Estimating Baseline Travel Times for the UK Strategic Road Network

Alvaro Cabrejas Egea
MathSys Centre for Doctoral Training
University of Warwick
Coventry, United Kingdom
a.cabrejas-egea@warwick.ac.uk

Peter De Ford
McKinsey & Company
San Jose, Costa Rica
pdeford@outlook.com

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

Abstract—We present a new method for long-term estimation of the expected travel time for links on highways and their variation with time. The approach is based on a time series analysis of travel time data from the UK's National Traffic Information Service (NTIS). Time series of travel times are characterised by a noisy background variation exhibiting the expected daily and weekly patterns punctuated by large spikes associated with congestion events. Some of these spikes are caused by peak hour congestion and some are caused by unforeseen events like accidents. Our algorithm uses thresholding to split the time series into background and spike signals, each of which is analysed separately. The the background signal is extracted using spectral filtering. The periodic part of the spike signal is extracted using locally weighted regression (LWR). The final estimated travel time is obtained by recombining these two. We assess our method by cross-validating in several UK motorways. We use 8 weeks of training data and calculate the mean absolute percentage error (MAPE) of the resulting travel time estimates for week of test data, repeating this process 4 times. We find that the MAPE is significantly reduced compared to estimates obtained by simple segmentation of the data and compared to the estimates published by the NTIS system.

Index Terms—Travel Time, Traffic Congestion, Prediction

I. INTRODUCTION

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) [1]. Highways England collects data of the speed, flow, travel time and other data 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, novel methods for generating profiles are investigated, such that they do not require segmentation, clustering or an underlying model, and that can be computed by only using statistical and spectral analysis of previous data.

The available literature on travel times is extensive but most recent research in focuses on short-term forecasting *** HERE BE LIT REVIEW***

Most of these methods have been specifically tuned for the conditions on which they have been developed and transferability to other sites is often not evaluated. To test the transferability of this method, it is tested on three different motorways.

II. DATA

The basic building blocks of the Strategic Road Network are the so-called links. They are segments of motorway, with a length varying between 500 and over 6000 meters. The links on the network incorporate thousands of induction loops (sensors) at different sites, which report their measurements to a centralised system.

NTIS data is being used to assign a traffic profile to each link on the network. A link's profile contains the expected travel time at any given time of day, understood as the average time to transverse said link. These profiles use historical travel time data within the same link to provide expected future travel times.

Currently, generating these profiles makes use of a combination of an Exponentially Weighted Moving Average (EWMA) and heavy time-dependent clustering, although the exact methodology is not public.

A. Data Selection

- Link-level data was extracted for sections of the M6, M11 and M25 motorways in England.
- The dataset for M6 comprises 90 days of data (07/03/2016-05/06/2016) across 14 links.
- The dataset for d M11 comprises 90 days of data (07/03/2016-05/06/2016) across 25 links.
- The dataset for M25 comprises 75 days of data (07/04/2017-20/06/2017) across 61 links.
- Links with more than 10% of missing data were discarded.
- Whenever missing data was detected for 10 or less minutes, it was linearly interpolated.
- Whenever missing data was detected for over 10 minutes, it was left as missing values.

The first two motorways were selected due to their high usage and 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:

- Traffic speed
- Traffic flow
- Traffic headway.
- Current travel time
- Profile travel time
- Free flow travel time
- Event flags for:
  - Spontaneous congestion events
  - Weather events
  - Other unspecified types of events.

C. Training and Testing Data

In the case of the M6 and M11, the data comprises 12 complete weeks of travel time. The previous 8 weeks are used to predict the following 4. However, in an attempt to more closely simulate real operation, it was assumed that new measurements were received at the end of every week, allowing for an updated prediction using the new measurements in addition to the last 7 of the original 8 weeks used for training, effectively working on a rolling basis.
In the case of the M25, the dataset comprised 9 complete weeks. Hence, 6 weeks of data were used to predict the following 3 in the same manner as described in the previous paragraph.

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 is the average time that the vehicles currently in it are taking since they enter the segment until they exit it.
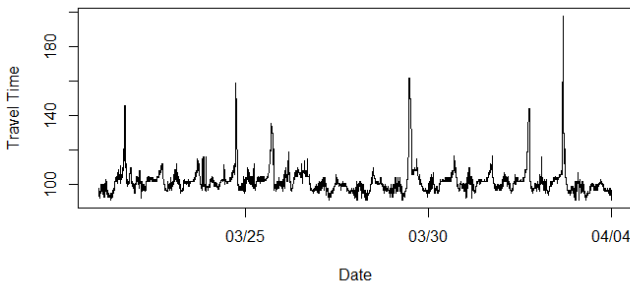

Fig. 1. Travel times on link 117007401 in the M6 over three weeks.

As it can be seen in Figure 3 the travel time remains within a vaguely predictable pattern most of the days with bounded minima corresponding to the free flow time, although there are some outliers below this value corresponding to speeding drivers. Very low highway travel times will be found during the nights, mostly being equal to the free flow time, these 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 the evening rush can be observed, which potentially, albeit less likely, can also create jams. After this, the travel times slowly decay towards the night period of free-flow regime.

However, as it also follows from Figure 3, it can be seen 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. These extreme oscillations are much less predictable both in intensity and inter-oscillation period than the recurrent congestion described previously.

IV. CURRENT TRAFFIC PROFILES

In this section, current approaches towards the generation of Traffic Profiles will be described, in line with how these are calculated in the United Kingdom.

A. Use of Exponentially Weighted Moving Average

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

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

The main problem found when 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 in the form of a traffic jam, 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 requires the use of heavy date based 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 this is combined with the shortcomings presented in IV-A, it is found 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
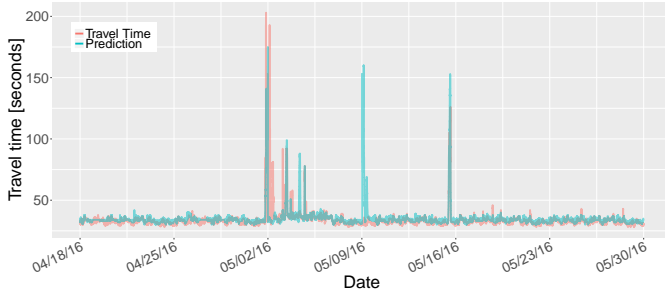
Fig. 2. Example of the current algorithm predicting increases in travel times for several weeks after a peak is seen in the measured travel times (25/04/2016)

necessary, since the needs for segmentation can geographically vary, given that the EWMA approach tends to better approximate endemic congestion. This dependence on experienced teams can lead to the creation of legacy systems which may not be well understood after a few years, decreasing their usability over time unless extra effort is put into transmitting this knowledge and continually training new staff.
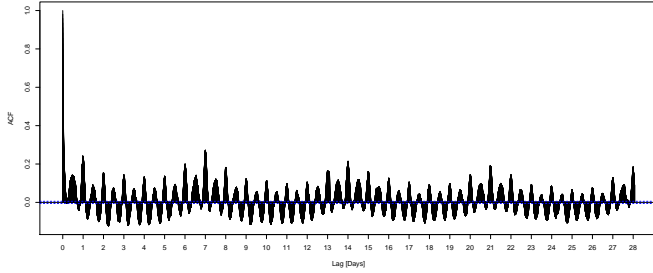
V. ALGORITHM

Fig. 3. Autocorrelation function of a link in the M6 over a period of 4 weeks, showing seasonal patterns on the daily and weekly periods.

Given the cyclic nature of traffic, the aim was a prediction algorithm that could account for the periodic variations and endemic congestion while being resilient to fluctuations and rare events. This algorithm also should:

- 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. Naive Segmentation

To obtain an accurate indication of the performance of the algorithm developed in the following 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 a week and using a training set composed of the previous of n weeks, the Naive Segmentation (NS) profile is:

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

B. Decomposition in Background and Spikes

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

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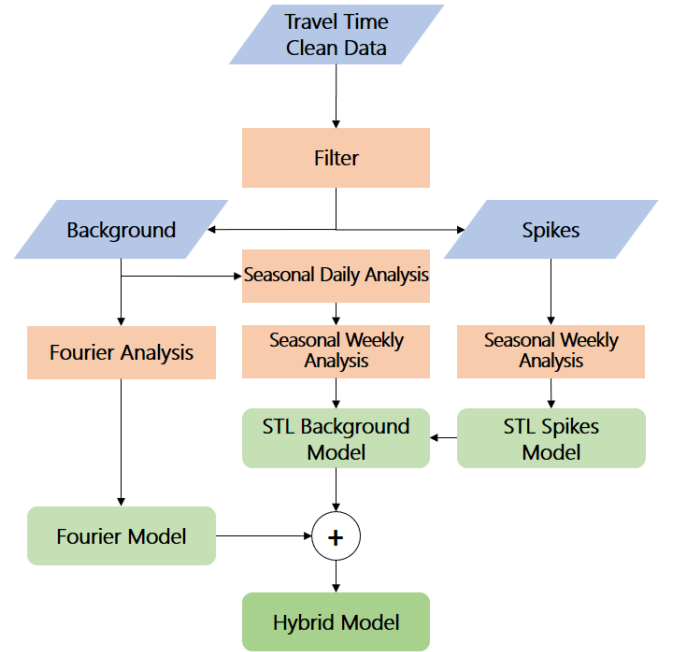 - Suitable for seasonal decomposition.

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

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. Assuming Gaussian noise ξ:

$$Travel\ Time_t = Background_t + Spikes_t + \xi \qquad (3)$$

The objective was to separate them in such a way that most of the recurring congestion is captured as part of the Background and used for spectral analysis, attempting 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, were 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 rare events deviating from the profile and white noise.

To prevent the differing lengths of the links from affecting this decomposition, for this step, 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 legal speed allowed by the motorway. However, this step only mitigates the non-regularity of the time series, since there are drivers who do not follow these limits.

A threshold was heuristically chosen to separate the two components in the different regimes from the normalised travel times. 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.
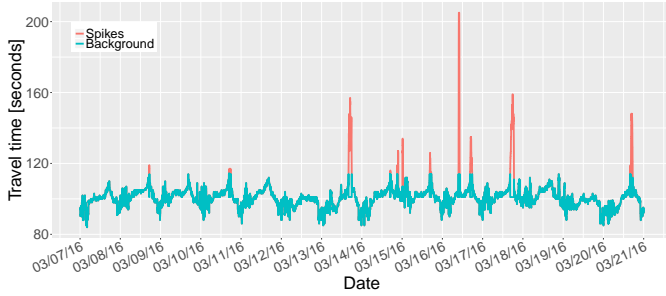

Fig. 5. Decomposition of time series in background and spikes

For this purpose, an indicator function was defined, taking for every minute the value:

$$\delta_t^{spike} = \left\{ \begin{array}{ll} 1 & : \text{Over threshold} \\ 0 & : \qquad \text{Otherwise} \end{array} \right. \qquad (4)$$

TABLE I
NORMALISED TRAVEL TIME THRESHOLDS

Motorway	Threshold
M6	1.1
M11	1.2
M25	1.4

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 by removing completely the frequency components from a certain frequency and up

while the information bearing low frequency components are retained. For this task, the Fast Fourier Transform (FFT) [4] 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-Trend Decomposition based on LOESS (STL) [2] 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" [3].

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 Step 1
3) STL decomposition for weekly seasonality
4) Comprobation of Background's Remainder and Trend from Step 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 Step 1 and Step 3 to obtain global seasonality
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 Step 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. Seasonal-Spectral Hybrid Profile

In order to create the final Seasonal-Spectral Hybrid profile, referred to as Hybrid in the following sections, one of the two forecasts generated in the previous points is taken, depending on what is the identified regime, as described in

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

Where δ_{spike} is the indicator function defined in Equation (4).

VI. ACCURACY RESULTS

In this section the accuracy of the algorithm described above is compared against the Published Profiles and the NS Model. For each temporal point i, the Mean Average Percentage Error (MAPE) is defined below:

$$MAPE = \frac{\left(\sum_{i=1}^{n} \frac{\|x_i - \hat{x}_i\|}{x_i} \right)}{n} \quad (6)$$

In the case of the Motorways the Root Mean Squared Error (RMSE) has been calculated for each temporal point i as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2}{n}} \quad (7)$$

In both cases the error for a Link is defined as the average MAPE or RMSE across all prediction points. The error for a Motorway is defined as the average of the error across all its links.

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 6, 7 and 8, the profile generated by combining the Seasonal and Spectral analysis has a higher accuracy than the Published Profiles and the NS 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.

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 9, 10 and 11, the Hybrid 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, all locations and training lengths. The most relevant improvement occurs during the morning and evening peak hours, where the algorithm presented in this paper does not accuse meaningful
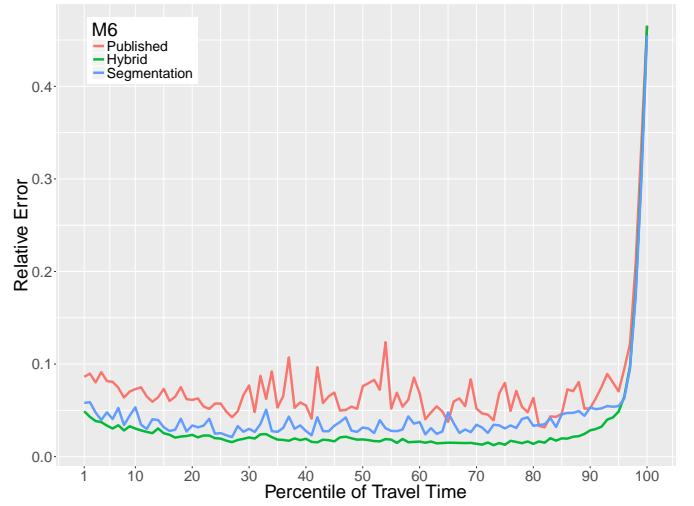

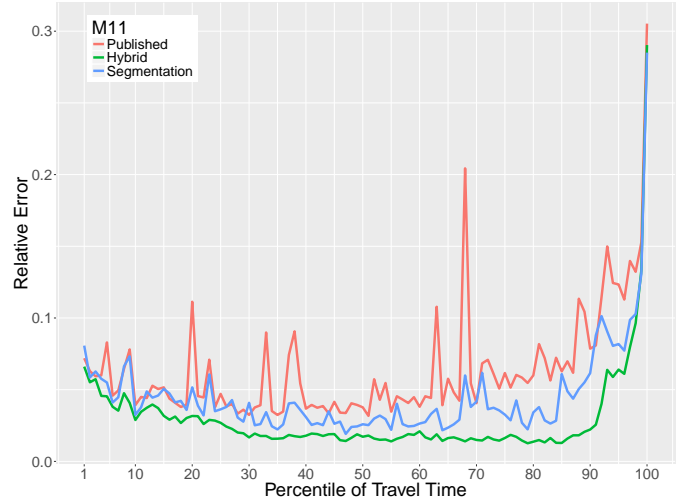Fig. 6. Average accuracy results in M6 across percentiles of travel time.

Fig. 7. Link average accuracy results in M11 across percentiles of travel time.
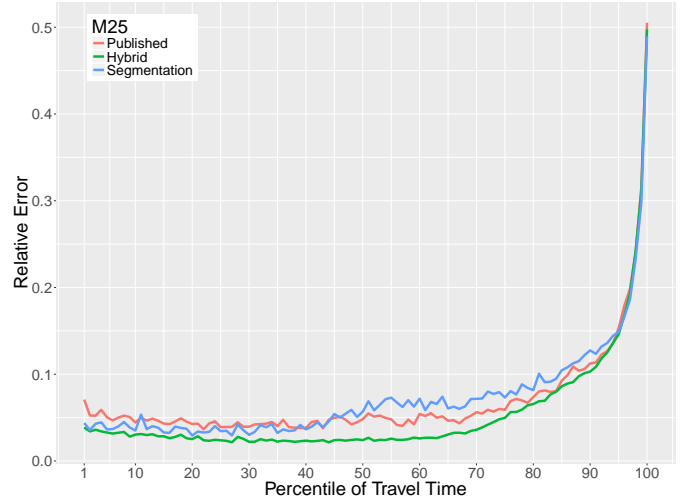

Fig. 8. Link average accuracy results in M25 across percentiles of travel time.

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 NS Model.
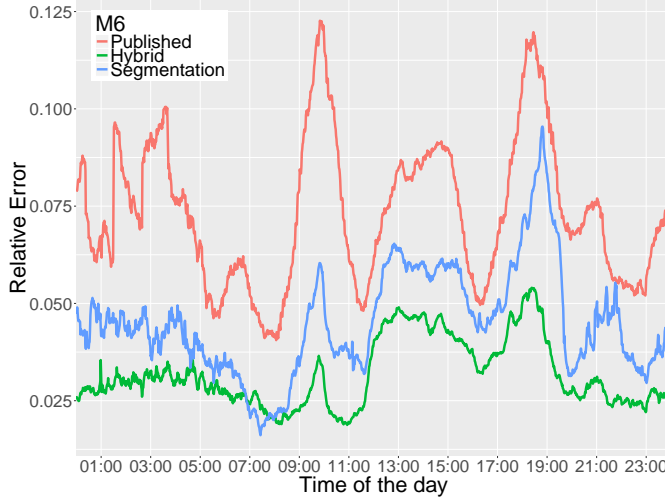

Fig. 11. Link average accuracy results in M25 across times of the day.

Fig. 9. Link average accuracy results in M6 across times of the day.

TABLE II
MAPE PER LINK ON M6

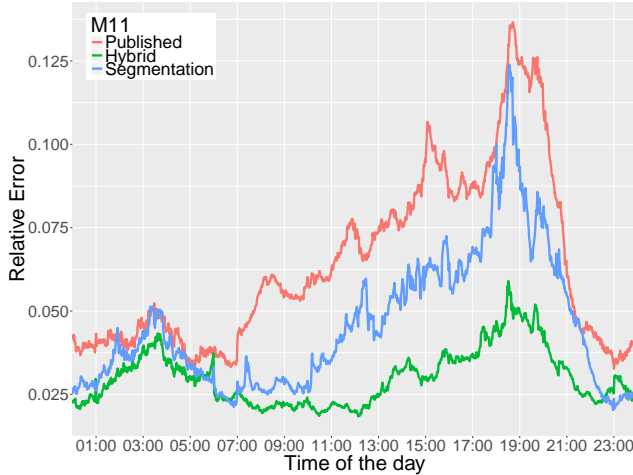| Link	MAPE
117007401	2.11%
117007501	4.21%
117007601	2.46%
117007801	4.80%
117007901	3.08%
117008401	3.31%
117009102	2.69%
117011901	2.35%
117012001	4.46%
117012101	2.62%
117012201	4.31%
117012301	2.75%
117016001	3.60%
123025901	1.89%

Fig. 10. Link average accuracy results in M6 across times of the day.

VII. FUTURE WORK

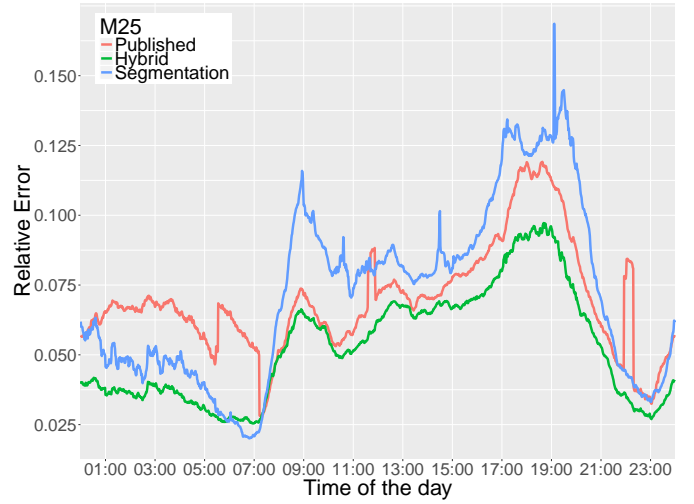The algorithm supplied above meets all the requirements described in Section V except for the fact that it requires a heuristically set threshold. One potential way of reaching total compliance with these requirements is to perform the decomposition between Background and Spikes by applying a Wavelet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

TABLE III
GLOBAL MAPE PER MOTORWAY

Motorway	MAPE [%]	RMSE [s]
M6	3.18%	3.68
M11	3.00%	3.55
M25	5.52%	6.67

minimum training data set, as well as maximum performance with increased training.

ACKNOWLEDGMENTS

REFERENCES

[1] The Highways Agency, "National Transport Information System Publish Services", Technical Report, 2011.

[2] 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.

[3] R. J. Hyndman and G. Athanasopoulos, "Forecasting: Principles and Practice", Otexts, 2013.

[4] 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.