Validation and applicability of floating car speed data for road safety

Chris Jurewicz[a], Dr. Ian Espada[a], Tariro Makwasha[a], Dr Clarissa Han[a], Dr Hafez Alawi[b], Dr Jiří Ambros[c],

[a]Australian Road Research Board, Australia; [b]Transport Accident Commission, Australia; [c]CDV – Transport Research Centre, Czech Republic.

Abstract

Traffic speeds are one of the most important factors influencing road safety outcomes. Arrival of ‘big data’ and powerful analytical methods opened a new avenue for obtaining and applying traffic speed data for road safety purposes. This paper briefly explores conventional speed data sources and the relatively new floating car data (FCD) as a source of speeds. It then compares the conventional spot-speeds with FCD speeds for selected parts of the road network in Victoria. Presented analyses shows a systematic relationship between the two data types, with FCD speeds being lower than spot-speeds. Availability of FCD speeds was much greater than spot-speeds, covering most of the public road network.

With appropriate understanding, FCD speeds can be used for many road safety purposes: network-wide speed monitoring, and for informing and evaluating speed management. Further exploration of specific use-case feasibility and data calibrations are planned. The paper points to new opportunities for safety monitoring and evaluation, and for development of more sophisticated speed-safety models than those currently available.

Introduction

Traffic speeds are one of the most significant factors in road safety performance, one which has been relatively well understood through research over the past twenty years. Changes in speeds are an indicator of success of some road safety policies, e.g. speed enforcement. If percentages of vehicles speeding by more than 10 km/h, 20 km/h, and 30 km/h are reduced, then such policies are contributing to overall safety improvement. Changes in personal crash risk can be estimated using relationships developed by Kloeden et al. (2001, 2002).

Typically, road safety practitioners are interested in mean speed changes for use with Elvik’s Power Model (e.g. Nilsson 2004, Elvik 2009, 2013). Operating speeds (85th percentile) are also sought as they are an important operational and road design factor.

In other uses speed changes are used as a proxy measure for success. For instance, effectiveness of traffic calming can be measured by speed reductions well before crash numbers change and can be evaluated. This enables early corrective actions to be implemented, if necessary.

This Transport Accident Commission (TAC) sponsored study sought to better understand conventional and emerging ‘big data’ sources of speed information, their strengths and weaknesses in the context of system-wide speed monitoring and evaluation for road safety purposes. The study objectives were to explore the relationship between conventional spot-speeds and floating car data (FCD) speeds, their road network coverage, latency and availability, and other questions. The study concluded with preliminary recommendations regarding use of FCD speeds in road safety monitoring and evaluations.

Spot vs. floating car data speeds

Conventionally, speeds of all vehicles are collected at a point location on the road for a defined period (e.g. two weeks). The methods of measurement vary and include pneumatic tube counters, electromagnetic or piezo-magnetic loops, or TIRTL laser technology. Radar gun is another form of data collection typically using small samples of vehicles over a short time.

The common aspect of conventional techniques is the assumption that the spot measurement is applicable to the whole road section of interest (TRB, 2011). This approach means that traffic speed data is available in a limited number of locations on the road network, rather than a as continuous map of speeds. Also, if speed information is needed at a specific location, it usually is measured at a considerable expense and delay.

The second feature of spot-speed data is that it is collected under established guidelines avoiding geometric constraints (i.e. on flat straights) and away from intersections. The industry standard seeks this information to be presented as free flow speeds, i.e. with headways between vehicles greater than 4 seconds (Austroads 2016). This approach seeks to strip off the effects of congestion, road geometry, and of other road users' presence on drivers' choice of speed.

In recent years, vehicle navigation devices began offering a sizeable sample of accurate global positioning data across the entire road network. This 'big data' source is called floating car data (FCD), or probe vehicle data. It comes from a variety of sources such as in-built navigation services, taxi instrumentation, commercial vehicle logistics and tracking devices, and from mobile phones. In most applications, speeds are calculated from distance covered along a route between time-stamped satellite 'pings' (10-60 sec apart). Thus, FCD speeds are based mainly on travel time between two known points (minority are reported instantaneous speeds). The data is assigned to road segments defined by navigation service providers' maps (30m – 2000 m). Unlike spot-speeds, there is no direct way to screen headways, as the data is sampled (2-10% of the traffic) based on vehicles proprietary navigation technology. Overseas studies have shown promising correlations of FCD speeds with spot-speeds (Bekhor et al. 2013, Aarts et al., 2015, Reinau et al. 2016, Ambros et al. 2017). These findings informed the methods used in this study.

Study methods

The study reviewed the speed data sources currently used in Victoria from the perspective of the following strategic objectives set by TAC:

1. Monitor trends in speeds across the network to assist in managing the State Road Safety Strategy.
2. Provide input into speed management programs such as speed limit setting.
3. Evaluate the effectiveness of broad programs and local deployments of speed management programs and speed enforcement.

Spot-speed samples were obtained courtesy of VicRoads' Traffic Information System for several different source types. The sources were assessed across several criteria driven by the above strategic objectives, such as spatial and temporal availability, availability of speed KPIs (e.g. mean speeds, 85th percentiles, etc.), and frequency of data collection. The authors then sampled FCD speeds from two commercial providers and subjected the data to the same assessment for comparison.

In the second stage of analysis, samples of conventional spot-speeds and FCD speeds were obtained for direct preliminary comparison at selected sites and time periods in Victoria (locations determined by availability). Data was sampled from urban freeways and arterials, and from a rural road. The mean and 85th percentile speeds, and statistical distributions were compared between the two speed data sets, drawing preliminary validation conclusions and qualifications about the FCD speeds.

Preliminary use-cases were then explored to better understand the FCD speeds in their different applications to safety.

Findings

Conventional spot-speeds data sources

Samples of VicRoads speed data sources were analysed, and the key custodians were interviewed to obtain information pertaining to aspects which could not be directly measured. The following assessment of the data sources for speed was made. The assessment criteria were agreed to by TAC and VicRoads based on informing the needs of a future speed distribution management system. Table 1 presents the results.

Table 1 Conventional spot-speeds sources, assessment vs. TAC data criteria

Spot-speed data source type	1. Availability everywhere across public network	2. Frequent updates	3. Long historical data available	4. All leading KPIs possible	5. Short time periods	Comments
Tube data (Metrocount)	×	×	✓	✓	✓[1]	Specific locations on **rural arterials**. Annual updates, or less often.
Telemetry sites (pavement loops)	×	✓	✓	✓[2]	×	Specific locations on **rural arterials**.
TIRTL	×	✓	×	✓	✓[1]	Several sites on **urban freeways**.
Freeway data stations (pavement loops and studs)	×	✓	✓	✓[2]	×	Specific locations on **urban freeways**. Data of better standard than telemetry sites.
Arterial data stations (pavement loops)	×	×	✓	✓[2]	×	Specific sites on principal **urban arterials**. Updated annually.
Radar gun	×	×	✓	✓	×	76 locations on **urban arterials**, sampled annually for a two-hour period.

1. Yes, if set up at new locations for the required periods. Expensive to deliver and manage.
2. Available but unclear how biased the means and 85th percentiles are, as distributions are based on aggregated samples and/or and speed bins. These sources cannot measure vehicle headways, and can only estimate free-flow speeds (e.g. speeds late at night).

The assessment confirmed neither spot-speed data source covers the entire road network. Rather, all sources provide data at selected points across the state-controlled road network (local government roads were not included in the VicRoads system).

Also, it was found that each source had unique technical limitations related to data collection technology, data storage and aggregation, and calculation of speed KPIs. These limitations make these sources incomparable. For instance, freeway data stations measure speeds in 20 sec averages, continuously. Since individual vehicle speeds are not reported, this source cannot provide headway. Hence, only general speed profile is available regardless of headways and congestion. Free-flow speeds can only be inferred (e.g. in the middle of the night). In contrast, radar gun studies are conducted twice a year, for two hours, and are based on a sample of 100 surveyor-selected free-flowing vehicles with headways exceeding 4 sec. Thus, speed KPIs obtained from such different sources and different parts of the road network cannot be directly compared.

The findings suggest that there was no single ‘ground truth’ speed data source representing all parts of the road network. Rather, some of the spot-speed data sources should be noted for their accuracy and versatility. TIRTL and tube count data are proven low-error techniques, measuring individual vehicles, and thus providing rich data sets for analysis and interpretation.

FCD speeds

The same assessment criteria were applied to three proprietary sources of FCD speeds. Some samples were obtained and the data provider representatives were interviewed to obtain further information. The data characteristics were similar across all three providers, with small variations in technical presentation of the data and in the sampled vehicle fleet composition. Table 2 shows the results of the assessment.

Table 2. FCD speeds, assessment vs. TAC data criteria

Data source	**1. Availability everywhere across public network**	**2. Frequent updates**	**3. Long historical data available**	**4. All leading KPIs possible**	**5. Short time periods**	**Comments**
FCD speeds	✓	✓	✓	✓[1]	✓	May not cover some very low volume roads (e.g. access lanes). Quality and availability of short-period KPIs (e.g. 15 min) will be better on high-volume roads. Long extraction time periods may be needed on low-volume roads.

1. No headways can be provided and speed data is available as KPIs only, e.g. mean, std. dev, and a 5-percentile increments. Two of the three providers can also provide a sample size.

It was evident that FCD can provide a consistent source of speed data across the entire state. Given that FCD speeds are measured over longer distances, they are expected to be different to spot-speeds (time-mean, see Austroads 2016 for definitions). The main expected effect would be that the FCD speeds would have lower values as they are averaged over a known road length, rather than representing a speed under idealised location and driving conditions.

Most critically, no headway information is available, hence free-flow speeds cannot be directly measured. FCD speeds are not available in raw format, but the data is quality assured by the provider. Providers are bound by privacy laws, and then exceed these requirements by destroying parts of the information which could be used to identify individual vehicles/users, before the data is released for use. This may pose issues to some analysts, as data quality cannot be directly assessed as it can be with tube counter or TIRTL data[1]. Instead, data providers offer information about data filtering and cleaning methods (e.g. how stationary vehicles, personal navigation devices, or multiple devices per vehicle are stripped out of the data).

FCD speeds are available back to 2008 for much of the road network, with the last three years of data being of the best quality due to rapid increase in use of navigation services. This means that the first strategic objective, i.e. speed trend monitoring across the entire network can be satisfied if the data is deemed to be acceptable by researchers and practitioners.

[1] Such options may also not apply to other conventional point-speed data sources such as loops, studs and radar due to automated data aggregation to reduce storage.

FCD speeds are available for all types of road locations, including steep grades, curves and intersections, during congested and uncongested periods, and in road sections with traffic calming. Figure 1 shows a historical snapshot of FCD mean speeds for Wednesday daytime off-peak period across part of the Melbourne road network. Speeds on local streets have been turned off for display clarity. Such features of the data mean that the second strategic objective (speed management program inputs) may be met, subject to satisfactory assessment of the FCD speeds. These questions of validation are tackled in the next section.

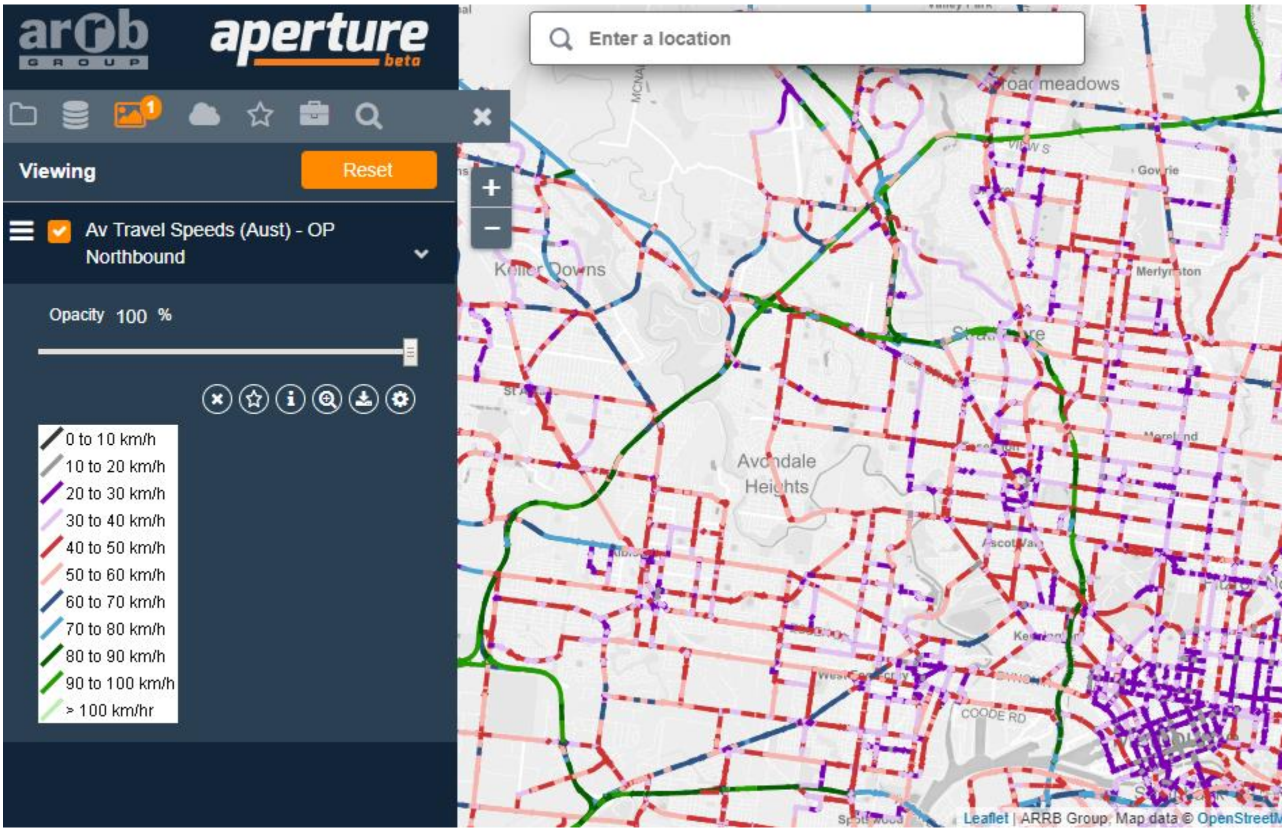

Figure 1. Snapshot of FCD mean speeds across the arterial network

Findings from data analysis and preliminary validation

The study's second and third strategic objectives sought to establish whether data could be used to inform speed management programs, and to evaluate their effectiveness. To answer this, preliminary validation of FCD speeds was undertaken against known conventional sources of spot-speeds.

The first test sought to establish if there was a general relationship between spot-speeds and FCD speeds. For this purpose, a publicly available set of VicRoads speeds was used and matched with FCD speeds. The VicRoads spot-speeds were based predominately on inductive loops, measured at metropolitan freeway and arterial data stations (see Table 1 for characteristics, both data sets used all-traffic speeds, i.e. no headways were available). The finest level of spot-speed data available were hourly mean speeds sampled during 2012 (dates unknown). These data stations were geographically matched with corresponding FCD links. FCD speeds were extracted from HERE Traffic Analytics system for the available three-year period (2012-14), also presented as hourly means. Locations with insufficient FCD data quality were removed from further analysis[2]. Analysis was carried out on data from 235 locations, with speed limits between 50 km/h and 100 km/h.

Direct comparison of the hourly mean speeds provided little insight due to significant data scatter (24 hourly values x 235 locations, with differences in data periods). Figure 2 shows more aggregated level of analysis, comparing the average hourly mean speeds for each location using both data sources. Figure 2 shows that while there were location-based effects (different road types and operational categories), there was a strong overall relationship between averaged mean loop spot-speeds and FCD speeds (R2=0.83). FCD speeds, being space-mean speeds were lower than time-mean spot-speeds, as expected. The standard error was 9.7 km/h, similar to that found previously by Espada and Bennett (2015), while Hrubeš & Blümelová (2015) found FCD speeds to be similar but lower than loop speeds. The relationship was also statistically significant. This level of correlation was welcome yet surprising, given that the two speed data collection methods were so significantly different.

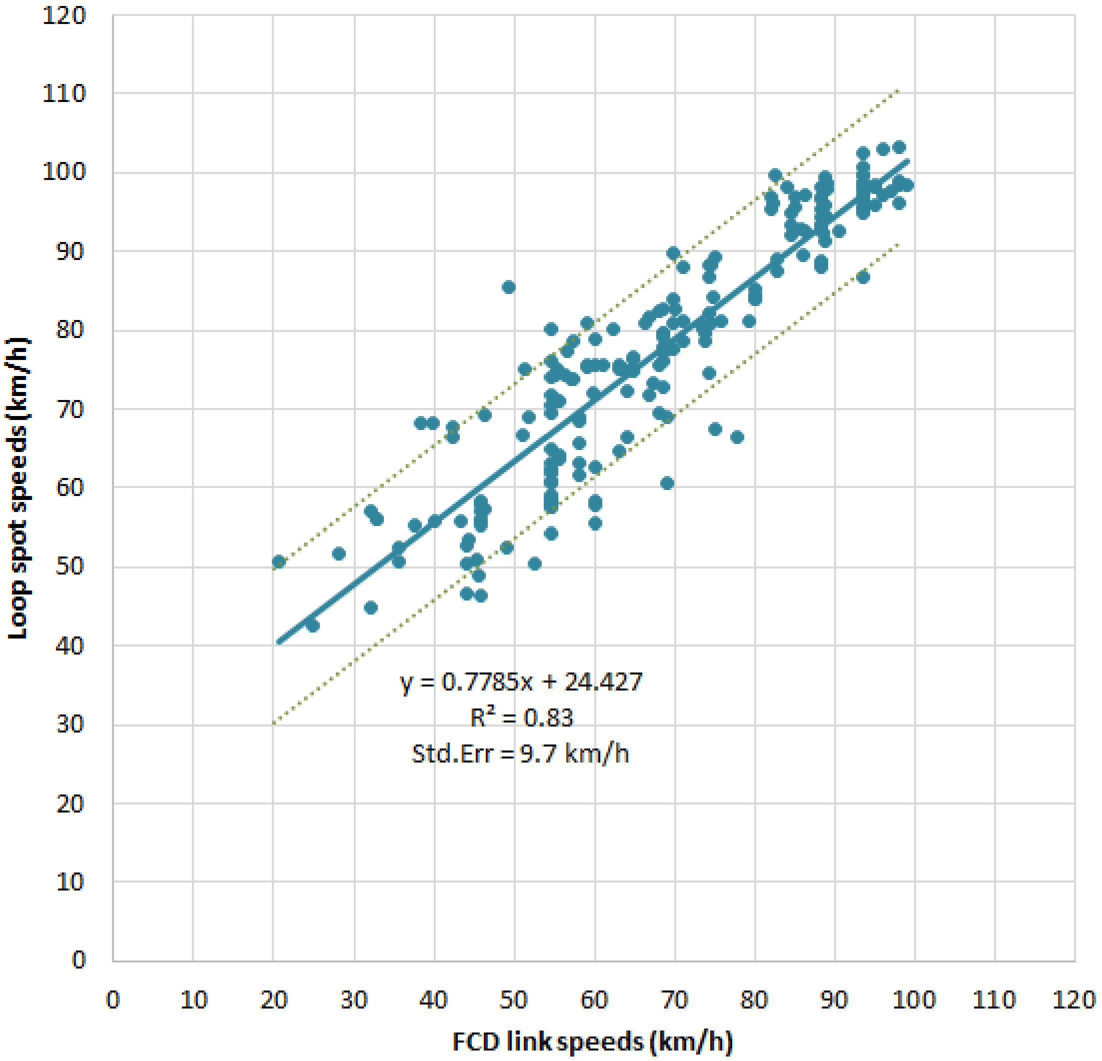

Figure 2. Averaged hourly mean speeds for every location

Further analysis was carried out, averaging hourly mean speeds across all similar locations, i.e. by speed limit. Results for different speed limits showed similar trends, and the results for 80 km/h roads are shown in Figure 3. It is clear, that both speed data sources show similar trends across the day in response to congestion. Also, FCD mean hourly speeds were typically 11 km/h lower than

[2] FCD data is improving continuously. Data quality issues present in 2012-14 may not be an issue in 2017.

loop mean hourly speeds (range 7 – 15 km/h for this speed limit). Figure 3 also shows the standard deviation for each data source; it is clear that both overlap, suggesting lack of statistically significant difference between mean speeds drawn from the two data sets for 80 km/h roads.

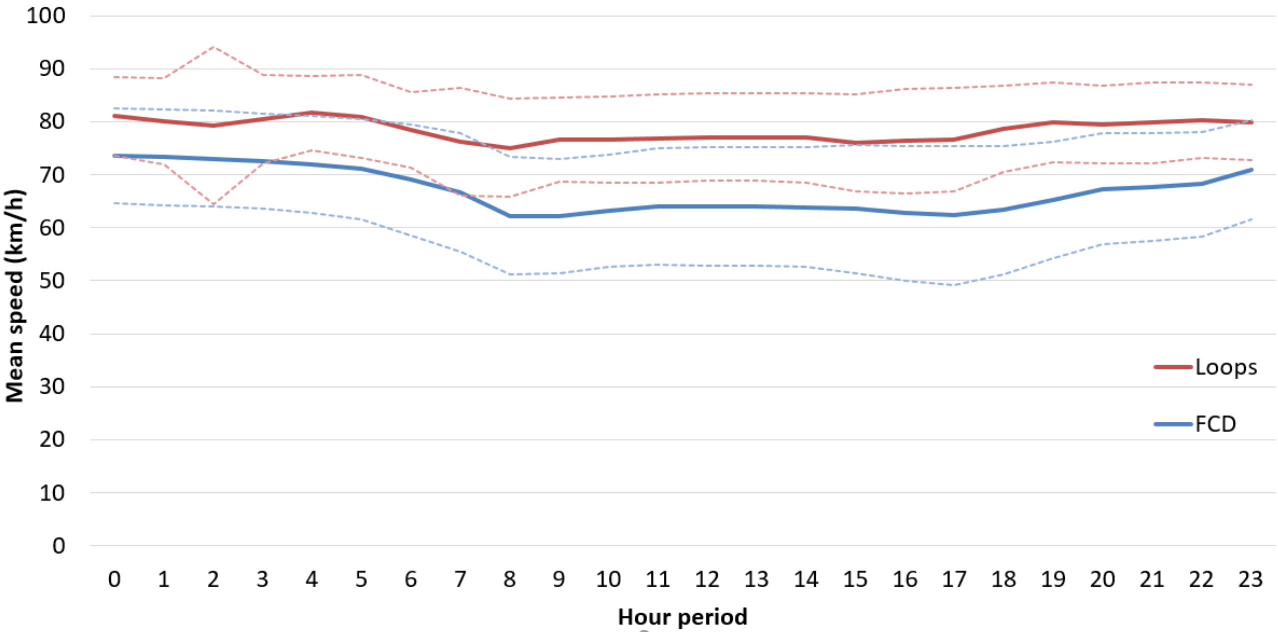

Figure 3. Comparison of hourly mean speeds averaged across all 80 km/h locations

Spot-speeds from loop data stations are just one of the data sets currently used in strategic speed monitoring and management in Victoria. TIRTL technology developed in Australia has been used more frequently in recent years to provide very precise spot measurements for traffic counting and classification, and for speed measurements (CEOS, 2017).

The analysis was limited to two locations, inbound and outbound at the same Monash Freeway chainage, to better understand if FCD data can relate sudden changes in speeds as well as a trusted conventional source.

All-traffic data was used in the analysis. Disaggregation of free-flow speeds was possible but difficult without access to proprietary TIRTL software. FCD data was extracted for links matching the two TIRTL locations. In this case both TIRTL and FCD speeds were extracted for the same April to June 2016 period. Given very high traffic flows on Monash Freeway, there were no data quality issues with either TIRTL or FCD speeds.

Figure 4 shows the results of the comparison for the AM peak analysis for inbound traffic (a) and outbound traffic (b). Given better access to the spot-speed data, analysis was possible for mean and 85th percentile speeds. Other speed percentiles were also compared. Figure 4 shows clearly the effects of peak flow: the inbound speeds (a) present a range of speeds expected before, during and after the congested period.

The outbound speeds (b) present largely uncongested flow. The outliers in the outbound data set represent times of incidents and temporary works on the freeway during the data period. Both data sources represented these temporary changes in traffic speeds equally well. This provides a degree of confidence that FCD speeds could provide a response to speed management projects.

In this analysis, Chi-square test was carried out showing that the two speed distributions were statistically different, both for mean and 85th percentile speeds (also confirmed by KS and ANOVA tests). The regression analysis showed consistent relationships between spot-speed and FCD speeds (see Figure 4). It is noted that the regressed relationship for mean speeds in Figure 4 a) was close to that found from the loop data in Figure 2. This only applied to inbound flow where a broad range of speeds was available.

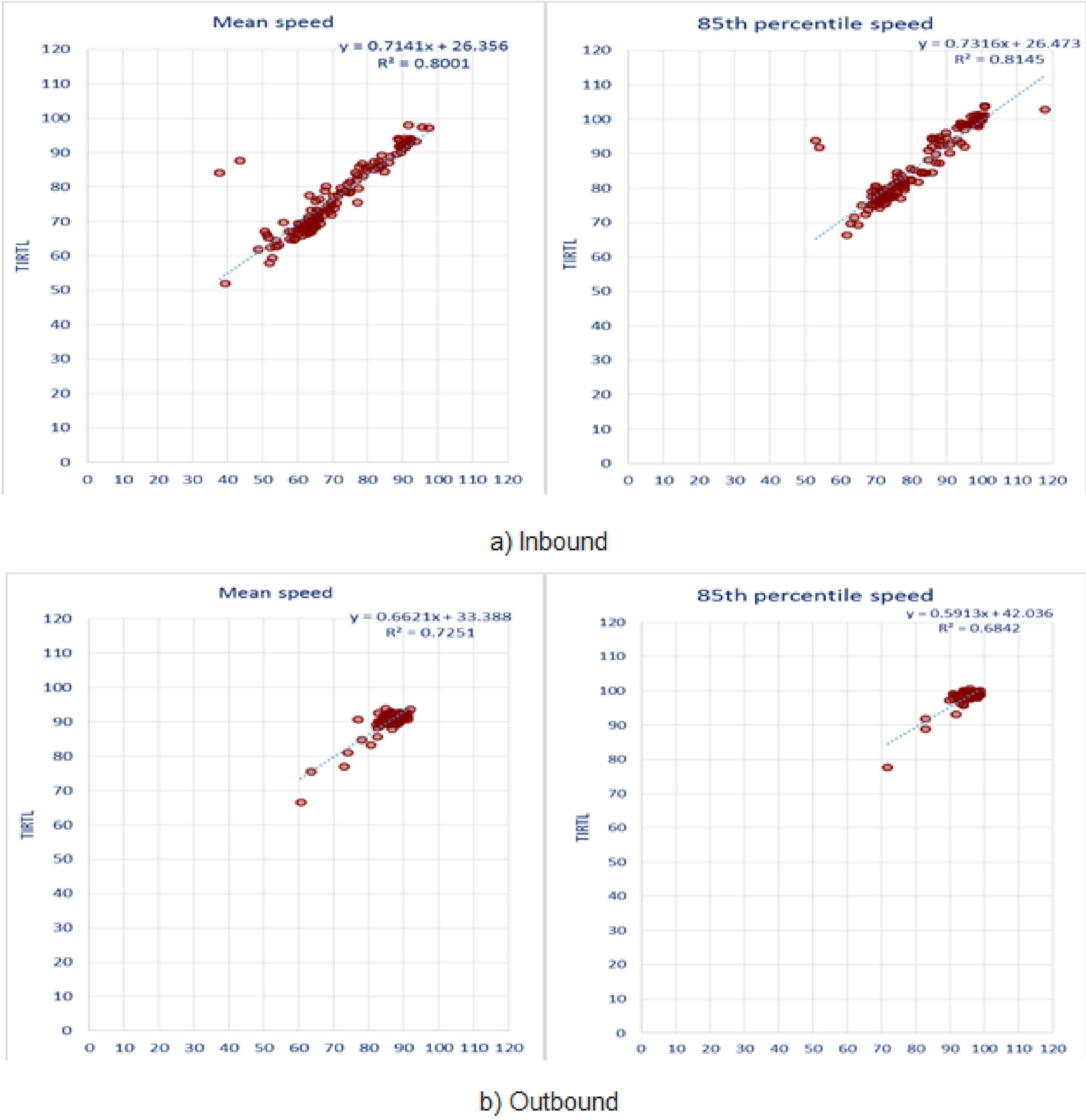

Figure 4. Comparison of hourly mean speeds by day of the week for the two locations

Use-cases of FCD speeds

Availability of FCD speeds across the road network suggests many new uses for policy makers and practitioners. Figure 1 showed basic mean speed mapping functionality. The following figures demonstrate some of other use-cases developed during the study.

Figure 5 shows the effect of reducing the speed limit in Bell St in the Melbourne suburb of Coburg, from 70 km/h to 60 km/h in early June 2015, to improve safety performance. FCD speeds were sampled for a midblock link between March and December 2015. Figure 5 shows mean and 85th percentile FCD speeds applicable between 10am and 2 pm.

Figure 5 shows the immediate effect of speed limit reduction on the 8 June, with a slight down-trend until September. Discounting this transitional period, the mean and 85th percentile speed before/after changes were in the range expected from a short follow-up study, even if the absolute values may require calibration (e.g. the relationship from Figure 2). Mean speed change can be used in Elvik (2009) to estimate the expected crash reduction due to speed limit reduction. Lack of speed change would prompt additional action (e.g. enforcement, and/or traffic calming measures)

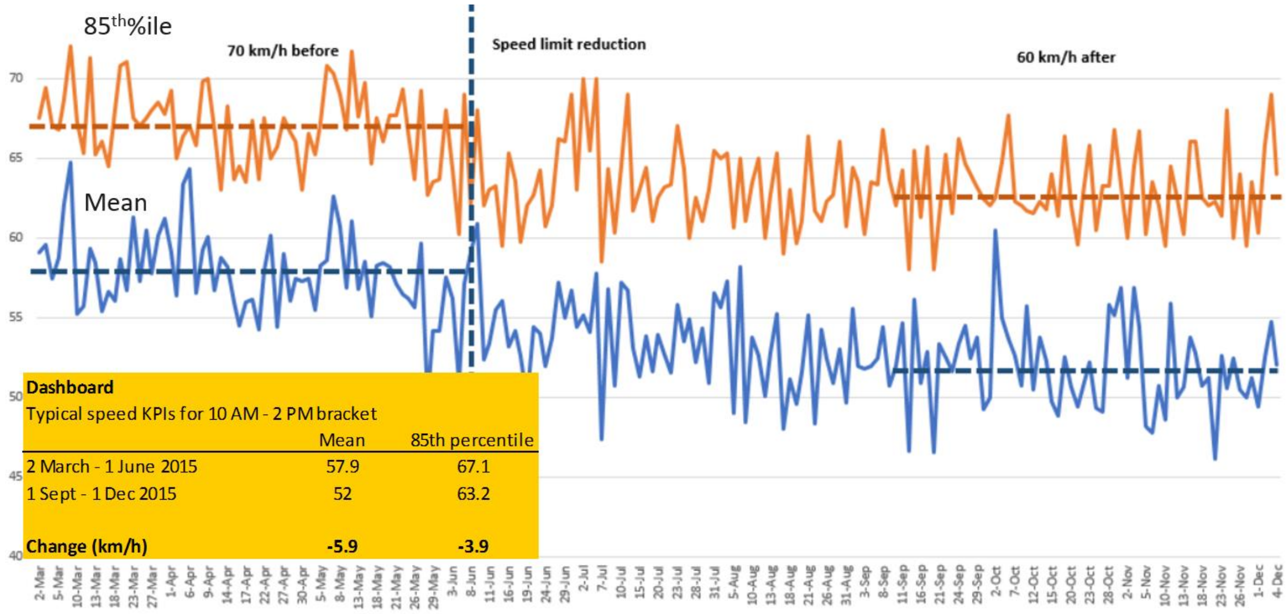

Figure 5. Bell St speed limit reduction, June 2015

The critical new development is that collection of before and after tube count data would have required road agency resources, if done at all. Placement of counters at this busy location would result in traffic management costs and delays. FCD speeds offer an opportunity for easy retrospective access to speed data, and thus enable monitoring and evaluation of speed management (strategic objective three).

Another use case involves observation of different types of time-trends for a rural location to inform speed limit enforcement programs. An FCD link was selected for Midland Highway in Bonie Doon, Victoria. Figure 6 shows hourly mean and 85th percentile speed plots for a period between 1 July 2016 and 23 March 2017. A clear trend is exhibited: speeds increase above the 100 km/h speed limit during the night, between 10 pm and 6 am. This is the time when majority of motorists may feel that the risk of police enforcement is at its lowest. Even without accurate calibration, the relative FCD speed data suggest that increased night-time enforcement would reduce risk of speeding-related crashes.

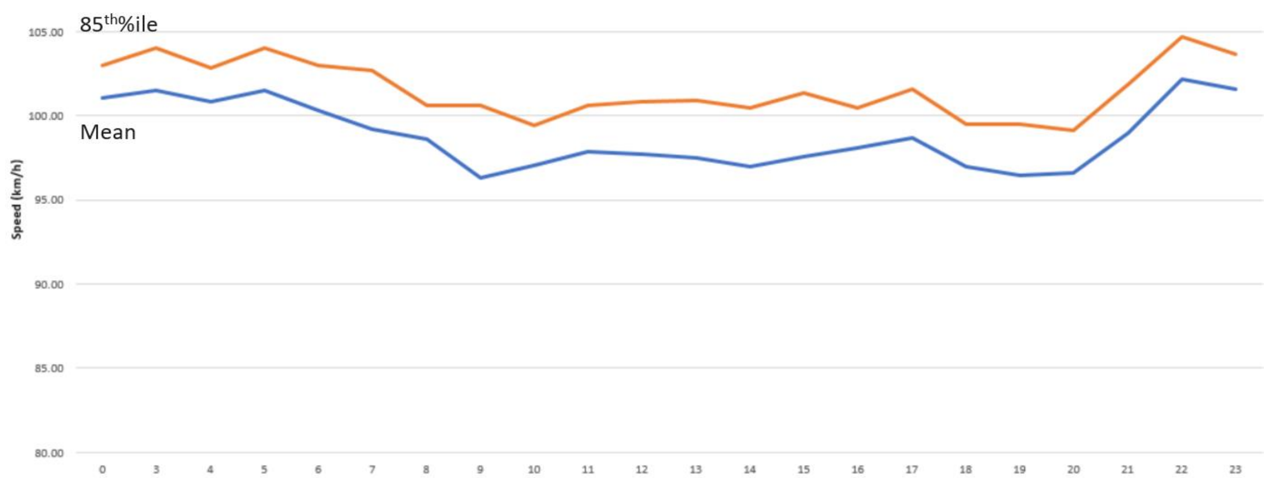

Figure 6. Mean and 85th percentile hourly speeds for the Bonie Doon link

There are many other proposed use-cases to be explored with FCD data. Some examples include:

- Area-based detection of traffic speed changes in response to traffic calming.
- Short-term evaluation of treatments, where speed is a proxy measure of safety.
- Input into enforcement planning through identification of temporal speeding hotspots, and road segments with high crash / severe injury risk due to speeding.
- Inputs into network-wide safety performance modelling and planning of investment programs (e.g. ANRAM).

These and other use-cases can be accommodated in a spatial analysis system for FCD speeds. Work currently under way is exploring online business analytics software for lean delivery of such a system. This approach would provide agility for data updates and evolving system functionality with the needs of its users.

Discussion

The following observations about FCD were made from the preliminary validation:

- FCD speeds are ideal for retrospective extraction for any link on the road network.
- Roads with high traffic volumes will generate sufficient FCD data samples more quickly – shorter data extraction periods can be accommodated (e.g. one month).
- Lower-volume roads will generally require longer data extraction periods.
- Obtaining speeds for short time periods (e.g. hourly) will require longer data extraction periods (weeks, months). Longer time periods (e.g. peak, off-peak, 24h) would require shorter extraction periods.
- Knowledge of 'probe' vehicle sample size is the best way to assure quality of the extracted data. Minimal sample size required for analysis varies with standard deviation. For relatively free-flowing traffic (std. dev < 5 km/h), a sample of 100 probes will produce speed KPIs accurate to 1 km/h. For more varied flows, or longer time periods, larger samples will be needed (around 700 vehicles for std. dev. of 13 km/h). Sample size and standard deviation are included in the data outputs provided by some of the FCD providers.
- FCD is becoming more plentiful as navigation services and devices proliferate via IoT phenomenon (Internet of Things). FCD speeds in 2017 are significantly more plentiful and of better quality than in 2014.

This paper proposed some preliminary calibration relationships between FCD and spot-speeds, based on available data (e.g. Figure 2). Such calibration may be useful as traffic practitioners will seek to sense-check FCD speeds against the expected spot-speed values collected using conventional methods. Current speed-safety models (e.g. Elvik 2009) rely on mean spot-speeds, so conversion of FCD speeds may be necessary.

Calibration models should be developed using large number of locations across different road stereotypes. One possible solution would be multivariate models for mean or 85th percentile spot-speeds considering input variables such as relevant FCD speeds, rural/urban environment, speed limit and road geometry.

This paper considered all-traffic speeds. Many traffic and design practitioners prefer using free-flowing speeds (headways greater than 4 sec), as they are intended to indicate driver response to speed limit and road geometry, rather than to presence of other road users. Views on this are evolving. Some propose that such free-flow conditions are rare for most of the travel undertaken on Australian roads. Consequently, these are not the traffic conditions under which majority of crashes and injuries occur. In this context, deeper understanding of speeds may be needed under different operational conditions. Mobility practitioners embrace this concept in their studies of operational network efficiency and congestion.

FCD speeds can be used to develop a new generation of speed-safety models to replace the Elvik (2009) Power Model. Some of the possibilities include intersection-specific models where approach speeds could be statistically related to crash outcomes. Also, specific relationships could be developed for pedestrian and cyclist safety outcomes given prevailing speeds. Temporal speed-safety models can be developed leading to better appreciation of the safety effects of congestion and speeding. Most importantly, use of FCD speeds would assist in application of proactive approach to road safety planning and Safe System implementation.

Conclusions

The study provided a significant leap in the understanding of FCD speeds and their application to road safety. FCD speeds provided by HERE were validated against samples of conventional spot-speeds (loops, TIRTL) showing FCD speeds to be systematically lower. It was demonstrated that FCD speeds could be calibrated to estimate mean spot-speeds, e.g. for use in Elvik's Power Model.

Use-cases in the paper showed that all three strategic safety objectives for network-wide speed monitoring, and for informing and evaluating speed management programs can be met using FCD. Many speed distribution indicators can be derived from FCD speeds, e.g. mean and 85th percentile speeds, standard deviation.

Further work is needed to understand functional data limitations in specific use cases, and to develop new practices, e.g. FCD speeds calibration, sampling technique guidance, and new speed-safety performance models.

Acknowledgements

The authors would like to thank VicRoads and HERE for provision of free data samples used in this study.

References

Aarts, L. T., Bijleveld, F. D., Stipdonk, H. L. (2015). Usefulness of 'floating car speed data' for proactive road safety analyses: Analysis of TomTom speed data and comparison with loop detector speed data of the provincial road network in the Netherlands. Report R-2015-3. Leidschendam, the Netherlands: SWOV. <https://www.swov.nl/sites/default/files/publicaties/rapport/r-2015-03.pdf

Austroads (2016). Guide to traffic management part 3 – traffic studies and analysis, AGTM03-13, Austroads, Sydney, NSW.

Ambros, J., Valentová, V., Gogolín, O., Andrášik, R., Kubeček, J., Bíl, M. (2017). Improving the self-explaining performance of Czech national roads. Transportation Research Record: Journal of the Transportation Research Board, No. 2635, pp. 62–70.

Bekhor, S., Lotan, T., Gitelman, V., Morik, S. (2013). Free-Flow Travel Speed Analysis and Monitoring at the National Level Using Global Positioning System Measurements. Journal of Transportation Engineering ASCE, Vol. 139, pp. 1235–1243.

CEOS (2017). TIRTL website, viewed 7 July 2017, <http://www.ceos.com.au/index.php/products/tirtl?id=10>

Elvik, R. (2009). The power model of the relationship between speed and road safety: update and new analyses, TOI report 1034/2009, Institute of Transport Economics, Oslo, Norway.

Elvik, R. (2013). A re-parameterisation of the power model of the relationship between the speed of traffic and the number of accidents and accident victims, Accident Analysis & Prevention, vol. 50, no. 1, pp. 854-60.

Espada, I., Bennett, P. (2015). ‘Probe data and its application in traffic studies’. Presented at 2015 IPWEA/IFME Conference, Rotorua, NZ.

Hrubeš, P., Blümelová, J. (2015). ‘Comparative Analysis for Floating Car and Loop Detectors Data’. Presented at the 22nd ITS World Congress, Bordeaux, France.

Kloeden, CN, Ponte G & McLean, AJ (2001). Travelling speed and the risk of crash involvement on rural roads, report CR 204, Australian Transport Safety Bureau, Canberra, ACT.

Kloeden, CN, McLean, AJ & Glonek G, (2002). Reanalysis of travelling speed and the risk of crash involvement in Adelaide, South Australia, report CR 207, Australian Transport Safety Bureau, Canberra, ACT.

Nilsson, G. (2004). Traffic safety dimensions and the power model to describe the effect of speed on safety, Bulletin 221, Lund Institute of Technology, Lund, Sweden, viewed 17 February 2017, <https://lup.lub.lu.se/search/ws/files/4394446/1693353.pdf>.

Reinau, K. H., Andersen, C. S., Agerholm, N. (2016). A New Method for Identifying Hazardous Road Locations using GPS and Accelerometer. In 23rd ITS World Congress, Melbourne.

TRB (2011). Modelling Operating Speed: Synthesis Report. Transportation Research Circular E-C151. TRB, Washington, D.C.