this analysis.

1.2.2 Bus journey time consists of vehicle travel time and dwell time at bus stops (for passenger boarding and alighting). Consequently, the data collection process included the recording of bus arrival and departure times, while the analysis which is intended to prove the data collection process and the in-principle use of such data has been carried out at the total journey time level. It would be possible to re-run the analysis to separate out dwell time, however this would result in a slightly smaller number of data points as enumerators were not always able to record both arrival and departure times accurately. Thus for this analysis, the bus travel time between any two stops (whether consecutive or not) is calculated by the difference between the values of departure time associated to the two stops.

1.2.3 Along the route, there are 37 stops in total from Bristol to Bath, which lead to 666 combinations of stop pairs. Ideally for data modelling, although not necessarily for bus operation, there would be 144 records of bus travel time for each of these 666 pairs of stops, to allow the average travel time to be worked out and also the standard deviation (SD) values together with coefficient of variation (CV) values. However, the enumerators were not always able to record the departure time for stops where the bus driver did not stop (as there were no passengers to get off and/or get on). Therefore for those bus stop pairs with only one stop having a departure or arrival time recorded, no travel time value is generated. As a result of this, a varied number of observations for different stop-stop combinations exist. For pairs of stops which are located in very densely-trafficked areas and which incur much more frequent stopping activities, the sample size tends to be large and close to 144, and in sparsely-trafficked areas the opposite is true. This defect could be remedied by introducing GPS data to provide values where enumerators have missed the records out however this is not necessary to prove the concept of creating a model and has therefore not been done at this stage. Although it may improve the analysis, it is not guaranteed to justify the additional data manipulation required at the Pilot Study stage.

1.3 Data for Analysis

1.3.1 Following data collection and processing, the dataset available consisted of up to 144 trips for each of the 666 (= (37x36)/2) stop-stop route sections. For consistency, in this report, each stop-stop combination is referred to as a 'route section' and each of the 144 instances is referred to as a 'journey'.

1.3.2 The following data was available and has been analysed:

- The travel time on the route section for each journey;
- The total number of passengers boarding and alighting the service during each journey;
- The total number of boarding passengers paying with cash/non-cash during each journey; and
- The number and types of junctions in each route section;
- A sample of TrafficMaster data showing sampled journey times along route sections in 15 minute segments.

1.3.3 Although some further data were available, the items above have been selected to demonstrate the potential of the method of analysis.

1.4 Initial Data Checks

1.4.1 The first step in the analysis is to check the impact of the varied sample size for the different route sections. 58 pairs of stops have fewer than 3 journey records and have been excluded from the analysis. For the rest of the route sections, it is possible to plot a scatter chart of CV values of travel time against the number of records, as shown in Figure 1a below.

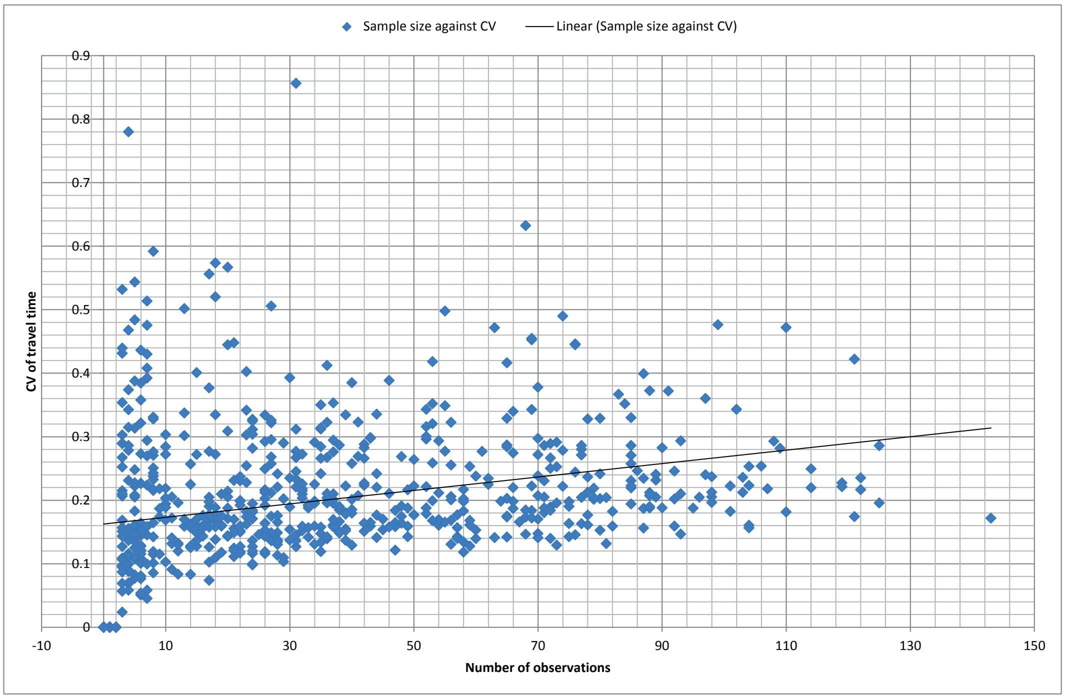


Figure 1a: Chart of scatter points for sample size VS CV of journey time

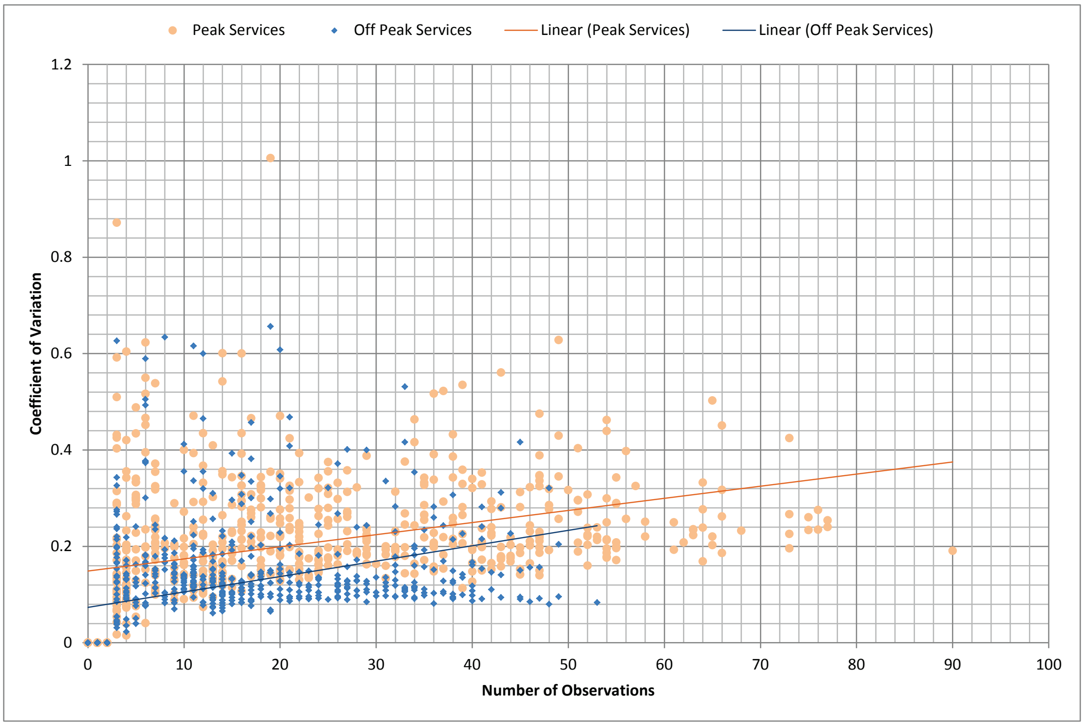


Figure 1b: Chart of scatter points for sample size VS CV of journey time including dimension of peak/off peak

1.4.2 As can be seen from the above figure (1a), if the sample size is small, the CV values tend to be greater and more varied. As the sample size increases, CV values become more concentrated. This demonstrates that a small sample size may introduce bias. Based on observation of this chart, the threshold for the sample size was initially set to 50, providing a reasonable number of route sections for analysis and avoiding the apparent fall in CV as the number of records falls. Therefore, the initial steps in the analysis only consider the 201 route sections with at least 50 records.

1.4.3 In later analysis, segmentation of the data into peak and /off peak[1] periods was introduced. Figure 1b shows the relationship between CV and observations once this segmentation is introduced. Based on observation of this chart, and to ensure a sufficiently large sample of routes to analyse, the threshold for the peak/off-peak analysis was lowered to a minimum of 10 observations.

1.4.4 As many of these combinations overlap by sharing the same section of route, bias could be introduced and make the results less representative. To allow some degree of sense-checking therefore, the sections between the operator-specified timing points[2] have been highlighted, as they are contiguous (without overlapping):

- Bristol Bus Station to Temple Meads;
- Temple Meads to Brislington Square;
- Brislington Square to Saltford Tyning Road;
- Saltford Tyning Road to Newton St Loe; and
- Newton St Loe to Bath Bus Station.

As there are only 5 such sections, conclusions cannot be drawn from these points alone, and indeed it would not be unreasonable to expect these to exhibit different characteristics to sections not involving timing points, as buses operating early would be more likely to wait at these points, in order to comply with the advertised timetable. Nevertheless, highlighting the sections is of relevance to the operator, in terms of the construction and performance of the timetable.

1.4.5 A map of the route is provided in Appendix B. It can be noted that the last two sections have a sample size of fewer than 50 journey records (35 and 42 respectively) and hence would otherwise have fallen outside of the group of the initial 201 selected route sections. For the purpose of highlighting the different characteristics of each route section, these stop pairs have been included intentionally. Therefore, in total, 203 stop to stop sections along the route are included in the analysis, spread over all 144 journeys departing during the period of the enumerator observations (14:26 until 18:15 hours).

[1] For simplicity, it was defined that if more than 50% of the whole bus journey (from the departure stop to the destination stop) takes place in the off peak period (10:00 am to 4:00 pm), the journey is treated as off peak. As the average bus travel time from Bristol to Bath is 1 hour, all the bus journeys departing before 3:30 pm are treated as off peak and those after 3:30pm as peak.

[2] A timing point is a bus stop along the line of route of a service at which the operator has specified the departure time. These are required to be specified in the timetable at regular intervals, as part of the service's registration with the Traffic Commissioner.

2 Analysis of Factors Influencing Bus Journey Time

2.1 Introduction

2.1.1 In the following sections, analysis is presented of the survey data on bus journey times and which investigates the influence different journey factors have on the variation recorded.

2.1.2 In the previous stage of this work, we conducted a literature review to identify what the most appropriate measures of variation would be. It had been noted that, for general traffic, while the standard deviation of travel time increases with journey distance, it increases less fast than the mean. This means that the ratio of standard deviation to the mean (defined as Coefficient of Variation or CV) falls with increasing distance.

2.1.3 For this reason CV is the standard measure for comparing different situations where mean values of travel time may be substantially different. It is also the recommended measure for urban travel time variability in general traffic (DfT 2013). Essentially, the CV indicates a journey becomes more reliable as the value tends towards 0.

2.1.4 We have therefore used CV as the standard measure of variation throughout this work, with some comparison against Standard Deviation and Average Journey Times where this provides useful comparators.

2.2 Influence of Total Journey Time

2.2.1 In appendix C, for each stop to stop combination, the average travel time, standard deviation value and CV value are calculated. This table replicates Table 2 of the Technical Report, but replaces the synthetic data with empirical data. The relationship of average journey time with the CV has been plotted in Figure 2, and with standard deviation (SD) in Figure 3.

Standard Deviation, Coefficient of Variation and Heterscedasticity

2.2.2 The different relationships displayed for SD and CV are very much in line with the findings for general traffic in Saltford in October 2012, using Trafficmaster data[3], that while the standard deviation of travel time increases with journey distance, it increases less quickly than the mean, so that the coefficient of variation [CV] (standard deviation divided by the mean) falls with increasing distance. This confirms that CV is a useful measure for considering the change in journey time variability with distance.

2.2.3 It appears however from Figure 2 that the variation in CV **decreases** as journey time increases, whereas the SD has a relatively constant level of variation.

2.2.4 This suggests that the relationship of CV to journey time may be heteroscedastic (having non-constant variance), whereas the relationship with SD is homeoscedastic (having constant variance). Similar patterns are seen later when comparing CV and SD with other independent variables.

2.2.5 This does raise some concern for regression modelling, as the estimation of a linear model inherently assumes that the variation is constant. This has been noted, but was not thought to invalidate the tests of relative strengths of relationships within this work. There are techniques available which can attempt to formally detect and correct for heteroscedasticity when estimating relationships. If proceeding to formally estimate an econometric model to explain journey time variability, such an approach may be justified. However, as this exercise is largely exploratory it is not felt that the presence of heteroscedasticity at this stage is likely to alter the overall conclusions.

[3] PBWSP Bus Journey Time Variability in Urban Areas Technical Report, tables 7 and 8

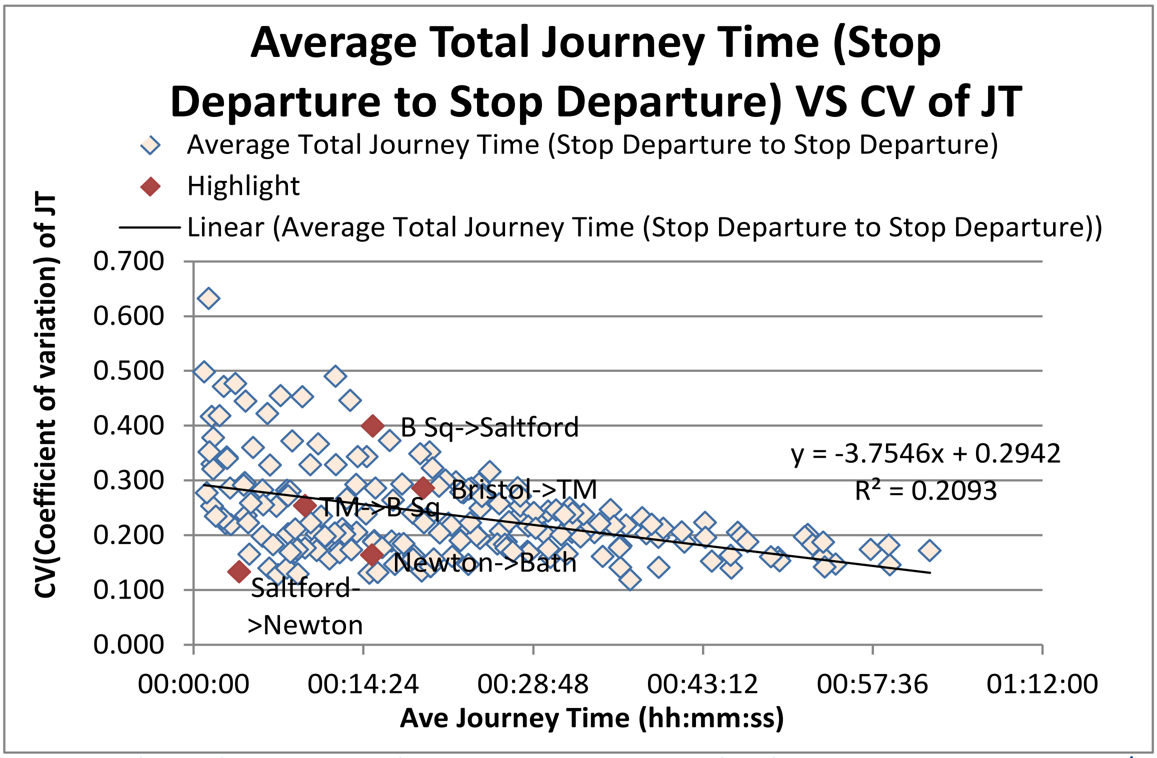


Figure 2: Chart of scatter points for average journey time vs CV of journey time, by route section[4]

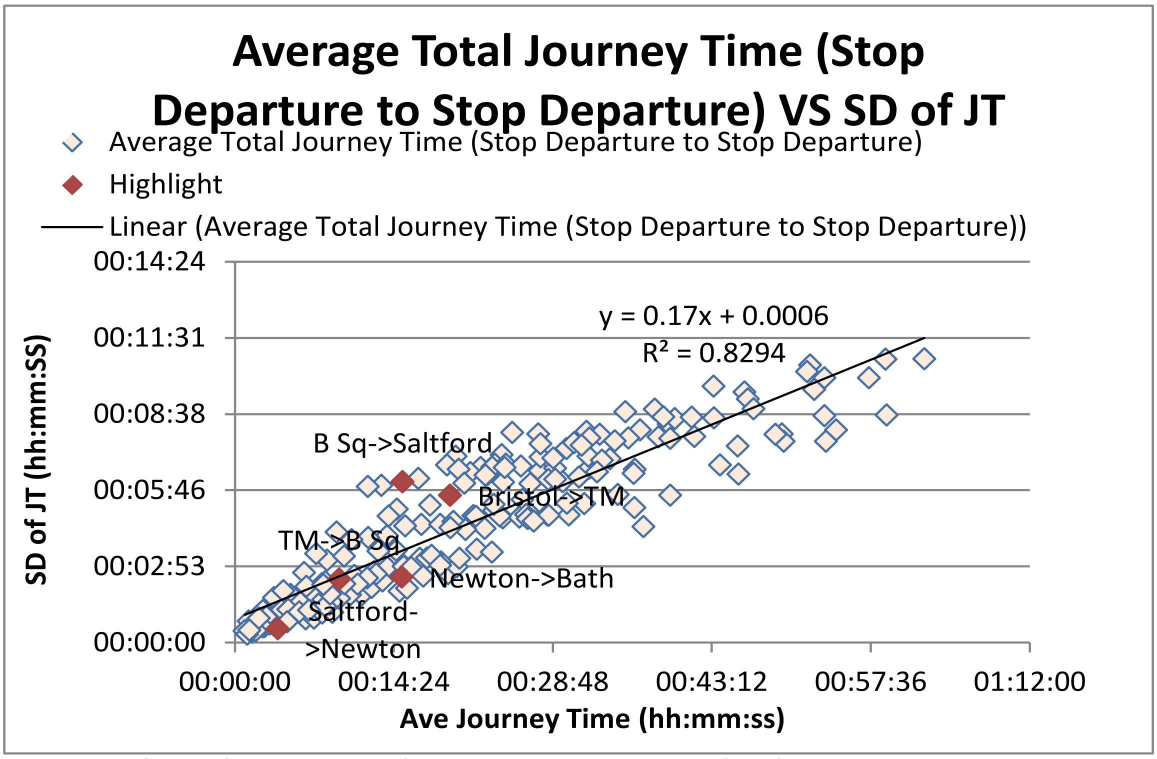


Figure 3: Chart of scatter points for average journey time vs SD of journey time, by route section

[4] Note that in this and all subsequent analysis using journey time, the actual time used in the equations is the time as a fraction of a 24 hour period. This is a convention in Microsoft Excel which is helpful for analysis. Axes have been labelled in hours, minutes and seconds (hh:mm:ss).

Conclusions for Bus Variability

2.2.6 Figure 2 shows that as average journey time becomes longer, the CV values tend to be more concentrated, and decline rapidly at first, and then more gradually. This supports the theory that buses make up lost time during a journey: although there may be delay on some sections of the route, it is possible within the route's schedule for a bus driver to catch the time up later.

2.2.7 The light coloured points in the chart are those route sections with 50 observations or more, and the darker points are the 5 timing point to timing point pairs. Among these 5 timing point sections, when average journey time increases, the CV becomes larger which is counter to the general theory, though it is perhaps not surprising that amongst a small subset external effects would be significant. It is worth noting that the Saltford to Newton St Loe and Newton St Loe to Bath sections, which have fewer than 50 observations, have the lowest CV values, which gives some reassurance that low sample size is not leading to high variability in these samples. The other possible reasons for this counter-trend are considered in more detail in the following analyses.

2.2.8 Figure 3 shows standard deviation (SD) values for each route section, again arranged by increasing average journey time. As can be seen from the chart, there is a strong linear relationship between average journey time and SD of journey time (regression equation: y = 0.1701x + 0.0006, where x= average journey time, y= SD of journey time, with a comparatively large R^2 value of 0.8294). Though this shows the highest absolute variation in journey time occurs for the longest route sections, the slope value of 0.17 confirms (as already suggested by Figure 2) that this variation is growing more slowly than journey time itself: i.e. when the bus travels for a longer time, the variation in time becomes proportionally smaller.

2.3 Influence of Number of Stops

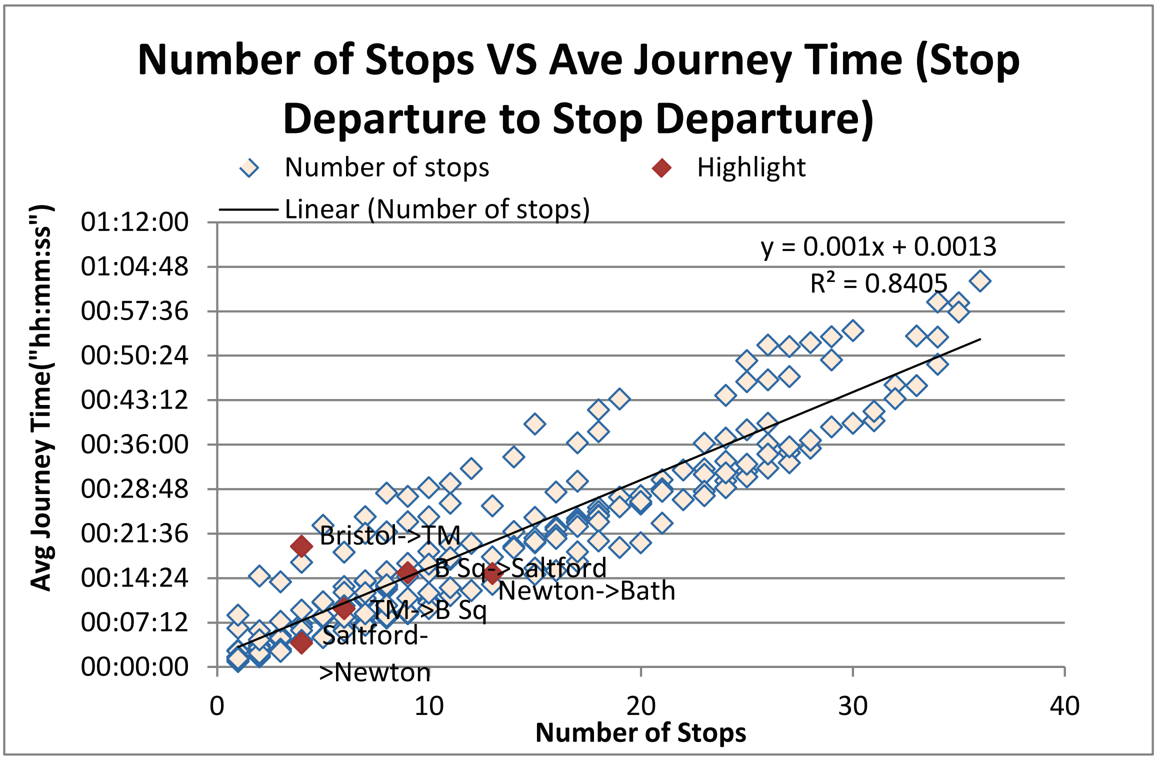


Figure 4: Chart of scatter points for number of stops passed vs average journey time by route section

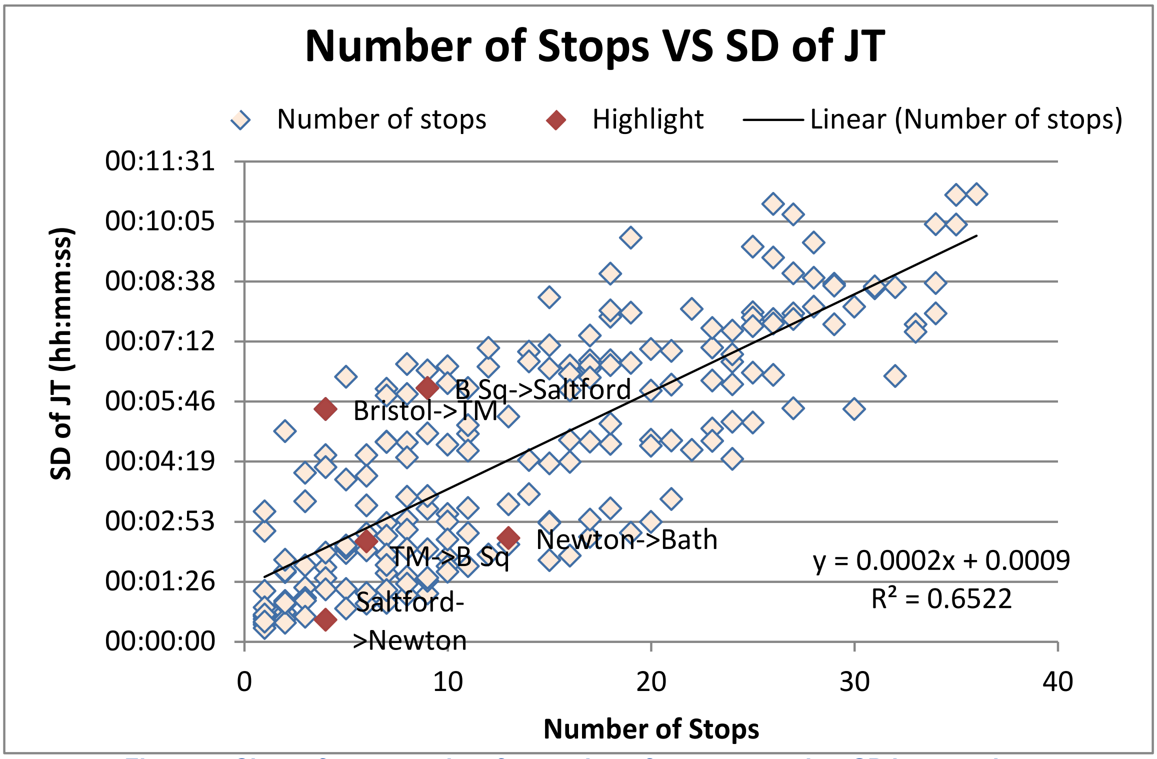


Figure 5: Chart of scatter points for number of stops passed vs SD journey time

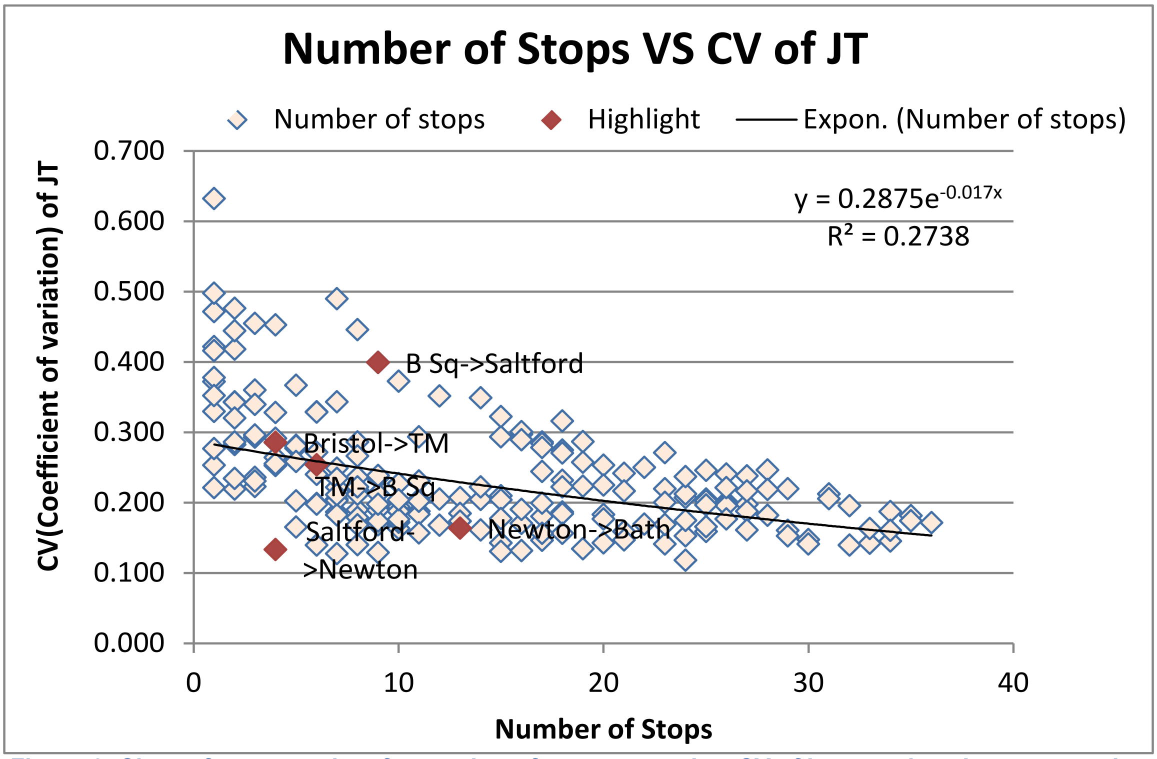


Figure 6: Chart of scatter points for number of stops passed vs CV of journey time, by route section

2.3.1 The above three charts examine the number of bus stops passed for each route section, against the average travel time, SD and CV values respectively. It should be noted that this analysis does not relate to whether the bus stopped at intermediate stops or not on each journey. As already discussed in paragraph 2.2.4, the relationship for CV also shows some sign of heteroscedasticity.

2.3.2 As would be expected, the number of stops is strongly correlated to the average travel time and standard deviation value. Significantly, the standard deviation is slightly less correlated with stops (Figure 5) than journey time (Figure 3), but the CV is more correlated (Figure 6, stops, and Figure 2, journey time). This suggests that the number of stops passed has more influence on 'catch-up time' than simply the journey time itself.

2.3.3 Amongst the timing point sections, Newton St Loe to Bath has the largest number of stops passed, which seems to offer some explanation for its lower CV value.

2.4 Influence of Number of Junctions Traversed

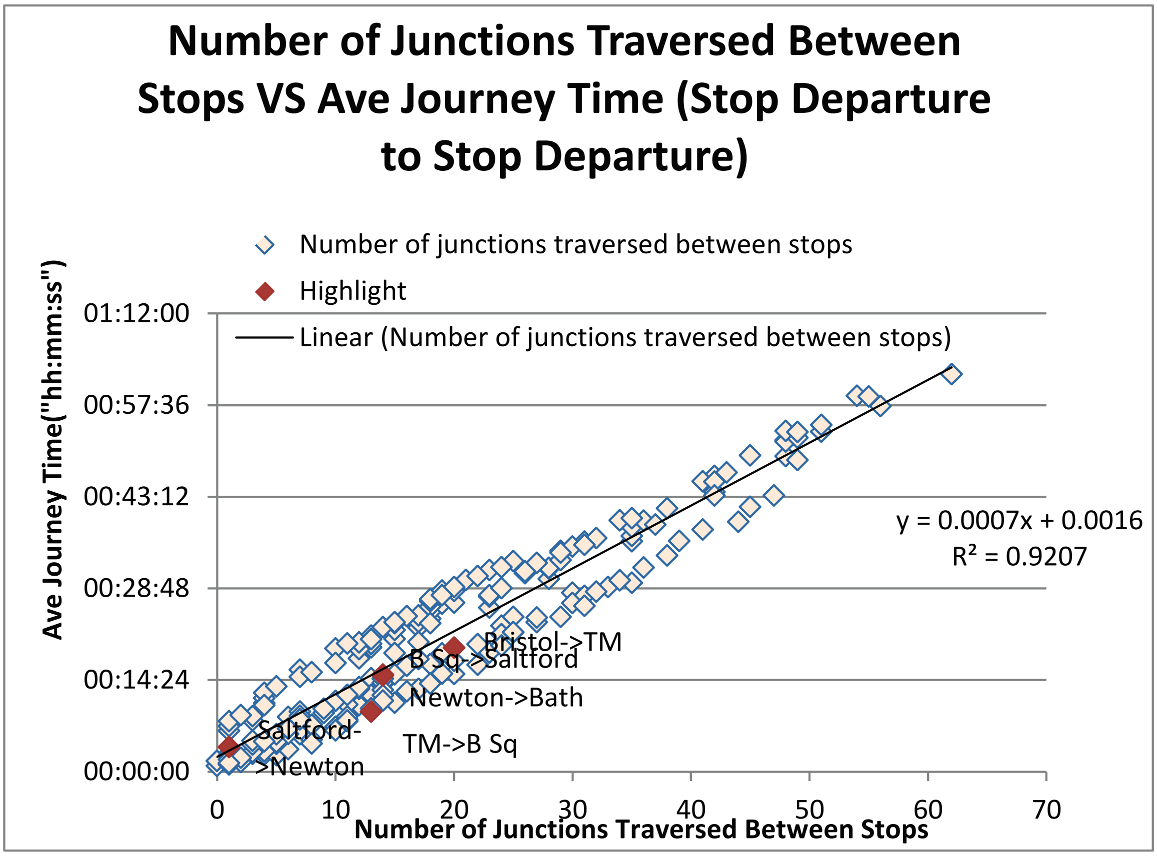


Figure 7: Chart of scatter points for number of junctions traversed vs average journey time

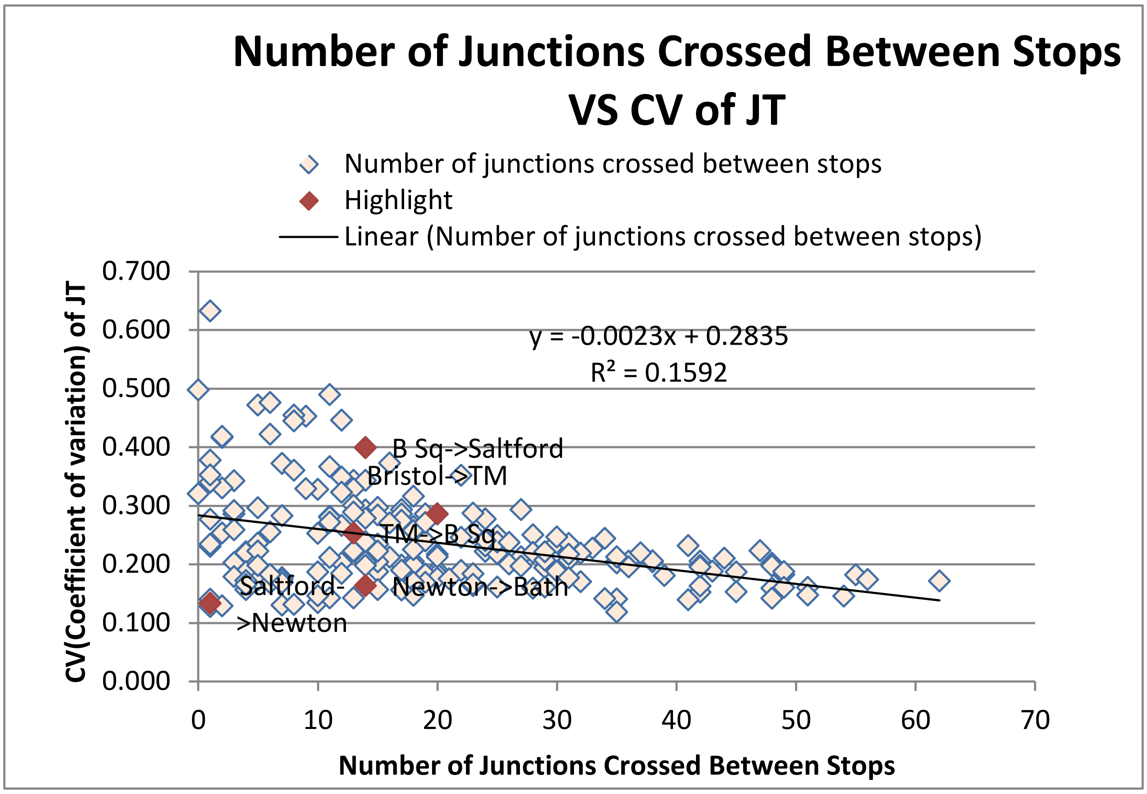


Figure 8: Chart of scatter points for number of junctions traversed vs SD journey time

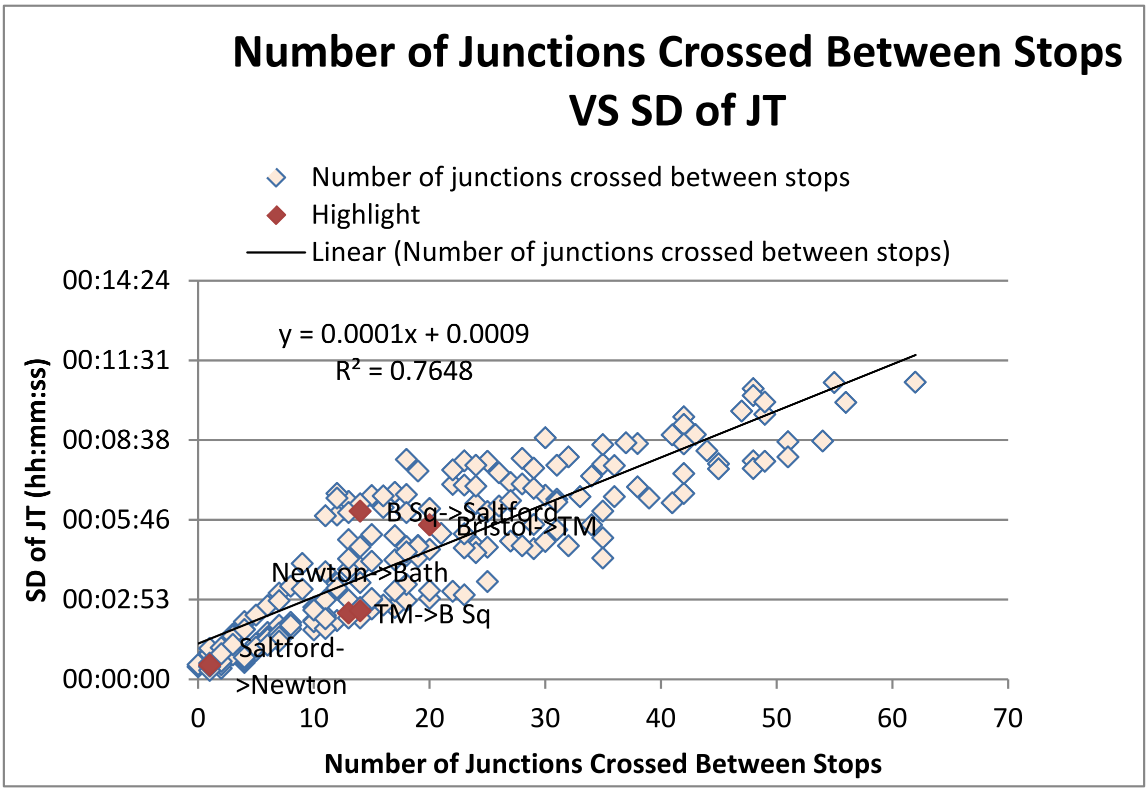


Figure 9: Chart of scatter points for number of junctions traversed vs CV of journey time

2.4.1 The relationship between the number of junctions traversed and bus travel time (average journey time, SD and CV) shows a similar pattern to the number of stops passed. However, the number of junctions is very highly correlated with average journey time (R-squared>90%), and the CV is only weakly correlated (R-square<20%). This suggests that the number of junctions for the total journey is not such a useful predictor of journey time variability which leads to the consideration of the type of junctions, including the presence of physical bus priority measures.

2.5 Influence of Junction Type

2.5.1 As shown above in Figures 7 to 9, the influence of the number of junctions traversed on journey time variability does not appear to be very strong. However, it was thought likely that the type of junction may influence the amount of journey time variability introduced by the junction itself. Indeed, any conventional bus priority scheme would be expected to improve both absolute journey time and variability.

2.5.2 To isolate this factor, the junctions have been grouped by type, and the number of junctions of each type along the route was summarised. Figure 10 below shows the junctions grouped into 8 separate types[5], and the number of these types between each adjacent stop-pair. The count of the number of each junction type occurring is shown in brackets on the key. Hence it can be seen that the number of controlled pedestrian crossings (36) without bus priority exceeds the total of all other types (26). There are 9 junctions in all with physical bus priority of some form.

[5] These categories of junction were available as part of the data collection. It has been subsequently suggested that the complexity of the junction be reviewed.

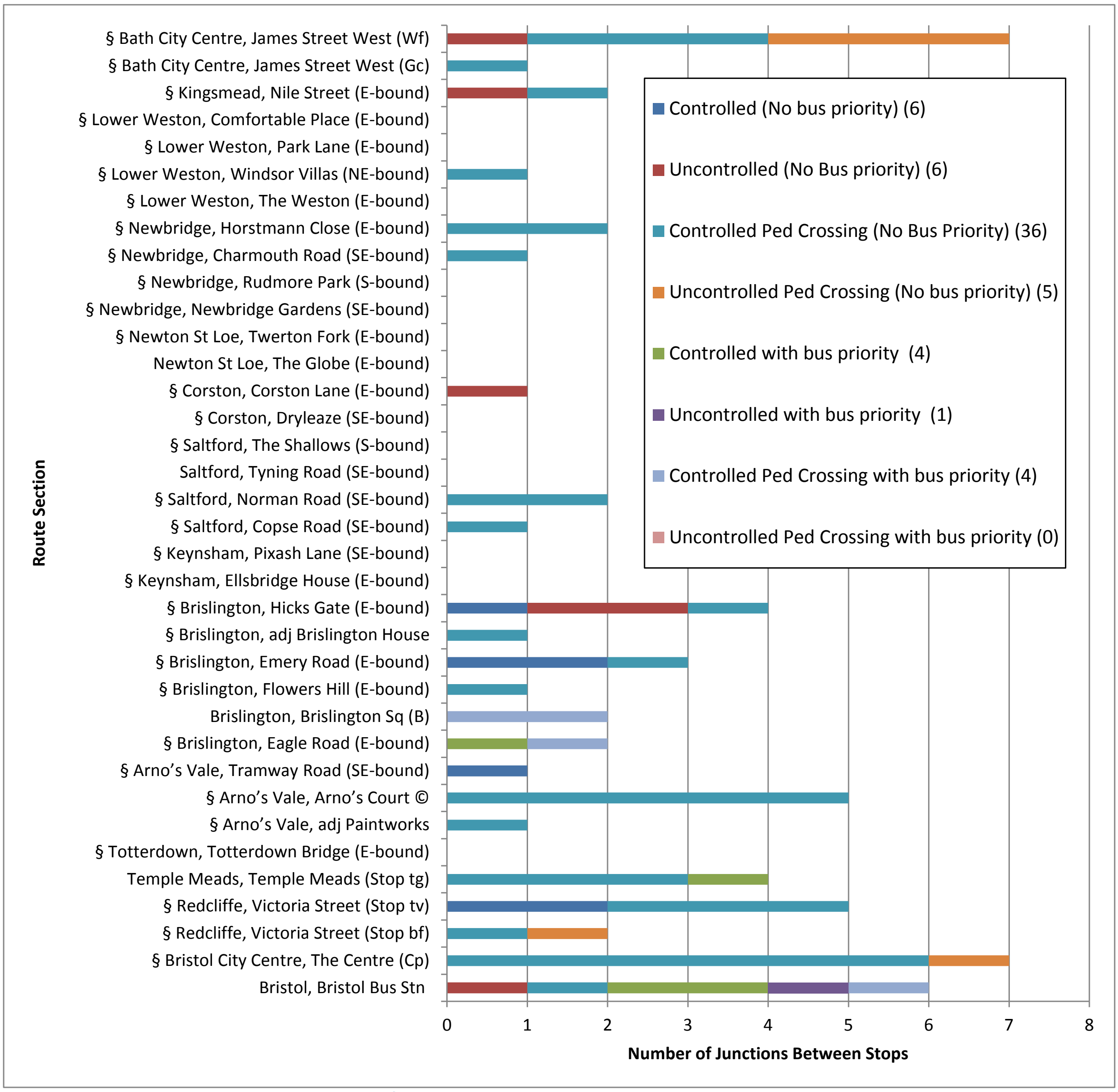


Figure 10: Number of Junctions between adjacent stops, by junction type

2.5.3 Figure 11 and Figure 12 below show the difference in speed and coefficient of variation split between junctions with and without bus priority. In this analysis, the time between adjacent bus stops has been focused on solely because any impact on bus journey time specifically from the junction would be expected to have a large impact from one bus stop to the next. Also, the measure of junction numbers which has been used is 'junctions per km', rather than the absolute number of junctions. This prevents the distance travelled between stops from biasing the results.

2.5.4 The speed chart (figure 11) shows that bus speeds are higher on average where there are fewer junctions. However, there are a number of route sections with few or no junctions present which have relatively low speeds. There is no evidence of a major difference with or without bus priority, although this may be because the bus priority is improving conditions to the level seen where there is no priority.

2.5.5 The coefficient of variation chart (figure 12) shows CV of journey time is lower for sections with fewer junctions, which seems intuitively correct although it has been seen above that the relationship is not strong. Again, the sections with few or no junctions do not have a consistently lower variation in journey time, and no pattern of difference is seen for junctions with or without priority.

2.5.6 Hence it can be concluded that a small number of junctions is necessary to achieve higher speeds and lower variation, but is not a sufficient condition in itself. On the evidence provided thus far, there appears to be no difference in impact according to whether there is bus priority or not. However, with such a small sample it is likely that other factors are involved which may explain this and it is not the objective of this study to consider or estimate the bus journey time which would occur if there were no bus priority.

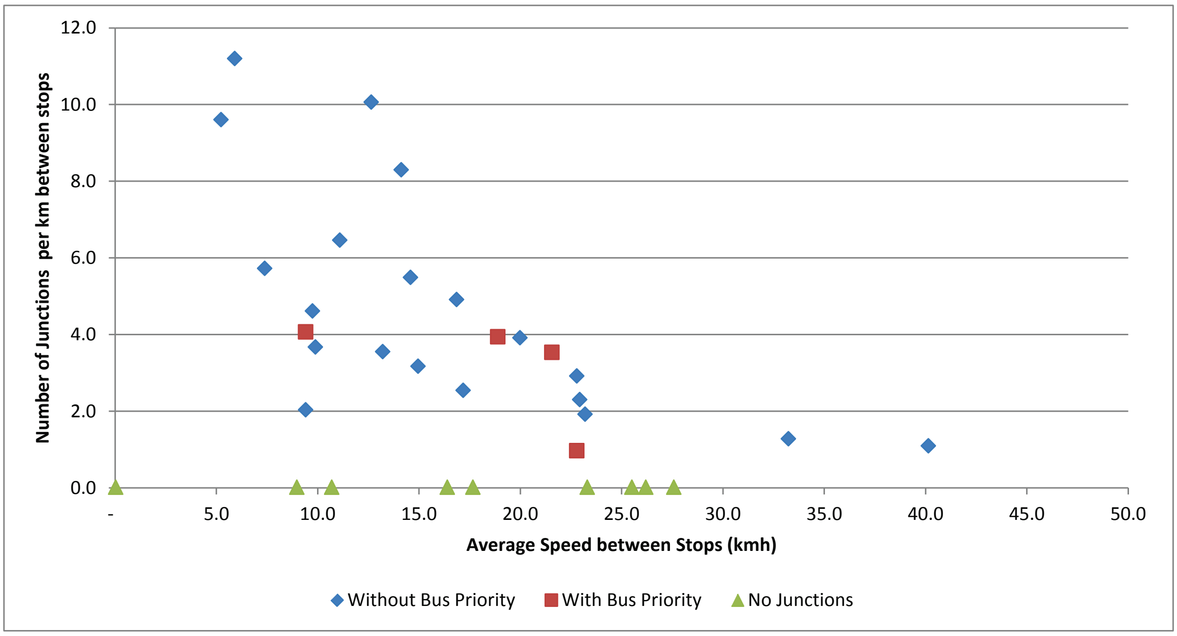


Figure 11: Number of Junctions per km by Average Speed between stops

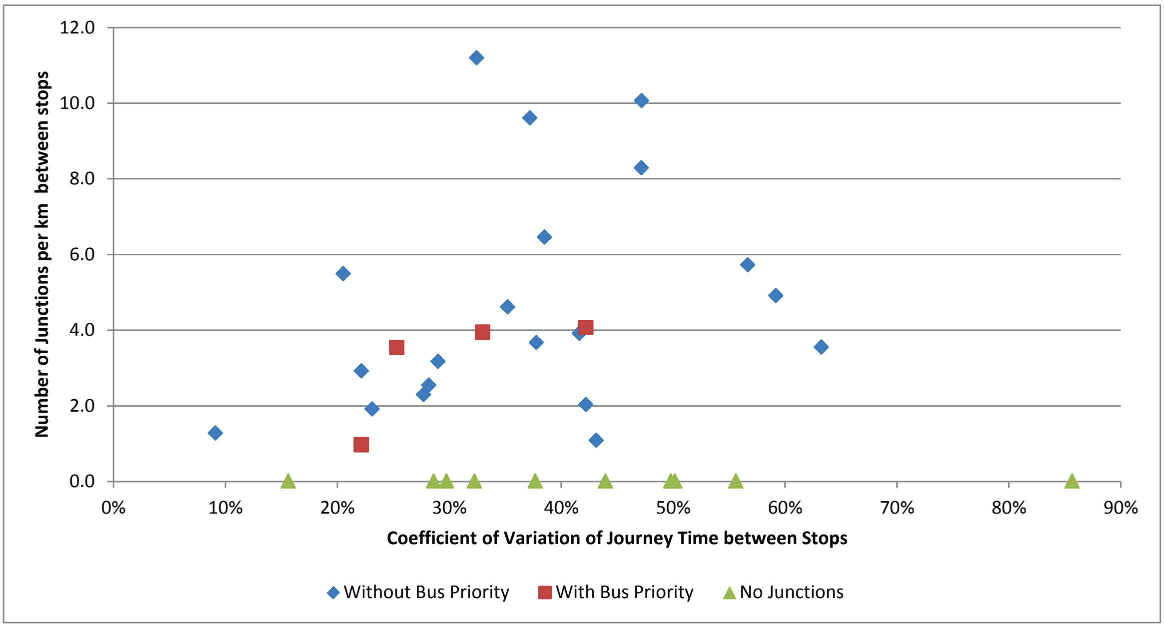


Figure 12: Number of Junctions per km by Coefficient of Variation between stops

2.6 Influence of General Traffic

2.6.1 As the data collection element of the study was concerned primarily with bus-specific data, such as passenger characteristics, Trafficmaster GPS data was processed to provide a sample of general traffic speeds and journey times for comparative sections of the bus route. The processing of this data is outlined in 'Data cleansing and preparation' in Appendix A, Pilot Study Data Collection. This data provided a total of 127 incidents where the Trafficmaster and bus journey time data are co-incident in terms of time and location i.e. the same route section is covered during the same 15 minute period. This covers 23 of the possible stop-stop route sections, of which 11 have either 1 or 2 observations, and only 5 have more than 10 observations. This makes variation in journey time extremely hard to determine on the basis of 'same time, same place'. This selective dataset has therefore been analysed alongside the full set of observations for each route section, considering bus and Trafficmaster data independently. The two approaches will clearly address slightly differing questions, as follows:

2.6.2 **'Same time, same place':** Demonstrating to what extent at a specific location variations in bus journey times are related to variability in the journey time of general traffic occurring at the same time. Given the fixed distances involved it may be clearer to consider this in terms of speeds: does the bus speed rise/fall on a section in line with the general traffic speed?

2.6.3 **'Different times, same place':** This approach yields more data, but can only consider whether a given location which exhibits variation in bus speeds also exhibits variations in speeds of general traffic. This would lead to conclusions about types of location. For this dataset, around 1,100 observations are available across 28 sites, with all but 4 sites having more than 10 observations.

2.6.4 It is unlikely that either comparison would establish a causal link between bus journey times and general traffic, however this information can give an idea of to what extent bus journey times and traffic speeds are related.

Same Time, Same Place analysis

2.6.5 Figure 13 below shows the comparison of bus and traffic speed at the four locations with the highest number of 'same time, same place' observations available. This shows very little direct correlation between bus speeds on a route section and the general traffic speed. There is a weak positive correlation in two locations, but no correlation at all in the remaining two locations (which have slightly more data points). This appears to suggest that for any given route section the lengthening of bus journey times is not directly caused by slow traffic. However, this may relate to the times of day surveyed or types of location. Further analysis with a greater number of data points could be expected to yield additional insight.

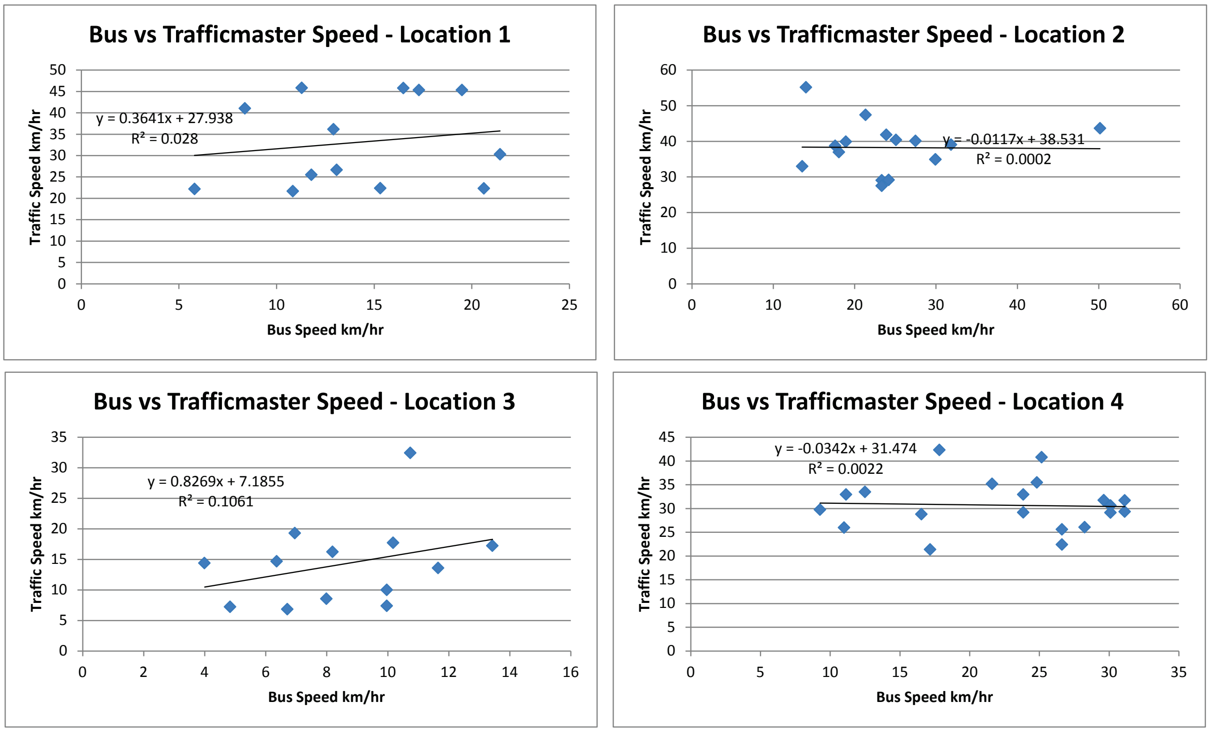


Figure 13 - Bus Speeds vs Trafficmaster speeds for 'same time, same place', locations with highest number of observations

Different Times, Same Place analysis

2.6.6 The charts in Figure 14 (below) compare characteristics of bus and general traffic journey time variation for comparable route sections. In each case, all available data for the comparable route sections are included, not just that collected simultaneously for both bus and general traffic. This leads to some conclusions which may provide greater insight.

2.6.7 The first chart (a) shows that the amount of variation in bus journey time observed for a route section is related to the variation in journey time for general traffic, though not strongly (R-squared = 0.23). The second (b) shows that the average speeds are more strongly related (0.56), though bus speeds as might be expected are consistently lower than general traffic speeds. This appears counter to the conclusion from the 'same time, same place' analysis above. However, it should be remembered that this information relates to the average speed and variation in journey time. It is therefore suggesting that when comparing locations, those where traffic speeds are low or variable, will also have lower/more variable bus speeds (and vice versa).

2.6.8 The 'same time, same place' analysis may begin to show similar patterns with more data and careful selection of route sections and survey periods. Alternatively, it may be that in the 15 minute time periods used for Trafficmaster data there is not a strong link between traffic and bus speeds, but that characteristics of a location lead to similar variations over a longer period.

2.6.9 Charts (c) and (d) show how the average speed of general traffic and buses respectively relate to variability in journey time. In both cases, it can be seen that speeds are generally lower where variation in journey time is high. This is intuitively correct, as the data range for speed is likely to be 'capped': a road with an average speed of 50km/hr would not be expected to exhibit cases of speeds significantly higher, whereas a road section with an average speed of (say) 10km/hr could easily have instances of speeds significantly above or below that figure. It also appears intuitively correct that general traffic speeds show a stronger relationship than bus speeds.

2.6.10 Charts (e) and (f) show that in each case the variation in journey time is <u>not</u> strongly related to the length of the route section. This seems to add weight to the earlier observation that it is the number of stops encountered by the bus which allows journey time recovery and thus smoothes out variation, not the distance driven. The fact that general traffic also does not show a relationship again suggests that the earlier pattern is related specifically to buses, and not a feature of general traffic – however longer route sections for general traffic should be studied to support this conclusion.

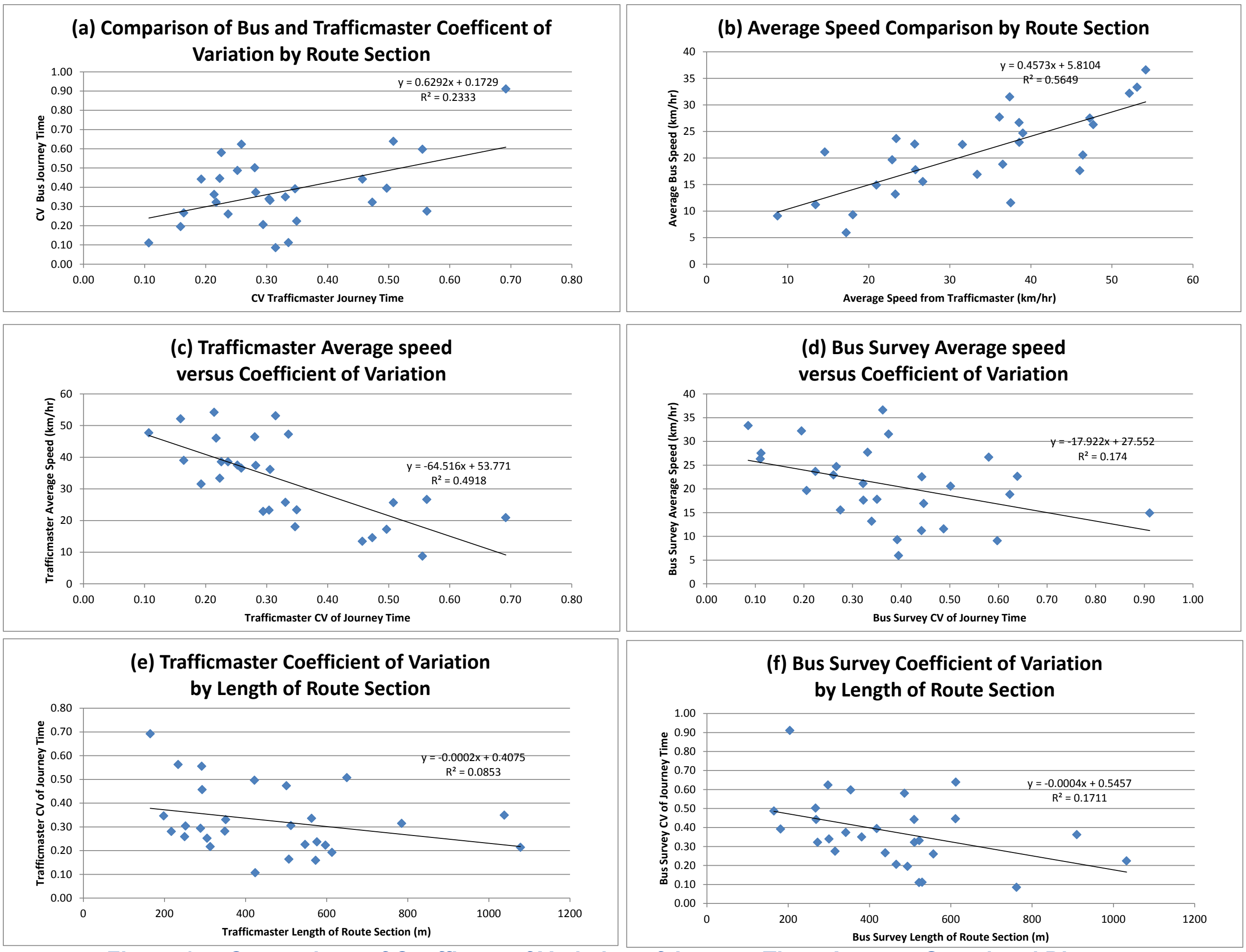


Figure 14 - Comparison of Coefficent of Variation of Journey Time, Average Speed and Distance for comparable route sections (Bus Survey and Trafficmaster data)

Conclusion

2.6.11 In conclusion on TrafficMaster data, it appears there are some interesting correlations between bus speeds and those of the general traffic. However, the relationships between CV of journey time for bus and traffic are not particularly strong. Given more data, it is possible further relationships could be found but at present this is not suitable for inclusion in a regression model.

2.7 Influence of Total Number of Passengers Boarding and Alighting

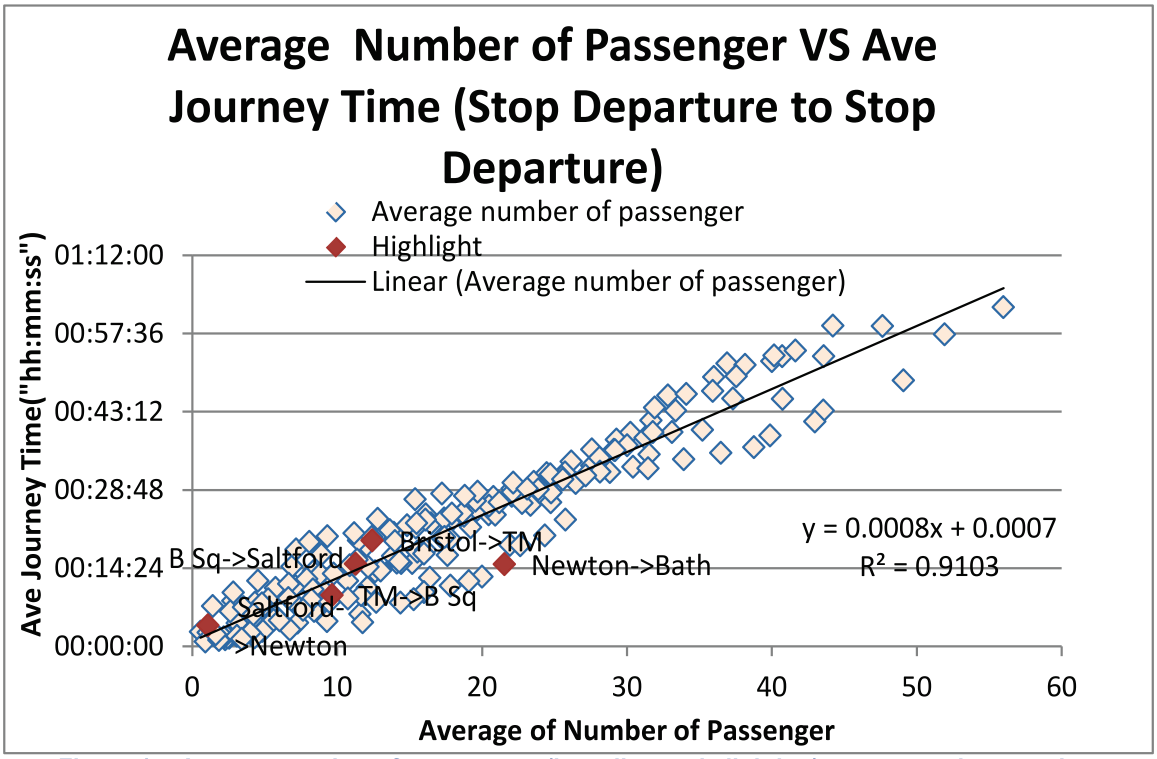


Figure 15: Average number of passengers (boarding and alighting) vs average journey time

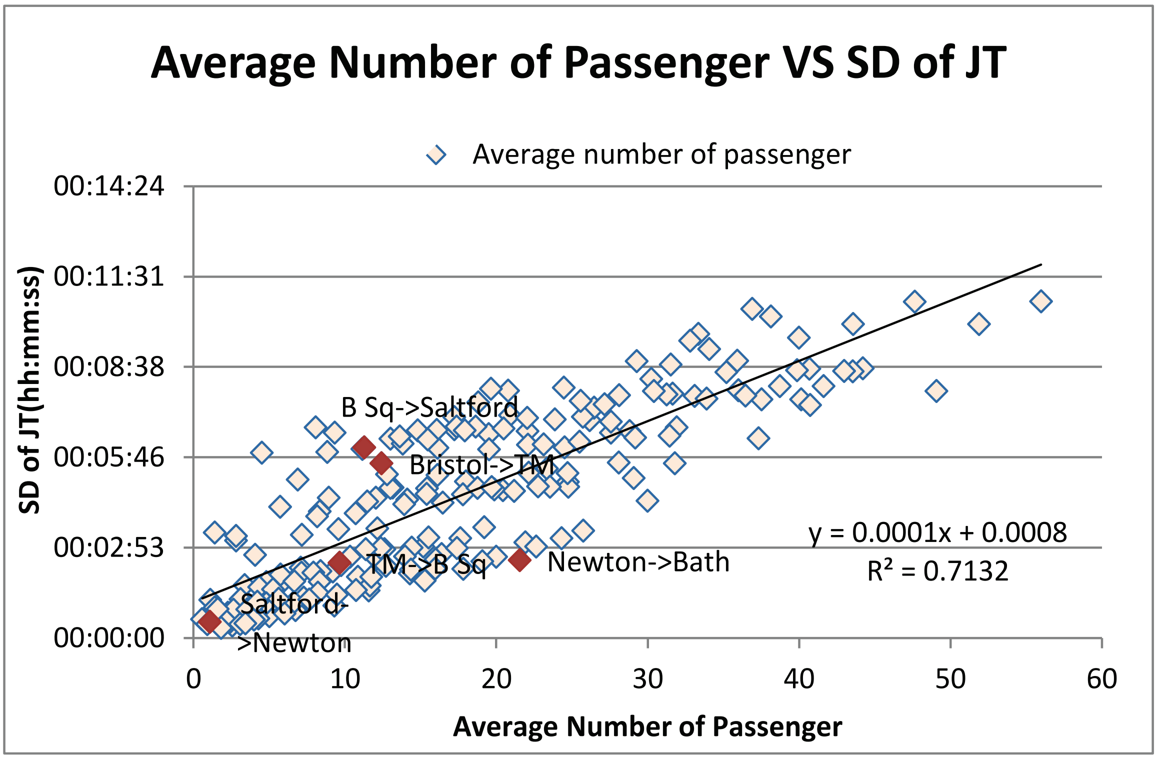


Figure 16: Average number of passengers (boarding and alighting) vs SD journey time

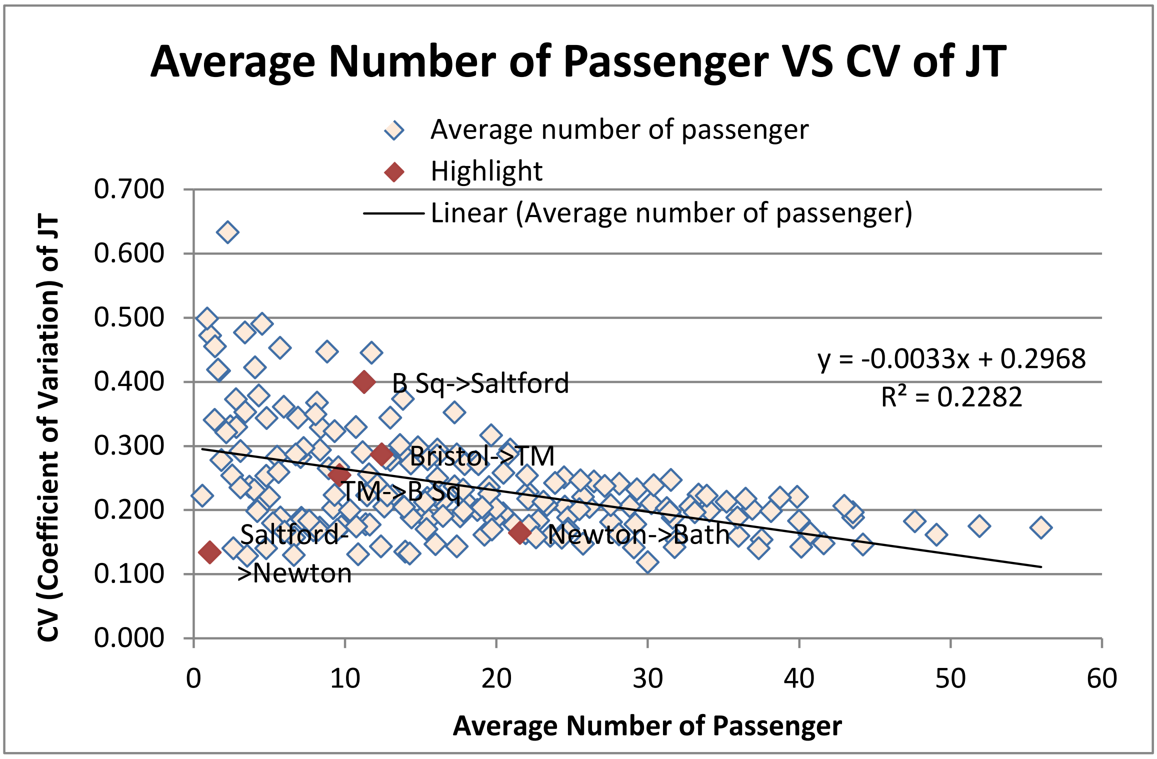


Figure 17: Average number of passengers (boarding and alighting) vs CV of journey time

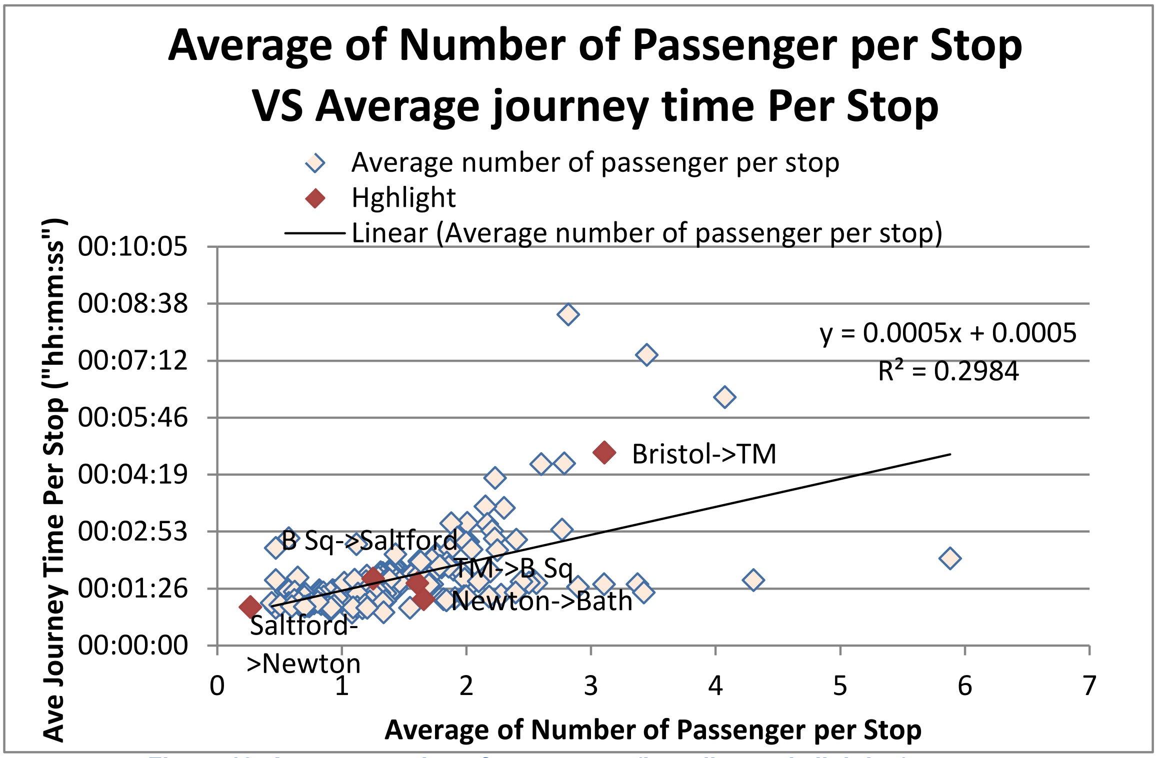


Figure 18: Average number of passengers (boarding and alighting) per stop vs average journey time per stop

2.7.1 It is of course the case that buses behave differently to general traffic in that they stop periodically to set down and pick up passengers. While it is expected that the bus schedule will take account of this, the effect of passengers, who will not all board the same bus every day and/or at the same time, is one of the most variable bus-specific elements (as opposed to junctions and general traffic conditions which will affect all vehicles). Figure 15, Figure 16 and Figure 17 consider the influence of the average number of passengers on the journey time. The total number boarding and alighting for each route section has been considered, averaged across the journeys[6]. This presents quite a similar pattern to the previous two variables, i.e. that the variable itself is very strongly related to journey time, and that therefore it is difficult to conclude whether the fall in CV with increasing passengers is actually related to the passenger numbers or the journey length. Logically, the latter seems more likely. Also, as with previous variables, there is some sign of heteroscedasticity in the CV chart.

2.7.2 A further plot has been made which shows the average number of passengers per stop against the average journey time per stop (see Figure 18). This chart suggests that when there are more passengers per stop, the travel time per stop (from arrival at one stop to the next, hence including both dwell time and time to pull away from a stop) becomes longer, though there is no strong linear trend. The section of Bristol Bus Station to Temple Meads stands highest with both large values of passengers and travel time, which is entirely logical for this service as Bristol is the principal departure point in the afternoon period for the whole route.

2.7.3 In order to further investigate the influence of passenger numbers on each journey, Figure 19, Figure 20, Figure 21, Figure 22 and Figure 23 examine how passenger numbers influence journey time for each valid journey record on the 5 route sections (Bristol Bus Station to Temple Meads, Temple Meads to Brislington Square, Brislington Square to Saltford Tyning Road, Saltford Tyning Road to Newton St Loe, and Newton St Loe to Bath Bus Station). As mentioned previously, the number of valid records varies for each section. The sections from Saltford Tyning Road to Newton St Loe and from Newton St Loe to Bath Bus Station only have 35 and 42 valid records respectively.

2.7.4 It can be seen that generally, the number of passengers correlates to the travel time on each route section: as more passengers are involved, the journey then takes more time. However, the relationship again is not linear. In particular, the outliers with journey times of greater than 20 minutes for the Brislingston Square to Saltford section have low passenger numbers. These observations may merit some further investigation.

2.7.5 In interpreting the influence of passenger numbers, it may be worth considering whether this variable operates in isolation. Higher passenger numbers are likely to correspond with peak periods where traffic levels are also high, so that traffic congestion may also be contributing to delays. As shown in the previous section which considers the available Trafficmaster data, sub-divisions of such datasets inevitably result in less reliable subsets and therefore no temporal sub-division has been made at this stage (however, see Section 3 for a comparison of peak and off-peak correlations).

[6] Note: It was decided to sum the boards and alights for these purposes as this represents the total 'churn' of passengers at stops, and avoids a measure which isolates only one element, or considers only the total number on the bus, which is unlikely to be relevant to journey time.

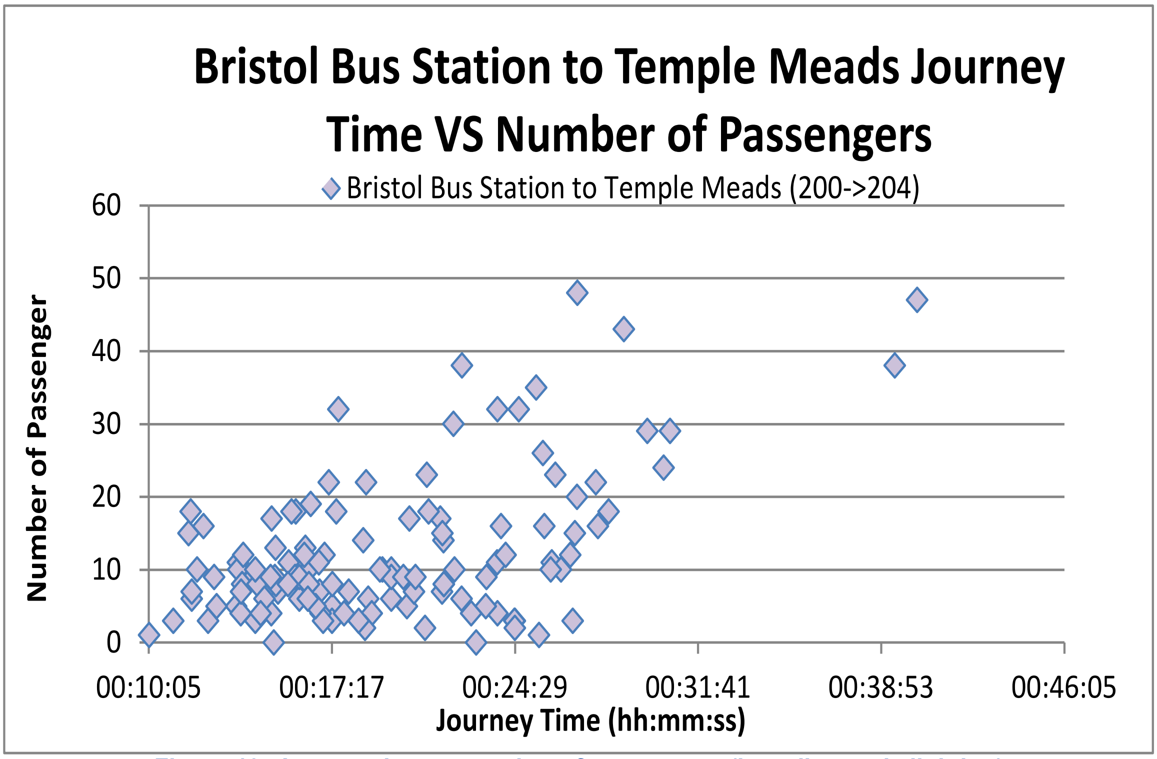


Figure 19: Journey time vs number of passengers (boarding and alighting) from Bristol bus station to Temple Meads

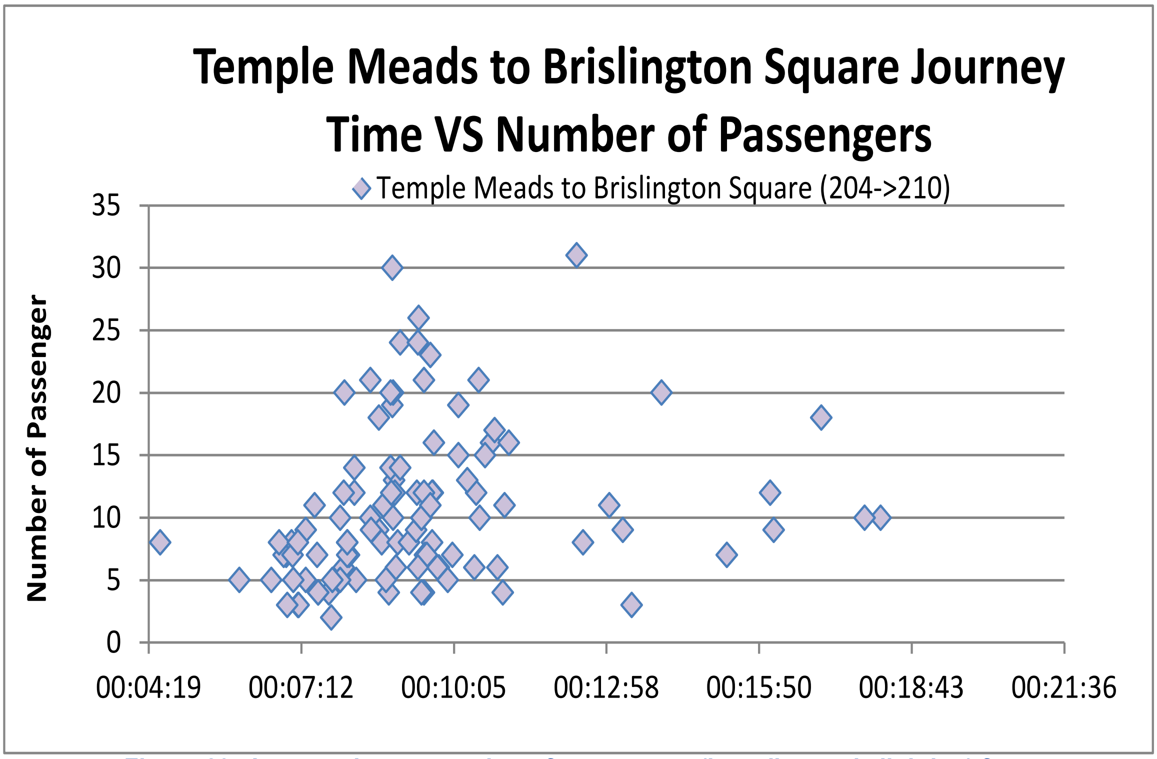


Figure 20: Journey time vs number of passengers (boarding and alighting) from Temple Meads to Brislington Square

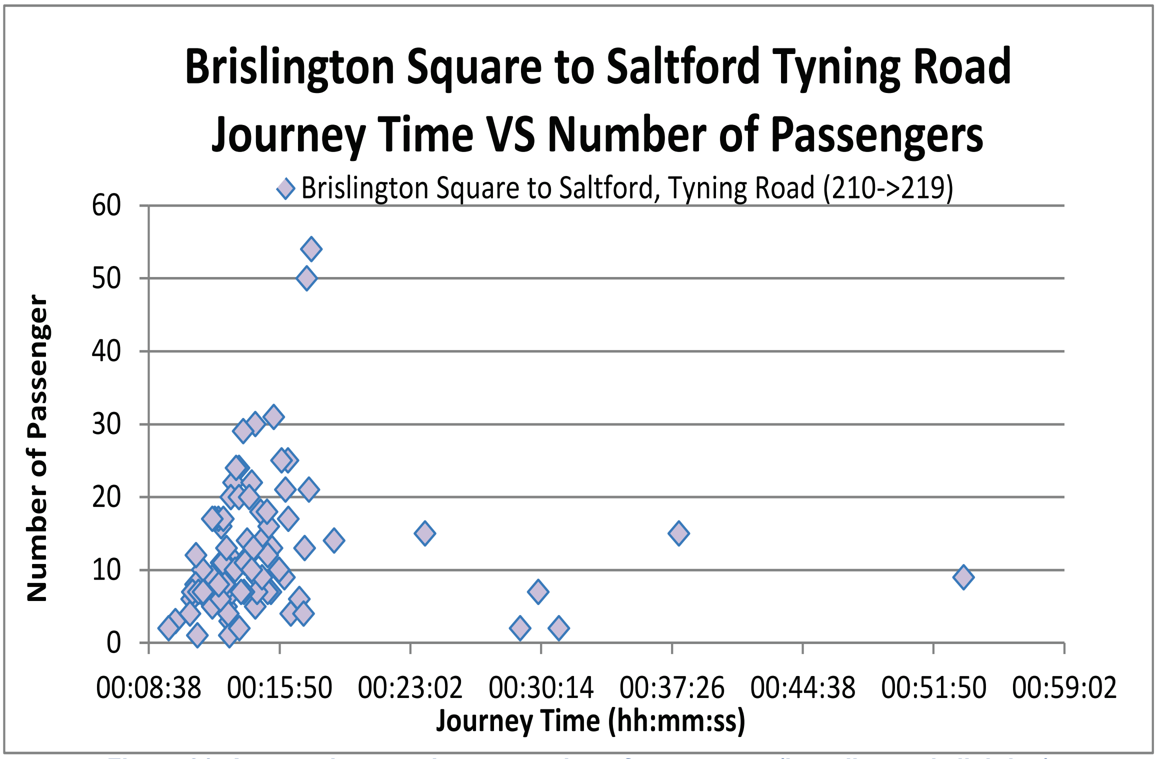


Figure 21: Average journey time vs number of passengers (boarding and alighting) from Brislington Square to Saltford Tyning Road

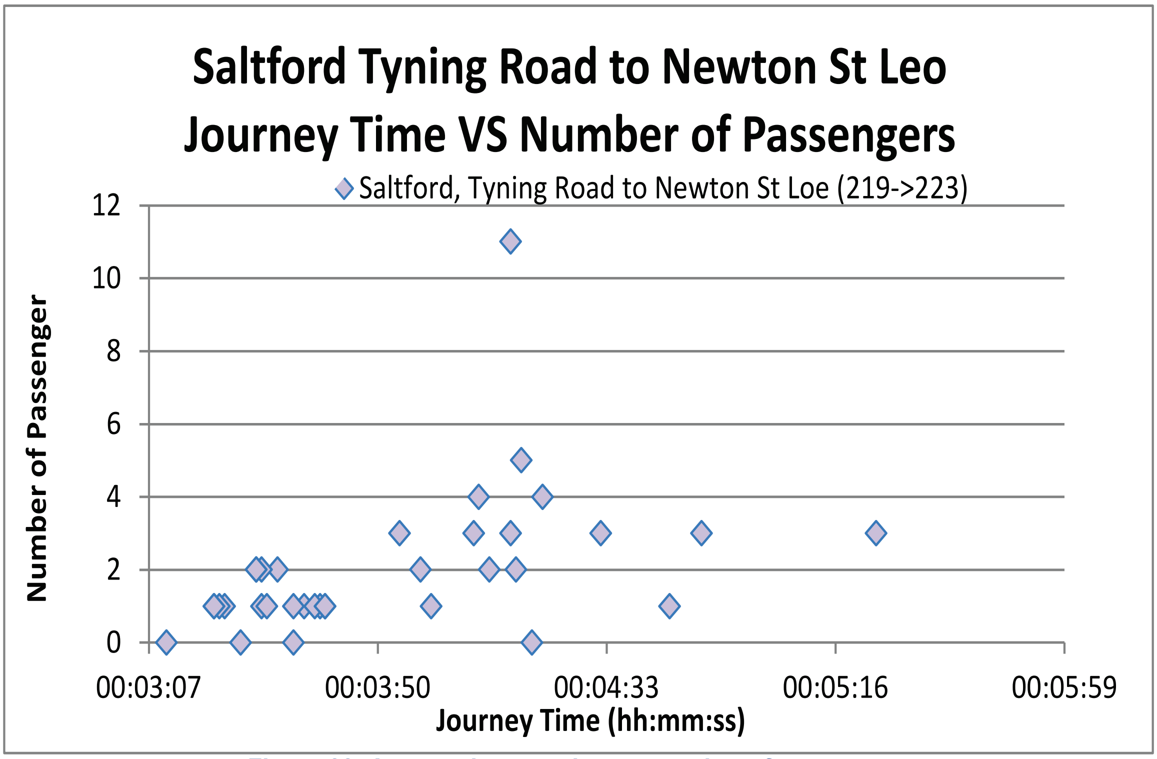


Figure 22: Average journey time vs number of passengers (boarding and alighting) from Saltford Tyning Road to Newton St Loe

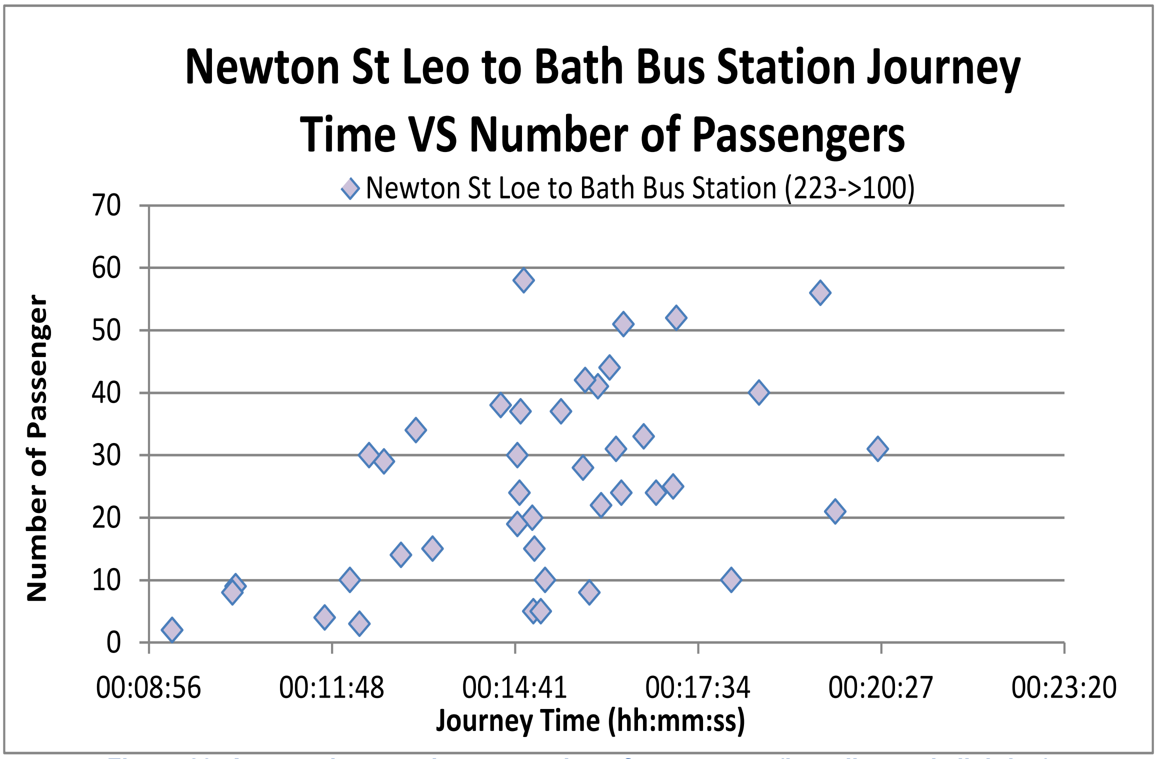


Figure 23: Average journey time vs number of passengers (boarding and alighting) from Newton St Loe to Bath Station

2.8 Influence of Cash or Non-Cash Payment

2.8.1 A further consideration in respect of passengers is the purchase/validation of tickets (including payment of a fare). From the literature review it was identified that ticket type was an element with medium importance to bus journey time variability, but one which the operator could control to a much greater degree than some of the high importance elements. While recording the specific ticket type was not practical, whether customers pay by cash or use non-cash means (i.e. return ticket portions, season tickets and other travelcards) was a recordable factor which could influence journey time. The variation in journey time has therefore been studied according to the total number of cash and non-cash payments made by boarders. The results are presented below, though it should be noted that the proportion of cash payments appears to be low on this route, making firm conclusions more difficult, although it indicates that there is limited scope on this service to further reduce variability based on this factor. This low proportion is however logical given that the survey focused on the afternoon and evening peak period, when many passengers will be returning home and are thus likely to be using tickets purchased earlier in the day (or be a regular commuter using a ticket or pass purchased previously). It should be noted that whereas previously all boarding and alighting passengers were considered, this analysis necessarily focuses only on boarders.

2.8.2 Figure 24 shows the relationship between the number of cash fare passengers (who pay cash as they board the bus) and the average journey time. The major feature to note is that the number of cash payments is typically very low: less than 1, and never more than 5. Furthermore, there is a jump in the journey time value between 1 and 2 passengers.

2.8.3 Manual investigations of the data were undertaken to better understand these patterns. It was concluded that the route sections with fewer than 2 cash fare passengers were in very different areas to those with 2 or more cash fare passengers. Essentially, there are two different groupings of the route sections, with the first group being in areas where few people boarded, and the latter in areas where more people board (as illustrated by the Bristol Bus Station to Temple Meads route section).

2.8.4 It appears that there is a difference in both absolute journey time and cash payment proportions for route sections which include the busiest stops. This seems logical as route sections are shorter in busy areas, and those areas are more likely to have 'casual' passengers who pay cash. However, importantly this distinction disappears when the coefficient of variation is considered (Figure 26): the CV chart does not show the same groupings, indicating that the cash paying passengers appear to have the same proportional impact on journey time variability, regardless of whether the route section includes these busy areas.

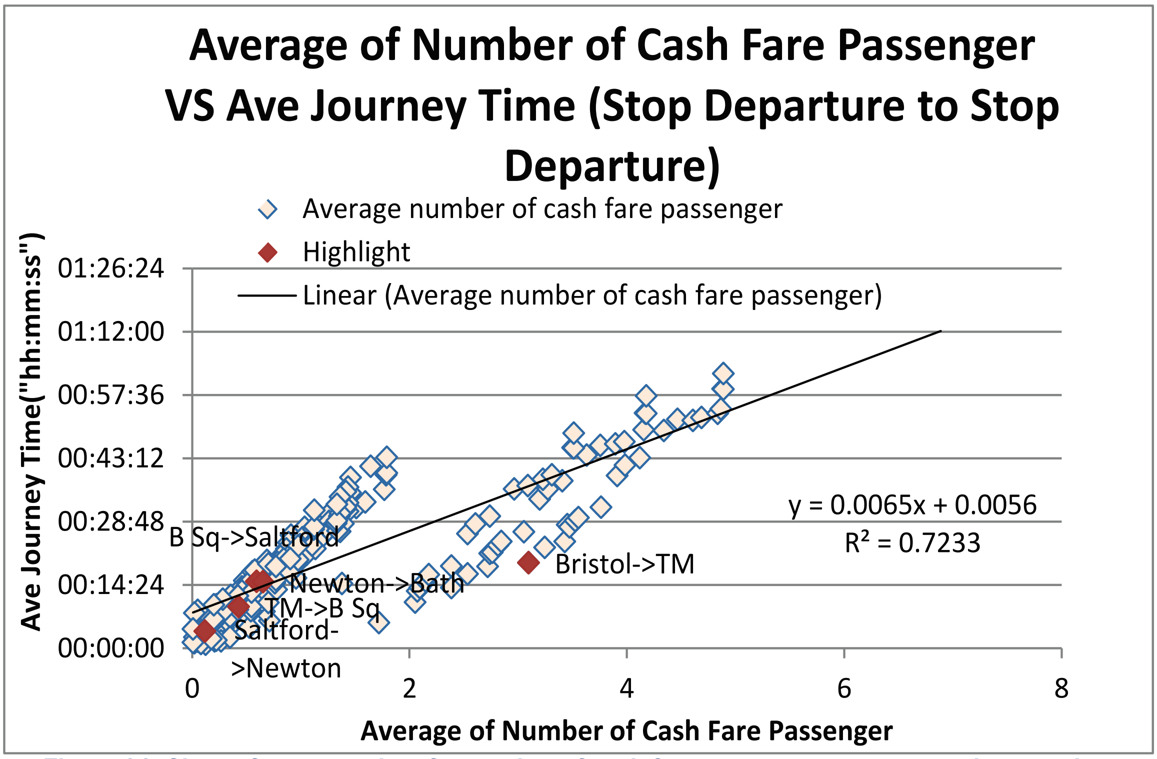


Figure 24: Chart of scatter points for number of cash fare passengers vs average journey time

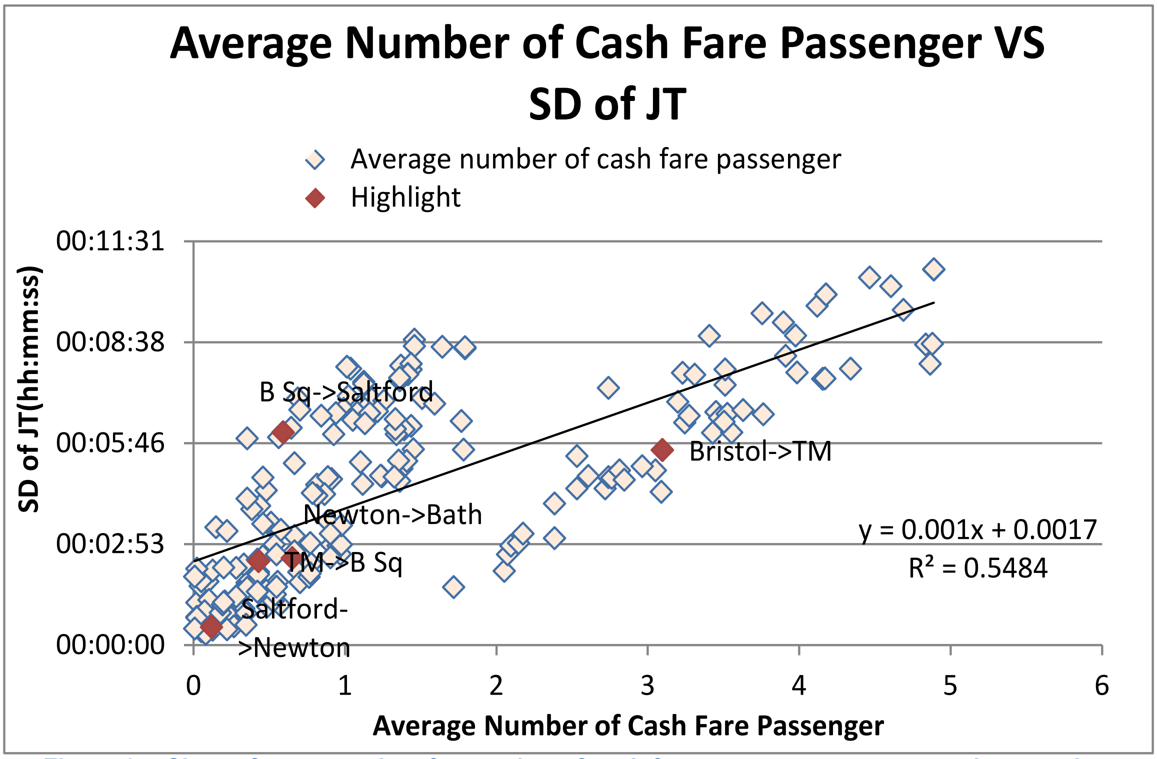


Figure 25: Chart of scatter points for number of cash fare passengers vs average journey time

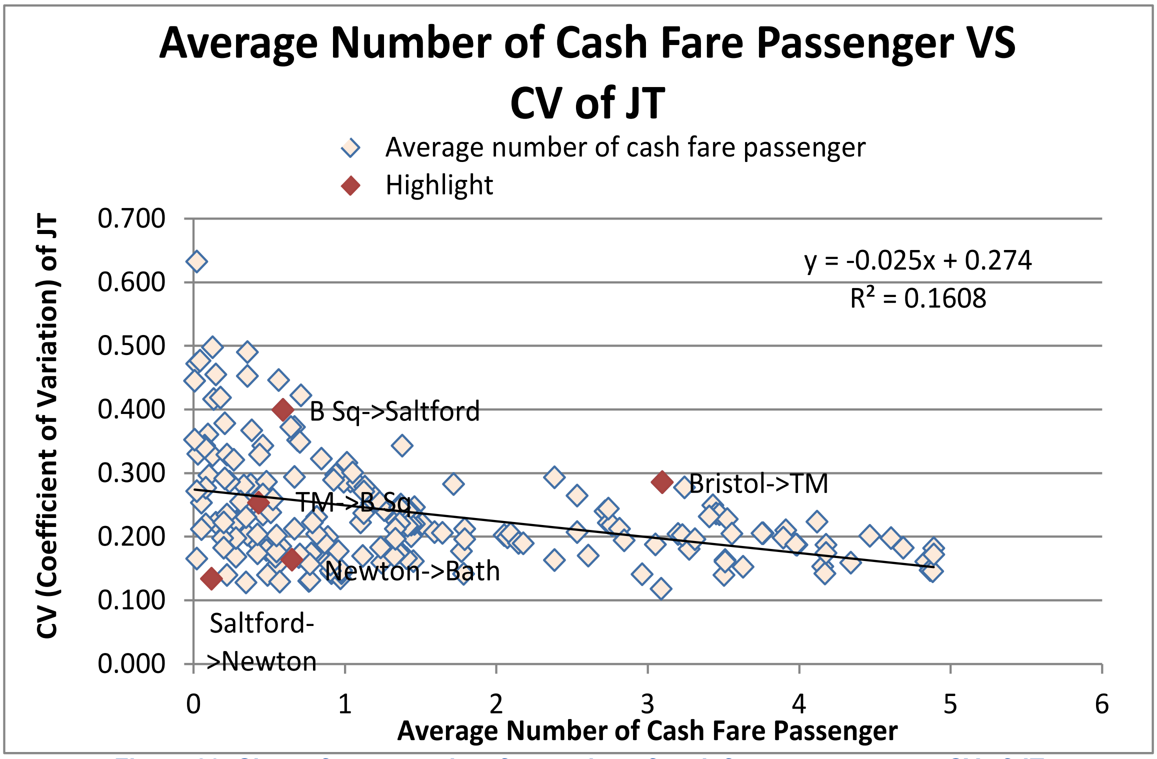


Figure 26: Chart of scatter points for number of cash fare passengers vs CV of JT

2.8.5 Figure 27 below shows the relationship between the number of non-cash fare passengers (who do not pay cash on the bus) and the average journey time and presents quite a similar pattern to other variables.

2.8.6 Generally, significantly more non-cash fare passengers were recorded in the survey than cash fare passengers. As with cash passengers, sudden falls in journey time are seen as non-cash passenger numbers increase. This can again be explained by considering the density of passenger boarding at certain stops, and the grouping of route sections with and without stops with heavy boarding.

2.8.7 Once again, the consideration of CV helpfully eliminates this pattern. It can be seen in Figure 29 that the non-cash passengers have a weaker correlation with CV of journey time than cash passengers (R-squared <15%), and that the slope is considerably lower (at -0.0069). This suggests that cash passengers do have a stronger influence on journey time variability than their non-cash counterparts, which appears intuitively correct.

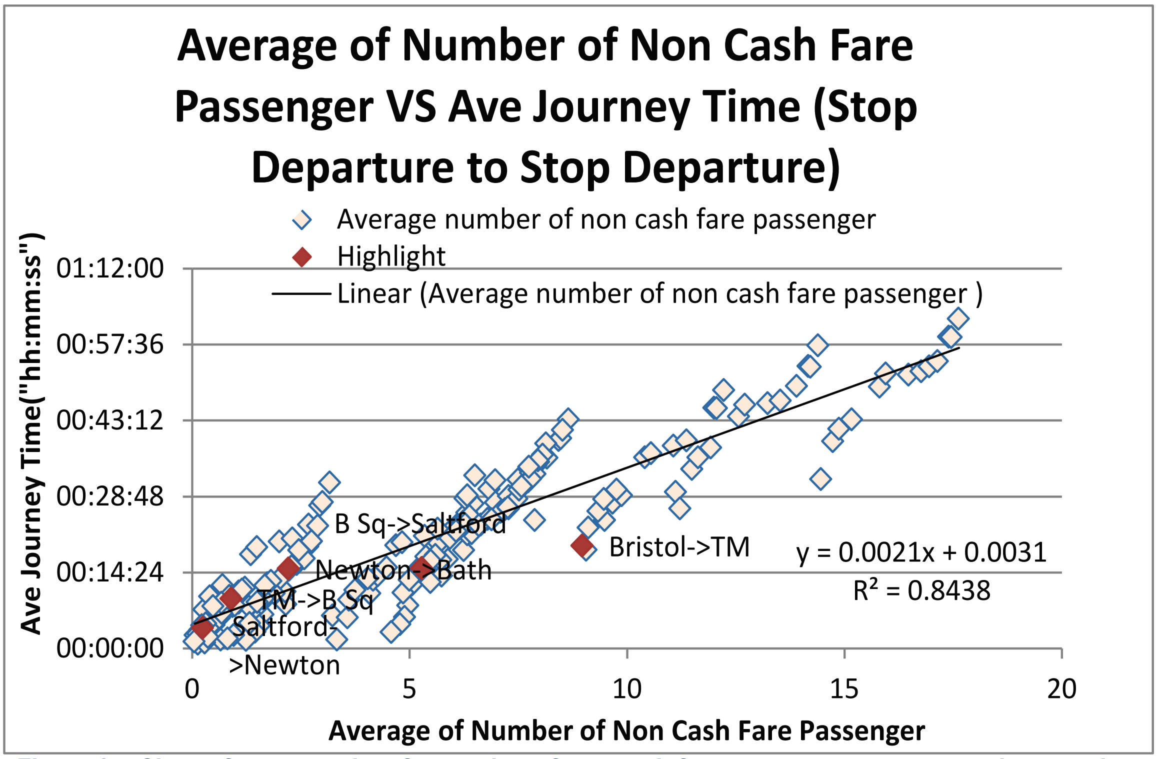


Figure 27: Chart of scatter points for number of non-cash fare passengers vs average journey time

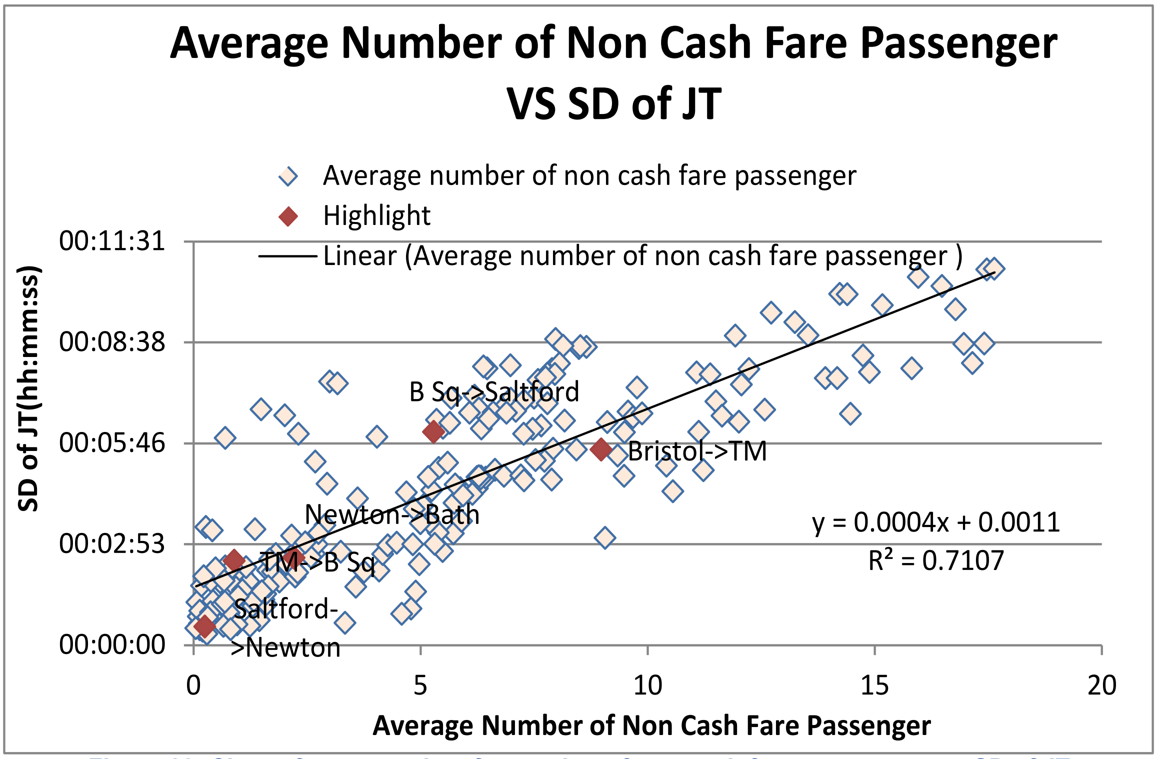


Figure 28: Chart of scatter points for number of non-cash fare passengers vs SD of JT

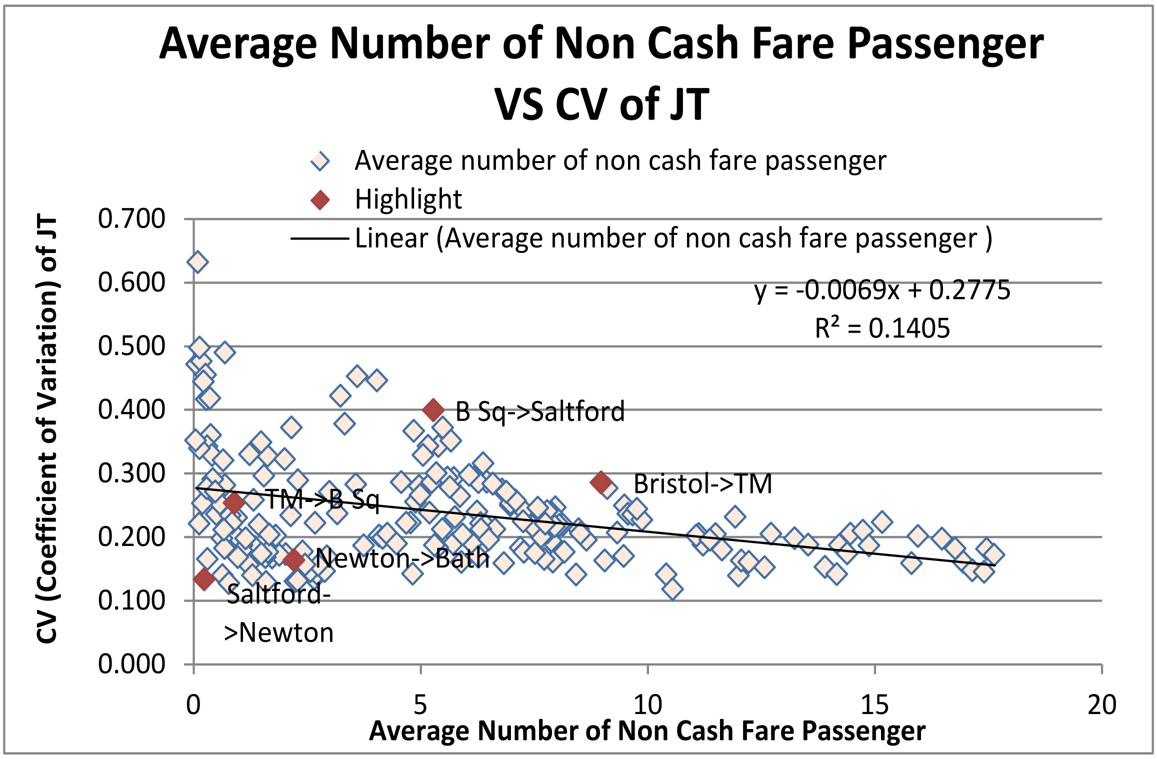


Figure 29: Chart of scatter points for number of non-cash fare passengers vs CV of JT

3 Correlation Coefficients and Regression Model

3.1 Analysis of Correlation Coefficients

3.1.1 One of the purposes of the study is to identify the factors having greatest influence on the variability of journey time. The linear correlation coefficient values are statistical measures of the strength and direction of the linear relationship between two variables. The value of the correlation coefficient is between -1 and 1, with higher absolute coefficient values indicating stronger linear dependency between two variables.

3.1.2 The following tables summarise the linear correlation coefficient values of all the potential variables against the CV of journey time values. This analysis has been carried out first for ALL journeys ('Overall' figure), then separately for peak and off-peak journeys. This provides a useful indication of how the nature and strength of the relationship varies by period.

3.1.3 It should be noted that for comparability, these analyses have all been conducted with a minimum number of observations per route section of 10[7].

Independent Variable	Explanatory Variables	Scenarios		
		All[a]	Peak	Off Peak
CV of Journey time	Number of junctions traversed on route section	-0.30	-0.23	-0.43
	Number of stops passed (the number of stops on the line of route, irrespective of whether the bus stops or not)	-0.41	-0.30	-0.59
	Number of stops called (the number of stops on the line of routes where bus has stopped to drop off/pick up passengers)	-0.40	-0.31	-0.53
	Ratio of number of stops called to number of stops passed	0.28	0.18	0.56
	Distance between stops	-0.38	-0.24	-0.61

Table 2: Summary of correlation coefficient values of all number of stop related variables against CV of journey time

Independent Variable	Explanatory Variables	Scenarios		
		All[a]	Peak	Off Peak
CV of Journey time	Average number of passengers per stop passed	0.23	0.19	0.33
	Average number of passengers per stop called	0.13	0.14	0.11
	Average number of passengers	-0.36	-0.29	-0.49
	CV of number of passengers (total boarding and alighting on route section)	0.51	0.38	0.52

Table 3: Summary of correlation coefficient values of Passenger related variables against CV of journey time

[7] see section 1.4 for a further discussion on sample sizes

Independent Variable	Explanatory Variables	Scenarios		
		All[a]	Peak	Off Peak
CV of Journey time	Average number of cash fare passengers	-0.38	-0.28	-0.47
	Average number of cash fare passengers per stop passed	0.00	0.02	0.05
	Average number of cash fare passengers per stop called	-0.06	-0.01	-0.06
	Average number of non-cash fare passengers	-0.39	-0.23	-0.52
	Average number of non-cash fare passengers per stop passed	0.03	0.08	0.02
	Average number of non-cash fare passengers per stop called	-0.06	0.07	-0.20

Table 4: Summary of correlation coefficient values of payment related variables against CV of journey time

3.1.4 The values referred to in the following text are relevant to the total survey period unless otherwise specified.

3.1.5 The results show that the CV of the number of passengers has the strongest positive influence on the CV of bus journey time with a correlation coefficient value of 0.51, which indicates that with the increase of CV of the number of passengers, the CV of journey time increases.

3.1.6 The number of stops has the strongest negative influence on the CV of bus journey time with a correlation coefficient value -0.41, which indicates that the more bus stops the bus passes, the smaller the CV of journey time tends to be, as already noted in paragraph 2.3.2.

3.1.7 This relationship probably explains some of the other high negative correlations with CV of Journey Time: all of the factors related to number of passengers, number of junctions etc. show this pattern. This is unlikely to be for any reason which is helpful for this analysis; it is simply the case that these represent longer route sections where 'catch-up time' diminishes the CV of journey time[8]. It is suggested that timetable structure may play a role here, as either the scheduled dwell time or other 'catch-up' time in the timetable at stops will allow variations in journey time to be smoothed out during each journey.

3.1.8 Due to this strong negative relationship of distance–related factors and CV of journey time, a further factor 'passengers per stop' was created, designed to test the correlation of the number of passengers, without the influence of distance. It was found that this has a lower correlation to CV of journey time (0.23). However, a similar relationship was not seen when testing cash/non-cash passengers per stop in Table 4.

3.1.9 Table 3 also differentiates between the total stops **passed** by each bus and those at which the bus **called**. It was felt that this might demonstrate that 'skipping' stops provides some additional benefit in terms of reducing the variation in journey time. However, there is comparatively little difference in the correlations when comparing the measures directly. The ratio of stops called to total stops was also tested, and found overall to have a weak positive correlation with CV (0.28), i.e. variability increases if the bus calls at a greater proportion of the stops passed.

[8] As we have seen in the previous section, the standard deviation (being an absolute measure of variation) still rises as these figures increase. The coefficient of variation follows a counter-pattern because it is a proportional measure.

3.1.10 When looking at the values in terms of peak period and off peak period, it can be seen that the correlation is usually much stronger in the off-peak than peak period, with the relationship often breaking down in the peak period. The 'overall' correlations are typically less strong than the off-peak correlations, indicating that the lack of a strong relationship in the peak is decreasing those figures. The correlation with stops called/stops passed is particularly high in the off peak, at 0.56.

3.1.11 The reason for this weaker peak correlation is thought to relate to the higher road traffic levels and increased congestion in the peak: we know that it cannot be caused by increased passenger numbers, as our passenger measures also show a weaker peak correlation.

3.2 Multi-regression for all time periods

3.2.1 Another purpose of the bus journey time variability study is to consider the elements required to build a model that shows the relationship between the factors having strong influence on bus journey time and the CV of bus journey time. Hence multiple-regression models have been tested. Models have been built based both on the whole period and separately for the off-peak, as this had significantly stronger correlations.

3.2.2 The factors for inclusion in the multi regression model have been selected based on:

- Strength of the correlation coefficient values, as shown in the tables above;
- Some judgement is used to select a package of factors which appear to have the most logical link to journey time variation for buses.

3.2.3 On this basis, the following factors were selected for the first model:

- Total number of stops passed: This is the 'journey length' factor with the strongest correlation to CV of Journey Time, with a stronger negative correlation than simply average journey time, distance or number of passengers, possibly due to timetable structure factors (e.g. catch-up time);
- CV of number of passengers: This has the strongest relationship to CV of journey time overall, with the logical implication that for the same journey, the trips with higher numbers of passengers will frequently take longer, and vice versa;
- Average passengers (boarding and alighting) per stop: Although this does not show as strong correlation as the other two factors, it is used in the model as it is a key factor of concern in our study.

3.2.4 It is generally good practice in regression model to minimise the number of variables added, as misleading results can be produced by adding poorly correlated factors. Models with large numbers of variables can easily achieve a high level of fit against the observed data, whilst being extremely poor at predicting outcomes, and hence not giving any useful information about the overall mechanisms involved.

3.2.5 The regression model produces the following equation to numerically describe the relationship between the CV of journey time and other factors. The model also tries to forecast the CV of journey time.

$$Y=\theta+\beta 1*X1+\beta 2*X2+\beta 3*X3$$

Where:
X1: The number of stops,
X2: CV of the number of passenger,
X3: the average number of passengers per stop
Y: dependent variable, CV of journey time

ϑ (intercept), $\beta 1$, $\beta 2$, $\beta 3$ (slope parameters) are the values to be determined by the multiple-regression model.

The regression output is listed as below:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.599583288
R Square	0.359500119
Adjusted R Square	0.349746314
Standard Error	0.068891675
Observations	201

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.13486	0.035093012	3.8429007	**0.00016**	0.065652768	0.204065
X1 (Number of Stops)	-0.00201	0.000736056	-2.724191	**0.00703**	-0.00345672	-0.00055
X2 (CV of number of passengers)	0.14014	0.033740564	4.1533883	**0.00005**	0.073598605	0.206677
X3 (Average Passengers per stop)	0.01932	0.005904517	3.2718519	**0.00126**	0.007674531	0.030963

3.2.6 With the above regression values, a regression model to forecast the CV of bus journey time values has been developed. The equation of the model is:

$$Y=0.13486-0.00201*X1+0.14014*X2+0.01932*X3$$

3.2.7 In the summary output, R Square indicates the goodness of fit, higher value of R Square indicates the model fits the data better. P-value indicates the probability of rejecting the null hypothesis. A low p value (<0.05) indicates that the factor is meaningful in the model and that changes in the dependent variable (Y) are related to changes in this factor.

3.2.8 The variables of X1, X2 and X3 all have estimated p-values smaller than 0.05. Therefore, they have significant correlation with the CV of bus journey time.

3.2.9 The overall R-Squared value of 0.36 indicates that the model fits the data to some extent (the inclusion of the number of stops, CV of the number of passengers and the average number of passengers per stop explain 36% of the variety of bus journey time), but there remains a large amount of variation not explained. This is not unusual for this type of model, and indicates that much variation is caused by factors we cannot capture within the datasets tested for this study.

3.2.10 The coefficients are intuitively correct, in that the variation in journey time will <u>fall</u> as the number of stops rises and <u>rise</u> if either the number of passengers becomes more variable <u>or</u> the number of passengers per stop rises. It should be noted that the regression model can say nothing at this stage about the physical conditions of the route section.

3.2.11 In conclusion, the regression model shows that the number of stops, CV of number of passenger, and average number of passenger per stop have reasonable correlations with CV of bus journey time. The multi regression model can to some extent predict the CV of bus journey time by using these variables. An increased dataset is likely to allow more robust estimates of such a model, allowing more factors to be taken into account.

3.3 Multi Regression Model for off peak period

3.3.1 As seen from the tables in Section 3.1, after introducing the extra dimension of peak and off peak periods, off peak is better explained by explanatory variables, having higher correlation values than both the peak period and the peak and off-peak periods combined. Therefore we have also constructed a model just for those journeys taking place in the off peak period.

3.3.2 All of the existing explanatory variables in the model for overall records have been kept in this model. In addition, an extra factor of “ratio of number of stops called to number of stops passes” has been selected due to its reasonably high correlation value 0.56 (see Table 2). This factor reflects how frequently the bus stops on each route section: a value of 1 suggests that the bus has called at all stops along the target section; a value of 0 suggests that the bus did not stop at all along the target section. The form of the regression model to off peak scenario is presented as follows:

$$Y=\theta+\beta 1*X1+\beta 2*X2+\beta 3*X3+\beta 4*X4$$

Where
X1: The number of stops;
X2: Ratio of number of stops called to number of stop;
X3: CV of the number of passenger;
X4: the average number of passengers per stop
Y: CV of journey time
ϑ (intercept), β1, β2, β3, β4 (slope parameters) are the values to be determined by the multiple-regression model.
The regression output is listed as below:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.773
R Square	0.598
Adjusted R Square	0.594
Standard Error	0.060
Observations	371

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.134	0.027	-4.927	0.000	-0.187	-0.080
X1 (Number of Stops)	-0.001	0.001	-2.693	0.007	-0.002	-0.000
X2 (*Ratio of number of stops called to number of stop*)	0.469	0.039	11.885	0.000	0.392	0.547
X3 (CV of number of passengers)	0.174	0.020	8.722	0.000	0.135	0.213
X4 (Average Passengers per stop)	-0.013	0.005	-2.508	0.013	-0.023	-0.003

3.3.3 With the above regression values, a regression model to forecast the CV of bus journey time values has been developed. The equation of the model is:

$$Y=-0.1365-0.0014*X1+0.4733*X2+0.176*X3-0.0125*X4$$

3.3.4 All the P-values are below the threshold value 0.05 which indicates that all the selected factors in this model are playing important roles to explain the response of the independent variable (CV). This model also achieves a higher value of R square than the previous model (0.59 vs. 0.36).

3.3.5 The improvement is likely to be due for the most part to the restriction of the model to the off-peak period. As noted above, it appears that in the off-peak the variation in journey time is more easily related to known factors. In the peak the variation is less correlated against the factors observed.

4 Conclusions and suggestions

4.1 Data Analysis and Regression

4.1.1 The work above has looked at how the variation in journey time shown in the survey is related to a number of factors collected during the survey. The major conclusions which can be drawn are as follows:

Analysis of Variation

4.1.2 As had been anticipated, the simple Standard Deviation (SD) of journey time behaves very differently from the Coefficient of Variation (CV). SD grows larger as most measures of journey length increase (time, number of stops, total passengers). The relationships for SD are also very strong, with R-squared values for regression usually in excess of 80%. In contrast, the CV always falls when considering statistics related to journey length. However, the relationship is much less strong and perhaps exponential rather than linear in nature.

4.1.3 Multiple regression techniques do improve the ability to predict CV for bus services. However, at best we are able to explain only 36% (or 59% in the off-peak) of variation using this approach.

4.1.4 One area of caution is that the coefficient of variation shows strong signs of heteroscedasticity (i.e. non-constant variation). Typically, for lower values of journey time, number of stops etc., there is considerably more variation in the value of CV, whereas the value is in a narrower band for larger values. If proceeding to formally estimate an econometric model to explain journey time variability, it may be necessary to compensate formally for the non-constant variance. However, it is unlikely that this process would fundamentally alter the relationships shown in the work, or their relative strengths. Therefore a separate treatment of heteroscedasticity would be unlikely to alter the overall conclusions of the current work.

4.1.5 It is also important to note that the current approach is based on pairwise stop comparisons, and many of the combinations are overlapping: for example the route section from stop 201 to 202 is included in that from 200 to 202, 200 to 203, 201 to 203 etc. This may lead to correlation between observations: if stop 201 to stop 202 has a more variable journey time, then the other route sections will be affected, though proportionally to a lesser extent. To a large extent this is an inherent and deliberate part of the methodology agreed for this study. We have attempted to account for the impact by explicitly highlighting the non-overlapping route sections (see Figures 2-9, 15-18 and 25-29), and also by studying adjacent route sections when considering the impact of junction types.

4.1.6 One alternative to using the overlapping route sections would be to increase the sample size (across a longer time period or more routes) to allow the selection of a set of non-overlapping route sections whilst maintaining a reasonable sample. Alternatively, with the current dataset sub-populations of non-overlapping route sections could be constructed and analysed to investigate whether the overlap is biasing results unduly.

Link between Variability and Journey Length

4.1.7 The higher correlations shown do not necessarily imply a direct causation between journey time variability and journey length. However, one possible hypothesis for such causation is that services are able to ‘catch-up’ delay experienced over longer journeys. Services may be <u>proportionally</u> less likely to be variable when measured over a longer route section, because delays experienced earlier on in the journey are regained. This explanation is based on the fact that the operational instructions given to drivers mean that they are required to maintain scheduled times (as much as possible) at timing points, and when ahead of time will wait or extend dwell times, or conversely that they have some flexibility to ‘speed up’, rather than simply maintaining lateness, by shortening dwell times when late, or driving faster than the scheduled speed where road conditions permit.

4.1.8 It is also very likely that timetable design plays a major factor in this 'catch-up' effect, with timings adjusted to account for areas where delays often occur, and allow some catch-up. The extent of variation in the timetable was one of the key reasons for the selection of the route and operator, as it was evident that variation in journey times is currently experienced and it was known that the operator has a review process to adjust schedules accordingly. However, the brief for this work is to examine variability (i.e. the distribution of actual times), and not to consider schedule adherence (i.e. how late or early any individual bus is operating); therefore this hypothesis has not been explored against the data gathered.

4.1.9 A further potential factor is that buses which are full may not observe all scheduled stops: buses full to capacity with no passengers wishing to alight will not be able to stop and pick up further passengers, leaving the intending passengers for a following bus. As a side effect, this would allow the loaded bus to catch-up time. Obviously the cause and effect becomes complex, as it not known whether a full bus is more likely to have been delayed earlier in its trip (either from an unexpectedly large load of passengers or from a disruption to service) and whether a late-running bus is more likely to 'skip' stops.

Transferability of Results

4.1.10 Though it seems intuitively likely that similar relationships would exist in other areas, the current multi-regression model cannot be assumed to be immediately transferable. Further work would first be required to ensure that the relationships revealed would have similar values elsewhere. The inter-urban service X39 is probably not the most common type of bus route, as it serves two cities, rather than a city and a more rural hinterland, and is more frequent than most inter-urban services. Some of the relationships found here (e.g. the 'catching up' along a route) may also not hold for a shorter, more urban route (e.g. suburbs to town centre).

4.1.11 The finding for this dataset, that as average journey time becomes longer, the CV values tend to reduce is contrary to the anticipated outcome and can only be stated at this stage to be true for the journeys covered by this data collection exercise. It does however indicate that bus journey time variability in urban areas is inherently complex and therefore supporting data for investment proposals by local authorities needs to be based on the specific operating characteristics of the bus route and location(s) in question.

4.1.12 Regarding the impact of junctions, it does seem apparent that route sections with a small number of junctions achieve higher average speeds and lower variation in journey time, but the number of junctions is not a sufficient condition in itself. On the evidence provided thus far, there appears to be no difference in impact according to whether there is bus priority at junctions. However, with such a small sample it is likely that other factors are involved which may explain this, including the intended consequence that the bus priority measure has rendered the junction similarly variable to other junctions and without it, the variability would be worse.

4.1.13 From the data available for this study, no significant conclusions can be drawn about the relationship between journey time variability of general traffic and that of buses. Some of the data does however confirm intuitive and conventional logic about vehicle speeds and points to ways in which bus journey times are inherently different to those of general traffic.

4.1.14 There is some indication of a weak correlation between variation in bus journey time and general traffic, which may merit further investigation. However, there was not sufficient information or a strong enough relationship to justify an attempt at inclusion in the regression model at this point.

Further Analysis Work

4.1.15 From our review of the dataset, it appears possible that more insights and improved relationships could be achieved by continuing analysis. Such analysis of the existing data could include, amongst other options:

- examination of the opposite direction of travel (Bath to Bristol);
- consideration of whether any of the principal variables are affected by whether the bus began its journey on time or not, or is operating within a small window to another vehicle on the same route;
- more detailed consideration of the nature of the route sections and design of the timetable (e.g timetable journeys times vs distance and stops covered).

4.2 Suggestions for Data Gathering

4.2.1 Existing automated data collection systems (e.g. Automatic Vehicle Location and Electronic Ticket Machines) do not provide details of the causes of bus journey time variability, far less identify and quantify the benefit of solutions, and therefore it would be necessary to carry out a large amount of manual data collection in order to investigate the subject. Indeed, because of the multiple factors and the level and consistency of data collection required to carry out appropriate levels of modelling, it is apparent that trying to gather all data simultaneously can to some extent erode the quantum and quality of the data. Therefore in order for DfT to understand more about bus service journey time variability, it is suggested that future data collection exercises concentrate on discreet factors of bus journey time variability. While it would be important to have a consistent time series for the different factors, this would almost certainly require the allocation of multiple enumerators in order to gather the data which would be inherently impractical on buses, particularly at peak times. This may explain to some extent why existing literature on the topic can be considered to be incomplete.

4.2.2 The pilot study has however confirmed that it is possible in principle to gather relevant data (subject to real-life disruptions of bus operations) and to use that in a simple model. In order to increase further understanding of bus journey time variability, the same data collection exercise, using the learning points from the Pilot Study, could be conducted for the morning peak period, or for the afternoon peak period during school holidays, as these were both identified by the operator and the local authorities as having substantially different patterns of general traffic (and bus patronage). The results of this pilot study also identify the challenges faced by the bus operator and the local authorities in fairly typical urban areas and further work could usefully be done in considering the practicality of punctuality standards set by the Senior Traffic Commissioner.

4.2.3 The transcribed enumerator data and a presentation showing the summary headlines of variability have been provided to both the operator and the local authorities for their own reference and analysis. From these files, it will be possible for the local stakeholders to gain an understanding of the locations and extent of delays encountered during the 3 week data collection period. However, the data will not of itself be sufficient for the complete evaluation or design of any physical highway interventions and therefore such data gathering is not a substitution for a local authority's own monitoring or data collection to support their investment plans.

4.2.4 While capital investment in bus infrastructure may reduce the variability of bus journey time, it is not proposed that local authorities (and/or bus operators where the bid is in partnership) are required to gather bus journey time variability data in the way conducted for this study. The reason for this is that the identification of a principal root cause has been demonstrated to be difficult due to the combination of factors affecting bus journey times. Furthermore, the practicalities of gathering data in the on-bus environment mean that the data collection is itself inexact, with a reasonable amount of data requiring relatively significant staff resources.

4.2.5 In the event that journey distance and number of passengers are two of the primary universal causes of bus journey time variability, it is unlikely that local authorities will be in a position to be able to influence these in a positive manner in respect of variability. For example, it would not typically be in the wider interests of public transport to inconvenience passengers by breaking routes between key conurbations into two services or taking steps deliberately to reduce the number of passengers. Bus operators could theoretically improve boarding times of passengers in the longer term by revising the design of vehicles and/or altering their ticketing policies but, in a de-regulated bus industry environment, local authorities would only be likely to implement advanced payment schemes with the support of an operator.

4.3 Suggestions for appraisal tools/guidance

4.3.1 Prior to commencing this study, it was understood that the Department's intention was to consider whether the study outputs could be used as the basis for improving the Department's transport appraisal guidance and toolkit (WebTAG).

4.3.2 The actual development of the project and the scope of the Pilot Study in particular limits the scope for definitive proposals for the updating of WebTAG (and/or the development of other tools) as the guidance has to be transferable to a wide range of different locations and applicable to numerous applications. It is clear from this project that the issues and factors affecting bus journey time variability are not just complex in themselves; obtaining usable data is also inherently complex (which may explain the current limited body of evidence on the subject). This latter point therefore raises the issue of whether the cost of bus journey time data collection and analysis would be proportionate to the costs associated with any proposed interventions.

4.3.3 This study's assessment indicates that only 36% of the peak variation in bus-time, and 59% of the off-peak variation, can be explained by direct measurement. Although this might be improved upon through use of additional data gathering techniques and/or remote sources, it is not possible to comment on just how much variation might then be explained. Furthermore, as route length appears to be a major determinant, the actual benefit of other measures may remain hidden within these statistics. It would therefore be impractical and unhelpful to produce guidance based on this estimation method at this stage.

4.3.4 Nevertheless, in terms of the elements identified in Table 1 as having high importance on bus journey time variability, the following points are noted:

- General traffic – these impacts could be reviewed in light of speed data available from sources such as Trafficmaster or standard automatic traffic data collection methods (see 'Further Work' below);
- Junctions (including any bus priority) – the deployment of bus priority measures which vary according to the very local considerations and constraints of each individual junction (and the permutations which this gives rise to in terms of bus priority and pedestrian crossings in particular) is likely to require visual inspection of the route under consideration. In order to assess fully the impact which previously-installed bus priority measures have had, there would ideally be some 'before and after' data but, depending upon the date of implementation, such data may no longer relevant;
- Route length/number of bus stops – as noted above, the length of the bus route and the number of scheduled bus stops do not vary dynamically in the same way as the performance of general traffic and junctions and therefore these factual inputs would be 'fixed' elements of any appraisal as there are likely to be considerable negative passenger impacts of reducing route lengths or the number of bus stops.

4.3.5 Appraising the impact of elements assessed to have medium or low impact on bus journey time variability will be likely to depend on subjective, anecdotal or limited data (either from the bus operator(s) or local authority), as elements such as the ticket type (and thus the payment of cash fares) will be dependent upon operator-specific factors. As such, any available evidence of this type would be unlikely to be sufficiently reliable for an appraisal tool, although it may help to guide the development of interventions e.g. by identifying the location of sections of route between junctions which suffer from operational disruption.

4.3.6 Incidentally, in respect of the documents which comprise WebTAG, current TAG unit M1.2 (Data Sources and Surveys) already advises, in respect of establishing passenger demand for (bus-based) public transport that "Practitioners should contact bus operators in their area of interest to enquire about the availability of ETM data"[9]. While ETM (Electronic Ticket Machine) data is likely to remain the most comprehensive source, it should also be noted that the emergence of Smartcard ticket schemes typically offers much more detailed datasets, although these are not always as readily interrogable as ETM data.

4.3.7 Scheduled data relating to journey length (either by time or by distance) and the number of bus stops can be obtained fairly readily from either the operator and/or from the National Public Transport Access Node (NaPTAN) database. Timetable and route details are also available at the national level, although the January 2014 version of TAG unit M1.2 could be updated to reflect the development of the Traveline National Data Set (http://84.45.123.82/tnds-login-or-register.html) which has now superseded the National Public Transport Data Repository (NPTDR) database.

Further Work

4.3.8 Recognising that the current evidence base is insufficient to provide detailed guidance, but that some elements may be of interest to the DfT in progressing an appraisal tool and/or guidance, we feel that the most immediate area of investigation is that of identifying locations where high bus journey time CV values are likely to occur. Our analysis demonstrated there may be a role for TrafficMaster data as the primary source of evidence. We have demonstrated that there was a reasonable correlation between the CV of general traffic as estimated from TrafficMaster and the CV of bus journey times as observed in the surveys. Our analysis was limited to a small sample of the TrafficMaster data, we suggest that further analysis is undertaken to develop a tool that would enable routes to be isolated and CV values reliably calculated. This tool would provide insights into the locations where journey times are observed to be most variable, which would enable focussed examination of the route characteristics, such as junctions and layout.

4.3.9 This analysis tool would have to be complemented with bus journey time data. Our approach to collecting this data for the pilot study involved timing journeys by direct observation such that enumerators could simultaneously record other factors. Remote observation, such as from Real Time Information (RTI) systems, would be sufficient as a means to verify that a specific route or route section could be classified as more variable than normal.

4.3.10 It is accepted that the above commentary simply improves the method with which candidate routes and locations are identified, as opposed to providing a mechanism to demonstrate how a scheme would contribute to reducing variability. However, with the advantage of a more systematic approach to identifying locations, this improves the sample set and improves the ability to explore potential causes. This in effect is the approach adopted for accident remedial analysis whereby the first step is to identify a cluster of incidents and then classify the causes.

4.4 Summary

4.4.1 The project proposal identified that "in contrast to highway schemes, public transport modes are based around the existence of a timetable, with only discrete possibilities for departure. Due to this, the convention is for the focus to be on delay, rather than journey time variations in appraising reliability benefit". Reflecting the need for further study of bus journey time variability, current WebTAG guidance on User and Provider Impacts[10] deals solely with rail, where the industry performance management regime is constructed to measure, and attribute responsibility for, delays in an entirely different, and more comprehensive, manner than that which applies to deregulated bus services (outside London).

4.4.2 The methodologies for the Technical Report and the Pilot Study have proven to be sufficiently robust to achieve the requirements of the "proof of concept" – namely that data which indicates causes of bus

[9] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/370617/webtag-tag-unit-m1-2-data-sources-and-surveys.pdf , section 2.4.6
[10] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/313801/webtag-a13-user-provider-impacts-may2014.pdf, section 6.5

journey time variability can be gathered - but that the factors affecting overall variability are too numerous and/or complex such that it is impractical at best, and impossible at worst, to gather all such data in one data collection exercise. As identified in paragraph 3.2.9, "The overall R-Squared value of 0.36 indicates that the model fits the data to some extent (the inclusion of the number of stops, CV of the number of passengers and the average number of passengers per stop explain 36% of the variety of bus journey time)."

4.4.3 This outcome therefore endorses the finding of the Technical Report that "the body of literature we have reviewed addresses a number of issues that contribute to bus journey time variability but none of the papers present a systematic approach to isolating all of the contributory factors which have been identified. Most of the papers focus on particular issues and generalise other variables, such as the volume of traffic for example, which is generalised to "peak" or "off peak.""

4.4.4 While further understanding of the individual factors affecting bus journey time variability can help to develop guidance to scheme promoters, it is not proposed that DfT try to estimate the impact on reliability that making physical changes to the highway network may have. Collecting the data is relatively difficult and expensive – and most of the factors influencing reliability are outside the realms of physical infrastructure (e.g. passenger numbers and journey length). Furthermore, the cost of implementing any individual physical measure on the network is generally relatively small and so the return on that investment is not justified when set against the cost of even a modest manual data collection exercise.

Appendix A Pilot Study Data Collection

A.1. The project team evaluated potential locations and bus services for the pilot study by analysing Local Sustainable Transport Fund projects, supplemented by their existing knowledge of the UK bus market. After the preferred location was identified by the project team and endorsed by the DfT, a stakeholder meeting was convened to introduce the project to the bus operator (First), the local authorities (Bath & North East Somerset Council and Bristol City Council) and the sub-regional Local Sustainable Transport Fund project office (based in Bristol City Council).

A.2. This meeting helped to define the parameters for the pilot study by updating existing local knowledge and informed the project team of factors which could affect the results of the study, such as ongoing work to refine the phasing and timing of traffic lights on the corridor in order to enhance existing bus priority.

A.3. While it was the ambition of the project team and all stakeholders to conduct the pilot study during a period of ‘normal’ road conditions, the practicalities of conducting research on the corridor between Bristol and Bath meant that major road works in one road on the line of route in central Bristol, which were scheduled to be complete, but had overrun, could not be avoided during the project period. While these roadworks did not result in the bus service deviating from its usual route, it did reduce the available carriageway from two lanes to one lane in both directions. General traffic was not diverted during this period and therefore the actual bus journey time along Victoria Street (between Temple Meads railway station and the Centre) may be longer and more variable than would normally be the case, however this situation prevailed throughout the pilot study period and is therefore considered to be ‘constant’ for the purposes of this study. Such disruption was however anticipated to provide valuable insight to the degree of variability that can be experienced from such roadworks.

A.4. In order to start as soon as possible, consideration was given to splitting the data collection either side of school half-term, but for practical considerations such as avoiding straddling the end of British Summer Time (on October 27th) and arranging the availability of enumerators, the 3 week data collection period was commenced after half-term, in the first full week of November 2013.

Preparations for the pilot study

A.5. The enumerators were drawn from existing PB staff, which provided flexibility in terms of planning and availability, and more importantly, engagement and experience in conducting data collection for transport projects. While it was originally anticipated that the staff would be the same 5 individuals throughout the 3 week data collection period, to maximise consistency of data collection and take advantage of experience gained during the first week of the task, 8 enumerators were actually used, with all of them attending the training session which was arranged and conducted by PB/WSP for the enumerators and their supervisor.

A.6. The purpose of the training session was to introduce the rationale and remit of the project overall, explain the pilot study concept, define the data collection requirements, conduct some test (on-street) observations and communicate the practical arrangements to apply during the data collection period. The training session was designed to be engaging with the enumerators as the practice of gathering data on-bus, while a journey takes place as ‘normal’, requires slightly different skills than roadside observations for traffic counts. This training was supplemented with a single page of guidance notes for the enumerator’s reference while travelling on buses.

A.7. The enumerators were supported by their supervisor and by the PB/WSP project managers during the data collection, although this level of support is acknowledged to be greater than would be required if an operator and/or local authority were to conduct a similar exercise.

A.8. The period of journey monitoring spanned departures from Bath between 13:24 and 17:00 and, as a result, return departures from Bristol between 14:26 and 18:15. There is therefore not absolute consistency in the coverage of peak and off-peak journeys from each terminus, however the overriding purpose of the pilot study was to demonstrate the concept of data collection, rather than to draw absolute conclusions about how buses operating on the route at different locations at the same time may exhibit either similar or different traits.

A.9. As anticipated by the project team, the following data collection issues arose during the data collection period:

- Periodic gaps in the recording of data by the GPS devices – while the devices were all checked for functionality and accuracy prior to the data collection period, a few occasions arose of devices not recording and these were dealt with promptly by the supervisor as he was provided with a number of spare devices. These incidents were a combination of equipment failure (either total or partial) or unintentional changes in equipment function (e.g. buttons being caught).
- GPS device recording during times other than bus journeys – enumerators were instructed to turn the device on in order to be ready to record data from the start of each bus journey. In reality, this involved small periods of the devices recording the movement of the enumerator in Bristol and Bath (either before starting work, during a break between journeys or after finishing their shift) and the data cleansing process required these records to be disregarded by manual intervention, with a degree of subjective judgement.
- Difficulties in maintaining clear sight of passengers boarding and alighting – the main requirement for clear sight of the area around the driver's cab was to observe the ticket machine (to determine cash and non-cash fares). In reality, travelling on buses at peak times, or where there are pushchairs/wheelchairs gathered at the front of the bus, causes enumerators to take a less favourable position on the bus, in order that they themselves do not then become a cause of extended journey times by blocking the aisle or infringing passenger conduct regulations.
- Difficulties seeing the road outside the bus – as the data collection period fell during late autumn after British Summer Time had ended, and was conducted during the afternoon, it was not always possible for enumerators to be clear about the causes of delays, particularly if it was also raining, on unlit sections of road, or with misted-up windows on the bus. This also made it harder for some enumerators to track the sequence of stops (amplified further below). In the event that the enumerator was unable to keep track of the sequence of stops, they were instructed to pick-up recording again at the next definite bus stop which could be identified.

A.10. Disruption to planned bus journeys was also experienced by exceptional traffic conditions and during the data collection period, the following events occurred:

- a major fireworks display took place at the rugby ground in Keynsham (Tuesday 5th November). While this was not on the line of route, the volume of traffic heading for the display, which started at 6pm, and parking on surrounding streets was sufficient to cause tailbacks on the Keynsham bypass;
- a road traffic accident (at Newton St Loe) and a broken down car on a single carriageway section of route (Saltford) caused extensive delays on the same afternoon (Monday 11th November) of a combined total of nearly one hour; and
- temporary water main replacement works in Saltford (Thursday 21st and Friday 22nd November) caused service delays of 30-45 minutes and the resultant cancellation of some journeys during and after the evening peak as drivers reached the end of their available driving hours.

A.11. The analysis has included the data from the bus journeys affected by these events in order to reflect the impacts on bus journey times which, while sometime unpredictable, are part and parcel of the challenges of operating buses in urban areas.

A.12. Disruption to enumerator shifts was a consequential impact of journey time variability – while the enumerator shifts were planned according to broadly the same duty schedule as the buses and drivers, it quickly became apparent that the last of the enumerator shifts would routinely exceed the scheduled time (because of the issues highlighted in the bullet points above) and the shift was therefore changed after the first week of the period. This change resulted in 2 journeys in each direction being replaced by earlier journeys for the last 2 weeks of the monitoring period. Flexibility of shift finishing times was nevertheless required throughout and was successfully managed by the supervisor.

A.13. Aspects which future data collection training sessions could provide more of a focus on include:

- On-board practice with the GPS recording equipment and the recording forms and an explanation of the data to be collected and how this was relevant to the study.
- Improved understanding of the line of route and sequence of stops – this was addressed during the early days of the pilot study by the supervisor providing printouts of maps and additional identification of landmarks and side road names, rather than bus stop names, on the line of route.
- The greater need to record passenger numbers and transactions than the times of buses – the GPS device automatically recorded the time, although enumerators were also provided with the space to record arrival and departure times at bus stops on the data collection sheet, in order to focus their attention on the matter of the time spent at bus stops, and the reasons for this. As a result, recording of times was much more complete than recording of passenger numbers, and cash/non-cash fares.

A.14. Enumerators were provided with a GPS device and a clipboard with data collection sheets which listed all of the bus stops on the line of route, with the bus operator-specified registered timing points identified separately[11]. The data collection form template was revised at the end of the first week as a result of feedback from the enumerators and concerned the spacing and marking of key data entry boxes and minor simplification of data entry requirements, without altering the primary form and function of the form.

A.15. The stakeholder meeting had requested that 14 particular traffic management features (certain traffic light controlled junctions and pedestrian crossings), which were anecdotally considered to contribute to bus journey time variability, should also be documented and these were also highlighted on the data collection sheets.

A.16. These arrangements made the data collection a largely manual process, and required the completed sheets to be transcribed into spreadsheet format after the event. Clearly, for a longer-term programme of data collection, recording all details on a hand-held device such as a ‘tablet’, would avoid much of the downstream processing of data which involved expected issues of substantial administrative staff time and deciphering of hand-writing and risks such as transposition errors. The cost of providing each enumerator with a ‘tablet’ (if they did not already have a suitable device) would have to be weighed carefully against the cost of the transcription.

A.17. In order to provide sufficient space for enumerators to record the timings and events encountered by each bus on each journey, the data collection sheets included up to 97 lines for each direction of

[11] Enumerators were also provided with a letter of authorisation from the operator in order to show to the driver and/or any passengers who queried their presence on the bus. In practice, the arrangements to engage the operator in briefing its staff, and the training given to enumerators about how to conduct the data collection on-bus, did not result in the letter needing to be shown actively.

travel. When printed, each journey had 10 sheets in order to be legible for both the enumerator and the data analysts. The size and spacing of the forms was adjusted following initial feedback from the enumerators in the first week. By allocating 5 enumerators to the bus route for up to 4.5 hours on each weekday afternoon for the 3 weeks of the pilot study, it was planned to monitor 10 journeys in each direction, representing 31.25% of PM departures on the service, and 17.54% of the total number of journeys on the service as a whole[12].

Data Cleansing and Preparation

A.18. The first stage of the data analysis was to download the data from the GPS devices and import it into a GIS programme to validate the accuracy and completeness of the recordings against the scheduled route. This process involved:

- Resolving differences between the geo-referencing of the GPS device and the analysis package. Data was converted from degrees, minutes and seconds to a decimal format as the analysis package was unable to plot GPS data in its original format.
- Identifying individual bus journeys from the data provided by each GPS data logger.
- Specification of a 'geo-zone' radius around each bus stop (by plotting the GPS data onto a map using ArcGIS). From testing different radii, a 'geo-zone' of 7.5m around each bus stop was decided upon. This was to account for tolerances in the accuracy of the GPS data loggers attributable to the physical environment of the bus in which the GPS data was being recorded, the 3 second interval between each GPS datum being recorded by the logger and the possibility that the bus did not stop, or was forced to stop slightly before or after each bus stop due to road conditions (e.g. parked vehicles).
- Identifying the time each bus enters and exits the 'geo-zone' for each individual bus run using the Feature Manipulation Engine (FME) platform from Safe Software. This enabled the process of identifying the arrival and departure time from each 'geo-zone' to be carried out automatically for each individual journey between Bath and Bristol and between Bristol and Bath, exporting the data as a single .csv file. This permitted the size of the 'geo-zone' around each bus stop to be easily changed if there was a case for doing so for individual bus stops (e.g. bus stations).
- Coding exported data with text strings to improve the identification of data and reviewing it using a pivot table in Microsoft Excel.

A.19. The second stage of the data analysis was to conduct the transcription from the substantial number of individual pieces of paper into spreadsheet format. This process included a small amount of data cleansing, as it was completed by administrative staff rather than staff involved in the analysis of the data (who would have greater understanding of the techniques and requirements for transport modelling), and put each enumerator's daily records into one spreadsheet, with a separate tab for each direction of travel. This work included those fields which had been specified for enumerator observation:

- Whether the bus stopped at the bus stop or not;
- The arrival time at the bus stop;
- Number of alighting passengers;
- Of the alighting passengers, how many had physical difficulty;
- Duration of the bus being stopped at the bus stop;
- Of the boarding passengers, how many paid a cash fare;
- Of the boarding passengers, how many had a non-cash fare;
- Of the boarding passengers, how many had physical difficulty; and

[12] The amount of paper generated meant that it was unrealistic to scan or photocopy the forms in order to transfer them from one PB/WSP office to another. This posed more of a logistical issue than anything else, and would not be expected to occur if further data collection was done by a bus operator and/or local authority.

- The departure time from the bus stop.

A.20. In preparation for the modelling work, these records were then composited into one spreadsheet for all journeys, with a separate tab for each direction of travel on each day of the pilot study period and each bus stop and journey was given a unique identifier in order to facilitate the further analysis. This spreadsheet was devised to construct a matrix which calculated the time (in minutes and seconds) between individual bus stops, based on the time recorded by the enumerators. The data fields incorporated were:

- Direction of travel;
- Bus stop name;
- Stop reference number;
- Bus stop ID (SMS code);
- Scheduled departure time;
- If the bus stopped at the stop;
- Arrival time;
- Departure time;
- Date; and
- Day of the week;

A.21. It was at this stage that a more substantial amount of data cleansing was carried out in order to ensure strict consistency of data for the modelling phase, including formatting, completeness and normalisation. At the end of the data cleansing and preparation stage it was decided to use the manually recorded data instead of the GPS data for the detailed analysis, as a number of issues and limitations had been revealed in the processing including:

- Accuracy of GPS data provided by the data loggers – some GPS data loggers produced data points a considerable distance off the route of the bus. This meant that even with a 7.5m 'geo-zone', arrivals and departures at some bus stops were not being recorded or was inaccurate compared to on-board manual counts.
- GPS data points recorded outside bus stop 'geo-zones' – at bus stops nearer to the centre of Bath and Bristol it appeared likely from the cluster of data points that some buses were stopping before or after the bus stop outside of the bus stop 'geo-zone' (mostly likely due to multiple bus services serving the same single stop). There is the possibility that the data could be realigned to the bus route manually, however this was judged to be an extremely time consuming process.
- Difficult in identifying arrival and departure times at Bristol and Bath bus station from GPS data alone – enumerators kept GPS loggers on for entire day. This led to GPS data being recorded between bus journeys. This prevented a departure and arrival time being obtained from each bus station. A substantially larger 'geo-zone' around the bus station may have reduced this problem, however this was judged having a detrimental effect on the accuracy of the data.

A.22. In overall terms, although the GPS data provided journey times it did not accurately provide dwell time (any more than the manual data did) or easily provide comparable reasons for delay at each stop, such as passenger boarding/alighting movements, and so would have required cross-checking with the manual counts anyway, resulting in a more lengthy data analysis period.

A.23. Following completion of the bus-only data analysis, it was necessary to carry out further extensive data manipulation in order to facilitate the comparison between bus journey time variability and general traffic variability, which was based upon the contemporaneous Trafficmaster data made available via DfT. Given the vastly different purposes and sources of the two datasets, it was not unexpected that the resultant amount of directly comparable data from the 3-week period was relatively small (127).

A.24. Initially, a shape file in geospatial vector data format was drawn at a high level and as such, in places, did not fully trace the path of the Ordinance Survey Mastermap ITN (Integrated Transport Network) Layer for the roads in the travel direction of Bristol to Bath that the Trafficmaster GPS Data has been projected onto. Instances where this occurred tended to be along dual carriageway sections where the carriageway of the opposite direction was identified within the shape file. Where this occurred average journey times for the opposite direction have been used.

A.25. For security purposes, the DfT provided the Trafficmaster GPS data with an anonymised “Link ref” which corresponded to a “Link id”. A shape file was provided to allow for the “Link id” of relevant ITN links along the Bristol to Bath bus route to be determined and a “Link id” to “Link ref” lookup spreadsheet was also provided to match the Trafficmaster data with the relevant ITN Link. Trafficmaster GPS data for each ITN link was provided, where available, as 15 minute averages, by date and vehicle class.

A.26. To identify the “Link ref” of relevant ITN links the shape file was loaded into GIS software. From within this software ITN links not along the route of the X39 Bristol to Bath were removed. Relevant INT links were also coded with the preceding bus stop. One of the limitations of the ITN dataset is that it graphically represents single carriageway roads as a single link. For this reason Trafficmaster GPS data had been coded with a letter indicating the direction of travel:

- ‘A’ indicating the direction of traffic flow as digitised in the original ITN and
- ‘B’ indicating the flow direction is opposite to the original digitisation.

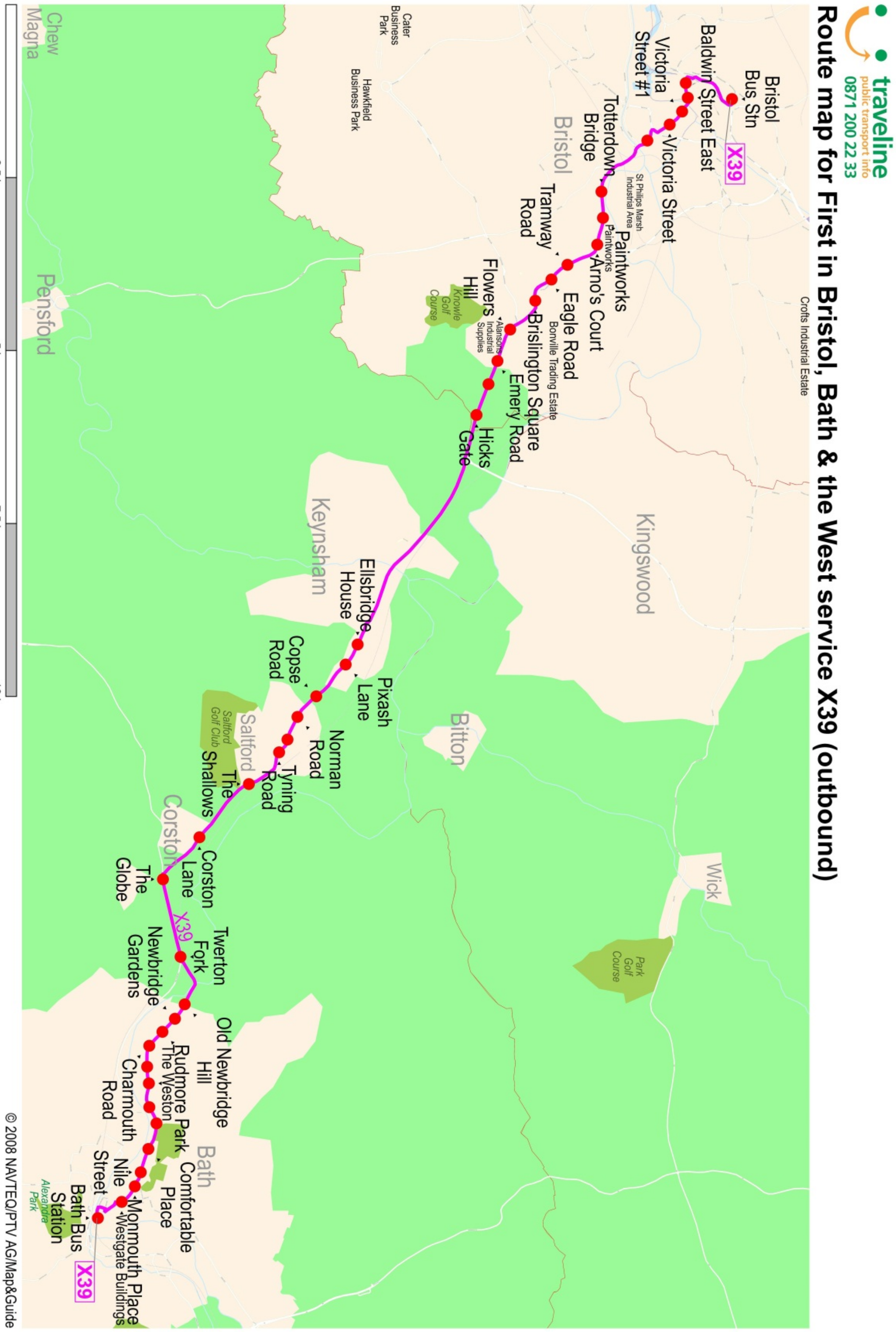
A.27. On dual carriageways, each direction of travel was specified by two separate links each coded with the letter ‘A’. The direction of the original digitisation was not specified in the data provided by DfT and nor could it be confirmed through any further communication with DfT. Therefore the directional coding of links was determined based on a manual review of the links provided within the GIS software. From this, the “Link Ref” for each ITN Link between each bus stop along the route could be obtained. This was then looked-up against the “link id’s” provided separately by the DfT.

A.28. The GIS analysis enabled journey times for each ITN link between bus stops along the route to be established. From this dataset additional matrices were produced summarising the total journey time between bus stops (as a sum of individual ITN links) by vehicle class, time and date. It is important to note that the way the Trafficmaster GPS data was summarised ensured that if data was not available for all of the ITN links between bus stops for a given 15 minute time period, no total journey time would be returned.

A.29. Clearly, for a valid comparison of the Trafficmaster GPS data with the bus survey data, the datasets are required to be in the same format. Enumerators recording the bus survey data recorded the cumulative journey time of each bus, noting down the arrival and departure time from each stop. The journey time between consecutive bus stops was thus calculated by the difference between the recorded departure and arrival times. In many instances enumerators only recorded a time when the bus stopped. This meant that no time was recorded for many stops. Therefore, it is only possible to make a comparison with the Trafficmaster GPS data where times were recorded for two consecutive stops, significantly reducing the amount of data available for comparison.

A.30. A total of 1147 manual bus survey journey times and 1079 Trafficmaster journey times have been identified from the data preparation process. From this, a comparison of the bus survey data has been made with the average ITN link journey time of cars (Class 1) and Light Goods Vehicles (class 2), for the same 15 minute segment. Trafficmaster GPS data was available in 127 instances where manual bus survey data had been recorded. It would be possible to increase the number of matches in the future by using an hourly average of Trafficmaster GPS data, however this is likely to be unrepresentative of observed bus journey time fluctuations, especially in the peak hours.

Appendix B Bus service route map



Appendix C Journey Time Data

Journey Number	From Stop	To Stop	Sample Size	Average Total Journey Time (Stop Departure to Stop Departure)	SD Journey Time	CV of Journey Time
2	200	201	121	00:06:16	00:02:39	0.422
3	200	202	102	00:14:41	00:05:02	0.343
5	200	204	125	00:19:29	00:05:34	0.286
6	200	205	55	00:22:54	00:06:21	0.277
8	200	207	114	00:24:19	00:06:04	0.249
9	200	208	78	00:28:07	00:06:39	0.237
10	200	209	122	00:27:37	00:06:30	0.235
11	200	210	119	00:29:01	00:06:36	0.227
12	200	211	96	00:29:42	00:06:04	0.205
13	200	212	98	00:32:06	00:06:35	0.205
16	200	215	93	00:39:18	00:08:15	0.210
19	200	218	87	00:41:38	00:07:47	0.187
20	200	219	104	00:43:23	00:09:41	0.223
26	200	225	92	00:49:34	00:07:53	0.159
27	200	226	79	00:52:06	00:10:29	0.201
28	200	227	75	00:51:51	00:10:14	0.197
29	200	228	101	00:52:30	00:09:34	0.182
30	200	229	54	00:53:25	00:08:35	0.161
31	200	230	70	00:54:28	00:08:02	0.147
35	200	234	76	00:59:03	00:08:36	0.146
36	200	235	110	00:58:58	00:10:42	0.182
37	200	100	143	01:02:27	00:10:43	0.172
39	201	202	91	00:08:22	00:03:07	0.372
41	201	204	108	00:13:48	00:04:02	0.293
42	201	205	50	00:16:54	00:04:28	0.264
44	201	207	97	00:18:36	00:04:28	0.240
45	201	208	74	00:21:36	00:04:47	0.222
46	201	209	107	00:21:52	00:04:46	0.218
47	201	210	103	00:23:29	00:04:59	0.212
48	201	211	85	00:24:18	00:04:43	0.194
49	201	212	90	00:26:30	00:04:59	0.188
52	201	215	77	00:33:59	00:06:57	0.204
55	201	218	71	00:36:15	00:06:32	0.180
56	201	219	89	00:38:04	00:08:49	0.232
62	201	225	80	00:43:58	00:06:42	0.153

Journey Number	From Stop	To Stop	Sample Size	Average Total Journey Time (Stop Departure to Stop Departure)	SD Journey Time	CV of Journey Time
63	201	226	71	00:46:11	00:09:28	0.205
64	201	227	66	00:46:28	00:09:12	0.198
65	201	228	88	00:47:00	00:08:50	0.188
67	201	230	60	00:49:42	00:07:36	0.153
71	201	234	65	00:53:33	00:07:36	0.142
72	201	235	95	00:53:24	00:10:00	0.187
73	201	100	121	00:57:26	00:10:00	0.174
76	202	204	90	00:05:51	00:01:39	0.283
79	202	207	80	00:10:27	00:02:07	0.202
80	202	208	64	00:13:02	00:02:35	0.199
81	202	209	89	00:13:50	00:02:50	0.205
82	202	210	88	00:15:20	00:02:55	0.190
83	202	211	72	00:16:47	00:03:11	0.190
84	202	212	75	00:18:41	00:03:03	0.163
87	202	215	70	00:26:03	00:05:24	0.207
90	202	218	70	00:28:18	00:04:49	0.170
91	202	219	76	00:30:03	00:07:20	0.244
97	202	225	70	00:36:12	00:05:06	0.141
98	202	226	58	00:37:01	00:04:22	0.118
99	202	227	58	00:38:24	00:07:46	0.202
100	202	228	73	00:39:26	00:07:43	0.196
106	202	234	60	00:45:38	00:06:22	0.140
107	202	235	77	00:45:34	00:07:25	0.163
108	202	100	104	00:48:59	00:07:52	0.161
144	204	205	52	00:02:42	00:00:36	0.221
146	204	207	103	00:04:26	00:01:03	0.236
147	204	208	73	00:07:02	00:01:47	0.253
148	204	209	109	00:07:55	00:02:14	0.282
149	204	210	106	00:09:27	00:02:24	0.254
150	204	211	87	00:10:49	00:02:32	0.234
151	204	212	88	00:12:37	00:02:40	0.212
154	204	215	85	00:19:49	00:04:34	0.231
157	204	218	78	00:21:55	00:03:32	0.161
158	204	219	93	00:24:11	00:07:06	0.294
164	204	225	82	00:30:16	00:04:49	0.159
165	204	226	72	00:31:53	00:07:58	0.250

Journey Number	From Stop	To Stop	Sample Size	Average Total Journey Time (Stop Departure to Stop Departure)	SD Journey Time	CV of Journey Time
166	204	227	66	00:32:03	00:07:03	0.220
167	204	228	88	00:33:16	00:06:53	0.207
169	204	230	57	00:36:07	00:06:23	0.177
173	204	234	63	00:39:28	00:05:34	0.141
174	204	235	98	00:39:51	00:08:28	0.212
175	204	100	125	00:43:25	00:08:30	0.196
207	205	100	56	00:41:24	00:08:31	0.206
240	207	208	63	00:02:34	00:01:12	0.472
241	207	209	99	00:03:32	00:01:41	0.476
242	207	210	97	00:05:03	00:01:49	0.360
243	207	211	78	00:06:29	00:02:07	0.328
244	207	212	77	00:08:11	00:02:18	0.280
247	207	215	77	00:15:25	00:04:25	0.286
250	207	218	69	00:17:25	00:03:12	0.184
251	207	219	84	00:20:00	00:07:02	0.352
257	207	225	72	00:25:46	00:04:44	0.184
258	207	226	65	00:27:29	00:07:53	0.287
259	207	227	59	00:27:42	00:07:01	0.253
260	207	228	80	00:28:51	00:06:58	0.241
262	207	230	56	00:31:11	00:06:15	0.200
266	207	234	59	00:34:42	00:05:35	0.161
267	207	235	86	00:35:22	00:08:43	0.246
268	207	100	114	00:38:51	00:08:32	0.220
270	208	209	68	00:01:17	00:00:48	0.632
271	208	210	69	00:02:49	00:00:58	0.343
272	208	211	52	00:04:21	00:01:17	0.296
273	208	212	56	00:05:55	00:01:31	0.255
276	208	215	52	00:13:56	00:04:47	0.343
279	208	218	52	00:15:18	00:02:52	0.188
280	208	219	54	00:17:41	00:05:11	0.293
286	208	225	53	00:24:01	00:04:47	0.199
287	208	226	52	00:25:08	00:07:56	0.316
289	208	228	58	00:26:23	00:04:50	0.183
296	208	235	60	00:33:04	00:07:52	0.238
297	208	100	79	00:36:42	00:08:02	0.219
299	209	210	104	00:01:34	00:00:24	0.253

Journey Number	From Stop	To Stop	Sample Size	Average Total Journey Time (Stop Departure to Stop Departure)	SD Journey Time	CV of Journey Time
300	209	211	85	00:03:10	00:00:42	0.219
301	209	212	85	00:04:38	00:01:02	0.223
304	209	215	80	00:12:04	00:03:58	0.329
307	209	218	70	00:14:38	00:03:29	0.238
308	209	219	88	00:16:38	00:06:12	0.372
314	209	225	75	00:22:39	00:04:18	0.190
315	209	226	71	00:23:34	00:06:44	0.286
316	209	227	66	00:24:33	00:06:44	0.275
317	209	228	85	00:25:55	00:06:40	0.257
318	209	229	50	00:26:33	00:04:41	0.177
319	209	230	58	00:28:24	00:06:09	0.217
323	209	234	67	00:31:40	00:05:15	0.166
324	209	235	89	00:32:09	00:07:44	0.240
325	209	100	122	00:35:39	00:07:44	0.217
327	210	211	85	00:01:37	00:00:32	0.330
328	210	212	85	00:03:06	00:00:53	0.286
331	210	215	83	00:10:34	00:03:53	0.367
334	210	218	72	00:13:00	00:03:28	0.266
335	210	219	87	00:15:13	00:06:04	0.399
341	210	225	81	00:20:53	00:04:16	0.204
342	210	226	70	00:22:15	00:06:37	0.297
343	210	227	65	00:23:20	00:06:37	0.284
344	210	228	85	00:24:27	00:06:37	0.271
346	210	230	62	00:26:46	00:06:01	0.225
350	210	234	65	00:30:06	00:05:16	0.175
351	210	235	92	00:30:44	00:07:33	0.246
352	210	100	119	00:34:26	00:07:37	0.221
354	211	212	70	00:01:39	00:00:37	0.378
357	211	215	69	00:09:14	00:04:10	0.453
360	211	218	66	00:11:07	00:02:05	0.187
361	211	219	76	00:13:18	00:05:56	0.446
367	211	225	69	00:19:32	00:04:21	0.222
368	211	226	57	00:20:05	00:02:52	0.143
369	211	227	52	00:21:21	00:06:25	0.301
370	211	228	74	00:22:43	00:06:20	0.279
376	211	234	54	00:28:27	00:04:48	0.169

Journey Number	From Stop	To Stop	Sample Size	Average Total Journey Time (Stop Departure to Stop Departure)	SD Journey Time	CV of Journey Time
377	211	235	78	00:29:01	00:06:10	0.212
378	211	100	98	00:32:48	00:06:27	0.196
382	212	215	69	00:07:23	00:03:22	0.455
385	212	218	65	00:09:54	00:03:15	0.329
386	212	219	74	00:12:02	00:05:54	0.490
392	212	225	68	00:17:49	00:03:17	0.185
393	212	226	55	00:19:14	00:06:43	0.349
394	212	227	56	00:20:17	00:06:33	0.323
395	212	228	72	00:20:49	00:06:01	0.289
397	212	230	50	00:23:28	00:05:13	0.222
401	212	234	54	00:27:05	00:04:36	0.170
402	212	235	77	00:27:43	00:07:31	0.271
403	212	100	98	00:31:26	00:07:27	0.237
454	215	218	66	00:02:51	00:00:58	0.340
455	215	219	73	00:04:16	00:01:14	0.291
461	215	225	71	00:10:46	00:02:01	0.188
462	215	226	57	00:11:31	00:01:48	0.157
463	215	227	53	00:12:23	00:02:05	0.168
464	215	228	68	00:13:23	00:02:19	0.173
466	215	230	50	00:15:59	00:02:50	0.177
470	215	234	58	00:19:21	00:02:36	0.134
471	215	235	75	00:20:07	00:02:52	0.142
472	215	100	93	00:23:17	00:03:25	0.147
515	218	219	65	00:01:32	00:00:38	0.416
521	218	225	65	00:08:01	00:01:38	0.203
522	218	226	52	00:08:41	00:01:33	0.179
524	218	228	69	00:10:27	00:01:48	0.172
531	218	235	68	00:17:04	00:02:30	0.146
532	218	100	87	00:20:22	00:03:11	0.156
539	219	225	72	00:06:25	00:00:54	0.140
540	219	226	59	00:07:10	00:00:55	0.128
541	219	227	57	00:07:55	00:01:06	0.140
542	219	228	73	00:08:52	00:01:09	0.129
548	219	234	56	00:14:55	00:01:57	0.130
549	219	235	81	00:15:36	00:02:03	0.132
550	219	100	104	00:18:37	00:02:55	0.157

Journey Number	From Stop	To Stop	Sample Size	Average Total Journey Time (Stop Departure to Stop Departure)	SD Journey Time	CV of Journey Time
627	225	226	55	00:00:53	00:00:27	0.498
628	225	227	53	00:01:40	00:00:32	0.320
629	225	228	71	00:02:31	00:00:35	0.231
631	225	230	53	00:04:51	00:01:15	0.259
635	225	234	56	00:08:35	00:01:27	0.168
636	225	235	78	00:09:26	00:01:40	0.176
637	225	100	92	00:12:49	00:02:35	0.202
640	226	228	62	00:01:54	00:00:27	0.234
647	226	235	65	00:08:46	00:01:31	0.174
648	226	100	79	00:11:58	00:02:26	0.204
650	227	228	61	00:01:08	00:00:19	0.277
657	227	235	59	00:08:12	00:01:23	0.168
658	227	100	75	00:11:09	00:02:12	0.198
661	228	230	53	00:02:12	00:00:55	0.418
665	228	234	58	00:05:51	00:01:09	0.198
666	228	235	82	00:06:44	00:01:14	0.182
667	228	100	101	00:09:57	00:02:13	0.222
675	229	100	54	00:08:38	00:01:50	0.211
681	230	235	55	00:04:44	00:00:47	0.165
682	230	100	70	00:08:02	00:02:11	0.272
699	234	235	53	00:01:20	00:00:28	0.352
700	234	100	76	00:04:24	00:01:57	0.445

WSP UK Limited
WSP House
London
WC2A 1AF
UK
Tel: +44 20 7314 4671
Fax: +44 20 7314 5111