Department for Transport

Longitudinal Analysis of Teletrac Navman Traffic Speed Data

- for the National Transport Model

May 2021

Department for Transport
Great Minster House
33 Horseferry Road
London
SW1P 4DR

This publication is also available on our website at www.gov.uk/government/organisations/department-for-transport

Any enquiries regarding this publication should be sent to us at www.gov.uk/government/organisations/department-for-transport

Contents

Executive summary

A longitudinal analysis of Teletrac Navman tracked vehicle data was performed in order to develop statistical models to predict urban area traffic speeds in forecast modelling years. The evidence-based models developed in the study could provide a feasible alternative to the current method of predicting speed changes on fixed speed links in urban areas in NTMv5.

This study involved the following:

- developing an approach to sampling of the Teletrac Navman database ensuring robustness of the analysis and modelling
- developing a method to measure link speed stability to ensure that link speeds are sufficiently stable over certain periods and can be used to estimate changes
- studying the impact of key factors hypothesised to contribute to urban speeds (such as road characteristics and growth-related variables)
- developing a set of statistical models that can be applied incrementally to predict speed changes on urban networks for a specified forecast year in relation to a specified base year

The final set of multivariate linear models developed (a total of six for different time periods and road types) were able to capture some of the variations in urban area speeds, given the limited availability of observed data on local variables. The final set of models use data on road classification, link length, car ownership, jobs, population density, and road density to predict speed.

In order to improve the model performance, it is recommended to use the models for predicting speed changes rather than for the direct prediction of speeds. Specifically, the method for predicting speed changes is recommended to be used for urban network slip roads, dual carriageways (excluding motorway dual carriageways), and single carriageways in combination with A roads and B roads. The use of the models on urban motorway dual carriageways and on local and minor roads single carriageways has resulted in relatively lower performance due to unexplained local variations which vary across years.

1. Introduction

Background

The Department for Transport (DfT) has developed the National Transport Model v5 (NTMv5) using VISUM, a macroscopic traffic modelling software. NTMv5 is a spatially detailed transport model covering England. The internal area of NTMv5 includes England and some cross-border zones to Wales, with most of Wales and Scotland treated as external areas.

Six use cases were identified that guided the detailed design of NTMv5 (Atkins 2019):

- strategic roads investment and resilience
- road user charging and other potential policy
- local investment and policy
- general support for DfT Teams (other than Roads/Local)
- scenario-based national traffic forecasting
- exploring the unknown (testing new policies/technical developments not modelled before)

Compared to previous versions of the NTM, the range of policies that NTMv5 can be used to test has increased, especially as a result of the increasing spatial detail in this version of the model, that is the increased representation of the national highway network.

The purpose of this study is to investigate one key limitation which arises in the network development. This is the treatment of Urban Area Speeds (UAS) in forecast years.

Forecasting urban speeds in NTMv5 makes use of the speed-flow approach on approximately 84% of the network links and a fixed speed approach on the remaining links (primarily urban network). To forecast changes in speed in the areas of the network with fixed speeds, a relationship has been derived between speeds and trip end growth. While the trip end responsive approach is straightforward and has the advantage of providing a stable representation of the speed in the central area of major cities and towns, with significantly reduced requirements in data collection and network coding, it has been developed for pragmatic purposes as an improvement to a fixed speed method but with limited understanding of its impact on speed forecasts.

This approach is not evidence-based and the estimation process does not take into account the differences in urban areas and road characteristics and the impacts of major interventions. This has limited the model's application, due to the impacts on generalised cost calculations and lacking flexibility to reflect the urban road speed changes in future year models.

UAS can be influenced by different factors (such as driving behaviour, road characteristics, traffic management schemes) and it can also influence demand levels, and consequently congestion and travel time. Little evidence is available in the literature on what these factors are and the extent of their impact on urban speeds.

DfT has engaged with Arup/AECOM consortium to work on deriving an evidence-based methodology to predict traffic's speeds on urban roads in future years. The methodology should make use of Teletrac Navman data and be applied on individual, bidirectional road links. The outcome of such a methodology will be used to inform the improvements of DfT's NTMv5 and highlight recommendations on further areas to be investigated on the analysis of traffic speeds on UK's road network.

DfT recognises that many factors could potentially affect urban traffic speeds over the years. Some of them are measurable, having long term impacts, or could be derived from published data, while some are temporary, inter-connected, or even very difficult to collect and/or quantify. Therefore, the resulting approach to predict future speeds should be pragmatic, stable and based on the best available information.

Two of the main criteria influencing the data being used in the analysis of traffic speeds are:

- current data availability
- data which can be obtained for future years, in order for the improved modelling approach to be used to forecast traffic speeds in future years

This document describes the methodology for the urban speed model development and validation. Specifically, the document provides the reader with the following:

- a method for processing and cleaning Teletrac Navman data, specifically outlier detection/removal
- a sampling approach of the Teletrac Navman database which ensures robustness of the research
- a method to measure the speed stability to ensure that link speeds are stable enough over certain periods and can be used to estimate changes
- a methodology which studies the impact of key characteristics potentially related to urban fixed speeds, and estimates urban fixed speeds and its changes at a specific confidence level
- the application of the developed method on NTMv5

This report starts by discussing the literature reviewed as well as the data sources collected and used in this study. The rest of the report is then dedicated to reviewing the exploratory analysis undertaken to understand the relationship between traffic speeds and other key variables, as well as the traffic speed predictive model development and validation and its final results.

It should be noted that all the analysis and the modelling undertaken in this study used R programming language (R Core Team 2020) and a number of R packages to facilitate the work.

Report Structure

This report is divided into five sections:

- Literature Review summarises the literature that has been reviewed as part of this study
- Data Collection outlines the data sources that have been collected and used for this study, including the data sampling methodology used for the extraction of Teletrac Navman data
- Exploratory Analysis discusses the exploratory analysis that was undertaken to understand the relationship between traffic speeds and other key variables
- Urban Speed Modelling presents and reviews the proposed methodology for forecast speed modelling and discusses how it can be applied in order to predict speed changes in future years in NTMv5
- Conclusion provides a conclusion on the outcome of the data analysis and modelling, along with key limitations of this study, and final recommendations

2. Literature Review

Introduction

This section discusses the literature reviewed to inform the analysis and modelling of urban traffic speeds, including literature on:

- the availability of data sources used in modelling traffic speeds
- the relationship between traffic speed and other key variables
- trend analysis of historical traffic speed data
- evaluation of current methods of modelling UAS
- emerging predictive or statistical methods of modelling traffic speeds

The review into the relationship between traffic speed and other key variables, such as growth variables, only found a limited amount of literature on this topic. Some studies have investigated the effect of physical measures on speed, finding that junction and link geometries are important determinant variables for speed and that on-street parking is effective at reducing speeds (Elliott et al. 2003; York et al. 2007). However, the available literature often relies on small data sets and with limited validation testing.

Several studies have been found that analyse large traffic speed data sets, typically tracked vehicle data, to understand trends over time. These studies have analysed trends in different locations and found differences in traffic speed by area type, speed limit, and across years (Atkins et al. 2018; Sharma et al. 2017; TfL 2013).

Data Sources

Information on traffic speeds can be extracted using four main data collection methods (DfT 2020):

- Moving Car Observer (MCO) surveys
- Automatic Number Plate Recognition (ANPR) cameras
- inductive loop detectors
- tracked vehicle data using Global Positioning System (GPS) devices

MCO surveys can be commissioned from traffic survey companies. While they allow for flexibility regarding survey location, they may understate average speeds where traffic exceeds speed limits and may require many observations in locations where journey times are variable.

ANPR cameras and inductive loop detectors have similar strengths and weaknesses: they have the potential to provide high sample rates and information on vehicle type but are limited to existing/possible camera locations. ANPR journey time data can be commissioned from survey companies, whereas Highways England's WebTRIS database provides data derived from inductive loop detectors on the Strategic Road Network (SRN).

Tracked vehicle data is beneficial in that it has the potential to provide a large sample size, wide geographical coverage, and capture day-to-day variability in travel patterns. However, its limitations typically include cost of the data, exclusion of short trips, and not providing fine spatial resolution and detailed segmentation. Teletrac Navman data can be accessed through DfT, whereas other sources can be purchased commercially.

Current Methods of Modelling UAS

Regional Traffic Models

The Regional Traffic Models (RTMs) make use of Teletrac Navman journey time data for their base year speeds. For forecasting future speeds, different approaches are used depending on road type and RTM focus:

- speed-flow curves are used on the SRN in all RTMs
- in the RTMs where the focus is on the SRN, future speeds in urban areas are modelled using the fixed-speed approach
- in the RTMs where urban areas are within immediate reach of the SRN, such as in the North RTM, the speed-flow curve approach is extended to cover urban areas (meaning full simulation coding of urban areas)

The speed-flow curve approach involves speeds responding to changes in traffic flow on links, hence capturing all demand responses at high level of spatial detail. To forecast fixed speeds, on the other hand, speed adjustment factors are calculated and applied to the base year speeds. These adjustment factors are calculated using the Road Traffic Forecasts available on the GOV.UK website, which have been derived from NTM and include speed estimates for England up until 2050 (Atkins and CH2M 2018; WSP 2018a). In the fixed-speed approach, speeds do not respond to any demand or network changes.

The RTMs have been used to investigate improvements to the forecasting of urban fixed speeds. Two studies developed adjustments to fixed speed coding that reflect delays as a result of flow changes in the forecasting years (WSP 2018a; WSP 2018b). Speed decay functions were derived to provide an appropriate static speed change in future years. These functions were developed for different area types rather than by individual links, as a result of lack of information on volume to capacity measures. The function uses a 'donor' model for the parameters to be estimated.

The two studies found that the use of fixed speed with decay functions generally produces plausible routeing and travel time outcomes, except for short links where even slight changes in delays can result in significant changes in flows (WSP 2018a; WSP 2018b). In addition, the studies found that assignment run times increase with the speed decay functions.

NTMv5

NTMv5 makes use of the RTMs for its base year speeds, which are derived from Teletrac Navman data. To forecast UAS, NTMv5 makes use of the speed-flow approach on approximately 84% of the network links and a fixed speed approach which is responsive to trip end growth on the remaining links (Atkins 2019). As such, the approach uses trip end growth, time of day speed adjustment base year factors, and base year free flow speeds to forecast speeds on the network. Equation 1 provides the formula used in calculating forecast speed for a specific time of day (v_{tod}^{F}):

$$v_{tod}^{F} = \frac{v_0}{1 + (\frac{v_0}{v_{tod}^{B}} - 1) \times (\frac{T^F}{T^B})^n}$$

Equation 1. NTMv5 Forecast Speed Formula (Atkins 2019)

Where:

- B is the base year
- F is the forecast year
- T is the total home-based trip ends (productions and balanced attraction weights/constraints) aggregated by area
- n is the exponent of trip end growth, assumed to be 1.5
- v_0 is the free flow speed using off-peak time period in the base year, derived from Teletrac Navman data
- v_{tod}^{B} is the base year fixed speed related to each modelled time period

This approach is straightforward and has the advantage of providing a stable representation of the speed in the central area of major cities and towns with significantly reduced requirements in data collection and network coding. The main challenge is identifying the links/areas where trip end growth and UAS have high correlation. The approach has been developed for pragmatic purposes, as an improvement to the fixed speed method but with an understanding of its limitation in speed forecasts.

The following assumptions are currently being applied to identify fixed speed links across the network within England (Atkins 2019):

- if the start node and end node constituting a link were coded as a type 4 dummy node in the RTM simulation networks, then these links were identified as fixed speed
- those links which are part of turns coded with infinite saturation capacity were also considered to be fixed speed links

It should be noted that the categorisation of urban area links in NTMv5 is derived from an earlier version of the RTMs.

Evaluation of Current Methods

Unrealistic routeing as a result of fixed speed coding has been confirmed in different studies using the RTMs (Atkins 2018; Atkins and CH2M 2018). One study investigated the impact of fixed speed on model results and economic appraisal by replacing the fixed speed coding in one urban area of the Trans-Pennine South RTM with speed-flow curves (Atkins 2018). The study found that the full simulation scenario had higher levels of delay in the urban area, saw changes in routeing to avoid the congested urban area, and had a higher Present Value of Benefits than the fixed speed scenario, leading to a higher Benefit-Cost Ratio (BCR).

A similar study replaced the full simulation coding in one urban area of the North RTM with fixed speed coding (Atkins and CH2M 2018). This study also found that the full simulation scenario better reflects increased delay in the urban area. However, both studies found that the more detailed representation in the full simulation scenarios came at the expense of increased assignment run time.

Emerging Methods

Several studies have been conducted that model traffic speeds using emerging predictive or statistical models. These models range in complexity, from regression analysis (Hooper et al. 2014) to machine learning methods (Huang and Ran 2003).

York et al. (2017) used regression analysis to better understand the impact of road characteristics (specifically road width, forward visibility, parking, and surface type) on traffic speed. The study developed two multi-linear regression models to model the logarithm of traffic speed: using manual count data and automatic count data, respectively. The models explained only between 20% and 22% of the total variation in the data, noting that all the variables included were significantly different from zero at 95% confidence level. The key disadvantage of this research is its time- and cost-consuming data collection, as it conducted site surveys to obtain detailed data on its independent variables. Hence the study is also limited in terms of validation.

Whereas several studies have been found on developing models for real-time short-term (such as 10 seconds ahead) prediction of traffic speeds (such as Vanajakshi and Rilett 2004; Wang et al. 2011; Ye et al. 2012; Chen et al. 2014), limited research is available on the use of predictive models for long-term forecasting. Additionally, these studies often focus on a particular area of interest with limited data points and, as seen in the York et al. (2017) study, validation testing is often limited.

3. Data Collection

Introduction

This section discusses the data sources used, the sampling methodology used to extract Teletrac Navman data, and the data fusion undertaken to conduct the explanatory analysis and traffic speed modelling.

In this study, the main data source is Teletrac Navman journey time data, which was collected from DfT for England for the years of 2013 to 2019. The data provides average journey times at 15-minute intervals. The following secondary data sources were collected:

- Ordnance Survey (OS) Integrated Transport Network (ITN) layer of highway links, including bidirectional information on road classification (A road, B road), road type (single/dual carriageway), and link length, collected from DfT for England and Wales for the years of 2013 to 2019
- ONS 2011 urban-rural classification of Middle Layer Super Output Areas (MSOA) for England and Wales, alongside map layers of MSOA and government regions, collected from the ONS Open Geography Portal
- National Trip End Model (NTEM) data including yearly total jobs, housing, population, population owning one or more cars, and average weekday home-based (HB) and non-home-based (NHB) productions and attractions by time period, extracted from CTripEnd for Great Britain for the years of 2011, 2016, and 2021

Data Sampling

Due to the size of the Teletrac Navman data, a two-step stratified sampling methodology was developed for the extraction of this data:

- step 1 produces a stratified random sample of 30 Local Highways Authorities (LHA) in England, weighted by the proportion of LHAs in government region and population categories
- step 2 produces a stratified random sample of ITN links in the sampled LHAs, weighted by the proportion of links in categories of population, housing, jobs, car ownership, road type, road classification, and link length

The sampling approach was based on 2018 ITN links, meaning that the total number of links will vary slightly by year once the sampled 2018 links are merged with other years.

This is as a result of changes in the road network between different years. Because of the large sample of links, the models will be able to explain the variation in speeds as a result of the variation in characteristics already present in the sample.

The first sampling step was conducted to ensure that the distribution of NTEM population in the sampled LHAs are representative of the population distribution in each government region across England. The second step ensured that the sampled links are representative of the distribution of the key variables used in this study, mainly: road type, road classification, population, car ownership, jobs, and link length.

Prior to conducting the second sampling step, all the MSOAs and links which are not the focus of this study were removed. This included removing rural areas and certain road types and road classification. Rural MSOAs were excluded since analysis of the Teletrac Navman data demonstrated that rural links have consistently higher average speeds in all years. In order to identify urban areas, the 2011 ONS rural-urban classification was used (ONS 2018). Assignments of MSOA to urban or rural categories are made by reference to the category to which most of their constituent output areas are assigned. Output areas are treated as ‘urban’ if they were allocated to a 2011 built-up area with a population of 10,000 or more. Urban and rural areas are then further sub-divided into broad morphological types based on the predominant settlement component, specifically relating to typical plot densities and neighbourhood density areal extents (Bibby and Brindley 2013). The urban area types included in this study were urban major conurbation, minor conurbation, and city and town.

As for road type and classification, the study focussed on those which were most likely to be modelled in highway models in general, and specifically in NTMv5. As such, the sample was limited to the road types of dual carriageway, single carriageway, and slip road; and the road classifications of motorway, A road, B road, minor road, and local street.

Based on the sample of LHAs, three independent sets of urban ITN links were randomly sampled for which Teletrac Navman data was extracted:

- sample 1 of ITN links for all government regions except for London was used for the model development, with Teletrac Navman data extracted for all years (2013 to 2019)
- sample 2 of ITN links for all government regions except London was used for model validation, with Teletrac Navman data extracted for 2018 and 2019
- sample 3 of ITN links for London was used for the model validation, with Teletrac Navman data extracted for 2018 and 2019

A total of 1,815 urban MSOAs were sampled for the extraction of ITN links for sample 1 and 2. These are visualised in Figure 1. Approximately 200,000 links were extracted for each sample. Sample 3 only covers London; these urban MSOAs are not included in the figure.

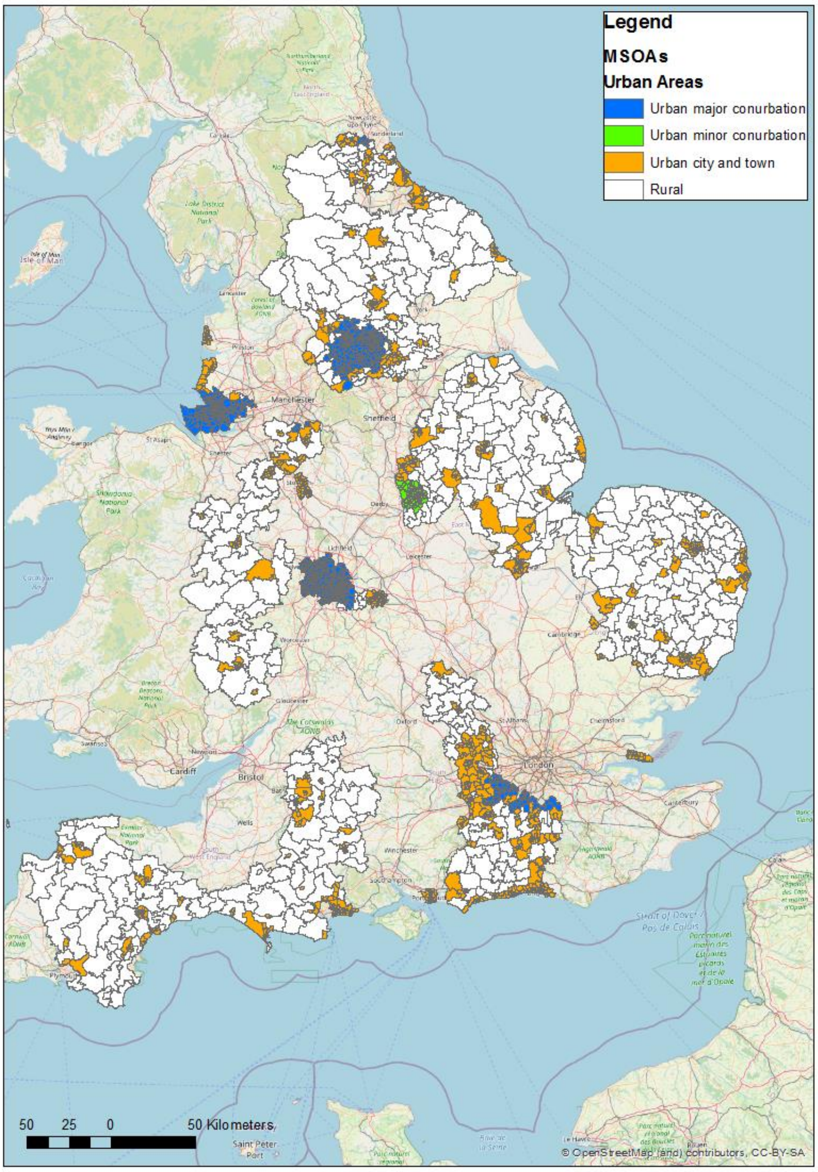

Figure 1. Sampled Urban MSOAs for the Extraction of ITN Links for Sample 1 and 2 (Sample 3 covers London only)

Throughout the sampling process, several verification checks were undertaken to ensure that the distributions of the sampled links were comparable to those of the original population (the entire data set).

Table 1 summarises the Teletrac Navman data extracted for sample 1.

Year	Number of Links	Number of Observations
2013	249,879	340,907,340
2014	255,047	461,165,461
2015	263,604	556,819,074
2016	262,570	653,436,304

2017	241,950	558,539,460
2018	194,517	344,725,495
2019	193,467	345,209,227

Table 1. Teletrac Navman Data Sampled for Model Development

Data Fusion

The next step was to fuse the data together, including the sampled links, Teletrac Navman data for the sampled links, and NTEM data.

The processing of the NTEM data involved using linear interpolation to obtain the population, car ownership, housing, and employment data by year. CTripEnd was run to get trip end data for each of the four time periods and purpose by year (2013-2019), after which they were aggregated into HB and NHB purposes. Throughout the processing of the NTEM data, the spatial and temporal resolution was maintained as MSOA level and period level, respectively.

The Teletrac Navman data underwent the following processing steps, resulting in data sets of mean speeds for each sampled link at 15-minute interval for all years:

- vehicle classes were limited to cars, LGVs, and HGVs
- time periods were obtained based on the definition of 07-10 for AM, 10-16 for IP, 16-19 for PM, and 19-07 for OP
- weekends and bank holidays were removed
- sampled links were selected
- weighted journey times in seconds were calculated using the number of observations recorded for each individual link in the case where multiple rows of data were recorded for the same link, assuming that the number of observations is a representative sample of traffic flows on the network
- average speeds were calculated in metres per seconds and converted to kilometres per hours

Speed distributions were analysed for the existence of outliers, if extreme values for speeds exist in the data such as 500 km per hour which is not achievable. Average and maximum speeds were studied at different levels of aggregation, finding that, across all years, when outliers are not removed, maximum speeds can reach unachievable levels for small percentage of data points. To mitigate the impact of outliers, a two-step outlier removal process was undertaken:

- step 1: 15 minutes data outside 3 standard deviation of average speeds at hourly and link level were removed, leaving 99.7% of the data
- step 2: 15 minutes data outside 3 standard deviation of average speeds at yearly, road type, and road classification level were removed, leaving 99.7% of the remaining data

The two-step process removes a small percentage of the tails of the data. This method assumes normality in average speeds which, considering the large amount of data available in this study, is plausible.

The method used for aggregating average speeds spatially and temporally involved using the number of observations recorded in the Teletrac Navman data, which assumes representativeness of the sample for the traffic on the link. For the exploratory analysis, the data was aggregated at different levels temporally (such as time period and month) and categorically (such as road type, road classification, and urban area type). For the modelling, link-based data was maintained but was aggregated temporally to yearly and time period level, as evidenced by the exploratory analysis. Yearly, monthly, and period aggregation level was also generated to study the modelling approach on monthly data.

Once the modelling data set was aggregated to yearly and time period link level, NTEM data was merged based on the year, time period, and MSOA of each link. Several variables were then derived and studied in terms of their potential to impact traffic speed. The main variables are:

- road density (km per individual) at yearly MSOA level: total kilometres of road network in an MSOA divided by the total population in an MSOA
- population density (individuals per m^2) at yearly MSOA level: total population in an MSOA divided by the total area (m^2) for an MSOA
- HB and NHB trip production rates based on HB and NHB trip productions and population in an MSOA at time period and yearly level: trip productions over population
- HB and NHB trip attractions per job based on HB and NHB trip attractions and employment data in an MSOA at time period and yearly level: trip attractions over jobs
- HB and NHB trip production and attraction growth at time period and yearly MSOA level: trip production/attraction for a specific year divided by the trip production/attraction of NTMv5 base year 2015
- HB and NHB trip production rate growth at time period and yearly MSOA level: trip production rate for a specific year divided by the trip production rate of base year 2015

4. Exploratory Analysis

Introduction

This section presents the conclusions from the exploratory analysis that was undertaken to better understand the relationship between traffic speeds and other key variables. This analysis serves to provide insights into how the variables will be used in the model estimation.

The following exploratory analyses were performed:

- temporal distributions of average speeds for each year
- yearly and temporal distributions of average speeds by urban area type, road type, and road classification
- correlation analysis using the coefficient of determination (R^2) to detect multicollinearity between the continuous variables and to inform their correlation with average speeds
- Analysis of Variance (ANOVA) testing for categorical variables to understand which are most important in explaining variation in average speeds

The exploratory analysis was conducted on sample 1. As discussed, the study focusses on the area types of urban city and town, urban major conurbation, and urban minor conurbation; the road types of dual carriageway, single carriageway, and slip road; and the road classifications of motorway, A road, B road, minor road, and local street. Outliers have also been removed.

Figure 2 provides an explanatory legend for how to read the graphs.

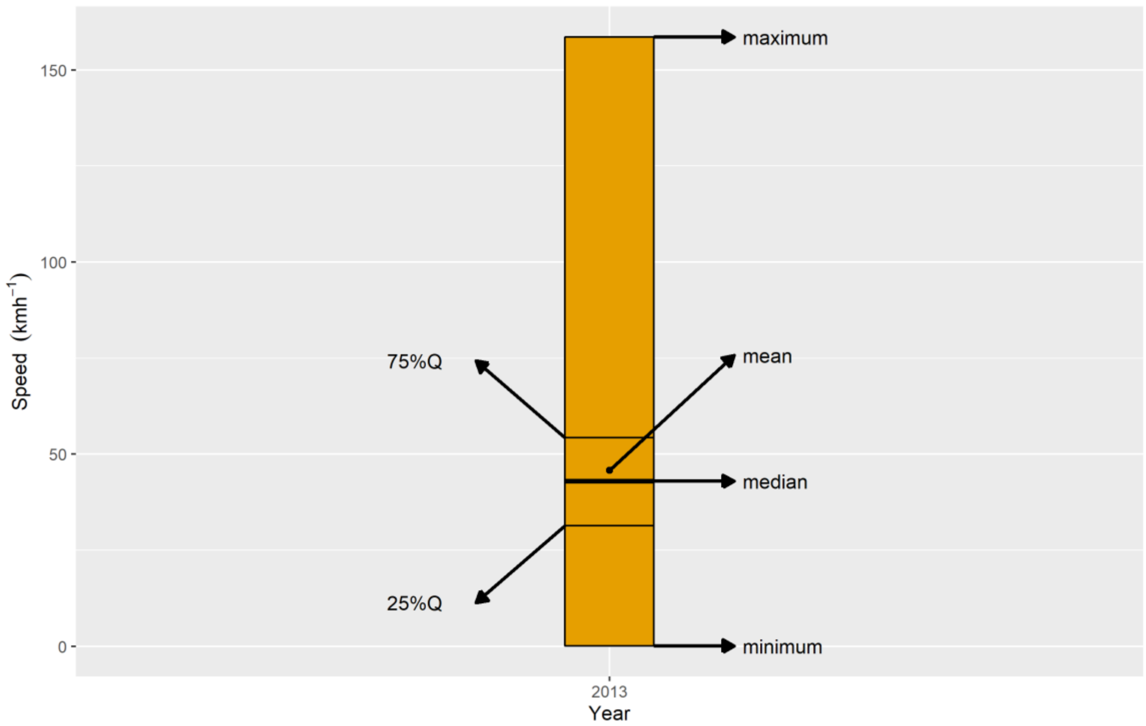

Figure 2. Exploratory Analysis Graph Legend

Results

The analysis concluded that the minimum level at which the traffic speed modelling should be made is at link level, yearly, and by time period.

While the yearly analysis does not show a specific trend in average speeds, there is a level of variation between years for different road types and road classification. This is demonstrated in Figure 3. The year variable is considered in the traffic speed modelling due to its relevance in highway forecast modelling generally.

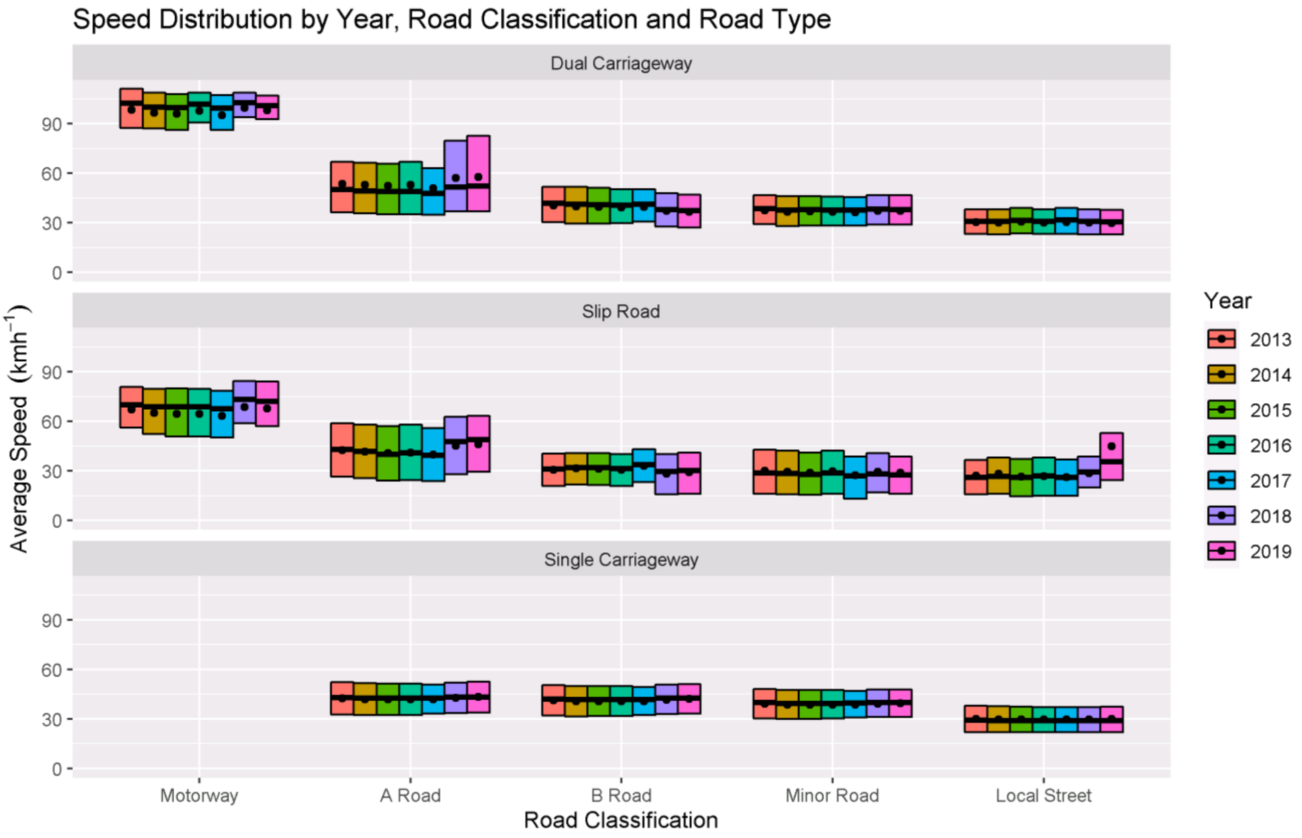

Figure 3. Speed Distribution by Year, Road Classification and Road Type

Systematic variation in average speeds was found between urban area types, specifically between the category of urban minor conurbation and the other two categories, with urban minor conurbation tending to have lower speeds. As a result, urban area type was taken forward as both a three-level and two-level variable (urban minor conurbation and urban city/town/major conurbation) in the model development.

Figure 4 shows that while there is no systematic variation between AM, IP, and PM periods, there is a clear trend of higher average speeds in the OP period. The lack of variation in the speeds between AM, PM, and IP periods could be explained by the level of activity normally observed on urban roads during the IP (such as commercial activities and trucks/goods vehicles making deliveries). A closer look into the variation in average speeds by hour demonstrated that hourly variation is only significant in the OP period.

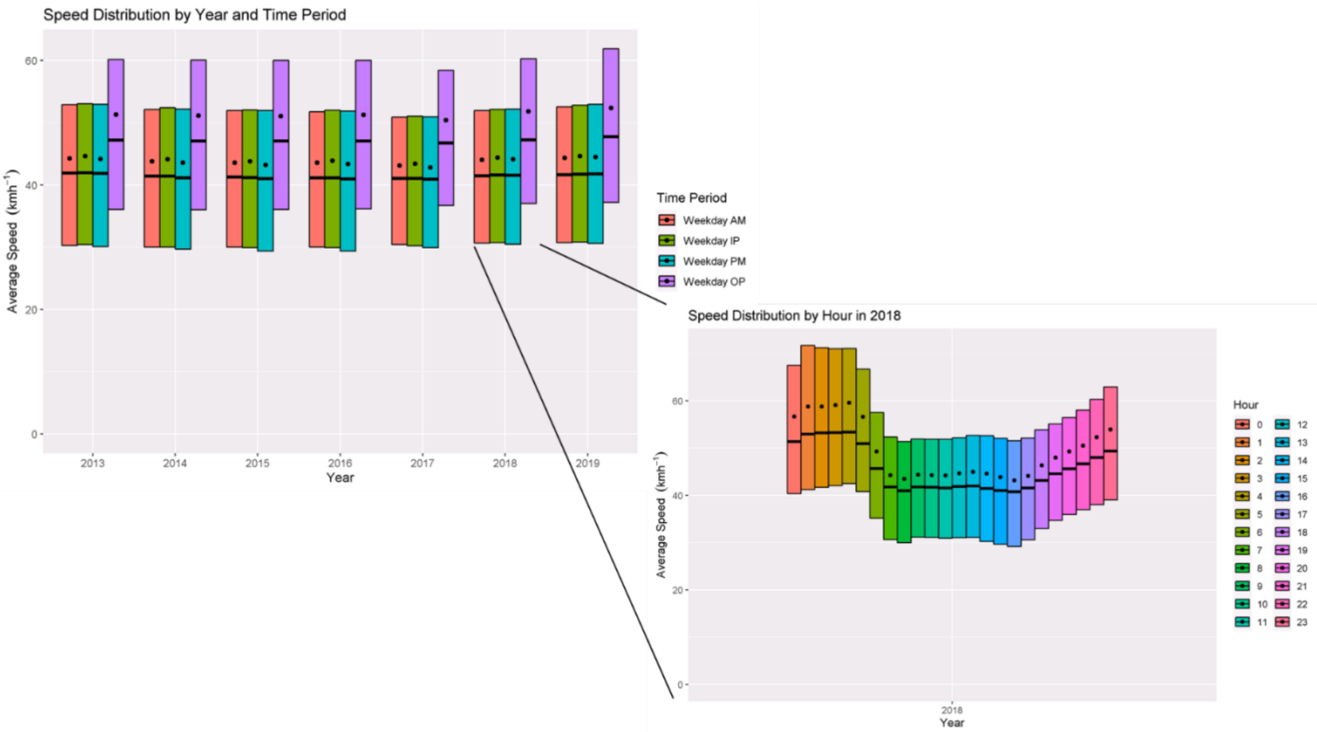

Figure 4. Speed Distribution by Year and Time Period with Zoomed in Speed Distribution by Hour of the Day for Year 2018

The analysis investigated variation in speeds by day of week, road classification, and road type but no systematic variation was found between Mondays to Fridays. Analysis into the variation in speeds by day of week and time period came to the same conclusion across all time periods.

While the analysis did not find any systematic monthly variation in average speeds in the years of 2013 to 2017, some anomalies were detected in 2018 and 2019. In both years, certain combinations of road type and road classification showed variation in average speeds by month, caused by a higher number of sampled Teletrac Navman observations in the relevant months. In 2018, anomalies were found in the month of September for several combinations, whereas 2019 included anomalies in the months of April to December for the combination of local streets slip roads. The urban speed modelling was conducted both with and without the identified anomalies, and while no changes in modelling results were detected between the two scenarios, the anomalies were removed in the final modelling exercise.

Correlation analysis was conducted on 2018 data only and found that link length, road density, and population density were the variables with the highest correlation with average speed. However, even the highest correlations are relatively weak, as seen in Figure 5. The crossed-out cells in the correlation matrix indicate that there is no significant coefficient for the two variables in question.

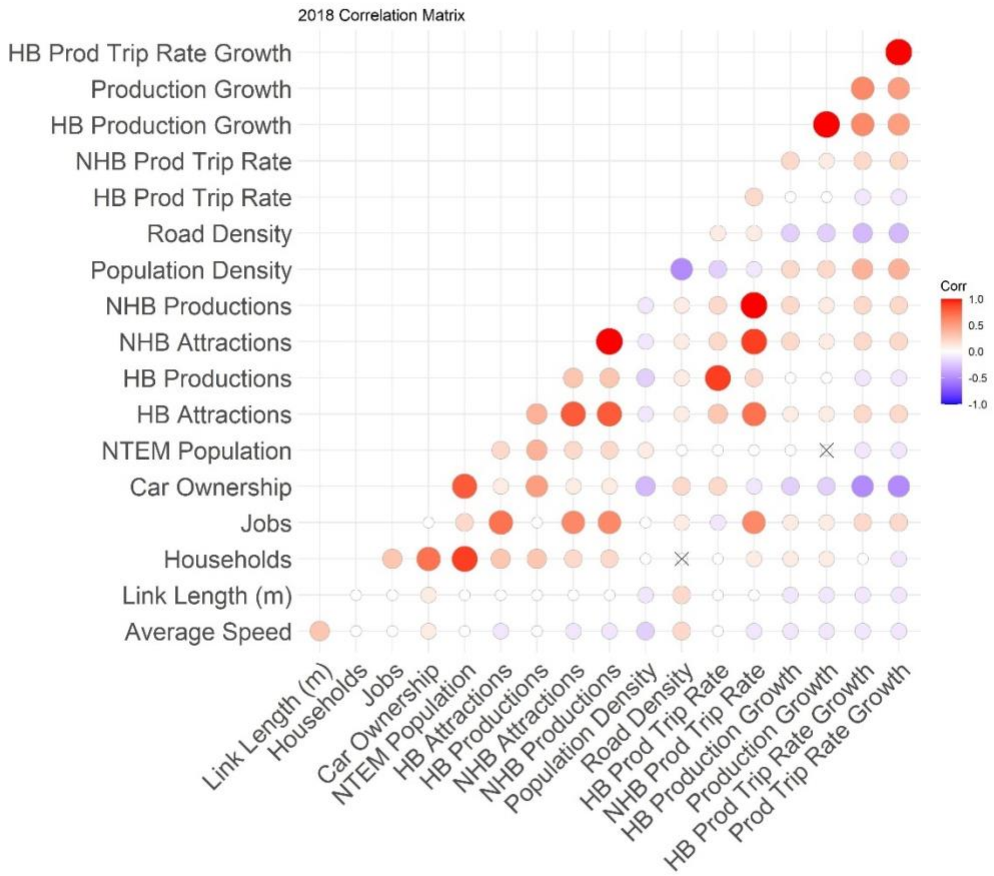

Figure 5. Correlation Matrix of Key Continuous Variables

ANOVA testing was conducted on the variables of road type, road classification, urban area type, time period, and year, with the following statistics extracted: generalised effect size (GES), degrees of freedom, and mean sum of squares. The highest explanatory power for the variation in traffic speed was obtained from road classification and road type. It was also found that time period with four separate levels (AM, IP, PM, OP) have almost the same effect size as time period with OP and AM/IP/PM combined. Therefore, the two-level time period variable was taken forward into the model development.

The exploratory analysis provided insights into the model development. Observations derived from the correlation analysis and ANOVA testing complement the modelling work presented in the following section.

5. Urban Speed Modelling

Introduction

This section discusses and presents the results of the proposed methodology for urban traffic speed modelling and model application.

Models were tested on two data sets: yearly and time period level, as well as yearly, monthly, and time period level. Using the yearly and time period data set is more appropriate in terms of its application to the NTMv5 model and to highway models in general, as they are normally developed for the AM, IP, and PM time periods for the base year and forecast years.

The study investigated link speed stability and how to treat unstable links. Coefficient of Variation (COV) was used as a measure of variability in speed, specifically its variation in comparison to its mean value. COV is calculated by dividing the standard deviation of a variable over the mean. A high value of COV reflects high variation in speeds observed for a specific link in a specific year and time period, whereas a low COV reflects lower variation in speeds. As such, the higher the COV, the less stable the link. The calculated COV were used as weights in the modelling framework in order to avoid any assumption on the definition of a stable link and cut-off point of COV. This means that links with high COV contribute less in the model development, while those with low COV contribute more.

The following model performance statistics were used:

- R^2 (coefficient of determination) which measures the percentage of variation explained by the developed model; the higher the value, the better the developed model
- RSS (residual sum of squares) and RMSE (square root of the mean of sum of squares of errors) where the lower the RSS and RMSE values, the better the developed model

Model Development Method

In this study, the following model development approach was undertaken:

- iterative process of model development with backward selection of variables/features and production of model verification statistics/plots

- best model selection through a four-stage iterative process: stage 1: best performing model in terms of model performance statistics with minimum number of variables selected and testing for possible log transformation of continuous variables; stage 2: addition of interaction terms based on explanatory analysis and selection of best performing model; stage 3: revisiting the data by using the monthly data set and updating the outlier detection method; stage 4: sub-setting the data into different categories and developing a model for each subset of data
- model validation on two data samples: sample 2 (same area as the calibration data set but different sample) and sample 3 (London area not included in the calibration data)

A backward selection of model variables means that all variables are initially used, with model verification statistics recorded, followed by removing one variable at a time to check whether the verification statistics change. If the verification statistics improve or do not change, the variable is removed as it is not adding additional explanatory and predictive power, and the process continues. If the verification statistics worsen, the variable is kept, and the process continues with another variable tested.

The model form that was mainly tested was the generalised linear model (GLM), of which linear regression was mainly used. Linear regression models the expected value of the continuous variable, average speeds Y, as a linear function of predictors, X, taking the form shown in Equation 2.

$$\hat{Y}_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi} + \epsilon_i$$

Equation 2. Generalised Linear Regression Model Form

Where $i = 1, \ldots, n$ is the number of observations and p is the number of predictors.

Its main model assumptions are that Y and errors ϵ are normally distributed and errors are independent with a constant variance. The unknown model parameters, β, are typically estimated using the ordinary least squares method, that is by the minimising the sum of squares function shown in Equation 3.

$$S = \sum_{i=1}^{n} w_i \left(Y_i - \hat{Y}_i\right)^2$$

Equation 3. Weighted Least Square Minimisation of Residual Sum of Squares

Where Y_i is the observed value $\hat{Y}_i$ is the modelled value, and w_i are the weights for each observation. In the unweighted scenario $w_i = 1$.

Model Development Results

The performance statistics for the best selected models from stage 1 to 3 are presented in Table 2. Around 30 models were tested in stage 1, with the best performing model including the variables of road type, road classification, link length, time period, road density, population density, car ownership, and log of jobs.

In stage 2, several interaction terms were studied, with the updated best model including interaction between road type, road classification, and link length. While the model is explaining 45% of the variation in speeds at an aggregate level, the model performs poorly at a detailed level, such as by road type and road classification. As a result, the following stages checked if the modelling performance could be improved at a more detailed level.

Stage 3 involved testing two different scenarios: using the monthly data set and revisiting the outlier detection method. Only the latter improved model performance. Specifically, any dual and single carriageway data with higher speeds than the cut-off points of 70mph plus 10% and 60mph plus 10%, respectively, were removed. Additionally, two standard deviation away from the mean, instead of three, was used to remove outliers.

Model	R^2	RMSE
Stage 1 Selected Model	0.441	15.451
Stage 2 Selected Model	0.452	15.313
Stage 3 Selected Model	0.460	15.016

Table 2. Summary of Model Performance for the Best Models – Stage 1 to 3

Stage 4 developed separate models for different subsets of data. Three scenarios were tested:

- the first scenario tested four sub-models by time period (AM, IP, PM, and OP periods separately)
- the second scenario tested two sub-models by time period (AM/IP/PM periods combined, and OP period on its own)
- the third scenario tested six sub-models by road type and time period, as outlined in Table 3

Developing separate models by time period and road classification did not improve model performance. As a result, the final set of models are split by road type and two-level time periods (AM/IP/PM periods combined, and OP period on its own). The best selected model form is shown in Equation 4. The performance statistics for the final models are presented in Table 3.

$$\hat{v} = \beta_0 + (\beta_{1k} \times road\ class_k) * (\beta_2 \times link\ length) + \beta_{3l} \times time\ period_l + \beta_4 \times road\ density + \beta_5 \times population\ density + \beta_6 \times car\ ownership + \beta_7 \times log(jobs)$$

Equation 4. Best Selected Model Form – Stage 4

Where $\hat{v}$ is the mean link speed, k is the number of road classification factor levels minus one, l is the number of time period factor levels minus one, β_1's are the parameters of each of these variables, and $*$ is used to represent the interaction term between road classification and link length.

Number	Sub-Model	R2	RMSE

1	Single carriageway – OP	0.461	14.806
2	Single carriageway – AM/IP/PM	0.403	14.324
3	Dual carriageway – OP	0.536	25.192
4	Dual carriageway – AM/IP/PM	0.518	26.632
5	Slip road – OP	0.590	22.434
6	Slip road – AM/IP/PM	0.569	22.535

Table 3. Summary of Model Performance for the Final Selected Model – Stage 4

The normality of residuals was investigated for the final models, finding that the residuals are normally distributed with mean value close to zero, as required. Residuals is the difference between observed and predicted average speeds.

The parameter values of the variables used in the six sub-models are shown in Table 4; the model numbers are described in Table 3. For each of the categorical variables, one category is considered as a reference and thus the parameter needs to be set to 0. The reference is the level to which all other levels in the variable are compared.

Generally, the parameters have sensible signs as expected. For instance, with reference to A roads (for single carriageway models), the B road, minor road and local street coefficients are negative. This means that there is a reduction in speeds as a result of a change from A roads to other road classifications. Similarly, link length has a positive sign throughout the six models reflecting what was observed in the correlation analysis; as link length increases, average link speeds increase.

Another example is population density which has a negative sign for single and dual carriageways. This indicates that as population per m^2 increases, these road types are expected to have a reduction in average link speeds. However, this is not the case for slip roads where an increase in population per m^2 is expected to have an increase in average link speeds. Population density for slip road models are one of the least contributing variables. The reason behind the positive coefficients can be a correction for the unexplained variation in average links speeds.

Car ownership is positively correlated with average link speeds with positive coefficients obtained. Car ownership is typically higher the further the locations are from city centres and city towns. As a result, car ownership is generally higher in less congested areas and as such it is positively correlated with average speeds.

Parameter	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	53.5103	48.8940	111.9823	111.6378	68.3547	68.0236
Road classification – motorway	0	0	0	0	0	0

Road classification – A road	0	0	-45.6638	-47.0781	-29.0690	-28.4255
Road classification – B road	-1.8248	-0.7406	-51.9644	-52.0467	-28.9717	-27.7830
Road classification – minor road	-5.5913	-2.8222	-55.0210	-53.8719	-35.3481	-34.2203
Road classification – local street	-18.3804	-14.4145	-61.6903	-59.1796	-32.0652	-30.5761
Road classification and link length interaction – motorway	0	0	0	0	0	0
Road classification and link length interaction – A road	0	0	0.0345	0.0340	0.0684	0.0625
Road classification and link length interaction – B road	-0.0033	-0.0036	0.0418	0.0400	0.0273	0.0205
Road classification and link length interaction – minor road	-0.0152	-0.0185	0.0332	0.0302	0.0853	0.0834
Road classification and link length interaction – local street	-0.0137	-0.0155	0.0159	0.0138	0.0101	0.0144
Link length (metres)	0.0300	0.0331	0.0015	0.0021	0.0266	0.0246
Time period – AM	0	0	0	0	0	0
Time period – IP	0	-0.2827	0	0.3804	0	0.4882
Time period – PM	0	0.5499	0	0.1372	0	-0.0987
Road density (km per individual)	0.0487	0.0571	0.5873	0.6560	0.4223	0.4842

Population density (individuals per m^2)	-358.9540	-332.7980	-820.4940	-907.6540	334.2634	375.4563
Car ownership (individuals with >1 cars)	0.0002	0.0002	0.0011	0.0011	0.0003	0.0003
Log(jobs)	-1.1852	-1.3340	-3.3823	-3.8392	-2.2085	-2.6280

Table 4. Coefficient Values for the Final Sub-Models

Figure 6 shows the observed versus predicted average link speeds based on the calibration data set. The comparison is made at the road type and time period level, which is the level at which the model was calibrated. R2 values range between 0.37 and 0.58 which is an improvement to the previous models at this disaggregation level. The figure's blue colour scheme and legend highlight the number of data points (density) of the calibration data set. Producing the same figure on the validation data sets – sample 2 and sample 3 – shows similar patterns to the calibration data.

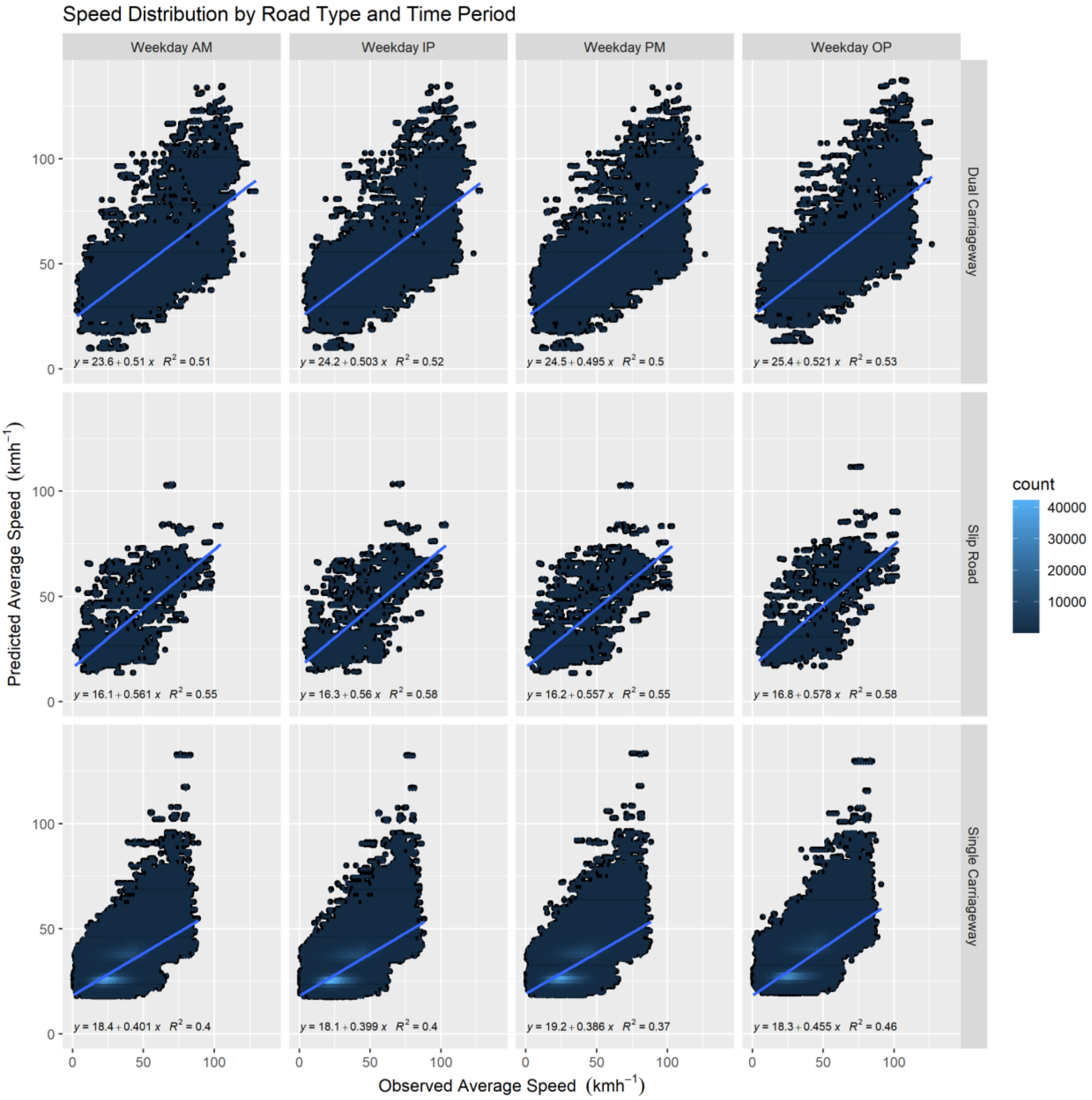

Figure 6. Observed versus Predicted Average Link Speed – Stage 4

Figure 7 shows the distribution of observed versus predicted average link speeds by road type and classification. This figure is a density plot, where density is the proportion of data points for each speed in the data set. At a more disaggregate level to which the model was calibrated (that is by road type and classification), the model is still lacking in its ability to capture the variability in average speeds. However, the models capture the peaks of the distributions, especially for single and dual carriageways.

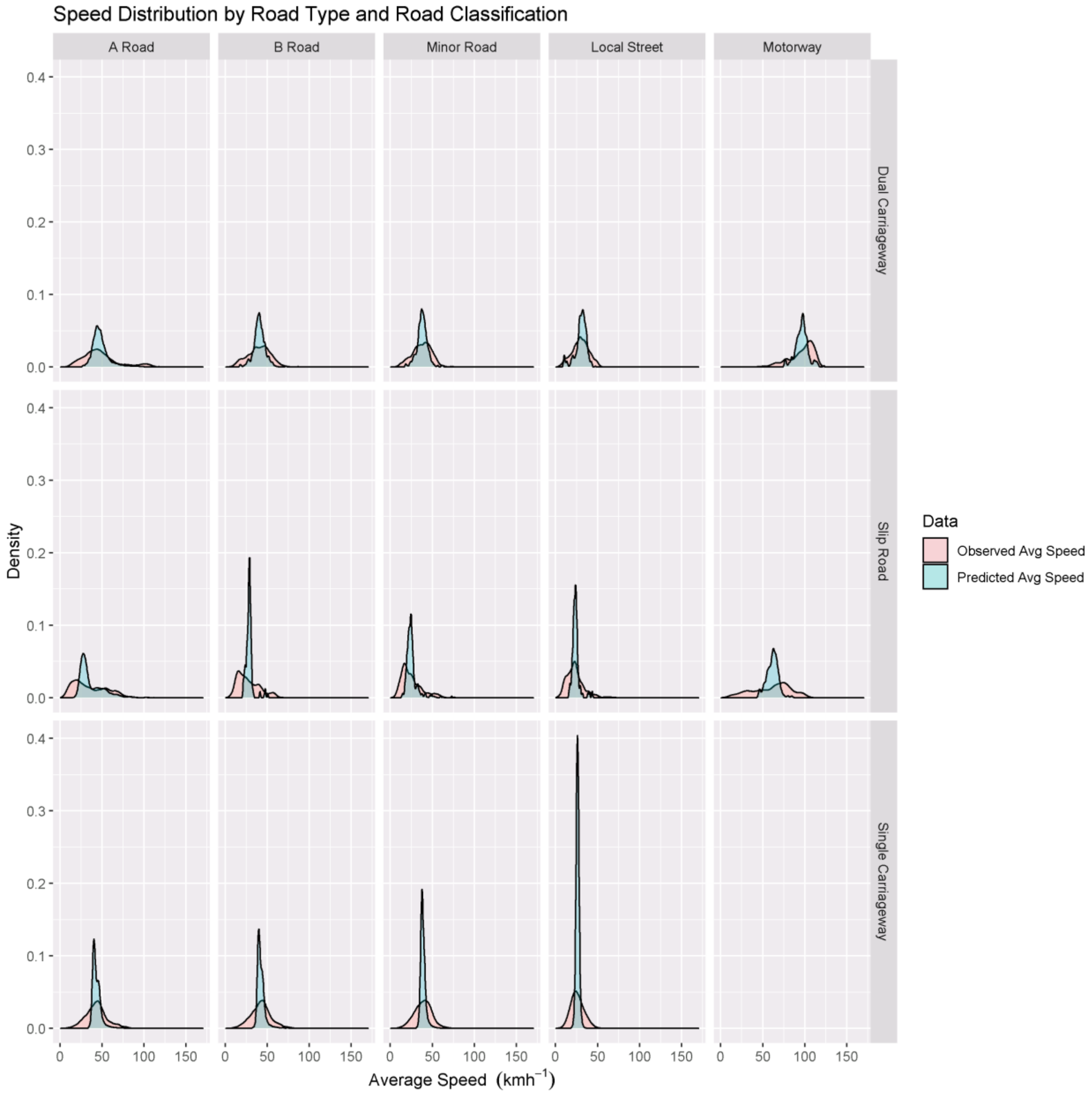

Figure 7. Observed versus Predicted Average Link Speed Distribution – Stage 4

The analysis presented in Figure 8 shows that for both validation data sets, the models are also generally able to capture the peaks of the distributions in average link speeds disaggregated by both road type and road classification. This indicates that while the model performs relatively poorly in predicting the variation in speeds at the local level, it is able to predict the mean value of average speeds for each combination of road type and road classification.

Figure 8 also shows that while speed data for London was not used in the calibration, the validation performance is similar to the areas used in the calibration. This is except for the case of motorways where generally London data performs poorer than other areas. However, it should be noted that there are relatively few motorway links in London and that drawing conclusions on the validation fit of the motorway category should be done with caution.

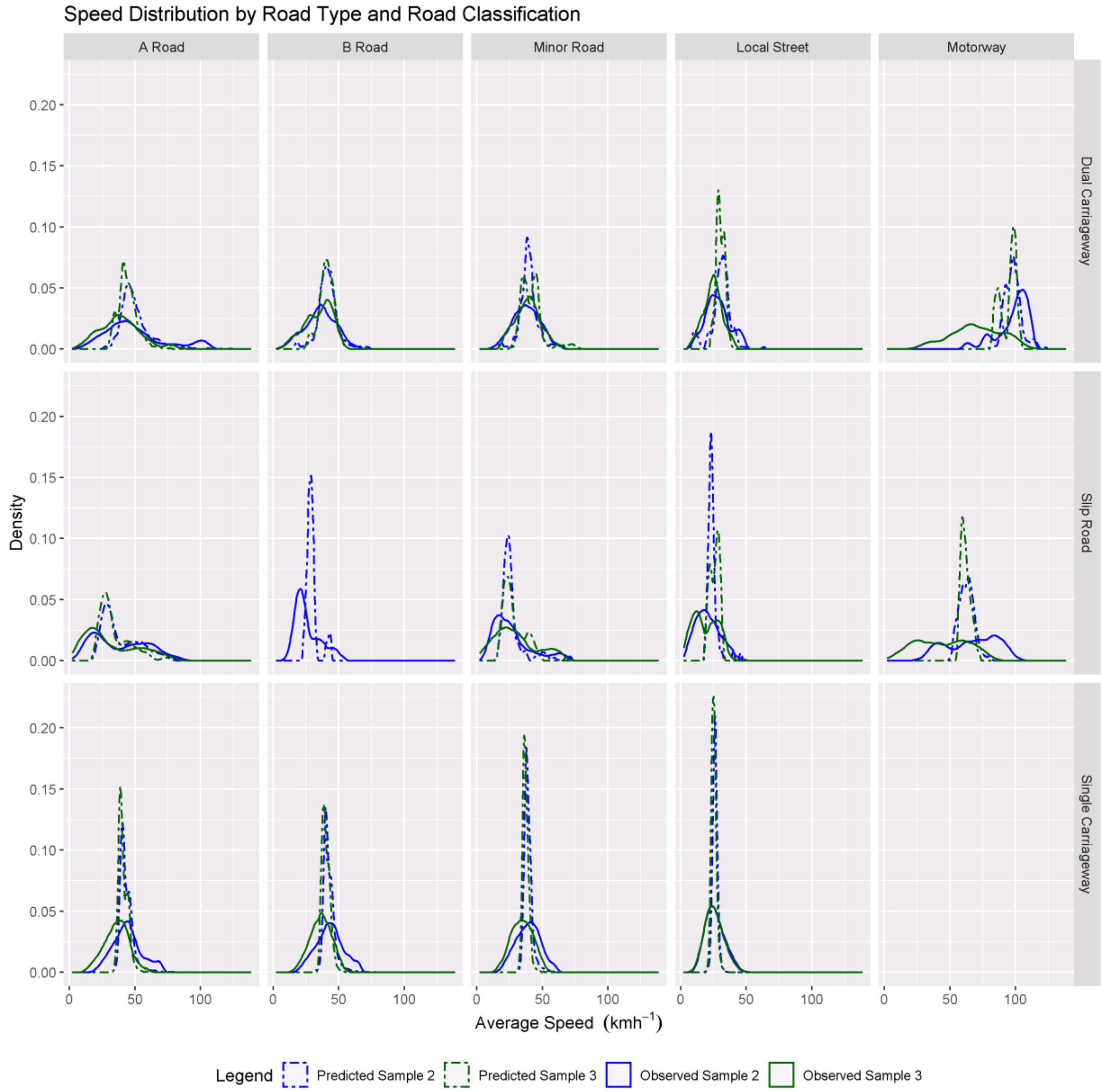

Figure 8. Observed versus Predicted Average Link Speed Distribution – Stage 4 – Validation Samples 2 and 3

The poor performance of the model at a detailed level is mainly a result of the lack of explanatory variables which are at link level. The main variables which are adding the most to the model are road type, road classification, and link length, all three of which are collected at a link level. However, the rest of the NTEM variables are all collected at an MSOA level or MSOA and time period level by year.

Final Model Application

Generally, there are four scenarios for model application:

- applying the model on a link for future year, where that same link exists in the base year without any changes in road characteristics and type
- applying the model on a link for future year, where its road characteristics and type have changed

- applying the model on a link for future year, which did not exist in the base year but does have an estimate of a 'base year speed', for example as part of the planning process
- applying the model on a link for future year, which did not exist in the base year and does not have a base year speed estimate

In the case of the last scenario, where a link does not exist in the base year and does not have any estimate of a base year speed, the developed models can be used to predict average speeds directly. Given the limitations of the lack of a wide range of data on local variables (such as speed limits), it is advisable to treat the direct application of the model with care by, for example, imposing local speed limits constraints.

However, an alternative approach to applying the models is needed to address the first three scenarios, which are the three most likely scenarios. This approach involves predicting speed changes relative to a specific year, preferably the latest year where observed data is available. Specifically, this is done at link level by applying the model on both a base and future year, calculating a factor for the modelled speed changes between the years, and applying the factor on the observed speed in the base year.

Such an approach is beneficial in that it would generally improve model performance as observed base year speeds are relied upon and errors are partially cancelling each other when calculating the factor. Equation 5 demonstrates how urban speed changes can be predicted.

$$\hat{v}_{c,f,i} = v_{i,b} \times \frac{\hat{v}_{f,i}}{\hat{v}_{i,b}}$$

Equation 5. Modelling Corrected Speeds Using Urban Speed Changes Predictions

Where:

- $\hat{v}_{c,f,i}$ is the final corrected modelled average speed for a future year, f, and link i
- $\hat{v}_{f,i}$ is the modelled average speed for a future year, f, and link i
- $\hat{v}_{i,b}$ is the modelled average speed for a base year, b, and link i
- $v_{i,b}$ is the observed average speed for a base year, b, and link i

The proposed method for predicting speed changes has been tested and the results are presented below. Figure 9 shows a comparison of observed and predicted link speeds by applying the proposed method on the calibration data set. The figure uses 2019 as forecast year and 2013 as the base year. As can be seen, R2 has significantly improved compared to the performance of the previous section, now reaching 0.8. Producing the same figure on the validation data sets – sample 2 and sample 3 – shows similar patterns to the calibration data.

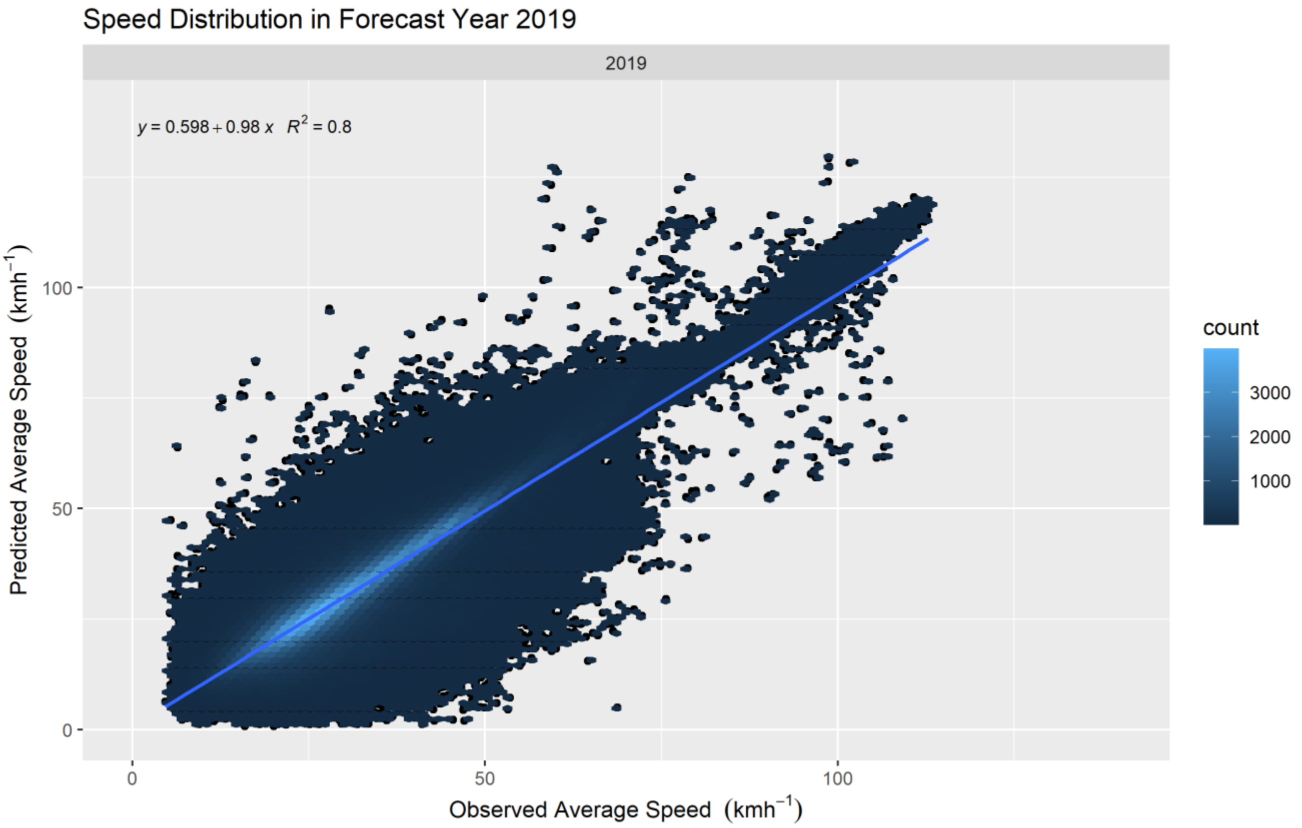

Figure 9. Observed versus Predicted Average Link Speed – Speed Changes Method

Using the proposed method, Figure 10 shows a comparison of observed and predicted link speeds at a more disaggregated level, by road type and road classification. The figure shows poorest fit for the dual carriageway motorway with an R^2 of 0.3. Variation in speeds on motorways can be attributed to different factors in the forecast year to that in the base year, for instance the changes in the use of variable speed limits. Single carriageway local streets and minor roads also perform relatively poorly due to local factors changing between the base and forecast years.

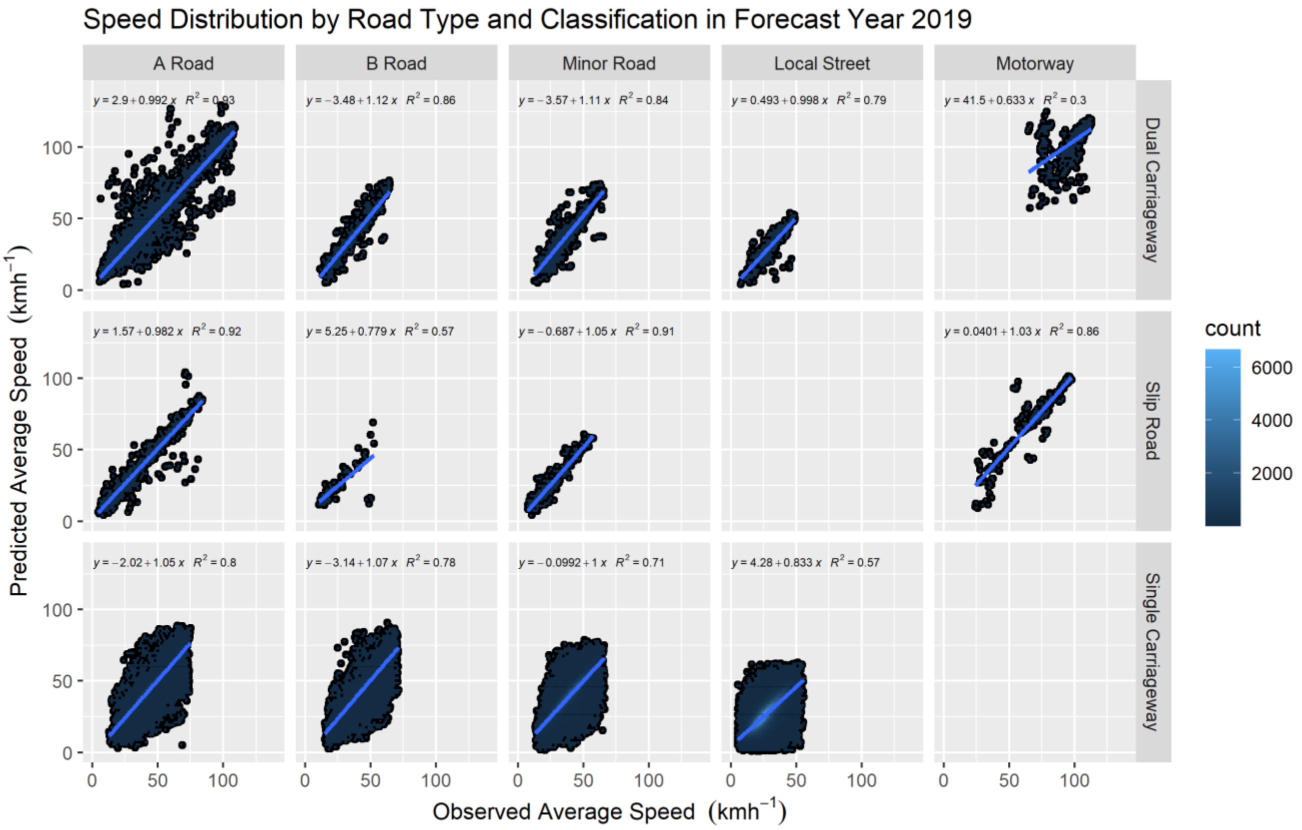

Figure 10. Observed versus Predicted Average Link Speed – Speed Changes Method

Figure 11 shows the distributions of observed and predicted average speeds comparing the two methods: using the models directly and using the models to predict speed changes. The proposed method of predicting speed changes shows significant improvements in capturing the variations of urban speeds. However, these need to be interpreted alongside Figure 9 which shows that there is both over and underpredictions in the data which is leading to similar distributions.

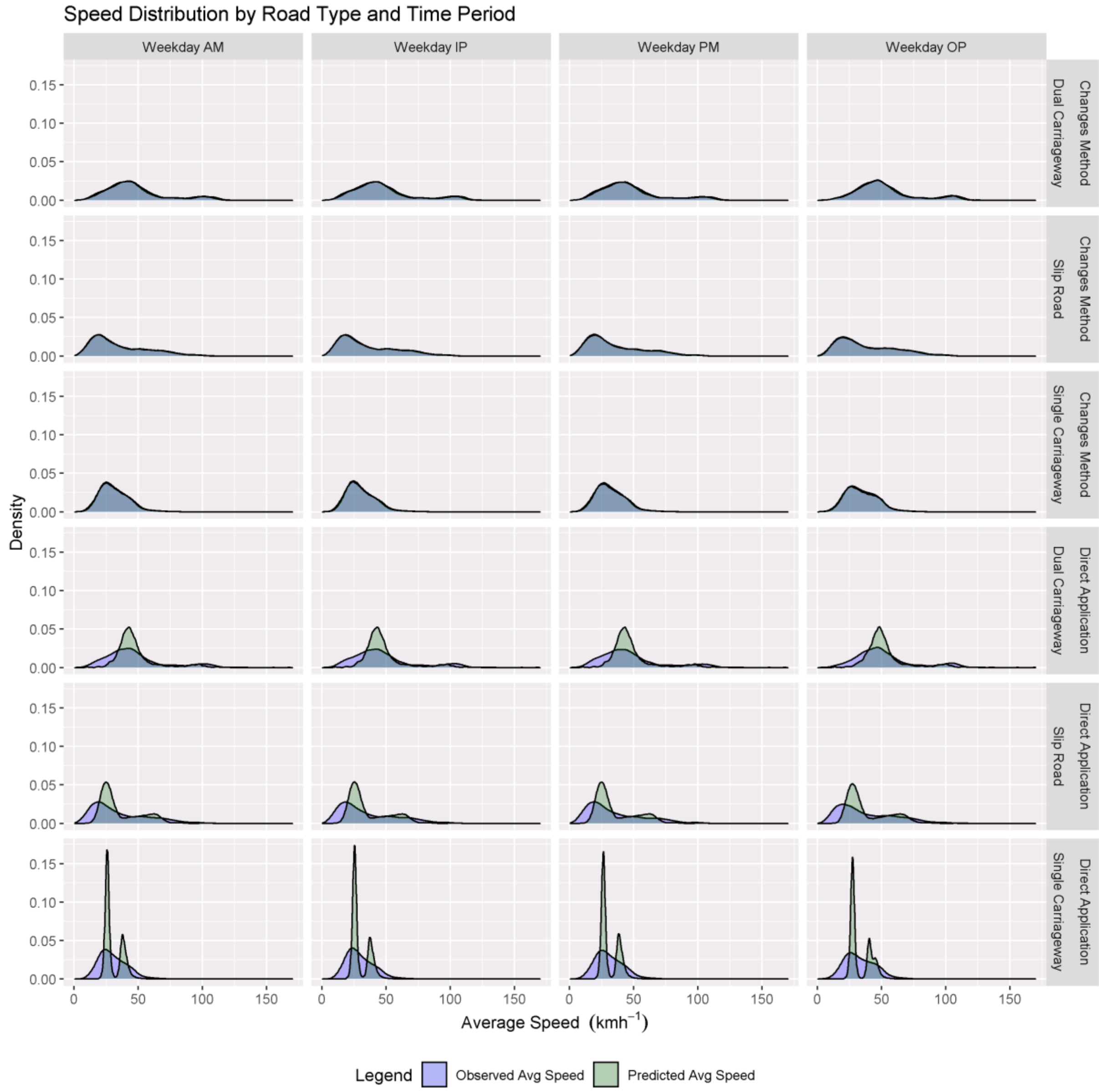

Figure 11. Observed versus Predicted Average Link Speed Distribution – Comparison of Speed Changes Method and Direct Application

Figure 12 and Figure 13 show the results of comparing the two methods on validation samples 2 and 3, respectively. The models are applied on future year 2019 for both validation data sets and, in the case of the predicting speed changes method, the base year is considered to be 2018. The validation results show similar conclusions to those of the calibration data set presented in Figure 11. As can be seen in Figure 13, using the proposed method produced plausible results when applied on London data. This gives more confidence in the developed method of predicting speed changes rather than predicting speeds directly.

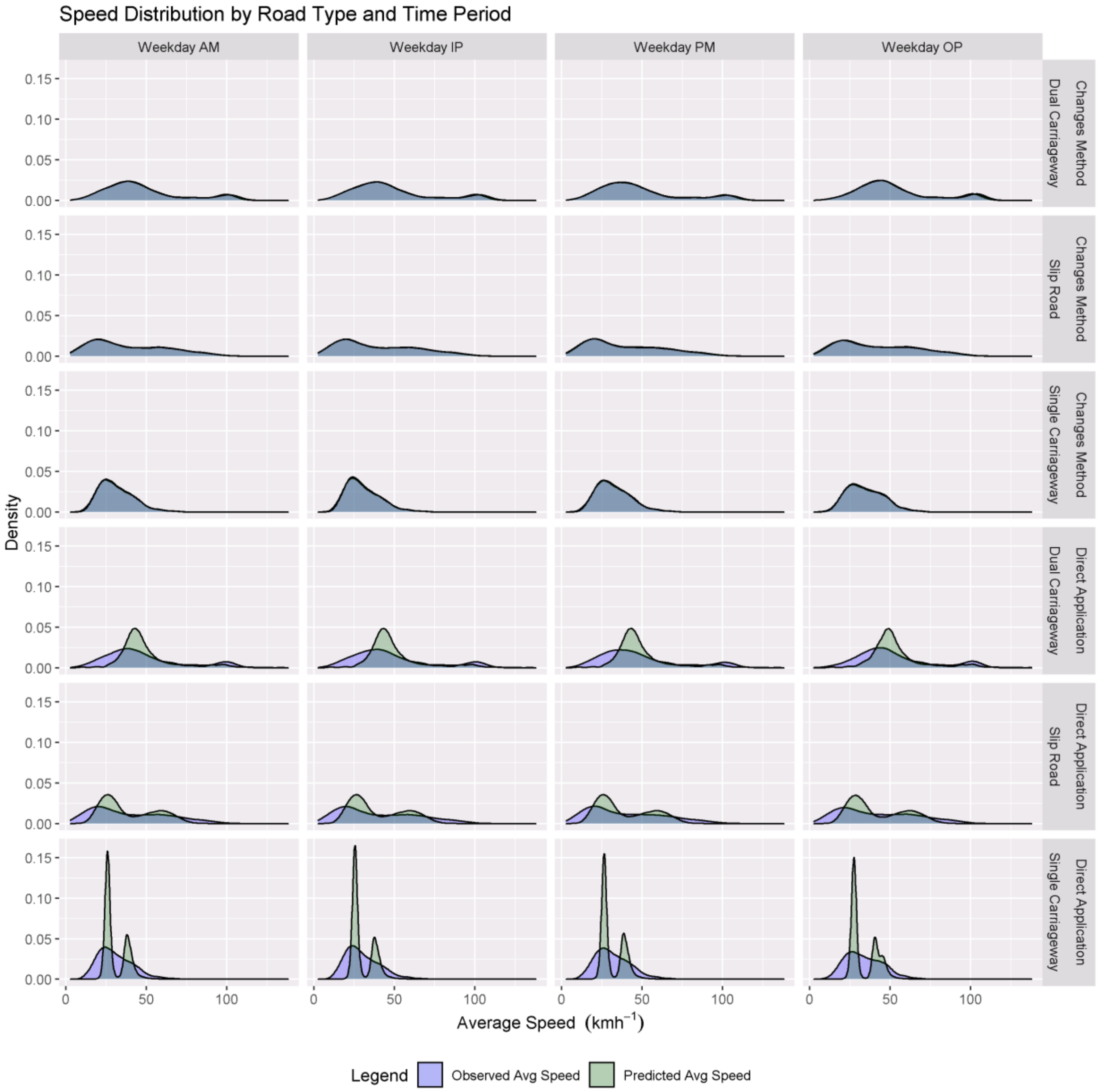

Figure 12. Observed versus Predicted Average Link Speed Distribution – Comparison of Speed Changes Method and Direct Application – Validation Sample 2

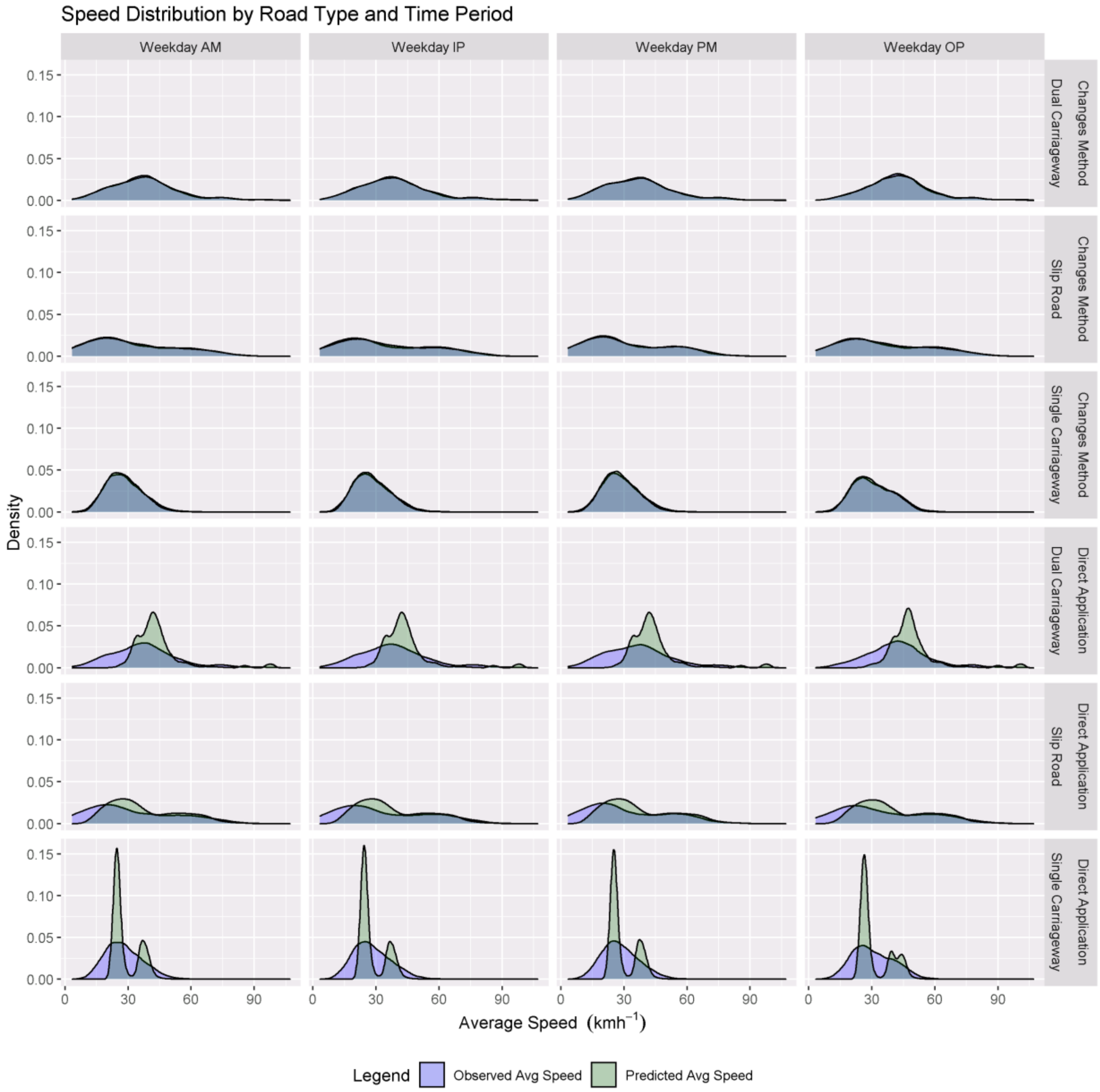

Figure 13. Observed versus Predicted Average Link Speed Distribution – Comparison of Speed Changes Method and Direct Application – Validation Sample 3

6. Conclusion

Summary

This technical note presented the results of a longitudinal analysis of Teletrac Navman data. The analysis aimed to meet the following objectives:

- developing an approach to sampling of the Teletrac Navman database which ensures robustness of the analysis and modelling
- developing a method to measure link speed stability to ensure that link speeds are sufficiently stable over certain periods and can be used to estimate changes
- a methodology which studies the impact of key characteristics potentially related to urban fixed speeds, and estimates urban fixed speeds and its changes at a specific confidence level
- the application of the developed method on NTMv5

Examples of key characteristics hypothesised to contribute to urban fixed speeds that were researched in this study include temporal variables (such as year, month, and time period), road characteristics, urban area type, and growth-related variables (such as jobs, population, population density, road density, trip ends, and trip rate growth).

This research has developed a two-step stratified sampling approach which allowed the generation of three sets of samples serving different purposes. The first sample was based on all the years collected (2013 to 2019) and was used for the explanatory analysis and model calibration. The second sample was used for model validation and was sampled for 2018 and 2019 only for the urban MOSAs of the first sample. The third sample was based on London urban MSOAs only and the years of 2018 and 2019, intended specifically for the purpose of studying the potential use of the model in London (which was not used as part of the calibration process).

The research has fused different data sets together and derived additional variables to be used in the explanatory analysis and model development stages.

As discussed in the Exploratory Analysis section, the study investigated the distribution of weighted average speeds temporally (such as by year, month, and time period) and by urban area type, road classification, and road type to better understand the relationship between traffic speeds and these variables. Correlation analysis and Analysis of Variance

(ANOVA) testing between average speeds and other key variables was also undertaken. The main findings from the exploratory analysis were as follows:

- the minimum level at which the modelling can be conducted is link level, yearly, and by time period
- at a time period level, OP average speeds are higher than those of AM, IP, and PM periods with consistent average speed for the AM, IP, and PM periods. In this study, time periods were defined as 07-10 for AM, 10-16 for IP, 16-19 for PM, and 19-07 for OP
- no systematic variation was found between days of the week (Monday to Friday)
- limited level of monthly variation was found, but some anomalies were detected in September 2018 and in local street slip road data in 2019 which were not included in the model development
- a considerable level of variation between hours was found in the OP period with limited level of variation for other periods
- limited level of variation was found between years without any evidence of a yearly trend in average speed, knowing that the analysis was made for a relatively small number of years
- while the correlation analysis showed that link length, road density, and population density are the variables with the highest correlation with average speed, these correlations are still relatively weak; no single variable was found to be strongly correlated with average speed
- ANOVA testing showed that the greatest explanatory power for the variation in traffic speed is obtained from using road classification and road type as categorical variables

The final data set used for modelling was average speed data for each of the sampled links for each year and time period.

The model development stage included an analysis of the sampled links' stability, measured by the coefficient of variation. To refrain from further assumptions on the definition of 'unstable' links, coefficient of variation was used as weights in the parameter estimation process, that is giving less weight to unstable links and more weight to stable links.

The model development process underwent a number of stages to reach the final set of models with the highest relative performance. Multivariate linear models were developed for each road type (single carriageway, dual carriageway, and slip road) and for the OP and AM/IP/PM time periods. This resulted in a total of six sub-models.

While the final set of models do not entirely capture the variation in speeds at the local level, they are able to predict the mean value of average speeds for each combination of road type and road classification.

The model performance was further enhanced by applying the model to predict speed changes rather than directly predicting speeds. This has led to significant improvement in the validation results. This is especially the case where there is no change in speed limits or other local factors between the base year and the forecast year.

Limitations and Future Work

Based on the analysis and modelling presented throughout this report, this research identified a number of limitations. These are discussed here alongside recommendations of possible future work:

- the study and modelling results present strong evidence that the models are capable of predicting the average speed for different combinations of road types and classifications. This was illustrated by comparing observed and predicted distributions of speeds for each road type/classification level, which showed that the models were able to capture the peaks of each of these distributions. Nonetheless, the results also present strong evidence that there is unexplained variation due to lack of data on the local level variables which can help better predict the local variability and micro-travel behaviour influencing average link speeds. This has led to a range of R^2 between 0.37 and 0.58 for the final six sub-models. This means that there is between 63% and 42% of unexplained variance in the final results. The performance was adjusted by applying the model using speed changes rather than directly applying them. The improvement is a result of propagated error partially cancelling each other for the base year and the future year. However, it is worth highlighting the limitations associated with limited availability of local level variables. Examples of local level variables that could be explored in the future include existence of on-street parking, number of lanes and speed limit
- the study has focussed on modelling speed using a multivariate linear modelling framework. Future work could involve studying more complex modelling structures such as the use of nonlinear models and the use of smoothing functions that have the possibility of improving the overall performance. However, more complex modelling structures (such as the use of smoothing functions) can lead to difficulties in interpretability and application which was avoided in the current research
- the study made use of disaggregate 15 minutes data in order to produce link level speed data averaged by year and time period. The aggregation process was weighted using the number of observations provided in the Teletrac Navman data. The analysis highlighted certain anomalies in certain months in 2018 and 2019 which requires further investigation with traffic flow data. However, it would require significant effort to match available traffic flow data with Teletrac Navman data, as well as significant assumptions on links where traffic flow data is not available
- the study has focussed on studying the variability by road characteristics as well as temporally (such as by time period, day of the week, and year). The modelling results attempted to explain spatial variation in average link speeds using spatial characteristics (such as population, road density, and car ownership). Exploring the variation in average speed spatially is also a potential further study using the data used
- the study makes use of NTEM data as it can be derived for future years. However, this means that the methodology is dependent on consistency between current and future versions of NTEM
- the models were validated for different factors across different road types/classification combinations. However, some of the validation results can be subject to relatively lower data points such as London motorways or local street slip roads.

Recommendations

This study does not recommend specific changes to the current method of link selection in NTMv5. It is recommended to keep using the speed-flow curve method in all instances where there is a detailed representation of junctions and junction delays. Where that is not applicable, applying the final set of models as described in the Final Model Application section is recommended.

The outcome of this research is a final set of models with their model variables and model coefficients, as well as a description of their application set out in the Urban Speed Modelling section. To summarise, the following has to be prepared in order for the model to be applied:

- identification of the link for which average speed needs to be forecast
- information on the link needs to be gathered. These are the final model predictors: road type, road classification, time period, link length in metres, population density in population per m^2, road density in kms per individual, population with one or more cars, and number of jobs. All of these (except for road type, road classification, time period, and link length) are to be calculated for the MSOA in which the link is situated
- the same information needs to be gathered for the links in the base year as for the forecast year in order to model base year speeds. Different options are available for the selection of the base year such as the model base year, or the latest year where average speed data for the links is observed
- observed average speed for the link in the selected base year needs to be collected and aggregated for the year and time period under consideration

Once these are prepared, the model structure and coefficients can be applied to determine the average link speed for the selected future year and time period.

The Final Model Application section has presented the validation results of applying the model incrementally (that is modelling speed changes). The results show that highest performance was obtained for slip roads, dual carriageways (excluding motorway dual carriageways), and single carriageway A roads and B roads. The remaining combinations of road types and classifications are subject to higher uncertainties in factors influencing their speeds and in the changes in speeds by year. It is thus recommended to use the developed model on those road types and classifications where the model performs best.

References

Atkins., (2019). NTM Future Development Quality Report. London: Department for Transport.

Atkins., (2018). Task 530 Trans-Pennine South Regional Transport Model – Impact of Fixed Speed Areas. London: Highways England.

Atkins., AECOM., and Maher, M., (2018). 20mph Research Study – Supporting Technical Appendix: Analysis of GPS Journey Speeds in Case Study Areas [online]. London: Department for Transport. [Viewed 26 June 2020]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/784905/technical-appendix-2-20mph-gps-journey-speed-analysis-report.pdf

Atkins and CH2M., (2018). Task 530 North Regional Traffic Model – Fixed Speed Area Assessment. London: Highways England.

Bibby, P. and Brindley, P., (2013). Urban and Rural Area Definitions for Policy Purposes in England and Wales: Methodology (v1.0) [online]. Department for Environment, Food & Rural Affairs, Department for Communities and Local Government, Welsh Government, and Office for National Statistics. [Viewed 15 September 2020]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/239477/RUC11methodologypaperaug_28_Aug.pdf

Chen, X., Pao, H. and Lee, Y., (2014). Efficient Traffic Speed Forecasting Based on Massive Heterogenous Historical Data. In: Institute of Electrical and Electronic Engineers (IEEE). IEEE International Conference on Big Data (Big Data), 23-30 October 2014, Washington DC, USA [online]. New York: Institute of Electrical and Electronics Engineers. pp. 10-17. [Viewed 23 June 2020]. Available from: doi: 10.1109/BigData.2014.7004425

DfT., (2020). TAG Unit M1.2: Data Sources and Surveys [online]. London: Department for Transport. [Viewed 18 June 2020]. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/888340/tag-unit-m1.2-data-sources-and-surveys.pdf

Elliott, M.A., McColl, V.A. and Kennedy, J.V., (2003). Road Design Measures to Reduce Drivers' Speed via 'Psychological' Processes: A Literature Review. Wokingham: Transport Research Laboratory.

Hooper, E., Chapman, L. and Quinn, A., (2014). The Impact of Precipitation on Speed-Flow Relationships along a UK Motorway Corridor. Theoretical and Applied Climatology [online], 117, 303-316. [Viewed 22 June 2020]. Available from: doi: 10.1007/s00704-013-0999-5

Huang, S.H. and Ran, B., (2003). An Application of Neural Network on Traffic Speed Prediction under Adverse Weather Condition. TRB Annual Meeting, 2003. [Viewed 23 June 2020]. Available from: https://www.researchgate.net/publication/265318230_An_Application_of_Neural_Network_on_Traffic_Speed_Prediction_Under_Adverse_Weather_Condition

ONS., (2018). Rural Urban Classification (2011) of Middle Layer Super Output Areas in England and Wales [online]. Office for National Statistics. [Viewed 15 September 2020]. Available from: https://data.gov.uk/dataset/9c0e093d-d267-4eb8-90d8-54475ab4d1ff/rural-urban-classification-2011-of-middle-layer-super-output-areas-in-england-and-wales

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Sharma, A., Ahsani, V. and Rawat, S., (2017). Evaluation of Opportunities and Challenges of Using INRIX Data for Real-Time Performance Monitoring and Historical Trend Assessment [online]. Lincoln: Nebraska Department of Transportation. [Viewed 26 June 2020]. Available from: https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1023&context=ccee_reports

TfL., (2013). Road Task Force – Technical Note 9: How Does the Road Network Perform in Terms of Speed, Congestion and Journey Time Reliability? [online]. London: Transport for London. [Viewed 26 June 2020]. Available from http://content.tfl.gov.uk/technical-note-09-how-does-the-road-network-perform-in-terms-of-speed-congestion.pdf

Vanajakshi, L. and Rilett, L.R., (2004). A Comparison of the Performance of Artificial Neural Networks and Support Vector Machines for the Prediction of Traffic Speed. In: Institute of Electrical and Electronic Engineers (IEEE). IEEE Intelligent Vehicle Symposium, 14-17 June 2004, Parma, Italy [online]. New York: Institute of Electrical and Electronics Engineers. pp. 194-199. [Viewed 23 June 2020]. Available from: doi: 10.1109/CCNC.2015.7157943

Wang, J., Shang, P. and Zhao, X., (2011). A New Traffic Speed Forecasting Method Based on Bi-pattern Recognition. Fluctuation and Noise Letters [online], 10(01), 59-75. [Viewed 23 June 2020]. Available from: doi: 10.1142/S0219477511000405

WSP., (2018a). TfN Strategic Development Corridor (SDC) Studies – Fixed Speed Adjustment for Urban Areas. Manchester: Transport for the North.

WSP., (2018b). Trans-Pennine Tunnel Study Plus – Fixed Speed Adjustment for Urban Areas. London and Manchester: Highways England, Department for Transport, and Transport for the North.

Ye, Q., Szeto, W.Y. and Wong, S.C., (2012). Short-Term Traffic Speed Forecasting Based on Data Recorded at Irregular Intervals. IEEE Transactions on Intelligent Transportation

Systems [online], 13(4), 1727-1737. [Viewed 23 June 2020]. Available from: doi: 10.1109/TITS.2012.2203122

York, I., Bradbury, A., Reid, S., Ewings, T. and Paradise, R., (2007). The Manual for Streets: Evidence and Research [online]. Wokingham: Transport Research Laboratory. [Viewed 24 June 2020]. Available from: https://trl.co.uk/sites/default/files/TRL661%283%29.pdf