improving public transport through technology

# Accuracy and Quality Of Real Time Predictions

RTIG Library Reference: **RTIGT041-1.0**

December 2020

Availability: Members Only

Price:

| | |
|---|---|
| Foundation Members: | Free |
| Full Members: | Free |
| Associate Members: | £500 |
| Non-members: | £1200 |

# List of contents



## Status of this document

This document is Published.

If there are any comments or feedback arising from the review or use of this document please contact us at secretariat@rtig.org.uk

# 1 Introduction

## 1.1 About this document

1.1 This document has been produced for the Real Time Information Group (RTIG). It provides RTIG members with guidance on how to measure the quality and accuracy of real time information in the form of predictions that are provided to the customer. It also provides recommendations for the presentation of predicted information to the customer.

## 1.2 Background and context

1.2.1 In recent years, there has been an increasing focus on delivering improved public transport information to passengers. For authorities, this is seen in part, as a means of achieving broader policy objectives such as increasing modal shift away from private car use and therefore easing congestion on the roads; as well as improving the environment. For bus operators, this is seen as a key part of improving the image of the public transport offer.

1.2.2 The result of this focus is that most bus operators are now providing real time data for customers. Indeed, in 2021 bus operators will be required under the Bus Services Act 2017 to provide location data to the Bus Open Data Service for the majority of their services. This presents a unique opportunity to ensure consistent provision of bus location data to customers.

1.2.3 Previous reports and specifications from RTIG have covered a wide range of topics, and a number of the reports have made passing reference to the quality and/or accuracy of real time information: in the form of predicted arrival and departure times. However, up to now, none have specifically covered the quality and accuracy of predictions.

1.2.4 With the near ubiquitous provision of location data for the bus fleet in the UK within reach, it is timely to consider the quality of data to ensure that the information produced is fit for purpose. This report sets out to address this gap through advice from RTIG.

## 1.3 Scope

1.3.1 This report has no statutory or other legal basis and is purely to provide advice to suppliers, authorities and bus operators who supply or use real time information systems.

1.3.2 All aspects of a real time system: from originating source data through to dissemination channels, are potentially impacted by this report.

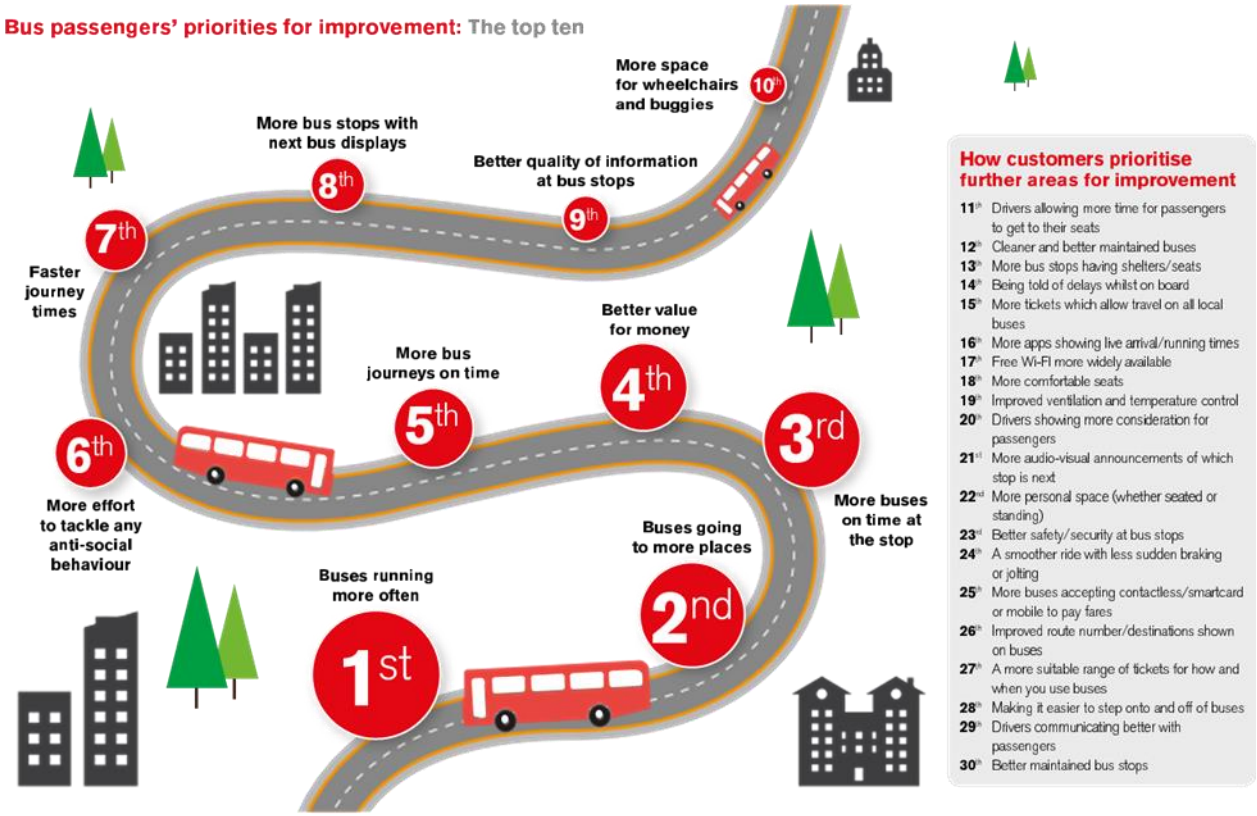
## 1.4 Acknowledgements

1.4.1 RTIG is grateful to the members of its Accuracy and Quality of Prediction Working Group for contributing to the construction and validation of this document - in particular: Swiftly Inc, TfL, SYPTE, WYCA, TfGM, Trapeze, VIX, ITO World, Transport API, Passenger and r2p UK.

# 2 Why are predictions important to customers?

2.1 Bus routes and networks are inherently unstable - making it difficult to maintain reliable schedules for passengers. To mitigate this, real time information (RTI) such as bus arrival times to each stop, can be used to update schedules - thereby increasing the perceived reliability of the system from a user perspective. However, a potential drawback of these real time traveller information systems is that they can provide a false sense of precision. That is, users expecting a certain arrival time based on RTI can develop increased negative feelings about bus services if the bus is earlier or later than expected. It is therefore, important to ensure that RTI is as accurate as possible.

2.2 Transport Focus[1] regularly survey passengers to understand their priorities. In their September 2020[2] report on bus passengers' priorities for improvement, they identify a top ten of passenger priorities which is set out in the diagram below:




1.1

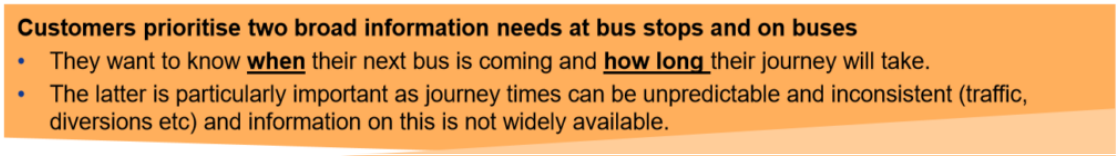
[1] https://www.transportfocus.org.uk/

[2] https://www.transportfocus.org.uk/research-publications/publications/bus-passengers-priorities-for-improvement-2/

2.3 Of the ten priorities identified, RTI can a have an impact on five:

- more buses on time at the stop;
- more bus journeys on time;
- faster journey times;
- more bus stops with next bus displays; and
- better quality of information at bus stops.

2.4 Providing accurate RTI helps passengers better plan their trips and minimise waiting times – both of which contribute towards a better customer experience. Passengers are typically interested in bus arrival times at the bus stop and the journey time to their destination. This is confirmed by the recent 'Digital Bus Innovations Report'




2.5 In the UK, there has been little published work carried out on the accuracy of predictions to either measure the performance of existing RTI systems or to help procure new systems. Elsewhere in the world, however, the requirement to measure the accuracy of predictions is much better defined. The UK market can learn from this experience and assist both suppliers and purchasers through improved clarity of purpose.

2.6 This document sets out to explain the core fundamentals of predictions and how their accuracy and quality may be measured.

# 3 What is a prediction and how are they created?

## 3.1 What is a prediction?

3.1.1 High quality passenger information is fundamentally dependent on high quality data; and accurate predictions are created from such data.

3.1.2 The aim of an RTI system is to describe the actual state of the bus or tram network to the passenger - rather than the planned (scheduled) *timetable*. For the operator, it is more about comparing the actual state to the planned *operation* (which includes the vehicle specific aspects). They are subtly different but use the same processes.

3.1.3 RTI is generated from data collected about the position of a vehicle on the road network at a given time, which can be compared to its *planned* position at that time; or used to generate predictions about where the bus will be later in its journey. The two data sets are combined in a process known as journey matching (see section 3.2 Journey matching). The RTI system then uses a prediction algorithm in order to create an inference on where the vehicle will be at a point in the future; and this data is published through standardised interfaces such as SIRI and GTFS.

## 3.2 Journey matching

3.2.1 Journey Matching is the process undertaken by the prediction engine to associate vehicle location data - normally provided as SIRI deliveries, with a journey in the scheduled data.

3.2.2 There are a number of linking dependencies that are required to ensure Journey Matching can occur and these are often subject to operational issues. The most common dependencies are:

3.2.3 **Out of date scheduling data** – if the departure times or ongoing schedules have changed, the prediction engine *needs* to know this. If it doesn't it will be attempting to match updates against journeys that have changed timings or even might not exist anymore. Indeed, a change in the scheduling data by even as little as a minute can throw a prediction out - pushing expected departure times out accordingly.

3.2.4 **No contact with the vehicle** – this is most commonly caused by 3G/4G blackspots, a faulty AVL system on the vehicle, or the driver not logging on to the AVL (ticket machine) in the correct manner.

3.2.5 You will note that the root of both of these issues can be found in the data creation (i.e. the beginning of the data flow) for schedule and AVL data. This makes the accuracy of this data of paramount importance as, if it is incorrect at the beginning of the chain, it will be incorrect throughout.

## 3.3 Creating a prediction.

3.3.1 The key focus of an RTI system is to create accurate predictions and provide them to your various endpoints: to reliably inform the travelling public about departure times from their chosen stop. The predictions created are exactly that: predictions; and as such, they will rarely be 100% accurate. However, the greater the sophistication of a prediction engine, the better the predictions will be.

3.3.2 The creation of a prediction can possibly be best described through the use of diagrams. A vehicle will progress along its route, aiming to maintain the times contained within the static scheduling data that has been set by the operator. Without any external influences, the timetabled arrival at each bus stop would reflect the static schedule.

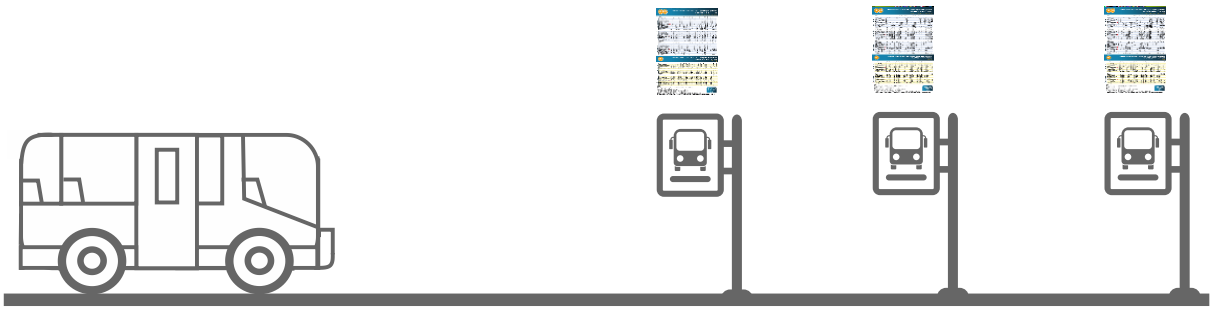


*Figure 3 Bus progressing along a road reflecting the timetable*

3.3.3 The progression along the route may be affected by a number of different external influences including:

- traffic congestion;
- the number of passengers boarding and alighting;
- roadworks and traffic incidents; and
- the prevailing weather.

3.3.4 The vehicle will report its location regularly to a prediction engine, which will calculate the alteration to the time at which the vehicle is expected to depart each stop. This prediction is then provided to the various endpoints employed within an information estate - where it will overlay the scheduled time with the actual predicted time of departure.

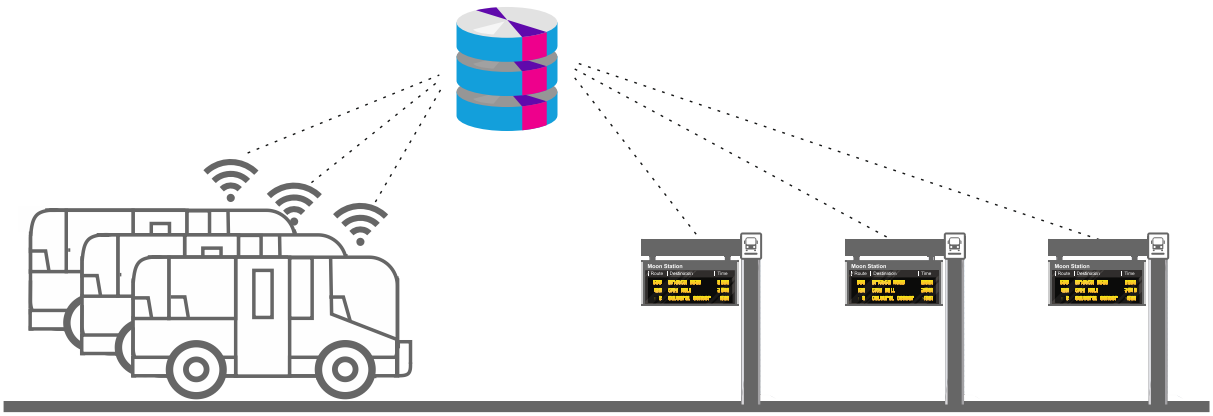


*Figure 4 Bus progressing along a road providing location data to enable predicted time to be created*

3.3.5 There are a number of factors that can improve a prediction but, at its most base level, a prediction can be best summarised as:

$$\frac{\text{Distance to next or upcoming scheduled points}}{\text{Predicted average speed between scheduled points}} = \text{Predicted time to scheduled point}$$

3.3.6 That is, the distance it has to cover in order to reach its destination - divided by the speed at which it is travelling, will equal the predicted time of arrival at that destination.

3.3.7 Obviously, this is incredibly crude and there are a number of factors that can improve a prediction. Indeed, most modern prediction engines take far more data into account when calculating a prediction. Dependent upon the RTI provider, this may or may not include:

- **Schedule periods between stops** – how long is the planned journey time between stop points?
- **Historical journeys** – how long has it *actually* taken in the past to cover the distance between stops?

- **Historical time and day journeys** – are the journey times influenced by the time of day – i.e. can it be reliably predicted that travelling into a city centre at 8am on a Monday morning will take longer than at 11pm on a Sunday evening?
- **Current traffic conditions** – does the prediction engine know anything of current traffic build up?

## 3.4 When does a journey start generating a prediction?

3.4.1 The timing of the start of prediction generation can vary depending on who is providing the feed and the sophistication of their prediction engine. Some providers will not provide a prediction until a particular service starts, whereas others ingest “running board” (service order of the vehicle) data and are able to provide what is known as a “cross journey prediction”. This is because they know what order a vehicle is running its services, when the driver is due to take a break (layover); and when there is a pause in service for driver relief. However, even if running boards are consumed, the first journey of a board will not be able to generate a prediction until the first journey itself has started. This can be overcome by the inclusion of what is known as dead run data, which provides the data on how the bus will get to the first stop of the running board - i.e. its initial journey from the depot or garage.

3.4.2 Should the RTI provider not give a cross journey prediction or use dead run information, there will not be a prediction at origination stops (such as bus stations and interchanges), and possibly for a number of following stops - dependent upon the distance/journey time between them.

## 3.5 Why don’t I get a prediction?

3.5.1 Once you have accurate scheduled data and a vehicle location, an RTI system will be able to create a prediction. If all the data is complete and accurate and all equipment is working, then every journey a vehicle operates will be able to be predicted. Notwithstanding, because of the inherent complexity and challenges, it is rare that 100% of journeys will be predicted.

3.5.2 There are many and varied reasons why predictions are not created for some journeys - these include:

- **Vehicles departing early** – if a vehicle arrives prior to its scheduled time it should, strictly speaking, hold at the stop until its scheduled departure time. Prediction engines are unable to predict such early departures. This is a particular problem at timing points - where it has to be assumed that a bus will not depart early;
- **Lack of connectivity** – all predictions rely on the original AVL data from the vehicle, so if there are 3G/4G/GPS black spots that prevent the vehicle from communicating back its position, reliable predictions cannot be generated. This can be particularly prevalent in rural and coastal areas (which may not have mast overlap); and high-rise parts of built-up cities that suffer from urban canyons. A total lack of connectivity will prevent the vehicle being tracked so that predictions cannot be generated; whereas intermittent connectivity may cause the vehicle to flip-flop between being tracked and untracked so that predictions are only created intermittently;
- **The failure of on-bus equipment** - for example, if the AVL components of a ticket machine are faulty, this will not always be noticed quickly if tickets are still being produced;
- **Drivers not logging in to their ETM/AVL system** – if the system that provides AVL data is not correctly initiated, it will not know what service it is operating, nor its running board, and so it will be unable to generate SIRI VM;
- **Broken SIRI feeds that are not sending data or the correct data** - this can happen between suppliers, and between suppliers and data consumers. All parties should work together to make these systems better at self-healing.

## 3.6 Ghost Buses

3.6.1 A ‘Ghost Bus’ occurs when a service is shown as predicting on on-line and/or electronic outputs (PIDS, on-line, SMS) but the bus (apparently) never arrives at the stop. These occur infrequently in most systems but can be particularly frustrating to a customer who does have some understanding of the difference between timetabled and predicted times (see the later section on presenting information to the customer).

3.6.2 There are many reasons for ‘Ghost Buses’ and the most common causes are discussed in this section.

3.6.3 **Intermittently Tracked or Untracked Vehicles** - Predicted or Scheduled time shown will disappear when the schedule time passes; meanwhile the bus might:

- have been cancelled physically but not electronically;

- have broken-down but has not been cancelled electronically;
- have been involved in an incident but has not been cancelled electronically;
- have been diverted from route at short notice;
- have gone off-route for longer and further than the prediction system can handle;
- become untracked and running late but will still call; or
- have faulty hardware leading to no data updates, no GPS, stuck GPS, or no ETM feed.

3.6.4 In addition, the following will affect all Running Boards in the Block:

- radio / GPRS black hole gaps in coverage lead to jumps in RTPI;
- more than one ticket machine is signed in with the same Journey details.

3.6.5 **Buses not assigned to journeys** – for a prediction to be generated, the data sent by the ticket machine to the real time system must match some or all of the expected data required for it to be assigned to a journey:

- the driver enters the wrong information into the ticket machine, so the bus is assigned to the *wrong* or *no* trip - resulting in no predictions;
- poor quality logons - the driver signs on to the ticket machine too early or late or when too far away from the start point of the journey. The bus will not then be successfully assigned to a trip;
- driving off-route because of a diversion due to roadworks or an incident -, causing the bus to be too far away from its expected stops along the route - resulting in predictions being removed from bus stop displays and outputs. If the bus subsequently returns to the route, a system will often identify this and start to create predictions again; or
- a planned journey not operating because of the weather, or operational reasons, for example, staff absence or no vehicle available for the journey.

3.6.6 **Cross-Journey Prediction Disjuncture** - due to a driver and/or vehicle on an incoming service scheduled to perform a follow-on journey, being unable to do so. This can be due to:

- the driver of the follow-on service being unavailable due to illness or roster issues;
- vehicle failure;
- the vehicle being reallocated to another (deemed as more important) service;

- the incoming service being so late that the follow-on service overlaps with a subsequent run of that follow-on service; or
- the incoming service being so late that it is truncated early (turned back).

3.6.7 **Customer misunderstanding** – disruption to high frequency services can lead to confusion as the prediction system cannot represent: 'bus bunching' effectively; service-run overlaps; and buses running the same service overtaking one-another (predictions on a display do not uniquely identify a bus journey - only the service).

3.6.8 **Overdue Cleardown** - 'Due' left up too long after the bus has departed because of:

- latency;
- incorrect coordinates for a stop;
- slow traffic;
- incorrect/incomplete dataset for journey times/line-of-route; or
- infrequent prediction update.

3.6.9 **Congestion/statutory layover affecting predictions** – variable traffic speed can 'freeze' predictions or even drive a prediction back up; or the requirement for a service to 'layover' may also affect predictions. Either of these occurrences can cause confusion to the customer.

3.6.10 **Service display order 'artefacts'** - as displays cycle through the next 'n' services, the displayed order of departure can change due to:

- latency differences between service providers;
- physically changed order of arrival (buses overtaking one another); or
- algorithmic and programming specifications.

3.6.11 **Faulty data sets** – incomplete and inaccurate data may also lead to misleading information being displayed. Examples include:

- configuration – the bus doesn't actually call at that stop;
- days of operation / holidays;
- diversions – often too small to be considered;
- TXC schema validation / missing elements;
- rejected data - speeding buses, times running backwards; and
- complex data – staff with the skills to ensure that the data is correct are not always available.

3.6.12 **Other faults** – sometimes other schedule/prediction providers (e.g. Google) will show a service is scheduled when it's not because the data is stale (i.e. it has taken too long to get to the receiving system to be regarded as accurate).

3.6.13 Whilst there is no easy solution to these undoubtedly frustrating situations, it is good practice to regularly review why journeys have not been predicted: to identify where there are repeating faults or errors that can be addressed.

# 4 Understanding Accuracy and Quality

## 4.1 How accurate is your RTI?

4.1.1 Whilst there is no single answer to this question, we can nonetheless, agree on a set of metrics to enable us to measure accuracy and quality; but to do so, we need to use a common framework to define those metrics.

4.1.2 A definition of Quality from ISO9000:2000 is “The degree to which a set of inherent characteristics fulfils requirements”

4.1.3 However, accuracy is often more subjective and measured accuracy can be very different to user-perceived accuracy. For example, a large error when the vehicle is 30 minutes away may be significant contractually and therefore deemed to be inaccurate, but it may not materially affect a user for whose purpose it is accurate enough.

4.1.4 The fundamental requirement of *quality* data is that it satisfies the requirements of its intended use. It is possible for data to be deemed as of poor quality for one purpose but of high quality for another. Data quality depends as much on the intended use as it does on the data itself. To satisfy the intended use, the data must be accurate, timely, relevant, complete, understood, and trusted. For example, a vehicle’s location update that took 12 hours to be made available would be timely enough to be classed as good quality if it were being used for historical reports; but if that data were needed to create a prediction for a current journey it would not be good quality, as it is not timely.

4.1.5 It is common for the words data and information to be used interchangeably but they are, however, fundamentally different, and for the purposes of measuring and reporting, they need to be used correctly. In this report we will use these definitions:

- Data is raw facts: numbers, words, dates, images, sounds etc. without context. An example within a real time system would be the *number* of journeys where a prediction was generated.
- Information is data that is put into context e.g. in a sentence or associated with field names/headings, for example, the *proportion* of journeys where a prediction was generated. This places the data about the number of journeys into the context of the total number of journeys.

4.1.6 In this report, we will identify suggested metrics for both data and information about RTI from the perspective of a real time system supplier or owner who needs to understand how a system is performing; and of the customers who use the information to make decisions about their travel.

## 4.2 Measuring quality

4.2.1 Various standards and frameworks exist for the measurement and communication of data quality. One that is particularly relevant to the work described here is *ISO19157: Geographic information — Data quality*, as specified by the International Organisation for Standards (https://www.iso.org/standard/32575.html). This provides a structured approach to describing data quality. Whilst its focus is on geographic data, it is nonetheless, applicable to all data types, and it has found to be very well suited to the transport domain.

4.2.2 The standard differentiates between distinct *dimensions* of data quality, including:

- completeness: is data missing, or erroneously present?
- positional accuracy: related to where things are;
- attribute accuracy: erroneous values associated with entities - the standard covers both quantitative and qualitative errors;
- temporal quality: how accurate temporal attributes and relationships are; and
- logical consistency: are logical rules adhered to?

4.2.3 It is temporal quality that is most relevant to Bus RTI data.

4.2.4 Data quality measurement often requires a comparison of a dataset against what is known as the ground truth - a directly measured data set. When measuring the accuracy of a prediction, the gold standard approach is to collect the data as though you were a customer - standing at a bus stop. Gathering ground truth in this way by physically visiting real world locations can be useful, but it is costly and does not scale well. An alternative source of ground truth can be a comparison of one test data set against an alternative version - to identify deviations between the two. In addition, a dataset can be compared with itself at a different time. In the context of bus RTI predictions, this provides a reliable and scalable approach for comparing predicted RTI against ground truth. One such observation to be tested is the prediction of when a bus would arrive at a stop - calculated at one point in time (the prediction at T1). The associated ground truth value is the subsequent point in time that the system observed the bus to reach that stop (T2).

4.2.5 ISO19157 defines broad measure types that can be applied within each domain. A measure type can either relate to errors (quantifying the problems with a dataset), or correctness (quantifying how accurate it is). In either case, measure types can be simple indicators (e.g. a specific error exists – true/false), counts (e.g. how many examples of a particular error exist) or rates (e.g. what percentage of the data set does not exhibit that particular error).

4.2.6 The working group has focused on being able to provide an indicator (true/false) for any RTI prediction that flags whether the difference between that prediction and its ground truth value was within an acceptable error threshold, or not. Rather than the more common practice of using a single threshold value (e.g. 2 minutes for all predictions), we have chosen a variable threshold - increasing for buses that are further away in time. Aggregating these indicators allows counts of the errors within a given threshold to be calculated. When compared to the size of the entire dataset, this provides a metric based upon the rate of errors within a given threshold. This can report the percentage of predictions that were considered *acceptable*.

## 4.3 Data Collection

4.3.1 The data needed to calculate any measure of quality can be collected in two ways – either by automated or manual processes.

4.3.2 The lowest cost and more scalable approach is to collect the data automatically from servers or data feeds using SIRI or GTFS-RT. This approach enables large volumes of data to be collected quickly, easily, and accurately; and is the most practical methodology for regular testing of prediction quality and accuracy.

4.3.3 Whilst manually collecting the data (as though you were a customer standing at a bus stop) is the gold standard and the best ground truth available, it does present significant challenges. This is particularly so in well serviced areas - with departures and passing vehicles every few minutes. It requires close attention to detail to ensure that the data is recorded accurately and cannot be sustained for more than perhaps 30 minutes without a break.

4.3.4 The process is time consuming, potentially cold and wet; and costly to carry out on any scale over a reasonable timescale and so is not often undertaken. It is, nonetheless, a highly worthwhile exercise to undertake because it replicates how customers actually experience the data. It will help highlight any delays and latency in the system that data collection from servers will not have; as well as helping to bring to light the causes of poor predictions: such as traffic lights just before a bus stop or poor cleardown performance. Such causes are not discernible when looking at aggregated large data sets.

4.3.5 Where there is a multiplicity of different digital outputs, it is possible to carry out manual data collection in an office environment. Whilst this will not capture the actual customer experience, it will highlight differences between outputs - which in of itself, is useful to understand from a customer perspective (which do I 'trust' the most?).

4.3.6 A hybrid approach can be adopted using streamed video cameras which, from the comfort of an office - allow easier comparison with multiple digital outputs.

4.3.7 Unless data is being collected at stop, it is likely that the "True" values may not be discernible but detecting deviation between two versions of events is, nonetheless, useful.

## 4.4 Sample Size

4.4.1 The size of data sample collected needs to be appropriate for the purpose to which the data will be put. The minimum appropriate sample size could be as low as a single stope event: 1 departure for 1 bus. However, it is likely that a much larger sample will often be required - especially if data is to be aggregated to provide measures reflecting the performance of multiple operators or services; or the overall performance of a supplier.

4.4.2 If a small data set is used - either because limited data is available or only a small proportion of the *total* data set has been sampled, care needs to be given to the interpretation of the data; and to the validity of any conclusions reached and resultant actions taken.

4.4.3 The location of captured data needs to be recorded as meta data for later review, and to ensure appropriate comparisons. For example, there are likely to be fewer predictions available for a route's origin when compared to bus stops nearer its destination.

4.4.4 The collection of data from feeds sent to signs where the signs have intelligence, can result in low volumes of predictions captured. This is because following an initial data update, the sign will only be sent data if the server determines a significant enough change to a vehicle's flight path (see 5.1 Prediction Accuracy). In this case, the use of another data feed or manual collection may be appropriate.

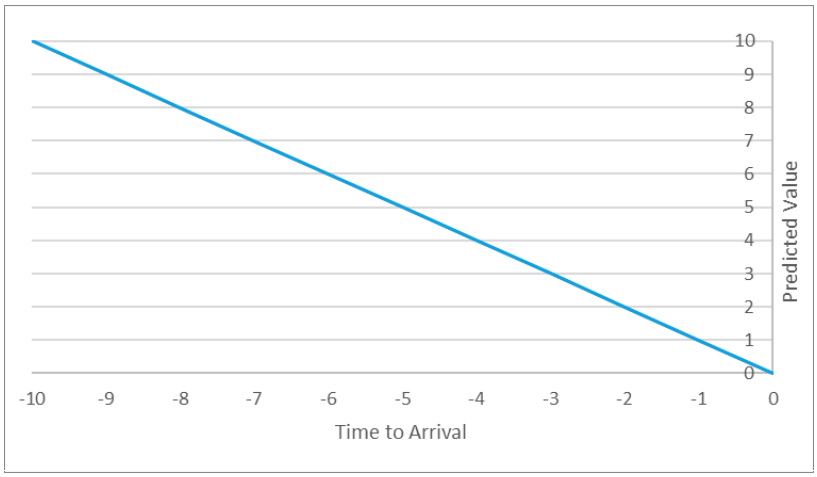
# 5 Measuring Prediction Quality & accuracy

## 5.1 Prediction Accuracy

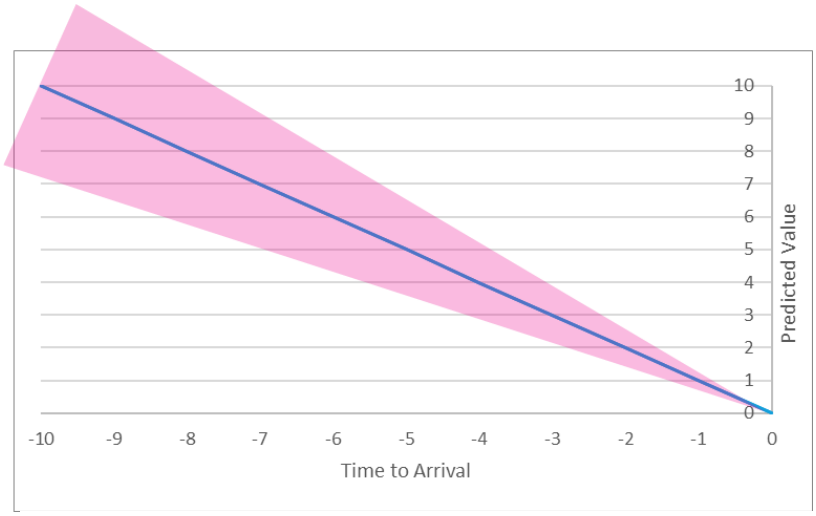
5.1.1 Prediction accuracy can be affected by several factors. Some will be within the control of a bus operator, some within the control of an RTI provider, and some neither. However, the most common causes of poor prediction accuracy will be poor data creation and poor data management.

5.1.2 For example, a prediction can only be generated when there is vehicle-centric data available that can be compared to schedule data. If a vehicle is not polling regularly - providing regular updates on its position, there is little that can be inferred about the vehicle's progression along a route and the prediction will be loose. Similarly, should static schedule data not be kept sufficiently up to date, there will be little chance of matching AVL data against service schedules - thereby rendering predictions inaccurate or, at worst, not possible.

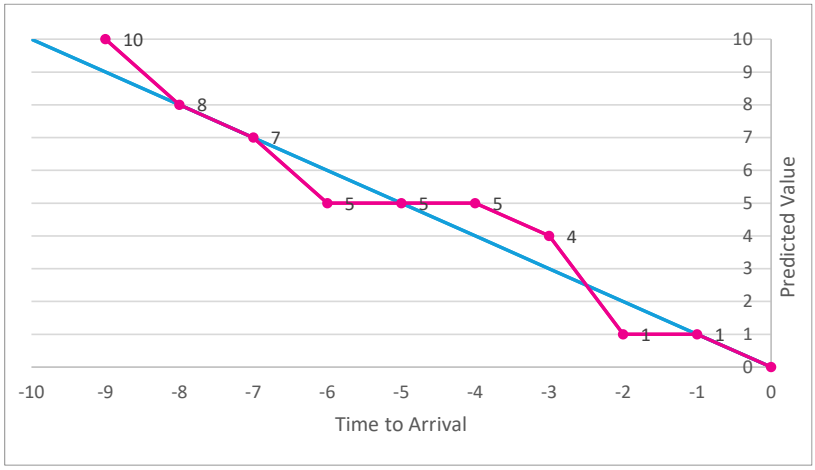
5.1.3 As a bus gets nearer to where the traveller wishes to join the service, the greater requirement there is for predictions to be accurate - with predicted time of departure (expressed in minutes) from a stop liable to go up as well as down: dependent upon the sophistication of the RTI system that has been implemented. It is, however, relatively safe to say that predictions will increase in accuracy the closer the service gets to your selected stops. This is because even though most prediction engines take historical and current road conditions into account, the closer the bus is to the stop - the less variables with the potential for journey disruption, there will be.



5.1.4 Predictions will generally get more accurate, the closer a service gets to the point for which the predicted departure is set. The further a vehicle is from that point, the greater scope there is for unforeseen variables to interrupt the vehicle journey. A flexible prediction engine will allow for this and allow a prediction to alter - dependent upon current road conditions.




5.1.5 In real life, the flight path of a prediction will never be smooth though:



5.1.6 Having reviewed a number of different approaches to measuring the quality of predictions, the working group identified time buckets as the most appropriate to use. An explanation of time buckets is set out in Section 6 below.

## 5.2 Other Measures

5.2.1 In addition to the use of time buckets as a measure for the quality of predictions, other useful measures to consider include:

- The proportion of journeys predicted: what proportion of journeys scheduled to run had a prediction generated for some or all of the journey? This is normally measured at a range of levels from system and operator level down to service and vehicle level - to help identify where there are problems.
- The percentage of SIRI VM messages that can be cross-referenced with schedule data.
- At a selected bus stop: what proportion of departures that were expected to have a prediction did not have one but were within 15 minutes of arrival at the bus stop.

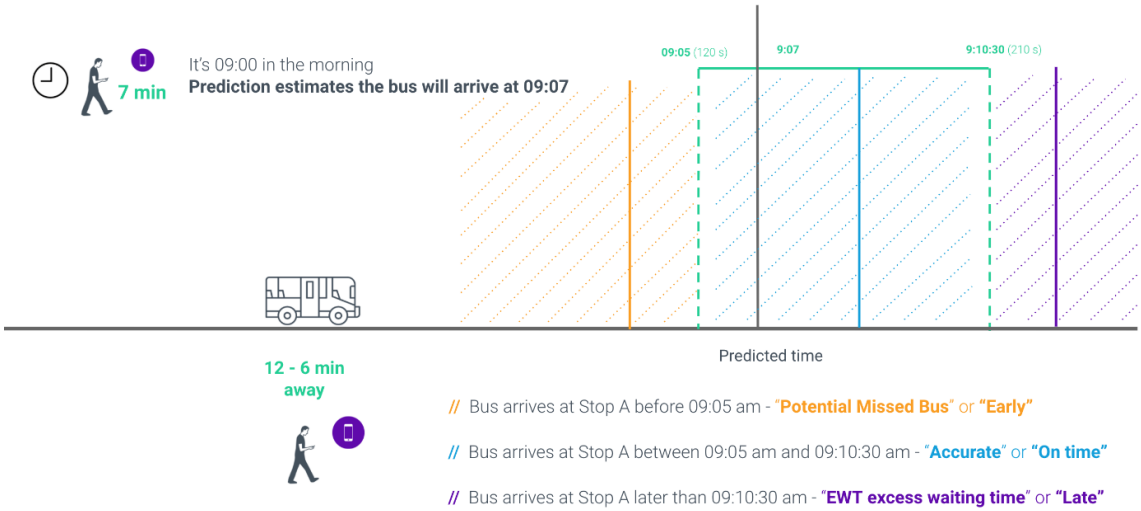
# 6 Time Buckets



6.1 It is very hard to create a perfect prediction that accurately tracks a vehicle on its “flight path” towards a bus stop; but the closer the vehicle is to a bus stop - the greater the ‘accuracy’ a customer expects. Because of this, when we measure predictions, we want to understand how accurate they are at different time scales. The definition of ‘accurate’ can then change depending on how far away a bus is.

6.2 To achieve this, an approach called time buckets is used. We can explain time buckets from the perspective of a customer wanting to catch a bus from a particular stop who uses a phone app to find out when the bus is going to arrive:


6.3 If we accept that a prediction is allowed a level of tolerance, or inaccuracy, of 120 seconds to +210 seconds from the 7-minute prediction.

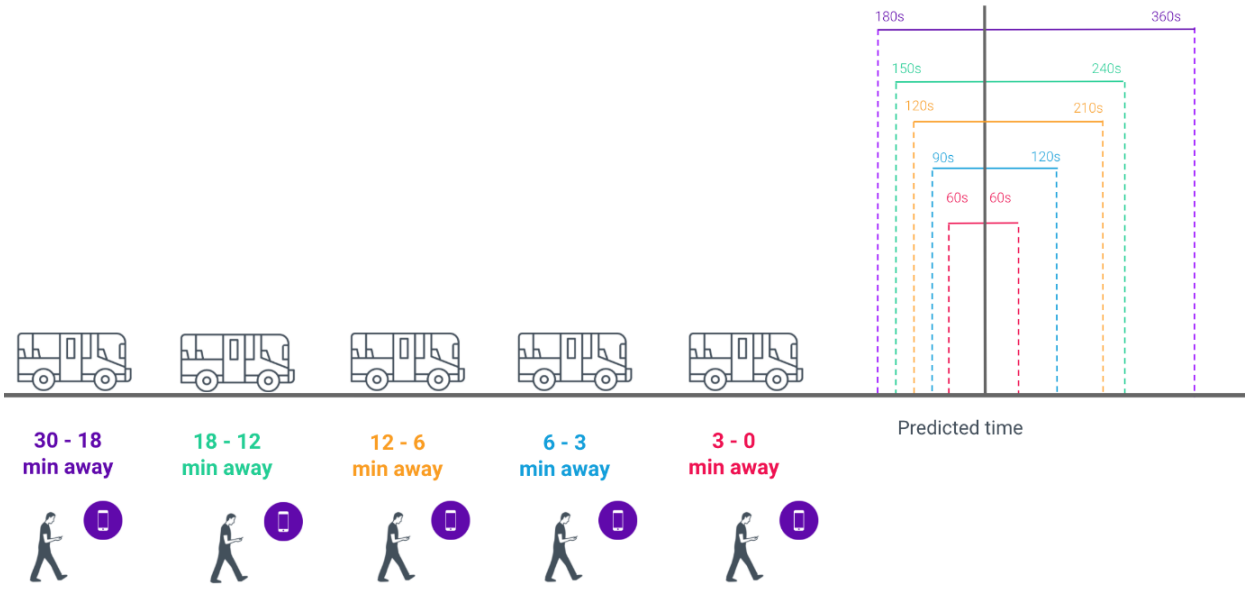


6.4 This example has described a single window of time – 12 to 6 minutes away, in which we measure the accuracy of a prediction.

6.5 We can use this approach and create multiple 'windows in time' or time buckets, to look at prediction accuracy. This allows each time bucket to have a different range for when we consider a prediction to be accurate. For example, if a bus is eleven minutes away and the prediction is inaccurate by one minute, that is less significant than if a bus is one minute away and the prediction says two minutes - so still inaccurate by one minute.

6.6 This approach allows for the relative accuracy to decrease as a bus approaches the stop, and this is the key reason why time buckets are used instead of a much simpler percentage error. For example, a vehicle that is 100% later than a 1-minute prediction is likely to be less significant to a customer than one that is 100% later than a 30-minute prediction.

6.7 Taking the previous single bucket approach, we can expand this:




6.8 The size of each bucket and the allowed margin of earliness and lateness - compared to the prediction, are flexible. However, the margins used need to be adjusted to reflect the operational or contractual requirements of the service. For example, it would be appropriate to have different bucket times for an urban high frequency service than for an hourly rural service.

6.9 The recommended starting point for defining your own time bucket and criteria is:

| BUCKET | ALLOWABLE EARLY | ALLOWABLE LATE |
|---|---|---|
| **0 – 3 minutes** | 60 seconds | + 60 seconds |
| **3 – 6 minutes** | -90 seconds | +120 seconds |
| **6 – 12 minutes** | -120 seconds | +210 seconds |
| **12 – 18 minutes** | -150 seconds | +240 seconds |
| **18 – 30 minutes** | -180 seconds | +360 seconds |

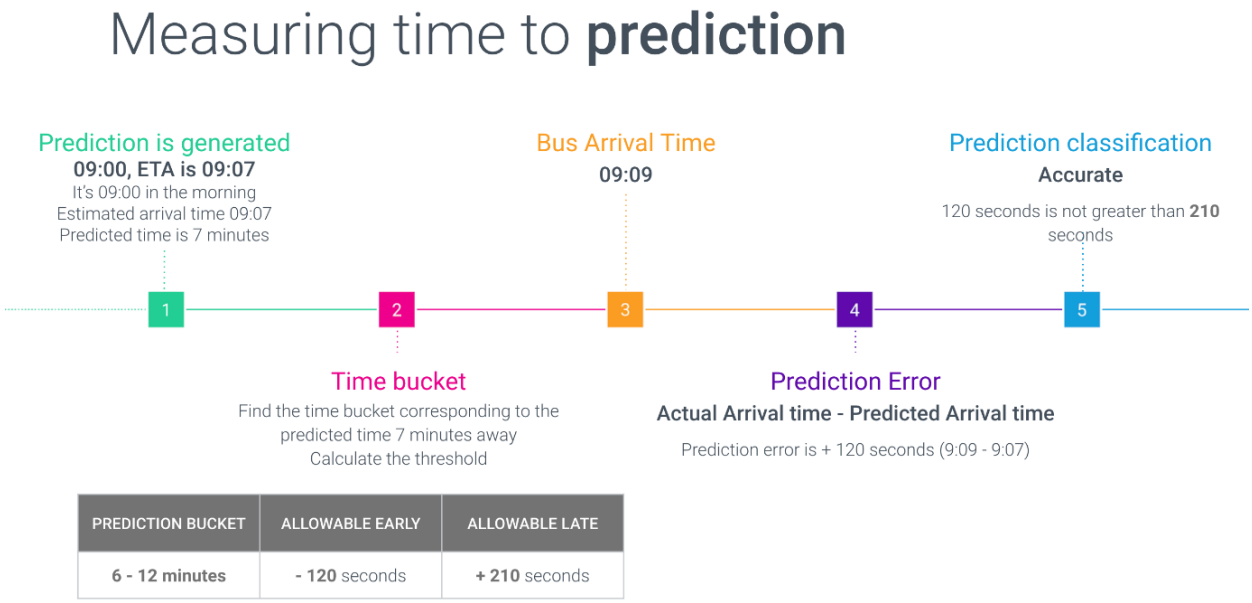
# 7 Calculating accuracy

7.1 There are two methodologies for measuring prediction accuracy. Firstly, it can be measured by comparing the predicted arrival time to the actual prediction; or secondly - the actual arrival time.

7.2 The sequence of events, continuing the example is:

- At 09:00 a prediction is generated.
- The estimated / predicted time for the bus at the bus stop is 09:07.
- The bus actually arrives at the stop at 09:09.

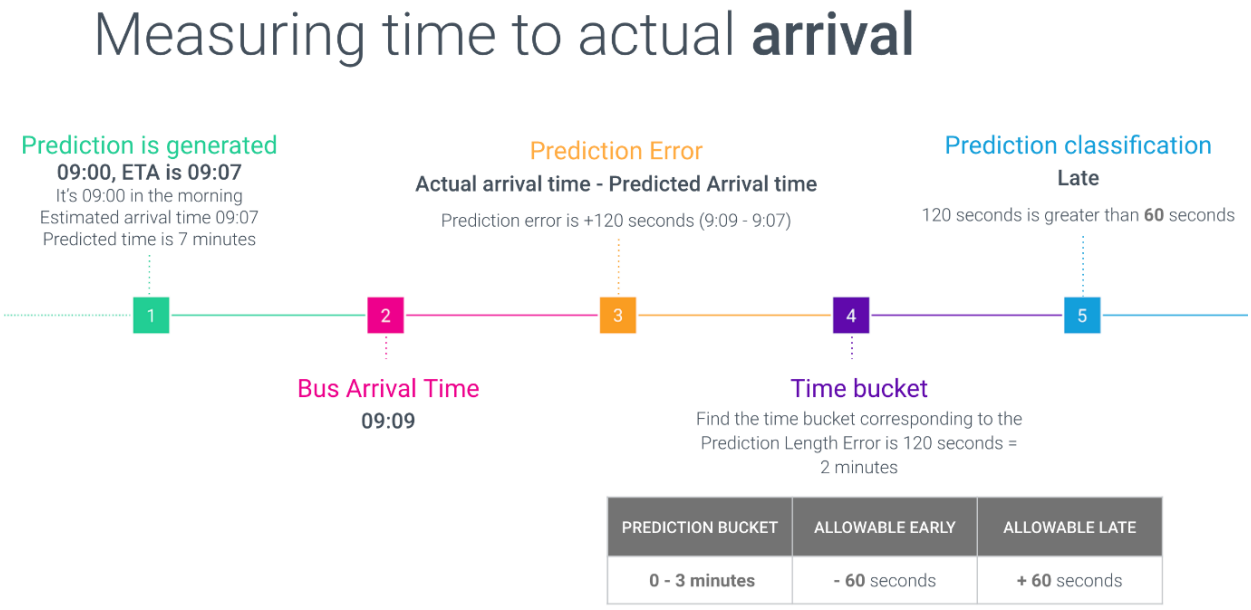
7.3 Measuring the time to prediction results in:




| PREDICTION BUCKET | ALLOWABLE EARLY | ALLOWABLE LATE |
|---|---|---|
| 6 - 12 minutes | - 120 seconds | + 210 seconds |

7.4 This method is the most popular as it measures the accuracy of the information that the passenger sees, and so provides information about the relative accuracy of what will be shown on a display or app.

7.5 Measuring the time to actual arrival results in:




7.6 This method evaluates the difference between the actual arrival time of the vehicle compared to the predicted time. This does not measure the accuracy relative to a passenger's perception but purely to prediction error.

7.7 The working group recommends that the time to prediction approach is used as this most closely reflects the customer experience.

# 8 Reporting prediction quality

8.1 Reporting of the quality of predictions can be done using raw data, the numbers of early and late predictions (helpful for those tasked with tackling problems in detail), or as a percentage - to provide some context.

8.2 The data should be analysed and reported at a level appropriate for the audience. For example, an operator's executive may be interested in the overall prediction accuracy to give an idea of the system performance as a whole; but the operations manager is likely to be more interested in each time bucket: to provide more information to target remedial work.

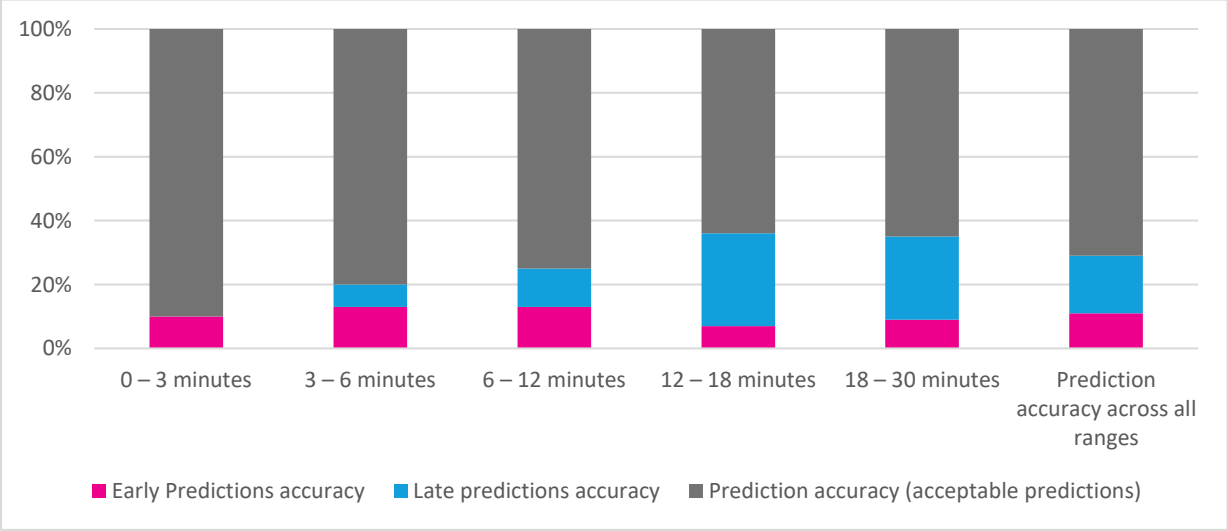
8.3 Example report showing the number of predictions that have been within each time bucket

| RANGES | 0 – 3 MINUTES | 3 – 6 MINUTES | 6 – 12 MINUTES | 12 – 18 MINUTES | 18 – 30 MINUTES | NUMBER OF PREDICTIONS |
|---|---|---|---|---|---|---|
| **Prediction Bucket** | -60 to +60 seconds | -90 to +120 seconds | -120 to + 210 seconds | -150 to 240 seconds | -180 to +360 seconds | |
| **Early Predictions** | 1 | 2 | 10 | 3 | 4 | 20 |
| **Late predictions** | 0 | 1 | 9 | 12 | 11 | 33 |
| **Total predictions** | 10 | 15 | 75 | 42 | 43 | 185 |
| **Acceptable predictions** | 9 | 12 | 56 | 27 | 28 | 132 |

8.4 Example report showing the percentage of predictions that have been within the time bucket:

| RANGES | 0 – 3 MINUTES | 3 – 6 MINUTES | 6 – 12 MINUTES | 12 – 18 MINUTES | 18 – 30 MINUTES | PREDICTION ACCURACY ACROSS ALL RANGES |
|---|---|---|---|---|---|---|
| **Prediction Bucket** | -60 to +60 seconds | -90 to +120 seconds | -120 to + 210 seconds | -150 to 240 seconds | -180 to +360 seconds | |
| **Early Predictions accuracy** | 10% | 13% | 13% | 7% | 9% | 11% |
| **Late predictions accuracy** | 0% | 7% | 12% | 29% | 26% | 18% |
| **Prediction accuracy (acceptable predictions)** | 90% | 80% | 75% | 64% | 65% | 71% |

8.5 Example graphics presentation showing the percentage of predictions that have been within the time bucket:



8.6 Every real time system and contract will have different technical and contractual requirements and constraints. Consequently, the accuracy requirement for predictions is likely to change between geographic areas or real time systems. It may be appropriate to set different expectations for accuracy for different time buckets - due to the increased customer sensitivity to inaccuracy as the predicted arrival time gets shorter. For example, a higher quality threshold could be expected for a 0-3 minute bucket than for an 18 – 30 minute bucket. Likewise, from a customer perspective, expecting fewer early predictions may be appropriate to ensure reduced likelihood of arriving at a stop after the bus has left.

8.7 Once a quality measure has been agreed, it should be used consistently to allow re-measurement to facilitate comparison over time: between suppliers or different versions of software.

8.8 A suggested starting point for measuring the accuracy of a prediction is:

- 90% and above is excellent;
- 80 – 90 % is good;
- 70 – 80% is satisfactory ; and
- 70% and below is unsatisfactory and needs improvement.

# 9 Improving Prediction Quality

9.1 Once attention is being paid to the quality of predictions in a real time system, the logical next question to ask is: how can I improve the quality?

9.2 Whilst some of the potential causes and solutions are covered earlier in the report, how easily these are identified is often down to good presentation of data. The greatest gains in quality can be made by tackling the issues that are causing the biggest problems. In a real time system of any scale, there is a large volume of data that needs to be reviewed; and to assist in such a review, it is advisable to use graphical analysis tools.

9.3 Many real time systems have reporting and analysis tools that can assist quality measurement and assurance by highlighting the most frequent errors. Indeed, the use of dashboards and graphical presentation to highlight where there are problems can be very effective.

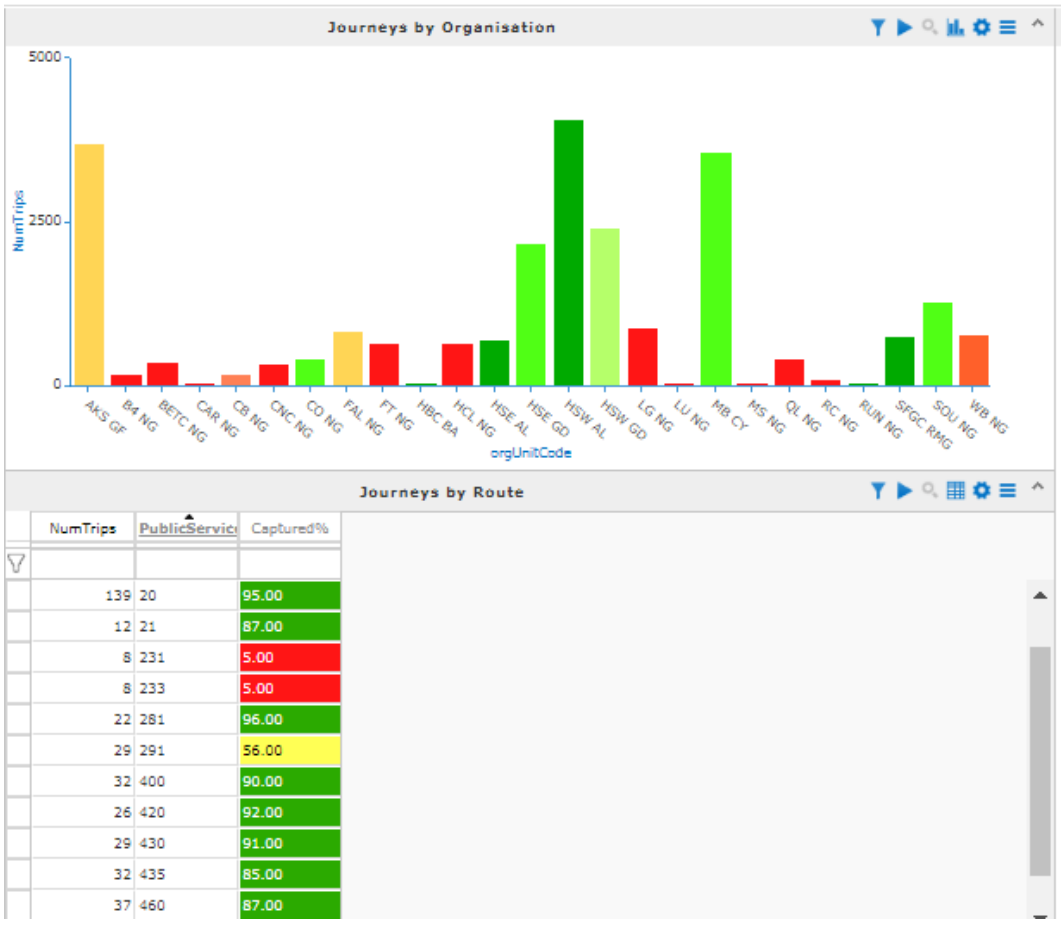



*Figure 8 A Trapeze system dashboard view showing number of trips by operator and then the % that of trips that were tracked*

# 10 Presenting predictions to customers

## 10.1 Countdown or Expected?

10.1.1 The ‘traditional’ approach to displaying bus RTI has been to show a 'countdown' to represent the predicted time at which a journey that was being tracked would depart the stop. Typically, this would be something along the lines of 5 mins, 4 mins, 3 mins, 2 mins then 'Due'.

10.1.2 There are, however, other approaches to presenting information to the customer. These are discussed in the document: ‘RTIGT037 1.0 Displaying Transport Information on Public Electronic Signs’.

10.1.3 Of particular note to this report is the ‘railway’ style presentation which includes an additional column to present the expected arrival time.

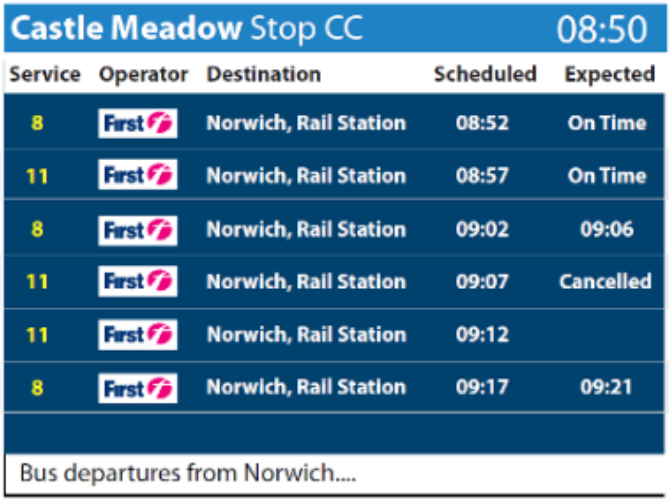



*Figure 9 An example display from Norwich showing Scheduled and Expected for bus services*

10.1.4 This style of presentation can make it easier for customers to see how accurate a predicted arrival time is - in that it is much easier to see when the service is expected and compare this against the clock – than to calculate what time it will be in 3 minutes.

10.1.5 Whatever approach to presenting predictions is chosen, it is important that the language and terminology used is easily understood by the customer. Consistent use of language is particularly important where customers travel between areas and real time systems. To facilitate consistency, RTIG has produced a document on language: ‘*RTIGT035-1.0 Language and terminology in Real Time Information systems*’. It is recommended that the terms and definitions described in this document are used whenever possible to aid customer understanding.

## 10.2 Arrival or Departure?

10.2.1 Typically, a customer asks one of two questions of predictions: ‘when does my bus leave?’ or ‘when does my bus arrive?’ So this is another consideration to take into account when deciding on the terminology to be used. It is the departure time that is displayed on a bus timetable and this is often how customer outputs from RTI systems are described. In reality, unless there is a layover or it is a timing point, the real time system is more often providing a prediction for the arrival time of the bus.

10.2.2 This is not a problem for the customer as the arrival and departure times are expressed in minutes. However, when looking at prediction accuracy, the data being analysed is in seconds and the difference between arrival and departure times may make a difference. SIRI data can contain both arrival and departure times for each stop. There is no right or wrong in this, but it is important the data utilised in calculations is clear and used consistently.

## 10.3 Use of Due.

10.3.1 As a bus approaches a stop the countdown time is often changed to say ‘Due’. The general use case is when a bus is predicted to arrive at a stop – which may vary locally but intending to mean: ‘too short a time for a minutes-countdown to be meaningful’ or ‘the bus is too close for the customer to consider wandering away’. The more frequent the location updates from a bus - the greater the likelihood of being able to use a shorter Due time. This short period is the most sensitive to disruption – through, for example, traffic lights or congestion; but it is also the time at which customers are most sensitive to the actual arrival of a bus.

10.3.2 The point of change depends on the real time system configuration and channel - varying from 90 seconds down to 30 seconds. This inconsistency can lead to customer confusion and misunderstanding – thereby reducing their confidence in the information. Therefore, we recommend that the ‘Due’ time should be standardised across customer channels in an area.