Department for Transport

Traffic Statistics Methodology Review

Quarterly Methods

Moving Britain Ahead

February 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 gives a detailed overview of one part of the review - investigating the methodology used to produce quarterly traffic estimates - and sets out the improvements identified. For a short summary of this project, its conclusions and all of the other projects within the review please refer to the Overview document[1].

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.

[1] Available at www.gov.uk/government/statistics/road-traffic-statistics-methodology-review

Executive summary

The Department for Transport (DfT) publishes estimates of traffic on Great Britain's roads[2]. Provisional estimates of road traffic in Great Britain are published quarterly. The Road Traffic Statistics Team have been investigating improvements to the methodology used to produce these estimates as part of a wider review of road traffic statistics.

This report presents the results of the Quarterly Methods project, which reviewed the methodology applied to produce the estimates. All aspects of this calculation process were reviewed by DfT Statisticians and an external statistical expert. Whilst the overall methodological approach was deemed suitable, potential improvements were identified for certain stages of the calculation process.

The improvements recommended can be summarised as:

- To apply an annually rebased methodology to the calculation of the preliminary traffic estimates.
- To simplify the method of constraining the preliminary figures to the final annual traffic estimates.

In testing, the impact of applying these improvements to the quarterly traffic estimates was relatively small, however, these recommendations result in significant improvements in the efficiency, flexibility, and transparency of their production.

The traffic statistics team are implementing the above methodological improvements into their processes. The results of this work are expected to be presented as part of the publication of *Provisional road traffic estimates for the year to end March 2018*, which is due to be released in July 2018.

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

1. Current Methodology

1.1 Provisional traffic estimates are published every quarter for the 12-month period up to the end of the quarter. These estimates are currently derived by using a network of around 200 automatic traffic counters to measure changes in traffic by vehicle type and road type.

1.2 The sample of automatic traffic counters is stratified by area, road type, and traffic flow levels. This stratification of the sample (called 'EFCat') was reviewed in a separate project, and a report on the results has been published[3]. The validation and imputation methods have also been reviewed separately, with the results included in the Overview Report for the Methodology Review[3].

1.3 Once data has been cleaned and validated, the calculation process begins. It is the review of the calculation process that is the subject of this report.

Current quarterly calculation process

1.4 The current calculation process has seven stages, and is set out in detail in Figure A. The stages of the process are:

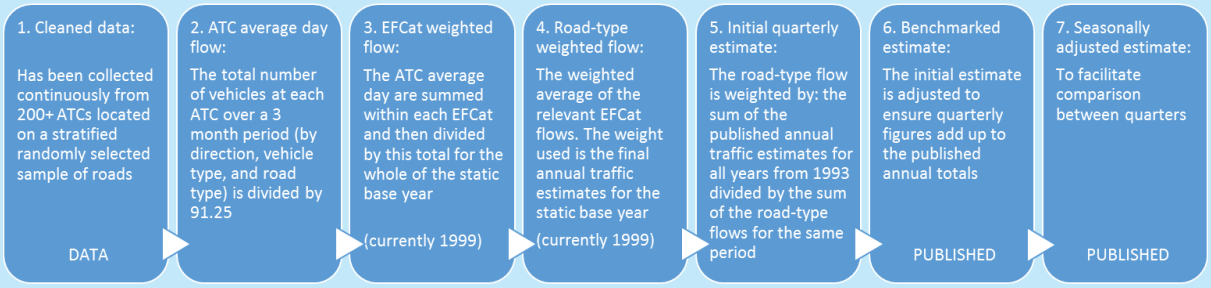
- Stages 1-2: aggregate each quarter's sample of traffic data up to the stratification level
- Stage 3: calculate the change in traffic levels since the reference year (the reference year is currently static and set as 1999)
- Stage 4: aggregate the change in traffic levels up to 'road type' level
- Stage 5: apply weights (currently based on the annual traffic levels in the reference year) to produce a 'preliminary traffic estimate' for the quarter
- Step 6: constrain the preliminary traffic estimates, so that the quarterly traffic estimates now sum to the final annual figures[4]. In the current methodology, this calculation is performed in the software 'Inter'.

1.5 An additional stage is applied at the end of this process to seasonally adjust the quarterly estimates, to facilitate the comparison across different quarters. The seasonal adjustment software 'x13 ARIMA SEATS' is used for this. This element of the process was reviewed previously to meet ONS good practice. Therefore, it is not included in this report.

[3] Overview Report and Stratification Report are available at www.gov.uk/government/statistics/road-traffic-statistics-methodology-review

[4] Annual traffic estimates are produced using a different process, using a large sample of roadside manual counts as a data source. Annual traffic estimates are regarded as more robust than the quarterly estimates. Therefore, when final annual traffic estimates are available for a calendar year (usually in May of the following year), the provisional quarterly figures for that year are revised ("benchmarked") so that they sum to the annual total. These revised figures are the final unadjusted quarterly estimates.

Figure A: Current process of calculating quarterly road traffic estimates



For each vehicle type:

- 2: If, for each ATC site, we have ATC counts z_i , the first step is to calculate the average daily flow in the quarter q by dividing the ATC count by 91.25. These are then totalled up for all the ATCs within each EFCat e.

$$x_{eq} = \sum_{e} \frac{\sum_q z_i}{91.25}$$

- 3: The next step is to create a ratio of the ATC daily totals for the current quarter, for EFCat e, to that for the annual total for 1999, multiplied by 4 to scale the quarter up to a year so that it can be compared to the annual total for 1999.

$$r_{eq}^{y,99} = 4 \times x_{eq} / \sum_{\varphi \in 1999} x_{e\varphi}$$

- 4: These ratios r are then multiplied by the 1999 EFCat weight, added up for each road-type (for all road types there is more than one EFCat), and then divided by the sum of the 1999 EFCat weights for that road type to produce a weighted ratio for each road category c.

$$r_{cq}^{w} = \frac{\sum_c W_e^{99} r_{eq}}{\sum_c W_e^{99}}$$

- 5: These are then used to create a 'Preliminary scaled up estimates' for this road-type using the existing annual totals A_{cy}:

$$Q_{cq}^{P} = r_{cq}^{w} \frac{\sum_{y'=1993}^{y'=y(q)-1} A_{cy'}}{\sum_{\varphi=1993q1}^{\varphi=(y(q)-1)q4} r_{c\varphi}^{w}}$$

The above formula is taking the total traffic for all years from 1993 to the previous year and scaling that by the sum of the weighted ratios for all the quarters of those years multiplied by the weighted ratio for the new quarter.

- 6: This is then modified to produce the publication estimates:

$$Q_{cq}^{F} = Q_{cq}^{P} + I_{cq} + \frac{1}{4} B_{cy(q)}$$

where I_{cq} is the adjustment output from the benchmarking program Inter, and

$$B_{cy(q)} = A_{cy} - \sum_{\varphi \in y} (Q_{c\varphi}^{P} - I_{c\varphi})$$

which is the extent to which the benchmarked quarterly preliminary estimates differ from the existing annual totals. For the latest year, where the final annual benchmark is not yet available, the Inter adjustment, I_{cq} is predicted by extrapolation from recent change. The latter adjustment, $B_{cy(q)}$ is zero for the current year. All quarters back to 1993 may be modified by this process.

- 7: The software 'x13 ARIMA SEATS' is used to produce seasonally adjusted outputs.

2. Office for National Statistics review

2.1 The Methodology Advisory Service at the Office for National Statistics (ONS), reviewed the current quarterly calculation methods and investigated alternative approaches for deriving the provisional quarterly traffic estimates. The findings from this review are set out in this section.

Stages 1 to 2 – Calculating stratification-level average daily flows

2.2 The first steps of the calculation process derive the average 24-hour traffic count for a given automatic traffic counter (ATC). These are derived for each of the three-month periods, and for each vehicle type. These average 24-hour traffic counts are then added together for all ATCs within each stratum.

2.3 **Review Finding:** No improvements were identified to these current steps for aggregating the sample of traffic data up to the stratification levels.

Stage 3 – Calculating stratification-level weighted ratios

2.4 The next step is to calculate the change in stratification-level flows. This ratio is calculated by dividing the current quarter's flow with the equivalent figures for the reference year. In the current methodology the reference year is static, set as 1999. This calculation is carried out for each stratum and each vehicle type. The resulting ratios are then multiplied by 400 to scale up the quarter to represent a full year.

2.5 The use of a static reference year (1999) in the ratio calculations was highlighted early on in the investigation as an area for potential improvement. The use of a static reference year introduces inflexibility to the sample of ATCs, as it has historically required the same sample of sites to have data for the current quarter in question, and in 1999.

2.6 The Department's sample of ATCs can change over time, for example, due to road layout and road classification changes, or because of equipment fault and/or 'downtime'. In addition, using a static reference year may also become more of an issue the further the series moves away from it, in terms of the level of adjustment required to ensure the accuracy of estimates.

2.7 **Review Finding:** The ONS review recommended an alternative ratio calculation: updating the reference year used every year. For example, instead of using data from 1999 to create the index in ratio calculation, it is possible to use data from the previous year (Yr – 1).

2.8 This would change the calculation equation

From: $r_{eq}^{y,99} = 4 \times x_{eq} / \sum_{\varphi \in 1999} x_{e\varphi}$ to: $r_{eq}^{y,y-1} = 4 \times x_{eq} / \sum_{\varphi \in y-1} x_{e\varphi}$

2.9 The main advantages for this change:

- It would allow a different set of ATC sites to be potentially used every quarter, requiring a contiguous period of 2 years of valid data from each ATC site to calculate the ratio. This in turn would mean that data from all functioning ATCs could be used in any given quarter, increasing the sample size. The increased flexibility of the ATC sample would make it relatively simple to introduce newly installed sites or remove non-functioning sites from the sample.
- It would reduce the amount of imputation necessary, as faulty ATC sites could be excluded from the sample. Under the current method, flow for temporarily or permanently broken ATCs in the fixed sample must be imputed. The result therefore would be more robust estimates.
- It is likely that it would reduce the level of adjustment needed between provisional and final annual estimates, increasing end-user trust in the statistics.

Stages 4 and 5 – Calculating preliminary quarterly traffic estimates

2.10 Once stratification-level ratios have been produced, weights are applied to gross-up the sample data and derive preliminary traffic estimates at the road-type level that are representative of all roads in Great Britain.

2.11 In the current methodology, two sets of weights are used:

- Final annual estimates for 1999 when aggregating the data up to road-type level.
- One calculated using the data for all years from 1993 to the latest available full calendar year i.e. comparing the sum of the final annual compared to the sum of the provisional estimates for this time period. This is applied to the road-type level figures to produce the preliminary estimate.

2.12 Figure B illustrates the make-up of the current road-type stratum.

Figure B: Current methodology: Road-types and stratification-level lookup table

Motorways	01	**Motorways** in holiday areas
	02	**Motorways** in other rural areas with an estimated AADF of up to 59,999
	03	**Motorways** in other rural areas with an estimated AADF of 60,000 or more
	04	**Motorways** in part rural and part urban areas and conurbations
	05	**Motorways** in mostly urban areas and Greater London
Rural A roads	06	**Rural A roads** in holiday and very rural areas with an estimated AADF of up to 4,999
	07	**Rural A roads** in holiday and very rural areas with an estimated AADF of between 5,000 and 7,999
	08	**Rural A roads** in holiday and very rural areas with an estimated AADF of 8,000 or more
	09	**Rural A roads** in all other areas with an estimated AADF of up to 13,999
	10	**Rural A roads** in all other areas with an estimated AADF of 14,000 or more
Urban A roads	11	**Urban A roads** in holiday areas
	12	**Urban A roads** in all other areas except Greater London with an estimated AADF of up to 19,999
	13	**Urban A roads** in all other areas except Greater London with an estimated AADF of 20,000 or more
	14	**Urban A roads** in Outer London
	15	**Urban A roads** in Inner London
	16	**Urban A roads** in Central London
Rural minor roads	50	**Rural minor roads** in holiday areas with an estimated AADF of up to 399
	51	**Rural minor roads** in holiday areas with an estimated AADF of 400 or more
	52	**Rural minor roads** in all other areas with an estimated AADF of up to 2,499
	53	**Rural minor roads** in all other areas with an estimated AADF of 2,500 or more
Urban minor roads	54	**Urban minor roads** in all areas except Greater London
	55	**Urban minor roads** in Greater London

2.13 **Review Finding:** Recommendations for improvements to both stage 4 and stage 5 were also highlighted in the ONS initial report. The two recommendations were:

- To replace steps 4 and 5 with a single step. This single step would use a simpler formula to produce preliminary estimates of traffic on all roads in Great Britain at the stratification-level directly.
- To weight the stratification ratios using the finalised annual traffic figures from the most recent year available only (i.e. the previous calendar year).

2.14 The resultant effect would place greater emphasis on more recent traffic estimates. This would help ensure that the sample size and representativeness of ATC sites is maintained (as the closer these are to each other in time, the smaller the impact in ATC turnover).

Stage 6 – Constraining quarterly traffic estimates to sum to final annual traffic estimates ("Benchmarking")

2.15 The annual traffic estimates produced by the Department are more robust than the provisional quarterly estimates. However, the sum of the four provisional quarterly traffic estimates for any given vehicle and road type will be slightly different to the annual figure. In order to maintain consistency between published quarterly and annual statistics, the quarterly figures in a given calendar year are adjusted up or down to sum to the final annual total. This adjustment is called "benchmarking".

2.16 The provisional quarterly estimates in a calendar year are benchmarked as soon as the annual traffic estimates are produced for that year, usually in May of the following year. At that point all four quarters go from being provisional estimates to final estimates.

2.17 The main focus of the review of the benchmarking process in the initial investigations by ONS, was the method and software used to implement it. The current benchmarking method uses a bespoke piece of software called INTER. This software was produced by ONS, and uses a regression model to apply quarterly adjustments which vary smoothly over time – i.e. it incorporates a degree of temporal autocorrelation into the adjustments applied.

2.18 There are several issues with the current use of INTER for benchmarking:

- The software is a black box. The team were not able to discover any documentation explaining exactly how it works. This introduces opacity to the methodology.
- It doesn't fit easily into the rest of the quarterly estimation process. For instance, files must be passed between programs.
- The outputs from INTER have 2 decimal places, so require a further adjustment to exactly match annual figures.
- The software is no longer supported. It cannot be maintained or repaired in-house or by the original developers.

2.19 The desirable features of a benchmarking method in traffic statistics are that it is:

- Understandable
- Well documented and "future proof"
- In keeping with GSS good practice

- Integrated into the rest of the process
- An appropriate method to give good results with traffic data
- A method using proportional adjustments, as it is more likely that traffic will be under- or over-estimated in quarterly figures by a certain percentage, rather than an absolute value

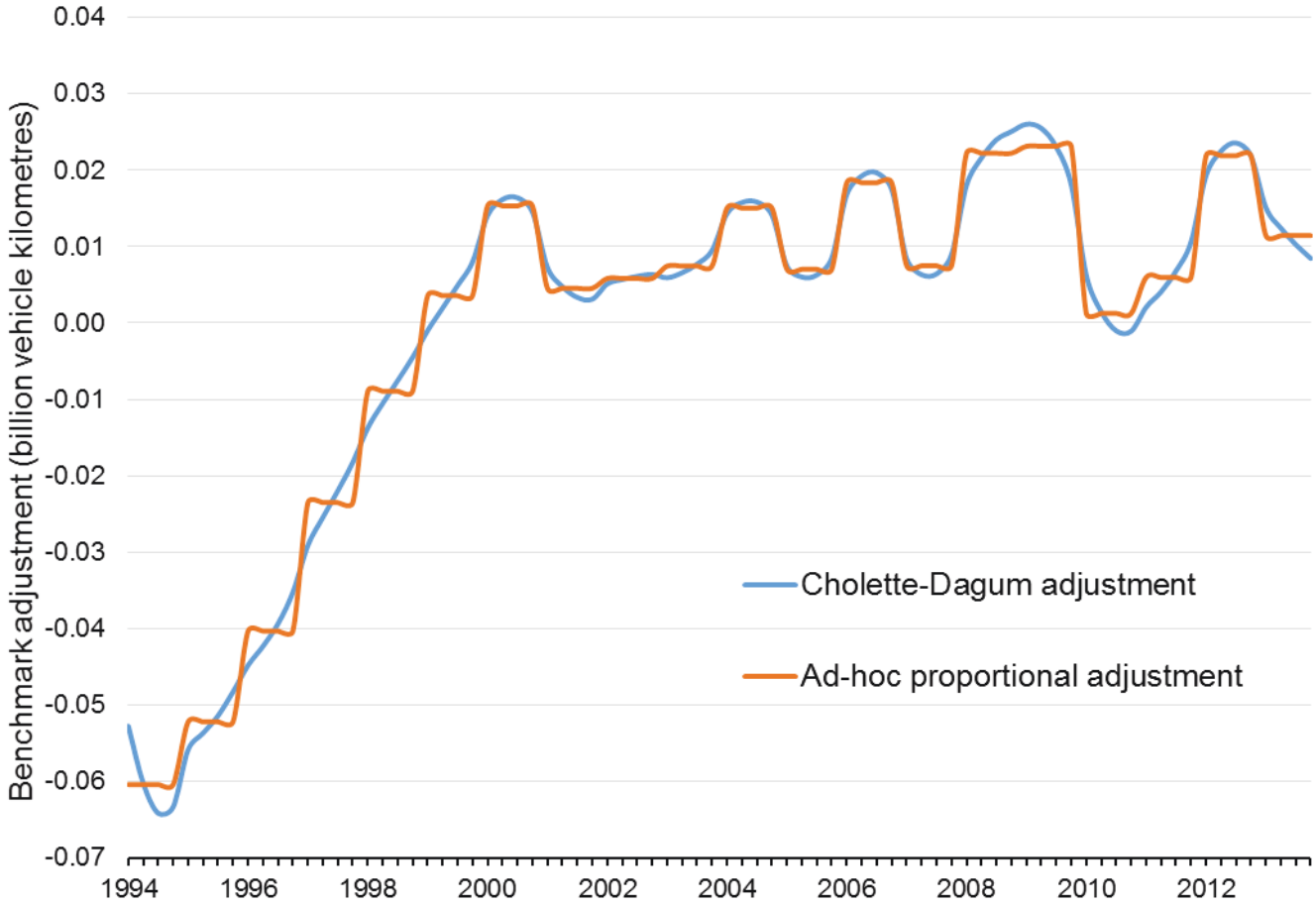
2.20 Two potential alternatives to the current method were tested in the initial investigations by ONS: an ad-hoc proportional adjustment, and a method of proportional adjustment developed by Cholette and Dagum using a regression model.

2.21 A comparison of the methods, advantages and disadvantages of the two alternatives are discussed below:

	Ad-hoc proportional adjustment	**Cholette-Dagum method**
Method	Finds the ratio of the final annual estimates and the sum of the four quarterly estimates – i.e. Annual total divided by the sum of the quarters. Each quarter is then multiplied by this ratio, producing quarterly estimates which sum to the annual total. As each quarter within a calendar year is adjusted by the same multiplication factor, there may be a large difference between the adjustment applied to the fourth quarter of one year and the first quarter of the next, producing step-like patterns of adjustments (see Figure C)	Fits a regression model through a time-series of the required total annual adjustments in percentage terms, ensuring that the fitted quarterly adjustments within a calendar year sum to the required annual adjustment in that year. The model is fitted assuming a correlation between one quarterly adjustment and the subsequent one – even when going from one year to the next – resulting in a smoothly varying adjustment over time (see Figure C)
Advantages	i. Can be implemented very easily and quickly in the same software as the rest of the process ii. Simple to explain and understand for traffic statistics users iii. Already in use elsewhere in DfT	i. Used by ONS for some time-series ii. Available as source code implemented in "R", so replicable and future proof iii. Takes into account autocorrelation between neighbouring quarterly deviations
Disadvantages	i. Less sophisticated than the C-D method. ii. Can produce unrealistic steps in the adjustment to be applied (see figure C)	i. Not easily explained, and not easily understood by most users of traffic statistics ii. Requires export and import of files between R and other programs

2.22 The two methods were tested by applying them to the preliminary traffic estimates for cars on motorways. Figure C illustrates the step-like pattern of adjustments that were applied over the time series with the ad-hoc proportional method, compared to the smoothly varying adjustments applied with the Cholette-Dagum method.

Figure C: Benchmark adjustment for car traffic estimates on motorways, using the Cholette-Dagum and the ad-hoc proportional methods



2.23 **Review Finding:** The review advised that the Inter software product should no longer be used as it is not supported, and that the ad-hoc proportional method be favoured as the replacement "benchmarking" method due to its simplicity and ease of integration. However, if the results of the two methods differ markedly when all vehicle types and road types are tested, then the ad-hoc proportional method could be considered too simplistic for the traffic time series, and the Cholette-Dagum method would then be preferred.

3. Testing methodology improvements

3.1 The review findings set out in the previous chapter were based on an analysis of car traffic on motorways. In order to determine the final recommendation(s) to improve the quarterly calculation methods, further testing was required for all other vehicle types and road types.

3.2 These tests sought answers to two key questions:

- How 'close' are the 'new' preliminary traffic estimates based on the recommended annually rebased method to the final annual estimates?
- How do the 'new' preliminary estimates produced by the annually rebased method compare to those produced by the current method?

3.3 The accuracy of both methods was tested by calculating, for each vehicle type and road type individually, the difference between the final annual traffic estimate in each year and the total of the four preliminary quarterly estimates in the same year. Both the current method and the annually rebased method were tested using data over a 20-year period (1993 – 2013) from the same (fixed) set of automatic traffic counters.

Stages 3 to 5 – Producing preliminary traffic estimates

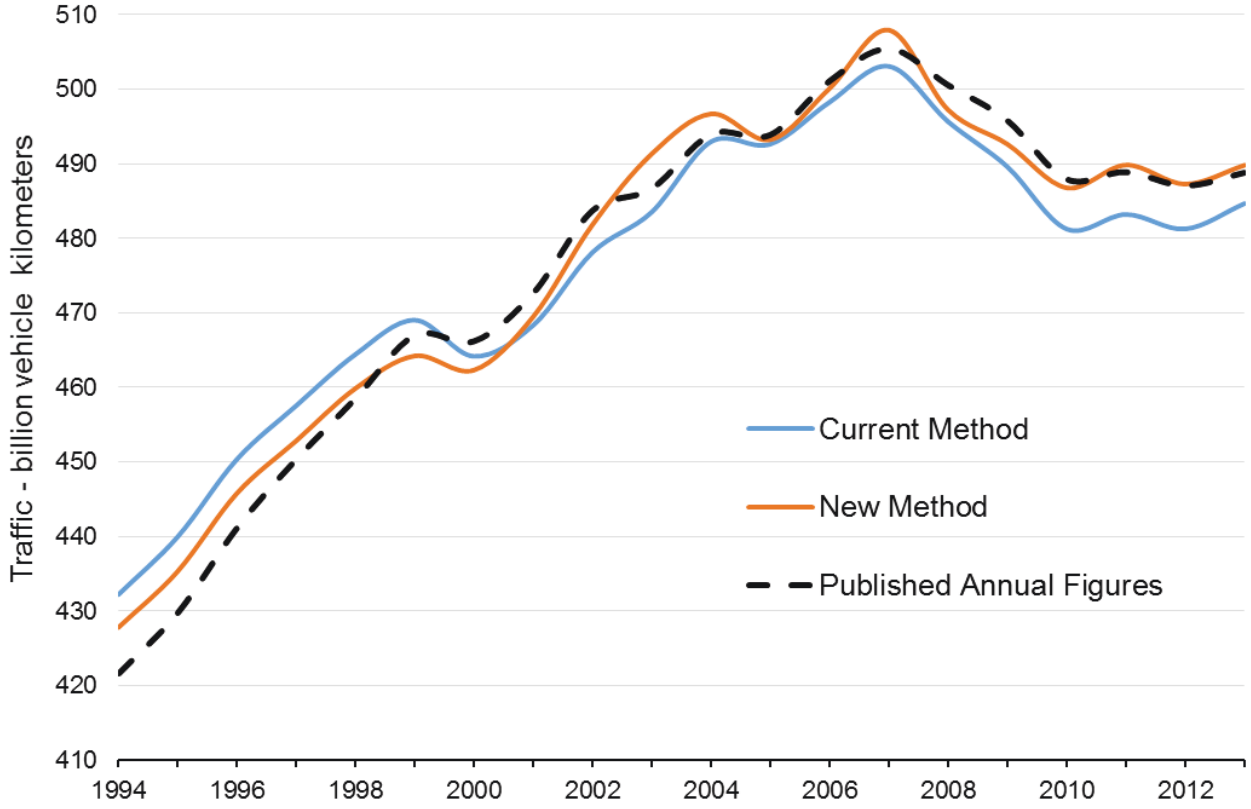
3.4 The key findings noted by ONS in their initial review, in relation to these three stages of the calculation process, were:

- Stage 3: Use ATC data from the previous year (Yr – 1) to calculate the change over time ratio at stratification level, instead of using a static reference year (1999).
- Stage 4 and 5: Combine and simplify the calculation of the preliminary quarterly estimates into a single process, using weights derived from only the most recent complete year's data.

Test results

3.5 Overall, the preliminary estimates of quarterly traffic produced using annual rebasing were closer to the final annual figures than those produced by the current methodology (see Figure D).

Figure D: Comparison of the methods for calculating preliminary estimates



3.6 The differences between the 'new' annually rebased preliminary estimates and the final annual estimates showed no strong trends over time (Figures E and F), instead fluctuating around 0. In contrast, the differences between the current preliminary estimates and final annual estimates show significant positive or negative trends over time, depending on vehicle and road type (see Figures E, F and G).

Figure E: Adjustment required by the methods (level)

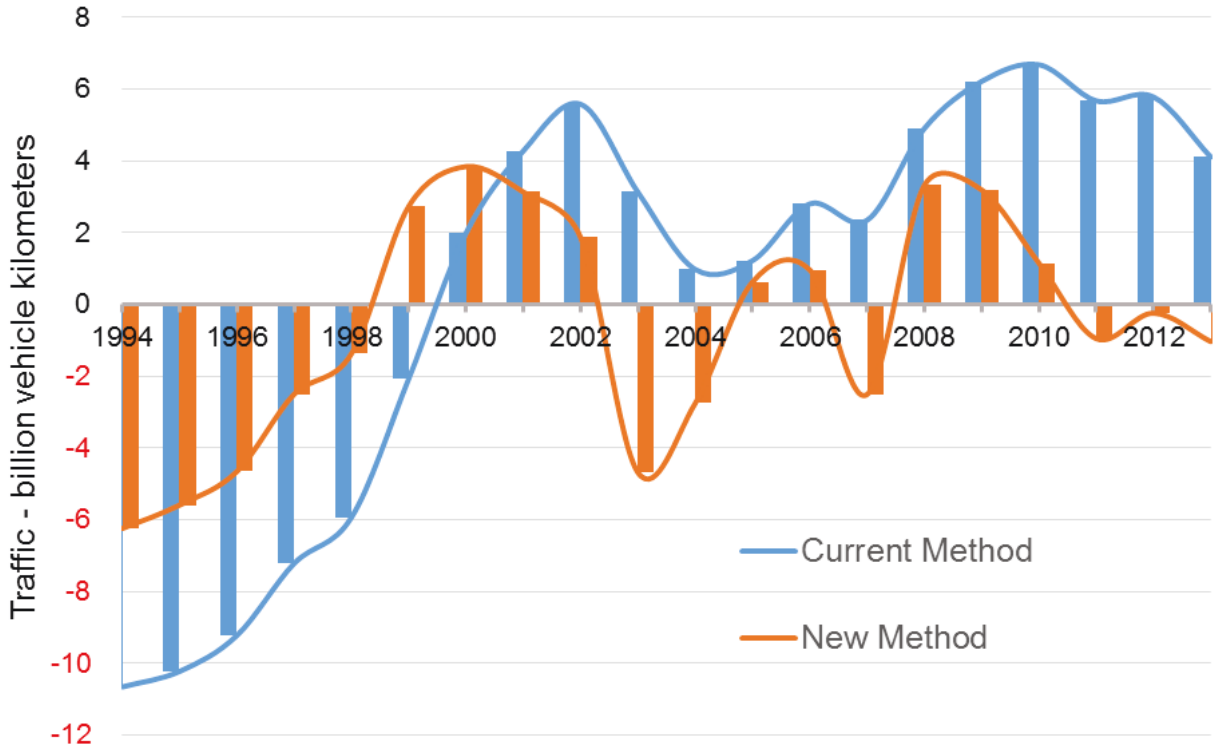
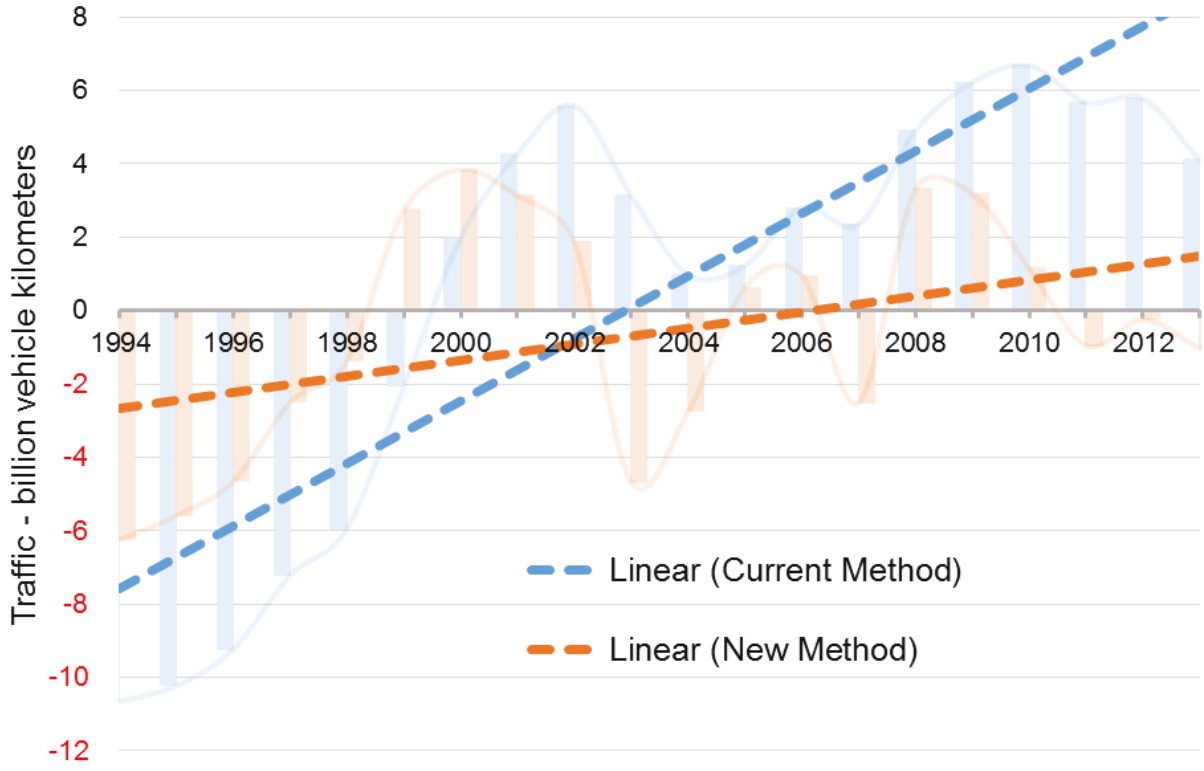


Figure F: Adjustment required by the methods (trend)



3.7 When comparing the annually rebased and current methodologies for each vehicle type and road type combination (Figure G), it is clear that in 18 of the 20 combinations the 'new' annually rebased method produces preliminary estimates closer to the final annual estimates.

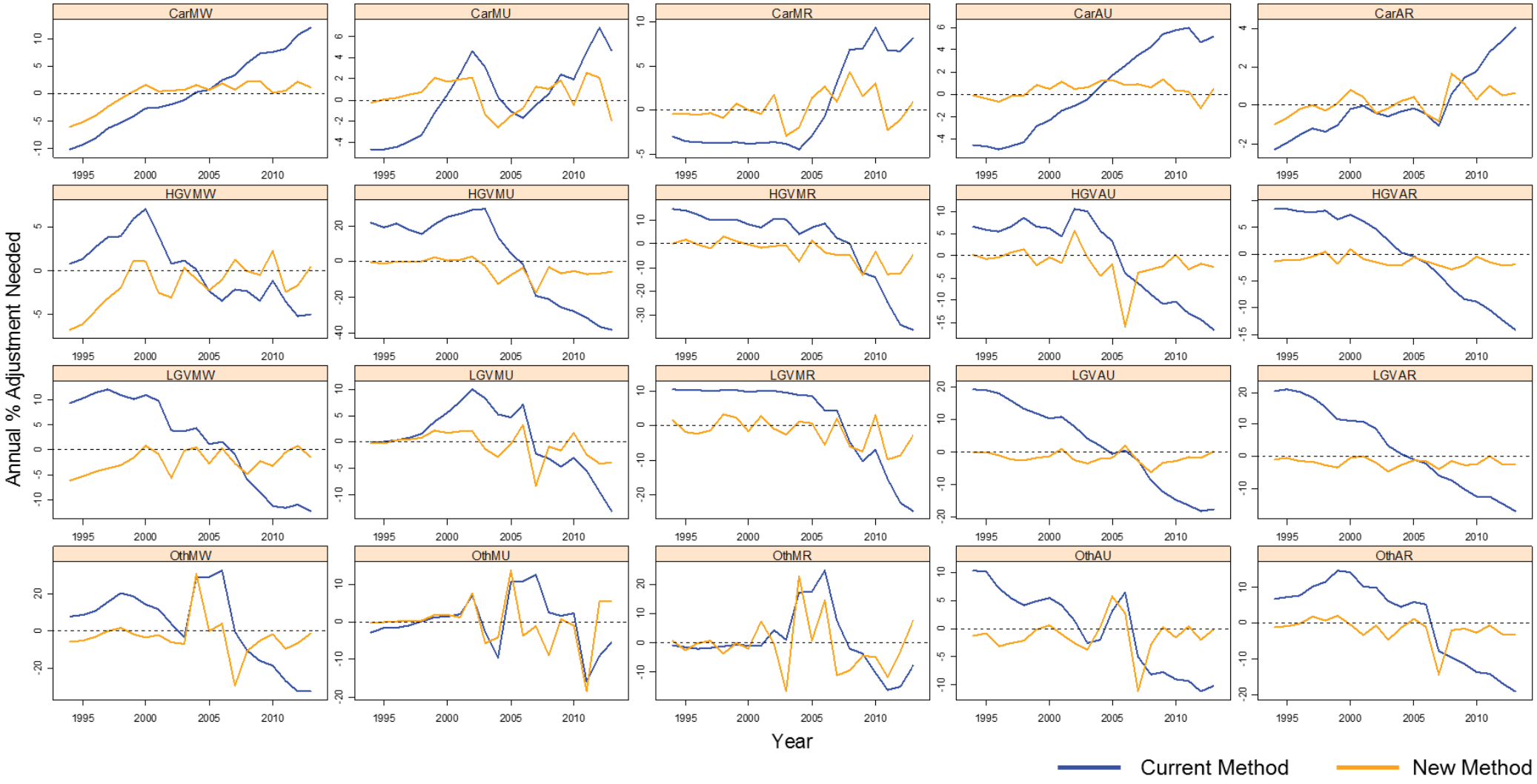
3.8 For two specific combinations - "other vehicles" (buses and motorcycles combined) on urban minor roads and on rural minor roads - the current method performs as well or better (see Figure G). However, it is important to note that other vehicle types on minor roads made up less than 1% of all traffic over the 20 year period (1993-2013) studied. Given this very small contribution to overall traffic levels, the comparatively poor performance of the annually rebased method for these vehicle type and road combinations does not affect the overall conclusion.

Recommendation

3.9 The proposed changes to the calculation methods should be implemented as they produce provisional estimates that are closer to final annual estimates than current methods.

3.10 The proposed changes simplify the calculation process, as they require fewer assumptions and making it more transparent for users. They also give an operational advantage of improved flexibility.

Figure G: Adjustment required for the methods, in percentage terms, for all vehicle and road type combinations



Stage 6 – Constraining quarterly traffic estimates to sum to final annual traffic estimates ("Benchmarking")

3.11 The key finding noted by ONS in their initial review, was that the simpler ad-hoc proportional method of "benchmarking" was preferred. However, this was based on analysis of car traffic on motorways only. An extension of the analysis to include all other vehicle types and road types was needed for a final recommendation.

3.12 ONS's guidance was that the choice of benchmarking method for traffic statistics depended on a single factor: how large are the differences between the results of the more sophisticated Cholette-Dagum method and the simpler ad-hoc proportional method?

3.13 Both benchmarking methods were applied to preliminary estimates of traffic for each of the 20 vehicle type and road category combinations between 1994 and 2013. The preliminary estimates used for this analysis were calculated using the annually rebased methodology described in the first section of this chapter.

Test results

3.14 The difference in final benchmarked estimates arising from the two methods was, in general, small. In 17 out of the 20 series, the maximum difference for any single rolling annual traffic total was less than 2%, and the average difference between rolling annual totals was less than 1% in all series (see Figure H).

3.15 Differences were greater for "other vehicle" (bus and motorcycle) traffic on motorways and minor rural roads, which are seasonally affected. However, as they make up around 0.5% of total traffic, they do not affect the overall pattern (very small differences between the two benchmarking methods).

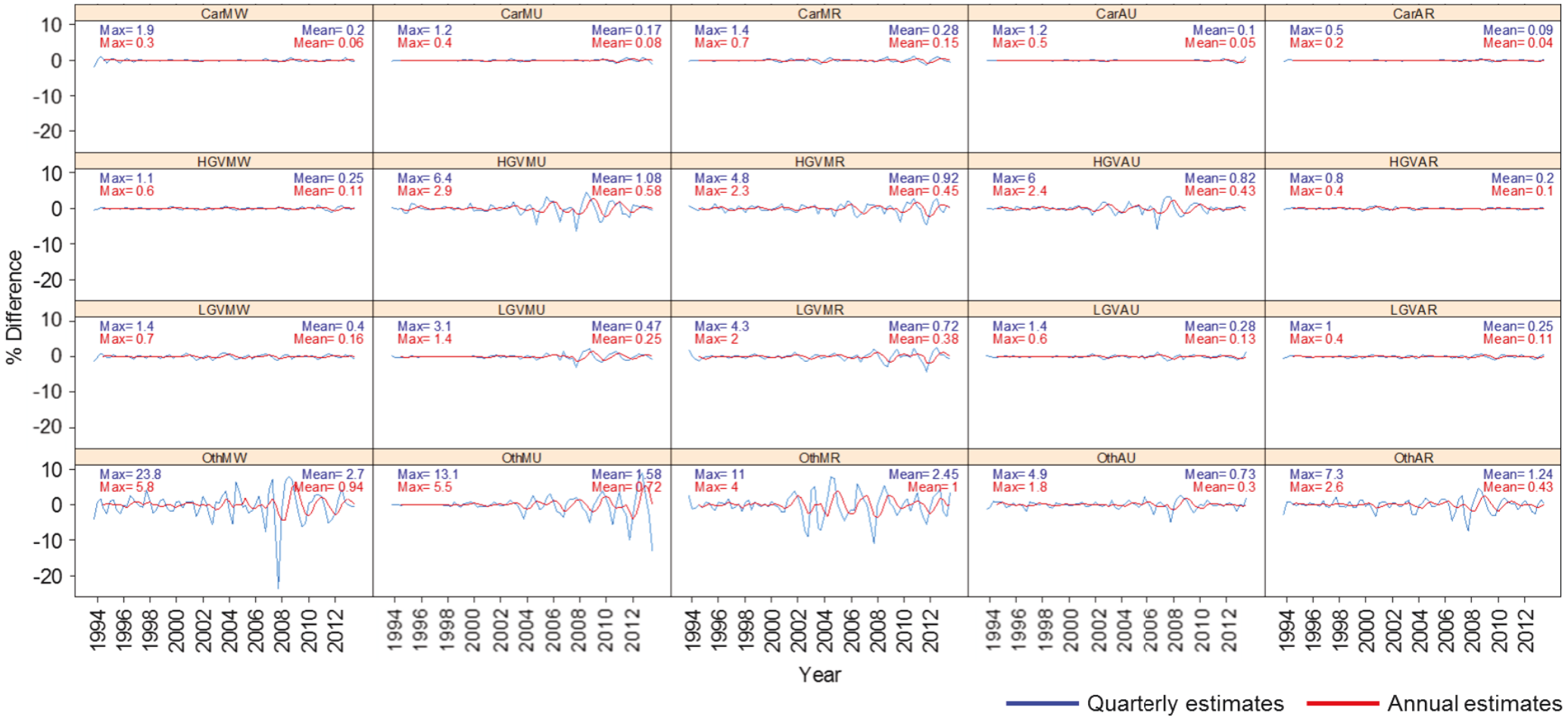
3.16 For all vehicle types and road types combined, the maximum difference for any single rolling annual total was 0.34%.

Recommendation

3.17 The analysis showed that the benchmarked quarterly and rolling annual traffic estimates produced by the Cholette-Dagum and ad-hoc proportional methods were very similar. This suggests that there is no clear benefit that the additional complexity the Cholette-Dagum method brings for benchmarking traffic time series.

3.18 In light of the advantages of the ad-hoc proportional method in terms of the ease of integration into the rest of the process and simplicity, it is recommended that this benchmarking process is adopted.

Figure H: Percentage difference between Cholette-Dagum benchmarked estimates and ad-hoc proportional benchmarked estimates, for each vehicle type and road category

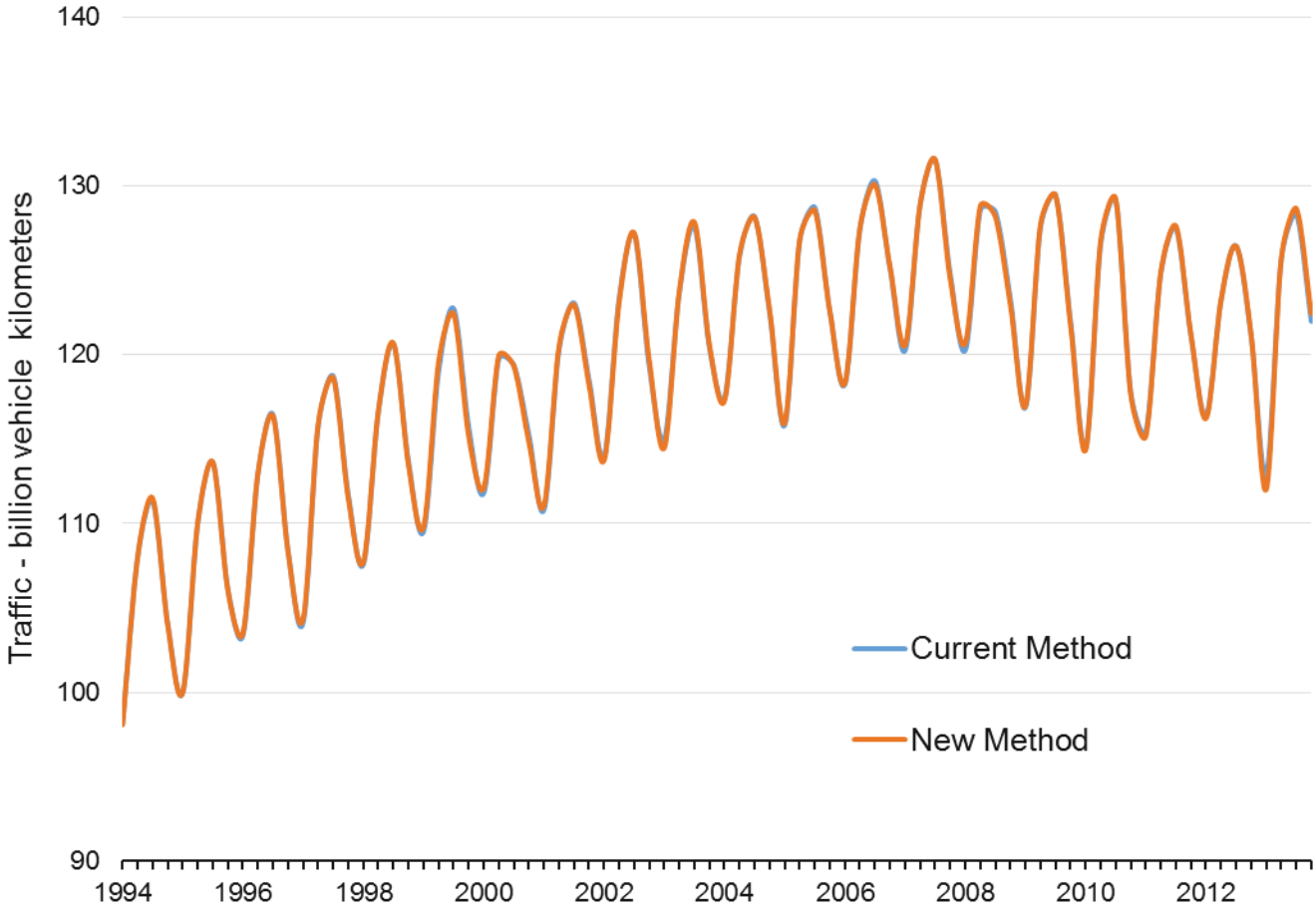


4. Conclusions and next steps

4.1 In the previous chapter, analyses relating to the calculation of preliminary estimates and the benchmarking method were presented separately. This leads to the question: how do quarterly traffic statistics calculated incorporating both of the recommendation compare with those calculated using the current methodology?

4.2 A comparison was made of the quarterly estimates calculated via the current methodology and estimates calculated via the recommended method. As can be seen in Figure I, the differences between the final quarterly statistics were negligible.

Figure I: Comparison of quarterly estimates for all motor vehicle traffic



4.3 The very small impact of the recommendations on final quarterly and rolling annual traffic estimates is reassuring. The recommendations outlined above would result in significant improvements to the flexibility and transparency of the process of producing provisional quarterly traffic estimates, without having major impacts on traffic estimates themselves.

Next steps

4.4 The traffic statistics methodology review is wide-ranging. Other areas of development that may benefit the production of provisional quarterly estimates include:

- The new roads stratification, containing fewer strata defined by slightly different criteria to the current stratification
- An upgrade of ATC sites in London.
- Development of a new method of data cleaning and imputation for ATC data
- Development of a new processing system based in SQL code

4.5 The traffic statistics team are implementing the above methodological improvements into their processes. The results of this work are expected to be presented as part of the publication of *Provisional road traffic estimates for the year to end March 2018*, which is due to be released in July 2018.