# Estimating Baseline Travel Times for the Strategic Road Network

Alvaro Cabrejas Egea
*Centre for Complexity Science*
*University of Warwick*
Coventry, United Kingdom
a.cabrejas-egea@warwick.ac.uk

Peter de Ford
*McKinsey\**
San Jose, Costa Rica
pdeford@outlook.com

Colm Connaughton
*Centre for Complexity Science*
*University of Warwick*
Coventry, United Kingdom
c.p.connaughton@warwick.ac.uk

*Abstract*—Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque eget orci quis neque efficitur fringilla in vitae nisl. Nam mollis pellentesque arcu. Donec laoreet sapien in dapibus semper. Mauris vel malesuada libero. Morbi cursus erat ac arcu mattis, id venenatis nibh commodo. Donec dui eros, volutpat et dignissim sit amet, vehicula sit amet urna. Fusce pharetra mi enim, vitae dignissim felis pharetra id. Mauris quis purus scelerisque, mollis diam sit amet, aliquet urna. Nullam a aliquet ex. Integer in nisi eget leo sodales iaculis quis eget leo. Vestibulum ornare ante ut sapien pretium, ac cursus magna blandit. Donec semper, nisi id luctus porttitor, purus nisi varius diam, quis porttitor justo magna id neque. Aliquam bibendum non purus sit amet lobortis. Nullam id pellentesque orci, vitae vestibulum velit.

Maecenas vulputate iaculis erat. Pellentesque blandit nisl at felis finibus, sed mollis ipsum varius. Pellentesque eros mi, eleifend sit amet dui eget, bibendum ultricies nisi. Ut non dui in nulla dictum finibus at et erat. Vivamus ultrices, arcu quis pulvinar pellentesque, sapien sapien rutrum ex, ac gravida erat nibh a mi. Nam ut.

*Index Terms*—Traffic, Congestion, Intelligent Algorithms

## I. Introduction

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.

Here, we investigate novel methods for generating profiles that do not require segmentation or an underlying model, and that can be computed by only using statistical and spectral analysis of previous data.

## II. Data

The UK National Traffic Information System (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 and an edge is placed following the motorway. The basic building blocks of the system are the so-called links. They are segments of these edges, with a length varying between 500 and 2500 meters, representing of carriageways in the edges between nodes connected by a road. The links on the network incorporate thousands of induction loops (sensors) at different sites, which report speed, flow and other data to a centralised system.

NTIS data is being used to assign a traffic profile to each link on the network. The profile gives the expected travel time at any time of day for the different sections of the Highway network, understood as the average time to transverse the links for the vehicles currently in it. These profiles use analysis of previous travel times within the same section of road to provide forecasts of expected future travel times and state of traffic.

Operating modern Smart Motorways in the UK revolves around the use of these traffic profiles, operators rely on them to identify deviations, take corrective actions and display accurate information on the overhead signs. Better profiles means more accuracy for the whole system.

Currently, generating these profiles makes use of a combination of an Exponentially Weighted Moving Average (EWMA) and heavy time-dependent segmentation.

### A. Data Selection

- Link-level data was extracted for the sections of M6, M11 and M25 shown in Figure **??**

- The dataset for M6 and M11 comprises 90 days of data (07/03/2016-05/06/2016).
- The dataset for M25 comprises 75 days of data (07/04/2017-20/06/2017).
- Links with more than 10% of missing data were discarded.

The first two motorways were selected due to their combination of recurrent and outstanding congestion. The M25 was selected on the base that it is the most used Motorway in England on a daily basis, suffering from chronic congestion.

### B. Data Contents

For each link on a specific date, the data consists of one entry per minute, containing, amongst others, 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.

## III. EXAMPLES OF TRAVEL TIMES

The most direct variable for measuring the state of traffic over a length of road is the vehicles' travel times. The travel time for a given segment of road as the average time that the vehicles currently in it are taking since they enter the segment until they exit it.

As we can see in Figure 1 the travel time remains within a vaguely predictable pattern most of the days. We will find very low highway travel times during the nights, mostly being equal to the free flow time, it will meaningfully rise as people leave to work and the add load to the motorways. This collective behaviour will create the morning traffic jams. As the morning rush passes, the travel times settle to a plateau that lasts until late afternoon when we can see the evening rush, which potentially, albeit less likely, can also create jams. After this, the travel times slowly decay until we are again in the night free-flow regime.
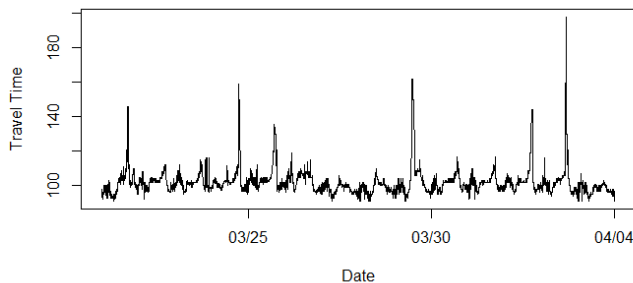


Fig. 1. Example of travel times on a link in the M6 over a period of three weeks.

However, as it also follows from Figure 1, we can see that there are a series of excursions out of this otherwise oscillating yet bounded typical travel time. In these occasions the travel time can increase several times fold the usual values.

## IV. CURRENT TRAFFIC PROFILES?

One of the main current approaches for highway operators is to generate profiles for each section of road. These profiles indicate what travel times can be expected in a given time and place.

### A. Use of Exponentially Weighted Moving Average

One currently used way for creating these Traffic Profiles consists in applying an Exponentially Weighted Moving Average (EWMA) on the same minute of every day, with the implicit assumption that we can expect similar road behaviour at the same time of the day. Whatever changes happen for a given time, are smooth enough so their trend will be appropriately approximated by the EWMA. In these algorithms we find that the travel time forecast $F(i, d)$ on the $i - th$ minute of a given day $d$ will be:

$$F_i^d = \alpha * t_i^{d-1} + (1 - \alpha) * F_i^{d-1} \qquad (1)$$

The main problem with using an EWMA is the manner in which the memory decays. Since more recent measurements are weighted more heavily than events in the past, if an extreme fluctuation occurs, the following forecasts will be biased, partially replicating this event and over-estimating the travel times in that section of road until enough new measurements have been taken to dissipate this effect.

### B. Segmentation

In addition, to acknowledge the specific differences between some special dates in the year, and the difference between days of the week, this family of methods require the use of heavy date segmentation. This segmentation means that the EWMA will be applied across dates which fall in the same category (i.e. Mondays, weekends, Christmas Day, ...) If we combine this with the shortcomings presented in IV-A, we find that some long reaching effects are generated, which can propagate for weeks into the future predictions, but do not have any reflection on the observed travel times, as it can be observed in Figure 2. Furthermore, in order to generate this segmentation in a valid manner, an experienced team is necessary, since the needs for segmentation can geographically vary. This dependence on experienced teams can lead to the creation of legacy systems which may not be well understood, decreasing their usability over time unless extra effort is put into transmitting this knowledge and continually training new staff.

## V. ALGORITHM

Given the cyclic nature of traffic, we aimed for a prediction algorithm that could account for the periodic variations while being resilient to fluctuations. This algorithm also should:
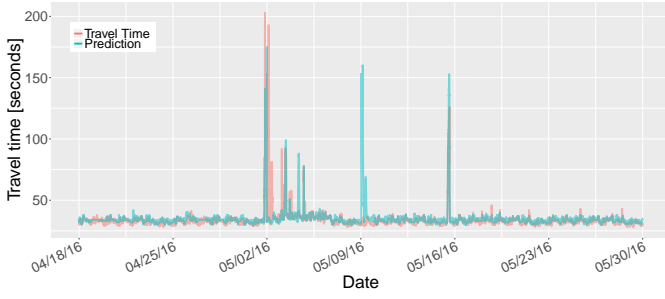
Fig. 2. Example of the current algorithm predicting increases in travel times for several weeks after a peak is seen in the measured travel times (25/04/2016)

- Mitigate the propagation of isolated events into future forecasts.
- Not require the use of time segmentation (experience) and be valid for regular and "special dates".
- Be location agnostic, the internal parameters should be set algorithmically based on the data.
- Have Gaussian, mean 0, uncorrelated residuals.
- Near flat Trend term, given the different time scales between seasonal cycles and changes in the general motorway flow.

### A. Null Model: Naive Segmentation

To obtain an accurate indication of the performance of the algorithm developed in the previous subsections, an example of basic segmentation was coded. This involved a weighted combination of the training data points using uniform weights. In this way, for the $i - th$ minute of the week and using a training set composed of the previous of $n$ weeks, the Null profile is:

$$Null(i, n) = \sum_{week=1}^{n} \frac{x_n^i}{n} \qquad (2)$$

### B. Decomposition in Background and Spikes

During the exploratory data analysis we found that, from the point of view of travel times, traffic operates in two clearly differentiated regimes.

- Background:
  - Stable around a mean value.
  - Small amplitude and high frequency.
  - Suitable for spectral filtering.
- Spikes:
  - Zero most of the time. Quickly go to extreme values.
  - Great amplitude and low inter-oscillation frequency.
  - Suitable for seasonal decomposition.

In order to stop the large deviations from affecting the underlying true profile, it was decided to separate these spikes from the rest and treat them in a separate manner.

$$x_t = Background_t + Spikes_t + \xi \qquad (3)$$

The objective is to separate them in such a way that the recurring congestion is captured as part of the Background

and used for spectral analysis, trying to mitigate the prediction error induced by the high frequency oscillations and obtaining a basic view of what can be daily observed. Meanwhile, the spikes, including the recurring component not captured in the background and the non-recurring congestion, are treated separately in an attempt to find seasonality in larger time scales than those in which the background oscillates. Ideally, after this step, an operator should only find isolated large events deviating from the profile and white noise.

To prevent the differing lengths of the links from affecting this decomposition, all travel times were normalized according to their corresponding link's free flow time, understood as the time to transverse the length of the link at the maximum speed allowed by the motorway.

A threshold was heuristically chosen to separate the two components in the different regimes. Intuitively, this threshold scales with the amount of recurring congestion in a link. Whenever a data point is above the threshold, it is flagged as belonging to a spike. For this purpose, we define an indicator function:

$$\delta_{spike} = \begin{cases} 1 & : \text{Over threshold} \\ 0 & : \qquad \text{Otherwise} \end{cases} \qquad (4)$$

TABLE I
NORMALISED TRAVEL TIME THRESHOLDS

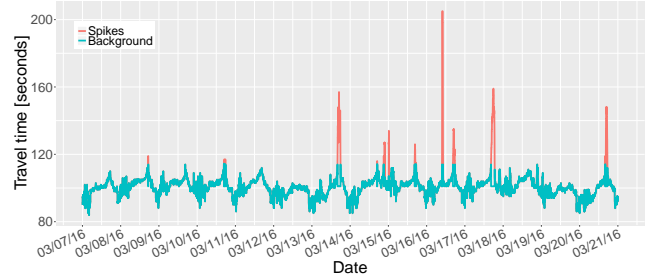| Motorway | Threshold |
|----------|-----------|
| M6       | 1.1       |
| M11      | 1.2       |
| M25      | 1.4       |



Fig. 3. Decomposition of time series in background and spikes

### C. Spectral Component

The main difficulty when dealing with the Background time series is the low amplitude, high frequency fluctuations that can be found almost ubiquitously. In these situations signal smoothing can be easily performed with removing completely the frequency components from a certain frequency and up while the information bearing low frequency components are retained. For this tast, the Fast Fourier Transform (FFT) [7] was used.

Below, the sequence of steps taken to process the Background part of the measured travel times is described

1) Calculate Background Power Spectrum using FFT
2) Remove frequencies corresponding to periods under 4 hours and over 1 week
3) Repeat for all $n$ weeks in training set
4) Apply and EWMA to the weekly predictions
5) Compute the Inverse Transform

### D. Seasonality Component

Seasonal Decomposition based on LOESS (STL) [5] was chosen for the seasonality analysis since "STL has several advantages over the classical decomposition method and X-12-ARIMA: Unlike X-12-ARIMA, STL will handle any type of seasonality, not only monthly and quarterly data. The seasonal component is allowed to change over time, and the rate of change can be controlled by the user. The smoothness of the trend-cycle can also be controlled by the user. It can be robust to outliers (i.e., the user can specify a robust decomposition). So occasional unusual observations will not affect the estimates of the trend-cycle and seasonal components" [6].

Below, the sequence of steps taken to process the Seasonality Component part of the measured travel times is described. Note that this should be applied to the $n$ training weeks as a single time series.

1) STL Decomposition of Background for daily seasonality
2) Extract and sum the series corresponding Trend and Remainder from 1)
3) STL decomposition for weekly seasonality
4) Comprobation of Background's Remainder and Trend from 3)
   a) Background's Remainder should be zero mean, Gaussian distributed
   b) Background's Trend should have a near zero slope
5) Add daily and weekly Seasonal components from 1) and 3)
6) Average seasonality across training weeks
7) Linearise Trend term from
   a) Select extreme values in the series and create a linear interpolation between them
   b) Extract the trend corresponding to the number of weeks for forecasting
8) Add linearised trend to seasonality obtained in 6)
9) STL Decomposition of Spikes for weekly seasonality
10) Comprobation of Spike's remainder and trend
    a) Spike's Remainder should be zero mean, Gaussian distributed
    b) Spike's Trend should have a near zero slope
11) Extract Spike's Seasonality corresponding to the number of weeks for forecasting
12) Add Spike's Seasonal component to the output of 8) to output from 9)

### E. Profile Generation

In order to create the final hybrid profile, we take one of the forecasts generated in the previous points depending on what is the identified regime, as described in

$$Hybrid = Seasonal * \delta_{spike} + Spectral * (1 - \delta_{spike}) \quad (5)$$

Where $\delta_{spike}$ is the indicator function defined in Equation **??**.
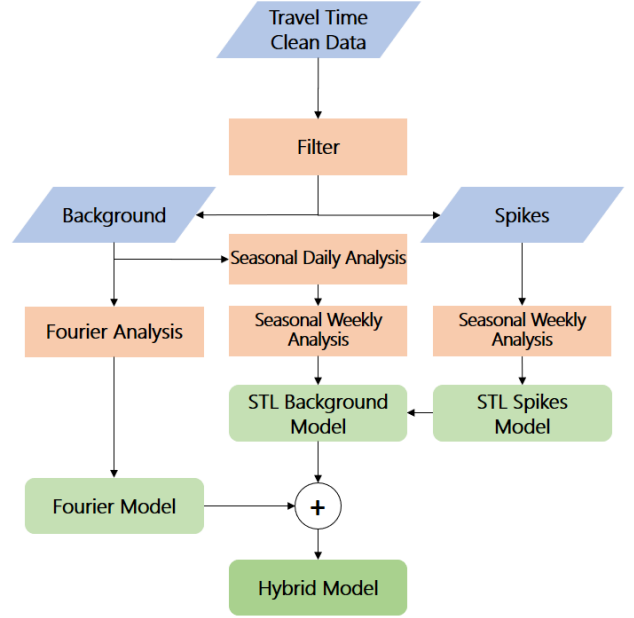


Fig. 4. Schematic of data streams in the algorithm.

## VI. ACCURACY RESULTS

In this section the accuracy of the algorithm described above is compared against the Published Profiles and the Null Model. The error is quantified as the average distance between the prediction and the measured travel times across all analysed links.

$$\text{Error} = \left( \sum_{\text{Link}=1}^{n} \frac{\|\text{Measurement} - \text{Prediction}\|}{\text{Measurement}} \right) / n \quad (6)$$

### A. Accuracy by Quantile

In this subsection, the accuracy of the algorithm is compared against the Published Profiles and the Null Profile across all percentiles of travel time. As it can be seen in Figures 5, 6 and 7, the profile generated by combining the Seasonal and Spectral analysis has a higher accuracy than the Published Profiles and the Null Model for all percentiles of travel time except for the most extreme values where they all perform poorly. The most relevant difference occurs between percentiles $[60 - 95]$, where the Published Profile starts to suffer from higher inaccuracy, due to some extreme events in the shorter links. In all cases, the Hybrid model is the one that requires greater travel times to reach a $20\%$ error in the prediction.

Fig. 5. Average accuracy results in M6 across percentiles of travel time.
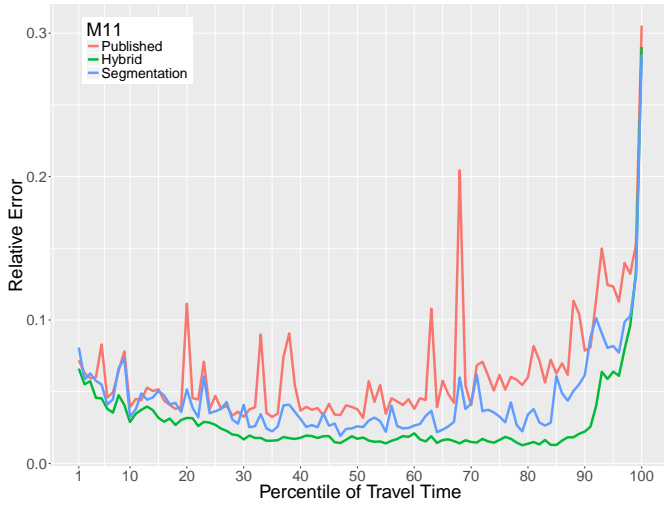


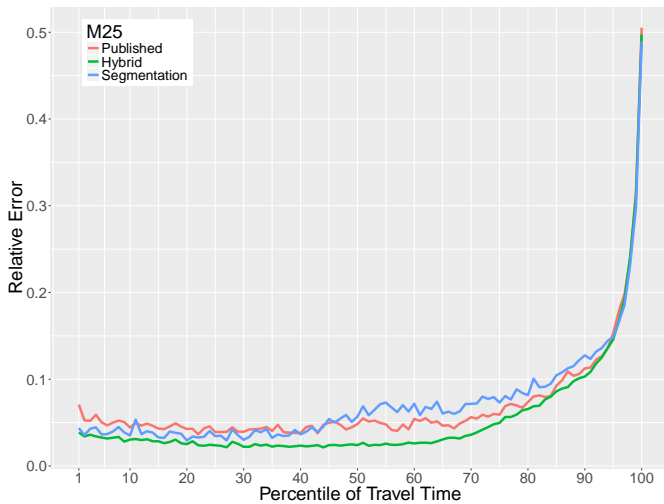Fig. 6. Link average accuracy results in M11 across percentiles of travel time.



Fig. 7. Link average accuracy results in M25 across percentiles of travel time.

## B. Daytime Error

In this subsection, the accuracy of the algorithm is compared against the Published Profiles and the Null Profile across the times of the day. As it can be seen in Figures 8, 9 and 10, the profile generated by combining the Seasonal and Spectral analysis has a higher accuracy than the Published Profiles and the Null Model for all times of the day. The most relevant improvement occurs during the morning and evening peak hours, where the algorithm presented in this paper does not accuse meaningful performance worsening relative to the morning plateau when compared with the other two profiles. In the case of the M6 and M11, where the training set is richer, the error at peak times is reduced by at least 50% in all cases, reaching as much as 68.7% in the case of the M6 morning rush. In the case of the M25, which is congested on a regular basis, the errors in the Published Profile during peak times are slightly lower than on the other cases, indicating that, given the methods used to calculate the Published Profile, this recurrent congestion is better captured by it, without the congestion being lighter in intensity than in the cases for the M6 and M11. Even in this case, the proposed algorithm performs significantly better than any of the other two, except for a brief window between 6-7AM when it is outperformed by the Null Model.
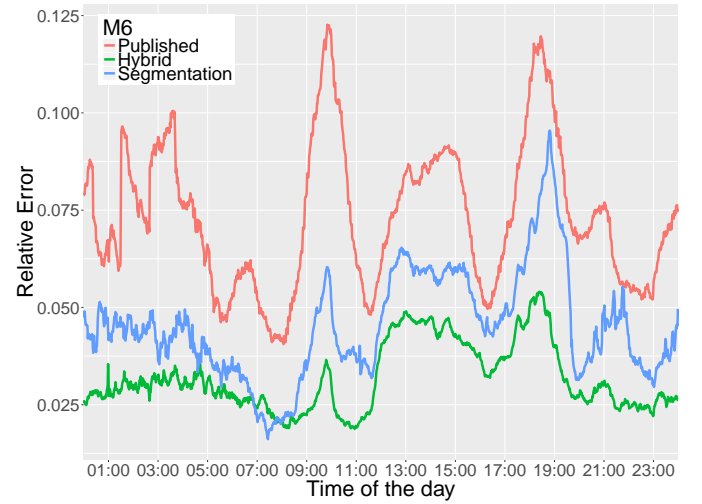


Fig. 8. Link average accuracy results in M6 across times of the day.

## C. Distribution of Errors

HERE BE MULTIPLOT(3) HISTOGRAM WITH THE DISTRIBUTION OF ERRORS, ONLY SHOW SHAPE

## VII. FUTURE WORK

The algorithm supplied above meets all the requirements described in Section V except for the fact that it requires a heuristically set threshold. One potential way of reaching total compliance with these requirements is to perform the decomposition between Background and Spikes by applying a Wavelet Transform. In the transformed time series, it is possible to see which time scales are more affected by fluctuations
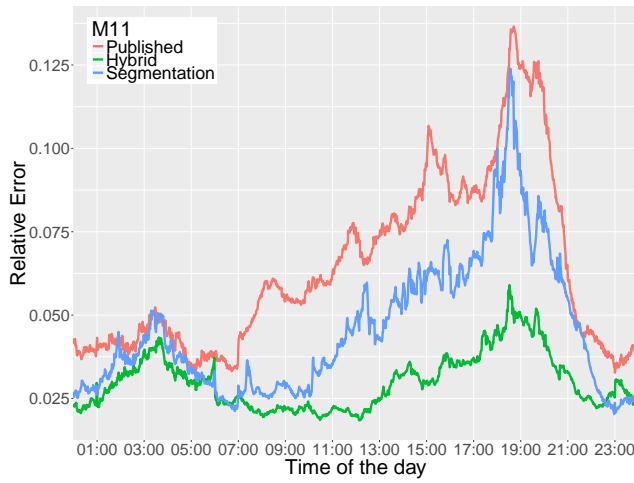
Fig. 9. Link average accuracy results in M6 across times of the day.
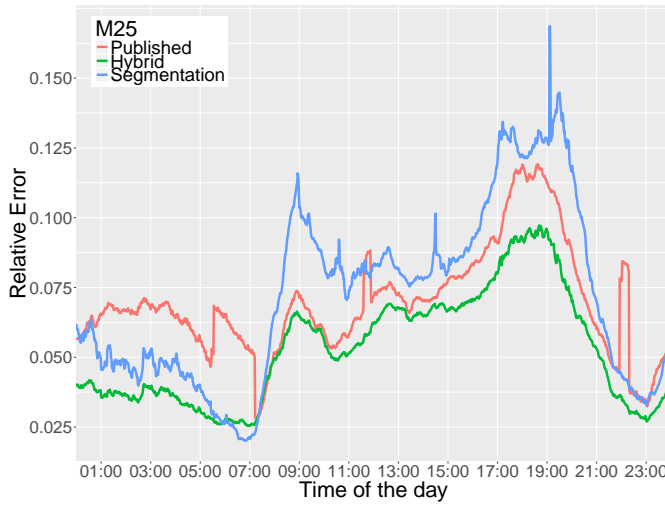


Fig. 10. Link average accuracy results in M25 across times of the day.

and rare events while these are ongoing. Then, the Wavelet and Scale coefficients for those combinations of time and scale can be set to mirror those where undisturbed background can be found and the difference between these newly adapted values and the original ones would conform the Spikes series. In the future a sensitivity analysis should be conducted to explore the limits of the algorithm in terms of minimum training data set, as well as maximum performance with increased training.

### 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 Transport Information System Publish Services", Technical Report, 2011.
[5] R. B. Cleveland, W. S. Cleveland, J. E. McRae and I. Terpenning, "STL: A Seasonal-Trend Decomposition Procedure Based on Loess", Journal of Official Statistics, vol. 6, 1990, pp. 3-73.
[6] R. J. Hyndman and G. Athanasopoulos, "Forecasting: Principles and Practice", Otexts, 2013.
[7] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series", Mathematics of computation, 1965, vol. 19, no 90, p. 297-301.