NTM EVOLUTION – LGV MODELLING AND FORECASTING

Analysis of Saturation Effects and Potential Use of Trafficmaster Data

PPRO 04/03/35 : Task 332(4-45-12)

20/03/2015

Quality Management

Issue/revision	Issue 1	Revision 1	Revision 2	Revision 3
Remarks	Initial draft to DfT	Revised	Final following DfT review	
Date	03/11/2014	19/12/2014	20/03/15	
Prepared by	Tim Gent	Tim Gent	Quang Nguyen	
Signature				
Checked by			Tim Gent	
Signature				
Authorised by			Tim Gent	
Signature				
Project number	70004128	70004128	70004128	
Report number	TN005	TN005	Final Report	
File reference			T:\10110012 DFT2_RD_PRJCTS\70004128 - DfT2 Task 332 LGV Modelling an\C Documents\TechNote\TN005 - Traffic Master Data Analysis_V2.5_WSPtemplate.docx	

NTM Evolution – LGV Modelling and Forecasting

Analysis of Saturation Effects and Potential Use of Trafficmaster Data

20/03/2015

Client

Liz Jacobs
DfT

Consultant

Tim Gent
WSP House
London
WC2A 1AF
UK

Tel: +44 77 1113 9387
Fax: +44 20 7314 5111

www.wspgroup.co.uk

Registered Address

WSP UK Limited
01383511
WSP House, 70 Chancery Lane, London WC2A 1AF

WSP Contacts

Tim Gent – 07711 139387

Table of Contents

1 Introduction

1.1 Background and purpose of note

1.1.1 WSP have been engaged by DfT to update and develop DfT's LGV traffic forecasting model. The LGV model is a sub-model of DfT's National Transport Model which is used to look at long term trends in transport demand.

1.1.2 Phase 1 of the work has delivered a revised model, documented in our Technical Note TN03, delivered in July 2014. At the same time, WSP proposed in a further note (TN04) options for Phase 2 of the work. After discussion with DfT the scope for this work was agreed and confirmed in the final TN04 as follows:

> *We will first prepare a brief narrative on the saturation effects and analysis of easily available data, with a maximum effort of 1-2 days (see section 2.2.3, first bullet);*
>
> *We will then consider the potential for TrafficMaster data to yield further insights into LGV traffic, in terms of origins, destinations, road types used and areas visited. This task will be dependent on gaining access to (1) further metadata on the TrafficMaster data, and (2) a usable sample of the data itself. The DfT will pursue this extra information prior to the work commencing.*
>
> *The scope and effort for this stage is uncertain, and will depend on the information available, and the interest generated. We are hopeful that some conclusions can be drawn within the remaining budget but cannot guarantee that these will be conclusive or necessarily yield positive results as to the usefulness of the data.*
> *(WSP TN04, 30 July 2014, paras 1.3.2 to 1.3.5)*

1.1.3 The work covered in these paragraphs has now been completed. As indicated above, the analysis carried out has been somewhat constrained by the budget available and the complexity of the datasets. In particular, the set-up time for obtaining and analysing the Trafficmaster dataset was considerably greater than anticipated due to the size and format of the data provided.

1.2 Content and structure of note

1.2.1 The contents of this note are as follows:

- Section 2: A discussion and some further investigation of factors which may limit the growth of LGV traffic in future, including saturation effects
- Section 3: Overview of Trafficmaster data received, metadata and initial analysis
- Section 4: Analysis of Trafficmaster data

2 Limiting Effects on LGV Usage and Ownership

2.1 Introduction

2.1.1 One of the key motivations for this project has been to allow DfT's traffic forecasts to account for the rapid growth in LGV usage in Great Britain, and to anticipate the proportion of national traffic which will be accounted for by LGVs in future.

2.1.2 To illustrate the importance of understanding LGV growth, Figure 2.1 below shows the historical and recent patterns of growth in LGV and all vehicle traffic. Though the trend from 1949 to the early 1980s was that the importance of light vehicles was falling[1], this reversed during the 1980s. LGV traffic showed its strongest rise relative to other vehicle types in the period from 2002-2007, rising from 11.3% to 13.3% of national traffic. Whilst absolute growth has slowed since 2007, the proportion of LGV traffic has continued to grow, accounting for 14% of all vehicle kms in 2013. To further put this in context, other Goods Vehicle traffic fell from 5.9% of total traffic in 2002 to 5.2% in 2013. Therefore LGVs are not only growing, but represent between two and three times the number of vehicle kms of HGVs.

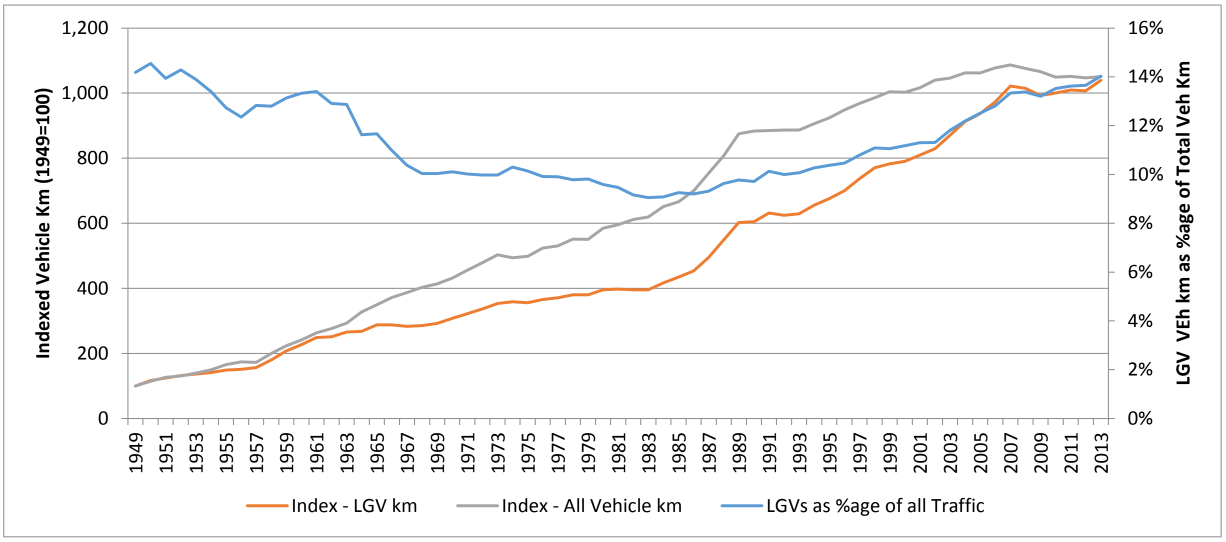


Figure 2.1- Vehicle km covered by LGVs vs Total All Vehicle distance, 1949-2014
(Source: DfT Transport Statistics 2013, TRA0101)

2.1.3 The delivered forecasting model (as documented in WSP's TN003, delivered 30th July 2014) used two key elements to project the future growth in LGVs:

- An econometric analysis of the relationship between GDP, fuel prices and LGV vehicle km, to produce projections of LGV traffic based on forecast growth in GDP and fuel price changes;
- An assumption that in future the elasticity to GDP would decrease, in line with a historic 'steady state' elasticity of approximately 1.0, and hence the rapid period of growth from 2002-2007 would not be the general trend when GDP is rising (see TN003, section 2.4).

2.1.4 The latter point has a critical impact on the forecasts themselves, and stems from the belief that it is unlikely that the overall trend in LGV vehicle km growth seen in the calibration dataset can be sustained into the long-term. This has been informally referred to as 'saturation'. However, in fact it may be more helpful to consider a series of possible reasons why future growth may be slower:

[1] This is noted for interest, though the precise reasons are likely to be historic and at present not considered to be within the scope of the current work.

1) **The historic rise was due to unique circumstances:** The recent rapid rise in LGV vehicle km may have been due to circumstances specific to that time period, which will not be repeated or continue unchanged into the future;

2) **Limiting supply side effects:** In future, limiting factors may apply to the growth in LGV usage and/or ownership, due to supply side effects which increase costs/decrease utility of LGVs. This would most likely stem from limitations in road space, though parking/storage space might also apply. Changes in the supply or cost of vehicles are also possible, but only due to external influences (e.g. regulation or taxation).[2]

3) **Limiting demand side effects:** Conversely**,** limiting factors could apply to growth in demand for LGV use or ownership. This is most likely to arise due to market saturation, i.e. the point at which there are no potential owners/users of LGVs who are without access to one, and hence growth only occurs as new owners/users appear. This is the effect most often implied by the term saturation, or more explicitly 'market saturation'.

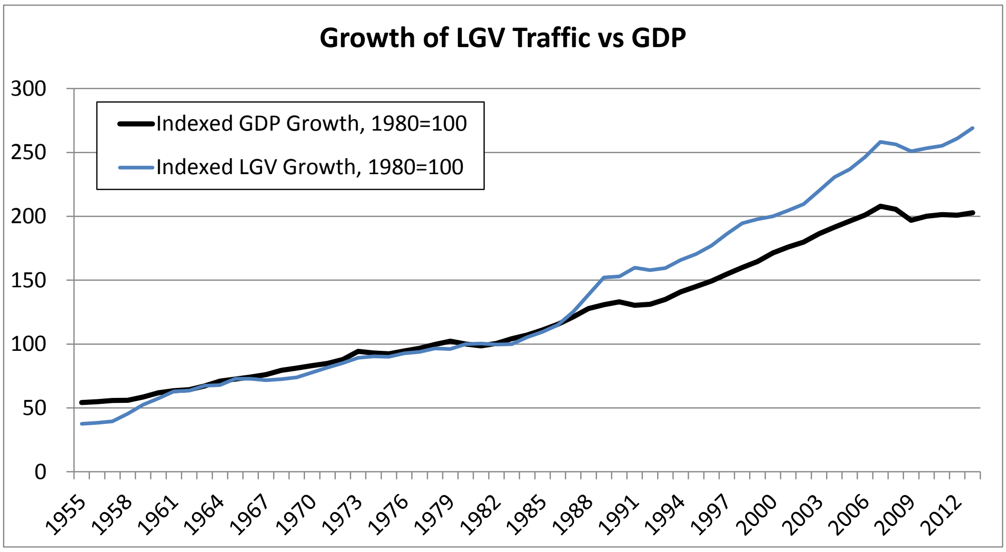
2.1.5 The approach taken in the current model can be seen as a combination of (1) and (3) above: we have observed that the GDP elasticity of LGV vehicle km was >1, so that the rate of LGV usage per unit of GDP was increasing. By considering the long-run relationship, we have assumed that the elasticity will return to 1 in the future, i.e. that LGV km will increase only as GDP increases.

2.1.6 In order to establish whether any (or all) of these limiting effects will take place, it is necessary to consider in more detail the reasons that LGV growth has been so high in recent years (or more explicitly, why GDP elasticity has been >1).

2.1.7 In the following section, we discuss some potential reasons that LGV use has been rising rapidly, followed in later sections by testing of some specific hypotheses using freely available data.

[2] It should be noted that the National Transport Model includes some response to limiting supply-side effects The LGV forecasts (along with car and HGV forecasts) are incorporated into the FORGE sub-model, which incorporates a series of responses to congestion including a 'volume response' or suppression elasticity. This work has not included a review of this response or its appropriateness for LGV traffic.

2.2 Potential Drivers of rapid LGV Traffic Growth



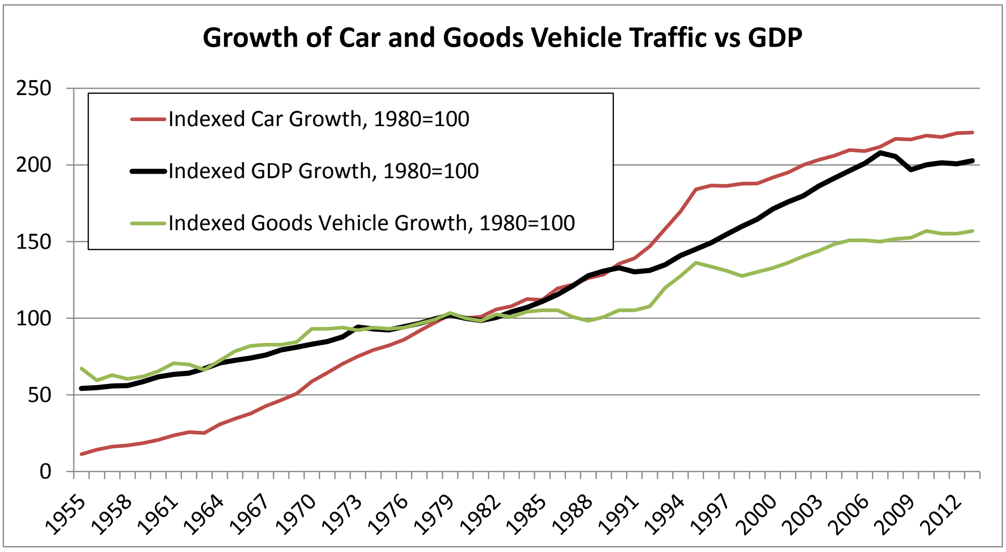


Figure 2.2 - Growth in GDP versus Vehicle Traffic (a) LGVs, and (b) Cars and Goods Vehicles
(Source: DfT Traffic Data, TRA101)

2.2.1 Our previous Technical Note TN003 (July 2014) included in Section 2.4 an analysis of the historic elasticity of LGV traffic to GDP. This relationship is further illustrated in **Figure 2.2** above. **Figure 2.2**a shows that LGV traffic grew strictly with GDP from 1960 until the mid-1980s, and then experienced a rapid growth above GDP in the late 1980s. Since then LGV traffic appears to have grown only slightly faster than GDP, other than the more rapid rise in the last few years. **Figure 2.2**b shows the same comparison for cars and goods vehicles. Most crucially, both car and goods vehicle traffic have grown more slowly than GDP since the mid-1990s, i.e. their GDP elasticity is less than 1. Therefore it may be helpful to consider the reasons why LGV traffic growth is stronger than that shown for cars and other goods vehicles over the last 15-20 years.

2.2.2 The DfT considered some of the potential reasons for this before commissioning the current work, and the following boxed section provides some key points from the project specification.

What are the main drivers of LGV traffic growth?
(extract from DfT Specification for TTEAR LGV modelling and forecasting , 1 April 2014)

Since the late 1990s, a number of **regulatory and operational** factors have served to increase the attractiveness of LGVs relative to small HGVs (3.5 to 7.5 tonnes). LGVs are subject to fewer regulations governing their operation and use than HGVs of 3.5-7.5 tonnes. Anyone operating a goods vehicle over 3.5 tonnes needs an operator licence involving the payment of a one-off application fee plus a fee to issue the licence (costing more than £600 in total). In January 1997, driving licence weight restrictions were changed making the standard driving licence only valid for vehicles up to 3.5 tonnes rather than 7.5 tonnes. Professional HGV drivers need to pass a Certificate of Professional Competence driving examination. There are additional regulations on drivers' hours and speed limiters that apply only to vehicles over 3.5 tonnes. Over the same time period, LGVs have become more versatile, with vehicle manufacturers making larger and longer models.

Taxation issues may also have affected the demand for vans over time, for example changes to vehicle excise duty and benefit-in-kind taxation. In 2008, personal use of a company van business-in-kind taxation rose from £500 to £3000 (policy announced in 2005).

Other studies[3] that have looked at the factors behind LGV traffic growth have noted a number of other **social and business changes** that will have increased the demand for vans and their use.

The growth in the number of households – 2.1 million between 2001 and 2012[4] - is likely to be an important factor. There has also been increased demand for home improvements and construction at residential addresses. The installation and maintenance of new telecommunications networks and equipment that requires installation, servicing and repair is another factor that has served to increase LGV traffic.

The rapid growth in internet shopping is likely to be a factor with internet sales doubling over the past five years alone (ONS[5]). Alongside this, there has been growth in the demand for express and parcel services.

Business changes will also have had an impact on van ownership and use, including the increase in the number of small and medium sized enterprises. The number of businesses with fewer than 50 employees increased by 41% between the start of 2000 and 2013[6]. Changes to businesses' supply chain management have resulted in more frequent, smaller, Just in Time deliveries. There has also been a trend towards the outsourcing of service functions.

This list of factors is not exhaustive but serves to highlight some of the complexities associated with disaggregating and quantifying the impact of different factors on LGV traffic growth.

[3] AEA Technology (2010) "Determining counterfactual CO2 emissions of new vans" and CfIT report July 2010: "Vans and the Economy".

[4] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/189965/AllTablesNonRegionalFinal__3_.xls

[5] http://www.ons.gov.uk/ons/rel/rsi/retail-sales/october-2013/sty-internet-sales.html

[6] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254554/bpe_2013_data.XLS#'Time-series'!A1

2.3 Evidence for Structural Changes in LGV Ownership

2.3.1 In this section we outline two specific hypotheses for the growth in LGV ownership which are related to LGV ownership and usage, based on the potential drivers in the previous section, and further discussion during the project.

2.3.2 The first hypothesis (H1) for the rapid growth in LGV ownership has been that some industry restructuring has taken place in non-distribution industries such as construction and utilities. These firms make use of goods vehicles to transport staff and tools to work locations. It is believed that skilled and semi-skilled workers are increasingly self-employed or employed by very small businesses, rather than directly employed by large firms as was the case in the past.

2.3.3 For example, historically construction firms may have directly employed gangs of workers, and provided transport to a work site and tools at that location. Today there is an increased likelihood of subcontracting to smaller firms, who have their own transport and tools, and therefore each require an LGV.

2.3.4 It should be noted that this phenomenon can be thought of as separate from the recession itself, and (again anecdotally) would have started much earlier. Therefore it can potentially account for the rapid rise in LGV usage prior to 2007.

2.3.5 An alternative hypothesis (H2) is that larger firms have adopted LGVs, for one or more of the following reasons:

- as part of the logistics chain (switching from HGV to LGV usage) (H2.1),
- as a result of growth in new services best delivered by LGVs (e.g. internet shopping, possibly some aspects of the telecommunications industry, and any growth in demand for home-based services, such as gardners or handymen) (H2.2);
- by consolidation of smaller firms (e.g. domestic trades such as plumbing being delivered by consolidated firms rather than self-employed tradesmen) (H2.3).

2.3.6 Whilst the former two factors might grow the overall numbers of LGVs, the latter might tend to maintain ‘parity’, simply replacing vehicles owned by small firms with fleet vehicles. This would of course act counter to our first hypothesis, and potentially mask it.

2.3.7 Finally, the increasing use of LGVs could simply be seen as the outcome of shifts in operation costs driven both by regulatory changes and economic/technology shifts (H3) as noted in the boxed section above. This effect could be seen simply as driving the processes covered by H1 and H2, and is not covered specifically here.

Relation to future LGV growth

2.3.8 Either hypothesis H1 or H2 could have a large bearing on our projections of future growth. The switch to small or self-employed firms (H1) is likely to be limited because (a) there are a finite number of ‘large’ firms which could outsource or otherwise restructure to form smaller companies and/or (b) the process will only take place where the economic / cost incentives encourage this change.

2.3.9 Likewise, each of the processes mentioned in H2 are limited: H2.1 and H2.3 by the natural limits of any restructuring, which once complete must return to strictly GDP-led growth; H2.2 is related to a subsector of GDP, which may or may not continue to rise in line with total GDP.

2.3.10 It is not currently clear that a forecasting model could be developed which would take account of these specific drivers, and allow these mechanisms to explicitly show the limits of LGV growth. Such a forecasting approach would require significant additional data collection and research.

2.3.11 However, with currently available data there are analyses which may give some indication as to whether these processes are (or are not) taking place. To investigate this further, we have considered:

- The corresponding changes in numbers of LGVs in private and company ownership, and the growth in LGVs vs HGVs;
- The growth in self-employed people in relevant trades/occupations, by comparing the 2001 and 2011 Census data;
- The growth in small firms/employers, from UK business registration data.

2.4 Structure of LGV Ownership

Private and Company Ownership

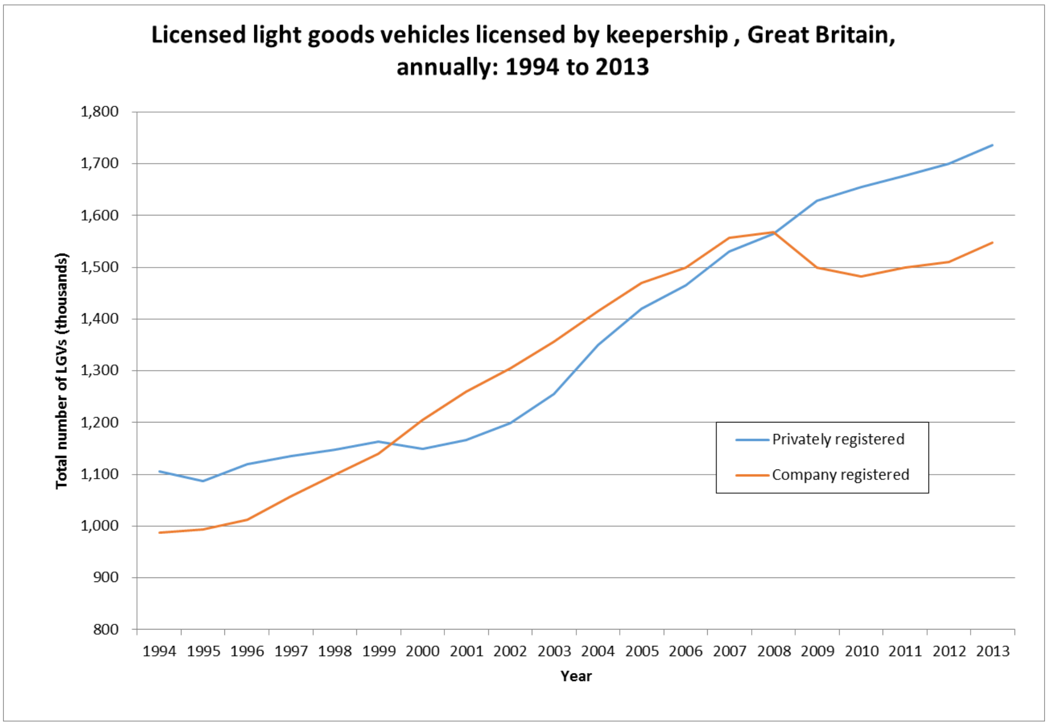


Figure 2.3 - Growth in LGVs in Great Britain, by Keepership[7] (1994-2013)
(Source: DfT VEH0402)

2.4.1 **Figure 2.3** above shows the number of privately and company registered LGVs in Great Britain. It is interesting to note that the numbers of each are broadly similar throughout the time period, showing that private registration has always been common. Prior to 1998, the number of privately registered vehicles exceeded company vehicles, but little growth occurred until 2002. There was a rapid rise in privately-registered vehicles in the period 2000-2007. From 2007, company registrations fell-off, whilst the rate of private registration was maintained[8].

2.4.2 Though no firm conclusions can be reached, this pattern is suggestive of the following phases of growth:

[7] The nature of private registration remains problematic, as it suggests that the owner is not engaged in any form of business. Some deeper investigations of these definitions, the types of vehicles and their usage may be worthwhile.

[8] It is worth noting that the total rate of registrations did fall substantially after 2007, and has only in 2013 recovered to 2006 levels, see Figure 2.4.

- Up to 2002, companies adopted LGVs in increasing numbers, whilst private registrations were stagnant, i.e. our second hypothesis H2 is dominant;
- From 2002-2007, both private and company registrations grew strongly. This may partly have related to a booming economy, with both the hypotheses above in action;
- After 2007, the recession has clearly affected company registrations, and any remaining action of H2 is hidden by the overall fall in company activity. Conversely, private registrations continue to grow strongly. This would seem to support hypothesis H1, as workers no longer employed by large firms may have obtained their own vehicles during this period.

LGV and HGV Ownership and Usage

2.4.3 **Figure 2.4** shows the growth in LGV licensing from 1994 to 2014, contrasted with the number of HGVs. This shows that the number of Heavy Goods vehicles has declined slightly over the period, whereas LGVs have continued to grow. Looking in more detail at the growth rates however, it is clear that up to 2000, the rate of growth was similar. After 2000, the HGV growth rate started to fall, but has recovered since 2009.

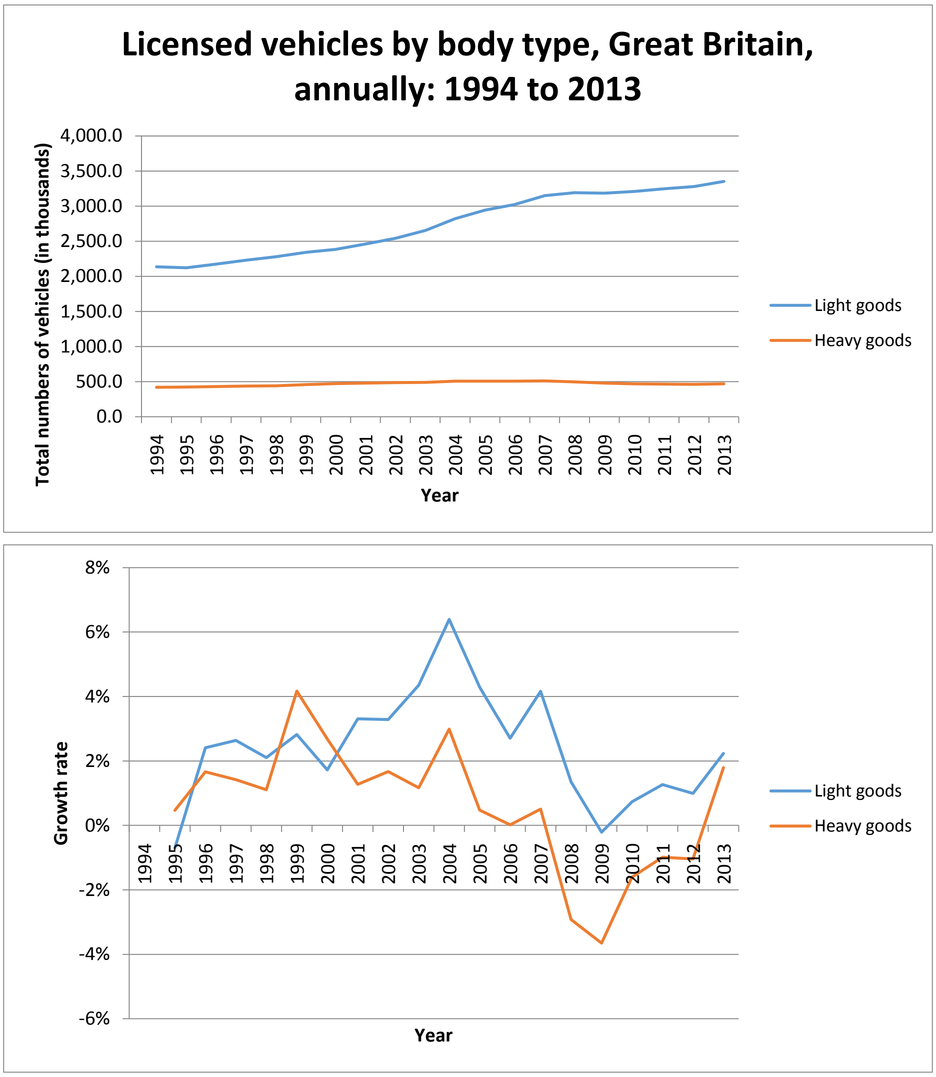


Figure 2.4 - Numbers of LGV and HGVs in Great Britain, absolute and growth rate
(Source: DfT VEH0102)

2.4.4 **Figure 2.5** shows the total vehicle kilometres driven on roads in Great Britain compared with the number of licensed vehicles, as an index from 1994 so that the trends for HGVs and LGVs can be studied[9]. It should first be noted that this is only a rough estimate of the distance covered per licensed vehicle, as the vehicle kilometres will include non-UK licensed vehicles visiting Britain. This may be significant for HGVs, but is likely to be much less so for Cars or LGVs[10].

2.4.5 The chart shows that all vehicle types have experienced a similar pattern, with rising distance per vehicle initially, but a consistent fall since 1998 (or earlier for cars). This suggests the reasons for the fall may be common between vehicle types, possibly linked to the reasons put forward elsewhere for the apparent 'peak car' phenomenon. This might include a general fall in the cost of vehicle ownership compared with vehicle use (i.e. vehicle purchase costs fell relative to fuel and value of time costs). Other possible explanations include:

- Structural changes in logistics and distribution to increase efficiency leading to each vehicle covering shorter journey legs, and therefore lower mileage (HGVs and LGVs);
- Companies may place greater value on having spare vehicles available to meet fluctuations in customer demand;
- Increased operation in urban areas meaning that the overall mileage requirements are lower (distances shorter), or that each vehicle is able to cover a smaller distance per day (lower road speeds). The potential influence of spatial development patterns is considered at greater length at the end of this section;
- Vehicles used by self-employed tradesmen are used chiefly for commuting, and not delivery trips, and hence naturally have lower mileage (LGVs).

2.4.6 It is likely that more than one of these factors is at play, plus additional unforeseen drivers.

2.4.7 For forecasting future vehicle kilometres it would be necessary to consider which has the largest impact, and whether each mechanism is likely to continue into the future.

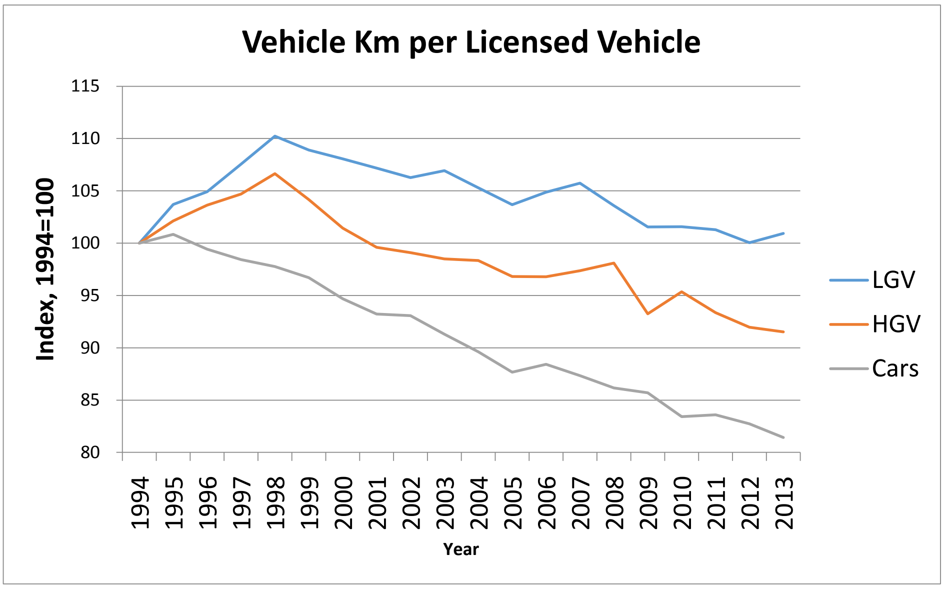


Figure 2.5 - Vehicle Kilometres per Licensed Vehicle
(Source: DfT VEH0102 and DfT Traffic Statistics)
Note: DfT Traffic statistics will include non-UK vehicles, and therefore patterns may represent mix of UK/non-UK traffic

[9] In absolute terms, the HGV figures range from 54,000 to 63,000 km/year, compared to 20,000 to 22,000km/year for LGVs

[10] DfT statistics on foreign registered vehicles have only been available from 2008. DfT TRA3201 estimates that in 2013 3.3% of HGVs were foreign registered, compared to 0.33% of light vans.

2.4.8 **Figure 2.6** shows the age-structure of the national LGV fleet, based on the number of years since first vehicle registration. During the period 1998 to 2007 the average vehicle age fell from 7.5 years to 6.6 years, and thereafter has risen to 8 years in 2013. The chart shows that in the first period the number of new registrations was consistently high, whilst vehicles in the 6-13 age bracket (but not 13+ years) were being scrapped at a higher rate. From 2007, the number of new vehicles drops and this feeds through to the national population. There is some evidence that the reduced rate of vehicle scrapping has contributed to growth of the fleet and its average age: the proportion of the fleet which we estimate was scrapped in 2005 was 7.9%, which has fallen to 5.7% in 2012. However, the age at which vehicles are scrapped has not grown, remaining at ~11 years, with only ~35% of vehicles scrapped being less than 10 years old (WSP estimates from DfT VEH0411).

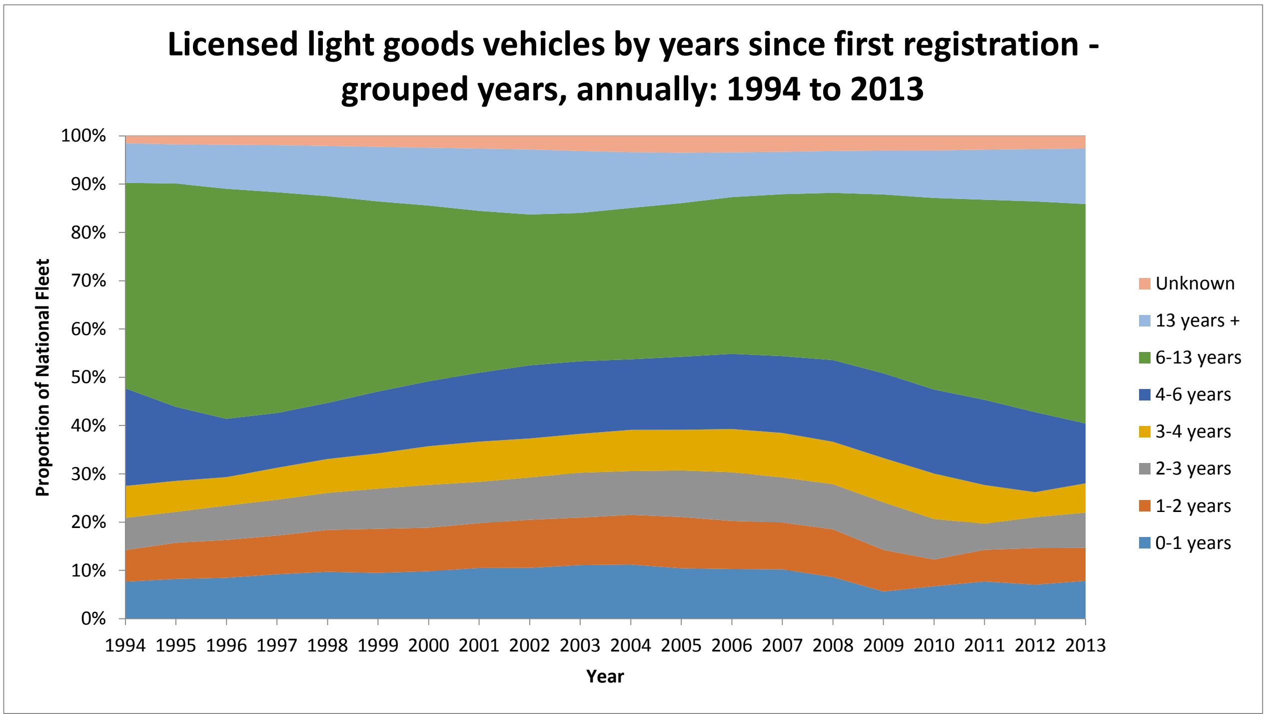


Figure 2.6 - Age of LGV Vehicle Fleet *(Source: DfT VEH0102)*

2.4.9 Taken together, the information in figures 2.3 to 2.6 does not appear to offer any major new insight, but supports previous evidence as to the growth of LGV usage pre-2007 and influence on the recession. It is clear that new vehicles were being added due to additional demand pre-recession, but since 2007 older vehicles are being retained in preference to purchasing new ones. This may relate to smaller firms having less regular vehicle replacement, and holding onto old vehicles, but this cannot be ascertained from the data.

2.4.10 Considering replacement of HGVs by LGVs, there is evidence that this may have occurred in the brief period from 2001 to 2005 when LGV numbers increased rapidly whilst HGV growth fell. The annual mileage of both HGVs and LGVs has been falling since 1998, which must presumably have a common cause.

Potential Influence of Spatial Development

2.4.11 A possible common cause for some of the trends seen above relates to the changing pattern of spatial development in England. Comparison of population Census data over the decades indicates that in the 1980s and 1990s the rate of growth in population was much greater in lower density rural areas than in larger cities. In contrast since 2001 this past trend of deurbanisation has been reversed so that there has been rapid population growth in the inner areas of most metropolitan cities as well as in the economically dynamic cities in the South of the country. Likewise, the rapid growth in out of town supermarkets and other stores seems recently to have largely run its course, with the major chains opening smaller stores within towns and mothballing their plans for the construction of further out-of-town superstores.

2.4.12 These recent trends influence goods vehicle traffic in two stages. Firstly, a substantial part of both LGV and HGV traffic arises to cater for the construction industry. If this construction activity moves from greenfield out-of-town locations within low density sub-regions (i.e. during the 70s to the 90s) to be concentrated instead within brownfield sites in congested higher density inner city areas, then there would be an associated slowing in traffic speeds due to urban congestion and speed limits and a reduction in the need to travel long distances to a dispersed range of sites. Secondly, once the construction has been completed and is occupied, the LGV and HGV movements to service the residents and employees in these new dense urban developments will also experience reduced road speeds and reduced need to travel long distances.

2.4.13 This urban densification trend would also encourage some shifting of deliveries from HGVs to LGVs due to the greater ease of local access and of short-term parking for the latter within congested urban areas,

2.4.14 This urban densification trend provides a possible explanation for why the miles travelled per goods vehicle have shown a downward trend over the last 15 years, Whether, it persists into the future will depend critically on the extent to which CLG removes the strong planning policy measures for densification that had been put in place at the end of the 90s.

2.5 Growth in Self-employed and Small Firms

2.5.1 In testing our hypothesis H1, it is helpful to consider whether the number of self-employed people has in fact risen. From public Census data this can be estimated for the period 2001-2011.

2.5.2 In 2001, there were 3.1m self-employed people in England and Wales, which rose to 4.1m by 2011, a growth of 31%. The increase included 263,000 additional self-employed in the Construction industry (2007 SIC F), and 195,000 in Transport and Communication (SIC H and J), which together accounted for 27% of the net growth[11]. The number of self-employed whose occupation is described as 'skilled trades' rose by 261,000, and 'process plant and machine operatives' by 106,000[12].

2.5.3 The size of registered businesses can also be considered using the UK Business: Activity, Size and Location publication from ONS[13]. This data summarises the number of UK businesses based on HMRC and Companies House data with employee numbers and turnovers. Considering employee numbers, there are some definitional issues with this data which have required some investigation:

- The Standard Industrial Classification (SIC) used was altered in 2009 (from the 2003 to 2007 system), and comparisons of company numbers by industry before and after this period are problematic (though not impossible);
- The dataset prior to 2008 included only VAT-registered businesses, but from 2008 onwards VAT and PAYE based enterprises;
- Prior to 2004, the number of employees was given only for the manufacturing industries;

[11] 2001 Census table S038 compared with 2011 table LC6602EW.

[12] 2001 S035 and 2011 LC6601EW.

[13] Available at: http://www.ons.gov.uk/ons/rel/bus-register/uk-business/2013/index.html

- The definition of a business unit does not relate to individual companies: instead employment size is quoted for 'Local Units' which are defined as individual sites within an enterprise, therefore presumably business restructuring or accounting practices may create new units.

2.5.4 The business turnover is provided in a little more detail, but suffers some of the same definitional issues. Our analysis of turnover however suggested that the fixed ranges used are unhelpful for longitudinal analysis (where correction for GDP or RPI would be required). Therefore we can only draw conclusions from a limited comparison of businesses by employee size, namely comparisons of 2004 with 2007 and 2009 with 2013, both of which provide 'like for like' comparisons:

- From 2004 to 2007, the number of Local Units grew by 77,970, 73% (57,000) of which had fewer than 5 employees. The majority of the growth was in the Construction industry (17,680 new units, 94% <5 employees) and Property (57,250, 82% <5). However, overall the number of small businesses remained very stable during this period (82% of firms with fewer than 10 employees in each year).
- From 2009 to 2013, the number of local units fell by 9,300. The number of small Construction firms fell by 23,000, though Property units of <5 employees again grew (by 26,000) and Post & Telecommunications grew by 15,000. In 2009 82.9% of units had <10 employees, though this fell to 82.7% by 2013.

2.5.5 This shows that the number of small business units grew in absolute terms from 2004 to 2007, when LGV numbers were growing, and fell from 2009 to 2013, when LGV numbers have been growing only slowly. This may support hypothesis H1, though the type of new businesses make this less likely: the 57,000 Property related business include some building services and security, but are otherwise largely administrative in nature; the additional small Construction firms account for at most 5.5% of the 326,000 additional LGVs in the period 2004 to 2007.

2.5.6 The evidence from the Census data is perhaps more encouraging, in that the scale of increase in self-employed from 2001 to 2011 could explain a large amount of the growth. If we assume based on the Census figures that there are 200,000 additional construction, transport and other small tradespeople who are self-employed, and that 50% of these have acquired an LGV, then this could account for 13% of the 787,000 additional LGVs from 2001-2011.

2.5.7 The impact could be more significant if we consider only the ADDITIONAL LGVs caused by the rise in LGV growth in the early 2000s: 2001 to 2011 there was an annual growth rate of 5.7%, compared with an annual rate of only 2.3% from 1995 to 2000. The higher growth caused an additional 246,000 LGVs to be registered. Therefore the additional self-employed could have contributed significantly to the additional LGV growth in this period.

2.5.8 However, such a relationship relies on the fact that the number of self-employed grew prior to the recession, in the period of most rapid LGV growth. It would be worthwhile confirming this by accessing more detailed business survey data for that period.

2.6 Discussion and Conclusions

2.6.1 In this section we have put forward some possible factors which might act to limit the future growth in LGVs. We have chiefly considered factors related to the industry structure and ownership of LGVs, which could explain why the GDP elasticity of LGV vehicle km has been high during the past 10 years. The decline of these influences may allow that elasticity to fall in future, and hence reduce future growth.

2.6.2 We have not however considered the cost and supply side factors put forward in paragraph 2.1.4.

2.6.3 Considering our two hypotheses, there is no absolute concrete data on which to estimate the influence of either. However, there is certainly circumstantial evidence that:

- Private registrations rose steeply in the highest growth period from 2002-2007, which indicates that a new type of ownership was stimulating the higher growth;
- The creation of additional small business units up to 2007 coincided with a growth in LGVs licensed in Great Britain;
- The number of additional self-employed tradesmen is sufficiently large that it could have contributed to the additional growth in LGVs in Great Britain;
- The rise in LGV ownership in 2002-2007 does appear to match a fall in HGV growth in this limited period, and prior to the recession, indicating possible restructuring;
- The fall in both LGV and HGV annual mileage from 1998 is also indicative of some change in vehicle usage, though the precise cause is not clear some potential explanations have been outlined.

2.7 Suggested areas for further investigation

2.7.1 There are a number of definitional issues with the data presented above, which may warrant further investigation and comparison with other studies. For example, it is not clear what type of activity privately registered vehicles are involved in, and whether their owners will be principally self-employed. The growth in the self-employed should also be specifically investigated for the period of most rapid LGV growth from 2002-2007.

2.7.2 It may also be helpful to review the definition of an LGV itself, both to ensure consistency across all the datasets being used, and also to look at changes in the mix of vehicles which are being treated as LGVs, and indeed to check whether vehicles NOT covered by the current LGV definition are now (or have in the past) been used for LGV-type activities. For example, it is known that the size of vehicles in general is increasing, and this has also impacted van: Mini-vans and van versions of estate cars were once common, but are now less often used, and the impact of this on both van use and measurement may require review. This issue is also raised in the RAC report on Van Travel Trends published in April 2014.[14]

2.7.3 The robustness of the analyses above could be greatly improved if it was feasible to repeat the LGV surveys that were last carried out by the Department from 2003-05. These surveys were very informative in identifying which industrial sectors made greatest use of which types of vehicles and of how this usage differed between privately-owned and company owned LGVs. It would be critical that the questionnaires adopted and the general conduct of the surveys were kept as close as possible to the earlier ones. This would ensure that trends over time could be examined consistently and so would maximise the reliability of the insights provided from the new survey. Although, a smaller scale Van Activity Baseline survey was carried out more recently in 2008, it was much simpler in structure and quite different in content from the earlier surveys. This unfortunately has greatly diminished its ability to provide useful insights into trends in LGV usage.

[14] Available at: http://www.racfoundation.org/assets/rac_foundation/content/downloadables/van_report_aecom_100414.pdf

3 Overview of the Trafficmaster dataset

3.1 Overview and introduction

3.1.1 The Trafficmaster dataset contains Global Positioning System (GPS) derived journey times of vehicles. These journey times are derived from the details of the time and vehicle location at which each sampled point along the journey has taken place. These data are used by the Department for publications relating to average journey times for persons and for vehicles on various road types and area types, as well as by the HA for journey time reliability measurement on the trunk road network.

3.1.2 This dataset also contains "origin-destination data collected automatically from on-vehicle GPS systems rather than through a survey. It may be biased towards newer vehicles (with, for example, built-in satellite navigation), but is believed to be the best origin data we have for LGVs, with over 40,000 vehicles in each annual sample." (para. 4.2.5, Sumner, 2014).

3.1.3 The remainder of this section includes:

- Section 3.2: An introduction to remotely sensed traffic data, their potential and limitations;
- Section 3.3: A discussion of the potential uses of Trafficmaster data, and the key tests required to confirm that the data is suitable for these purposes;
- Section 3.4: A summary of metadata received to date describing the data and fields.

3.2 Introduction to remotely sensed data collection

3.2.1 In general the potential to use data sources that are based on remote sensing of large samples of movements is of great interest to modellers. Such sources increasingly appear to have the potential to provide a much more cost-effective alternative to the expensive and infrequent RSI (road-side interview) surveys that traditionally have been used to construct Origin-Destination (O-D) vehicle matrices for use in implementing, calibrating and validating transport models.

3.2.2 These data collection methods include: smart phone navigation; conventional Global Positioning System (GPS) based sat-navs; Automatic Number Plate Recognition (ANPR) systems (e.g. Trafficmaster cameras); fleet management systems; etc. Each of these technologies has certain differences in the range of information that they collect which may be important for certain types of analyses. They all tend to have a number of attractive features in common:

- Coverage of the whole road system, with spatial detail down to individual links or sequences of links;
- Continuous coverage through the hours of the day, days of the week and weeks of the year;
- Large samples, at least before the temporal and spatial disaggregation is applied.

3.2.3 Many of these data sources also have a number of features that may complicate their successful usage within modelling studies:

- **Guaranteeing sufficient accuracy** – the raw data is collected for different purposes to those of modelling studies so that the processing and successful use of this data for modelling analysis is still at an early stage. Accordingly, we may not yet have an adequate knowledge for our purposes regarding the accuracy of the data and the types of errors that may lurk within the data source. Furthermore, the vendors of the data may not be the best people to rely on for unbiased information on data quality so its accuracy and freedom from bias should be checked independently.

- **Understanding the biases** – even if there were no accuracy issues in the data, there might still be challenges in scaling up from the sample obtained (irrespective of how large this sample is) to the relevant vehicle population of England. Information about the characteristics of each sampled vehicle (suitably anonymised) would provide a very good starting point to indicate whether there were major biases in the sample (e.g. to check whether the sample comprises newer and larger vehicles than the fleet as a whole). Provided the nature of any such biases can be quantified, they can be partially off-set by use of differential sample expansion weights. This could then increase the confidence in our population estimates of the vehicle movements of interest to this LGV study.

3.3 Potential of the Trafficmaster dataset and aims of analysis

3.3.1 The current analysis of Trafficmaster data aims to assess the extent to which the data available to the Department can be used to fill some of the major gaps in the evidence base about LGV patterns of movement. In particular the dataset may have potential to provide:

- A spatially detailed analysis of LGV origins and destinations, including vehicle routes;
- An indication of the length of a typical LGV trip, the number of trips made per day and the diurnal profile of trips (e.g. peak, interpeak and night-time movements);
- The continuous nature of the data **may** reveal historic trends related to the increase in LGV use, or at least have potential to track trends going forward.

3.3.2 However, if this data is to be used to represent movements of vehicles in analysis and policy-making, it is necessary to ensure that the dataset provides a representative cross-section of LGVs or at least provides sufficient information to enable any biases to be addressed through reweighting the sample. In the absence of this, the usefulness of the data to measure either growth trends or the spatial patterns of vehicle movements between zones is likely to be limited.

3.3.3 This may not exclude the data from other bespoke applications where bias/representativeness is less likely to be an issue. The data is already in use for measuring average vehicle speeds on given road sections, an application where the ownership, trip purpose or business type are unlikely to bias the overall result. It has also been proposed that the data could be used selectively to identify ‘hotspots’ for LGV movements, e.g. depots or other destinations with a high concentration of LGV activity (noting that this could only ever identify a location which is such a hotspot, but never identify ALL such locations).

3.3.4 However, the current questions focus on whether the data provides information about LGVs which is sufficient to support national forecasting. The key questions to be examined in this respect are:

- to measure how representative the data is w.r.t. the population of LGVs and
- to investigate whether we can access sufficient data about the residential location and other characteristics of the van owner to enable intelligent use to be made of the data.

3.3.5 As we will see in Section 4, the analyses which it has been possible to carry out do not fully address these questions: they are effectively a necessary but not sufficient test of the data’s suitability.

3.4 Metadata for the Trafficmaster dataset

3.4.1 An important early task was to assemble and examine the metadata describing the Trafficmaster dataset. The aim was to understand what this database covered and how the variables within it were defined, captured and processed. We also requested to see any analyses that may already have been carried out by DfT or its sub-contractors relating to the reliability of this Trafficmaster dataset and of its main fleet characteristics and trends.

Overview Information

3.4.2 Some broad information on data collection is available from the Department's Statement of Administrative Sources: Inter-urban Congestion Statistics:

> ***"Trafficmaster Global Positioning System (GPS) tracking data***
>
> Anonymised data from vehicles fitted with Global Positioning System tracking and communications equipment is provided to the Department for Transport by Trafficmaster for use in HATRIS. The data includes vehicle location reports corresponding to the physical location of SRN roads for every 10 seconds of relevant journeys. Speed and journey times are estimated by matching vehicle locations to the links of the HATRIS road network, and then calculating the time taken to traverse the link."

3.4.3 There is also a summary of the metadata within the "Statement of Administrative Sources: Journey Time Data System" as listed in Appendix A, this confirms the following but does not provide any major further insights:

- The main purpose of the system is stated as being to "monitor live congestion levels and allocate the fastest moving routes to vehicles equipped with Trafficmaster's satellite navigation system";
- The data relates to "Vehicles equipped with Trafficmaster's satellite navigation system";
- Intended coverage is "All roads within England" (**i.e. excluding Scotland and Wales**);
- It is stated that: "The Department specifies the version of the road network to which the data are mapped and how the data are validated by Trafficmaster."
- It is further stated that: "The data are checked by both Trafficmaster and the Department before being used for statistical analyses. All statistical outputs are thoroughly quality assured before release".[15]

3.4.4 It may be of interest to separately review the validation steps indicated by the latter two points, however that information has not been received at this stage.

3.5 Background information supplied by DfT and Trafficmaster

3.5.1 This section summarises the metadata and other data descriptions supplied specifically for this project.

3.5.2 We queried the DfT and the Trafficmaster company as to how the data has been collected, the reliability of the data and the coverage. The main conclusions of the discussion were as follows:

- The vehicle age, engine size, type (Petrol or Diesel) and the Vehicle identifier consistent with the DVLA data are **not** available in standard DfT data. It is possible the data could be provided by cross-referencing with the DVLA database (which Trafficmaster state has been done before), but an additional charge would be incurred. It is not known if the registered keeper's geographical location is available, but may be unlikely for data protection reasons.
- Annual mileage from odometer readings at previous MOTs is not available in the standard DfT data. DfT suggests that similar data can be calculated by using Trafficmaster OD data. The Trafficmaster OD data comes with a Vehicle ID field, which make it possible to calculate the annual mileage for each vehicle. This is not the same as odometer readings, but will show the total distance driven by any vehicle in the Trafficmaster vehicle sample over any period.

---

[15] DfT have confirmed that validation steps are carried out to ensure there are now obvious errors in the data (i.e. observations > 0, journey times > 0 etc.). There are some additional rules which look at the relationship between values (such as the relationship between sum of square of journey times vs. average speeds and observations, which was formulised using the Cauchy-Swartz Inequality.

- The contracts for Trafficmaster fitment are long term, but the uniformity of the fleet over time cannot be guaranteed, i.e. the mix of vehicles may change in ways that impact on the statistics derived.
- In Trafficmaster data, a trip is defined as 'ignition on to ignition off'. Although there is no additional validation process, the Trafficmaster data processing ensured that the trip will not be broken up when they pass through the poor GPS coverage areas.
- Trafficmaster GPS vehicles typically achieve high levels of geographical coverage of the Motorway, A roads and B roads in England, with coverage levels in any one month being: M Road Coverage: 99.99%, A Road Coverage: 99.92%, B Road Coverage: 99.83%, C Road Coverage: 51.85%. DfT have not set minimum road coverage criteria within the DfT-Trafficmaster contract.
- Trafficmaster ensure that the fleet remains at over 50,000 vehicles (all types). In December 2013 the TM fleet had 102,800 vehicles in their fleet (53% car, 45% LGVs, 1.4% HGVs and 0.6% other vehicle types) and in June 2014 there were 115,300 vehicles in their fleet (50% cars, 48% LGVs, 1.4% HGVs and 0.6% other).
- The LGVs are defined as vehicles not over 3.5 tonnes in Trafficmaster data.
- The samples of LGV used in Trafficmaster data are derived from the Trafficmaster customer base including Citroen (where the system is standard on all LGV's), BMW and Jaguar Land Rover (where the system is supplied as the manufacturer approved stolen vehicle tracking product). A proportion of customer base from Trafficmaster's Fleet Director GPS Fleet Tracking product sold to vehicle fleet operators are derived as well.

3.5.3 The Citroen and BMW contracts have been in place for more than 5 years, the Jaguar Land Rover contract circa 3 years. Trafficmaster have stated that "Generally vehicles are less than 3 years old", but no other indications of contract dates are available due to confidentiality restrictions.

3.5.4 From the information available, we can therefore conclude that:

- It may be possible to identify the age or detailed type of the vehicle, but this will incur additional cost;
- It is unlikely that the vehicle owner or point of registration would be released by DVLA without considerable efforts to ensure data protection;
- The coverage of roads is very high, and the proportion of LGVs within the dataset is extremely high : vehicle licensing statistics show 9.6% of all registered vehicles were LGVs in 2013 (DfT Table VEH0102), compared with 45% of Trafficmaster vehicles;
- However, the dataset is limited to a small number of vehicle types, and based on indications from TrafficMaster may contain disproportionate numbers of vehicles in large fleets and newer or 'high end' vehicles which are more likely to have security systems fitted. This may therefore not give a true un-biased sample of the national fleet, as these vehicles are likely to have different patterns of usage to vehicles owned by (for example) smaller firms and the self-employed.

3.6 Data Available from Department for Transport

3.6.1 The DfT hold two separate Trafficmaster datasets:

- **Road Link Level GPS Journey Time data: spans across England, data from Sep 2006 to June 2014.** In addition to the 15-minute ITN link-level aggregated GPS journey time data and the Origin-Destination data, Trafficmaster also provide the full route sequence (i.e. which ITN links a vehicle traversed during its trip) for every trip within the English Boundary and for every vehicle in TMs fleet, with the first and last 500 metres of every journey anonymised. This would essentially mean any trips < 1km, would not be included in Trafficmaster trip-level dataset.[16]

[16] This restriction would be very problematic for the significant post and courier section of the market, where distances between successive drops are small. The treatment of these trips would need to be further clarified with Trafficmaster before undertaking any analysis of this data.

- **Origin-Destination Data: spans across the whole of Great Britain (Scotland, Wales, England), data from Sep 2011 to August 2013.** Includes all Origin-Destination trips from ignition on-ignition off at Lower Super Output Area, with Vehicle ID so trips by a specific vehicle can be tracked. For this dataset, the start and end points of each journey are fully captured, so that trips of <1km are included. However, this means that the trip itself and length are reported: the reporting is for aggregate areas which will report most short trips as internal to a reporting area.

3.6.2 It was decided that within the scope of this work only the Origin-Destination data should be obtained and processed. This was partly for reasons of time and budget, as the Link Level data would be much larger and more onerous to analyse. It was also felt that studying the Origin-Destination data first is the correct analytical sequence, as it allows an initial view to be taken on the potential biases in the data before proceeding further.

3.6.3 Furthermore, though data for the entire period 2006-2013 was provided, the format differed before and after 2011, and it was decided for reasons of time to focus on the post-2011 data. Subsequently it became clear that the volume of data produced for 1 year was too large to easily manipulate in the formats available from the DfT[17], therefore it was agreed that further analysis should focus on a single calendar month.

Structure and Content of the Origin-Destination data

3.6.4 0B contains a record of the Origin-Destination files provided by the DfT and the format of these files. For our purposes the key points are as follows:

- Each record represents a single trip (ignition-ignition) by a vehicle, with a vehicle ID which allows trips by a single vehicle to be tracked through the dataset;
- The OD data contains the start and end zones (based on Lower Super Output Areas), associated start and end date/time, distance travelled and time taken for trip;
- The vehicles are divided into types, of which type 2 is defined as 'LCVs (up to 3,500kg)' (our analysis has therefore extracted Type 2 vehicles only);
- The distance travelled is divided between the Strategic Road Network (SRN) and Local Roads (non-trunked A roads, B roads and unclassified).

[17] DfT were able to export a single 'flat' text file containing a single year of data, but this file was 14GB and 140million records. It is impractical to process such a file in standard Windows software available during this project, though some packages are available and could be obtained should this become necessary.

4 Analysis of Trafficmaster Sample

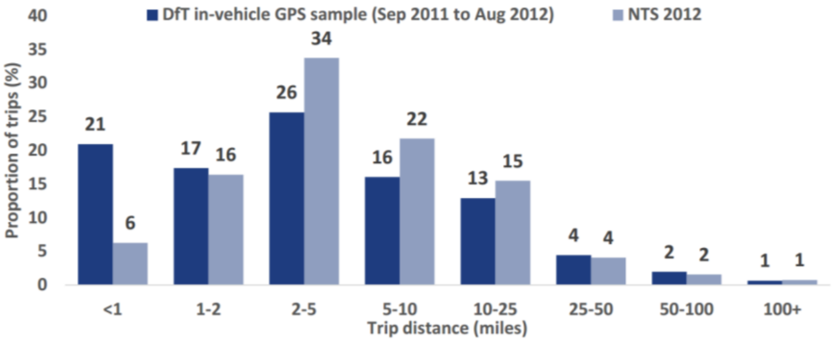
4.1 Introduction

4.1.1 This section of the report contains our analysis of a sample of the Trafficmaster Origin-Destination dataset. The final dataset supplied by DfT was a single month of this data from 16 September to 15 October 2012. The dataset contained 5,712,542 trip records for LGVs, across 32,537 unique vehicles.

4.1.2 The next section presents some further top level analysis of Trafficmaster undertaken by the DfT which is helpful to understand the Trafficmaster data. Thereafter, we present a series of summaries of the LGV dataset supplied for this study.

4.2 Comparison of Trafficmaster with other national data sources

4.2.1 The DfT has undertaken some analysis to assess the representation of Trafficmaster vehicle driving behaviour relative to the general vehicle population for England [18]. **Figure 4.1** below shows a comparison of trip distances between Trafficmaster and NTS survey trips for Car/Van drivers and passengers. This shows clearly that the Trafficmaster data has an excess of trips <1 mile compared with NTS, and a smaller proportion of trips of 2-25 miles.



Sources: DfT in-vehicle GPS data, 78.6 million trips between Sep 2011 and Aug 2012 (GB vehicle coverage); DfT National Travel Survey 2012, 217,091 trips as 'Car/Van driver' and 'Car/Van passenger' (GB coverage)

Figure 4.1 - Comparison of Trip Distances : DfT Trafficmaster data vs DfT National Travel Survey 2012
(extracted from 'Use of the Strategic Road Network', TRADFT006, DfT August 2014)

4.2.2 In their summary to the report, DfT suggest that NTS respondents may round-up the length of short trips, and conversely have access to more accurate GPS-based data for longer trips. Considering the Trafficmaster data, it is also very likely that the ignition-ignition recording of trips may artificially 'create' extra short-trips, should the journey pause and the engine be turned off (e.g. shopping, school drop-offs, or even vehicle stalls).

4.2.3 It is also worth noting that the proportion of vans in the Trafficmaster dataset in 2014 (48%) is greatly above that in the national vehicle fleet as a whole. There is therefore a likelihood that the comparison including both Cars and LGVs in Figure 4.1 includes some bias.

[18] https://www.gov.uk/government/statistics/use-of-the-strategic-road-network

4.2.4 **Figure 4.2** shows a comparison of the Trafficmaster and NTS annual mileage data specifically for LGVs. This shows a reasonably good comparison of mileages record by Trafficmaster and NTS, but an excess of vehicles with very high mileage (<21k per annum) in the Trafficmaster data.

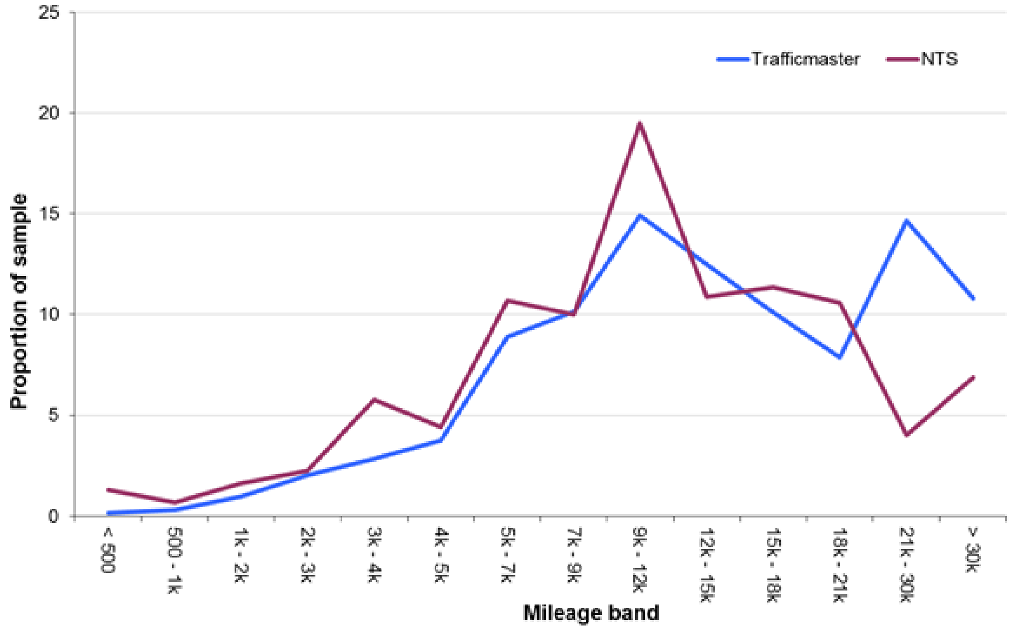


Figure 4.2 - Comparison of Annual Vehicle Mileage for LGVs - Trafficmaster vs NTS
(supplied by DfT, derived from Trafficmaster and 2011 NTS Table NTS0904)

4.2.5 These national comparisons are useful confirmations that the Trafficmaster data has strong similarities to NTS, with possible biases in its excess of short trips (possibly due to the collection method) and a bias towards high annual mileage vehicles (possibly due to the sample selection).

4.3 Trafficmaster Sampling method vs National Fleet

4.3.1 Only anecdotal information on the sample of vehicles in the Trafficmaster data is available. This advises that the vehicles are in general <3 years old, and either high end specifications, part of company fleets or made by Citreon, BMW or Jaguar Land Rover.

4.3.2 Figure 2.6 of this report shows the age structure of the national vehicle fleet, indicating that only 20-25% of vehicles are typically less than 3 years old. Furthermore, the discussion there suggests that the ownership and usage of vehicles may differ widely by age.

4.3.3 DfT Vehicle registration data (Table VEH0124) also allows the number of LGVs by manufacturer to be judged. Citroen vehicles account for 7% of all LGVs, whereas BMW and Jaguar Land Rover account for only 4% between them. (Ford and Vauxhall together account for 39%).

4.3.4 On this basis, it appears that Trafficmaster are drawing their data from a sample of the national fleet which may not be representative of the national fleet.

4.4 LGV data sample analysis

4.4.1 In this section we describe a series of tests carried out on the Trafficmaster LGV sample, to answer specific questions about the data.

4.4.2 In what follows, all data presented is based on LGV movements only in the DfT supplied Trafficmaster dataset for 16 September-15 October 2012. The analysis is presented by English Region, plus Scotland and Wales (though these are simply for notational convenience counted as 'regions').

Test 1: Does the Trafficmaster data provide reasonable samples of trips in all English regions, and plausible region-region movement patterns?

Table 4.1 - Total LGV Origin-Destination Movements recorded by Region (yellow bars indicate magnitude of cell entry)

	N East	N West	York&H	E Mids	W Mids	East	London	S East	S West	Wales	Scotland	Total
N East	362,379	794	3,906	225	69	42	13	15	4	13	766	368,226
N West	814	674,318	6,527	2,877	5,271	203	124	308	157	4,144	1,503	696,246
York&H	3,853	6,625	493,516	9,853	997	409	128	219	85	123	121	515,929
E Mids	228	2,829	9,779	382,852	11,064	6,308	903	4,217	371	160	29	418,740
W Mids	87	5,341	1,022	10,906	486,843	1,159	833	2,639	3,911	2,405	55	515,201
East	56	176	435	6,373	1,157	554,336	31,687	11,927	357	65	9	606,578
London	8	113	98	883	712	31,809	711,824	41,435	546	59	1	787,488
S East	14	253	204	4,269	2,727	12,004	41,127	811,195	9,285	296	6	881,380
S West	6	141	75	376	3,978	402	565	9,162	517,235	2,382	2	534,324
Wales	7	4,094	102	151	2,304	53	78	299	2,485	176,687	8	186,268
Scotland	775	1,485	129	40	52	10	2	5	5	6	199,653	202,162
Total	368,227	696,169	515,793	418,805	515,174	606,735	787,284	881,421	534,441	186,340	202,153	5,712,542

4.4.3 **Table 4.1** shows the total Trafficmaster LGV movements on a Region-Region basis. This demonstrates immediately a reasonable spread of coverage across all English regions, as well as Wales and Scotland. As might be expected, the majority of movements appear to be within the regions themselves. This is confirmed by **Table 4.2** and **Table 4.3** below, which show the distribution of trip origins and destinations for each region. In each case, in excess of 90% of trips start and end in the same region. This is the case both for the English regions and Scotland and Wales, demonstrating that the Trafficmaster sample includes vehicles operating exclusively in these areas.

4.4.4 As might be expected, the least self-contained regions are London, the South East and East of England, due mainly to the high number of London trips starting or ending in the other two regions.

Table 4.2 - Distribution of Trip Destinations, by Origin

	N East	N West	York&H	E Mids	W Mids	East	London	S East	S West	Wales	Scotland	Total
N East	98%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	100%
N West	0%	97%	1%	0%	1%	0%	0%	0%	0%	1%	0%	100%
York&H	1%	1%	96%	2%	0%	0%	0%	0%	0%	0%	0%	100%
E Mids	0%	1%	2%	91%	3%	2%	0%	1%	0%	0%	0%	100%
W Mids	0%	1%	0%	2%	94%	0%	0%	1%	1%	0%	0%	100%
East	0%	0%	0%	1%	0%	91%	5%	2%	0%	0%	0%	100%
London	0%	0%	0%	0%	0%	4%	90%	5%	0%	0%	0%	100%
S East	0%	0%	0%	0%	0%	1%	5%	92%	1%	0%	0%	100%
S West	0%	0%	0%	0%	1%	0%	0%	2%	97%	0%	0%	100%
Wales	0%	2%	0%	0%	1%	0%	0%	0%	1%	95%	0%	100%
Scotland	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	99%	100%
Total	6%	12%	9%	7%	9%	11%	14%	15%	9%	3%	4%	100%

Table 4.3 - Distribution of Trip Origins, by Destination

	N East	N West	York&H	E Mids	W Mids	East	London	S East	S West	Wales	Scotland	Total
N East	98%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	**6%**
N West	0%	97%	1%	1%	1%	0%	0%	0%	0%	2%	1%	**12%**
York&H	1%	1%	96%	2%	0%	0%	0%	0%	0%	0%	0%	**9%**
E Mids	0%	0%	2%	91%	2%	1%	0%	0%	0%	0%	0%	**7%**
W Mids	0%	1%	0%	3%	95%	0%	0%	0%	1%	1%	0%	**9%**
East	0%	0%	0%	2%	0%	91%	4%	1%	0%	0%	0%	**11%**
London	0%	0%	0%	0%	0%	5%	90%	5%	0%	0%	0%	**14%**
S East	0%	0%	0%	1%	1%	2%	5%	92%	2%	0%	0%	**15%**
S West	0%	0%	0%	0%	1%	0%	0%	1%	97%	1%	0%	**9%**
Wales	0%	1%	0%	0%	0%	0%	0%	0%	0%	95%	0%	**3%**
Scotland	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	99%	**4%**
Total	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

4.4.5 A further check was carried out to determine the 'symmetry' of vehicle movements in the dataset: whether the number of movements starting/ending in each region matches the total movements in the opposite direction. This is a necessary sense-check because it confirms that each outward trip has a corresponding return trip, and vice versa. A non-symmetric matrix would suggest either that a significant number of trips are missing, or that many vehicles are making one-way trips.

4.4.6 In fact, **Table 4.4** confirms that the number of mis-matched trips is low. The highest absolute difference is 308 for South East-London trips, which is 0.7%: this is well within the limits if a small number of vehicles were to make two-day trips at one end of the time period, and hence not return. More significant may be the mis-match of 121 from London-West Midlands (17%) or 30 from London-Yorkshire and Humberside (30%). However, the number of trips involved is so small as to be immaterial for the wider analysis.

Table 4.4 - Symmetry Test: Destination-Origin subtracted from Origin Destination Trips

	N East	N West	York&H	E Mids	W Mids	East	London	S East	S West	Wales	Scotland
N East	0										
N West	20	0									
York&H	-53	98	0								
E Mids	3	-48	-74	0							
W Mids	18	70	25	-158	0						
East	14	-27	26	65	-2	0					
London	-5	-11	-30	-20	-121	122	0				
S East	-1	-55	-15	52	88	77	-308	0			
S West	2	-16	-10	5	67	45	19	-123	0		
Wales	-6	-50	-21	-9	-101	-12	19	3	103	0	
Scotland	9	-18	8	11	-3	1	1	-1	3	-2	0

Test 2: Does the Trafficmaster data provide a meaningful sample of vehicles in each region?

4.4.7 There are 32,537 unique LGVs in the Trafficmaster sample, compared with 3,209,000 LGV registrations recorded by the DVLA in 2012. This means that the Trafficmaster data provides a sample of 1% of the total registered LGVs.

4.4.8 We next wished to ensure that the Trafficmaster data is representing vehicles equally in each of the regions. As no vehicle registration data is available for the Trafficmaster data, it was assumed that each vehicle's 'home' region is the location of its most frequent trip origin. As we will see, this is an arbitrary assumption not necessarily reflecting vehicle registration, though it should reflect where the vehicle spends the majority of its time[19].

[19] Inter alia, it was found that for more than 99% of vehicles the most frequent trip was within a single region, whereas only 94% of the Trafficmaster trips as a whole are contained within one region, further re-inforcing the view that travelling outside the home region is unusual for the vast majority of vehicles.

4.4.9 **Figure 4.3** shows the 'home' vehicles in each region as a proportion of the DVLA registrations in 2012. This suggests that the 'sample' is above 0.75% in all English regions, though slightly lower in Scotland and Wales. The latter may not be surprising, as Trafficmaster's own information suggests they are more focused on vehicles registered in England. The higher figure for London and lower figure for the East and West Midlands are discussed below.

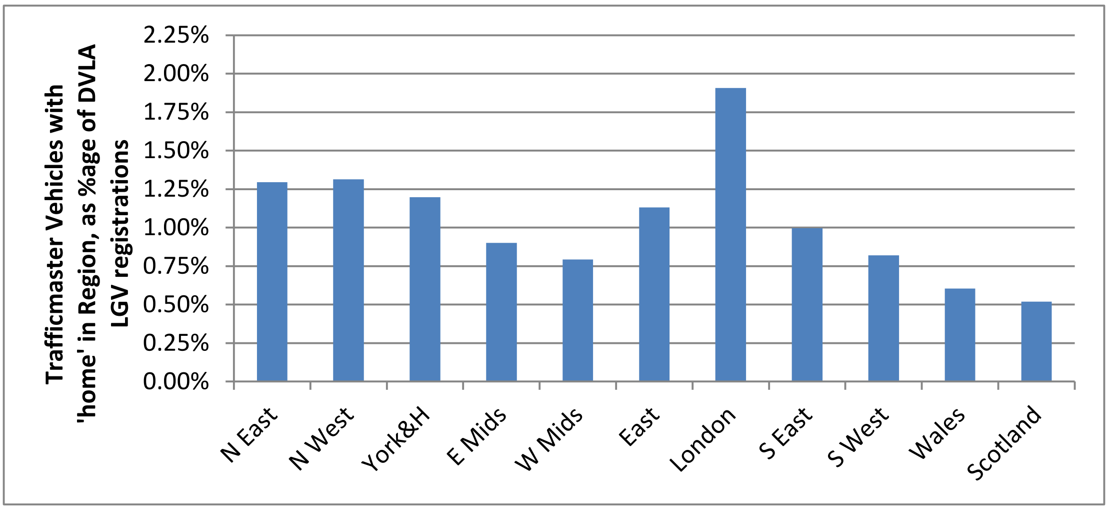


Figure 4.3 - Trafficmaster 'home' vehicles as a percentage of the DVLA registrations in each region
(Source of DVLA data: DVLA veh0404, 2012)

4.4.10 **Figure 4.4** shows a comparison of the distribution of trip origins, the assigned 'home' regions for Trafficmaster LGVs, and DVLA registrations. There is some disparity between the distribution of trip origins and 'home' regions for vehicles. This is most obvious in London: it is very likely that many vehicles occasionally visit London, but are not based there, which would easily explain this.

The overall pattern of Trafficmaster 'home' regions and LGV registrations is reasonably similar which is again reassuring, but there are obvious exceptions. London has only half the share of vehicle registrations compared with 'home' vehicles in the Trafficmaster dataset (6.3% vs 11.9%). Therefore at least 50% of the vehicles operating within London are not actually registered to owners in London. This is a very large proportion, but may not be surprising for business fleet vehicles which may commonly be registered at company depots in cheaper areas outside Greater London.

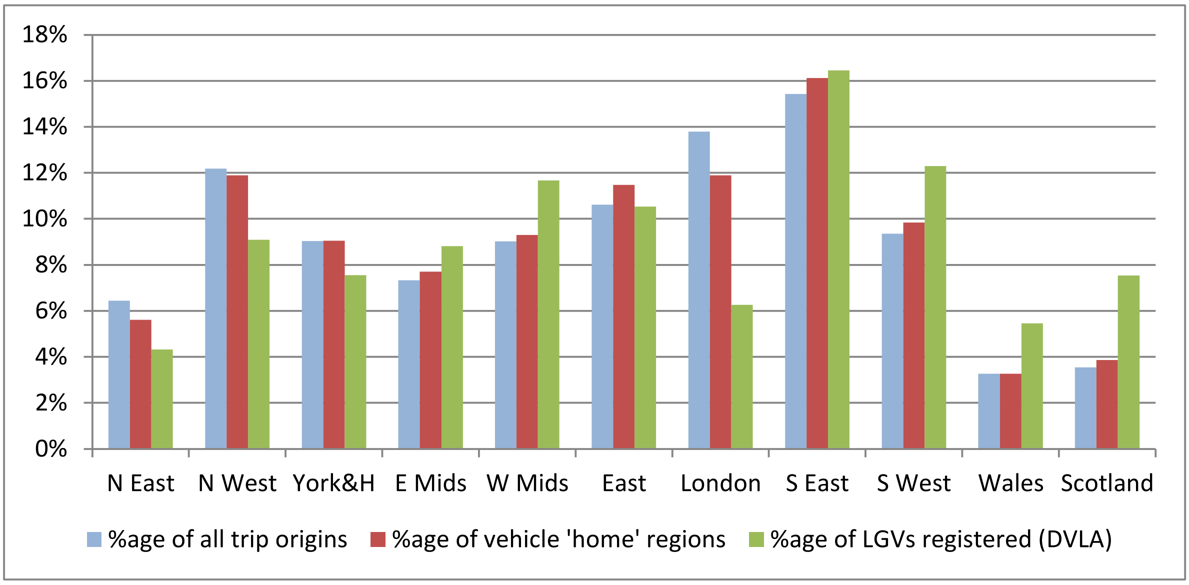


Figure 4.4 - Comparison of Trafficmaster trip origins, Vehicle 'home' regions and vehicle registrations

4.4.11 There are similar discrepancies in the North West, Yorkshire & Humberside and the North East. This is balanced by the reverse effect in the West Midlands, South West, Wales and Scotland, i.e. there are a greater proportion of vehicles registered than apparent from the Trafficmaster data. As would be expected, these findings replicate the pattern in **Figure 4.3.**

4.4.12 There may be varying explanations for these discrepancies:

- the Trafficmaster company may have limited reach in Wales and Scotland, and greater market penetration in London and the North of England;
- Vehicle registrations may be concentrated in the Midlands and South West. The former may be sensible based on the focus of the logistics industry in the Midlands;
- Given that the Trafficmaster system is often fitted to fleet vehicles, it may be that a higher proportion of vehicles in the Midlands and South West are operated by small companies, who will tend not to have Trafficmaster systems fitted (and conversely elsewhere).

4.4.13 Some further investigation was carried out to ascertain the behaviour of the vehicles whose most frequent trips were within London. It was hypothesised that a proportion of these may be based outside London, or on its outer fringes, and travel into London at the start of the day. There is some evidence in the dataset to support this view:

- LGVs which make mainly London-London trips have 82.8% of their trips within London, and 4.7% into London from outside (mainly the South East or East). However, of the first trips made by these vehicles after 6am, only 76% are within London, and 9.5% are into London. There is also a slight increase in the number of trips travelling out of London, from 4.7% for all day trips, to 7.2% for the first trip after 6am.
- The first trip of the day WITHIN London is also longer, rising from 5.6km for all day trips, to 8.5km for the first trip of the day.

4.4.14 Taken together, these results support the view that LGVs may travel into London for the working day, though the extent of the effect is unlikely to be enough to entirely explain the discrepancy between DVLA registrations and London operation in the Trafficmaster dataset. It must therefore still be the case that there are many vehicles in the Trafficmaster dataset which are registered to non-London owners but spend the vast majority of their time in London.

Test 3: Sense-check of Number of Trips per Day and Average Trip Length

Table 4.5 - Average Trips per Day, most frequent O-D pair of each vehicle

	N East	N West	York&H	E Mids	W Mids	East	London	S East	S West	Wales	Scotland	Total
N East	6.3	0.0	0.5									**6.3**
N West		5.5	0.0	0.5	0.7					0.9		**5.5**
York&H			5.2	0.1	0.0							**5.2**
E Mids			0.5	4.6	0.2	0.3	0.0	0.3				**4.5**
W Mids					5.0			0.7	0.1			**5.0**
East				0.0		4.5	0.8	0.2				**4.4**
London				0.0		1.4	5.5	0.9	0.6			**5.5**
S East				0.1	0.6	0.0	0.1	4.7	0.3			**4.6**
S West								0.4	5.0			**5.0**
Wales					0.5					5.0		**5.0**
Scotland											5.0	**5.0**
Total	**6.3**	**5.5**	**5.2**	**4.5**	**4.9**	**4.5**	**5.5**	**4.6**	**5.0**	**5.0**	**5.0**	**5.0**

Table 4.6 - Average LGV Trip Length (km) for Trafficmaster dataset (all trips)

	N East	N West	York&H	E Mids	W Mids	East	London	S East	S West	Wales	Scotland	Total
N East	8	131	80	240	312	357	434	473	446	364	108	**9**
N West	134	10	79	83	87	306	335	304	283	52	126	**12**
York&H	81	79	9	58	171	234	298	305	321	224	313	**13**
E Mids	239	81	58	9	53	74	159	70	199	233	417	**14**
W Mids	315	89	171	54	10	171	193	124	79	84	412	**14**
East	374	306	232	73	167	11	42	64	231	310	567	**15**
London	443	341	307	160	195	42	6	41	176	255	631	**9**
S East	441	297	289	67	121	64	41	11	67	206	573	**14**
S West	477	272	318	204	81	228	175	67	11	91	489	**14**
Wales	406	51	223	226	81	312	260	207	87	12	463	**15**
Scotland	108	128	324	422	431	590	578	563	503	457	14	**16**
Total	**9**	**12**	**13**	**14**	**14**	**15**	**10**	**14**	**14**	**15**	**15**	**13**

4.4.15 Based on the total vehicles and trips in the dataset, we can calculate that on average vehicles make 5.7 trips per day. By assigning each vehicle to its most frequent O-D pair, we can also estimate how this varies according to the trips being made. **Table 4.5** shows the average number of trips per day made by LGVs, calculated by considering only the trips made by vehicles in their MOST FREQUENT O-D pairs. The overall trip average is 5.0, reflecting the fact that vehicles make 88% of their trips on the most frequent O-D pair. Overall, the results appear sensible: within region trips are more frequent, and longer distance trips less frequent. There is no strong frequency difference between the regions.

4.4.16 **Table 4.6** shows the average trip length for each O-D pair, considering all LGV trips in the dataset. This again seems reasonable, with trip lengths varying plausibly according to O-D distance, and London-London trips being shortest of all categories.

Test 4: Proportion of SRN and Local Road Traffic

4.4.17 The Trafficmaster O-D dataset includes the distance covered in each trip on the Strategic Road Network (SRN) and locally-managed roads[20]. This provides the opportunity to compare the proportion of trips on the SRN with that estimated by DfT as part of their traffic statistics. Overall, the DfT statistics show that 32.5% of LGV vehicle mileage in England was on the SRN in 2012 (DfT Traffic Statistics TRA0104). For the Trafficmaster dataset this figure is 42% (counting trips from England, including those to Scotland and Wales). This indicates a much higher proportion of Trafficmaster mileage on the Strategic Road Network than reported by DfT.

[20] Note: though the DfT traffic statistics identify SRN roads within Scotland and Wales which are not Highways Agency managed, the Trafficmaster data appears to count all roads outside England as 'Local'. Therefore Scotland and Wales have been explicitly excluded from this analysis.

4.4.18 To investigate this further, **Figure 4.5** provides a comparison of the figures by region. The general pattern shown here appears reasonable, in that London has the lowest proportion in both datasets, and the North West the highest. Generally there is a good correlation and the figures are quite close.

4.4.19 However, it must be noted that the DfT statistics are network-based and therefore include all distance covered on roads within a region, even those passing through. Conversely, the Trafficmaster figures shown are for trips with both origin and destination within the region. Such a comparison would be expected to indicate a higher proportion of SRN distance for the DfT dataset than Trafficmaster, as the longer-distance trips between regions would be more likely to use major roads. This is the case in most regions, though the difference is not great: a maximum of 9% (in the East Midlands), with the North West and South West having a greater proportion of SRN mileage in the Trafficmaster dataset. This evidence may indicate that the Trafficmaster data is biased towards trips on the Strategic Road Network, whether because of the vehicles and trips recorded, or issues with definitions and allocations.

Figure 4.5 - Proportion of LGV Vehicle Km on Strategic Road Network for Trafficmaster and DfT Traffic Statistics (2012)

Test 5: Does Average LGV Annual Vehicle Kms match DfT figures?

4.4.20 As a final test, we have compared the estimated annual vehicle kilometres from Trafficmaster with that which can be obtained from DfT and DVLA figures. For the latter, we have taken the total DfT LGV kilometres used in Test 4, and divided by the DVLA LGV registrations used in Test 2. These annual figures have been compared with a pro-rata scaling of the Trafficmaster data to give an annual estimate.[21]

4.4.21 The overall result showed 25,538km per year on average from the DfT dataset, and 26,890km per year for the Trafficmaster dataset, a 5.3% difference%. This may be sufficiently different to warrant further investigation.

[21] It would undoubtedly be better to take an entire year of Trafficmaster data for this comparison. The proportion of weekends and bank holidays in the 1 month sample was checked to ensure that this did not introduce any bias when scaling up.

4.4.22 **Figure 4.6** provides a regional view of this data, though again there is difficulty in comparing the two datasets, and in this case it is more difficult to draw conclusions. The DfT data is based on network vehicle km and places of registration, whereas the Trafficmaster data can only provide a rough indication of these based on trip origins and destinations. To account for this, two estimates of vehicle mileage from Trafficmaster are provided: based on trips within the region only (a low estimate), and based on all trips originating in the region (high). The results show that the figure estimated using DfT statistics lies mainly between these estimates, though there is no obvious regional pattern. This at least indicates that no specific region is contributing to the overall difference, but it seems no further inference can be made at this point.

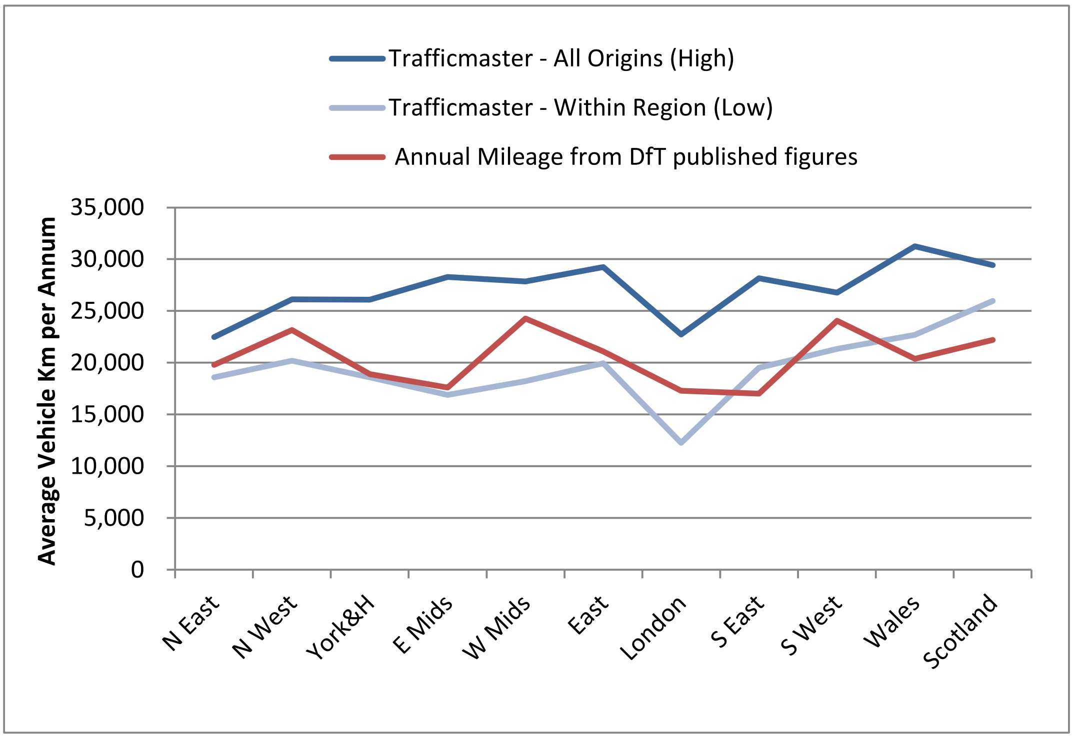


Figure 4.6 - Comparison of DfT and Trafficmaster estimates of LGV Vehicle Annual Vehicle Km

4.5 Conclusions on Trafficmaster data

4.5.1 The tests described above are designed to constitute a necessary, but not sufficient, set of checks on the Trafficmaster data. It is important to note that we have assessed only whether the data is helpful for understanding levels of LGV ownership and traffic and NOT to measure speeds or precise local movements of LGVs.

4.5.2 The most significant conclusion relates to the sample of vehicles present within the dataset. Comparison against the national fleet of the age distribution and of the manufacturers indicates that the Trafficmaster fleet is much newer on average than the national fleet. Furthermore, because it is focused primarily on just three manufacturers it does not cover adequately the two largest selling manufacturers Ford and Vauxhall, which together comprise 39% of the national fleet. There are also good reasons to believe (though this cannot be demonstrated absolutely), that the Trafficmaster fleet is biased towards larger firms and high-end vehicles. For all these reasons, substantial doubts arise over how the data gathered from the Trafficmaster data could safely be generalised to the national level.

4.5.3 More encouragingly, the 5 tests described in section 4.4 do not indicate any major issues or gaps in the data: all of the regions appear to be represented, with reasonably sensible numbers of trips and journey lengths. These conclusions may be encouraging when using the Trafficmaster data for national tracking of certain LGV trends. However, the bias in the sample of vehicles would be of significant concern were this data to be used in place of other existing sources.

4.5.4 There is some indication that the journey lengths are longer than indicated in other DfT datasets. Corresponding to this, there appears to be a significantly high proportion of distance travelled on the Strategic Route Network by Trafficmaster vehicles than DfT's own data would suggest. Both of these may be explained by the greater proportion of newer and fleet vehicles present in Trafficmaster

4.5.5 Furthermore, issues of definition of the home-location make many further analyses difficult. As outlined in the introductory remarks, knowledge of the vehicle owner and their home location are important for forecasting, as this information can be used as an input to drive forecasting.

4.6 Suggested Further Work

4.6.1 This analysis has aimed only to provide a series of 'necessary' tests to indicate possible issues with the Trafficmaster data, and has thus far considered only the regional level of aggregation.

4.6.2 Given that only very minor gaps or bias at the regional level have been observed, it would be worthwhile proceeding to analyse the data by Local Authority (LA), and perhaps lower levels of detail. Vehicle ownership and traffic statistics are published by the DfT at the LA level, which would mean that Tests 1 and 2 in Section 4.4 can be completed. However, the other tests would be more problematic. However, the major advantage of mapping the data at LA level or below would be to identify whether the low level pattern of vehicle origins and destinations is plausible, or obviously biased.

4.6.3 It may additionally be feasible to interview samples of the Trafficmaster LGV customers, and a random sample of van operators from the general population, to measure more directly the differences in types of owner (size of business, company vs personal etc.) and the usage of the vehicles.

4.6.4 Linked to the study of vehicle types suggested in Section 2.7, there may also be merit in analysing the types of vehicles supplied by Citroen, BMW and Jaguar Land Rover, to compare against the Vauxhall and Ford brands which constitute a larger part of the LGV market. The detailed body types and numbers of vehicles registered are available through DVLA statistics, and could be linked either intuitively or through further primary research to types of vehicle use.

4.6.5 Each of these exercises would give an improved understanding of the limitations of Trafficmaster data, and would indicate what information can be extrapolated from the dataset, and where drawing conclusions may prove unhelpful.

Appendix A TrafficMaster data outline with "Statement of Administrative Sources: Journey Time Data System"

"Administrative/Management Source or System

Name/Title of the original administrative/management data source

Trafficmaster journey time dataset

Name of the organisation responsible for the original data source

Trafficmaster

The purpose of the administrative/management system/source

To monitor live congestion levels and allocate the fastest moving routes to vehicles equipped with Trafficmaster's satellite navigation system

Unit of Inquiry (Claimants, Taxpayers, Households, Accidents, etc)

Vehicles equipped with Trafficmaster's satellite navigation system

Intended coverage of administrative/management system/source (here if this was a survey what would have been the sampling frame)

All roads within England

Completeness i.e. actual coverage (e.g. take-up rate)

The exact coverage varies from month to month based on the movements of Trafficmaster's subscribers

Geographical coverage of statistical product (eg UK, GB, England and Wales, England)

England

Lowest level of geographical coverage (eg local Authority, Postcode area, Postcode, etc)

Individual road links on Ordnance Survey's Integrated Transport Network

Extent to which statistical end-producers can influence the system (do you have much control over the information collected)

The Department specifies the version of the road network to which the data are mapped and how the data are validated by Trafficmaster.

Data definitions used

Vehicle Journey Time – the average time it takes a vehicle to travel one mile on the specified routes and time periods.

Person Journey Time – the average time it takes a person to travel one mile on the specified routes and time periods.

Classification systems used

Ordnance Survey's Integrated Transport Network

The data collection process

- Periodicity/Timing

  GPS data are collected by Trafficmaster on a constant basis. The data are processed and provided to the Department every month.

- Validation procedures

  The data are checked by both Trafficmaster and the Department before being used for statistical analyses. All statistical outputs are thoroughly quality assured before release "

Appendix B Data files received from Department for Transport

The files received from DfT were in two parts, with slightly different formats and zoning systems

- Data from September 2006 to August 2011: which was formatted using the National Transport Model (NTM) zones, and
- Data from Sept 2011 to Aug 2013: which was formatted using ONS Local Super Output Areas (LSOAs).

List of files received:

Data from Sept 2006 to Aug 2011:

- Data Defintions for OD Data Sep 2006 to Aug 2011.doc
- OD Data Sep 2006 to Aug 2011.zip contains 5 CSV files:
  - OD Data Sep 2006 to Aug 2007.csv
  - OD Data Sep 2007 to Aug 2008.csv
  - OD Data Sep 2008 to Aug 2009.csv
  - OD Data Sep 2009 to Aug 2010.csv
  - OD Data Sep 2010 to Aug 2011.csv
- PASS3 NTM OD Zones used in OD Data from Sep 2006 to Aug 2011.zip containing a shape file format of PASS3 NTM OD zones used for defining OD grids used in 2006-2011 data.

Data from Sept 2011 to Aug 2013:

- Two CSV files “OD Data Sep 2011 to Aug 2012” and “OD Data Sep 2012 to Aug 2013” containing Trafficmaster OD data for all trips associated with Trafficmaster’s fleet of vehicles for academic years 2011-12 and 2012-13 respectively.
- A zip file “ODZones_NetworkFile” containing a shape file format of the LSOA Zoning System used for defining the OD grids used in the data.
- Guide To OD Data Delivery Sep 2011 to Aug 2013.doc
- OD Data Field Definitions Sep 2011 to Aug 2013.doc

Origin-Destination Fields in the 2011-12 and 2012-13 Data Sample (*Source: from the DfT file 'ODDataFieldDefinitions.doc' supplied by e-mail 27 August 2014***)**

<u>Variable list:</u> Each record in both OD datasets (2011-12 and 2012-13) relates to a single trip from a Trafficmaster vehicle. Note: a trip in the Trafficmaster data set is defined as being from "ignition on" to "ignition off" status for the vehicle.

Field Name	Format	Units / Resolution	Description
VehID	Integer	1 to 999999	A random integer assigned to each vehicle uniquely. The vehicle ID remains with the vehicle until the life-time of it's subscription.
VehType	Char(10)	"car", "lgv", "hgv", "other"	A description of the body type of the vehicle (e.g. car, LGV, HGV or Other vehicle type)
OriginLSOAZone	Char(9)	Lower Super Output Area (2011 census)	Origin Lower Super Output Area of Trip
DestinationLSOAZone	Char(9)	Lower Super Output Area (2011 census)	Destination Lower Super Output Area of Trip
OriginArea	Char(20)	Larger Area (Local Highway Authority)	Origin Area as defined by polygons within the LCR.tab, Glasgow.tab and Edinburgh.tab MapInfo files.
DestinationArea	Char(20)	Larger Area (Local Highway Authority)	Destination Area as defined by polygons within the LCR.tab, Glasgow.tab and Edinburgh.tab MapInfo files.
OriginDateTime	"dd/mm/yyyy hh:mi:ss"	Seconds	Date/Time when the vehicle started its journey
DestinationDateTime	"dd/mm/yyyy hh:mi:ss"	Seconds	Date/Time when the vehicle ended its journey
TotalDistance	Integer	metres	Distance travelled by vehicle from origin to destination
DistanceLocalRoads	Integer	metres	Distance travelled by vehicle on locally managed roads (Local A roads, B and unclassified roads)
DistanceSRNRoads	Integer	metres	Distance travelled by vehicle on Strategic Road Network (Highway Agency Managed Roads: Motorways and A-Trunk Roads)
TotalTravTravelTime	Integer	seconds	Time taken by vehicle from origin to destination
TravelTimeLoLocalRoads	Integer	seconds	Time spent travelling on locally managed roads (Local A roads, B and unclassified roads)
TravelTimeSRNRoads	Integer	seconds	Time spend travelling on Strategic Road Network (Highway Agency Managed Roads: Motorways and A-Trunk Roads)
TotalStraightDistance	Integer	metres	Total Straight Line Distance from Origin to Destination

The vehicle types are classified as :

Class	Description
1	Cars
2	LCVs (up to 3500kg)
3	HGVs (over 3500kg)
4	HGVs (over 7500kg)
5	Mini-Buses and Coaches
6	Taxis
7	Motorised caravans
8	Other vehicles
9	Unknown

Format of the 2006 to 2011 Data Sample

The 2006-2011 data was not analysed as part of this exercise. Information on data definitions is however included below for completeness. From Sept 2006 to Aug 2011, the data is provided as a single comma-separated file containing one record for each start zone, end zone pair occurring in a given combination of vehicle type, period and data source as listed in the table below. The origin and destination zoning systems are defined by the PASS3 NTM zones. These have been supplied in Shape format as a set of non-overlapping polygons that together cover the whole of Great Britain.

Field	Contents	Units	Data type
1	Vehicle type	n/a	integer
2	Period	n/a	integer
3	Data source	n/a	integer
4	Start zone	n/a	integer
5	End zone	n/a	integer
6	Number of journeys	n/a	integer
7	Total journey distance	metres	integer
8	Total straight-line distance	metres	integer
9	Total journey time	minutes	integer
10	Sum of squares of journey distances	metres squared	integer

Appendix C Completeness of Trafficmaster Tests to date

As part of the preparation of this work, Ian Williams has provided a Project Note detailing stringent investigations into Trafficmaster data. There has not been opportunity within this work to apply all of these tests. We present here an edited version of those investigations, with commentary on the extent which they have yet to be answered.

Data Requirements for Quality Assessment (from PN009, Section 3.2)

Data Requirement	Current Status
a) The **method of selecting the sample** of vehicles that are sensed, including as much background information as possible about which types of LGVs are most and which are least likely to subscribe to the Trafficmaster service. Are there particular classes of vehicles with a low or zero probability of subscribing? Have the main types of subscribers changed over time? This information is used in order to identify the more likely sampling biases and to determine how these biases might be offset.	This is currently known only anecdotally from e-mail exchanges, and no firm documentation has been identified. It seems unlikely that changes over time can be tracked. Indications to date are that the dataset is biased towards newer vehicles and those owned by larger operators.
b) Whether it can provide a **linked series of all movements** of specific individual vehicles throughout an extended time period, or whether there is no linkage (as yet) between observations on different trips of the same vehicle?	This linkage exists and can be used.
c) The **spatial coverage** of the system, does it have known geographical areas of low levels of coverage or of non-coverage? For example, what happens to vehicles in tunnels or in mountainous valleys – do their trips get broken up when they pass through areas with poor GPS coverage? Are LGV owners in some parts of England less likely to be signed up to the Trafficmaster service?	Some biases do exist when comparing with the DVLA registration locations of LGVs. However, it is not clear whether these arise due to bias in the spatial coverage, or due to differences between locations of registration and operation.
d) The **temporal coverage** of the system, does it in principle always cover all time periods for a full recent year? Can it guarantee that this universal coverage has in practice been achieved everywhere for all subscribers? Can those whose subscription has lapsed in the course of a calendar year be distinguished from those who just ceased to use the vehicle due to say, holiday or mechanical problems? Failing that can it provide clear information on where and when system failures led to gaps in the provision of accurate data and clear information on which classes of vehicles will have been more or less likely to have been missed?	From the commentary received, it appears likely that some subscriptions only operate for part of the year. However, in analysing 1 month of data it has not been possible to test this.
e) **Does it guarantee to provide data for** each sampled vehicle trip of **the full journey from origin to destination?**	In the O-D dataset analysed, the distance travelled is split only between the SRN and

Data Requirement	Current Status
For each such trip does it provide the distance travelled on each individual road link or just on each class of road link (e.g. built-up and non-built-up, for motorway, dual and single carriageway, for the Strategic Road Network (SRN), A, B and C roads, for each region)?	Local Roads, though the larger GPS dataset is believed to provide more detail. In the O-D dataset the journey O-D is at LSOA level, so some spatial detail is exclud-ed.
f) What **method and rules does it adopt to build up whole trips** from the sequential individual link legs of the trip? Is there any documentation to validate this approach to trip definition and to compare the resulting trip length profile in in aggregate with that obtained from alternative national sources such as the NTS or the earlier LGV Sur-vey results? The definition of a trip may be particularly problematic for LGVs that are in use for collection / deliv-ery activities (e.g. post office vans) which my comprise a continuum of short individual movements throughout the day that might be difficult to distinguish from the patterns of movement observed within congested urban areas that have large numbers of signalised crossings and junctions.	The definition of a trip is strictly from Ignition On-Ignition Off. Hence journey stages could frequently be broken into short legs not iden-tified as trips in surveys. Some could be mis-leading (e.g. petrol re-fills and vehicle stalls), whilst others may be helpful (e.g. delivery stages).
g) Does it have a vehicle identifier that would **enable com-plementary vehicle data to be extracted from DVLA** of the type described below?	This is believed to be possible, but still prob-lematic for cost and confidentiality reasons.
h) The analysis of the potential biases in the vehicle move-ment database and the weighting methods that are avail-able to mitigate biases depend crucially on access to (suitably anonymised) DVLA data that is matched to the individual vehicles in the Trafficmaster sample. The **vehi-cle characteristics data** required for this sample include the following. Information on the registered keeper of the vehicle: i. geographical location to whatever geographical level of detail is consistent with confidentiality requirements; ii. whether a private (mr., mrs, etc.) or business (based on a name of a firm) registration? iii. number of other vehicles registered at the same ad-dress, to provide indirect information on the scale and nature of the business activity. Information on the vehicle itself , including: iv. age; v. engine size and type (petrol, diesel, …); vi. vehicle type (car, LGV, …); vii. annual mileage from odometer readings at previous MOTs for all of the vehicles old enough to have already required at least one MOT.	Not available without DVLA data.

Testing for biases in the dataset (from PN009, Section 3.3)

This Section explains three groups of tests that could be carried out to test the accuracy and representativeness of the sample of vehicles and movements in the Trafficmaster dataset. Our starting null hypothesis is that this sample is representative of the vehicle population for England. We than examine the evidence to determine whether this null hypothesis should be overturned.

Proposed Test	Current Status
We can use DVLA data to determine whether the vehicle sample is already representative of the national LGV fleet, through comparing its characteristics with those published for the English fleet as a whole. If it is not representative, then the ratios of sampled to national vehicle totals for each vehicle segment could be used to calculate weighted expansion factors that would act to increase the representativeness of the expanded sample, provided that the biases observed are not catastrophic.	A very simple version of the test carried out regional level, but inconclusive without direct vehicle link to DVLA registrations.
Having checked whether the LGV fleet in the sample is representative, the next set of checks to be carried out is to examine the representativeness of the resulting pattern of vehicle movements. We compare for each individual sampled LGV, the difference between its annual total vehicle kilometres as observed from the GPS procedure versus the annual mileage measured from the difference of its two most recent odometer readings (or where there is just one reading, use the current mileage divided by the age in years of the vehicle). There are analyses that could be carried out that would enable these odometer readings to be aged by a further part year to ensure that we are comparing like with like for each age of vehicle. Any systematic differences remaining, after the vehicles have been averaged up into large groups corresponding to each LGV type segment, would suggest either that the sensed kilometres are in error or that the odometer readings have a systematic problem. If there are not systematic differences, then this comparison would provide convincing evidence of the quality of the sensed dataset of movments.	No progress has been made on this test. Odometer readings cannot be sourced for the Trafficmaster vehicles : Trafficmaster themselves advise that the distance travelled in their dataset is a proxy, but this is what the test is proposed to determine. There is some evidence in our tests that average mileages are higher in the Trafficmaster dataset than in general, but whether this is related to the sample bias or over-reading of mileage is not known.
Having successfully expanded the trip data, through use of weighted expansion factors specific to each of the vehicle segments, it is then feasible to compare the national vehicle kilometres by road type with those published from the RTS, in order to confirm whether the dataset is spatially representative. This latter step can still be carried out through use of a simple unweighted expansion factor that is applied identically to all trips but this simplified approach may be less likely to produce an acceptable match to published totals and would provide little guidance on how to recover from any biases that had been observed.	An unweighted comparison has been carried out at this stage.

Tel: +44 77 1113 9387
Fax: +44 20 7314 5111
www.wspgroup.co.uk

UNITED
BY OUR
DIFFERENCE