Department for Transport

Traffic Statistics Methodology Review

Stratification Project

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
General enquiries https://forms.dft.gov.uk
Website www.gov.uk/dft

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 the stratification project which was undertaken as part of the review, including statistical methods and results. It has been updated since original publication in August 2015 to reflect the results of implementing the new stratification in the July 2018 road traffic estimate publications.

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

Using this Document

This document is laid out in chronological order of the analysis carried out, with subsections for each variable considered. At the end of each chapter is a summary of what was concluded from that analysis, some variables are revisited during each piece of analysis until a firm conclusion is reached.

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

Executive summary

The Department for Transport (DfT) publishes annual and quarterly estimates of traffic on Great Britain's roads here.

DfT traffic estimates are based on observations of traffic on samples of road links, defined as a stretch of road between two junctions. The traffic observations are collected by a panel sample of Automatic Traffic Counters and by an annual sample of around 7,000 manual counts.

A fundamental component of the traffic estimation process is the stratification of the sample of Automatic Traffic Counters. The aim of stratification is to minimise sampling error by grouping together roads with similar traffic patterns.

The main aspect of this project was to explore whether the current stratification groups are the best categorisation to group together road links with similar traffic flow patterns. The approach taken by the project was to look at the effects of certain road attributes on fluctuations in traffic flow. Those attributes found to affect traffic flow could be used as divisors for the stratification categories.

This document provides a detailed overview of the stratification project, including statistical methods and results.

Current Stratification

The current stratification was introduced in the early 2000s. It comprises of a large number of categories, some of which have quite a small sample of Automatic Traffic Counters (ATCs) within them meaning that estimates for these categories can be quite noisy.

Analyses

The project used data from the DfT's network of 200 ATCs, allowing comparisons across different times of the day, different times of year and between different years to be made.

Exploratory analyses aimed to determine what road characteristics are important with regards to variations in traffic flow. The analyses are split into two main sections: the first looked at variations within the year, which relates to expansion factors; and the second at variations between years, which related to growth factors.

The analysis investigated the level of variation in traffic observed at each ATC, as well as variations between day and night, weekday and weekend, seasonal variations and variations in flow between years.

All analyses were peer reviewed by an external methodological expert from the Office for National Statistics.

Findings

The analyses found that different road attributes were significant for within year variation to between year variations. Therefore, two different stratifications are necessary.

The next step is to allocate each of the ATCs to a stratum and calculate expansion factors and growth factors and assess the effect that these have on overall traffic estimates.

The recommendation from the project is a stratification that can be aggregated up in two different ways for the calculation of expansion factors and growth factors. This is set out in Box 1 below.

Box 1: Road Traffic Statistics Stratification

01 Trunk Motorway
02 Principal Motorway
03 Trunk Urban 'A' roads <20,000
04 Trunk Urban 'A' roads >20,000
05 Trunk Rural 'A' roads <20,000
06 Trunk Rural 'A' roads >20,000
07 Principal Urban 'A' roads <20,000
08 Principal Urban 'A' roads >20,000
09 Principal Rural 'A' roads <20,000
10 Principal Rural 'A' roads >20,000
11 Outer London 'A' roads
12 Inner London 'A' roads
13 London 'B' roads
14 London 'C' and Unclassified roads
15 Urban 'B' roads
16 Urban 'C' and Unclassified roads
17 Rural 'B' roads
18 Rural 'C' and Unclassified roads

Expansion Factor Stratification

01 Motorways
02 Urban 'A' roads with flows over 20 thousand vehicles a day
03 Urban 'A' roads with flows under 20 thousand a day
04 Rural 'A' roads with flows over 20 thousand vehicles a day
05 Rural 'A' roads with flows under 20 thousand a day
06 Outer London 'A' roads
07 Inner London 'A' roads
08 Urban Minor roads
09 Rural Minor roads
10 London Minor roads

Growth Factor Stratification

01 Trunk Motorway
02 Trunk Urban 'A' roads
03 Trunk Rural 'A' roads
04 Principal Urban 'A' roads & Principal motorways
05 Principal Rural 'A' roads
06 London 'A' roads
07 London 'B' roads
08 London *'C' and unclassified* roads
09 'B' Roads outside London
10 'C' and unclassified roads outside London

1. Current Methodology

Introduction

1.1 The Department for Transport's traffic estimates are based on observations of traffic on samples of road links, defined as a stretch of road between two junctions. These are used to derive quarterly and annual road traffic estimates as published here.

1.2 A fundamental component of the Traffic Statistics estimation process is the stratification of the sample. This groups together roads that have been shown to have similar traffic patterns based on road characteristics such as road classification and flow level. The sample is then stratified across these groups to increase how well it represents the road network as a whole.

1.3 This sample stratification applies to both manual counts, which are conducted over a 12 hour period for one neutral day of the year, and Automatic Traffic Counters (ATCs) which are static and count traffic continuously throughout the year from a national network of around 200 ATCs.

1.4 Not every link is counted every year, a sample of links are counted and then expansion factors and growth factors are applied to get an annual traffic estimate of the whole network in a given year. The stratification is also applied when producing both expansion factors and growth factors, therefore, it needs to take into account both variation within the year and variation between years.

Current Stratification

1.5 The current stratification was introduced in the early 2000s (see Annex A). It comprises of a large number of categories (22), some of which have quite a small sample of Automatic Traffic Counters (ATCs) within them, meaning that estimates for these categories can be quite noisy. The characteristics used for categorising roads into the 22 groups are:

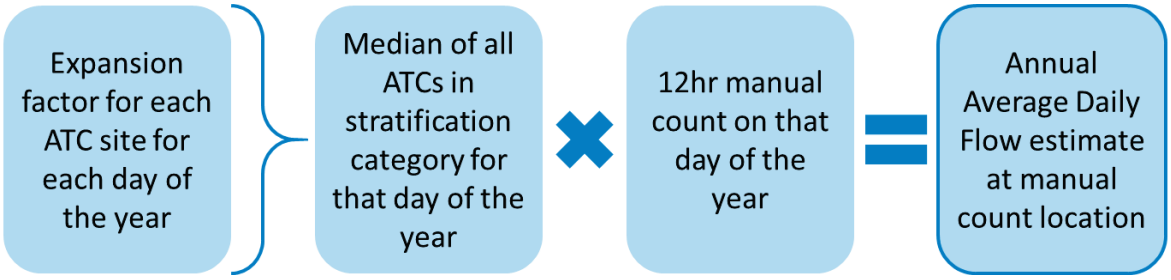
- Whether the road is in London, and one instance of whether it is in outer, inner or central London
- The Road Class (Motorway, 'A' Road, Minor Road)
- Whether the road link directly passes through Urban or Rural areas. Additionally the classification makes reference to different levels of area types such as ‘mostly rural’.
- Whether the road is in a Holiday area (based on Local Authority classification which takes into account things such as whether there is a national park or a seaside town within the authority).
- The level of vehicle flow (AADF) on the road link.

2. Within Year Analysis

2.1 The aim of the stratification project was to identify what road or area type attributes are important with regards to explaining variation, in order to group roads together that display similar variation. It is important to understand variation within the year (between weekdays and weekends for example).

2.2 Manual counts at locations on major roads are undertaken for a 12-hour period on one weekday in a neutral month during the course of the year. This is converted to an estimate of annual average daily flow at that location through the application of an expansion factor (for the stratification category to which the road location belongs). The expansion factor is calculated from Automatic Traffic Counter (ATC) data and captures the temporal variation in traffic on different road types: between night and day, different days of the week, and across months of the year.

Current method



2.3 Flow, and variations in flow, from ATC data was combined with geographical information about the road network from Ordnance Survey's Integrated Transport Network and other sources. Therefore, the road characteristics that could be examined include all of those listed in Chapter One that are currently used, plus:

- Road management status (principle vs trunk 'A' roads)

2.4 Although ATCs could be stratified by their seasonal, daily, and weekly flow variations, this information is not available for other road locations. Other road locations will only have traffic observations for 12-hours on a neutral day in a given year from a manual count. Therefore, it is not appropriate to use these characteristics to define strata.

2.5 The analysis in this project was limited by the data available. There is not an ATC on every road link in Great Britain, however, the ATC data was taken to be a representative sample. This paper aims to explain variations in flow based on this sample of sites.

Investigating flow variation using ANOVA and coefficient of the variance

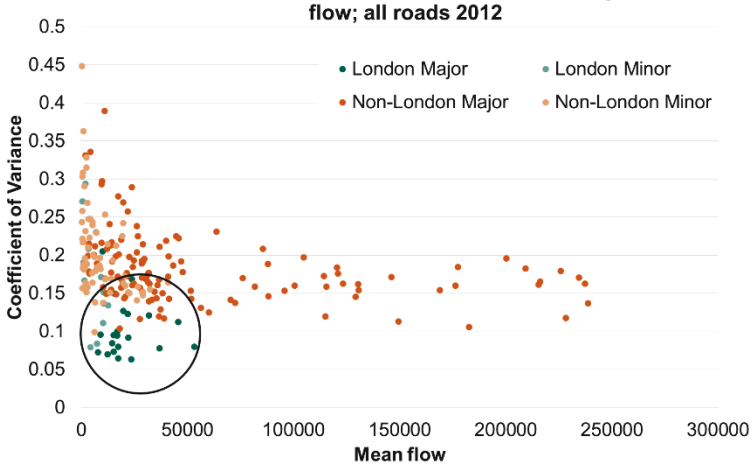
2.6 In order to identify which road or area attributes are important with regards to explaining variation, a measure of variation in flow was calculated that was independent of the size of traffic flow. This meant that analyses could be carried out

to test for correlations and significant differences between the variations on different roads, without being confounded by the size of the traffic flow.

2.7 The measure used was the coefficient of the variance (CV), calculated for each ATC site. The idea of this was to look at the effects of certain attributes (e.g. London/ outside London) on fluctuations in traffic flow and therefore what attributes should be used to define strata.

2.8 Chart 1 shows evidence of a relationship between coefficient of the variance and the mean flow. The correlation between the two is significant at one per cent, suggesting that as flow increases the relative variation decreases. However, the spread of the coefficient of the variance is quite large on lower flow sites.

Chart 1: Relationship between variance and average vehicle flow; all roads 2012



London

2.9 The cluster of sites in the bottom left hand corner of Chart 1 (circled) are largely London sites with particularly low levels of relative variation. This accounts for the some of the variation at lower levels of flow and suggests that London is something of an anomaly and therefore perhaps should have its own stratum.

2.10 To empirically test this, ANOVA analysis was carried out. The results found that London is significantly different (based on CV) from the rest of the country. To further investigate the different areas within London, Post hoc Tukey's HSD results comparing Outer London, Inner or Central London, and Other urban 'A' roads were carried out. They revealed that Outer London is significantly different from other urban areas but not from Inner or Central. This suggests that London should be combined into one group (see Annex B for the ANOVA output).

Metropolitan areas

2.11 Correlation and ANOVA analysis did not find a significant difference between metropolitan and non-metropolitan areas regardless of whether or not London was included or excluded from the test. This suggests that metropolitan area status should not be separated into a different stratum.

Holiday areas

2.12 Holiday areas are an element of the current stratification. The current stratification uses a local authority level classification to define holiday areas. Examining the coefficient of the variance found that although a small group of ATC sites in holiday areas do have higher levels of variation the vast majority did not.

2.13 No other area classification has been found that could be used identify road links that would have high levels of seasonality.

2.14 Looking at international methods, New York State have three category types; commuter, winter peak and summer peak, acknowledging the importance of seasonal variations. The road links have been allocated to one of these three categories based on knowledge of the local areas and data collected from ATCs along those routes. Unfortunately, the DfT traffic statistics team do not have this level of knowledge for all roads in Great Britain, and our network of ATCs is not sufficient

for this purpose. As a result, the application of tourist or holiday areas to road traffic statistics is outside the scope of this current project, to possibly be reviewed if an appropriate geography becomes available.

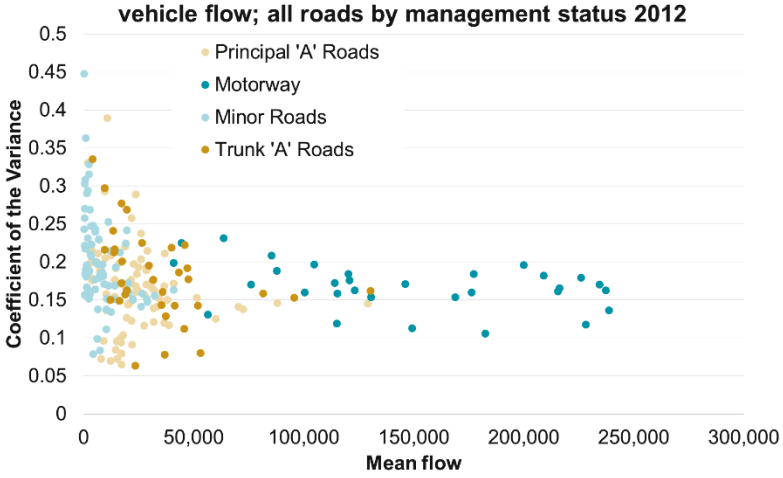
Urban/rural

2.15 Correlation and ANOVA analyses found that there is a significant difference between rural and urban areas. This excluded London sites, because London has been shown to be different from other urban sites, and also excluded motorway sites because the vast majority of these are in rural areas and somewhat separate from local traffic patterns.

Road classification and management

2.16 The DfT road traffic statistics methodology is split into major (motorway and 'A' roads) and minor roads. Looking at correlations and ANOVA analyses suggests that major and minor roads are significantly different from each other with regards to relative variance.

Chart 2: Relationship between variance and average vehicle flow; all roads by management status 2012



2.17 Looking at a further road class breakdowns in Chart 2, it appears that motorways are relatively unique with regards to their patterns. They occupy a narrow and relatively low band of variation with the majority of sites having a coefficient of the variance between 0.15 and 0.20. The road management status of 'A' roads (i.e. principal or trunk) appears to have little or no influence over the within year variation displayed by the traffic on them. 'A' roads have a particularly large variation in both flow and coefficient of the variance. In particular, the CV appears to get higher, and more diverse, at higher levels of mean flow compared to the other road types (the points are more spread along the y-axis at flows between 20,000 and 40,000 than minor roads are at that level).

2.18 Comparing all of the means via Tukey's HSD, looking at road classification individually (i.e. comparing motorways, 'A' roads, 'B' roads 'C' roads and unclassified roads to each other individually) 'C' and unclassified roads are significantly different from both 'A' roads and motorways. However, 'B' roads are not different from any other group based on CV.

Conclusions

The Coefficient of the Variance (CV) and ANOVA analysis suggests that:

- London should be separate from the rest of the country
- Metropolitan areas do not need their own category
- It is not currently possible to include holiday areas due to difficulties defining geographical areas
- Urban and rural areas are distinct
- Major and minor roads are distinct but distinction between road classifications is unclear, particularly for 'B' roads.

Investigating seasonal, weekly and daily variation

2.19 The next step in this project was to look at different aspects of variation within the year. The coefficient of the variance calculated above is based on daily level data. As a result, the variation does not take into account the variation between day and night which is important with regards to calculating expansion factors. The expansion factor calculation was broken down into three components;

1 A ratio between the average 24 hour and the average 12 hour period (7AM till 7PM), where a ratio of 1 would signal that all of the traffic is during the daytime and a ratio of 0.5 would signal that there is an even split between day and night.

2 The ratio between the average working weekday (Monday to Friday) and average weekend day, where a number greater than 1 signals that the average weekend day is busier compared to the average weekday and visa-versa.

3 A crude seasonality factor was calculated by dividing the three months with typically the highest flow (July, August, September) by three months that typically have the lowest flow (December, January, February), where a larger number signals greater seasonality.

2.20 These three ratios and the mean daily flow for each site were examined using cluster charts and ANOVA techniques in order to isolate important road characteristics that could be used for identification of strata.

London and Metropolitan areas

2.21 All three metrics found a significant difference between London and the rest of the UK's road traffic patterns. This confirms the conclusions from the previous analysis.

2.22 To test for the differences between area groups within London, Tukey's HSD was used as part of ANOVA analysis on all three metrics for three area groups; Outer London, Inner or Central London, and other Urban areas. All three groups were significantly different from each other on day-night and weekend-weekday metrics while none of them were significantly different according to the seasonal metric.

2.23 Testing the differences between road classes and flow within London were shown to be largely insignificant with the exception of the differences between 'A' and unclassified roads.

2.24 London was significantly different from metropolitan areas on all three metrics while there were no significant differences between metropolitan areas and other urban areas, confirming previous analysis that suggested London should be separate but further distinction based on metropolitan area status is not necessary.

Road Classifications

2.25 Analysis found a distinction between major roads and minor roads on two of the three metrics (weekday-to-weekend ratio was not found to be significantly different).

Major Roads

2.26 Outside London, 'A' roads and motorways were only significantly different for the 24 hour-12 hour ratio. This makes intuitive sense, a higher proportion of the total traffic on 'A' roads takes place during the daytime than for motorways where traffic is more evenly distributed throughout the day and night (i.e. the level of variation is higher for 'A' roads).

Minor Roads

2.27 'B', 'C' and unclassified roads also appear to be poor indicators of variation, each minor road type was not significantly different from the others on any of the metrics used suggesting a minor road grouping makes sense. Additionally, 'A' and 'B' roads appear not to be distinct from each other based on these three metrics.

2.28 Based on the above analysis and the analysis that has been carried out using the coefficient of the variance, road classification appears to represent more of a continuum of types of roads with non-distinct boundaries between neighbouring road types. While broad categories are different from one another (major and minor) and some roads are different from others ('A' roads compared to 'C' and unclassified roads, for example) there is no clear point at which to break up the sample by classification.

2.29 To investigate further, Tukey's HSD was used as part of ANOVA analysis on all three metrics (Day and night, Weekend and weekday and seasonal metrics) and for five groups (Motorway, 'A' road, 'B' road, 'C' road and unclassified road). Very few differences were found to be significant, the results of this Tukey HSD are summarised in the table in Annex B.

2.30 From this we can conclude that both motorways and minor roads need to be distinct groups, but it is unclear whether 'B' roads should be combined with 'A' roads or 'C' and unclassified roads. Also, the evidence is not conclusive as to whether 'A' roads should be combined with motorways or have their own group.

Road management

Principal 'A' roads and trunk 'A' roads outside London were only found to be significantly different on the seasonal metric suggesting that road management should not be separated out. However, some distinction was found between motorways and 'A' roads of both types which provides some evidence, but not conclusive evidence, that 'A' roads and motorways could be separated.

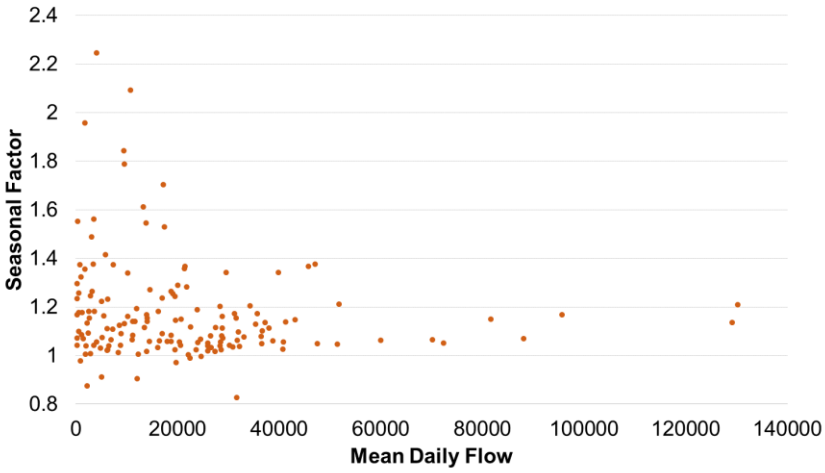
Definition: Road management

The trunk road network, consisting of most of the motorways and some 'A' roads in England, is managed at the national level by Highways England and is collectively referred to as the Strategic Road Network. All other roads are local authority managed (sometimes called principal).

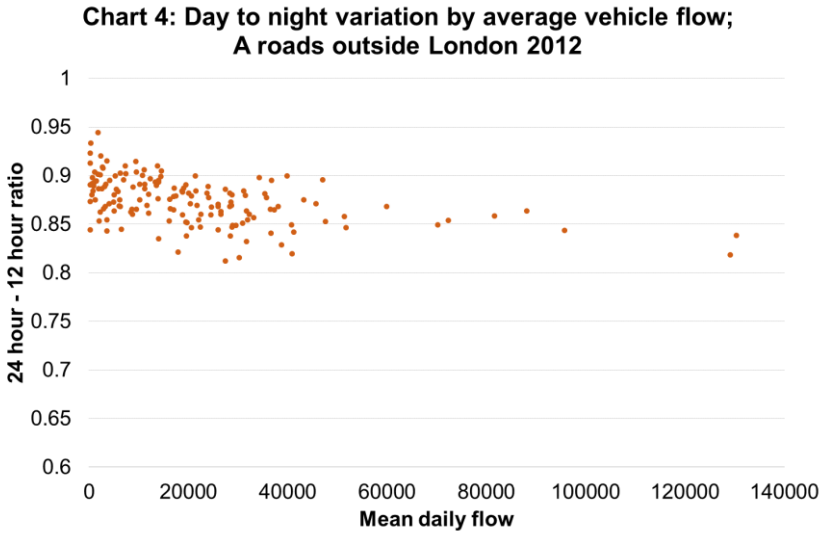
Level of flow

2.31 Looking purely at motorways and 'A' roads, and also 'A' roads on their own, dummy variables were created to compare different levels of flow: above 10 thousand vehicles a day versus below 10 thousand vehicles, and so on for splits at 20 thousand, 30 thousand, 40 thousand, and 50 thousand vehicles a day

Chart 3: Seasonal variation by average vehicle flow A roads outside London 2012



on average. These dummy variables were tested by comparing the means of three metrics.



2.32 The results found both flows above 10,000 and flows above 20,000 were significantly different than roads with flows below 10,000 and 20,000 respectively for all metrics. Looking at the test results, the 20,000 flow level appeared to be the more appropriate split based on the strength of the significance and the number in each group. Charts 3 and 4 place the seasonal factor and the 24 Hr - 12Hr ratio against the average daily flow of that site for 'A' roads outside of London. Both show a slight upwards trend for flows under 20,000. The spike in variation is particularly pronounced for the seasonal factor. Therefore based on the pattern, the significance and the number in each group it makes sense to split 'A' roads into those with flows above 20,000 and those below 20,000.

2.33 When replicated with minor roads the sample sizes are too small for conclusions when split at 10,000 or 20,000 indicating that minor roads do not need to be split by flow as there aren't enough roads with significantly different flows to justify a separate stratum.

Urban and Rural definition

2.34 Urban and rural definitions are used as part of the current estimation process. Traditionally they have been applied to road traffic statistics for both 'A' roads and minor roads. For both 'A' roads and minor roads, urban and rural groups were significantly different on two of the three metrics (12hour-24hour and seasonal). Urban roads show low levels of seasonality and a smaller proportion of their total traffic take place during the day. Both of these patterns are as expected.

Conclusions

2.35 These analyses suggest that:

- London should be distinct, further analysis into areas found that metropolitan areas outside London are not different from non-metropolitan areas.
- 'A' roads and Minor roads within London are distinct.
- For London 'A' roads these analyses are inconclusive on whether Outer London is distinct from Inner or Central London.
- Motorways should be distinct from minor roads but it is not clear if they are distinct from 'A' roads.
- Major roads and Minor roads should be distinct – 'A' roads are distinct from 'C' and unclassified roads but 'B' roads are distinct from neither. It is unclear whether they should be combined with 'C' and unclassified roads or in a category on their own.

- Roads with flows above 20,000 should be split from roads with flows below 20,000 for 'A' roads only, and 'A' roads do not need to be split by management status.
- For both major and minor roads, urban and rural areas should be separated.

Testing of preliminary strata

2.36 At this stage, if the roads are organised into groups, keeping any groups for which the evidence is so far inconclusive separate for now, there are 12 groups, shown in the table below.

1.	Motorways
2.	Urban 'A' roads with flows over 20 thousand vehicles a day
3.	Urban 'A' roads with flows under 20 thousand a day
4.	Rural 'A' roads with flows over 20 thousand vehicles a day
5.	Rural 'A' roads with flows under 20 thousand a day
6.	Outer London 'A' roads
7.	Inner or Central London 'A' roads
8.	Urban 'B' roads
9.	Rural 'B' roads
10	Urban 'C' and Unclassified roads
11.	Rural 'C' and Unclassified roads
12.	Minor roads ('B', 'C' and Unclassified) London

2.37 The next stage of this process was to calculate daily expansion factors at site level and compare them to each other within and between the proposed groups, to make decisions on any of the questions that were inconclusive based on ANOVA analysis and also to validate the ANOVA analysis.

Should 'B' roads be combined with 'C' and unclassified roads, or with 'A' roads, or kept separate?

2.38 Both the median and mean expansion factors for urban 'B' roads and urban 'C' and unclassified roads were graphed and the patterns throughout the year were found to be very similar to each other (see Annex C). This suggests that the two groups are very similar with regards to variations in traffic patterns and therefore could be merged. This was repeated for rural 'B' roads and rural 'C' and unclassified roads and similar results were found, again persisting throughout the year.

2.39 'B' roads were then graphed against 'A' roads with flows over 20,000 and below 20,000 for both urban and rural roads (see Annex C). From these it was clear that 'A' roads and 'B' roads represent distinct patterns, this supports the suggestion that 'B' roads and 'C' and unclassified roads can be merged.

How many strata are needed for London 'A' roads?

2.40 The main question with regards to London 'A' roads is whether to include outer London with inner and central London, with urban 'A' roads or on its own. When the means and medians of these three groups were graphed, unlike the 'B' and 'C' and unclassified roads where the patterns were very similar, these show different patterns for all three groups (see Annex C). The patterns observed suggest that these three categories should be kept separate as long as the number of ATCs allow this.

2.41 Looking at the contributions that each 'A' road category makes to the total major road traffic figures, the London categories are a lot smaller in traffic levels than the others, in particular, the inner or central London 'A' roads category contributes around 1.6% of all major road traffic while outer London 'A' roads contributes 3.3% (see table below). Therefore, one London 'A' road category is more appropriate given the small traffic contribution.

EFCat	Description	Major road traffic proportion (%)
1	Motorways	31.8
2	Urban A roads, over 20,000	9.4
3	Urban A roads, under 20,000	10.0
4	Rural A roads, over 20,000	19.8
5	Rural A roads, under 20,000	24.1
6	Outer London A roads	3.3
7	Inner or central London A roads	1.6

Major roads outside London

2.42 If motorways and 'A' roads were combined the majority of the motorways would be combined with rural 'A' road with flows over 20 thousand vehicles a day. Hence, mainly these two categories were compared. Looking at charts of both the median and mean expansion factors, motorways and rural 'A' roads (with flows over 20 thousand vehicles a day) appear to show subtly different patterns. In particular, motorways have a higher expansion factor during the week, as a result of a higher proportion of traffic over night, than rural 'A' roads (see Annex C). This, and the differing trends in traffic over the years between these road types, suggests that these road types should be kept separate.

Conclusions

2.43 These analyses suggest that:

- 'B' roads should be combined with 'C' and unclassified roads
- London 'A' roads should be combined as one category
- Motorways should have a separate category

2.44 Based on the results of these analyses above the new suggested stratification strata are now:

1. Motorways
2. Urban 'A' roads with flows over 20 thousand vehicles a day
3. Urban 'A' roads with flows under 20 thousand a day
4. Rural 'A' roads with flows over 20 thousand vehicles a day
5. Rural 'A' roads with flows under 20 thousand a day
6. London 'A' roads
7. Urban Minor roads
8. Rural Minor roads
9. London Minor roads

Peer Review

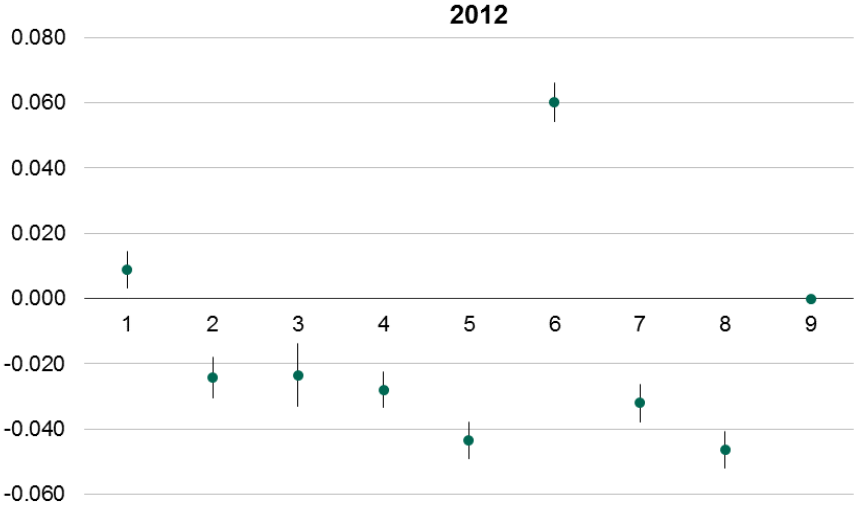
2.45 An independent methodological advisor from the ONS reviewed the within year analyses set out in this Chapter, and carried out one final check on this categorisation. This was to run a regression with fixed effects that takes into account the effect that day of the week and month have on the expansion factors (region was also tested and not found to be significant).

2.46 The fitted model had an R-squared of 43%, and so only goes some of the way in explaining the variation in Expansion Factors. Other factors, such as the weather and the timing of school holidays will also play a part, but have not been factored into the model.

2.47 The proposed strata were run through the model. The resulting coefficients for the proposed strata are shown in Chart 5.

2.48 It can be seen that the proposed categorisation works well. Groups 1, 6 and 9 are clearly defined with their 95% confidence interval range not overlapping any other category; groups 5 and 8 are very similar but separate from other groups; group 7 only overlaps slightly with groups 2, 3 and 4 which are all very similar. Looking at the definition of the road

Chart 5: Proposed new strata, regression coefficients, 2012



categories, these groupings seem intuitively reasonable. (Note that the estimate for category 9 is set to zero (anti-logged, this is 1) and the coefficients for the other categories are relative to this.)

2.49 When replicated by DfT statisticians with 2012 data, and then reproduced with 2013 data, the same pattern was found.

2.50 A slightly more detailed breakdown of strata was also tested with Motorways split between urban/London and other; and minor roads classified according to class of road ('B'/'C'/unclassified). From this, three further category separations were suggested for further investigation. Following replication with both 2012 and 2013 data and discussions between DfT and MAS these were not pursued.

Implementation

2.51 During 2017, DfT statisticians tested the implementation of the stratification for expansion factors on the full 2016 dataset.

2.52 This testing confirmed the above findings, with one exception. A separate category was appropriate to distinguish between Inner and Outer London 'A' roads. As a result, the final stratification for within-year variation and for expansion factors is set out in Box 2.

Box 2: Road Traffic Statistics Stratification: within year

Expansion Factor Stratification

01 Motorways
02 Urban 'A' roads with flows over 20 thousand vehicles a day
03 Urban 'A' roads with flows under 20 thousand a day
04 Rural 'A' roads with flows over 20 thousand vehicles a day
05 Rural 'A' roads with flows under 20 thousand a day
06 Outer London 'A' roads
07 Inner London 'A' roads
08 Urban Minor roads
09 Rural Minor roads
10 London Minor roads

3. Between Year Approach

3.1 The aim of the stratification project was to identify what road or area type attributes are important with regards to explaining variation, in order to group roads together that display similar variation. It is important to understand variation between years.

3.2 Some manual counts on major roads are done annually, others at 2, 4 and 8 year intervals. For those done less frequently than annually, the previous year's annual average daily flow (AADF) figure is converted to an AADF for the year in question. This is done through the application of a growth factor for the relevant stratification category, to convert the estimate from the year before to the year in question. Therefore, it is important to group roads together that display similar variation between years for the application of growth factors. The growth factor is calculated from Automatic Traffic Counter (ATC) data.

3.3 In the same way as for the within year analysis, flow and variations in flow from ATC data was combined with geographical information about the road network from Ordnance Survey's Integrated Transport Network and other sources.

3.4 The use of manual count data for the between year analyses was considered. Growth factors were calculated from links that are counted every year and the same methodology was applied to these growth factors as for the ATC growth factors. However, links with higher flows are more likely to be counted every year and this underlying bias meant that the sample size available was not large enough to provide conclusive results and this analysis was disregarded.

Investigating average growth in traffic using ANOVA

3.5 An average growth factor was calculated from ATC data for 1993 to 2013. This average was linked to observable characteristics of the road links as outlined in the introduction to Chapter 2. Growth factors for individual years and 5-year averages were also calculated and investigated, but it was found that there was too much variability in growth, and extreme outliers skewed the analysis. These alternatives are discussed further in the peer review section.

3.6 Annex D provides a table that summarises which variables were found to have a significant effect on growth in traffic based on this ANOVA analysis.

Road Classification

3.7 Major and Minor roads were found to be distinct. Looking into more detailed classification; Motorways were found to be clearly distinct in traffic growth patterns to all other road types. However 'B' and 'C' roads were not clearly distinct from any other road type. To try to establish which groupings of classification are more appropriate, ANOVAs were run with 'B' roads grouped with 'A' roads, and 'B' roads grouped with 'C' and unclassified roads separately. Both of these ANOVAs came out as significant. This suggests that the relationship between traffic growth and road classification follows more of a continuum than distinct groups and further

investigation is needed to decide how to split by road class. In the data that the traffic team use as a sample framework for minor roads, 'C' and unclassified roads are currently grouped together. For this reason it makes sense to group 'C' and unclassified roads together, although this may be reviewed further if the sample frame data were to change.

Road management

3.8 Unfortunately there are no ATCs on local authority managed motorways. Therefore, local authority managed motorways are excluded from this ATC analysis (see the next section using graphical representations of traffic flows to investigate principal motorways). Trunk motorways were not significantly different from Trunk 'A' roads, however, both trunk 'A' roads and Motorways were significantly different from both minor roads and principal 'A' roads. This suggests that Trunk 'A' roads and principal 'A' roads have distinct traffic growth patterns, it also suggests that trunk 'A' roads could be combined with motorways to form an SRN stratum.

Area type: Urban and Rural

3.9 Urban areas, rural areas and motorways were found to be significantly different from one another when comparing all roads. Further breakdown suggests that 'A' roads have significantly different traffic growth in urban and rural areas, but that minor road traffic does not differ between areas.

Area type: Metropolitan vs non-metropolitan

3.10 Metropolitan areas (excluding London) were found to be significantly different from non-metropolitan areas. However, by nature metropolitan areas are more likely to be classed as urban areas, therefore, it is possible that the differences between metropolitan areas and non-metropolitan areas are driven by the amount of urban and rural within the areas. Filtering by urban and then by rural and examining the differences between average growth factors in metropolitan and non-metropolitan areas found no significant differences. Therefore this suggests that it is not necessary to split roads by both urban and rural and metropolitan and non-metropolitan areas.

Flow

3.11 Dummy variables were created to compare different levels of flow, these were above 10 thousand vehicles a day versus below 10 thousand vehicles, and so on for splits at 20 thousand, 30 thousand, 40 thousand, and 50 thousand vehicles a day on average. Although some significant differences were found for all roads and major roads, when split by road classification, no differences were significant. Therefore, it appears that differences in growth seen between roads of different flows are largely accounted for by the differences in flow between different road classifications, for example motorways generally have a much larger flow than 'A' roads. Therefore, this does not suggest that splitting by flow will provide any extra indication of traffic growth trends over grouping by road class.

Conclusions

3.12 These analyses suggest that:

- London should be separate from the rest of the country
- Major and minor roads are distinct but distinction between road classifications is unclear, particularly for 'B' roads. 'C' and unclassified roads should be grouped together due to the framework used.

- Trunk and principal 'A' roads may be distinct from each other but it is unclear whether principal and trunk motorways should be grouped.
- Growth on 'A' roads in urban areas is distinct from 'A' roads in rural areas, but growth in rural and urban areas does not differ for other road classes.
- Metropolitan areas do not need their own category.
- It is not necessary to split roads by flow.

Investigating growth in traffic by examining graphical representations of traffic growth

3.13 Following ANOVA analysis of ATC data there are some unanswered questions:

1 Are LA managed motorways different from trunk motorways in traffic trends?

2 Are trunk 'A' roads different to motorways?

3 Are the differences between principal and trunk 'A' roads driven by the differences between urban and rural areas?

4 Are 'B' roads more similar to 'C' and unclassified roads, or 'A' roads?

1. Are LA managed motorways different from trunk motorways in traffic trends?

3.14 There are currently no ATCs on LA managed motorways, therefore, annual traffic estimate totals (based on ATCs and manual counts) were used to graph and examine patterns in traffic trends. Trunk motorways have displayed a steady increase in traffic whereas principal motorways have fluctuated but the trend has been broadly flat over time.

3.15 However, because there are currently no ATCs used in traffic estimates on principal motorway roads, they wouldn't be able to form a stratum by themselves. Additionally the relative contribution of principal motorways is very small (0.3% of all traffic on LA managed roads), therefore it is not necessary to dedicate a whole stratum to principal motorways.

Chart 6: Traffic on major roads, (index 2000=100)

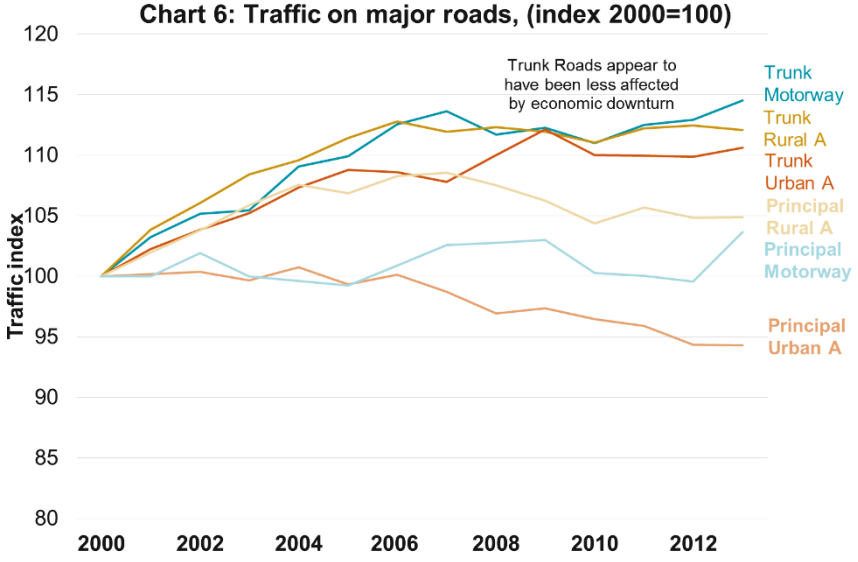
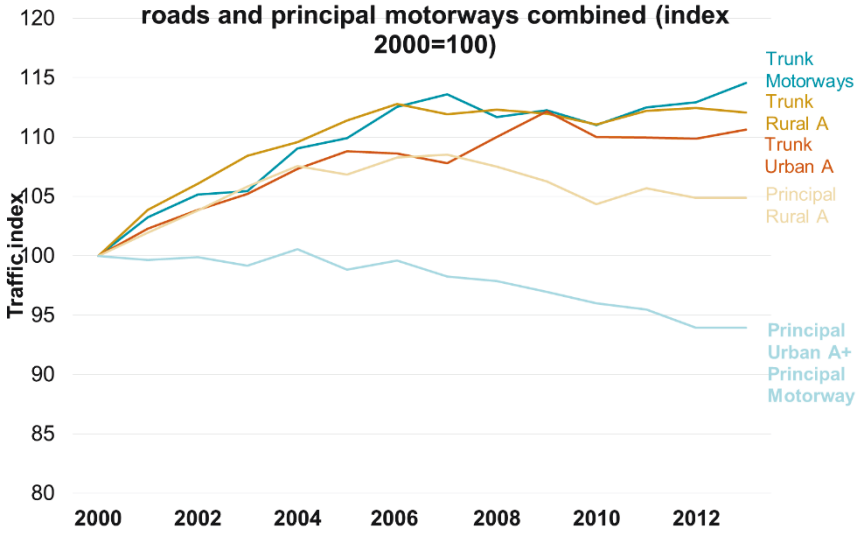


Chart 7: Traffic on major roads with principal urban A roads and principal motorways combined (index 2000=100)



3.16 Principal Motorways most closely match the trend of principal 'A' roads. To investigate whether to combine principal motorways in a stratum with rural or urban principal 'A' roads, all principal motorways were identified on a map and whether they were in an urban or rural area was determined, all 39 links identified were in urban areas therefore principal motorways should be combined with principal urban 'A' roads as demonstrated in Chart 7.

2. Are trunk 'A' roads different to motorways?

3.17 It appears from the graphs that growth on trunk motorways and trunk 'A' roads is very similar over the past decade. This is true for rural more than urban trunk 'A' roads, which makes sense as trunk motorways are predominantly in rural areas. Based on these graphs and the insignificant ANOVA between motorways and trunk 'A' roads, there is a case for combining all trunk roads into one SRN stratum. However, there is a user interest in differences between motorway traffic and 'A' roads traffic, therefore, it may be sensible to keep trunk 'A' roads and trunk motorways separate for growth factors. '

3. Are the differences between principal and trunk 'A roads driven by the differences between urban and rural areas?

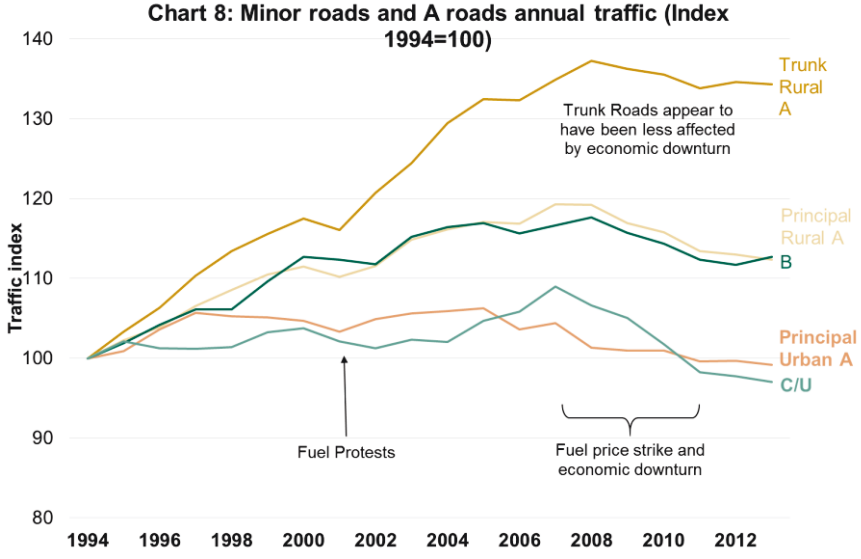
3.18 Trunk 'A' roads are more likely to be in rural areas than principal 'A' roads, therefore it is possible that, in the same way as for metropolitan areas, the differences between principal and trunk 'A' roads are accounted for by the differences between traffic trends in urban and rural areas.

3.19 An ANOVA on trunk 'A' roads, principal 'A' roads and minor roads in rural areas was significant[2], post hoc Tukey analysis found that principal and trunk 'A' roads in rural areas were significant from each other but neither from minor roads in rural areas. This suggests that there are differences in traffic trends on principal and trunk roads that are not completely accounted for by the differences between urban and rural roads. This conclusion is supported by Chart 6 above, which shows the different trends in 'A' roads split by both area type and management status. Principal rural 'A' roads have not shown as much growth as trunk rural 'A' roads in recent years. In addition, principal urban road traffic has declined, whereas principal rural road traffic is levelling off, suggesting that a combined growth factor is not appropriate.

4. Are 'B' roads more similar to 'C' and unclassified roads, or 'A' roads?

3.20 The ANOVA results for road class are not significant and looking at a road traffic index chart there isn't enough justification to separate minor roads by urban and rural areas. Therefore, 'B', 'C' and unclassified roads (not split by area type) were compared to 'A' roads split by area type and road management status (as these distinctions have already been established).

Chart 8: Minor roads and A roads annual traffic (Index 1994=100)



[2] It was not possible to perform an ANOVA to look at trunk vs principal within urban because there were no trunk 'A' roads in urban areas in the ATC sample used.

3.21 Chart 8 shows that the trend of traffic on 'B' roads is most similar to rural principal 'A' roads. This was investigated further examining different combinations of road groupings and although the similarity with rural principal 'A' is striking, it looks to be a coincidence rather than anything systematic. Therefore, it is recommended that 'B' roads are a separate growth factor category.

Conclusions

These analyses suggest that:

1. Local authority managed motorways should be combined with rural local authority managed 'A' roads.

2. Trunk 'A' roads and Motorways should be kept separate.

3. Principal and Trunk 'A' roads are distinct even within urban and rural areas.

4. 'B' roads should have a separate growth factor category.

Peer Review

3.22 An independent methodological advisor from MAS (ONS) reviewed the analysis carried out by DfT. They recommended not to focus too much on the results of the ANOVA using GFs averaged over a number of years and suggested looking at graphs indexed to different years to ensure that trends persisted.

3.23 The peer reviewer ran the annual growth rates through a fixed effects model, with road class, area (urban/rural) and year as a fixed effect. The inclusion of year dramatically increases the R-squared, from close to zero to around 0.7, clearly indicating that the general trend in traffic growth is the most important factor. This is intuitive, but necessary to prove.

3.24 The data was also run through a model regressing annual growth rates against year by road class by area. It was found that:

- London generally has lower growth rates than other urban areas, which in turn tend to be lower than for rural areas (note that the coefficients used in this analysis are a simple average of the three road types in each area) supporting the recommendation to split by rural, urban and London.
- Rural 'A' roads have consistently lower growth factors than other rural roads and urban 'C' and unclassified roads have consistently higher growth factors than other urban roads providing evidence for splitting by road classification within area.
- London 'C' and unclassified roads have lower growth factors in most years than other London roads[3].
- 'C' and unclassified roads in rural and urban areas are clearly distinct from 'B' roads and 'A' roads.
- It was also found that road management status makes a difference.

3.25 Taken together, these support the proposed stratification of road type by area proposed for the derivation of growth factors.

[3] However, the London 'B' road results are erratic, reflecting the small number of count points. DfT are upgrading the London network and are currently re-sampling for more sites, which aims to increase the number of ATC sites on 'B' roads. The stratification of links within London will be reassessed when the new network is completed and running.

Implementation

3.26 During 2017, DfT statisticians tested the implementation of the stratification for growth factors on the full 2016 dataset.

3.27 This testing confirmed the above findings. As a result, the final stratification for between year variation and for growth factors is set out in Box 3.

Box 3: Road Traffic Statistics Stratification: between year

Growth Factor Stratification

01 Trunk Motorway
02 Trunk Urban 'A' roads
03 Trunk Rural 'A' roads
04 Principal Urban 'A' roads & Principal motorways
05 Principal Rural 'A' roads
06 London 'A' roads
07 London 'B' roads
08 London *'C' and unclassified* roads
09 'B' Roads outside London
10 'C' and unclassified roads outside London

4. Conclusions

4.1 The final conclusion is to assign each road to one category and these categories will be aggregated up in different ways for calculating and using expansion factors and growth factors. The full categorisation and two different ways that this will be aggregated up are shown in the diagram in Box 4 below.

4.2 This final stratification has been applied to production of annual road traffic estimates for 2016 onwards.

Box 4: Road Traffic Statistics Stratification

01 Trunk Motorway
02 Principal Motorway
03 Trunk Urban 'A' roads <20,000
04 Trunk Urban 'A' roads >20,000
05 Trunk Rural 'A' roads <20,000
06 Trunk Rural 'A' roads >20,000
07 Principal Urban 'A' roads <20,000
08 Principal Urban 'A' roads >20,000
09 Principal Rural 'A' roads <20,000
10 Principal Rural 'A' roads >20,000
11 Outer London 'A' roads
12 Inner London 'A' roads
13 London 'B' roads
14 London 'C' and Unclassified roads
15 Urban 'B' roads
16 Urban 'C' and Unclassified roads
17 Rural 'B' roads
18 Rural 'C' and Unclassified roads

Expansion Factor Stratification

01 Motorways
02 Urban 'A' roads with flows over 20 thousand vehicles a day
03 Urban 'A' roads with flows under 20 thousand a day
04 Rural 'A' roads with flows over 20 thousand vehicles a day
05 Rural 'A' roads with flows under 20 thousand a day
06 Outer London 'A' roads
07 Inner London 'A' roads
08 Urban Minor roads
09 Rural Minor roads
10 London Minor roads

Growth Factor Stratification

01 Trunk Motorway
02 Trunk Urban 'A' roads
03 Trunk Rural 'A' roads
04 Principal Urban 'A' roads & Principal motorways
05 Principal Rural 'A' roads
06 London 'A' roads
07 London 'B' roads
08 London *'C' and unclassified* roads
09 'B' Roads outside London
10 'C' and unclassified roads outside London

Annex A: Current Stratification

A.1 Road categories used in the calculation of expansion factors prior to this analysis and proposed changes.

EfCat	Category Description
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
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
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
50	Minor rural roads in holiday areas with an estimated AADF of up to 399
51	Minor rural roads in holiday areas with an estimated AADF of 400 or more
52	Minor rural roads in all other areas with an estimated AADF of up to 2,499
53	Minor rural roads in all other areas with an estimated AADF of 2,500 or more
54	Minor urban roads in all areas except Greater London
55	Minor urban roads in Greater London

Annex B: Within Year ANOVA results

B.1 Post hoc Tukey's HSD results comparing Outer London, Inner or Central London and Other urban 'A' roads revealed that Outer London is significantly different from Other urban areas but not Inner or Central suggesting that London should be combined into one group.

*=significant at 95% confidence level
Dependent variable: Coeffictient of Variance
Tukey HSD

	Outer London	Inner or central London	Non London
Outer London	_	.169	*.000
Inner or Central London	.169	_	*.000
Non London	*.000	*.000	_

B.2 Tukey's HSD was used as part of ANOVA analysis on all three metrics (Day and night, Weekend and weekday and seasonal metrics) and for three groups (Motorway, 'A' road, 'B' road, 'C' road and unclassified). The significant differences are summarised in the table.

*=significant at 95% confidence level

Day Night

	M	A	B	C	U
M	_	.344	*.016	*.004	*.012
A	.344	_	.212	.065	.184
B	*.016	.212	_	.974	1.000
C	*.004	.065	.974	_	0.959
U	*.012	.184	1.000	0.959	_

Weekend Weekday

	M	A	B	C	U
M	_	.177	.975	1.000	.962
A	.177	_	.669	.302	.645
B	.975	.669	_	.975	1.000
C	1.000	.302	.975	_	.963
U	.962	.645	1.000	.963	_

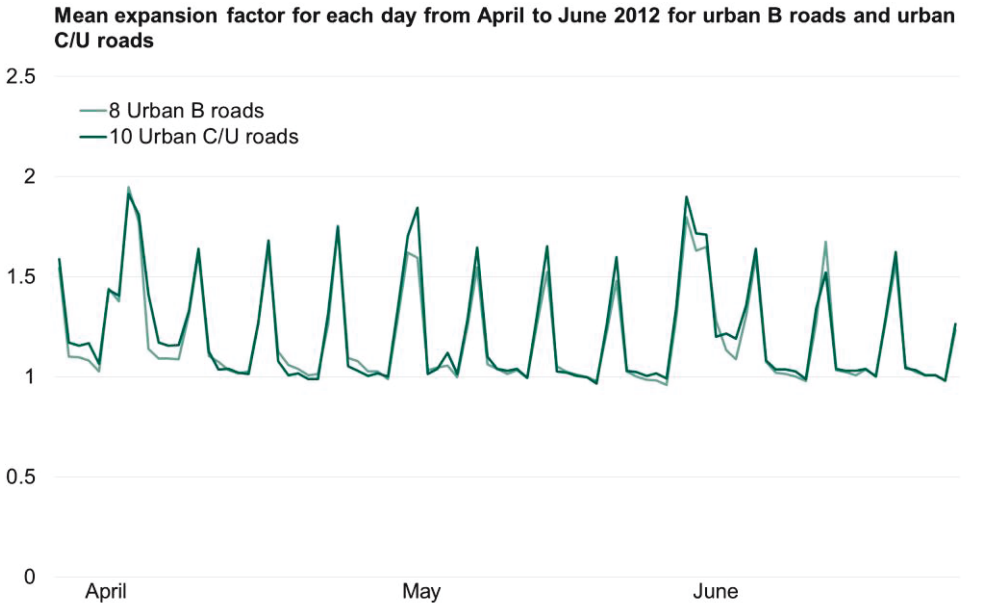
Seasonal

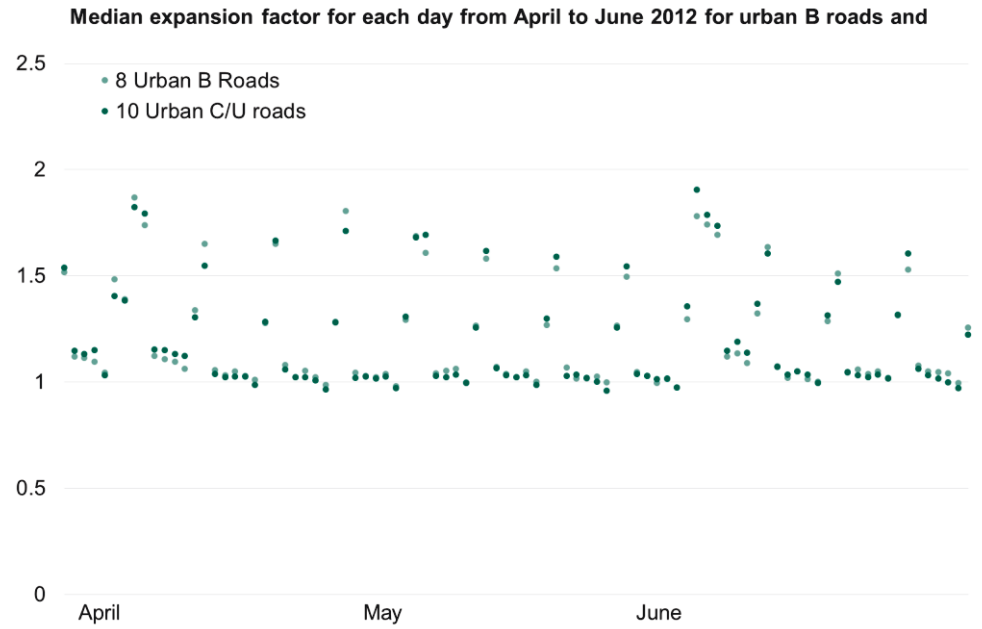
	M	A	B	C	U
M	_	.991	.318	.469	*.043
A	.991	_	.317	.520	*.021
B	.318	.317	_	1	.938
C	.469	.520	1	_	.919
U	*.043	*.021	.938	.919	_

Annex C: Within Year graphical analysis detail

'B' roads

C.1 Both the median and mean of daily expansion factors for urban 'B' roads and urban 'C' and unclassified roads were graphed and were found to be very similar to each other at each point throughout the year 2012. This suggests that the two groups are very similar with regards to traffic patterns and therefore could be merged. Comparing the mean and median expansion factors for rural 'B' roads and rural 'C' and unclassified roads in a similar fashion to the urban sites above produces similar results. Again these patterns continue throughout the year (although the graphs shown just demonstrate one quarter).

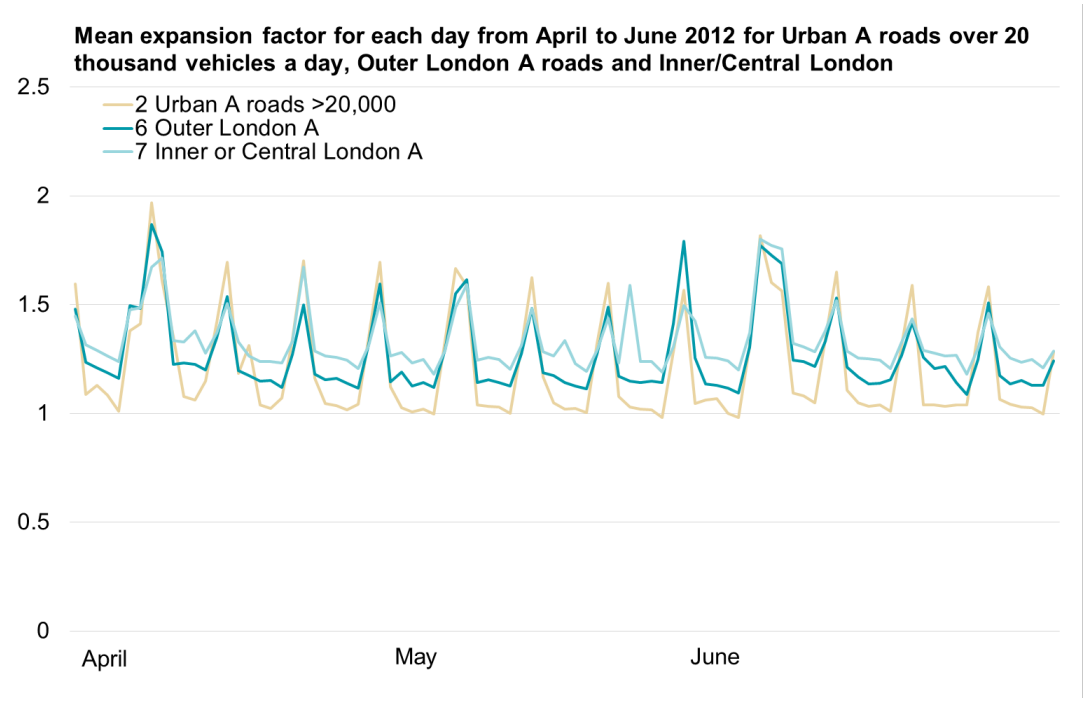


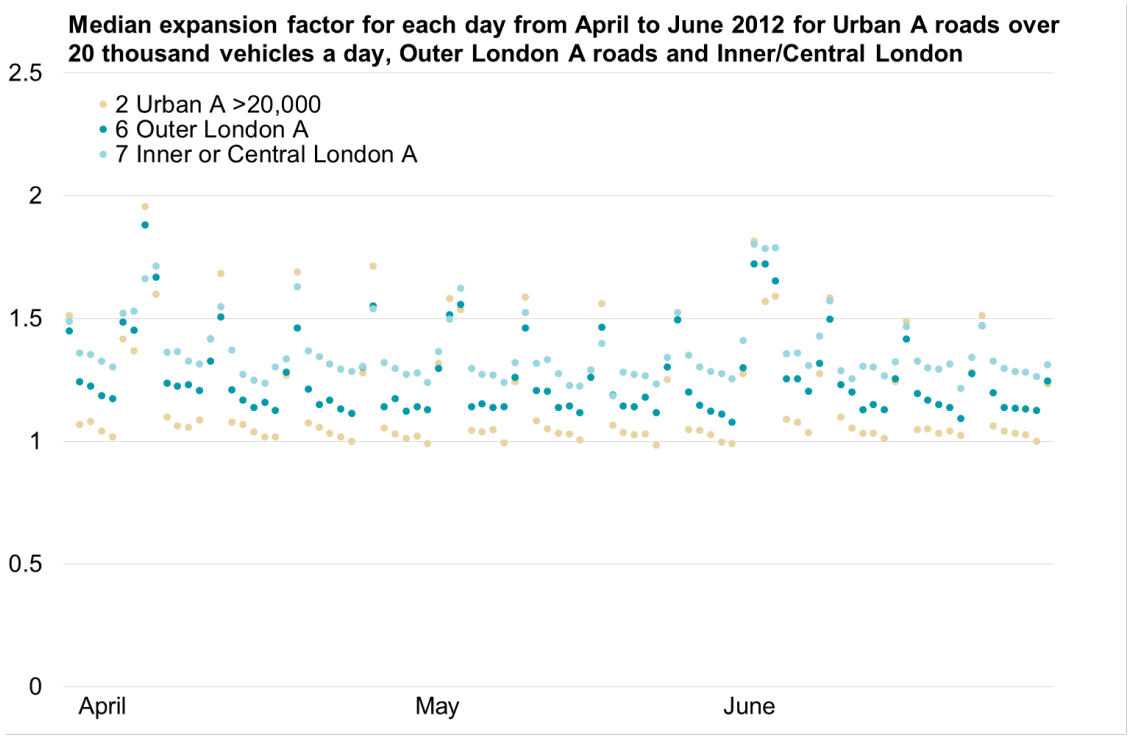


C.2 'B' roads were then graphed against 'A' roads with flows over 20,000 and below 20,000 for both urban and rural roads. From these it was clear that 'A' roads and 'B' roads represent distinct patterns, for example urban 'A' roads with flows less than 20,000 a day have a larger level of expansion during the week and a smaller level of expansion at the weekend than urban 'B' roads , again this pattern persisted throughout the year.

London 'A' roads

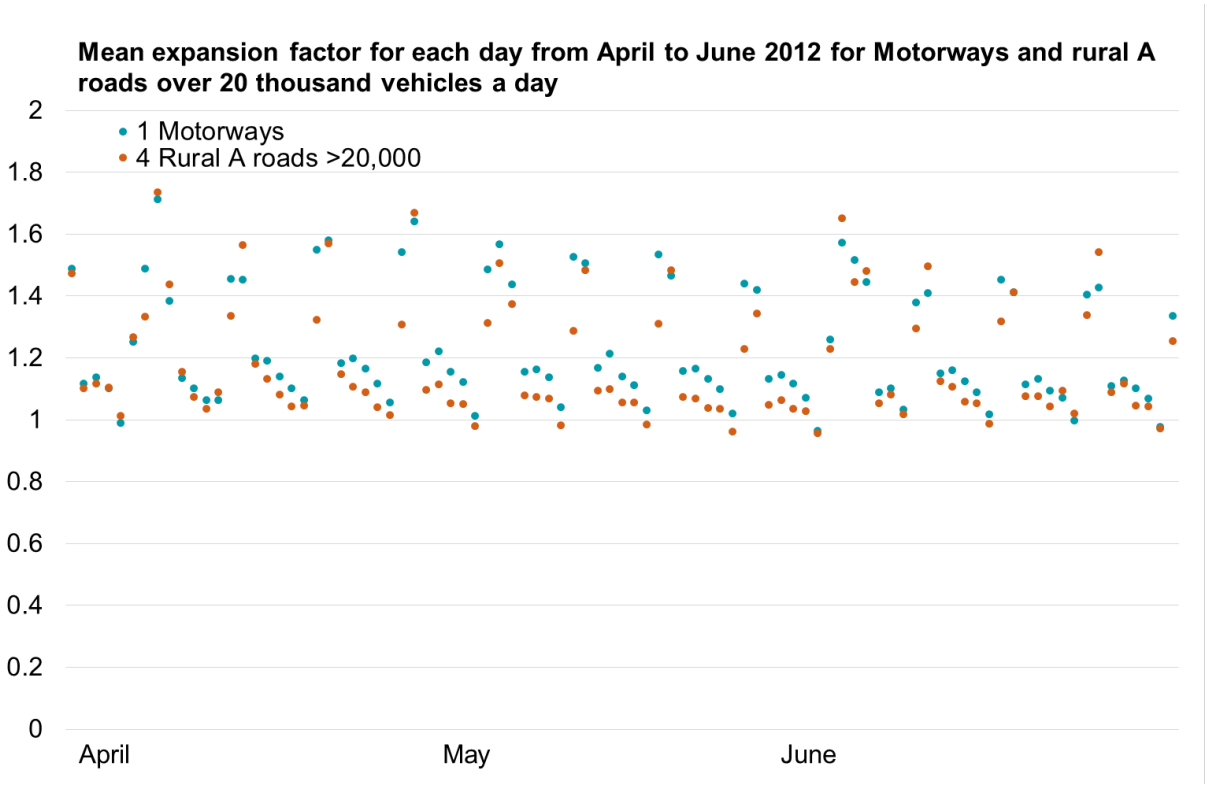
C.3 Outer London was compared to inner/central London, these showed different patterns for all three groups. In particular, the level of expansion during the week is lower in outer London than inner and central London but higher than urban 'A' roads over 20 thousand. This is a result of a higher proportion of traffic occurring during the week in inner and central London compared to on urban 'A' roads.

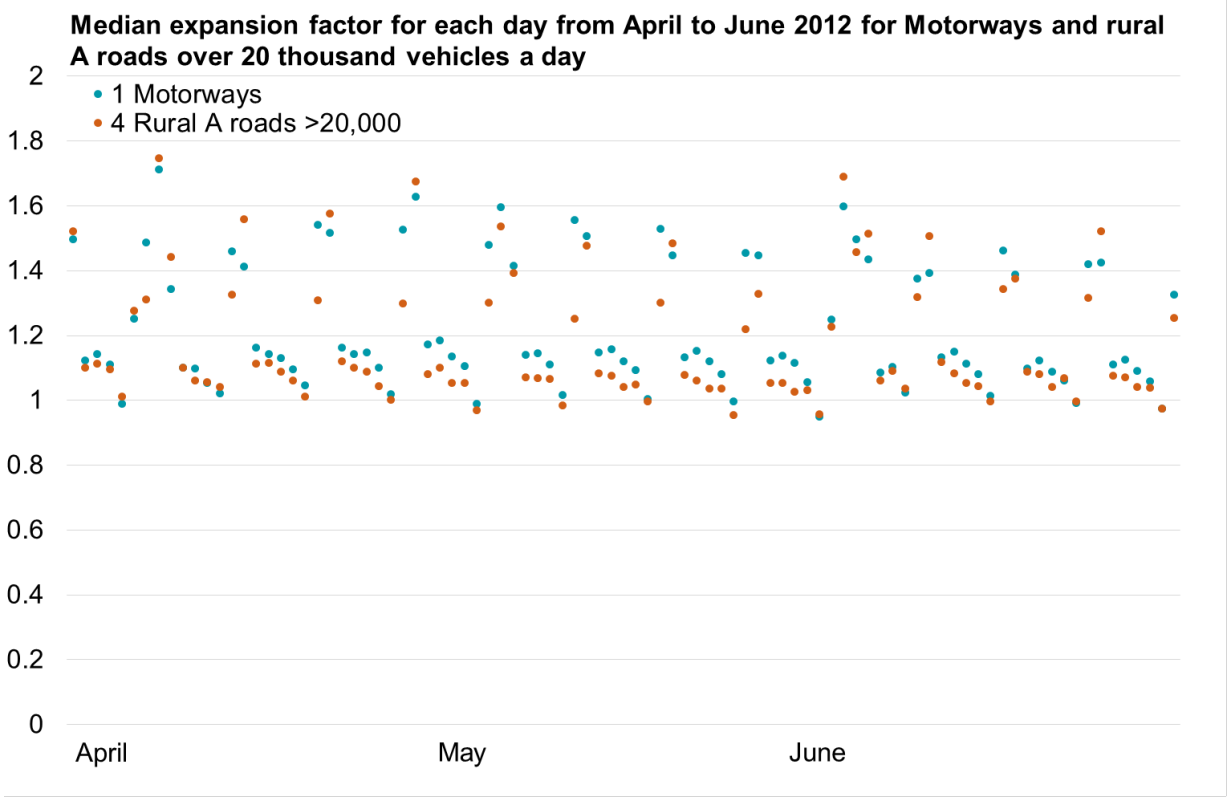




Major roads

C.4 Comparing motorways to rural 'A' roads in excess of 20 thousand vehicles a day shows subtly different patterns. For example the difference in expansion factor between Saturday and Sunday is much larger for Motorways than for rural 'A' roads.





Annex D: Between Year ANOVA results

D.1 ANOVA results, * indicates a significant ANOVA result, footnotes give Tukey post hoc significant results where there are three or more groups.

	Variables in ANOVA	ATC data 93-13
Area type	Metropolitan area vs non metropolitan area	*
	Urban vs rural (excluding Motorways)	*
	Urban vs rural ('A' roads)	*
	Urban vs rural (minor roads)	
	Urban vs rural ('B' roads)	
	Urban vs rural ('C' roads)	
	Urban vs rural (Unclassified roads)	
	Urban vs rural vs Motorways regions	*1
Flow	<10k vs >10K	
All roads	<20k vs >20K	*
	<30k vs >30K	*
	<40k vs >40K	*
	<50k vs >50K	*
Major roads	<10k vs >10K	*
	<20k vs >20K	*
	<30k vs >30K	*
	<40k vs >40K	*
	<50k vs >50K	*
'A' roads	<10k vs >10K	
	<20k vs >20K	
	<30k vs >30K	
	<40k vs >40K	
	<50k vs >50K	
Road class	'A'	*
	'B'	
	'C'	
	Unclassified	*
	Motorway	*
	All	*2
	M vs 'A' vs BCU	*
	M vs AB vs CU	*
	Major vs Minor	*
Road management	M vs Principal 'A' vs Trunk 'A' vs Minor	*3

1 all combinations of urban rural and motorway sig

2 M vs 'A', M vs C, M vs U

3 M vs PA, M vs Minor, TA vs PA, TA vs Minor