Department for Transport

Traffic Statistics Methodology Review
Alternative Data Sources

Moving Britain Ahead

July 2018

The Department for Transport has actively considered the needs of blind and partially sighted people in accessing this document. If you have other needs in this regard please contact the Department.

Department for Transport
Great Minster House
33 Horseferry Road
London SW1P 4DR
Telephone 0300 330 3000
Website www.gov.uk/dft
General enquiries: https://forms.dft.gov.uk

Contents

Foreword

The Department for Transport's Road Traffic Statistics Team have conducted a review of the traffic estimates for Great Britain. The aim of the review was to seek opportunities for innovation and efficiencies in the production of traffic statistics, without degrading their quality in terms of accuracy and reliability, timeliness and meeting user needs.

This document provides more detail on one of the main aspects of the review, investigating alternative data sources for collecting information on levels of road traffic within Great Britain.

User Feedback

We are keen to receive user feedback on the issues covered in this document. This can be given via the Road Traffic Statistics Team inbox: roadtraff.stats@dft.gov.uk.

Acknowledgments

DfT is grateful to Jim O'Donoghue from the Office for National Statistics (ONS) Methodology Advisory Service (MAS) for his input and helpful advice, and to the UK Statistics Authority for providing this support for the project via the Quality Improvement Fund.

Executive Summary

The Department for Transport (DfT) publishes estimates of traffic on Great Britain's roads[1]. The Road Traffic Statistics Team have investigated alternative data sources for collecting information on levels of road traffic within Great Britain, as part of a wider review of the road traffic statistics.

This report presents the results of the Alternative Data Sources project. A data source review was essential to ensure that the road traffic data collection continues to produce robust estimates of traffic, to keep up with technology developments, and to meet stakeholder needs.

The focus of the Alternative Data Sources project was to investigate and determine feasible options for collecting traffic count data, which would replace or supplement the current sources.

The project was split into three work streams, and a summary of the findings is presented below, with more detail in the remainder of this report.

Road-side data sources: feasibility trials of alternate technologies

Some of the technologies tested showed promising accuracy of total vehicle counts, however, none provided the accuracy of vehicle classification required. Therefore, none were deemed suitable for the wholescale replacement of the existing automatic traffic counters and manual counts combination for DfT traffic statistics production.

The DfT team will continue to explore the potential for the use of the total count data (i.e. without vehicle classification) from the technology types tested to estimate traffic statistics for all motor vehicles only.

As the market is continually changing, the team will continue to investigate technologies under development that may have applications for traffic counting in the future. Examples that have arisen since the completion of this project include new video analytics technologies.

Non-road-side data sources: synthesising estimates and MOT test data

The methodology trialled for synthesising road traffic estimates was not deemed reliable for large-scale implementation at this time. The team will continue to investigate if it could be used for individual road links, such as those where it is difficult or dangerous to conduct traffic counts.

The development of the use of odometer readings from MOT test data to estimate road traffic continues, with experimental statistics currently planned for 2019.

Data sharing with other organisations

Automatic traffic counter (ATC) data from other organisations would be a useful additional data source.

Since this project was completed, DfT have included Highways England, Transport Scotland, and Transport for London ATC data in the publication "*2017 Road Traffic Estimates for Great Britain*".

[1] Available at: www.gov.uk/government/collections/road-traffic-statistics

1. Current data collection methods

1.1 To produce DfT's road traffic statistics, road traffic observations are currently collected via three main methods:

- counts performed by trained enumerators standing by the side of the road for a 12 hour period, known as manual counts
- a small number of counts performed by trained enumerators watching a video recording of traffic, known as manual video counts
- automatic traffic counter (ATC) data, which are permanent installations embedded in the road surface. These combine Inductive Loops with Piezoelectric Sensors in a 'Loop – Piezo Sensors – Loop' array, and record information about vehicles passing over them, including vehicle length and wheelbase, to classify vehicles.

1.2 Around 8,000 manual and video traffic counts are conducted each year across roads in Great Britain. The mobility and flexibility of manual and video counts enables the Department's road traffic estimates to have the breadth of geographical coverage it has and to produce accurate and detailed vehicle classification. The counts are carried out between March and October, due to weather and light considerations at the count locations. The counts are of 12 hours duration – 7am to 7pm.

1.3 The ATC data provides continuous traffic count data for over 200 locations. These locations are a stratified panel sample. They provide sufficient observations so that in-year traffic variations can be estimated by road type and vehicle type.

Road type:

1.4 Half of all manual and video counts take place on the ***major road network*** (motorways and 'A' roads) - currently around 4,000 locations per year. This enables detailed road-level traffic estimates to be produced for motorways and 'A' roads.

1.5 The 2012 review of uses and users of traffic statistics established that the existing traffic data for each junction-to-junction link of the major road network was essential for a number of customers of road traffic data, in particular for producing road traffic forecasts and road transport emission statistics. Since 2012, the use of this road-level data has grown. Around 350,000 customers accessed this data in 2017 via our traffic counts website (http://www.dft.gov.uk/traffic-counts/) or via data.gov.uk (http://data.gov.uk/dataset/gb-road-traffic-counts). Therefore, it is desirable that any alternative data sources are either of a sufficient sample size or are flexible and mobile, so that the traffic estimates can continue to be produced at this detailed geographic coverage, either on their own or in combination with other sources.

1.6 For ***minor road*** ('B', 'C' and Unclassified roads) traffic estimates, the traffic on the same roads across Great Britain are counted each year via manual and video counts - over 4,000 locations. This panel sample approach enables robust regional and national level traffic estimates to be produced for minor roads.

Vehicle type:

1.7 The traffic estimates are split by vehicle classification, which is captured by both the manual and video counts and the automatic traffic counters. The vehicle classification split available is:

- Pedal Cycles
- 2-wheeled Motor Vehicles
- Cars and Taxis
- Buses and Coaches
- Light Goods Vehicles
- Heavy Goods Vehicles
    - 2-axle Rigid
    - 3-axle Rigid
    - 4-or 5-axle Rigid
    - 3- or 4-axle Articulated
    - 5-axle Articulated
    - 6- or more axle Articulated

1.8 More detailed explanations of the current data and methods used to produce traffic estimates are available via the websites below.

Links

Quarterly and annual statistics:
www.gov.uk/government/collections/road-traffic-statistics

Street level traffic counts (traffic estimates for each link on the major road network):
www.dft.gov.uk/traffic-counts/

More detailed guidance in the road traffic section of this note:
www.gov.uk/government/uploads/system/uploads/attachment_data/file/270083/contents-page.pdf

2. Alternative road-side data sources

2.1 One of the aspects of the Alternative Data Sources project was to investigate other road-side possibilities for collecting traffic count observations. These could either replace or supplement the current manual and video count sources. This chapter sets out the options explored.

2.2 The alternative road-side data sources investigated in this aspect of the project are those that provide geographically detailed traffic data, and so are those that are mobile and can be deployed flexibly.

2.3 The scope of the project on road-side technologies was to conduct feasibility trials to investigate the available road-side technology options for the production of road traffic statistics. The aim was to determine how well different equipment types performs in a range of ways, including:

- potential accuracy and reliability
- the range of vehicles classified
- the ease of deployment and operation, including health and safety issues
- how well the equipment works in different environments, such as heavy rain.

Technologies Trialled

2.4 A list of the available alternative road-side technologies was drawn up, see Figure 1. The statistics team borrowed a range of different brands for each of the technologies from a variety of contacts.

2.5 It should be noted that although a range of technologies were tested (including different brands of the same or similar technology), the trial did not seek to test all of the possible technology brands for all traffic counting purposes. The aim was to determine a technology type's suitability for DfT traffic statistics use.

Figure 1: Technology types

Rubber Tubes

Two pneumatic tube sensors are fixed on to the road surface a specific distance apart. The sensors register axle (tyre) hits from vehicles travelling over them. Based on the information from the two tubes, an estimate of the speed and the vehicle classification can be made.

Rubber tubes require a team of two to set up, and health and safety issues would need to be taken into account when the tubes are being installed and removed. Traffic management would be required. After the counting period, the tubes are taken up and can be redeployed at a different count site. Four different brands of rubber tube were trialled.

Radar

Radar uses radio waves to detect vehicles, and can record a basic vehicle classification.

A radar systems device is usually attached to bridge, gantry, lamp-post, or similar. The system needs to be at a certain angle to the road that it is monitoring. If this angle is incorrect, then the data quality decreases. Radar systems do not usually require traffic management[2]. There are fewer Health and Safety issues associated with Radar systems than for some other sources. Two different brands of radar and a bluetooth technology that works in the same way were trialled.

Video Counts (Automatic Video Recognition)

Video cameras are set up at the side of the road, which record passing vehicles. The camera data are then analysed using automatic recognition software to classify vehicles observed.

Video cameras can struggle to count/classify vehicles in adverse weather (e.g. fog, heavy snow). The use of video counts would not usually incur the costs associated with traffic management. Two different types of video technology were trialled.

Automatic Number Plate Recognition (ANPR) Cameras

Mobile ANPR cameras are currently used by the DfT Vehicles Statistics team, for the purposes of Vehicle Exercise Duty Evasion Statistics. The camera is usually situated at the side of the road and reads a vehicle's number plate which, when linked to the DVLA database, provides vehicle information that can be used to classify vehicles.

The camera has a certain degree of accuracy in terms of capturing all vehicles passing and capturing a readable number plate. This accuracy can be improved by subsequent human processing. Two different types of ANPR camera technology were trialled.

[2] Traffic management involves directing vehicular and pedestrian traffic around a construction zone, accident or other road disruption, thus ensuring the safety of emergency response teams, construction workers and the general public.

Analysis Summary

2.6 All of the technologies were trialled at several sites. A great deal of practical knowledge about deployment and reliability of the technology types was gained.

2.7 The traffic counts recorded by each of the technologies were directly compared to a count estimated from video footage that had been manually enumerated twice. Ten hours of observations were sampled per site. These hours were selected to capture a mixture of day, night, busy and quiet periods.

Comparing performance

2.8 The main findings from these feasibility trials were:

- Accuracy: Some of the technologies tested showed promising accuracy of total vehicle counts (within 1% deviation from the video count).
- Vehicle classification comparison: The analyses for different vehicle classifications found that none of the technologies provided detailed and accurate enough vehicle classification counts for DfT's statistical purposes.
- Sensitivity: all of the technologies tended to under count when compared to the manually enumerated video count. Some technologies were less sensitive than others but, even within the same type of technology, different brands had differing levels of sensitivity.
- Costs: some of the equipment could potentially be more expensive and less flexible than manual counts. Some types of equipment require local authority permissions for their temporary installations.
- Reliability: some of the technologies failed at some of the sites, but this was not always obvious until the analysis stage. Exploring boxplots by site and by technology revealed some set-up issues that had not been identified at the time of the trials.
- Security: some of the equipment was vandalised, some were more likely to be vandalised.

Conclusion

2.9 Some of the technologies tested showed promising accuracy of total vehicle counts, however, none provided the accuracy of vehicle classification required. Therefore, none were deemed suitable for the wholescale replacement of the existing automatic traffic counters and manual counts combination for DfT traffic statistics production.

2.10 The DfT team will explore the potential for the use of the total count data (i.e. without vehicle classification) from the technology types tested to estimate traffic statistics for all motor vehicles only.

2.11 As the market is continually changing, the team will continue to investigate new technologies under development that may have applications for traffic counting in the future. Examples that have arisen since the completion of this project include new video analytics technologies, and drone or satellite images.

3. Alternative non-road-side data sources

3.1 Non-road-side options for obtaining or synthesising traffic data have developed in recent years. Following a review of available data and the methods used to produce traffic estimates in other countries, two types of sources were investigated: synthesised traffic estimates using, for example, in-vehicle GPS; and administrative sources, such as MOT test odometer readings.

3.2 This chapter sets out the investigations carried out on these two types of non-road-side data source.

A. Synthesised traffic estimates investigation

3.3 Synthesised estimates would use data on traffic flows from other sources, such as in-vehicle GPS data (or, in the longer-term, possibly from mobile phone data), to model how traffic flows around the road network. When combined with traffic count observation data from a subset of locations on the given network, an estimate of traffic on each of the roads on the network could be calculated. Whilst this method still requires observations of traffic from short- and/or long-term traffic counts, it has the potential to reduce the sample size or frequency of these counts.

3.4 The DfT congestion statistics team use in-vehicle GPS data to produce average speed and delay estimates for motorways and 'A' roads in England. As this data source was already available, it was used by the traffic statistics team to test the feasibility of synthesising traffic estimates.

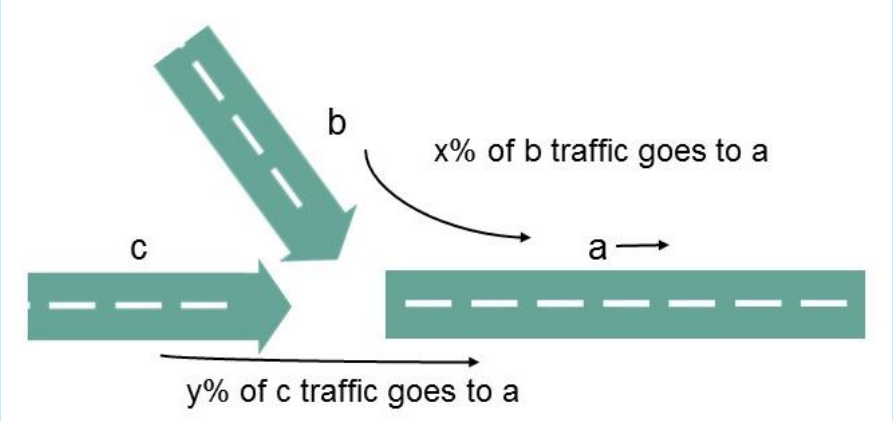
Methodology

3.5 The methodology applied is set out in Figure 2.

Figure 2: GPS-based analysis method A – adjacent links

The car traffic flow for road-link 'a' in the direction shown, is derived by:

- Using in-vehicle GPS data to calculate the percentages of vehicles that join road-link 'a' from road-links 'b' and 'c'.
- Applying these percentages to the observed counts on road-links 'b' and 'c'.

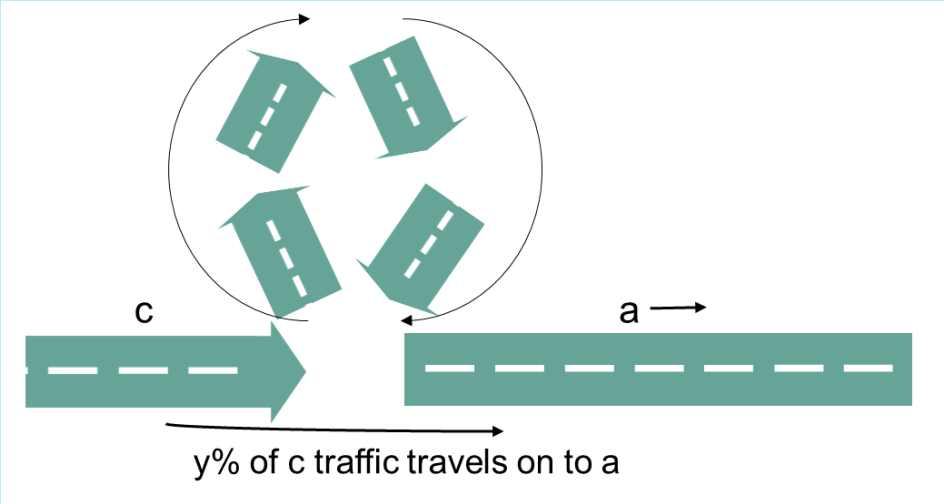


3.6 The in-vehicle GPS data was analysed in a database package to determine the percentage of traffic that flowed between adjacent links (as set out in Figure 2). Due to the complexity of the algorithms, the method was tested for just 30 paired road-links. To enable a larger sample to be tested, an alternative method (see Figure 3) was also tested for 8,000 paired road-links.

Figure 3: GPS-based analysis method B – trip analysis

The vehicle traffic flow for road-link 'a' in the direction shown, is derived (instead of being counted) by:

- Using in-vehicle GPS data to calculate the percentages of vehicles that go on to travel on road-link 'a' from road-link 'c'.
- Applying this percentages to the observed counts on road-link 'c'.



3.7 Method B uses the information from a vehicle trip to determine if the vehicle passed through the two road links (in chronological order) as it traverses the road network.

3.8 If the two methods produced results with similar levels of accuracy, then Method B would be preferable given that it is more efficient to process.

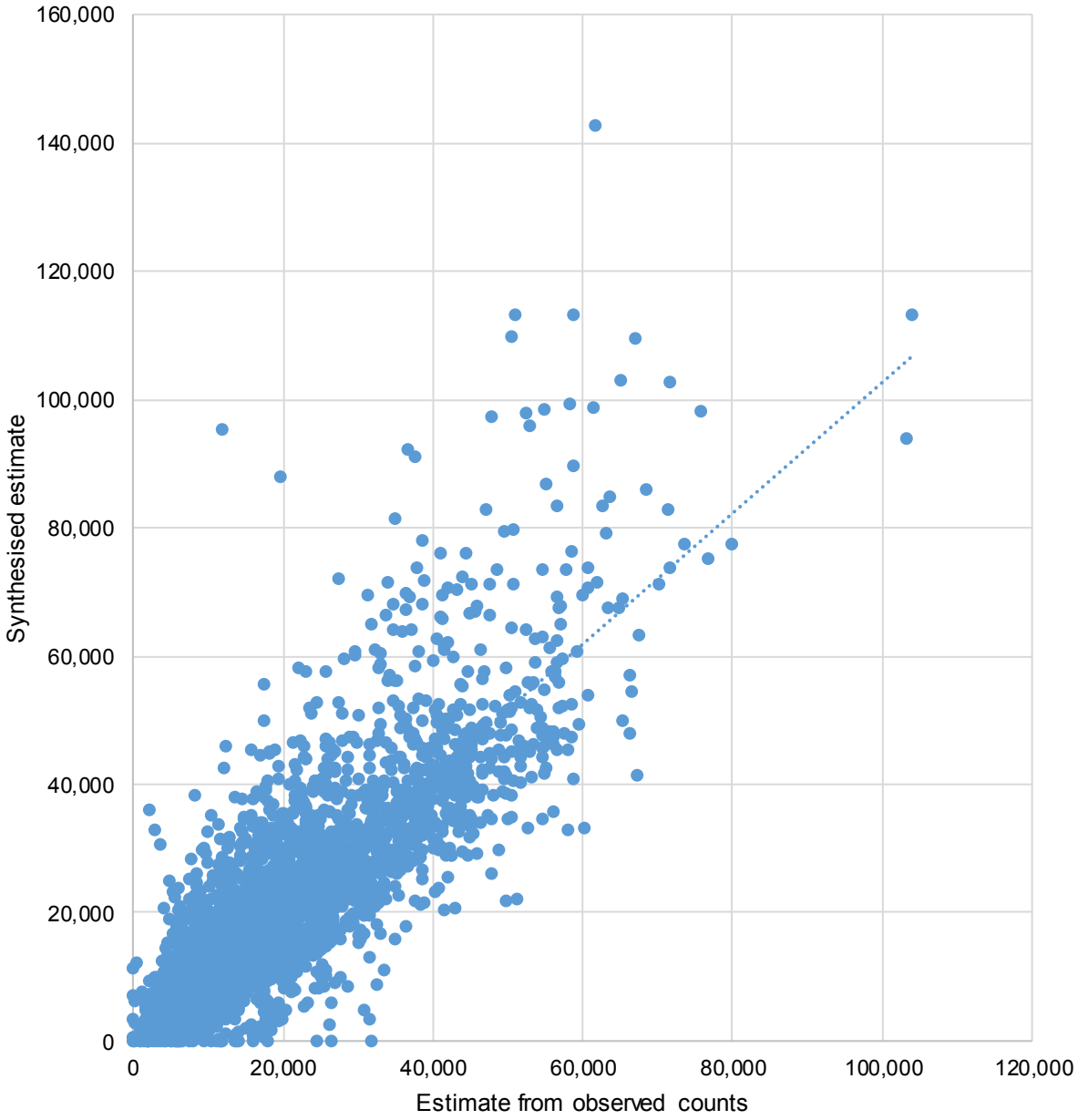
Analyses

3.9 Based on the small sample of 30 paired road-links, both Methods were run and compared against the traffic estimate from observed counts. The average percentage discrepancy between the observed count estimate and the synthesised estimate were close (20% for Method A and 23% for Method B). Given this relatively small difference, all other analyses were run using Method B as it could use a larger sample of road links.

3.10 Figure 4 shows a scatter plot of the estimate from observed counts (x-axis) compared to the synthesised estimate based on Method B (y-axis). The synthesised estimate was often higher, and there were a large number of outliers.

3.11 Cars produced the best comparison (as set out in Figure 4), followed by LGVs then HGVs. The HGV sample is a lot smaller (in both Method A and Method B datasets), which will have affected the analyses. For cars, the synthesised estimate was within 30% of the observed count estimate for 59% of the links; and 23% of links had less than 10% difference between the two estimates.

Figure 4: Comparison of estimates derived from observed counts and synthesised estimates (using Method B), for cars



3.12 To better understand the reasons for the differences in synthesised and observed estimates, a set of 20 case studies were selected, which were examined in more detail.

3.13 A number of reasons for the discrepancies between the estimates were found. Of the 20 case studies selected:

- Four had estimates that had a difference of less than 10%, and the difference was generally even lower for estimates produced using Method A. These road links tended to be motorways, or 'A' roads with few junctions.
- Eight had large discrepancies in the estimates produced due to missing data for roads feeding into the calculation. Often these road links had a major traffic contribution from one or more minor roads, for which either a count or in-vehicle GPS flows were not available.
- Five had discrepancies in estimates produced due to the different road network layouts used by the two methods, which caused mismatches in calculation of the

synthesised estimate. If this methodology was adopted, this aspect could be resolved in the longer-term.

- Three had discrepancies where the reason couldn't be identified.

Conclusion

3.14 The methodology trialled in this analyses was not deemed reliable for large-scale implementation at the current time.

3.15 The team will continue to investigate if it could be used for individual road links, such as those where it is difficult or dangerous to conduct traffic counts.

B. MOT test data

3.16 DfT statisticians have been working with the Driver & Vehicle Standards Agency to develop the MOT test dataset so that it can be robustly analysed. That project is outside this Methodology Review.

3.17 During an MOT test a record is made of the vehicle's odometer reading. The MOT test data, including the odometer reading, have been collated since the MOT system was computerised in 2005.

3.18 Using this information, annual mileage could be estimated for each vehicle included in the MOT test system. This includes cars, vans, and motorcycles but excludes buses and lorries.

3.19 The MOT test data has the potential to add great value to the traffic estimates, as it would provide a variety of additional vehicle information (such as age and propulsion type). As the MOT test dataset doesn't have information about where a vehicle is being driven, it can't be used as a replacement source for the traffic estimates.

Next steps

3.20 The road traffic statistics team are working closely with the statisticians leading the project developing the MOT test data, with experimental statistics currently planned for 2019.

4. Data sharing with other organisations

4.1 The final source of alternative traffic data under consideration in this review is traffic data collected by other organisations. A range of traffic observation data is potentially freely available to the Department for Transport from other public bodies. Within Great Britain, traffic count observations are also collected by:

- Highways England (responsible for trunk motorways and 'A' roads in England)
- Devolved administrations (trunk roads outside England), and
- Local authorities (all other public roads).

4.2 The potential benefits of using these sources are that they can provide a regular data supply for a subset of locations, and so reduce the number of locations that the Department for Transport needs to collect data on. However, like other administrative data, the Department would have no control over any changes made to these sources. In addition, careful consideration is needed as to the effect of adding a different source of data into the traffic estimates methodology.

4.3 The Methodology Review has looked at the use of automatic traffic counter data owned by the Scottish Government and by Highways England. This chapter sets out the outcomes from the investigations.

A. Data sharing with Scottish Government

4.4 The annual traffic estimates currently use data from a subset of the automatic traffic counters owned by the Scottish Government. These data are used to estimate traffic growth on certain motorway and 'A' roads sections in Scotland (approximately 100 locations) using the method set out in Figure 5.

Figure 5: Current Scottish Government traffic data method

To estimate the average number of motor vehicles that travel on a particular junction-to-junction road link (the 'AADF') for the latest year, the following formula is used:

$$\text{Annual Average Daily Flow (AADF) for latest year (for a given vehicle type)} = \text{Previous year's AADF (for a given vehicle type)} \times \frac{\text{ATC total for latest year}}{\text{ATC total for previous year}}$$

The 'ATC total counts' above are taken from the Scottish Government's automatic traffic counters located on the particular junction-to-junction road link. They are the total number of all motor vehicles counted.

4.5 This methodology was reviewed by the Methodology Advisory Service. They proposed one change to the above methodology. To investigate if an appropriate vehicle split is possible from the Scottish Government's automatic traffic counter data, and to use that in the formula above instead of using the 'all motor vehicle' total count for all vehicle types. This will be considered for future investigation.

4.6 Since this project was completed, DfT have tripled the number of locations that use Transport Scotland's ATC data in the publication "*2017 Road Traffic Estimates for Great Britain*".

B. Data sharing with Highways England

4.7 Highways England have over 10,000 automatic traffic counter sites across their network of motorway and 'A' roads in England. However, there are differences in the equipment and how it is set-up to the method used by DfT traffic statistics, as set out in Figure 6.

Figure 6: Differences in automatic traffic counters and their data

	DfT Traffic Statistics automatic traffic counters	Highways England automatic traffic counters
Vehicle classifications	11 (12 including pedal cycles) vehicle classifications based on vehicle length and axle configuration	4 vehicle type classifications based on vehicle length
Original purpose	Maintained for statistical purposes	Maintained for operational purposes of network management

4.8 To investigate the potential use of Highways England data, the data for all sites, for 15-minute intervals for all days in June 2015 were obtained. This amounted to almost 28 million rows of data, over 60 times more data than the DfT traffic statistics team currently process each month. This has implications for resources and database capacity.

4.9 After importing these data into a database, and performing some data checks, the data were aggregated to daily and to weekly levels for the exploratory analyses.

Analysis of results

Data matching and cleaning

4.10 Two initial location matches were done using GIS spatial software:

- Firstly, a match of the locations of Highways England automatic traffic counters with the DfT traffic statistics team's automatic traffic counter locations. This found 237 Highways England counters that corresponded to 31 DfT counters.
- Secondly, DfT's manual count locations in June 2015 were matched to Highways England's automatic traffic counter locations. This resulted in 1,502 Highways England counters matched to 160 DfT manual count locations.

4.11 A quick exploratory comparison of the DfT and Highways England data for these matched locations found a relatively high number of outliers showing large discrepancies of flows between Highways England counters and DfT counters for the matched locations.

4.12 A second phase of data cleaning manually reviewed the spatial matching, and conducted more validation of the Highways England flow data. These investigations found that some matched sites were not recording the same thing. Some of the difference was caused by counters being out of action for part or all of the month.

However, the main reason for differences were the different set-up of the counters as set out in Figure 7.

Figure 7: Differences in automatic traffic counter set-up

	DfT Traffic Statistics automatic traffic counters	Highways England automatic traffic counters
Locations	Only on the main carriageways, between junctions	Can be anywhere on the network including within junctions and on slip roads
Road/Lane coverage	Counters cover all lanes and both directions at a given location	Counters cover a single direction at a given location, and sometimes only one lane

4.13 Removing the sites affected by the above factors reduced matched locations for the sites in the investigatory work to:

- 152 Highways England counters matched to 24 DfT counters.
- 735 Highways England counters matched to 122 DfT manual count locations.

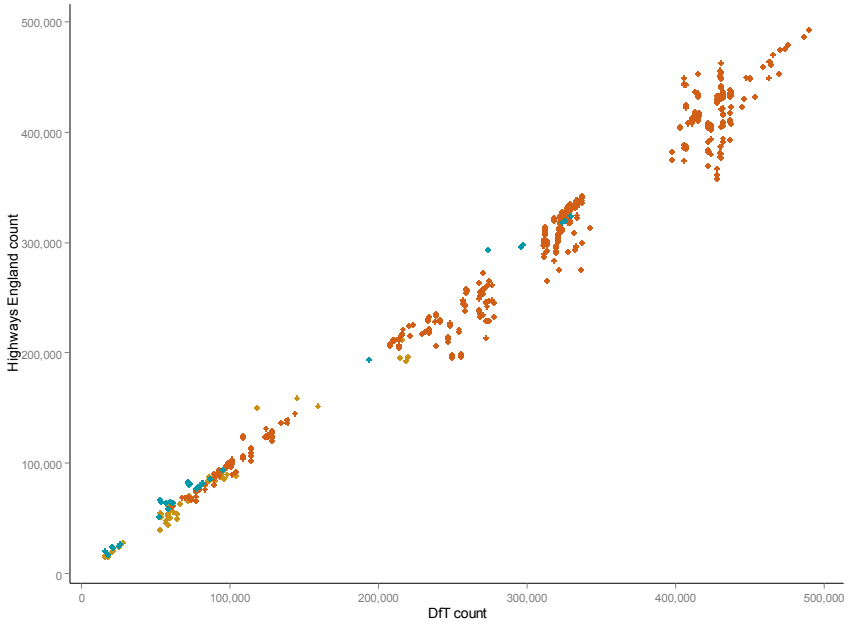
4.14 This second phase of data cleaning was resource intensive, as it required considerable manual inspection of the datasets. For future roll-out, this would be a substantial set-up resource for reviewing the full Highways England dataset of over 10,000 sites.

4.15 Using this cleaned count data, a final comparison was made across the datasets.

Comparing DfT and Highways England ATC data

4.16 In Figure 8, each point corresponds to a week's data at one Highways England site matched to the corresponding data for the DfT site. The different colours represent the different types of HE counter. A reasonable match can be seen between Highways England data and DfT data, after the second phase of data cleaning.

Figure 8: Weekly vehicle counts at matched Highways England and DfT traffic statistics' automatic traffic counter sites[3]



[3] Different colour dots represent a different type of Highways England traffic counter.

Comparing DfT manual count data to Highways England's ATC data

4.17 Figure 9 shows the hourly data from Highways England counters compared against the matched hourly manual counts. There are some Highways England counters that are undercounting compared to DfT manual counts, implying that a further stage of data cleaning and matching is needed on the data.

Figure 9: Hourly vehicle counts at matched Highways England and DfT traffic statistics' manual count sites[4]

Peer Review

4.18 The ONS Methodology Advisory Service (MAS) peer reviewed the analyses and supported the conclusions.

4.19 MAS also conducted feasibility analysis into estimating the percentage split of traffic by vehicle type using DfT counter vehicle type splits and Highways England ATC totals. The reason for this additional analyses was because, as outlined above, Highways England ATCs don't classify vehicles in the same way as DfT traffic statistics. The analysis found the confidence intervals to be quite small, which suggests that it should be possible to derive estimates of traffic by vehicle type and by road type by combining the two data sources.

Conclusion

4.20 Highways England's automatic traffic counter data would be a useful additional data source, subject to obtaining the resources required to match the network of counters to the road network used for the DfT road traffic estimate calculation.

4.21 Since this project was completed, DfT have undertaken work to improve the data cleaning and the matching between the DfT network used and Highways England's ATC locations. As a result, DfT have included used the total count data from Highways England (and Transport for London) ATC data in the publication "*2017 Road Traffic Estimates for Great Britain*".

[4] Different colour dots represent a different type of Highways England traffic counter.

5. Conclusions and next steps

5.1 This note has provided an overview of the work carried out in 2015-16 on the alternative data project of the Department for Transport's review of the traffic estimates for Great Britain. A summary of the conclusions and next steps is set out below.

Road-side data sources

5.2 Some of the technologies tested showed promising accuracy of total vehicle counts, however, none provided the accuracy of vehicle classification required. Therefore, none were deemed suitable for the wholescale replacement of the existing automatic traffic counters and manual counts combination for DfT traffic statistics production.

5.3 The DfT team will explore the potential for the use of the total count data (i.e. without vehicle classification) from the technology types tested to estimate traffic statistics for all motor vehicles only.

5.4 The team will continue to investigate new technologies under development that may have applications for traffic counting in the future. Examples that have arisen since the completion of this project include new video analytics technologies, and drone or satellite images.

Non-road-side data sources

5.5 The methodology trialled for synthesising road traffic estimates was not deemed reliable for large-scale implementation at this time. The team will continue to investigate if it could be used for individual road links, such as those where it is difficult or dangerous to conduct traffic counts.

5.6 The development of the use of odometer readings from MOT test data to estimate road traffic continues, with experimental statistics currently planned for 2019.

Data sharing with other organisations

5.7 Automatic traffic counter data from other organisations is a useful additional data source.

5.8 Since this project was completed, DfT have included Highways England, Transport Scotland, and Transport for London ATC data in the publication "*2017 Road Traffic Estimates for Great Britain*".