# Even busier than usual: modelling excess congestion on the Strategic Road Network

Ayman Boustati*, Alvaro Cabrejas Egea*, Colm Connaughton†, Peter De Ford González*, Laura Guzmán Rincón*, Steve Hilditch ‡, Guillem Mosquera Doñate*

**Abstract**

The National Traffic Information System (NTIS) provides real-time measurements of traffic volume and speed on the Strategic Road Network. NTIS data has been used to assign a traffic profile to each section of road or "link" on the network. The profile gives the expected travel time at any time of day. Thales UK are members of a consortium operating the NTIS system on behalf of Highways England. One of their functions is to provide real-time information on deviation-from-profile events (DPEs) caused by congestion in excess of that expected from profile. This report examines simple data-driven algorithms for predicting the duration of DPEs and discusses ways in which the skill of such predictions can be quantified. A large dataset containing 3 months of traffic data from the M11 and M6 motorways is used. The DPE events are extracted using modern cloud-based data analytics methods and their statistical properties are characterised. Simple heuristic models, including the one currently used by Thales, are implemented and benchmarked against the scoring system used by Highways England to assess prediction skill. Our preliminary results indicate that improved skill scores are possible with minimal increase in model complexity. The best performance is obtained using a weighted average of a group of simpler models. The Highways England prediction skill score is sub-optimal from a user perspective. It can be demonstrated that this sub-optimality could be exploited to provide better predictions to users without significantly compromising the skill score.

**Keywords**

Transportation systems — Data science — Predictive modelling — Intelligent mobility — Time to clear — Complex Adaptive Systems

*Mathematics of Real-World Systems Centre for Doctoral Training, University of Warwick, Coventry CV4 7AL, United Kingdom
†Centre for Complexity Science, University of Warwick, Coventry CV4 7AL, United Kingdom
‡Thales UK, Dolphin House/Ashurst Dr, Stockport SK3 0XB, United Kingdom

## Contents

## 1. Introduction

Transport is found at the very essence of many physical phenomena occurring in the universe. From quantum field theory to galaxy collisions, the process of exchanging mass, energy and momentum is crucial to understand the world we live in. From a more mundane perspective, although macroscopic human transportation involved walking or swimming from the beginning, elaborate modes of transports — such as domestic animals, canoes or aircraft — have been crucial for the development of society. In particular, here we are concerned with a phenomena born in the 20th Century that still prevails as a very annoying and costly problem affecting all layers across society: the traffic jam.

In fact, investment in the infrastructure of transport systems is an essential driver for the economy [1]. In developed countries such as the UK, there is limited capability to increase the physical capacity of the transport infrastructure. This is specifically the case for the Strategic Road Network, comprising approximately 4,400 miles of motorways and major trunk roads across England [2]. Current transportation policy and research is focused on Intelligent Mobility rather that the construction of new infrastructure: in the context of road transport, Intelligent Mobility aims to utilise new technologies and real-time data to improve the efficiency of existing physical infrastructure [3].

The UK is considered world-leading in its ability to collect and process real-time data from its road network. Highways England are responsible for making this data available through the National Traffic Information Service (NTIS) [4]. Highways England collects data of the speed, flow and travel time on the Strategic Road Network, using sensors on the road and in vehicles. These are operated by Thales UK in collaboration with other partners.

As a member of the consortium operating the NTIS, Thales UK is tasked to provide real-time information about the so-called "deviation-from-profile events" (DPEs), defined as the continuous exceedance from usual (averaged) observations in terms of travel time through delimited road sections. These events usually signify severe congestion on the road and can cause major disturbances to the network operations. For this reason, Thales UK supplies predictions on the lengths of these events to aid the network operators. The aim of this project is to use the data provided by NTIS to better understand the DPEs and to develop simple data-driven methods that refine the predictions that Thales UK currently makes.

There are two clearly distinct approaches to this problem. The first is based on the physical modelling of vehicular traffic, usually involving over-simplified microscopic models or computationally expensive fluid-like descriptions. Although providing useful theoretical insight and stimulating research, the application of such models to real-time prediction is too involved for the scope of this report. The second approach involves the statistical analysis of several measurable magnitudes present in the system (e.g. travel time series) and making use of heuristic and statistical forecasting methods. This last approach has been taken for the development of the algorithms in the report, for its computational simplicity and its low dependency on physical principles.

It is worth noticing that most of the existent work on this field falls in the first approach: many efforts have been done since the 1950s to understand the open problem of vehicular traffic modelling, although this is still subject to debate and cannot provide many solid applications. Usually, the mathematical work produced concerning vehicular traffic can be categorized in either macroscopic or microscopic modelling. For example, Interacting Particle Systems (IPS) studies, Kinetic Theory models and Car-Following models fall into the microscopic description [5] [6]. IPS successfully explain how conservation laws [7] arise from the microscopic stochastic process associated [8]. On the other hand, Kinetic models — based on Boltzmann-like equations for the Kinetic Theory of Gases — allow for a better understanding of the impact of different physical and behavioural considerations such as car length, safe distance, reaction times or driver-desired velocities. Finally, Car-following models [5] are good for establishing a deterministic link between driver behaviour and stability of the flow. Different driver behaviours allow to study the sensitivity to traffic jams. Macroscopic or fluid-like models — the fundamental being Lighthill-Whitham model [9] — are useful to derive long-term behaviour given certain initial conditions [10], in the form of Cauchy or Riemann problems. In general, realistic approaches involve solving stochastic versions of Burgers' equation [7], this being the origin of the aforementioned computational complexity.

As mentioned before, the approach used in this project differs from the mentioned literature for its principle-free or heuristic flavour. Note that as far as the authors are concerned, none of the data-driven models and algorithms presented here have ever been published up to this date, so they can be considered novel contributions to the field.

This report is organised as follows: Section 2 contains a concise explanation of the process undertaken to acquire, select and process data from the NTIS; Section 3 performs general statistical analysis on the selected dataset, including clustering, summary statistics of main observables and speed/flow trajectory analysis; Section 4 reviews the existent approach used by Thales to tackle the DPE duration forecasting problem; the core of the report is presented in Section 5, where we propose five heuristic algorithms, one statistical-based method and a weighted multi-model approach for forecasting travel times and DPE durations; finally, Section 6 discusses results and gives an overview of project limitations, future directions and other final remarks. An appendix can be found at the end of the document with supporting information for the report.

## 2. Data: Selecting and Processing

The NTIS employs a directed network model to represent the major roads in England; wherever a road encounters a diversion, a node is placed on the network. The basic building blocks of this network are the so-called *links*: they are portions of carriageways representing edges between nodes connected by a road. The links on the network incorporate thousands of induction loops (sensors) at different sites, which report speed and flow data to a centralised system.

NTIS data is used to calculate and assign traffic profiles to the links across the network. A traffic profile for a link reports the average time to traverse the link under expected road conditions for a specific time and day [11].

The data was acquired from the NTIS web service for the months of March, April and May. Upon acquisition, the data was cleaned, formatted and stored in a database for easy access (for more information refer to Appendix A).

Linear interpolation was performed only where there was missing data for 10 or less consecutive minutes. Otherwise, a missing value flag was raised.

### 2.1 Data Contents

For each link on a specific date, the data consists of one entry per minute, containing the following:

- Average values of traffic speed, traffic flow and traffic headway.
- Average values of current travel time, profile travel time and free flow travel time.
- Event flags for spontaneous congestion events, weather events and other types of events.

### 2.2 Data Selection

The steps of process of selecting and processing the data are sequentially described in the following subsections and summarised in Figure 1.

#### Location Dependent Data Selection

Link-level data was only extracted for the sections of M6 and M11 shown in Figure 2. These have a fixed speed limit throughout the whole section, a desirable property which avoids the introduction of additional noise or perturbations in the traffic measurements.

#### Training and Testing Datasets

To perform the analysis over the extracted data, it was divided into two subsets:

- **Training Dataset:** corresponds to 70% of all data. The exploratory analysis and the calibration of algorithms is performed on this subset.
- **Testing Dataset:** corresponds to the remaining 30% of all data. To avoid overfitting, this data set is used only for testing the performance of algorithms.

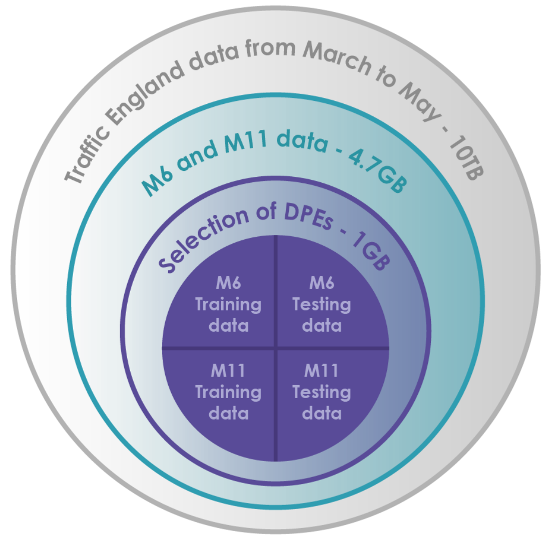



**Figure 1.** Selection and processing of Traffic England data with approximate sizes.

3

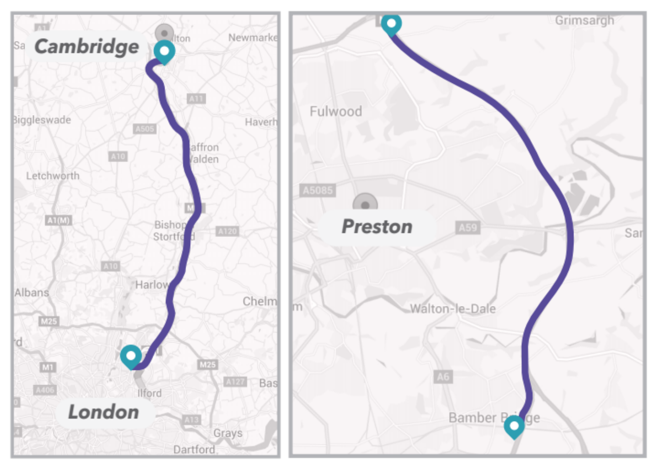
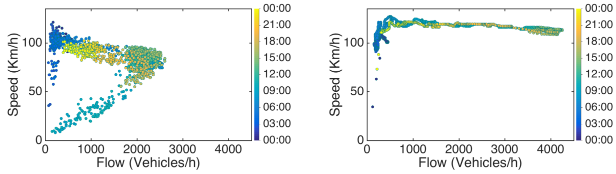



**Figure 2.** Sections of the M6 and M11 motorways selected for the analysis.

#### Time Dependent Data Selection

To ensure the quality of the analysis the data was selected based on the following criteria:

- Discard weekends and bank holidays, focusing on days with a ”normal” traffic regime. Qualitative difference can be observed for both regimes in Figure 3.
- Only consider data points between 5am and 11pm, to avoid diluting the data with entries which are non representative of the problem due to low occupancy.
- Discard time periods containing an accident flag, as the behaviour of these depends heavily on type of event, making them significantly different from spontaneous congestions [12], placing their dynamics beyond the scope of this report.




**(a)** Traffic flow and speed evolution over a Monday.

**(b)** Traffic flow and speed evolution over a Sunday.

**Figure 3.** Samples of the evolution of traffic.

# 3. Exploratory Data Analysis

## 3.1 Deviation From Profile Events

First, the attention was focused analysing what was labelled as the Deviation from Profile Events (DPE). These events are defined as the continuous exceedance (5 minutes or more) of the profile travel time by more than six seconds. They are specified by a time series of the Deviation from Profile Intensities $x_t$, which are defined as follows:

$$x_t = c_t - p_t - 6 \tag{1}$$

where $c_t$ and $p_t$ are the current travel time and profile travel time respectively. Here $x_t$ can be positive or negative, but since a negative deviations implies lower travel times, these were not taken into account. The DPEs were obtained from the raw travel time data using a threshold based approach that was provided by Thales through private communication [13]. It was decided to discard any DPE that is shorter than 20 minutes or longer than 360 minutes in duration. Furthermore, events only considered where the maximum $x_t$ was greater than or equal to 20. Given the noisy nature of the DPE time series, it was decided to smooth the observations using a low-pass filter (See Appendix C for more details).

The rest of this section details the exploratory data analysis that was performed on the DPE data.

## 3.2 Clustering

Testing was conducted to check if the DPEs can be categorised based on their raw time series data. To this end, we applied an unsupervised clustering algorithm, Agglomerative Hierarchical Clustering with Complete Linkage [14] on a collection of different DPE time series. As a measure of distance, Dynamic Time Warping (DTW) [15] with a Manhattan local cost was used. The algorithm was able to detect clear outliers; however, it was unable to produce visually distinguishable clusters. This is suspected to be caused by the presence of a significant number of noisy observations that skewed the results of the clustering algorithm.

## 3.3 Summary Statistics at Link Level

Six distinct statistics were calculated at the link-level. These calculations were performed separately for the M6 and M11. Three of them were *unnormalised*, while the rest were *normalised* by re-scaling both the time and the intensity scales.

In the following subsections, we present descriptions of the statistics, along with the results of their calculations.

### Description of the Unnormalised Statistics

The following three unnormalised statistics of the DPE were calculated:

- **Duration:** corresponds to the duration of a DPE, measured in minutes.
- **Global Maximum of the Intensity:** is defined as the greatest deviation from profile in a single DPE. The maximum is measured in seconds.
- **Size:** is defined as the area under the curve of the time series of a DPE. Physically it represents the aggregated lost time by all vehicles traversing the link for the duration of the DPE.

The duration, the global maximum intensity and the size are graphically illustrated in Figure 4.

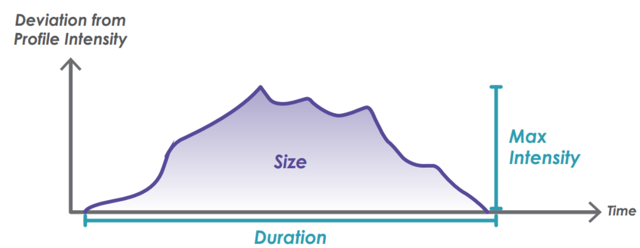


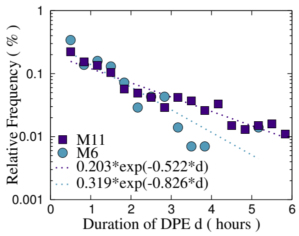

**Figure 4.** Graphical representation of the duration, maximum intensity and size of a DPE.

Figure 5 summarises the unnormalised statistics for M6 and M11, where the distribution of each quantity is shown. As observed in Figure 5a, the DPE duration distribution follows an exponential law with a characteristic length of 1.20 hours (72 mins) for M6 and 1.92 hours (115 mins) for M11. Similarly, according to Figure 5b, the distribution shows an exponential decrease and the characteristic maximum intensity is calculated. For M6 it is 0.88 mins (52 sec) while for M11 it is 0.95 mins (57 sec). Finally, the characteristic size is 43 and 41 for M6 and M11 respectively, as shown in Figure 5c.

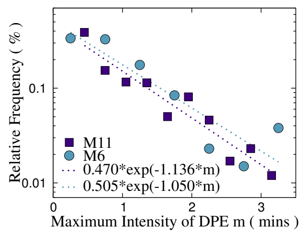
### Description of the Normalised Statistics

The normalised statistics calculated were:

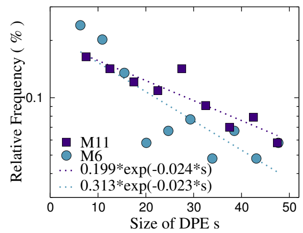
- **Location of the Maximum:** is the normalised time at which the global maximum of a DPE occurs. This is represented by a number between 0 and 1.
- **Symmetry Factor:** measures the symmetry of the time series of a DPE around its global maximum. The symmetry factor is defined as the ratio of the decline to the growth, where the decline is the time from the global maximum to the end of a DPE and the growth is the time from the beginning of a DPE to the global maximum.
- **Trapezium Parameters:** a trapezium can be specified by four parameters $(a,b,c,h)$, where $a$ is the length of




**(a)** Distribution of DPEs durations.




**(b)** Distribution of maximum intensity of DPEs.




**(c)** Distribution of size of DPEs.

**Figure 5.** Unnormalised statistics showing the distribution of sizes, maximum intensities and durations of DPEs.

projection of the left leg onto the longer base, $b$ is the length of the projection of the shorter base onto the longer base and $c$ is the length of the projection of the right leg onto the longer base and $h$ is the height.

Assuming a DPE consists of: build-up, plateau and clearance, we can model the event as a trapezium with $a$ representing the time in build-up, $b$ representing the time in plateau and $c$ representing the time in clearance. The remaining parameter $h$ is heuristically fixed to $0.8\times$Global Maximum Intensity. This is illustrated in Figure 6.

Once the parameters that define the shape have been obtained, the length of the base and maximum intensity are normalised to allow comparison between different DPEs.

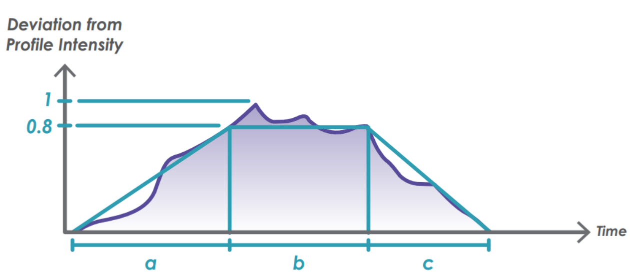
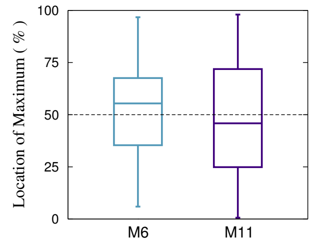
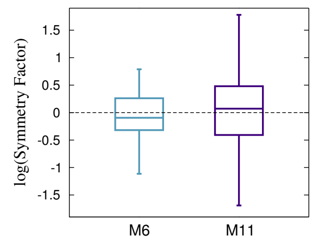



**Figure 6.** A schematic graphical representation of the trapezium model.

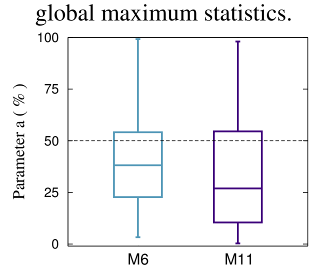
The normalised statistics are summarised in the box plots presented in Figure 7. A box plot representation of each statistic is shown, comparing the overall patterns for M6 and M11. Although the distribution of the position of global maximum differs between motorways, it is generally located near the middle of the DPE, as shown in Figure 7a. A similar result is obtained for the symmetry factor, observed in Figure 7b. Finally, the distributions for the trapezium parameters are presented in Figures 7c, 7d and 7e.



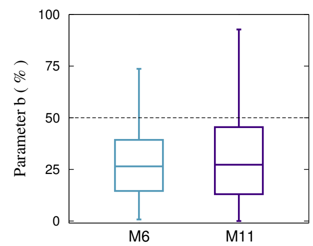

**(a)** Distribution of position of global maximum statistics.



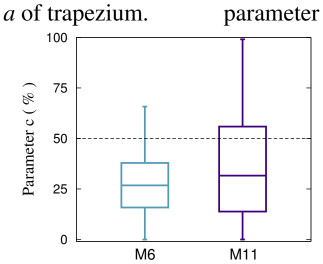

**(b)** Distribution of symmetry factor statistics.




**(c)** Distribution of size of parameter $a$ of trapezium.




**(d)** Distribution of size of parameter $b$ of trapezium.




**(e)** Distribution of size of of parameter $c$ of trapezium.

**Figure 7.** Box-plots of the normalised statistics.

### 3.4 Speed/Flow trajectories

Examining different Speed vs. Flow plots of DPEs a dichotomous pattern can be noticed. In some of them, the trajectory of the transition from a normal regime to maximal congestion is very different to the one in the opposite direction. This contrasts with other plots in which these trajectories are very similar. An example of this behaviour is shown in Figure 8.

It was suspected this pattern is related to either the link length or the symmetry factor. To test this, a distance metric was developed that quantifies the difference between any two trajectories $u$ and $v$. The trajectories were fixed to 100 points each using interpolation:

$$u: \{p_1^u, p_2^u, \ldots, p_i^u, \ldots, p_{100}^u\}$$
$$v: \{p_1^v, p_2^v, \ldots, p_i^v, \ldots, p_{100}^v\} \qquad (2)$$

where a point $p_i^\bullet$ is a coordinate $(flow, speed)$ in the Speed vs. Flow graph. The distance metric developed is denoted by

$D$:

$$D = \frac{1}{100} \sum_{i=1}^{100} ||p_i^u - p_i^v|| \quad (3)$$

$D$ was calculated for the two trajectories in every DPE, the plot obtained can be seen in Figure 9.

It was not possible to find a meaningful threshold to cluster the DPEs. Moreover, after some experiments, we were unable to find any relationship of $D$ with link length or symmetry factor. This dichotomous pattern can be better analysed using physical models and taking into account adjacent links.

It was also hypothesised that the trajectory difference was the result of hysteresis in the system. Hysteresis is a time-based dependence of a system's output on present and past inputs that causes a lack of reversibility when varying the input [16]. Further analysis is needed to determine if the system actually exhibits hysteresis. However, taking into account that this is a system far from equilibrium, this seems unlikely.

## 4. Current Approach

Thales UK calculates real-time predictions of the time it takes to return to profile during a DPE. The complete behaviour depends on a multitude of factors, most of which are very difficult to measure. A solution to this problem is to use heuristics and statistical methods to make the predictions. This section presents the prediction and scoring algorithm currently used.

### 4.1 Existing Algorithm

Through private communication [13] Thales UK provided the algorithm they currently use to predict the time it takes to return to the profile at each time step during a DPE. For every newly observed point in the DPE time series, the algorithm makes a prediction on the duration of the DPE based on a *prediction rule*. This rule uses all the observations from the beginning of the event until the current point to make the prediction. The algorithm assumes a minimum threshold $B$ for the duration of a DPE, under which any predictions will be given the value $B$ (typically set to 20 minutes). The algorithm is detailed in Algorithm 1, where the `Predict` function is the prediction rule. The prediction rule that Thales UK currently uses assumes that a DPE is symmetric around its global maximum; therefore it predicts that the event will last twice the time it takes to reach the observed global maximum. This is given in Algorithm 2.

### 4.2 Existing Scoring

The skill of the prediction of an algorithm for this problem is scored according to the Average Prediction Error at the Midpoint $E_{\text{Mid}}$, defined below for a sample of $N$ DPEs:

$$E_{\text{Mid}} = \frac{1}{N} \sum_{i=1}^{N} E_{\text{Mid}}^{(i)} \quad (4)$$

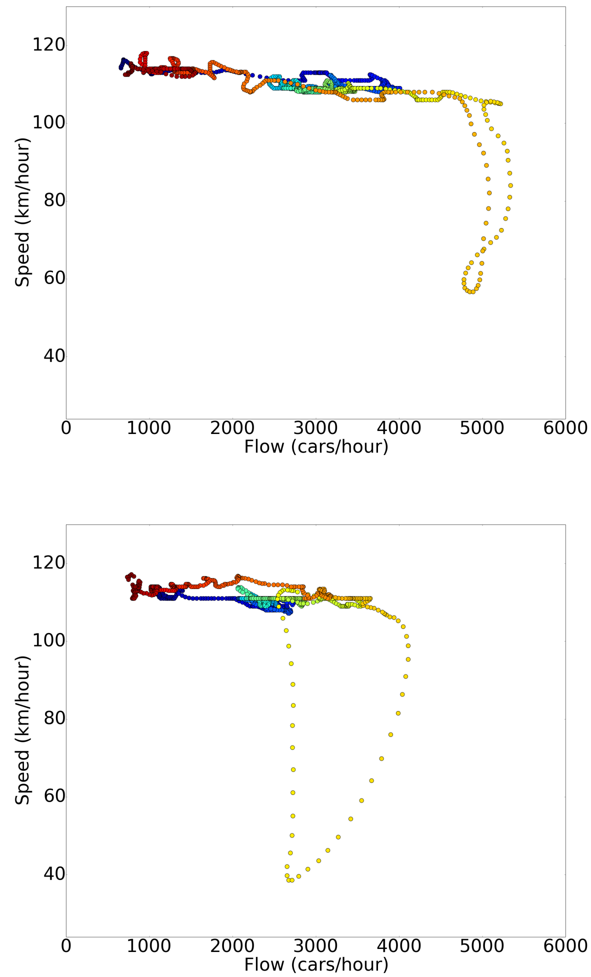



**Figure 8.** Two dichotomous behaviours. The colour gradient represents the time of the day from red, signifying early hour starting from 5am, all the way to 11pm, represented by dark blue. The top figure shows that the trajectory going from the normal regime to maximal congestion is similar to the trajectory going in the opposite direction. This contrasts with the bottom figure where there are two distinct trajectories.

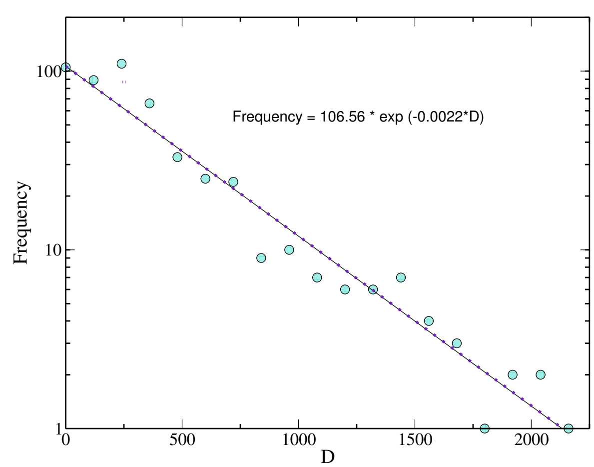



**Figure 9.** Histogram of $D$ for all DPEs

**input** : A time series of intensities: $\{x_1, x_2, \ldots, x_T\}$
**output** : A series of predictions for every point in time: $\{y_1, y_2, \ldots, y_T\}$
initialise B;
**for** $t$ in $1{:}T$ **do**
    $y_t \leftarrow \texttt{Predict}(x_{1:t})$;
    **if** $y_t < B$ **then**
        $y_t \leftarrow B$;
    **end**
**end**

**Algorithm 1:** General algorithmic paradigm provided by Thales for predicting the return to profile at every point in a DPE.

**Function** *Predict*($x_{1:t}$)
    buildup_time $\leftarrow \arg\max_{i\in\{1,\ldots,t\}}(x_{1:t})$;
    prediction $\leftarrow$ 2*buildup_time;
    **return** prediction

**Algorithm 2:** Prediction rule that Thales UK currently uses.

where $E^{(i)}_{\text{Mid}}$ is the percentage error for the $i^{th}$ DPE. $E^{(i)}_{\text{Mid}}$ is calculated as follows:

$$E^{(i)}_{\text{Mid}} = 100\frac{|y^{(i)} - \hat{y}^{(i)}_{\text{Mid}}|}{y^{(i)}} \tag{5}$$

where $y^{(i)}$ is the true value of the duration of the $i^{th}$ DPE and $\hat{y}^{(i)}_{\text{Mid}}$ is the predicted value at the midpoint of the $i^{th}$ DPE.

## 5. Proposed Prediction Methods

In this section we will discuss some algorithms that were developed to solve the Return to Profile problem, i.e. to give real-time forecasting for the duration of DPEs.

### 5.1 Heuristic Algorithms

#### Null Model

An assumption can be made that the length of DPEs has a narrow inter-quartile range. Therefore, the median DPE length is a good estimation of duration of any DPE. This corresponds to the prediction rule in Algorithm 3.

**Function** *Predict*($x_{1:t}$)
    prediction $\leftarrow$ median length of all DPEs on the motorway;
    **return** prediction

**Algorithm 3:** Prediction rule that corresponds to predicting the median.

#### Relative Maximum

An improvement one can make on the original algorithm is to use a relative maximum instead of a global maximum for prediction. Hence, we estimate the time to clear from the current point by the time to reach the latest observed local maximum. This is given in Algorithm 4.

**Function** *Predict*($x_{1:t}$)
    buildup_time $\leftarrow \max_{s\in\{1,\ldots,t\}} s \quad \text{s.t.} \quad x_{s-1} \leq x_s$;
    prediction $\leftarrow$ 2*buildup_time;
    **return** prediction

**Algorithm 4:** Prediction rule when using a local maximum as a reference point.

#### Midpoint Prediction

Naively, one can assume that every newly observed point in a DPE occurs at the midpoint of that event. While this assumption seems outright wrong at the beginning, it turns out that it provides the best prediction results based on the scoring criterion. This assumption simply means that the algorithm should predict that a DPE will last the same time that it has already elapsed, i.e. the duration of the event is just double the current time. This is given by to the prediction rule in Algorithm 5.

**Function** *Predict*($x_{1:t}$)
    buildup_time $\leftarrow t$;
    prediction $\leftarrow$ 2*buildup_time;
    **return** prediction

**Algorithm 5:** Prediction rule corresponding the assumption that every new point is the midpoint of a DPE.

#### Multiplying by a Constant Factor

One method that empirically outperforms the original algorithm is multiplying its build up (time to the observed global maximum) by a constant factor of 2.4. This is shown in Algorithm 6.

**Function** *Predict*($x_{1:t}$)
    buildup_time $\leftarrow \arg\max_{i\in\{1,\ldots,t\}}(x_{1:t})$;
    prediction $\leftarrow$ 2.4*buildup_time;
    **return** prediction

**Algorithm 6:** Prediction rule corresponding to multiplying the build up of the original algorithm by 2.4.

#### Intensity Scaling

Assuming that the time remaining to return to profile is proportional to the intensity, we can predict the time to clear based on the current value of the intensity by scaling it by factor $C$ as in Algorithm 7.

```
Function Predict(x_{1:t})
    initialise C;
    time_to_clear ← C*x_t;
    buildup_time ← t;
    prediction ← buildup_time + time_to_clear;
    return prediction
```

**Algorithm 7:** Prediction rule based on the proportionality assumption.

### Dynamic Trapezium

In the Section 3.3, we explained how we modelled a DPE as a trapezium. Here we use this model to predict the duration of a DPE. We assume that the DPE is shaped like an isosceles trapezium with parameters $a$, $b$, $c$ and $h$, where $a = c$ by the definition of an isosceles trapezium and $h$ heuristically set to 0.8 times the observed global maximum of the intensity (this is described in more detail in Section 3.3). Therefore, $a$ is set as the minimum time it takes to reach the height $h$ and $b$ is the time from that intersection until the current temporal point. Based on these assumptions, the total duration of a DPE event will be $a+b+c = 2a+b$. This corresponds to the prediction rule in Algorithm 8.

```
Function Predict(x_{1:t})
    mx ← max(x_{1:t});
    a ← min_{s∈{1,...,t}} s   s.t.   x_s ≥ 0.8mx;
    b ← t − a;
    prediction ← 2a + b;
    return prediction
```

**Algorithm 8:** Prediction rule based on modelling the DPE as a trapezium.

## 5.2 Linear Regression

This algorithm is based on the statistical class of linear regression models. The idea behind it is to find statistically significant relations between congestion-related observables and the duration of a DPE. Historical data is used to train the model, and the adjusted parameters are used for real-time forecasting while observing new data from the links.

### Linear Regression Model

As a first step towards predicting the return to profile of a DPE, we chose the Symmetry Factor $S$ (defined in Section 3.3) as the response variable of a linear regression model. This is a convenient observable, since it is independent of the scale of DPEs and can be used to infer their duration.

The predictors for the regression are derived from the features of the DPE time series, beginning with the time to maximum deviation $t_m$. We firstly postulate that there exists a linear relationship on the logarithmic scale between $S$ and $t_m$:

$$\log(S) = \beta_0 + \beta_1 \log(t_m) \tag{6}$$

where $\beta_0$ is an intercept, $\beta_1$ is the slope of the line. Although $S$ have been chosen as a response variable, prediction purposes demand information about the duration of the DPE $y$. In fact, it can be easily seen that $y$ is related to $S$ through $t_m$ and its complement, the time to return-to-profile after the global maximum, $t_r$. In particular, using the definition of $S$ given in Section 3.3:

$$y = t_m + t_r = t_m(1+S) \tag{7}$$

The data shows strong evidence for the relation in Equation 6; however, the variance is too large to be explained by a single regression. This variability can be properly addressed by considering multiple regressions through the use of a categorical variable. We found that the adjusted R-squared coefficient of the regression is significantly improved by considering the number of peaks $P \in \{1,2,\dots\}$ for each DPE as a factor in the regression. To simplify this, we discretise $P$, i.e. we divide the data into $k$ different bins (categories) with equal depth (frequency) based on $P$. Each category $j$ can be viewed as a regression on its own having a different intercept $\beta_0^j$ for $j \in \{1,\dots,k\}$ and slope $\beta_1$ common among all the considered categories. Hence, the linear relationship for category $j$ will be give by:

$$\log(S) = \beta_0^j + \beta_1 \log(t_m) \tag{8}$$

Here, $k$ is a parameter of the model. A large number of bins can produce over-fitting, whereas a small number can reduce the capacity of the model to explain the variance. We have both the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) [17] as heuristics to help us choose the value of $k$ that produces in the most parsimonious model. These criteria are information theoretic quantities that measure the relative quality of statistical models taking into account the trade-off between goodness-of-fit and model complexity. Note that other criteria might be used for this purpose, and the performance of the algorithm may change.

### Regression Results

We present the results of the regression on the data for k = 6 for the M6 and k = 7 for the M11. These are shown for the M6 and M11 separately: in Tables 1 and 2 one finds the least squares estimators for the linear coefficients of the model in Equation 6 complemented with the categorization of $P$, i.e. intercepts for each category $\beta_0^1,\dots,\beta_0^k$ and the common slope $\beta_1$; Figures 10 and 11 respectively represent the adjusted model for each considered category.

Note that Tables 1 and 2 also show information about the significance of each predictor and the quality of the fit. The p-values measure the significance of each of the predictors based on a $t$-test, where a low p-value implies a highly significant predictor. The adjusted R-squared score reports the proportion of the variability in the data that the model is able to capture.

Interestingly, the results also show that both motorways share a similar slope, i.e. the relation $\log(S) \simeq -1.3\log(t_m)$ holds

| Coefficients | Estimates | p-values |
|---|---|---|
| $\beta_1$ | -1.308 | $< 2e-16$ |
| $\beta_0^1$ | 3.515 | $< 2e-16$ |
| $\beta_0^2$ | 4.002 | 0.00561 |
| $\beta_0^3$ | 3.654 | 0.42141 |
| $\beta_0^4$ | 4.399 | $4.88e-06$ |
| $\beta_0^5$ | 4.598 | $2.55e-08$ |
| $\beta_0^6$ | 5.537 | $< 2e-16$ |
| Adjusted R-squared | 0.5958 | |

**Table 1.** Results for applying the linear regression model on M6 data.

| Coefficients | Estimates | p-values |
|---|---|---|
| $\beta_1$ | -1.337 | $< 2e-16$ |
| $\beta_0^1$ | 3.352 | $< 2e-16$ |
| $\beta_0^2$ | 3.831 | 0.00015 |
| $\beta_0^3$ | 4.090 | 9.94e-09 |
| $\beta_0^4$ | 4.595 | $< 2e-16$ |
| $\beta_0^5$ | 5.140 | $< 2e-16$ |
| $\beta_0^6$ | 5.580 | $< 2e-16$ |
| $\beta_0^7$ | 6.140 | $< 2e-16$ |
| Adjusted R-squared | 0.7963 | |

**Table 2.** Results for applying the linear regression model on M11 data.

in both cases. This suggest that it might be worth investigating the universality of such relationship using more data from the same and other motorways. If the mentioned universality is real, it is reasonable to assume that the lack of adjustment for the M6 in comparison to the M11 is due to the scarcity of data and not a consequence of a poorly chosen model. This would indicate that its performance would increase when incorporating more and more historical data.

### Linear Regression Algorithm

The previous analysis uses *a posteriori* information for DPEs to obtain the values of the predictors, i.e. the total number of peaks and the time to maximum. However such information only becomes known at the end of the DPE; therefore, in addition to the variance of the regression, there is some uncertainty that comes from estimating the values of the predictors , especially at the early stages of an event. Hence, the operational procedure for the algorithm is to simply update the values of the predictors at each time step during the event and recalculate $S$ accordingly. With new values of $t_m$, $P$ and $S$, one can produce a forecast using Equation 7.

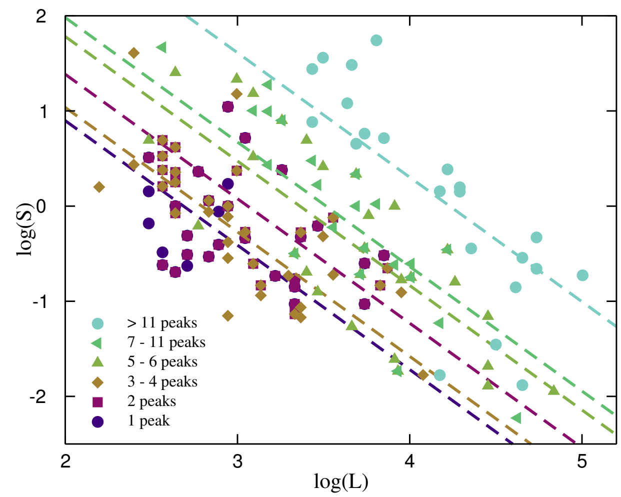



**Figure 10.** Fit of the linear regression on M6 data.

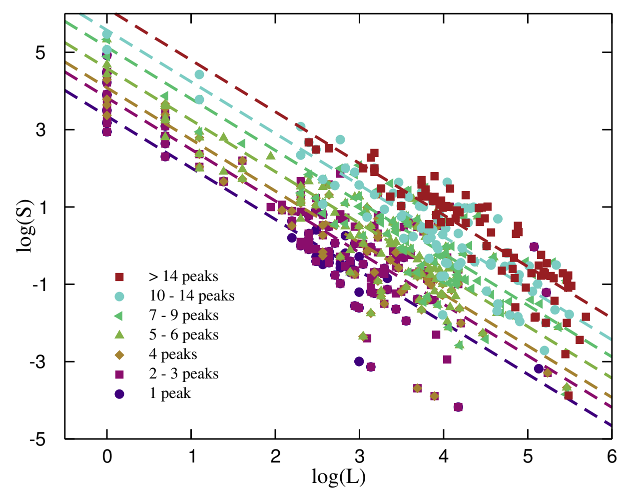



**Figure 11.** Fit of the linear regression on M11 data.

## 5.3 Weighted Multimodel

One can often improve the prediction power in a task by averaging predictions from an ensemble of models. Here, we propose a method to obtain a weighted average for the duration of events predictions of different algorithms. We define a weighted multimodel as the weighted sum of the predictions of its constituent models. Consider a set of $N$ predictive models that provide predictions at time $t$, $\tau_m(t) \in \{\tau_1(t), \ldots, \tau_N(t)\}$, the weighted multimodel $\tau(t)$ is given by:

$$\tau(t) = \sum_{m=1}^{N} w_m \tau_m(t) \tag{9}$$

where $w_m \in \{w_1, \ldots, w_N\}$ is the weight given to model $\tau_m(t)$. Note that $\sum_{m=1}^{N} w_m = 1$. The weights are hyperparameters of the multimodel and can be set either by experimentation or by derivation. We chose the latter as it provides a statistically sound explanation of the multimodel.

Assuming that among the set of $N$ model, there exists a true model that can predict the correct duration of the DPE. Let $\mathbb{P}(M=m)$ signify the probability that model $\tau_m(t)$ is the true model. Hence, the "expected" model can be given by:

$$\tau(t)=\mathbb{E}(\tau_m(t))=\sum_{m=1}^{N}\mathbb{P}(M=m)\tau_m(t) \tag{10}$$

Hence, comparing Equations 9 and 10 one can set $w_m=\mathbb{P}(M=m)$. Of course, $\mathbb{P}(M=m)$ is not known and not trivial to calculate; however, it can be estimated empirically from the data (see Appendix F for more details).

For the purpose of this project we chose a weighted multimodel consisting of three constituent models: the Midpoint Prediction, the Dynamic Trapezium and the Linear Model.

### 5.4 Benchmarking

To compare the performances of the collection of algorithms we developed, we applied them on the M6 and M11 testing datasets. As a measure of skill, we extended the scoring criterion presented in Equations 4 and 5 to allow for measuring the Average Prediction Error at any percentile of the DPEs. This is defined below:

$$E_p=\frac{1}{N}\sum_{i=1}^{N}E_p^{(i)} \tag{11}$$

where $E_p^{(i)}$ is the percentage error at the $p^{th}$ percentile of the duration of the $i^{th}$ DPE. $E_p^{(i)}$ is calculated as follows:

$$E_p^{(i)}=100\frac{|y^{(i)}-\hat{y}_p^{(i)}|}{y^{(i)}} \tag{12}$$

where $y^{(i)}$ is the true value of the duration of the $i^{th}$ DPE and $\hat{y}_p^{(i)}$ is the predicted value at the $p$ percentile of the $i^{th}$ DPE.

Figures 12 and 13 show the Average Prediction Error for values of $p$ between 0 and 100 for M6 and M11, respectively. The results in both figures show very similar tendencies for M6 and M11. In the following, we will comment quantitatively the results for M6:

- In the first 20% of the events duration, all algorithms have a very similar Average Prediction Error, which decreases at almost a constant rate from ∼100% to ∼70%. This result shows the difficulty to make an accurate prediction in the beginning of an event.
- After ∼20% of the event duration, the Null Model remained nearly constant at ∼68% error. This value sets a threshold from which it can be determined if an algorithm has prediction capabilities or is just comparable to a random guess.
- The Midpoint Prediction algorithm performed impeccably at the 50% of the DPEs. This is an obvious result, since every minute the algorithm always predicts the time to clear to be equal to the current time of the DPE – so at the 50% -the midpoint- it will always hit the target perfectly. However, it can be seen that the strength of this algorithm at the midpoint is its weakness near the starting or ending points of the DPE.
- The four other algorithms' performances had a similar tendency: a decreasing average prediction error from the beginning until ∼80% of the event duration followed by a small error increase until the end. It should be noted that the Dynamic Trapezium and the Weighted Multimodel performed much better than the Existing Model in every event duration percentile.
- In the inner plot of Global Error vs. Middle Inaccuracy, it can be seen which algorithms are best in overall terms (Global Error) or will minimize the Middle Inaccuracy. The Global Error is defined as the sample mean of the Average Prediction Error throughout all the duration of the events. The Middle Inaccuracy is defined as the percentage of predictions at the midpoint that had an error of more than 20%.
- The Midpoint Prediction algorithm is by definition the best algorithm regarding Middle Inaccuracy, achieving 0% in it. So this should be the algorithm of choice if the main concern is to score highly on the Average Percentage Error at the Midpoint. The Dynamic Trapezium and the Weighted Multimodel achieved the lowest Global Error. These two should be especially considered if the purpose is to obtain an overall good prediction throughout all the DPE.

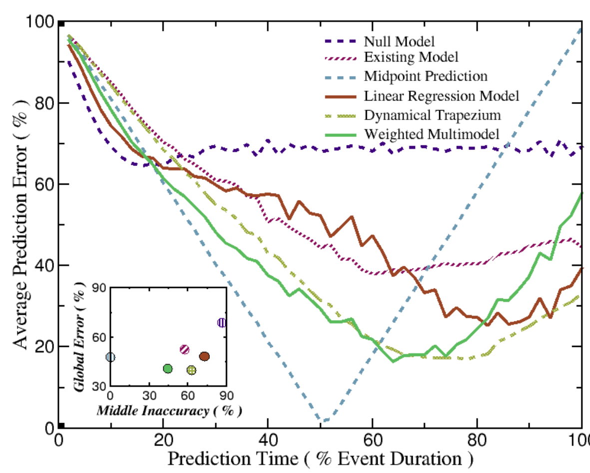



**Figure 12.** Results of the benchmark for the collection of the algorithms for M6

## 6. Discussion and Conclusion

Due to time constraints on the project, it was only possible to build small a database for the period between March and May 2016 for the M6 and M11 motorways. These motorways were selected due to their desirable qualities - fixed speed limits

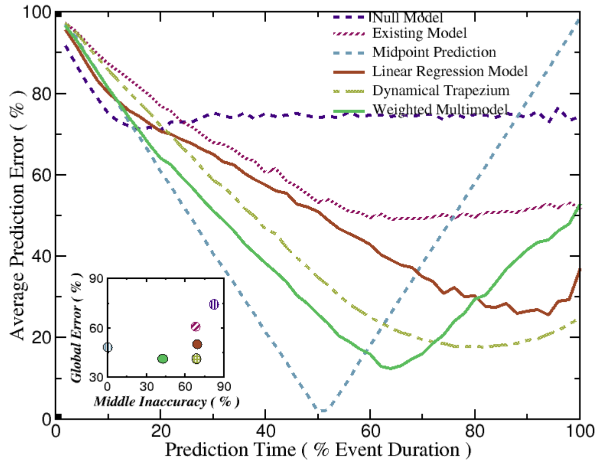



**Figure 13.** Results of the benchmark for the collection of the algorithms for M11

and low quantity of accidents. The fact that the database only covered a period of three months, did not allow for the possibility of studying seasonal tendencies. Yet, it was possible to extract valuable information from the data. A characterisation of the the nature of the DPEs was performed using exploratory data analysis; interestingly, power laws were found in the DPE duration and intensity statistics, despite the fact that the shapes of the DPE time series displayed very diverse dynamics.

In the quest of achieving good forecasting skill, statistical and heuristic methods were preferred over physical models, due to their computational simplicity and low reliance on physical principles. Five heuristic methods and one statistical method were proposed, along with a mixed multimodel. The Midpoint Prediction algorithm was able to achieve perfect accuracy by exploiting the way the scoring criterion was defined revealing the sub-optimality of this performance measure. The Linear Regression Model demonstrated a very good overall performance, but was beaten by the Dynamic Trapezium, showing the strength of heuristic methods. The Weighted Multimodel comprising of a weighted average of the predictions of a group of simpler models had the best overall performance.

It was established that there was a great room for improvement when it comes to the scoring system. It is suggested that a good scoring system evaluates the predictions throughout the whole event using different weights for each percentile. One idea is to make the weights proportional to the average intensity of the event at the given percentile of duration of a DPE.

An analysis of the Speed/Flow trajectories in the transition from a normal regime to maximal congestion (and vice-versa) was made. No cause or important relationship was found that could explain why trajectories were very similar in some cases, and very different in others. However, it was hypothesised that this behaviour could be explained by hysteresis in the system.

There is a wide spectrum of physical models and machine learning algorithms that can be applied in the prediction DPEs and traffic jams duration. The huge datasets available through the NTIS could provide a good playground for experimenting with these algorithms.

## Acknowledgments

We would like to thank Colm Connaughton for his guidance during this project. We would also like to extend our gratitude to Steve Hilditch and the rest of the Thales team for proposing such an interesting problem and providing support throughout this project's course.

## References

[1] A.J. Venables, J. Laird, and H. Overman. Transport investment and economic performance: Implications for project appraisal. Technical report, United Kingdom Department for Transport, October 2014.

[2] N. Peluffo. Strategic road network statistics. Technical report, United Kingdom Department for Transport, January 2015.

[3] P. Wockatz and P. Schartau. Traveller needs and UK capability study. Technical report, Transport Systems Catapult, October 2015.

[4] The Highways Agency. National Traffic Information Service Publish Services. Technical report, 2011.

[5] D. Chowdhury, L. Santen, and A Schadschneider. Statistical physics of vehicular traffic and some related systems. *Phys. Rep.*, 329:199, 2000.

[6] D. Helbing. Traffic and related self-driven many-particle systems. *Reviews of modern physics*, 73(4):1067, 2001.

[7] C. M. Dafermos. *Hyperbolic Conservation Laws in Continuum Physics*. Springer Berlin Heidelberg, 2010.

[8] M. Garavello and B. Piccoli. *Traffic Flow on Netwoks*. American Institute of Mathematical Sciences, 2006.

[9] M. J. Lighthill and G. B. Whitham. On kinematic waves. ii. a theory of traffic flow on long crowded roads. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 229, pages 317–345. The Royal Society, 1955.

[10] A. Kotsialos, M. Papageorgiou, C. Diakaki, Y. Pavlis, and F. Middelham. Traffic flow modeling of large-scale motorway networks using the macroscopic modeling tool metanet. *IEEE Transactions on intelligent transportation systems*, 3(4), 2002.

[11] M. McMullen. External inteface design document: DATD. Technical Report WA119-08-007-002-03-02-33 v1.0, Thales UK, March 2014.

[12] R. E. Wilson, Shrestha B. P., and Lees-Miller J. D. Prediction of time-to-clear: a report for the national traffic information service (ntis). Technical report, Transportation Research Group, University of Southampton, UK, January 2012.

[13] Private Communications - Thales Algorithm, 2016.

[14] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer-Verlag New York, 2009.

[15] M. Müller. *Information retrieval for music and motion*. Springer-Verlag Berlin Heidelberg, 2007.

[16] S. H. Strogatz. *Nonlinear Dynamics And Chaos*. Studies in Nonlinearity. Perseus Books Publishing, MA, USA, 1994.

[17] C. M. Bishop. *Pattern Recognition and Machine Learning*, volume 4. Springer, New York, 2006.

## 1. Data Acquisition

UK motorways incorporate thousands of induction loops and other types of sensors at different sites throughout the road network. These sensors report different measurements for all parts of the road network to a centralised system managed by the NTIS [4]. This data is made available publicly in the form of Daily Aggregated Traffic Data (DATD) publications. The DATD publications are published at the end of each day and contain multiple datasets pertaining to various measurements within the road network for that day. The DATD publications for the last three months are freely available to subscribers to the NTIS system and are provided in XML format. The specifications of the DATD publications are presented in Table 3.

## 2. Data Architecture

The NTIS employs a publish-subscribe messaging pattern in its system, whereby, the published data is made available on the internet. Subscribers to the NTIS system are able to receive the published data in real-time or in the form of DATD by sending the relevant request to the NTIS system via a web service (Figure 14) [11]. To this end, a subscriber system was built in Python. The subscriber system is of capable of interfacing to the NTIS web service, and contains code that requests and downloads the DATD publications. The specifications of the subscriber system are provided below:

- The subscriber system sends an HTTP GET request to the NTIS web service containing the relevant headers. If the request is successful, the module receives a response from the web service that contains the a zipped file of the DATD XML publications. The module is then able to download this file.
- The subscriber system is capable if unzipping the downloaded DATD publication and scanning the XML files for the relevant data.
- While scanning for the required data, the system is able to insert it into an SQL Relational Database.
- The system is run on two Linux virtual machine hosted on Microsoft's cloud service Azure, which also hosts the database.

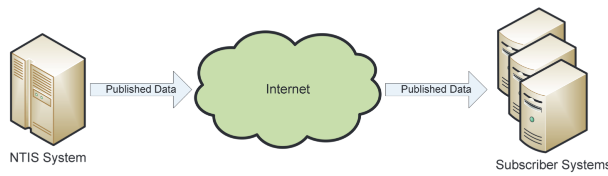



**Figure 14.** A simplified schematic representation of the physical interface of the NTIS system.

## 3. Finite Impulse Response Filter

To smooth the data, we chose a discrete time Finite Impulse Response filter. The output sequence of this filter is a weighted sum of the most recent input values. This is described in the following:

$$y_n = \sum_{i=0}^{N-1} b_i x_{n-i} \tag{13}$$

where $n$ is the current point in time, $N$ is the filter order and $b_i$ is the $i$th filter coefficient for $i \in 0,\dots,N-1$. Not that $\sum_{i=0}^{N-1} b_i = 1$. For this project, we selected a fourth-order low pass filter with the following coefficients:

$$\{0.5, 0.25, 0.125, 0.0625, 0.0625\}$$

The frequency response magnitude of this filter is shown in Figure 15.

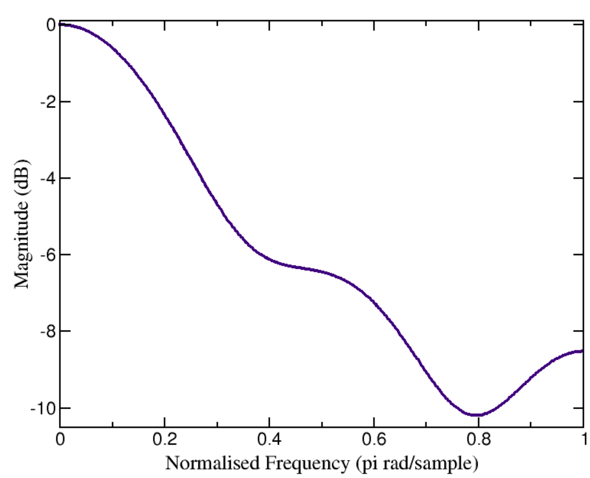



**Figure 15.** The frequency response magnitude of the chosen FIR filter.

## 4. Technical Description of the Scoring Criterion

Following the occurrence of a Deviation From Profile Event, an estimate shall be provided of when the relevant section of the NTIS Network shall return to Profile.

For a sample of $N$ DPEs, the following information shall be provided:

1. $Return\,to\,Profile\,Time_{ARP}$ $(RTPT_A)$: the time at which the point where the DPE occurs returns Traffic Flow values which are determined to be within the Profiled value.
2. $Event\,Mid\,Point$ $(EMP_A)$: defined as the time midway between the $Event_A$ Publication Time Stamp and the $Return\,to\,Profile\,Time_{ARP}$.
3. $Estimated\,Event\,Time_{MP}$ $(EET_{MP})$: defined as the estimated time to Return to Profile published at the Event Mid Point $(EMP_A)$.
4. $Actual\,Event\,Time_{MP}$ $(AET_{MP})$: defined as the difference between the $Return\,to\,Profile\,Time_{ARP}$ and the $Event_A$ Publication Time Stamp.

The percentage error on the return to profile is calculated as follows:

1. $EstError_A = 100\frac{AET_{MP}-EET_{MP}}{AET_{MP}}$ if $AET_{MP} \geq EET_{MP}$
2. $EstError_A = 100\frac{EET_{MP}-AET_{MP}}{AET_{MP}}$ if $AET_{MP} < EET_{MP}$

The Average Error on the Return to Profile is calculated using the following Performance Measure:

$$PM = \frac{1}{N}\sum_{A}^{N} EstError_A \tag{14}$$

## 5. Statistical Concepts

### Linear Regression

Given a sample of $n$ observations $\{Y_i, X_{i1}, \dots, X_{ip}\}_{i=1,\dots,n}$, composed of a response variable $Y_i$ and $p$ different independent variables or predictors $X_{ij}$, $j = 1, \dots, p$, the model is formulated as follows:

$$Y_i = \beta_0 + \beta_1 f_1(X_{i1}) + \dots + \beta_p f_p(X_{ip}) + \varepsilon_i \tag{15}$$

for $i \in \{1, \dots, n\}$, where $\varepsilon_i$ are the error terms modelling uncorrelated noise affecting the underlying system. The error terms are independently and identically distributed with $\varepsilon_i \sim N(0, \sigma^2)$, i.e. the errors are normally distributed with a constant variance. Note that $f_1, \dots, f_p$ might be non-linear functions of the predictors, i.e. the linearity of the model is referred to the relation between response and model parameters $\beta_0, \dots, \beta_p$[17].

### Information Criteria

As has been mentioned in the report, we have made use two information criteria for model selection purposes: Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Both can be derived from statistical principles and provide a measure of the trade-off between goodness of fit and model complexity.

In particular, AIC derives from Information Theory, and describes the process of loss of information when using a certain model to describe a particular dataset. Mathematically, its is defined as:

$$\mathrm{AIC} = 2k - 2\ln(L) \tag{16}$$

where $k$ is the number of fitted parameters and $L$ is the maximum value of the likelihood function of the model. Similarly, BIC derives from Bayesian arguments and is defined as:

$$\mathrm{BIC} = -2\ln(L) + k\ln(n) \tag{17}$$

where $L$ and $k$ are the same as before, and $n$ is the size of the dataset.

## 6. Weight Calculation for the Weighted Multimodel

Consider the weighted average at time $t$, $\tau(t)$, defined as:

$$\tau(t) = \mathbb{E}_m(\tau_m(t)) = \sum_{m=1}^{N}\sum_{q=1}^{100}\mathbb{P}(M = m, Q = q)\tau_m(t) \tag{18}$$

Where $Q$ is a random variable with support $\{1, 2, \dots, 100\}$ representing the percentile in a DPE, $M$ is a random variable with support $\{1, 2, \dots, N\}$ representing the true predictive model (algorithm) and $\tau_m(t)$ is the predicted time to return to profile at time $t$ from model $m$.

Using the law of conditional probability, one can write:

$$\mathbb{P}(M = m, Q = q) = \mathbb{P}(M = m|Q = q)\mathbb{P}(Q = q, t) \tag{19}$$

The probabilities on the right hand side are explained in words below:

**$\mathbb{P}(M = m|Q = q)$:** Probability that $m$ is the best model given that we are in the $q^{th}$ percentile of the DPE.

**$\mathbb{P}(Q = q, t)$:** Probability of being in the $q^{th}$ percentile at time $t$.

These probabilities can be calculated as follows:

$$\begin{aligned}\mathbb{P}(M = m|Q = q) &= \frac{\#\text{ times } m \text{ is best at } q}{\#\text{ DPEs}}\\ \mathbb{P}(Q = q, t) &= \frac{\#\text{ times we are in } q \text{ at time } t}{\#\text{ DPEs}}\end{aligned} \tag{20}$$

The intuition behind this ensemble is that during a DPE, at a certain time of day $t$, the proportion of the event that we are currently in is unknown. Furthermore, the best model for that point in the DPE is also unknown. Therefore we estimate a probability distribution for each of these unknowns, namely $\mathbb{P}(M = m|Q = q)$ and $\mathbb{P}(Q = q, t)$. We use these distributions to obtain the expected value of prediction from all $N$ models.

| **Published Data Type** | **Description** |
|---|---|
| ANPR Travel Times | Raw travel times, measured using number plate recognition between pairs of Automatic Number Plate Recognition (ANPR) camera sites. |
| Events | Events affecting the traffic status of the road network; manually created and modified by operators of the NTIS system or received from external systems. |
| MIDAS Loop Data | Traffic data, measured from roadside loop sensors monitored by Motorway Incident Detection and Automatic Signalling (MIDAS) Gold outstations. |
| TAME Loop Data | Traffic data, measured from roadside loop sensors monitored by Traffic Appraisal, Modelling and Economics (TAME) outstations. |
| TMU Loop Data | Traffic data, measured from roadside loop sensors monitored by Traffic Monitoring Units (TMU) outstations. |
| Processed Traffic Data – Fused FVD and Sensor Data | Traffic data, derived from fusing and processing raw Fused Vehicular Data (FVD) and sensor (MIDAS/TMU loop, ANPR) traffic data. |
| Processed Traffic Data – Fused Sensor-only Data | Traffic data, derived from fusing and processing raw sensor (MIDAS/TMU loop, ANPR) traffic data. |
| VMS and Matrix Signal Data | Variable Message Sign (VMS) and Matrix signal display and status information. |
| NTIS | The NTIS Network and Asset Model. The NTIS Model contains reference data that enables the traffic and event data included in the DATD to be mapped to the road network. |

**Table 3.** Data types included in the DATD [11]