New Generation Transport in Leeds

Modelling Service Reliability

November 2009

Prepared for:

Leeds City Council

and

Metro

Prepared by:

Steer Davies Gleave
28-32 Upper Ground
London
SE1 9PD

+44 (0)207 910 5000
www.steerdaviesgleave.com

1 The background

Introduction

1.1 Metro and Leeds City Council are developing a proposal for a major urban public transport scheme in Leeds. The proposed scheme, named New Generation Transport (NGT), is intended to provide high level of service along the main radial corridors leading to the Leeds city centre.

1.2 Steer Davies Gleave is assisting in this study by undertaking the demand forecasting work for the scheme and assessing its business case.

1.3 Traditionally, the impacts of public transport unreliability were not explicitly accounted for in public transport demand forecasts, and the benefits from improved reliability did not have a formal role in the appraisal of transport schemes.

1.4 However, the fact that travellers do respond to the level of reliability, and the existence of economic benefits or costs associated with this response, have recently been acknowledged in the *Eddington Report*; in *Towards a Sustainable Transport System*; and in early work on the *NATA Refresh*.

1.5 In addition, there is now already some draft WebTAG guidance on the treatment of reliability in modelling and forecasting, although at this stage this has not yet formalised into an integral part of the widely-implemented modelling and appraisal practice.

1.6 At an early stage of our work, it has been agreed with Metro and LCC that the improvement of public transport reliability should be one of the key goals for the NGT scheme, and that potential benefits from improving reliability should be revealed and assessed.

1.7 This has raised various methodological challenges, both theoretical and practical. A separate work stream under the broader NGT study is therefore dedicated to tackling these challenges and to putting together a comprehensive approach that meets all that is required to incorporate reliability considerations in the demand forecasting and appraisal work. This report summarises all the work done under this stream.

A Definition of Reliability

1.8 Reliability can be defined in various ways, and there are various discussions of the *predictability*, *punctuality* or *regularity* of travel conditions on public transport services. Although there are differences between the various terms, a fundamental trait of a reliable transport system that resides in all of them is a low level of day-to-day variability in travel times.

1.9 Note that we are *not* particularly interested in travel time variability between different vehicles making similar journeys at the same time (because of waiting times at signals, conflicts with pedestrians, differences in driving style and so on). We are also *not* focusing here on how travel times vary between different times of day. Rather, the main indicator of unreliability is the difference in travel times between similar journeys made at a similar time on different days.

1.10 There are many reasons for day-to-day fluctuations of public transport travel times. These include fluctuations in travel demand, variation in driving behaviour, changes in the amount of roadside activity, weather conditions, accidents and incidents, and other reasons. Some of these can be expected because of the difference between weekdays and weekends, or because of special travel patterns on holidays.

1.11 *We are only interested in the variability that remains unexplained after variations due to these predictable reasons have been subtracted; otherwise it is likely to be accounted for elsewhere in the model and the appraisal already*. The variability we are interested in is random in the sense that its various causes are too subtle or too complex to be expected by travellers.

1.12 The key measure often used to indicate levels of variation is the standard deviation. It is not uncommon to use the standard deviation of travel times across different days as a measure of unreliability. In this note we also use other measures, for reasons which are explained in more detailed later. We also show later how we undertake the conversion from one measure to another, which has been one of the main technical problems we had to solve.

How Reliability is Modelled

1.13 To account for the level of reliability in the forecasts which will later feed into the cost-benefit analysis, we should first undertake two separate tasks: one which involves demand modelling work and another that has to do with network (supply) modelling.

1.14 The demand side analysis is aimed to estimate how a *given change in the level of reliability* is valued by travellers, i.e. to give reliability a unit value. The unit values used for the appraisal of NGT are based on several sources, including a Stated Preference (SP) survey we held in Leeds and some other sources. We present the approach used for demand side in Chapter 2.

1.15 The change in the level of reliability for future scenarios is not given, and for this reason the network (supply) side of the modelling framework is needed. This component has the objective of estimating how unreliable each element of the public transport system will be in each of the scenarios we investigate, so that this can then be used as input for the demand model. We present the supply analysis in Chapter 3.

1.16 Combining the demand and supply analysis is a challenge in itself. It involves the various tasks that need to be undertaken so that the appraisal of the NGT scheme incorporates potential benefits from improved reliability. The way we prepared these inputs for the appraisal is described in Chapter 4.

2 Modelling reliability: the demand side

Introduction

2.1 We are interested in looking at the reliability of travel times from a demand perspective because this would enable us to ascribe a value for each minute of unreliability from a traveller's perspective. With this unit value we can then include the level of reliability in the appraisal of NGT just like we do with the average travel time.

2.2 In early 2008 we undertook a major Stated Preference (SP) survey in Leeds. The survey examined the attitudes to a large number of attributes of public transport services. The aim of the survey was to estimate a mode choice model with sufficient sensitivity to these various attributes, and to derive the willingness to pay for changes in the levels of these attributes. Reliability attributes were included in the survey, and the design of our approach for including reliability considerations in the broader evaluation methodology therefore started concurrently with the design of the survey.

2.3 It was later agreed that valuations from the survey were not the only demand inputs required for the reliability work, and therefore other sources of information were used, too. We describe the way we determined the unit values of reliability in this chapter.

The Stated Preference Study and the Scheduling Approach

2.4 The entire SP study, including the work described in this chapter, has already been described in more detail in a separate report, but we highlight here key issues related to the introduction of reliability attributes to the NGT study.

2.5 SP studies often help us learn about the attitudes of travellers to features of a transport service. An SP survey consists of a series of choice tasks, each tasks presents several alternative services to respondents. Each alternative has several attributes, hence by examining how respondents trade-off between the different attributes we can reach conclusions about how important they find each attribute.

2.6 When we focus on attributes such as travel time or the fare of a bus journey, the way of presenting their levels in the survey is relatively straightforward. But this is not the case when the attributes represent the level of reliability, because there are different ways of indicating how good the level of reliability is.

2.7 Analysts often describe different levels of reliability using statistical terms, such as the standard deviation of travel times. But many travellers are not sufficiently familiar with these terms, and we would not be able to have confidence in valuations of reliability derived from a survey which was not well-understood.

2.8 We carried out a comprehensive review of recent studies in this area, to see how they deal with this issue. A list of the studies we reviewed is brought towards the end of this report. Most studies found that the best way to capture the attitudes of travellers to the reliability of public transport services is via the effect of reliability on the extent of *early or late arrival*. Much evidence shows that travellers have strong feelings towards the chance of arriving early or late, which directly affects their travel behaviour. This is often referred to as *The Scheduling Approach* for the valuation of unreliability.

2.9 We therefore followed this approach in our survey and in the entire study. Hence our unit values are a value for each minute of lateness and another value for each minute of earliness. When the generalised cost of travel is calculated, one component is the amount of lateness per journey (in minutes) multiplied by the value of lateness (in pence per minute), and another component is the amount of earliness multiplied by the value of earliness. The sum of these two components is the cost of unreliability for the journey.

2.10 Note that this is an indirect approach in the sense that the attributes presented in the survey do not capture the level of reliability but the late or early arrival that result from this level. But most research in this area found that this approach is more powerful in explicitly revealing the aversion of travellers to unreliability, all the same.

2.11 The SP study included 12 different survey exercises. Each respondent participated in four of these, and in each exercise there was a series of choice situations, focusing on a small set of service attributes. Two the SP exercises focused on reliability that looked separately at unreliability of the time of departure and at unreliability of the time of arrival. Therefore, in principle the valuations derived from the survey could include two values for earliness (one at the origin stop and one at the destination) and similarly, two values for lateness.

2.12 One of the key conclusions from the reliability exercises was that the best way of capturing the effects of unreliability is with two variables only: one representing the mean extent of early arrival at the origin and the other - the mean level of late arrival at the destination. This is convenient as it penalises both earliness and lateness but with just two variables.

2.13 Monetary values for these attributes are presented in Table 2.1. To put the value of lateness and earliness in context, the table also indicates the estimated values for in-vehicle time, walk time and service headway. All values are in pence per minute.

2.14 Note that the presented values are for travellers that do not have a car available for their journey. For travellers with an available car, values were found to be on average 8% higher.

TABLE 2.1 THE WILLINGNESS TO PAY (TRAVELLERS WITH NO CAR AVAILABLE)

Attribute	Commuting trips	Trips on the employer's business	Other trip purposes
In-vehicle time	2.8	6.3	3.1
Walk time	3.1	3.9	3.6
Headway	4.0	4.9	3.1
Mean earliness at the origin	**6.4**	**21.0**	**7.2**
Mean lateness at the destination	**13.2**	**16.4**	**11.4**

2.15 The relativities between the values for different trip purposes are consistent with our expectations. Business travellers exhibit a much higher willingness to pay than travellers on any other purpose; commuters are willing to pay slightly more than travellers on other trip purposes (such as shopping or leisure) for most attributes.

2.16 For the purpose of the appraisal of NGT we need to use standard values of time rather than values from the survey. To retain the relative importance of reliability, we calculate for each journey purpose the ratio of the values of lateness and earliness to the value of average in-vehicle time. These are summarised in the following table.

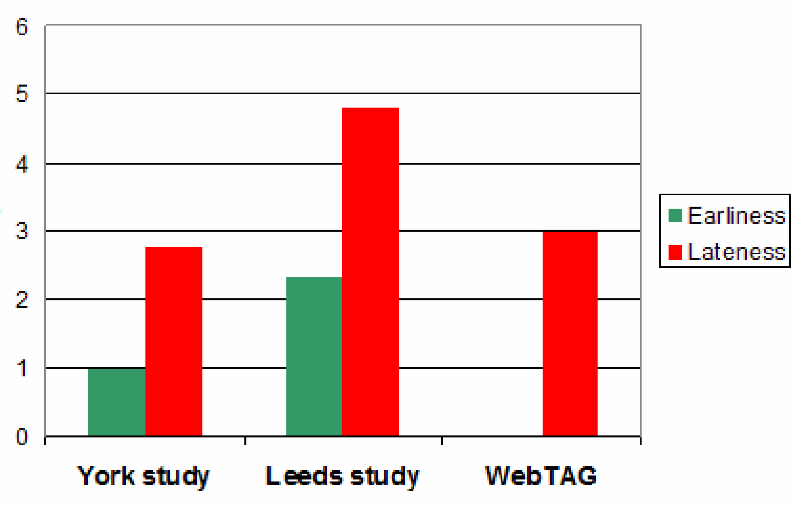
TABLE 2.2 VALUES OF LATENESS AND EARLINESS, RELATIVE TO THE VALUE OF AVERAGE TRAVEL TIME

Attribute	Commuting trips	Trips on the employer's business	Other trip purposes
Mean earliness at the origin	2.3	3.3	2.3
Mean lateness at the destination	4.7	2.6	3.7

A Discussion of the Value of Earliness

2.17 Figure 2.1 shows the values of earliness and lateness, measured relatively to the value of the average in-vehicle time, based on three sources. The middle bars are based on 'commuting' and 'other' trips in the current study. The leftmost bars are based on a study of bus users in York, documented in the article "Direct versus Indirect Models for the Effects of Unreliability" (see reference list later). The rightmost bars are the values from WebTAG.

FIGURE 2.1 VALUES OF EARLINESS AND LANTESS FROM THEE SOURCES (RELATIVE TO THE VALUE OF AVERAGE TRAVEL TIME)



2.18 There are differences in the relative value of lateness, but all the values show that a minute of lateness is valued as several minutes of average travel time. The differences seem to be logical variations between different cities, and all the values are within the range of acceptable values mentioned in WebTAG. Therefore, although the value of lateness calculated in our SP study is higher than in the two other sources, it still looks suitable for direct use in the appraisal of NGT.

2.19 The case with the value of earliness is more complex. The WebTAG approach to valuing earliness is that this can be completely discarded. The logic behind this is that people are more averse to late arrival at their destination than to early arrival, and hence it is believed that considering lateness alone will only lead to a minor error.

2.20 Unfortunately we find this reasoning flawed, since it is based on an assumption that travellers will simply accept the level of lateness approximated by the supply model. In reality, since travellers normally dislike late arrival, they bring their departure forward in order to avoid late arrival. By fully penalising lateness and not accounting for such response, we are likely to seriously overestimate the cost of unreliability.

2.21 In other words, we must also penalise earliness because otherwise the shift to an earlier departure is not capped by any counter-penalty. It is the balance between the values of earliness and lateness that determines how early travellers depart and what the real cost of unreliability is.

2.22 The question is which value we should use for a minute of earliness. As explained earlier, in our SP study we have covered all types of unreliability, either at the origin or at the destination. We found that a model penalising unreliability at the origin stops is more statistically significant. This penalty has to do with the chance that the public transport service is missed before the passenger arrives at the departure stop after it has already left. This was found a more dominant concern than the concern that the service departs on time and then arrives too early at the destination.

2.23 A concern that was raised in the context of the appraisal of NGT was that penalising earliness at the origin would overestimate the cost of unreliability, because in practice such earliness can be easily avoided, and therefore the cost never applies. The operators can avoid early departure from the departure stop by determining that drivers have to wait at the stop so that they do not leave before their scheduled time.

2.24 If we assume that such rule is operated, we only need the value of earliness to account for early arrival to the destination due to fluctuations in the journey time. This will be a more realistic value, and clearly lower than the value we derived from the SP study. But since we did not estimate such value in this study, we need to base it on other sources.

2.25 The York study mentioned earlier (and shown in Figure 2.1) has a relatively value of 1.0 for earliness. Namely, it finds that a minute of earliness has about the same cost as a minute of the travel time itself. Other studies we reviewed show values around 1.0 or slightly lower. To adopt a prudent approach we have decided to use a relative value of 0.75 for earliness in the appraisal of NGT.

Conclusion

2.26 In this chapter we explained that we estimate the cost of unreliability through its effect on lateness and earliness. We explained how we determined the unit values of lateness and earliness to be used in the appraisal of NGT. The values we use later in this report are 4.7 for lateness and 0.75 for earliness. Both values represent the number of average in-vehicle time minutes that are equivalent to one minute of lateness or earliness, respectively.

3 Modelling reliability: the network side

Introduction

3.1 In the previous chapter we presented main findings concerning the value of each unit of early or late arrival of public transport services. Since earliness and lateness are a direct result of the unreliability of travel times, these valuations can later be converted into a cost of unreliability.

3.2 However, these are only *costs per unit*. For a complete analysis of the level of reliability and its implications, we also need information on how reliable or unreliable transport services are. Namely, we need additional information on the *number of units* of the same reliability measure we have a cost per unit for.

3.3 This chapter describes the work done to analyse how reliable the transport system is, and how reliable it would be with the NGT service. The work described in this chapter concerns the performance of the transport network rather than the preferences of its users. It is therefore a supply issue rather than a demand issue.

3.4 The different sections of this chapter cover several streams of work which should help us have a clear picture of the levels of reliability in Leeds at present and in future scenarios. The first section describes the extensive data processing work we have undertaken. The second section presents the format for reliability reports which have been generated for existing bus routes in Leeds, as an intermediate product from the study.

3.5 Subsequently we describe how geographical information on streets and roads in Leeds was collated and used in an attempt to explain differences in the level of reliability. Then we explain how we defined several different scenarios of reliability, to be used in the appraisal of the NGT service. Towards the end of the section we describe the process of converting information on reliability from a stop-to-stop link format to an origin-destination (OD) format.

Data Handling

3.6 All the tasks described in this chapter are based on information of day-to-day variations in travel times of bus routes in Leeds in the recent past. This information was provided to us by Metro in the form of a large number (above 600) of files containing travel times automatically recorded at different days.

3.7 Each file is for a single route and for a period of time that varies in length between one week and one month. Overall, the information received covers a three-month period in 2006 (October till December) and the equivalent period in 2007; namely, six months of data in total.

3.8 We excluded information for weekends and for the Christmas period from the data, as our main concern is the analysis of reliability on standard weekdays.

3.9 Note that throughout this text, the term ‘route’ refers to a single sequence of stops with a preset order *in one direction*. Information on route 1 going northbound, for example, is stored in separate files from the information on route 1 going southbound. The analysis of these two directions is therefore done here independently and they count as two different routes. However, this is merely a matter of terminology.

3.10 The information we received on each route is split into several files. We assume that this was done in order to keep each individual file at a manageable size. However, this has considerably complicated our data handling work, since we wanted to pool together all the information on each route across all periods.

3.11 In addition, we needed the data to be restructured for various other reasons, which we describe below. We therefore embarked on an extensive process to convert the data in its original format into the desired format.

3.12 The restructuring process is programmed largely as a series of Visual Basic macros. The automation of the process is aimed to reduce the amount of work involved, enable consistency control and error checking, and to allow the process to be easily repeated if a need to introduce any changes arises.

3.13 Figure 3.1 illustrates the format of the original data files. Within each file, each block of rows describes one departure, and within this block, each row stands for one stop. The stops are sorted alphabetically, and there is no direct information on their geographical order along the route.

3.14 The leftmost six columns provide general information about the route and about individual departures. To the right of these first six columns, each set of four columns provides information on a specific day, which is then repeated for many different days. The four columns for each day include a flag for whether information for that day is available; the scheduled run time (in minutes); the deviation from the scheduled time (in minutes); and the dwell time. The scheduled run time is repeated for each day, although for most days the same information is provided in this column.

3.15 Note that the information about scheduled time can be generally used to infer the order in which the bus arrives at the stops where the scheduled times are different from each other. In each route there are some stops that have the same scheduled time; for these stops, no information is provided about their geographical order. As discussed below, this created further difficulty in the analysis.

FIGURE 3.1 THE ORIGINAL DATA FORMAT

[redacted]

FIGURE 3.2 DATA FORMAT AFTER STAGE 1

[redacted]

3.16 Due to the complexity of the reformatting process, it was done in several separate stages. The first stage has the following objectives:

- Move all the information on each route to single file, so that we have one file per route;
- Replace the multiple columns describing the scheduled time with a single column;
- Convert the information stored as 'deviation from scheduled time' into a more straightforward 'time from departure', by adding the scheduled time at each stop to the deviation for that stop on each day;
- Remove unnecessary information, such as the columns containing a '1' flag where information is available;
- Sort the data by geographical order where the scheduled time information allows it (namely, in all cases but those when several stops have the same scheduled time);
- Replace the structure where each departure is formatted as a separate table with a variable number of rows, which is more difficult to handle automatically, with a simpler structure where the information on all departures is presented in one table; and
- Facilitate easy reading of information by having a single column for each day, containing the deviation from the scheduled time on that day.

3.17 The data format after the first stage is illustrated in Figure 3.2. This format is the starting point for the next reformatting step.

3.18 The objective of the second reformatting step is as follows to merge information from departures which are very close to each other (e.g. 07:12 and 07:14). The need to group together information on departures which are very close to each other came up since there were some revisions of route timetables during the period covered by the data; for example, the data included information for a 07:12 departure in October 2006 but a 07:14 departure in October 2007.

3.19 It is clear that the 07:12 and 07:14 departures can be seen as one. We only allowed such merging where there were only a few minutes difference between the departures. If the gap between departures is larger than this, the performance and reliability of each of them should be analysed separately.

3.20 The reliability analysis, presented later in this document, examines a series of travel time measurements for each departure of each route across many days. For the results of this analysis to be statistically significant, we need to have data on each departure from a large enough number of days. As a minimum, we need the series for each departure to include information from 20 days. Since the second reformatting stop joins some adjacent departures, it reduces the chance that we lose information on certain departures which were recently subject to minor timetable changes.

3.21 The objective of the third reformatting step is to correct the geographical order of the stops, which may be incorrect where several stops along the same route have the same scheduled time.

3.22 The correction of the order of stops required running a sub-process in which a list of stops per route, in the right order, was imported from an external file (one per route).

The format of the external list of stops imported into our database process is illustrated in Figure 3.3.

FIGURE 3.3 EXTERNAL LIST OF STOPS IMPORTED TO OUR PROCESS

Route	Position	Stop Code	Stop
56_in	1	45013720	Whinmoor Shop Ctr
56_in	2	45013721	Whinmoor Terminus
56_in	3	45025282	Oakdale Meadow
56_in	4	45010577	Naburn Approach
56_in	5	45013711	Marsett Way
56_in	6	45013712	White Laithe Road
56_in	7	45010784	Naburn Road
56_in	8	45010785	Naburn Close
56_in	9	45010786	Baildon Chase
56_in	10	45010579	Sledmere Square
56_in	11	45010527	Sherburn Road
56_in	12	45010526	Grimes Dyke School
56_in	13	45010529	Stanks Drive
56_in	14	45012836	Ash Tree Bank
56_in	15	45010528	Stanks Gardens
56_in	16	45010534	Stanks Lane North

3.23 The output from the third restructuring phase is therefore a series of individual files by route with appropriately merged departures, and sorted in the correct geographical order.

3.24 The objective of the fourth reformatting step is to convert the information stored as 'arrival time at the stop' to a format that applies to a stop-to-stop segment, i.e. 'time between stop A and stop B'. This is needed because later we will try to understand how features of the stop-to-stop segment affect the level of reliability on that segment.

3.25 In addition, the same (fourth) reformatting step is also used to identify stop-to-stop segments for which there is insufficient data (less than 20 days) and remove them from the database. The data format after the fourth restructuring process is illustrated in Figure 3.4.

FIGURE 3.4 THE DATA IN A STOP-TO-STOP FORMAT

[redacted]

3.26 Since information from some days has either been removed during the process, or had been missing from the start, it can be seen that the dataset has gaps in it in some columns. There may cause some computational problems in the analysis later. Therefore, the fifth stage of the reformatting process is meant to remove these gaps by shifting any cell that contains data to the leftmost column with available space.

3.27 By doing this we move from a format where each column represents a specific date to a format where each column stands for ‘day 1’, ‘day 2’ and so on, but is no longer identified with a specific date.

3.28 At the end of the fifth restructuring step, we have individual files for each route which no longer include any gaps in the data; there are neither days nor stops where information is missing.

3.29 The sixth (and final) restructuring includes the following actions:

- Merge all individual files into a single file covering all routes;
- Filter for errors whereby a specific stop-to-stop segment time looks irrationally high;
- Calculate the stop-to-stop speed and filter for errors whereby the speed seems unrealistic.

3.30 The result of this final stage is a single file containing all available information on stop-to-stop segments across all routes, with any erroneous data points removed.

Route Reliability Reports

3.31 The output from the processing described above is a comprehensive dataset which allows us to study the level of reliability of many bus routes in Leeds. Although this is not directly related to the evaluation of NGT, it seemed valuable to also use this information to inform the day-to-day management of the current bus network in Leeds.

3.32 We have therefore developed a ***route reliability report (RRR)*** for each bus route. RRR's are straightforward and informative summaries of the information we processed, and they follow a similar format for all routes.

3.33 Much of information about the level of reliability using a single measure. This measure was chosen to retain consistency with work done previously by the Leeds City Council. The reliability measure is calculated as follows:

$$R = \frac{P_{95} - P_5}{M}$$

Where:

- R is the value of the reliability measure for a stop-to-stop segment.
- P_{95} is the 95^{th} percentile of travel times along the segment. Namely, 95% of the travel times observed on that segment are below this value.
- P_5 is the 5^{th} percentile of travel times along the segment. Namely, 5% of the travel times observed on that segment are below this value.
- M is the mean travel time along the stop-to-stop segment.

3.34 The lower the value of this measure, the more reliable travel times are on the relevant segment. Note that when calculating the percentiles and the mean, each observation (i.e. each individual travel time) is from a different day, as we focus on day-to-day variations of travel time.

3.35 The difference between P_{95} and P_5, which forms the numerator in the formula above, indicates how wide the range of travel times is. Where this difference is high there is a lot of variation between days, i.e. a low level of reliability, and where it is low there is little variation.

3.36 Indeed, we could simply use the difference between the maximum travel time (P_{100}) and the minimum (P_0). Compared to (P_{95}- P_5) this would be more sensitive to outliers, i.e. to the very highest and very lowest travel times. The measure presented is slightly less sensitive to extreme cases since travellers are generally aware that infrequent occurrences of a very high or very low travel time do not necessarily represent the normal level of reliability of a bus route.

3.37 In the formula above, the difference between the high and low percentiles is divided by the mean travel time in order to make the reliability measure comparable across segments of different lengths. The numerator, if not divided by the mean, would have higher values on longer road segments, and would therefore not be suitable for examining how the level of reliability varies between different locations along a route. Once divided by the mean travel time, this measure can be interpreted the same way for short segments and long segments.

3.38 This measure of reliability is used for three of the five types of graphs presented in the RRRs. Graph 1 in each RRR presents how the level of reliability varies between different departures. For each departure, the graph presents the average reliability measure across all sections of the route. Hence, it provides evidence on temporal (rather than spatial) evolution of the level of reliability. This is illustrated in Figure 3.5, with the example of route 1 going southbound. It can be observed that for this route there is little difference in levels of reliability across the day.

3.39 Graph 2 in the RRR shows how reliability changes between different sections of the route. The reliability measure is presented for each stop-to-stop segment separately but is averaged across different departures. The segments are sorted not by their geographical order but in an ascending order of the reliability measure. Namely, the sections displayed at the right end of the graph are those that suffer from the poorest levels of reliability.

3.40 The graph also shows the average travel time on each segment. The reliability measure is displayed on the left hand side y-axis and the travel time in minutes on the right hand side. This is illustrated in Figure 3.6, again for route 1 going southbound. From this we can see that there is no obvious link between travel time across a segment and the level of reliability.

3.41 Graph 3 presents the same information as graph 2, namely the reliability measure for each stop-to-stop segment, averaged across all departures. However, now it is sorted in a geographical order of the stops along the route. From this graph it is easy to notice where along the route reliability levels are the best or the worst. It can also be observed whether unreliable areas are bunched together or scattered in various locations along the route. Again, the average travel time across a segment is also presented. This is illustrated in Figure 3.7.

3.42 Graph 4 presents the average speed implied by the travel time information across each stop-to-stop segment. It should be noted that the accuracy of the information we have on the distance between stops is low; distances are rounded to the nearest 0.1 mile. Given that many segments are quite short, it was necessary to group consecutive segments together in order to provide a slightly more robust estimate of speed. This is illustrated in Figure 3.8.

3.43 Graph 5 is in fact not a single graph but a series; a separate graph is presented for a departure in each hour of operation. The series of graphs does not use the reliability measure presented in graphs 1 to 3, because it shows travel times on each day as a separate curve, and does not aggregate data from different days into a single measure.

3.44 Each of the curves presented in graph 5 shows the full itinerary of a bus route that on a single day; the same departure is presented for all days. The full set of curves forms a band, the width of which implies how reliable the service is. If the multiple curves join into a relatively narrow band, it means that there is little variability between days, hence the service is reliable. Similarly, a wide band signifies an unreliable service.

3.45 The width of the band varies between different stop-to-stop segments along the route. This allows us to compare levels of reliability between different locations. This is illustrated in Figure 3.9 for a departure in the hour 15:00 to 16:00 of route 1 southbound.

3.46 The reason why there are gaps in the graph is that we wanted the x-axis (presenting the segments) to be uniform across the whole series of graphs 5 presented in one RRR. Therefore, a segment for which information is available for some parts of the day but missing for other parts is still displayed on the x-axis for all graphs.

3.47 Graphs 1, 2, 3, 4 and a series of graphs 5 form a full RRR for one route. **A disk containing all RRRs we created is attached to this report.**

Geographic Information Handling

3.48 To estimate the levels of reliability of the NGT service (and other future services) we need to be able to associate reliability with some explanatory variables.

3.49 Some of the reasons why travel times vary from one day to another are too complex or too subtle for us to model (for example, day-to-day changes in weather conditions, accidents, journey cancellation because of workforce issues, vehicle breakdown, and a variety of other operational matters). But we still wanted to analyse whether some of the variation between days is due to factors which can be identified, and which have to do with the physical features of the urban road network.

3.50 Before making any attempt to examine this, we needed to collect data on those physical features which may, or may not, help us later to explain differences in the level of reliability between different part of the network. We therefore embarked on a process of creating a dataset, which include both the information on the different levels of reliability (from the analysis above) and the data that will be used to try to explain these differences.

3.51 The following are various characteristics of street segments which we felt may have an effect on the level of reliability of bus routes using them:

- Whether a bus lane is operation, and at what times of day;
- If a bus lane exists, what type of physical segregation from general traffic is used;
- Whether or not a bus lane is also used as a cycling lane or adjacent to a cycling lane;
- The number of lanes for general traffic;
- Whether or not bus stops are in lay-bys;
- The density of pedestrian crossings;
- The type of approach for buses before junctions; and
- The density of minor road intersections.

3.52 A large data collection exercise was undertaken using Google StreetView to compile the dataset of the above physical features of the all road sections we have reliability information for. The datasets covers most major streets in Leeds. The bus stops and street segments covered are illustrated in Figure 3.10.

FIGURE 3.5 GRAPH 1 IN THE ROUTE RELIABILITY REPORT

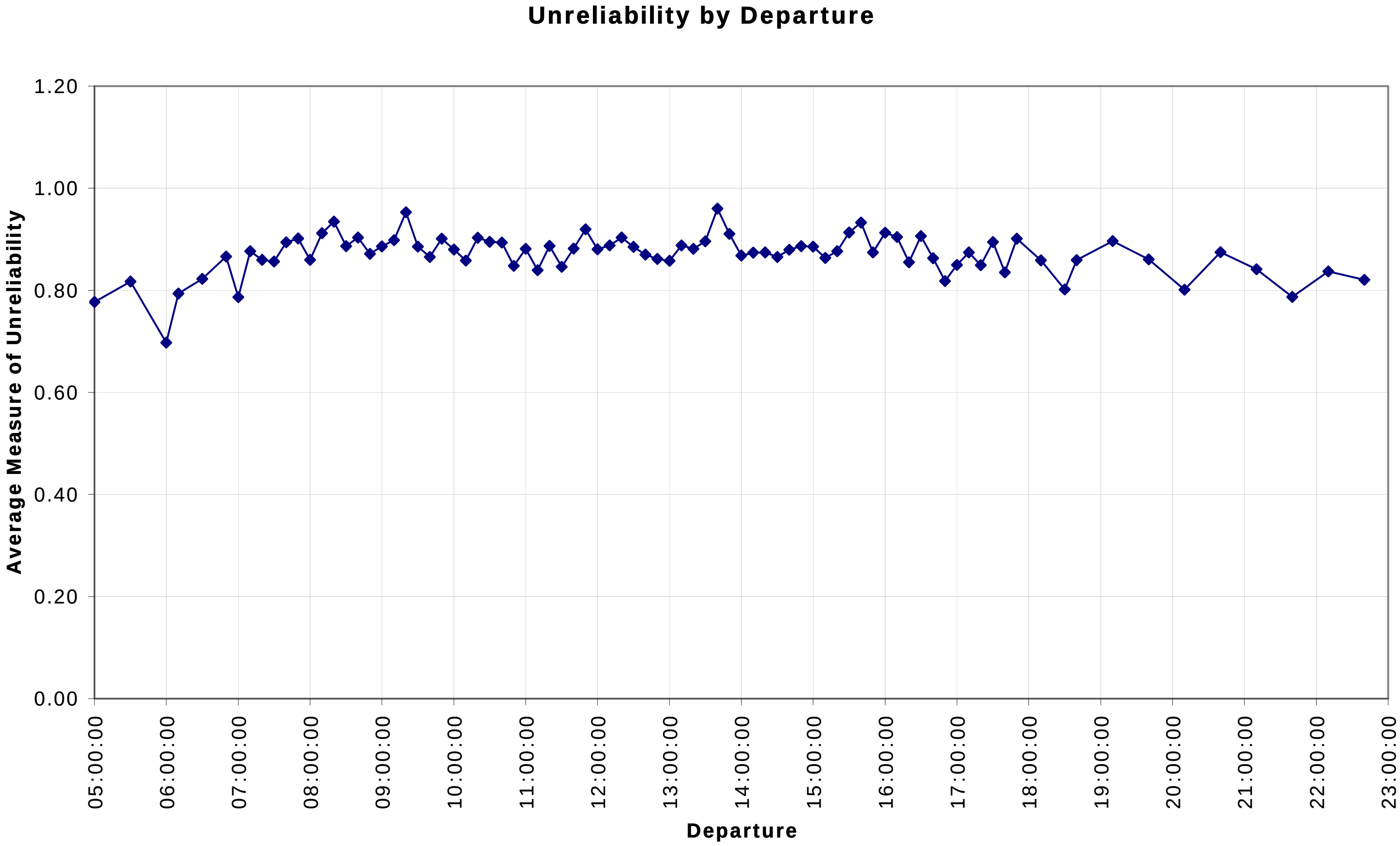


FIGURE 3.6 **GRAPH 2 IN THE ROUTE RELIABILITY REPORT**

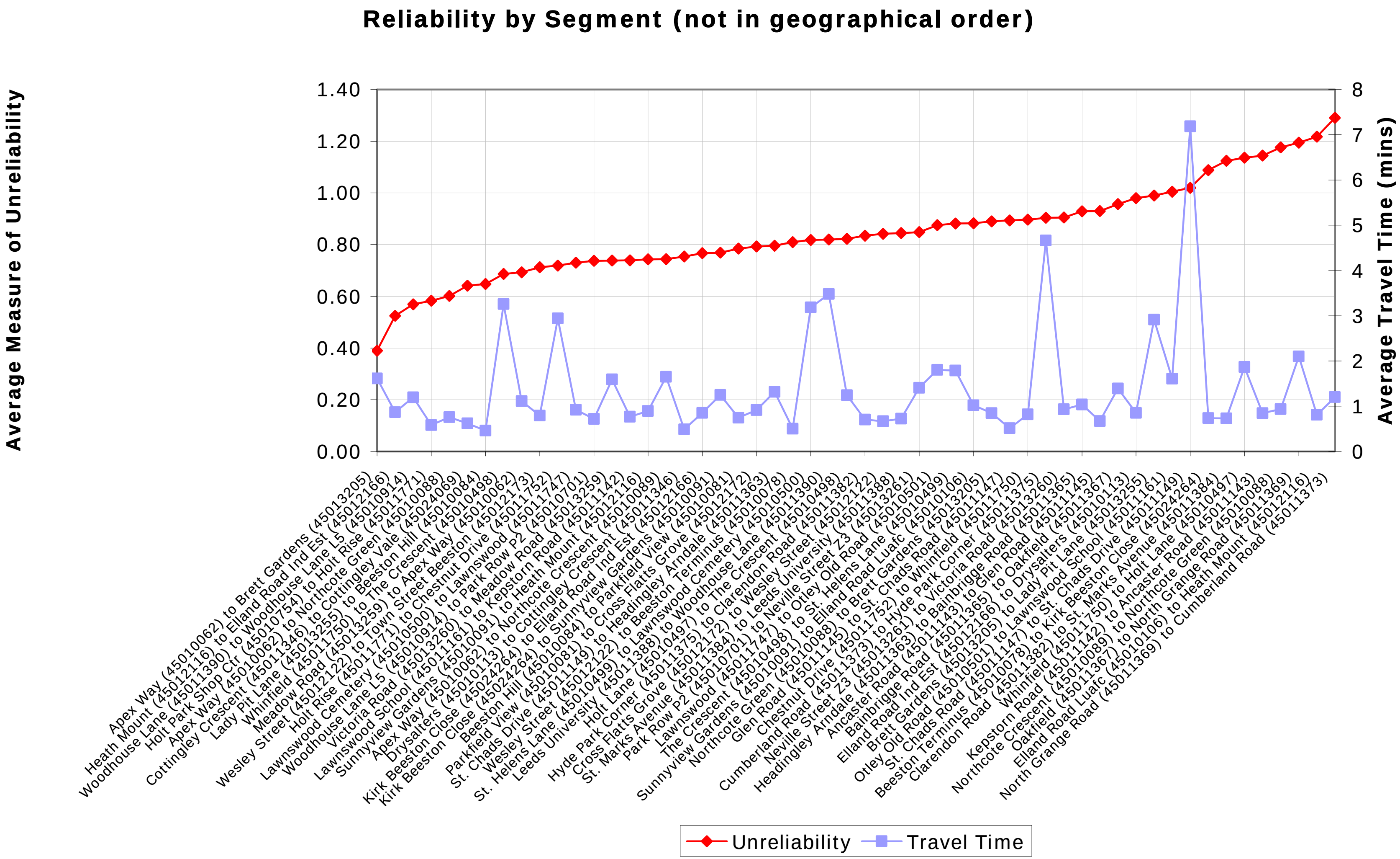


FIGURE 3.7 GRAPH 3 IN THE ROUTE RELIABILITY REPORT

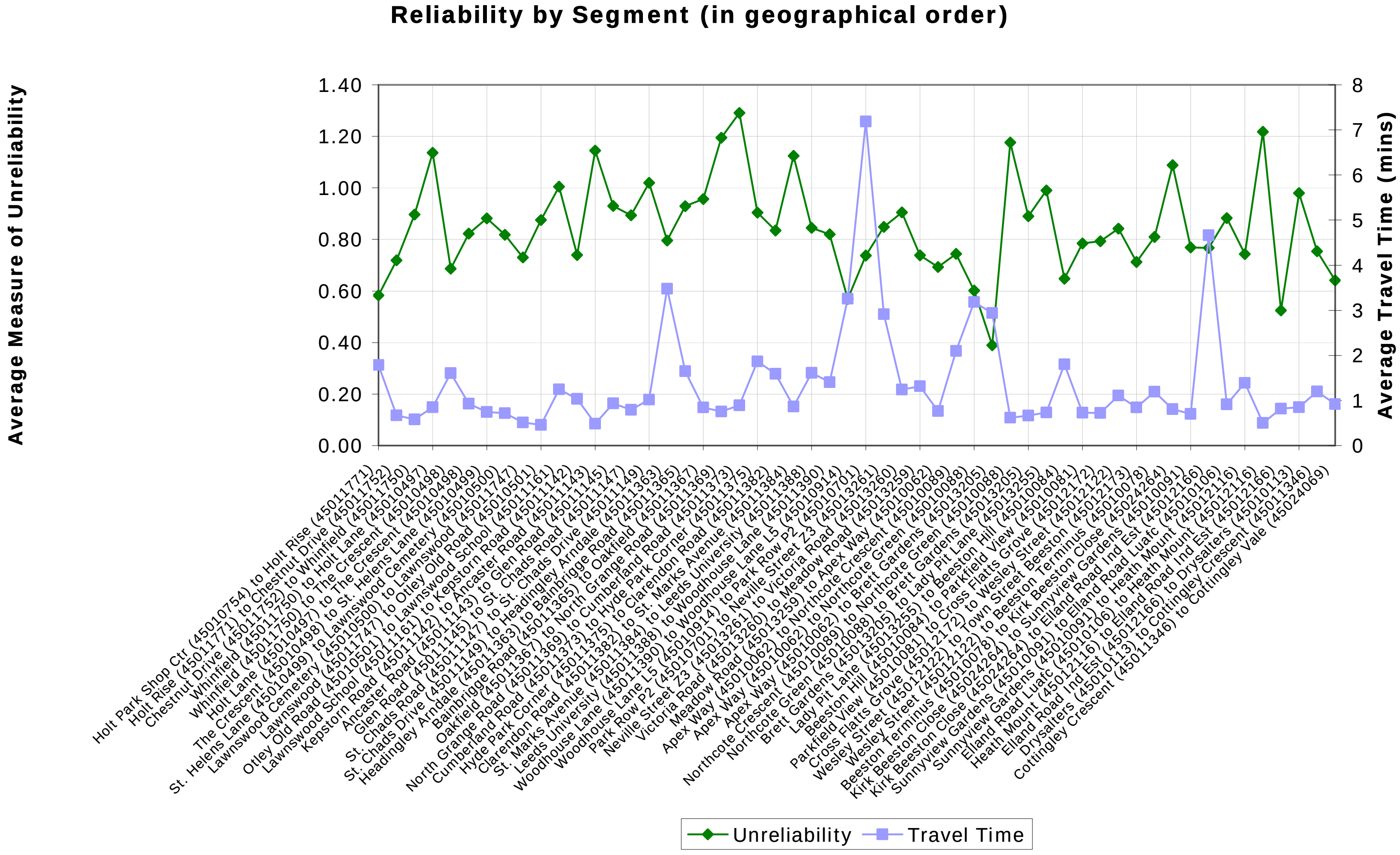


FIGURE 3.8 GRAPH 4 IN THE ROUTE RELIABILITY REPORT

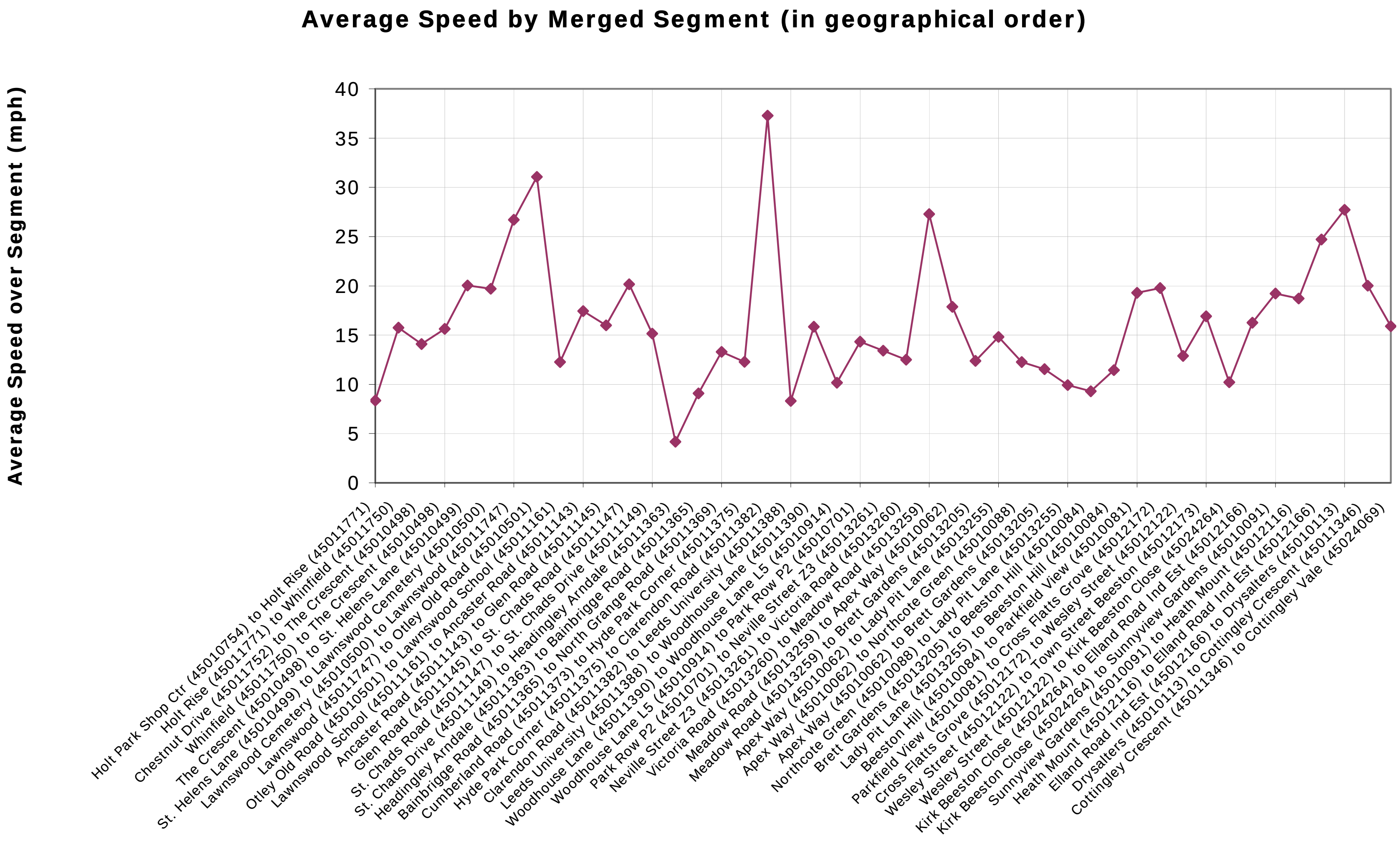


FIGURE 3.9 GRAPH 4 (ONE OF A SERIES) IN THE ROUTE RELIABILITY REPORT

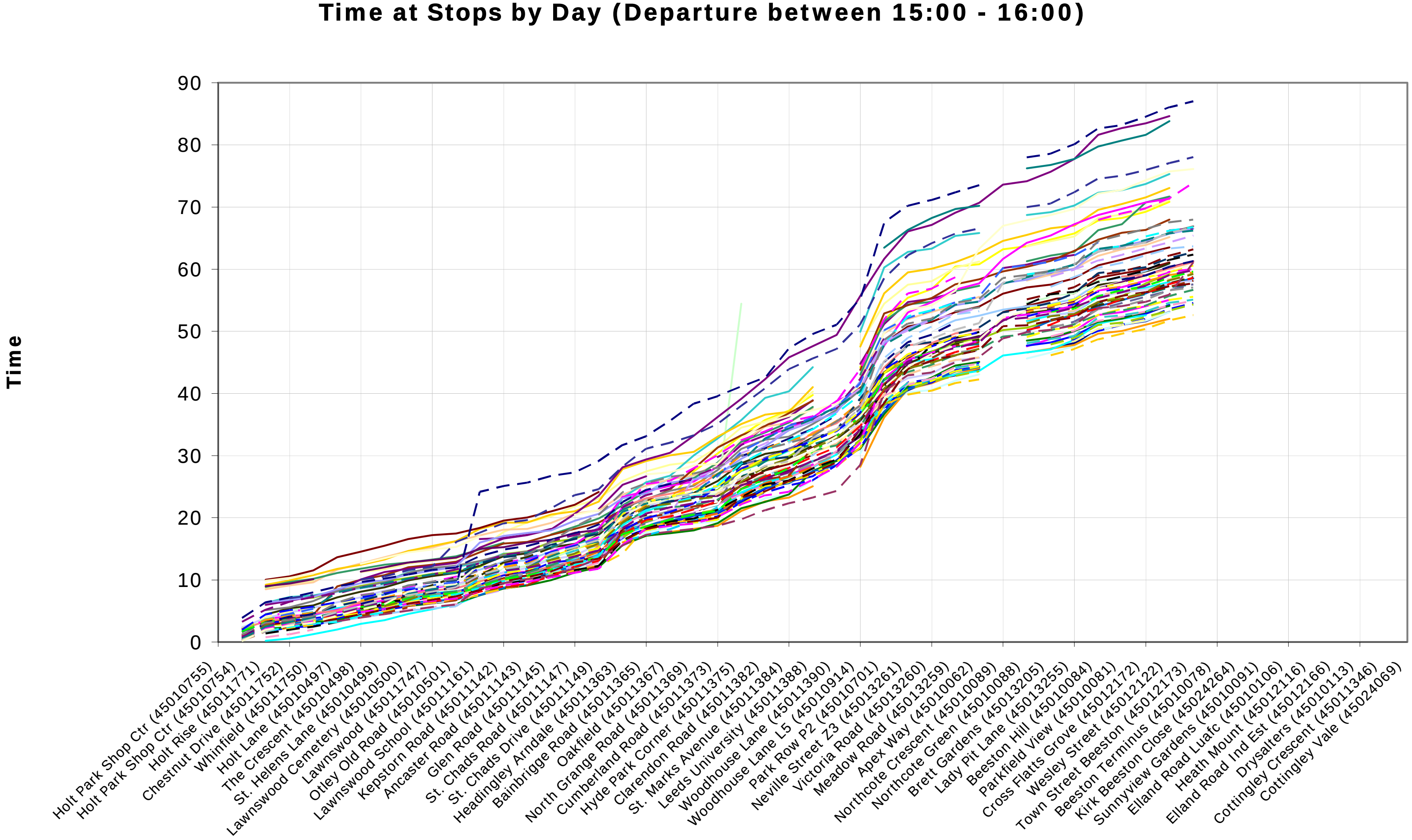
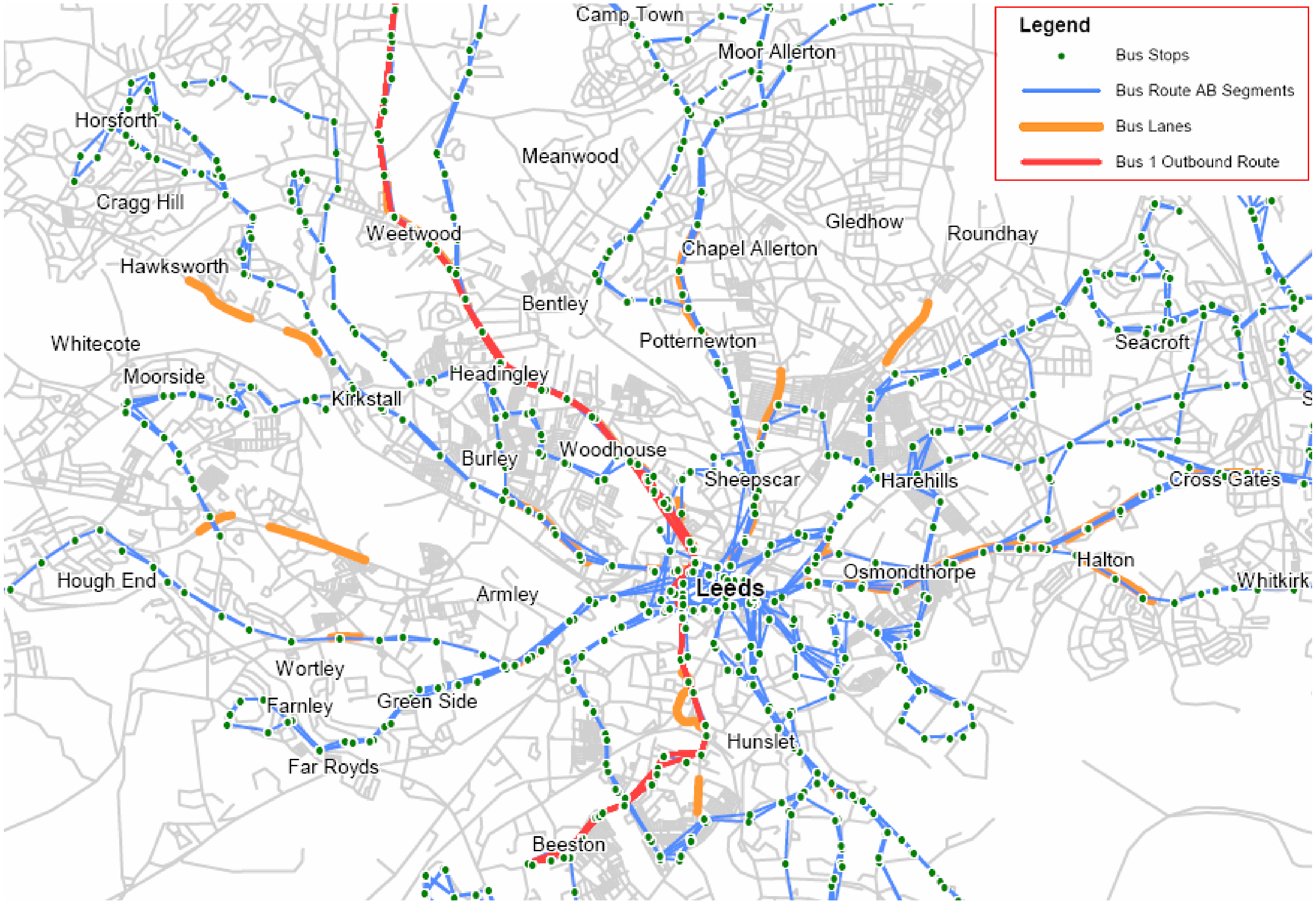


FIGURE 3.10 THE GEOGRAHPICAL COVERAGE OF THE NETWORK ANALYSIS



3.53 We added to the dataset some additional fields of information that seemed important to test as factors that affect the level of unreliability. The primary additional field was the level of demand at a stop, although it should be highlighted that we only had information on the average demand, while it is likely that the variations in demand are a more likely cause of fluctuations in times between days.

3.54 With this dataset we tried to estimate a regression model that expresses the level of unreliability on a stop-to-stop segment as a function of different attributes of that segment. A very large number of model forms were tests, with different combinations of explanatory variables from the dataset.

3.55 The modelling experiments showed that there is a significant correlation between the level of reliability and various features of the road. This is indicated by high values of the t-test. Some examples of variables identified as affecting the level of reliability with high significance are the segment length (t-stat = -24.3), the provision of bus priority (t-stat = -6.8), and dummies for highly-congested time periods (t-stat = 6.5).

3.56 However, the amount of 'noise' in the data is still very high. Therefore, even if we could reach models with a high t-test, they overall had a very low R^2(below 0.1), indicating low explanatory power. This in itself is not surprising, given that it is very ambitious to build a powerful model for a random phenomenon. Many previous attempts to do so in much larger academic studies have not been very successful.

3.57 We therefore chose a more pragmatic approach for our current purposes. It is highly likely that the main reason why NGT will be more reliable than current bus services has to do with the level of priority it will get in traffic, manifested by high-quality segregation along some sections, in addition to priority in traffic signals.

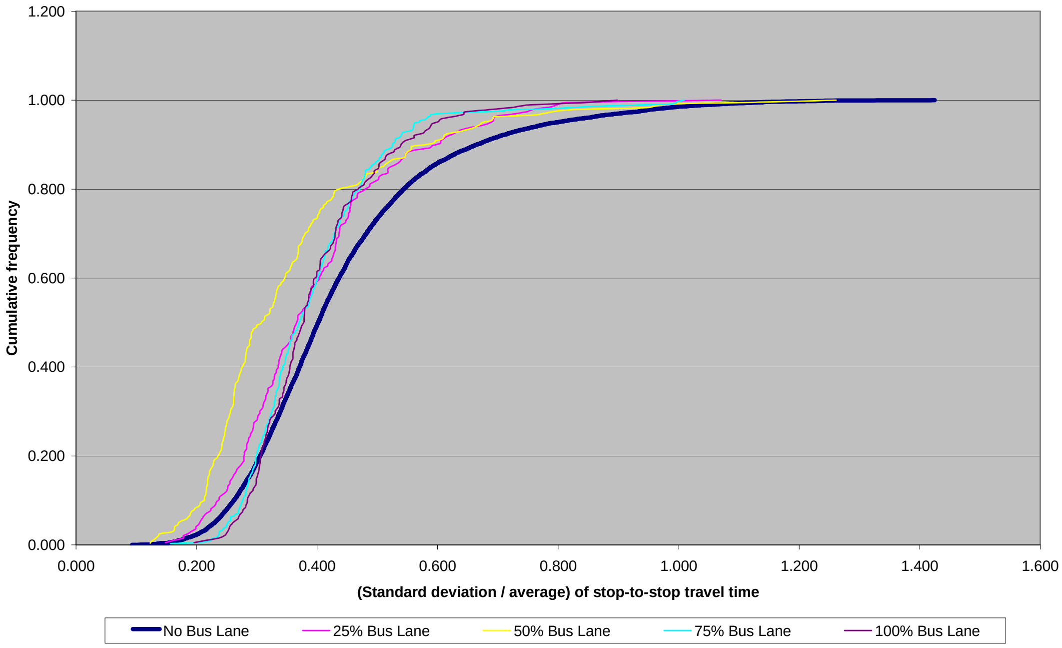
3.58 This means that we can base our estimate of the NGT level of reliability on some analysis of the difference between roads segments with different levels of bus priority in the existing network. Some analysis of these differences is demonstrated in Figure 3.11.

3.59 The figure is presented as a cumulative frequency curve. The curve is made of various data points, all taken from the dataset described earlier. The x-axis displays the ratio of the standard deviation of travel times to the average travel time. We have used this ratio, rather than the reliability measure used in the RRRs, for various computational reasons, but the relationship revealed is very similar, and the shape of the graph would be almost identical if we presented the same reliability measure.

3.60 The importance in the cumulative frequency curves is in the difference between the rightmost curve and the other ones. The rightmost curve is based on data from road segments with no bus priority. The other curves are based on data from segments with different levels of priority. These are not the only curves we have looked at; we created different curves for bus lanes with different features. They show various relationships, which fed into our specification of scenarios, as described further below.

3.61 The general understanding is that priority in traffic clearly affects the level of reliability. But priority measures given in different ways, in different parts of Leeds, vary from each other widely in their ability to improve reliability. Detailed analysis of the data clearly indicates that there are a small number of bus lane segments with very poor levels of reliability, which skew the results and make the average impact of bus lanes much lower than that of other bus lanes. This is likely to be due to various local factors and the specific geometrical configuration of each such segment, which we are not aiming to pick up in our analysis.

FIGURE 3.11 CUMULATIVE FREQUENCY OF RELIABILITY IN SEGMENTS WITH VARYING LEVELS OF BUS PRIORITY



Reliability Scenarios

3.62 Following the analysis of the effect of different bus lanes on the level of unreliability, we specified two 'reliability scenarios'. The feature that defines each scenario is the percentage reduction in the standard deviation of travel times in the corridor with priority (namely, in our case, the NGT corridor) compared to the other roads.

3.63 Scenario A is defined using the data from bus lanes in Leeds, excluding those that perform poorly. The improvement in reliability (i.e. reduction in standard deviation of travel times) in this scenario is 17.5%.

3.64 Scenario B is not based on the analysis above, but on a literature review. The review sought evidence in publications from other cities about the extent to which successful priority schemes improved the reliability of public transport services. The following are links to websites presenting some of the main sources reviewed:

- http://www.konsult.leeds.ac.uk/private/level2/instruments/instrument011/l2_011c.htm
- http://www.itsbenefits.its.dot.gov/its/benecost.nsf/ID/F1A967F0D3CD5DE58525733A006D4B07?OpenDocument&Query=BApp
- http://bhls.eu/IMG/pdf/Manchester_UK_Dublin_meeting_QBC_presentation19.02.08.pdf
- http://www.straphangers.org/reports/Final_M96_report_2.PDF

3.65 The conclusion from the review was that careful design can improve reliability by 50%. Clearly, the justification for the big difference from the levels found in Leeds need to be explained. The key point is that bus priority in Leeds today, even where considered successful, makes limited use of advanced priority in signal settings and extensive enforcement of misuse by private vehicles, while there advanced measures will form an integral part of the NGT scheme.

3.66 Since the use of new technology to control and monitor priority is considered an integral part of the NGT scheme, it was decided to see a 50% improvement in reliability as the central case. For proper disclosure it should be stated here that the reliability benefits calculated later, which are included in the business case, will not materialise if the level of priority of the NGT routes is at a similar level to that provided in bus lanes in Leeds today.

Conversion to Origin-Destination Format

3.67 The broader appraisal study of NGT is undertaken at an origin-destination (OD) level, with changes in the generalised cost affecting the demand in a different way for each OD. By contrast, the format of the reliability data is at a stop-to-stop segment level, and our forecasts of the improvement also apply on this basis.

3.68 It was therefore necessary to create a procedure to convert information on unreliability between these formats. The procedure generally applies to any scenario, but we implement it separately for the scenario with NGT and the scenario without NGT, since there are different levels of reliability in each of them.

3.69 To carry out this conversion, we first created a new table that lists all the stop-to-stop segments used when travelling by bus or NGT from each origin to each destination. This is an output from the assignment model developed in Voyager. Since many ODs are served by more than one possible route, or combination of routes with transfers, this list also includes that proportion of demand for each OD using each stop-to-stop segment.

3.70 There were segments that are used by some ODs for which we did not have reliability data from the route used, but we did have reliability data from other routes using the same segments. In such cases, we have used average level of reliability experienced by other routes along the same segment.

3.71 There were also segments used by some ODs for which we did not have reliability data at all. In such cases, we used the average level of reliability across all segments in the Leeds network.

3.72 Note that this in-filling process does not change our estimates of travel time, because our measure of reliability is normalised by time. The assumption that lies behind this process is that any segment with missing data experiences the average amount of reliability seen across all other segments.

3.73 As a product of this work, we can go through all segments in the network and assign their levels of unreliability to the relevant OD's only, and thus convert the segment-based reliability to OD-based reliability. However, a major difficulty in applying this process is the fact that all measures of reliability are not linear-additive. The level of reliability across a series of segments is not simply the sum of the levels of the individual segments.

3.74 To account for this we have set up a process that can add up different levels of reliability from different segments, taking their non-linear nature into consideration. The process works as follows.

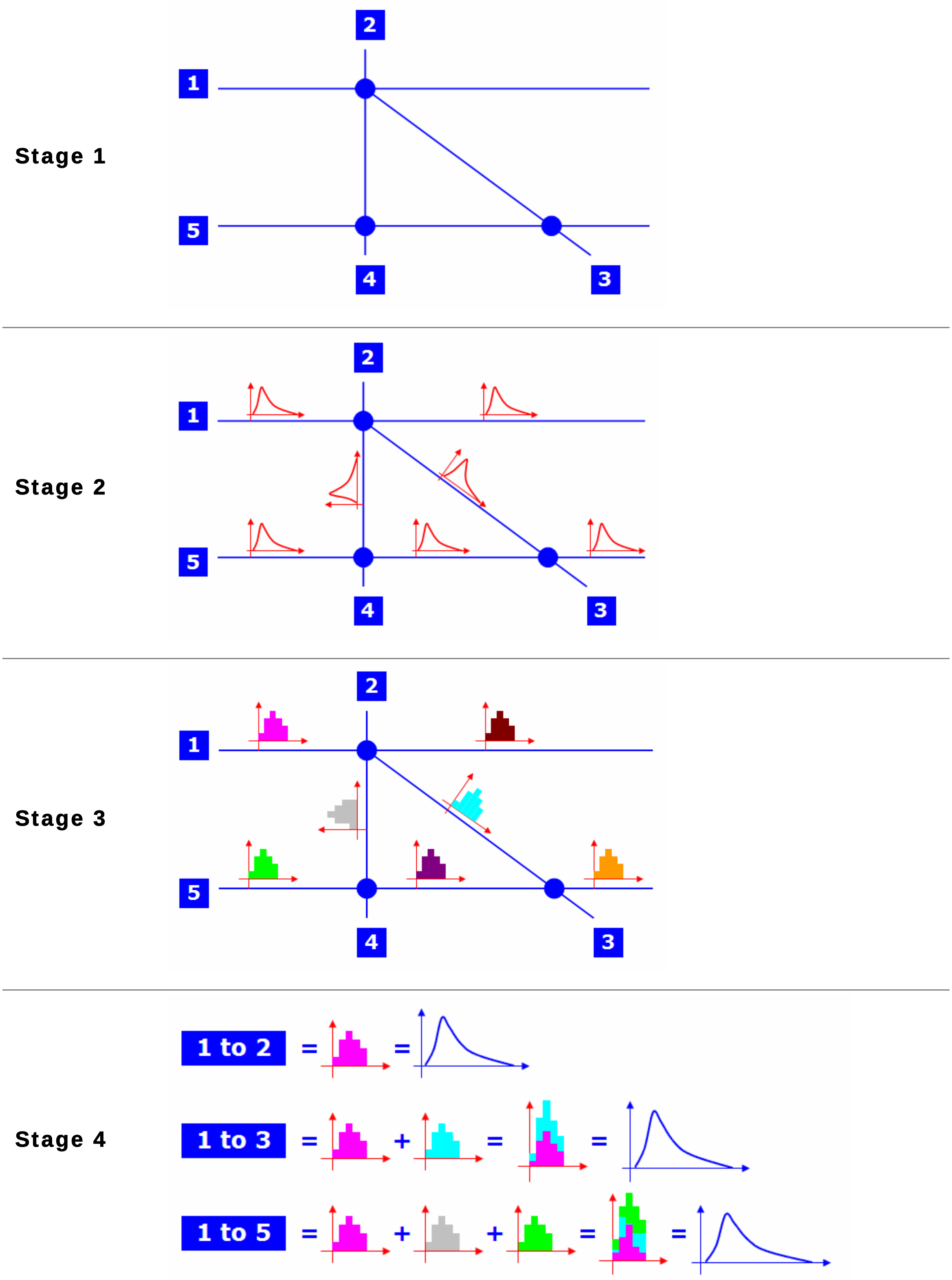
3.75 For each segment that has a given average travel time and given level of reliability, this can be described as a distribution of travel times. We can therefore use this distribution to generate a set of 'realisations', i.e. a large number of travel times that together have the given mean and standard deviation. As the 'large number' we use 50 realisations for each distribution.

3.76 We assume here that travel times follow the Gamma distribution. This is a commonly acceptable assumption, which is suitable for our needs because just like travel times, the Gamma distribution is always positive and it skewed to the right. It has a long tail thus allowing a proportion of times to be high.

3.77 Once we have 50 travel times for each segment, randomly-generated based on a given mean and standard deviation, we add times from all the segments used by a certain OD (based on the table created earlier) to get 50 different total journey times for this OD. The 50 total times for this OD form a distribution of travel times for the OD, from which we can calculate any measure of reliability at the OD format.

3.78 To clarify this process, note that we add together all the 1st realisations (from many distributions of 50 segment-based times) across all segments to get the 1st travel time in the distribution of the OD-based travel times. We add together all the 2nd realisations from the many distributions, across all segments, to get the 2nd travel time in the distribution of the OD-based travel time. Once we do this 50 times, we have a full distribution of travel times for this OD. This process is demonstrated in figure 3.15 below.

FIGURE 3.12 CONVERTING RELIABILITY ESTIMATES FROM SEGMENT TO ORIGIN-DESTINATION FORMAT



3.79 When this process is done separately for the scenarios with and without NGT it will give different results, because in the scenario with NGT some stop-to-stop segments have improved levels of reliability. After the conversion to OD format, this will have a major effect on ODs adjacent to the NGT route, and a small effect on other ODs if they have a certain chance of using this route. OD where no trips are made by NGT will not be affected.

3.80 There were some cases where there was no sufficient information on the level of reliability on all relevant segments (or routes) for a specific OD pair. In these cases, if there was an adjacent OD pair that does have all the required information, we sometimes 'borrowed' the results of this analyse between from that OD. This was only done if the ODs are close enough to each other to assume that they use similar travel itinerary.

Conclusion

3.81 We have presented in this chapter various processes undertaken to obtain forecasts of the level of reliability with and without the NGT scheme. We carried out detailed analysis of information on the levels of reliability on bus routes in Leeds in 2006 and 2007. These result in a comprehensive set of Route Reliability Reports (see attached disk) which are to be used by Metro and Leeds City Council for a wider range of purposes than the NGT business case.

3.82 We estimated the level of reliability in Leeds on streets where bus priority measures are in operations. While these do improve the level of reliability when compared to streets without priority measures, the extent of improvement is not sufficient to be translated into reliability benefits in the appraisal. The expected levels of NGT reliability are assumed to be at similar levels to those reported in other places where a similar scheme is in operation.

3.83 We also showed here how we convert the estimates of stop-to-stop reliability to values at an origin-destination level. Further conversion of outputs from the demand and supply analysis into inputs for the appraisal is described in the next chapter.

4 Inputs for the Appraisal

Introduction

4.1 In the previous chapter we outlined our methodology for calculating the level of reliability for each OD affected by NGT. To estimate the costs and benefits associated with the scheme, we need to combine the levels of reliability with the unit values determined in Chapter 2. This requires a separate methodology, because the unit values are for earliness and lateness, i.e. they are not given directly in units of reliability.

4.2 We introduce the impact of reliability as a proportion increase in the cost of travel time. This is convenient for two reasons. First, the appraisal software does not have a readily-available functionality to deal with reliability, and by adding the cost of unreliability to the time costs we avoid any major difficulty this may cause. Second, our unit values for earliness and lateness are convenient to use this way, because we have calculated them as a ratio of the standard value of time.

4.3 This chapter discusses the methodology for generating the required inputs for the appraisal.

Calculating Costs While Allowing for Re-Scheduling

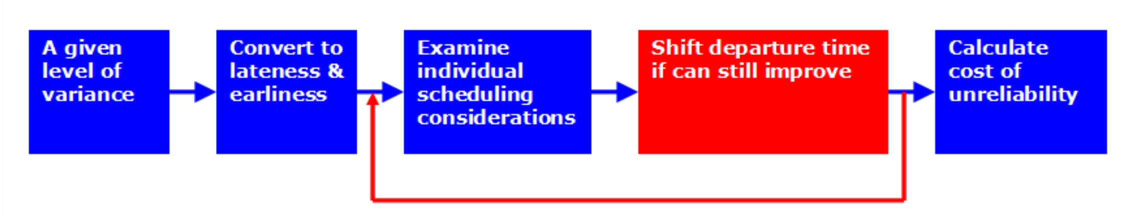
4.4 Earlier in this report we expressed the opinion that a realistic estimate of the generalised cost of travel, including the impact of unreliability, cannot ignore travellers' trip re-scheduling behaviour. That is, it needs to account for the fact that travellers change their departure time to trade-off between amounts of earliness and lateness.

4.5 A full iterative 'loop' between the level of reliability and travellers' choice of a departure time (and vice versa) is not supported by standard modelling packages, and would be difficult to set up as a robust bespoke application. Instead, we propose here a simplified version of the same approach, i.e. a process whereby travellers still shift their departure time, even if not with a full iterative process. This approach is easier to set up and appears relatively realistic in behavioural terms despite its simplicity.

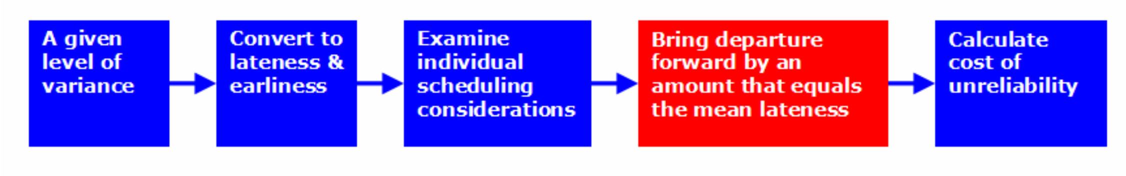
4.6 The difference between the 'full loop' approach and this shortcut version is presented in the figure below. While the full loop repeats the calculation of the level of earliness and lateness as many times as required for the behaviour of multiple travellers to stabilise, the alternative approach only corrects the initial estimate once. Once there is an estimate of the mean lateness and the mean earliness, we assume that each traveller brings their departure forward by an amount that equals their mean lateness. The mean earliness and lateness are not calculated again although they would clearly change.

FIGURE 4.1 A SIMPLE APPROACH TO ALLOW FOR TRIP RE-SCHEDULING

Departure time choice loop approach



Shortcut approach



4.7 The justification for carrying out such calculation is that, since the value of a minute when arriving early is lower than the value of a minute when arriving late, travellers will always be willing to shift their average arrival time so that it falls before their preferred arrival time. Note that the base assumption of the shortcut approach does not imply that travellers avoid any lateness; it is very likely that the mean lateness after such shift will remain higher than zero.

4.8 This is demonstrated by the following example, for the case of a single traveller:

- Assume that the given departure time is 08:00, the desired arrival time is 09:00 and the average journey time is 55 minutes.
- Assume also that based on the level of reliability estimated separately, likely arrival times on 5 consecutive days are 08:50, 09:03, 08:48, 08:50 and 09:04.
- The amount of lateness on these 5 days would be 0 on the first, third and forth days; 3 minutes on the second day and 4 minutes on the fifth day.
- The mean lateness is therefore the average of 0, 3, 0, 0 and 4, that is 1.4 minutes.
- We can be quite confident that this traveller would agree to bring his or her departure forward by 1.4 minutes because then each minute of mean lateness would be traded for extra earliness.
- We would not be confident that this traveller would accept a larger shift, as this might bring the amount of earliness to a level that has a higher cost than that of late arrival.
- The shortcut method does not re-examine this as it does not involve a departure time shift loop. It will therefore still result in an inaccurate estimate of the cost of unreliability. However, this is the best estimate we could reach without an iterative correction loop, and it does seek to balance to costs of earliness and lateness even if in a simplified manner.

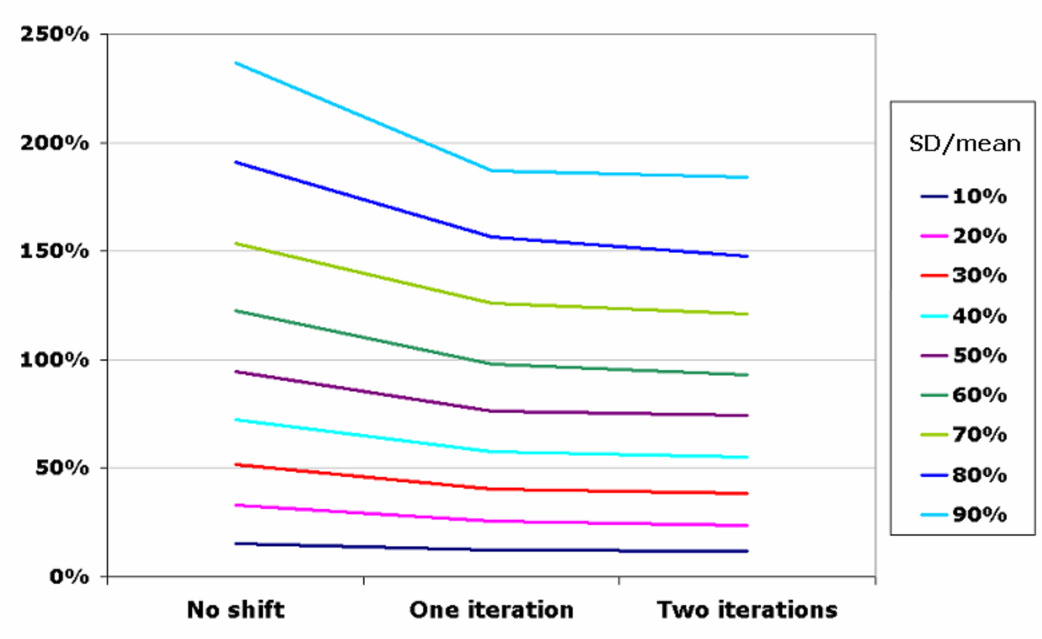
4.9 The conversion of the level of reliability per OD to the percentage increase in costs therefore follows these steps:

- Random travel times are generated for OD pair, based on its level of reliability (calculated as described in Chapter 3). The generation of the random times follows the same concept of 'realisations' described earlier, for 50 days.
- The earliness and lateness to the destination are calculated separately for each realisation. For example, if the average travel time is T and on a specific day the realised travel time is 1.1T, then the lateness on that day is 0.1T.
- The mean lateness and earliness across all days are calculated.
- We assume that the departure time of travellers on each OD is shifted so that they leave earlier by an amount that equals the mean lateness.
- The new mean earliness and lateness are multiplied by the respective appraisal values and added together to form the cost of unreliability.
- The ratio of the cost of unreliability to the cost estimate before this addition forms the relative impact of reliability for the respective scenario.
- Benefit-cost ratios can then be calculated in the standard way.

4.10 The following figure illustrates the effect of doing this with the 'shortcut' approach. The vertical axis shows the percentage increase in cost due the inclusion of reliability costs. This is done for several cases that differ from each other in the level of reliability. It shows that when departure time shift is not allowed, the cost increase is about 30% higher; but this is clearly an overestimation of the real impact of reliability.

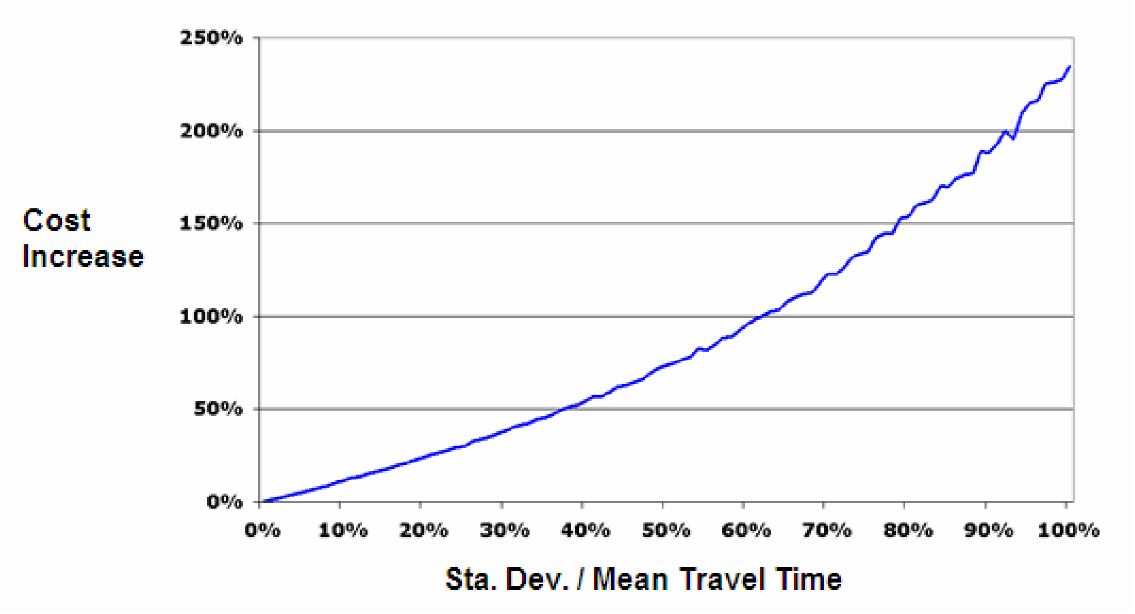
4.11 Once there is an allowance of the departure time shift, one iteration seems sufficient, and the effect of the second iteration is marginal.

FIGURE 4.2 A SIMPLE APPROACH TO ALLOW FOR TRIP RE-SCHEDULING



4.12 The overall increase in time costs due to the inclusion of reliability in the appraisal is shown graphically below. Note that the percentages shown depend on the relative values of earliness and lateness defined earlier.

FIGURE 4.3 COST INCREASE FOLLOWING THE INTRODUCTION OF UNRELIABILITY COSTS



The Impact of Reliability on the NGT Business Case

4.13 The final output of the process described here is a table indicating the percentage increase in cost for each OD in the study area, either with NGT or without. This is used to increase the time costs by the relevant proportion and then feeds into the TUBA process in the standard way.

4.14 The detailed appraisal process and outputs are described in the Major Scheme Business Case.

5 References

5.1 The following is a list of articles and publications reviewed while developing the approach presented here:

Abdel-Aty, M. A., Kitamura, R. & Jovanis, P. P. (1995), Investigating Effect of Travel Time Variability on Route Choice Using Repeated-Measurement Stated Preference Data, Transportation Research Record, No. 1493, pp. 39-45.

Atkins Consultants LTD (1997), Bus Reliability Study – Stated Preference Research, Great Britain.

Bates, J., Dix, M. & May, T. (1987), Travel Time Variability and its Effect on Time of Day Choice for the Journey to Work, Transportation Planning Methods, Proceedings of seminar C held at the PTRC Summer Annual Meeting, University of Bath, Vol. P290, pp. 293-311.

Bates, J., Polak, J., Jones, P. & Cook, A. (2001), The Valuation of Reliability for Personal Travel, Transportation Research E, Vol. 37, pp. 191-229.

Batley, R., Ibanez, J. N., Wardman, M., Shires, J. & Whelan, G. (2007), A Discrete Choice Study to Assess the Impact of Reliability on Passenger Rail Demand, Paper presented at the European Transport Conference in Noordwijkerhout, The Netherlands.

Black, I. G. & Towriss, J. G. (1993), Demand Effects of Travel Time Reliability, Centre for Logistics and Transportation, Cranfield Institute of Technology, Great Britain.

Gaver, D. P. Jr. (1968), Headstart Strategies for Combating Congestion, Transportation Science, Vol. 2, No. 2, pp. 172-181.

Hall, R. W. (1983), Travel Outcome and Performance: The Effect of Uncertainty and Accesibility, Transportation Research B, Vol. 17, No. 4., pp. 275-290.

Hollander, Y.(2009), What Do We Really Know about Travellers' Attitudes to Travel Time Variability? Paper presented at the International Choice Modelling in Harrogate, UK.

Hollander, Y. (2007), The Cost of Bus Travel Time Variability. PhD thesis, Institute for Transport Studies, University of Leeds.

Hollander, Y. (2007), The Cost of Bus Travel Time Variability, research report, Institute for Transport Studies, University of Leeds.

Hollander, Y., Mackie, P. J. & Liu, R. (2007), Estimating Traveller Costs with a Departure Time Choice Loop, Paper presented at the European Transport Conference in Noordwijkerhout, The Netherlands.

Hollander, Y. (2006), Direct Versus Indirect Models for the Effects of Unreliability, Transportation Research A, Vol. 40, No. 9, pp. 699-711.

Hollander, Y. (2005), The Attitudes of Bus Users to Travel Time Variability. Paper presented at the European Transport Conference, Strasbourg, France.

Hollander, Y. (2005), Travellers' Attitudes to Travel Time Variability: Inter-modal and Intra-Modal Analysis. Paper presented at the 3rd international SIIV congress - People,