Australasian Transport Research Forum 2023 Proceedings
29 November – 1 December, Perth, Australia
Publication website: http://www.atrf.info

Impact of using hybrid data on travel time reliability reporting

Ronny Kutadinata[1], Kevin Wu[1], Kenneth Lewis[1], Frans Dekker[2], Merle Wood[2]

[1] Mobility Futures Group, National Transport Research Organisation / ARRB, Australia

[2] Infrastructure Management and Delivery, Queensland Department of Transport and Main Roads, Australia

Email for correspondence (presenting author): ronny.kutadinata@arrb.com.au

Abstract

This paper presents the results of a study undertaken to understand the impact of using hybrid data for travel time reliability reporting. The travel time reliability measure reflects the day-to-day variability of travel time experienced by motorists. It is computed based on the National Performance Indicator (NPI) methodology developed by Austroads. Route travel times calculated with three traffic data sources plus one hybrid data were compared: inductive loops (STREAMS traffic management system, the current system being used), Bluetooth, GPS probe (HERE) and the blended set of these three sources (referred as Hybrid in the study). When the values from these four data sources are tabulated, STREAMS data yield the lowest reliability value (indicating lowest variability of travel time). Hybrid data, with the advantage of improved data coverage, is the most comparable to STREAMS. Bluetooth and HERE data show a higher range of variability of travel time. It is hypothesised that Bluetooth and HERE can capture delays on arterial roads better under saturated traffic conditions. However, these emerging data sources on its own also have a limitation - the low market penetration rate results in highly variable travel times. In turn, the results also indicate that the network-level reliability reporting will be adversely impacted when the data source is changed from STREAMS to Bluetooth or HERE.

1. Introduction

While the use of hybrid travel time and volume data was proven to improve coverage (Johnston, 2021), the appropriate application of it should be evaluated, due to the potential variability it may introduce. One of these applications is the well-used travel time reliability measure. The measure calculates the day-to-day variation in travel time on motorway and arterial routes. Due to the importance of the measure, it is crucial that it accurately reflects the variability as experienced by road users, as opposed to the variability due to data blending process, which could use different combinations of data sources from one day to the next.

Previously, a hybrid dataset for the Gold Coast road network has been generated (Johnston, 2021). The hybrid data is made up from STREAMS National Performance Indicator (NPI) modelled layer, HERE and Bluetooth data. The technology in which the data is collected differs by the source, which are: STREAMS data is collected from inductive loop detectors installed at signalised intersections and motorways; HERE is a probe speed data provider, they collect

GPS location-based speed data from the fleet which either has HERE navigation map installed (as part of the onboard unit) or using their mobile phone app; Bluetooth data is collected from scanning Bluetooth devices and pairing the MAC (Media Access Control) address at sequential intersections. Note that both STREAMS and Bluetooth are stationary recording of passing traffic, whereas HERE is a vehicle tracking GPS data.

The aim of the hybrid dataset is to maximise the strengths of each dataset and minimise their weaknesses. However, it had been shown that different data sources have different characteristics (Saxena *et al.*, 2021) and data source selection affects the travel time reliability calculations (Haghani, Zhang and Hamedi, 2014).

The aim of this work is to determine if hybrid travel time data and other sources are comparable for calculation of route travel time reliability when compared with STREAMS data (the status quo). Specifically, this project is to:

1. Determine how the computed route travel time reliability differs when using various data sources (hybrid, STREAMS, HERE and Bluetooth)
2. Understand under which conditions the other data sources can comparably be used for route travel time reliability calculations in replacement of STREAMS.

This paper is structured as follows. Section 2 provides a review of the relevant literature on travel time reliability and the use of hybrid data for this purpose. Furthermore, Section 3 outlines the methodology used for the data processing and analysis, followed by the presentation of the results and discussions in Section 4. Finally, Section 5 concludes this paper.

2. Literature review

Travel time reliability has gained widespread attention in recent years. A guide on travel time reliability published by the Federal Highway Administration (FHWA) within the United States Department of Transportation defines travel time reliability as “the consistency or dependability in travel times, as measured from day-to-day and/or across different times of the day” (FHWA, 2017).

This literature review particularly focuses on the day-to-day variation of travel time for different times of the day, addressing the following key topics: best practice in measuring travel time reliability; and impacts of hybrid or other data sources on travel time reliability calculations.

2.1. Measuring travel time reliability

Although the term ‘travel time reliability’ was defined in similar fashion among various studies, several different ways quantifying travel time reliability have been proposed. The travel time reliability metric can typically be grouped into four major categories, namely:

- statistical range measures: travel time window, percent variation, standard deviation, reliability, semi-standard deviation, variability index, and width of travel time;
- buffer time measure: buffer time, buffer time index, planning time index, travel time index, 50^{th} or 80^{th} percentile travel time index;
- tardy trip indicators: Florida reliability method, on-time arrival indicator, misery index, modified misery index, congestion frequency, and reliability rating;
- other measures: congestion measure, skewness statistic.

Austroads (2020) suggested that the planning-time index was the best single reliability measure for urban conditions, while acknowledging different reliability measures may be appropriate for different audiences or analyses. FHWA (2017) recommended four common reliability measures: 90th or 95th percentile travel time, buffer index, planning time index, and frequency that congestion exceeded some expected threshold, while NZ Transport Agency (Denne *et al.*, 2013) and ATAP (2021) recommended the statistical ranged approach – coefficient of variation. FHWA (2017) discouraged using statistical measures such as standard deviation or coefficient of variation as travel time reliability performance indicator as they are not readily understood by non-technical audiences nor easily related to everyday commuting experiences. On the other hand, NZ Transport Agency (Denne *et al.*, 2013) argued that travel time reliability relates to the variability of travel time, or in statistical terms the variance of travel time (i.e. the square of the standard deviation).

Furthermore, it was recommended to use a time interval (i.e. the size for binning the travel time data) between 15 min to 1h for reliability evaluation, where shorter duration in this range is typically used for operational analyses and longer durations used for planning analyses (TRB, 2016).

2.2. Impacts of data sources on travel time reliability

Hybrid data integrates multiple data sources into one single hybrid form of data by selecting, filtering, and aggregating the data readings from multiple sources. These processes can be referred to as data fusion and data patching processes. The quality of hybrid dataset is highly source dependent (e.g. how data is aggregated). At this stage, there is no agreed approach for blending data sources to develop corridor or regional performance measures (Fontaine *et al.*, no date), although some guidelines exist in the literature (Kessler, Rempe and Bogenberger, 2021). However, it is noteworthy that data aggregation tends to smooth the data variation (Fontaine *et al.*, no date), which could cause the reliability estimates to lose accuracy.

For instance, a Perth case study (Saxena *et al.*, 2021) found that calibrating a travel time reliability model using hybrid data underestimates the reliability values (i.e. underestimates variability) as compared to using probe data. This can mean two things: (1) Probe data source can produce higher measured travel time reliability than the hybrid data source due to lower sample size, and (2) hybrid speed fusion and patching processes may potentially reduce the measured route travel time variability.

Haghani et al. (2014) assessed the impact of data sources, namely Bluetooth and INRIX (data fusion between probe data and traditional sensor data), on travel time reliability for two freeway routes (26 and 29 kms in length) using the 12 months data in 2012. They summarised that different data sources may provide different values for the same travel time reliability index, and the accuracy of measurements of reliability highly depends on the data collection technologies and the target application (e.g. segments with or without high-occupancy lanes). Another key finding of Haghani et. al (2014) was that among all the reliability performance measures, the standard deviation and performance variation measures are more sensitive to the data source.

3. Methodology

This section details the steps taken for the data processing and analysis to compute and compare the day-to-day travel time reliability of a route obtained from the various data sources considered in this study.

In this study, there are four data sources that are being considered, namely: STREAMS NPI data, Bluetooth data, HERE data, and hybrid data. The reader is referred to the original report (Johnston, 2021) for the detailed description of the hybrid data. The hybrid dataset consists of a blend of speed data of each link from each data source in 5-minute intervals. Additionally, there is volume data for both STREAMS and hybrid data (STREAMS volume is used for the hybrid), whereas the Bluetooth and HERE data has sample counts. Figure 1 shows a snippet of the hybrid dataset.

Figure 1: Excerpt of the hybrid dataset

	NPI_Link_ID	localDate	localTime	Speed	Volume	Length	NPI_Speed	NPI_Volume	Here_Mean_Speed	Here_Link_Count	BT_Speed	BT_Samples
1	14851576	02/09/2019	00:00.0	40.62	30	317	26.59	30	NaN	NaN	54.65	3
2	14851576	02/09/2019	05:00.0	41.60	14	317	30.00	14	35	1	55.40	3
3	14851576	02/09/2019	10:00.0	45.26	23	317	24.28	23	NaN	NaN	61.00	4
4	14851576	02/09/2019	15:00.0	41.70	11	317	24.55	11	NaN	NaN	54.56	4
5	14851576	02/09/2019	20:00.0	32.29	33	317	24.85	33	NaN	NaN	39.72	3
6	14851576	02/09/2019	25:00.0	25.70	17	317	21.15	17	NaN	NaN	39.34	1
7	14851576	02/09/2019	30:00.0	25.64	23	317	25.73	23	NaN	NaN	25.37	1
8	14851576	02/09/2019	35:00.0	32.66	29	317	23.62	29	NaN	NaN	37.18	6
9	14851576	02/09/2019	40:00.0	26.60	20	317	23.06	20	NaN	NaN	28.37	6
10	14851576	02/09/2019	45:00.0	40.35	14	317	22.00	14	46.5	1	46.46	7
11	14851576	02/09/2019	50:00.0	39.89	41	317	20.68	41	NaN	NaN	49.50	6
12	14851576	02/09/2019	55:00.0	41.57	18	317	20.00	18	NaN	NaN	54.51	5

Notes:
1. NPI_Link_ID (Col 1) is the "National Performance Indicator" link identifier, whereas Length (Col 6) is the link length.
2. localDate (Col 2) and localTime (Col 3) specifies the data timestamp. Note that localTime is the start time of a 5-minute interval.
3. Speed (Col 4) and Volume (Col 5) are the hybrid speed data and volume respectively.
4. NPI_Speed (Col 7) and NPI_Volume (Col 8) are the STREAMS speed data and volume respectively.
5. Here_Mean_Speed (Col 9) and Here_Link_Count (Col 10) are the HERE speed and sample count data, respectively.
6. BT_Speed (Col 11) and BT_Samples (Col 12) are the Bluetooth speed and sample count data, respectively.

As discussed above, the planned analysis requires the route travel time (instead of link) and in 15-minute time intervals/epochs. Thus, there needs to be some data processing (Section 3.1) prior to the calculation of the reliability measure and the subsequent comparison analysis (Section 3.2).

3.1. Obtaining the route travel time

This subsection details the process to calculate the route travel time for each 15-minute time interval from 1 January 2019 to 29 February 2020 (excluding weekends and public holidays) between 6am–7pm. This study used 15-minute time intervals as it is deemed to be the most appropriate for operational perspective, and aligning with the study done by Haghani et al. (2014). This means, for each route, there will be a total of 13,104 travel time values in 2019 (52 time epochs x 252 days) and 2,184 values in Jan-Feb 2020 (52 time epochs x 42 days). There are two steps involved, as follows.

As the dataset is originally stored in 5-minute intervals, the 5-minute data is aggregated into 15-minute time intervals. This is done by using the volume-weighted mean of the link travel time values within that 15-minute epoch (i.e. aggregating three data points into one). Note that the authors have loosely used the term “volume”, because the Bluetooth and HERE data have sample count information, rather than volume.

Next, for a given 15-minute time interval, the route travel time is calculated by adding the travel time of all the links making up the route. When there is missing data from some of the links within that route, their travel times are derived. This implies that it is assumed that the links with missing travel time data have the same speed as the average speed on all the links with data. The calculation is as follows:

$$\tau_t = \frac{\sum_i L_i}{\sum_{i \text{ with data}} L_i} \cdot \sum_{i \text{ with data}} X_{i,t}\,, \qquad \forall\, t\,, \tag{1}$$

where τ_t is the travel time of the route at time interval t, L_i is the length of link i, and X_i is the travel time of link i.

3.2. Travel time reliability comparison

In this study, the reliability measure used is the NPI (National Performance Indicator) reliability measure R (Austroads, 2018), as follows:

$$R = 1 + 1.6449\, C\,, \tag{2}$$

where C is the coefficient of variation (which is the standard deviation divided by the mean). Note that the NPI reliability measure R is simply a statistical method to estimate the buffer time index (by assuming a normal distribution). The authors selected a measure that is based on standard deviation as it is the most sensitive to data source selection (Haghani, Zhang and Hamedi, 2014).

This study looks at the day-to-day variation at each time interval, instead of the intra-day variation. For each 15-minute time interval, these reliability measures are computed on a monthly basis, which is only done when there are at least five data points for that month (i.e. five typical weekdays worth of data). The range of reliability measure is typically between 1.0 and 1.5. A reliability value above 1.5 is considered extremely unreliable.

In the next step, the travel time reliability computed from the four data sources are compared. The comparison is done using the STREAMS data as a benchmark (the status quo). Note that the STREAMS data is just a reference point (since the Queensland Department of Transport and Main Roads currently uses it to determine and corporately report on travel time reliability), as opposed to the "ground truth". The aim is to confirm the hypothesis of the emerging data sources result in a higher value of reliability, and to discover any patterns on these differences. Uncovering these patterns will provide some insights on the conditions at which the other data sources serve as a comparable data source with the STREAMS data when computing travel time reliability.

This study employs the *student's t-test* to test whether there is statistically sufficient evidence that the reliability values obtained from other data sources are higher than STREAMS. For this purpose, the difference in reliability values obtained from two data sources is firstly computed (e.g. STREAMS vs hybrid). Then, a right-tailed one-sample t-test is performed on these difference values. Specifically, let $x = R_{\text{OtherSource}} - R_{\text{STREAMS}}$. Then, the null hypothesis states that the mean of x is less than or equal to zero, i.e. $\mu_x \leq 0$; whereas the alternative hypothesis states that $\mu_x > 0$. The t-statistic is computed as follows:

$$t = \frac{\mu_x - \mu_0}{\sigma_x / \sqrt{n}}\,, \tag{3}$$

where $\mu_0 = 0$, σ_x is the standard deviation of the data x, and n is the number of data points. In this study, a confidence level of 95% is used.

In this study, 36 routes are selected with 8 being motorways and 28 being arterial roads (Figure 2). This study uses travel time data from 1 Jan 2019 up to 29 Feb 2020, between 6:00 and 19:00, excluding public holidays and weekends.

4. Results

The data processing has produced the reliability values for each route obtained from the four data sources. Then, the STREAMS reliability values are used as the benchmarks and the difference between the reliability values obtained from STREAMS data and obtained from the other data sources (Hybrid, HERE, or Bluetooth) are computed.

Figure 2: Map of the studied routes with the IDs denoted (with a unique ID for each direction)

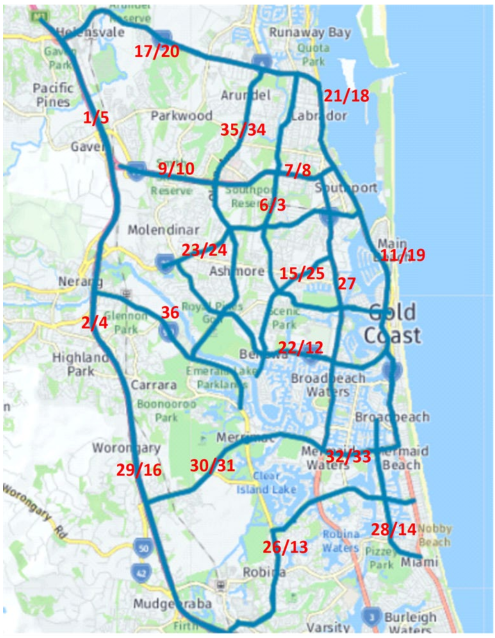
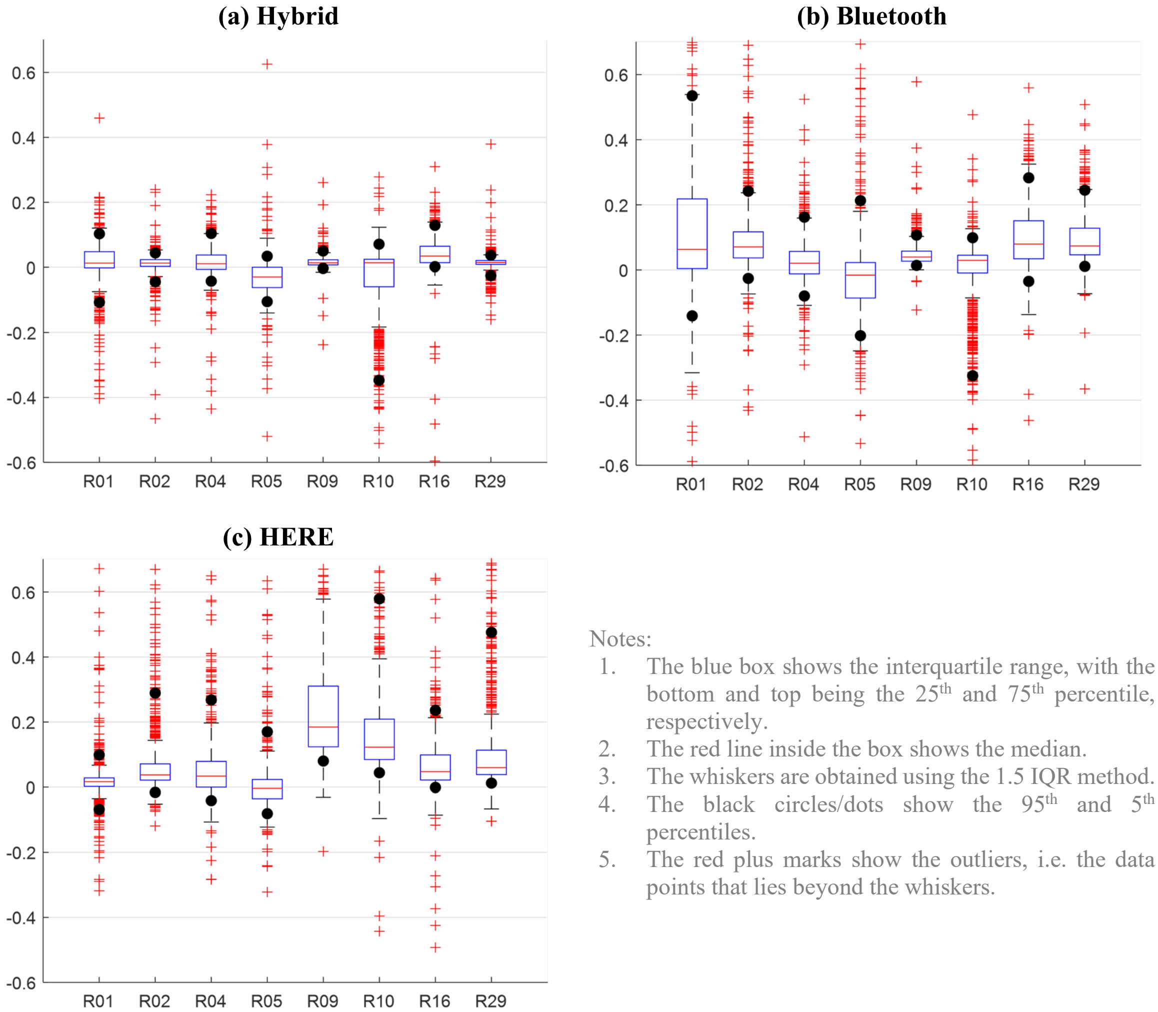


Table 1 shows the distribution of these calculated difference values, aggregated by road type. Note that a positive value means the reliability value obtained from the other data source (e.g. hybrid data) is higher (therefore travel time variability is higher) compared to that obtained from STREAMS data. It can be seen that the hybrid data tends to be higher but showed very little difference when compared to STREAMS, with the median lying at 0.01 and interquartile range (75th percentile minus 25th percentile) width of approximately 0.03 and 0.04 for motorways and arterials, respectively. Bluetooth and HERE data yield reliability values that are markedly higher compared to STREAMS, with HERE data having more prominent differences. Furthermore, the differences are bigger on arterials for Bluetooth and HERE; and the patterns seem to be consistent between the time periods.

Table 1: Percentile values for the reliability differences between other data sources and STREAMS

		AM			Interpeak			PM			All		
Type	**Data**	**Hy-brid**	**BT**	**HERE**	**Hy-brid**	**BT**	**HERE**	**Hy-brid**	**BT**	**HERE**	**Hy-brid**	**BT**	**HERE**
Motorways	**Max**	0.29	1.48	1.85	0.63	1.41	3.31	0.23	1.45	3.40	0.63	1.48	3.40
	95% tile	0.08	0.22	0.33	0.10	0.27	0.39	0.08	0.30	0.53	0.09	0.26	0.42
	75% tile	0.03	0.08	0.11	0.03	0.10	0.12	0.03	0.11	0.15	0.03	0.10	0.13
	50% tile	0.01	0.04	0.04	0.01	0.04	0.05	0.01	0.04	0.06	0.01	0.04	0.05
	25% tile	0.00	0.02	0.01	0.00	0.01	0.02	-0.03	-0.02	0.02	0.00	0.01	0.02
	5% tile	-0.05	-0.04	-0.04	-0.07	-0.11	-0.04	-0.22	-0.22	-0.05	-0.10	-0.13	-0.04
	Min	-0.37	-0.52	-0.29	-4.25	-4.25	-2.24	-2.44	-2.36	-2.27	-4.25	-4.25	-2.27
Arterials	**Max**	0.36	1.27	4.70	0.38	1.49	4.81	0.42	1.28	4.71	0.42	1.49	4.81
	95% tile	0.08	0.41	1.45	0.08	0.51	1.42	0.09	0.45	1.76	0.08	0.47	1.54
	75% tile	0.03	0.16	0.63	0.03	0.17	0.62	0.03	0.17	0.79	0.03	0.16	0.68
	50% tile	0.01	0.08	0.40	0.01	0.08	0.39	0.01	0.09	0.51	0.01	0.09	0.43
	25% tile	-0.01	0.04	0.26	-0.02	0.04	0.27	-0.01	0.04	0.35	-0.01	0.04	0.29
	5% tile	-0.07	-0.05	0.14	-0.06	-0.03	0.16	-0.05	-0.02	0.20	-0.06	-0.03	0.16
	Min	-0.58	-0.61	-0.48	-0.15	-0.14	-0.02	-0.21	-0.26	0.03	-0.58	-0.61	-0.48

Figure 3: Box plots showing the reliability differences against STREAMS for motorway routes



Notes:

1. The blue box shows the interquartile range, with the bottom and top being the 25th and 75th percentile, respectively.
2. The red line inside the box shows the median.
3. The whiskers are obtained using the 1.5 IQR method.
4. The black circles/dots show the 95th and 5th percentiles.
5. The red plus marks show the outliers, i.e. the data points that lies beyond the whiskers.

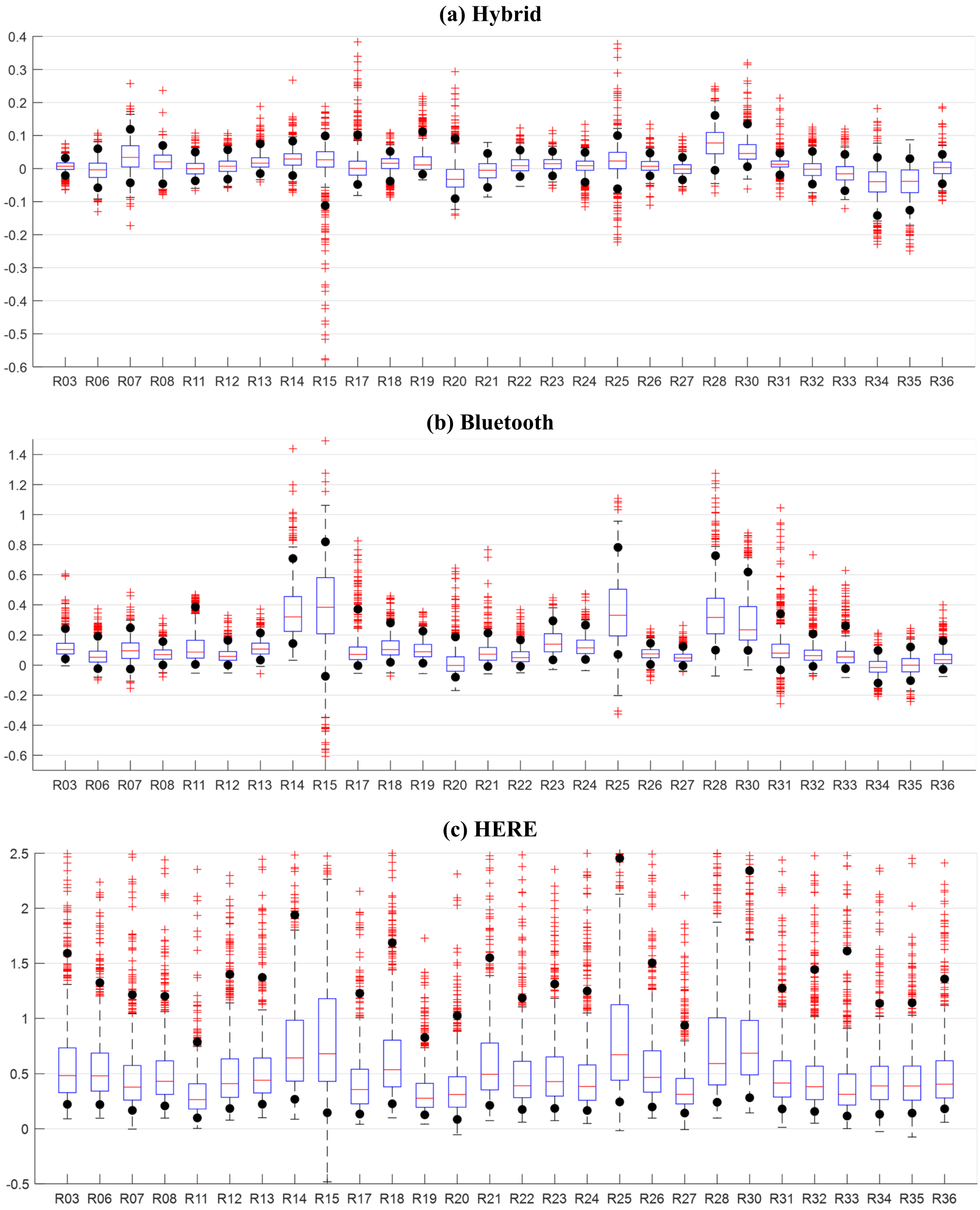
A series of box-and-whisker plots are produced to provide more insights of the results at the route level, as shown in Figure 3 and Figure 4. Some observations and further information are presented below.

- For hybrid data (Figure 3(a) and Figure 4(a)), it can be calculated that the average median for motorways and arterials are very close to zero (0.01 and 0.00 respectively), while the average interquartile range for motorways and arterials are at 0.03 and 0.04 respectively. Only 6 routes (1 of the 8 motorway routes and 5 of the 28 arterial routes) have their median lower than zero. Half (18 out of 36) of the routes have the 25th percentile at or above zero. This is much more than the 4 routes which have the 75th percentile at or below zero. When motorways and arterials are compared, a higher percentage of motorway (than arterials) support the general finding that hybrid data produce higher reliability values than STREAMS.
- For Bluetooth data, the difference is more prominent as shown in Figure 3(b) and Figure 4(b). Only 3 routes (2 arterials and 1 motorway) have the median found below zero and the rest has the 25th percentile above or very close zero. There is even one route where the minimum difference is zero and another where it is still positive.
- HERE data exhibits the biggest differences with STREAMS, with most routes have the 5th percentile located above zero. Four motorway routes have the 5th percentile above zero; and the minimum differences of 13 routes are still positive. It is also worth noting

that a lot of upper outliers are found on arterial routes, indicating some of the reliability values are unrealistically high ($R > 2.0$).

Furthermore, the mean difference values along with the results of the t-tests are presented in Table 2, with further detailed results presented in Table A1 in the Appendix. It can be seen that, for all but two cases, there is statistically significant evidence to reject the null hypothesis.

Figure 4: Box plots showing the reliability differences against STREAMS for arterial routes



Notes:

1. The legends are the same as Figure 3.
2. The scale of the y-axis is varied for each subfigure for better visibility.

In other words, it can be concluded with 95% confidence level that the other data sources produce higher reliability values compared to STREAMS. The exceptions to this are for hybrid data on motorways during PM peak, and when considering the whole study period ('All'). Furthermore, aligning with the previous observations from Table 1, Figure 3, and Figure 4, it can be noted from Table 2 that the mean difference values are highest for HERE and lowest for hybrid data; as well as higher for arterial routes.

Table 2: The mean of the difference values between other data sources vs. STREAMS at various time periods for different road types.

Type	AM			Interpeak			PM			All		
	Hybrid	BT	HERE	Hybrid	BT	HERE	Hybrid	BT	HERE	Hybrid	BT	HERE
Arterial	0.01	0.12	0.54	0.01	0.14	0.54	0.01	0.13	0.68	0.01	0.13	0.58
Motorway	0.01	0.06	0.09	0.01	0.06	0.10	-0.02	0.04	0.14	0.00	0.05	0.11

Note: The red shaded cells indicate the case where the null hypothesis cannot be rejected (i.e. p-value greater than 0.05).

5. Discussion

5.1 Possible causes of difference in reliability reporting

Among the three alternative data sources, the hybrid data is the closest to STREAMS data in terms of the reliability values. This is as expected as the hybrid data blends with high percentage of STREAMS data due to the lack of coverage (data gaps) from Bluetooth or HERE data, which sit at 68% and 45% respectively; in comparison to 85% and 93% coverage for STREAMS and hybrid data respectively.

Additionally, the differences on arterials are larger compared to the differences on freeways. This might be explained by the lack of data coverage for HERE and Bluetooth on arterials, but the larger range of reliability values can also be because HERE and Bluetooth are able to capture delays on arterial roads better under saturated conditions. This aspect of HERE and Bluetooth data overcomes the limitation of STREAMS NPI methodology, which, in the case of arterials, derives delay from detectors before the stop bar and signal settings.

5.2 Impact on travel time reliability reporting

As part of the operation of the road agency of the state of Queensland in Australia, a network-level reporting is carried out to a provide a snapshot of the reliability of the state-controlled road network. For this purpose, an NPI reliability threshold of 1.2 is used to categorise routes with 'good' vs. 'bad' reliability for each month and for each 15-minute time period. Table 3 shows the impact of changing the data source of this reporting from STREAMS to another data source (hybrid, Bluetooth, or HERE) on the network-level reliability numbers. Note that the percentage shows the proportion of bad reliability occurring for a route considering all 15-minute periods (within the study period of 6am-7pm) and all 14 months (Jan 2019 to Feb 2020).

It can be noted that, in general, using data sources other than STREAMS will likely lead to an increased proportion of the network with bad reliability. For hybrid data, only 3 of the 36 routes will experience changes above ±10% in overall reporting. On the other hand, both Bluetooth and HERE data will introduce very large increases in bad reliability for arterials. For motorways, however, the changes are not as excessive.

6. Conclusions

In summary, this study confirms that the data source used has an impact on travel time reliability calculations, more so on arterials compared to motorways. Among the three other data sources, hybrid data is the most comparable to STREAMS; whereas HERE data is the

Table 3: Changes of 'bad' reliability proportion when switching from STREAMS to other data sources

Arterial ID	Data source: Hybrid	BT	HERE
3	0%	26%	100%
6	0%	20%	100%
7	12%	39%	90%
8	3%	26%	97%
11	2%	48%	94%
12	1%	12%	99%
13	0%	33%	100%
14	1%	98%	100%
15	1%	80%	89%
17	5%	28%	95%
18	-1%	38%	96%
19	7%	37%	96%
20	7%	18%	93%
21	-1%	34%	98%
22	2%	13%	99%
23	0%	59%	98%
24	0%	49%	99%
25	4%	83%	87%
26	0%	8%	99%
27	0%	6%	97%
28	14%	95%	100%
30	5%	83%	98%
31	1%	19%	91%
32	0%	28%	95%
33	-1%	25%	97%
34	-11%	-3%	86%
35	-7%	0%	92%
36	1%	9%	99%

Motorway ID	Data source: Hybrid	BT	HERE
1	-1%	25%	-1%
2	-1%	11%	10%
4	3%	5%	5%
5	-8%	-8%	-4%
9	0%	2%	55%
10	-7%	-7%	21%
16	10%	27%	14%
29	-1%	12%	16%

source with largest differences. On average, the use of the tested hybrid data (instead of STREAMS data) has a small impact on the travel time reliability calculations and network reporting (using a threshold value of 1.2) for arterials and motorways during peak periods. Yet, the use of Bluetooth and HERE as data sources yields statistically significant higher reliability values on all periods on both road types. Furthermore, on some individual routes, the reliability values seem unrealistically high, and the difference will severely impact the network-level reporting.

Based on these findings, it can be concluded that practitioners should expect some differences in reporting the reliability values when switching to other data sources. The use of hybrid data may be considered appropriate, as the reliability values produced were found to be within a reasonable range. However, this does not necessarily infer that STREAMS data is the most accurate. It does, however, pose the question whether STREAMS underestimate the travel time variability or whether the dynamic aggregation mechanism of the hybrid data introduces additional variability. Furthermore, the coverage for HERE data may have been improved in recent years, which will potentially impact the results if this study is to be repeated by using more recent datasets. These questions are good candidates to become the basis of future works.

Acknowledgement

This paper is based on the research outcomes of the National Asset Centre of Excellence (NACOE) Project R117: The Impact of using Hybrid Data Versus Individual Data Sources for Travel Time Reliability Measurement. NACOE is a collaborative research agreement between the Queensland Department of Transport and Main Roads (TMR) and the Australian Road Research Board (ARRB), which is under NTRO.

References

Austroads (2018) *Guide to Traffic Management Part 3: Traffic Studies and Analysis*. AGTM03-17.

Austroads (2020) *Innovation and best practice in performance measurement and transport outcomes*. AP-R638-20. Sydney, NSW: Austroads.

Denne, T. *et al.* (2013) *Blueprint for a best practice measurement indicator set and benchmarking*. Research Report 522. NZ Transport Agency.

FHWA (2017) *Travel time reliability: making it there on time, all the time*. US Department of Transportation. Available at: https://ops.fhwa.dot.gov/publications/tt_reliability/ttr_report.htm (Accessed: 19 September 2022).

Fontaine, M. *et al.* (no date) *Travel time speed and reliability data*, *Traffic Forecasting Resource*. Available at: https://tfresource.org/topics/Travel_Time_Speed_and_Reliability_Data.html (Accessed: 19 September 2022).

Haghani, A., Zhang, Y. and Hamedi, M. (2014) *Impact of data source on travel time reliability assessment*. UMD-2013-01. Washington, D.C., USA: Mid-Atlantic Universities Transportation Center.

Johnston, D. (2021) *Intelligent hybrid data model: description of methods and results*. Contract report. Intelligent Transport Service.

Kessler, L., Rempe, F. and Bogenberger, K. (2021) 'Multi-sensor data fusion for accurate traffic speed and travel time reconstruction', *Frontiers in Future Transportation* [Preprint].

Saxena, N. *et al.* (2021) *Australian transport assessment and planning guidelines: road reliability measurement – research report*. Canberra, ACT: Transport Infrastructure Council.

TRB (2016) *Highway capacity manual*. 6th edn. Washington, D.C., USA: TRB.

Appendix – Detailed route results

Table A1: The mean of the difference reliability values for each route and the t-test results.

Type	ID	AM			Interpeak			PM			All		
		Hybrid	BT	HERE	Hybrid	BT	HERE	Hybrid	BT	HERE	Hybrid	BT	HERE
Arterial	3	0.00	0.11	0.56	0.01	0.13	0.57	0.00	0.11	0.78	0.01	0.12	0.63
	6	-0.01	0.06	0.55	-0.01	0.06	0.51	0.00	0.06	0.70	0.00	0.06	0.58
	7	0.04	0.10	0.39	0.03	0.10	0.45	0.04	0.09	0.67	0.04	0.10	0.50
	8	0.02	0.08	0.61	0.02	0.06	0.44	0.02	0.08	0.58	0.02	0.07	0.53
	11	0.01	0.10	0.33	0.00	0.16	0.30	0.00	0.09	0.44	-0.01	0.12	0.36
	12	0.01	0.08	0.62	0.01	0.06	0.45	0.01	0.06	0.63	0.01	0.07	0.56

Type	ID	AM			Interpeak			PM			All		
		Hybrid	BT	HERE	Hybrid	BT	HERE	Hybrid	BT	HERE	Hybrid	BT	HERE
	13	0.02	009	0.57	0.02	0.11	0.53	002	0.14	0.62	0.02	0.14	0.57
	14	0.04	0.37	0.84	0.03	0.41	0.82	0.02	0.28	0.78	0.03	0.36	0.81
	15	-0.03	0.26	0.86	0.04	0.49	0.98	0.04	0.38	1.01	0.02	0.39	0.95
	17	0.02	0.11	0.36	-0.01	0.10	0.39	0.02	0.09	0.71	0.01	0.10	0.48
	18	0.02	0.14	0.62	0.01	0.11	0.67	0.01	0.12	0.77	0.02	0.12	0.69
	19	0.02	0.10	0.36	0.02	0.10	0.30	0.02	0.10	0.40	0.02	0.10	0.35
	20	-0.05	0.00	0.36	-0.03	0.01	0.32	0.02	0.06	0.56	-0.02	0.02	0.41
	21	0.01	0.12	0.72	-0.02	0.06	0.56	0.00	0.08	0.72	-0.01	0.08	0.66
	22	0.02	0.06	0.37	0.00	0.04	0.46	0.02	0.08	0.71	0.01	0.06	0.51
	23	0.01	0.12	0.46	0.02	0.15	0.44	0.02	0.18	0.78	0.01	0.15	0.55
	24	0.01	0.15	0.51	0.01	0.12	0.65	0.00	0.12	0.50	0.01	0.13	0.37
	25	0.01	0.35	0.89	0.04	0.40	0.98	0.02	0.31	0.86	0.02	0.36	0.92
	26	0.01	0.06	0.58	0.01	0.09	0.57	0.01	0.08	0.70	0.01	0.08	0.61
	27	0.00	0.06	0.33	0.00	0.05	0.35	0.00	0.06	0.54	0.00	0.05	0..40
	28	0.07	0.25	0.84	0.07	0.39	0.91	0.09	0.39	0.89	0.08	0.35	0.88
	30	0.05	0.25	0.86	0.06	0.31	0.94	0.05	0.30	0.87	0.06	0.29	0.89
	31	0.01	0.11	0.58	0.02	0.11	0.50	0.01	0.11	0.55	0.01	0.11	0.54
	32	0.01	0.09	0.43	-0.01	0.07	0.51	0.00	0.08	0.64	0.00	0.08	0.52
	33	-0.02	0.07	0.44	-0.02	0.07	0.38	0.00	0.07	0.65	-0.01	0.07	0.48
	34	-0.06	-0.03	0.37	-0.04	-0.01	046	-0.03	0.01	0.61	-0.04	-0.01	0.48
	35	-0.05	-0.02	0.42	-0.04	0.00	0.43	-0.03	0.03	0.59	-0.04	0.00	0.47
	36	0.00	0.04	0.44	0.00	0.06	0.54	0.01	004	0.63	0.00	0.05	0.54
Motorway	1	0.03	0.13	0.02	0.01	0.10	0.02	-0.01	0.16	0.04	0.01	0.13	0.03
	2	0.01	0.08	0.06	0.01	0.09	0.06	0.00	0.09	0.09	0.01	0.09	0.07
	4	0.01	0.03	0.03	0.03	0.04	0.07	0.00	0.01	0.07	0.02	0.03	0.06
	5	-0.03	0.01	0.00	-0.02	-0.01	0.01	-0.05	-0.03	0.06	-0.03	-0.01	0.02
	9	0.02	0.06	0.22	0.01	0.04	0.29	0.02	0.05	0.38	0.02	0.05	0.30
	10	0.03	0.06	0.22	-0.08	-0.06	0.12	-0.17	-0.14	0.22	-0.07	-0.05	0.18
	16	0.03	0.04	0.05	0.06	0.13	0.10	0.04	0.11	0.08	0.04	0.10	0.08
	29	0.01	0.08	0.10	002	0.12	0.11	0.00	0.08	0.16	0.01	0.09	0.12

Note: The red shaded cells indicate the case where the null hypothesis cannot be rejected (i.e. p-value greater than 0.05).