



# Predictor fusion for short-term traffic forecasting

Fangce Guo\*, John W. Polak, Rajesh Krishnan

Centre for Transport Studies, Imperial College London, London SW7 2AZ, UK

## 1. Introduction

Short-term traffic prediction can be defined as the process of estimating the anticipated traffic conditions in the short-term future given historical and current traffic information (Vlahogianni et al., 2014). A wide variety of Intelligent Transport Systems (ITS) applications require predictive information regarding aspects of the transport network state. For example, in the case of network management applications predictive information enables the development of proactive (rather than reactive) network and incident strategies (Van Arem et al., 1997). Similarly, individual travellers can use the predictive information to plan their mobility decision more efficiently (Dia, 2001). Therefore, accurate short-term traffic prediction is one of the key components in ITS applications, and has been the subject of extensive research in recent years (Vlahogianni et al., 2014).

Traffic conditions in urban areas exhibit recurrent patterns over time (Williams and Hoel, 2003). Most traffic prediction methods make use of this periodicity. However, recurrent traffic conditions are affected by planned incidents such as road works, sports events and unplanned incidents and accidents, resulting in a deviation from the recurrent patterns. Short-term prediction is arguably more important during such abnormal conditions because of uncertainty about how the traffic state will evolve into the future. These factors also make it substantially more challenging.

The challenges of short-term prediction during disruptions resulting in abnormal conditions are not unique to transport. In a number of other fields including bioengineering, computer science and power networks, an effective approach to such problems has been found in the idea of prediction fusion (Bonissone et al., 2011; Ishida and Kinoshita, 2008; Loh and Henry, 2002; Mizianty and Kurgan, 2009). The central concept underlying prediction fusion is based on the observation that different prediction methods will have different strengths and weaknesses in different contexts of normal and abnormal operation (Chan and Stolfo, 1997). Therefore rather than relying on a single prediction methods, it is potentially desirable to combine or fuse the predictions arising from different methods, in a manner that is analogous to sensor data fusion. A key issue in prediction fusion is clearly the method used.

This paper applies similar ideas to the problem of short-term prediction in the transport domain. In particular, the main objective of this paper is to develop a fusion-based framework to improve the accuracy of traffic prediction based on a combination of multiple stand-alone predictors that work accurately under different traffic conditions. We evaluate fusion-based frameworks using three different fusion strategies: averaged, weighed and k-Nearest Neighbour (kNN) methods, applied to three different machine learning methods, Neural Networks (NN), Support Vector Regression (SVR) and Random Forests (RF). We select these three methods as the stand-alone predictors, not only because they are widely used in short-term traffic prediction under different conditions (Guo et al., 2017; Guo et al., 2010) but also because they are very different techniques resulting in model diversity (Mitchell, 1997). The stand-alone and fusion based methods are compared against each other using traffic flow and travel time data from Central London under different traffic conditions.

## 2. Related work

### 2.1. Previous short-term traffic prediction studies

Short-term traffic prediction is an active research area. However, most of the literature only considers prediction under normal

\* Corresponding author.

E-mail address: [fangce.guo@imperial.ac.uk](mailto:fangce.guo@imperial.ac.uk) (F. Guo).

traffic conditions, using recurrent traffic information. A common characteristic of these studies is that the recorded traffic data that had been influenced by abnormal conditions are detected and removed from the database used in traffic prediction applications (Castro-Neto et al., 2009). A wide range of different statistical and machine learning prediction models have been applied to such scenarios. For example, many early traffic prediction models use the basic Auto-Regressive Integrated Moving Average (ARIMA) (Box and Jenkins, 1970) method (Hamed et al., 1995; Szeto et al., 2009; Van Der Voort et al., 1996). More recently, machine learning methods have been widely used to predict future traffic variables. Commonly used methods include Neural Networks (NN) (Abdulhai et al., 1999; Ishak and Alecsandru, 2004; Park and Rilett, 1999; van Hinsbergen et al., 2009; Zheng et al., 2006), Support Vector Regression (SVR) (Wu et al., 2004), k-Nearest Neighbours (kNN) (Clark, 2003; Habtemichael and Cetin, 2016; Krishnan and Polak, 2008; Smith and Demetsky, 1997; Smith et al., 2002) and Random Forest (RF) (Guo et al., 2017; Guo et al., 2014; Leshem and Ritov, 2007). More recently still, deep learning methods such as Convolutional Neural Networks (CNN) and stacked auto-encoders (SAE) have been applied in short-term traffic prediction (e.g. Lv et al. (2015), Polson & Sokolov (2017)).

Comparison of prediction accuracy between these above methods was made in many studies (e.g. Guo et al. (2010); Guo et al. (2017); Smith & Demetsky (1997); Smith et al. (2002); Vlahogianni et al. (2004)). In general, these comparisons show that there is no single method that can best predict traffic states in all types of datasets and under different conditions (Tan et al., 2009; Vlahogianni et al., 2014; Zheng et al., 2006). Some studies discussed the reasons why a combination of individual forecasting methods might improve accuracy (Clemen, 1989; Makridakis, 1989). In recent years, researchers have begun to investigate strategies to combine predictors for traffic forecasting to increase accuracy under normal traffic conditions. The Bayesian approach was used in Zheng et al. (2006) to combine short-term predictive results from two single neural network predictors. Both training and testing data were collected from four locations under typical normal traffic conditions on an expressway in Singapore. A neural network based approach was used in Tan et al. (2009) to aggregate the results of three time-series forecasting methods. They used one-hour traffic flow data collected from one site on a highway in China. The results showed that the aggregation method could improve one-step ahead prediction accuracy. The authors stated that the developed model in their research only worked under normal traffic conditions because only simple time series prediction methods were used. More recently, a probabilistic method was used in Djuric et al. (2011) to fuse results from four single traffic speed predictors. Data used in their study was collected from a loop detector on a freeway in the USA. However, the developed aggregation models could not provide accurate prediction results during abnormal traffic conditions caused by incidents, sports events and severe weather. More recently, Tselentis et al. (2015) tested both linear regression and Bayesian combination methods with individual time series methods for short-term freeway traffic speed prediction. Both spatio-temporal and exogenous information such as rainfall and volume were considered in the proposed models. Similarly, Qiu et al. (2016) proposed an integrated precipitation-correction model for freeway traffic flow prediction using fusion techniques with four basic forecasting models. Vlahogianni (2015) proposed the surrogate model using three prediction methods for combination in short-term freeway traffic speed prediction. These studies have demonstrated that the combination of different predictors can improve the final accuracy of traffic prediction under normal traffic conditions on freeways and motorways. Table 1 shows the summary of the key features of traffic fusion literature reviewed above, under a number of headings covering the characteristics of the prediction context (such as urban/freeway), the fusion method and the traffic conditions (normal/ abnormal) within which the models were implemented.

The challenges of prediction under abnormal traffic conditions (such as non-recurrent traffic congestion that is caused by planned events such as road works or unplanned events such as incidents or accidents) have received much less attention in the literature, despite their significance, especially in urban areas. However, in the past decade interest in prediction under abnormal traffic conditions has grown. For example, Tao et al. (2005) used NN to predict short-term travel time during incidents using data collected from a highway corridor in the United States. Castro-Neto et al. (2009) proposed an Online-Support Vector Regression (OL-SVR) model to forecast traffic flow variables during holidays and traffic incidents using data collected from the United States ENREF.5. Random Forests (RF) was used in Guo et al. (2014) to predict travel time under incident conditions using data collected in urban areas in the United Kingdom.

Because different machine learning tools use different strategies to learn relationship from the training dataset, different predictors have differential performance in different circumstances. There is no single method that best all traffic variables under all traffic conditions. The strategy of predictor fusion is widely used to improve prediction accuracy in many fields, such as power (e.g., Bonissone et al., 2011), computer science (e.g., Loh & Henry, 2002) and biology (e.g., Chan & Stolfo, 1997). Motivated by this, a fusion-based framework is proposed to leverage the strengths of different machine learning tools using the same inputs for traffic prediction under a range of traffic conditions.

**Table 1**  
Categorisation of available literature in existing traffic prediction fusion models.

| Author                  | Context    | Input data resolution (min) | Prediction step | Input variables             | Traffic condition  | Fusion model             |
|-------------------------|------------|-----------------------------|-----------------|-----------------------------|--------------------|--------------------------|
| Zheng et al. (2006)     | Highway    | 15 min                      | 1               | Flow                        | Normal             | Bayesian                 |
| Tan et al. (2009)       | Expressway | 60 min                      | 1 and multi     | Flow                        | Normal             | NN                       |
| Djuric et al. (2011)    | Freeway    | 5 min                       | Multi           | Speed                       | Normal             | Probabilistic            |
| Qiu et al. (2016)       | Expressway | 5 min                       | Multi           | Flow and Rainfall           | Normal /Heavy rain | Precipitation-correction |
| Tselentis et al. (2015) | Freeway    | 5 min                       | 1               | Speed and Flow/<br>Rainfall | Normal /Heavy rain | Statistic and Bayesian   |
| Vlahogianni (2015)      | Freeway    | 5 min                       | 1               | Speed                       | Normal             | Surrogate model          |

## 2.2. Fusion-based methods

Fusion is used in machine learning to combine individual learning processes (Chan and Stolfo, 1997). The main advantage of fusing multiple methods is to leverage their complementary predictive characteristics. In order to combine multiple predictors, several different strategies can be used. The average method is the simplest method to combine multiple predictors. This method calculates the mean predictive values among different individual predictors. A variant of the simple average method is the weighted method (Chan and Stolfo, 1997). In this method, each predictor has a weight, which is calculated on the basis of how accurately the predictor performs using a validation dataset. The final prediction is the sum of weighted results of multiple predictors. However, the values of weights are predetermined and constant during forecasting. In recent studies, machine learning methods such as kNN are also used to fuse different predictors (Bonissone et al., 2011). The kNN method, as a typical lazy learning method (Hastie et al., 2001), does not involve any explicit model construction and can locate nearest neighbours of observations from a historical dataset. The weights of each predictor are dynamically calculated and updated with new observations and are not pre-determined. Given the characteristics of kNN, unexpected traffic conditions such as incidents and accidents can easily be incorporated into the fusion framework. kNN is selected as one of the fusion methods for traffic forecasting under different traffic conditions in this paper.

## 3. Fusion-based framework

### 3.1. Fusion-based methods

The basic idea behind a fusion-based framework is to generate a final prediction result by combining the output of two or more stand-alone predictors. Fig. 1 shows a prediction framework that combines  $m (m \geq 2)$  predictors.

### 3.2. Fusion strategies

In this paper, three different fusion strategies, average fusion, weighted fusion and kNN fusion, are used within the framework to fuse multiple predictors.

- **Average fusion-based method:** This simple method uses the average of prediction results of multiple predictors as the final result. Given a set of predictors,  $\hat{y}^i, i = 1, \dots, m$ , we seek to compute final prediction  $\hat{y}$ . The simple average method is

$$\hat{y} = \frac{1}{m}(\hat{y}^1 + \hat{y}^2 + \dots + \hat{y}^m) \quad (1)$$

- **Weighted fusion-based method:** In this method, weights of different predictors are not the same. Weighted hybrid method is written as

$$\hat{y} = \alpha_1 \hat{y}^1 + \alpha_2 \hat{y}^2 + \dots + \alpha_m \hat{y}^m = \sum_i \alpha_i \hat{y}^i \quad (2)$$

$$\alpha_1 + \alpha_2 + \dots + \alpha_m = 1 \quad (3)$$

$\hat{y}^i$  is the prediction result using  $i^{th}$  predictor and  $\alpha_i$  is the weight of  $i^{th}$  predictor. Weights are calculated using training dataset. Weighted fusion-based method is the same as simple average method when all the weights are equal to  $1/m$ . In this paper, the weights are calculated by the inverse of Mean Absolute Percentage Error (MAPE),  $w = 1/MAPE$  and  $\hat{y} = \sum_i w_i \hat{y}^i / \sum_i w_i$ .

- **The kNN fusion-based method:** The kNN method is highly unstructured and does not require any pre-determined model specification. The basic idea of kNN fusion-based method is, given a current traffic state, to search nearest neighbours to this state in the training used historical datasets, calculate prediction errors of nearest neighbour set, estimate weights of each predictor, and combine the final predicted outputs of individual predictors based on these weights. Fig. 2 depicts the flowchart of the kNN fusion-based method. There are two steps to summarise the kNN fusion-based method.

Step 1: Neighbourhood searching procedure:

The search procedure finds the nearest neighbours, which are the historical observations that are most similar to the current observation. Euclidean distance is used in this paper to determine the distance between the current input feature vector and historical observations. More details of the kNN parameter selection can be found in Guo et al. (2017).  $k$  is the number of historical observations with the nearest distances to the input feature vector. The set of  $k$  nearest neighbours of the input feature vector  $\mathbf{x}_c$  can be written as  $J_{\mathbf{x}_c} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k]^T$  and  $\mathbf{x}_j = [x_1, x_2, \dots, x_n]$ ,  $j = 1, 2, \dots, k$ , and  $n$  is the dimension of feature space. The trial and error method introduced in Guo et al. (2012) is used for setting the parameters of kNN.  $k$  is set to 5 in this paper.

Step 2: Weighted parameter estimation process:

This step is used to calculate weights of each predictor. For each vector  $\mathbf{x}_j$ , the predicted values  $\hat{y}_j^i = f_i(\mathbf{x}_j), i = 1, 2, \dots, m$  and errors  $e_j^i = \hat{y}_j^i - y_j$  of each predictor can be calculated. Thus, the weights of each predictors at current time can be estimated using the errors,  $w_i = 1/MAPE_i$ , where MAPE is calculated based on the selected nearest neighbour dataset. The main different between weighted and

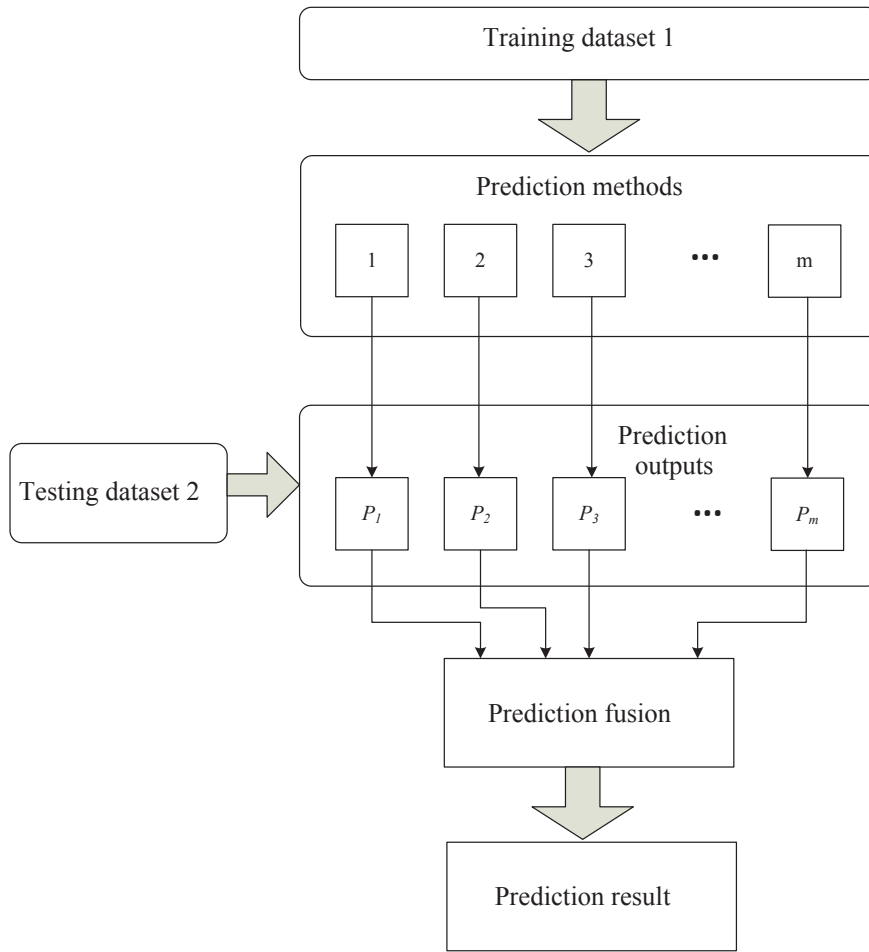


Fig. 1. A framework for fusion-based prediction.

kNN fusion methods is that the weights used in kNN fusion are dynamically updated every step.

### 3.3. Machine learning methods used in prediction

Three commonly used machine learning methods, Neural Networks (NN), Support Vector Regression (SVR) and Random Forests (RF), are selected as individual predictors in this paper. These selected machine learning tools use different algorithms for learning the relationships between two related data series. Hence, these machine learning methods are selected in the proposed framework to achieve model diversity.

- **Neural networks (NN):** Neural networks (NN) (Mitchell, 1997) capture complex relationship between multiple inputs and multiple outputs. As a data driven method, NNs do not require the prior specification of an explicit physical relationship between data. The relationship between inputs and outputs is inferred from a training dataset. NN has been widely used in traffic engineering for tasks such as traffic prediction and incident management. In this paper, a well-known feedforward NN (Mitchell, 1997) with a backpropagation algorithm is selected to predict traffic data. Backpropagation has a well-developed theoretical background and has good capability for modelling relationship between continuous variables (Smith and Demetsky, 1997). Moreover, it has been proven to be effective in a wide range of applications that require the representation of complex relationships, including in the context of traffic prediction (Park and Rilett, 1999). Cross-validation for the selection of parameters is used in this paper. After minimising the Mean Absolute Percentage Error in the process of optimisation, the number of hidden layers is set to 8 in this paper.
- **Support Vector Regression (SVR):** Support Vector Regression (SVR) was developed by Vapnik (1995) based statistical learning theory. SVR can solve the non-linear regression problems by mapping the input data into a higher dimension feature space where a linear regression model can be used in the calculation (Vapnik, 1995). A detailed explanation of SVR and its applications can be found in Vapnik (1995). In this paper, a Radial Basis Function (RBF) is used as the kernel function because it is shown to be more suitable for traffic prediction under different conditions (Guo et al., 2017). There are many methods introduced in the literature to

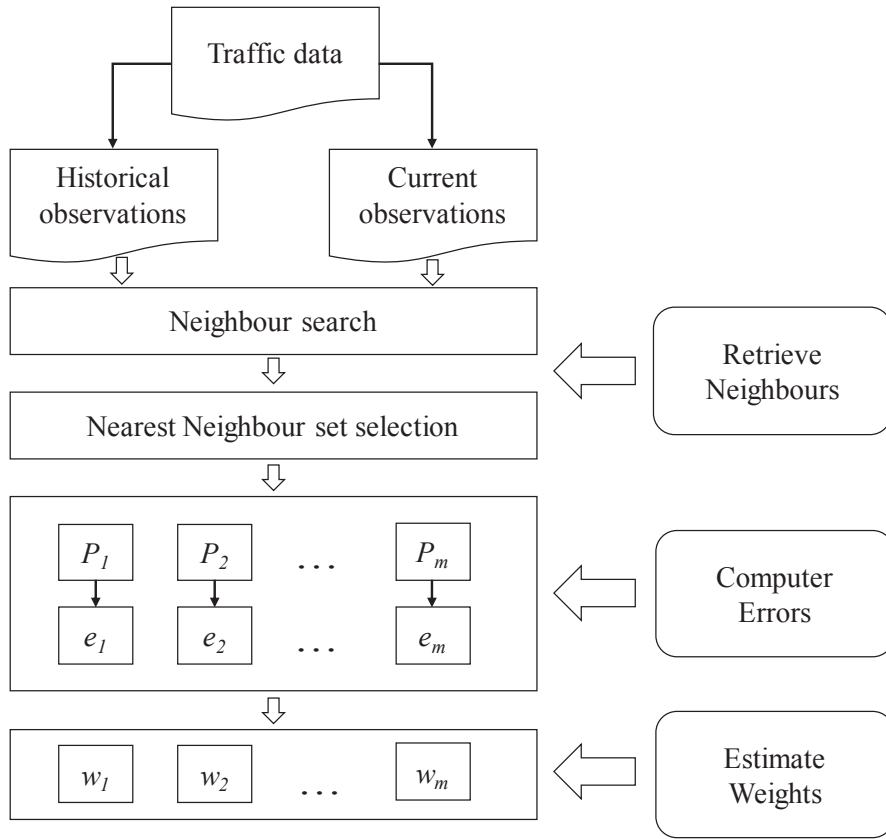


Fig. 2. Flowchart of kNN fusion based method.

select the optimised SVR parameters (e.g. Cherkassky and Ma, 2004). The process of parameter optimisation of SVR has been described in Guo et al. 2012. After optimising, the capacity value  $C$  of SVR is set to 80 and the  $\epsilon$ -insensitive loss function ( $\epsilon = 0.05$ ) is used.

- *Random forests (RF)*: Random forests (RF) was proposed by (Breiman, 2001) based on ensemble learning methods (see, e.g. Hastie et al., 2001). The main difference between RF and traditional tree method is that RF randomly selects variables to split each tree and node, which changes the methods of tree construction and avoids overfitting problems. In the application of regression, the predictive result is calculated by taking the average of all corresponding trees. More details of RF are described in Verikas et al. (2011). One of the key parameters is the number of trees. In the optimisation process, the trial and error method is used. In this paper, the number of trees is set to 500.

### 3.4. Prediction accuracy measurement

In this paper, there are three criteria, Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) used to evaluate the final prediction accuracy. MPE is used to evaluate the prediction bias at the disaggregate level. MAPE is used to calculate the mean of the absolute differences between predictive and observed traffic variables. RMSE gives additional weights to larger absolute errors. Therefore, these three measures are used to evaluate the accuracy and precision of prediction results. The definition of these measures are as follows:

$$MPE = \frac{1}{n} \sum_{i=1}^n \left[ \frac{t_i - \hat{t}_i}{t_i} \right] * 100\% \quad (4)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left[ \frac{|t_i - \hat{t}_i|}{t_i} \right] * 100\% \quad (5)$$

$$RMSE = \sqrt{\sum_{i=1}^n (|t_i - \hat{t}_i|^2) / n} \quad (6)$$

where  $t_i$  is the observed traffic variables;  $\hat{t}_i$  is the predicted traffic variables and  $n$  is the total sample size.

#### 4. Experiments

The context of this paper for short-term traffic prediction is the urban road networks in the UK. Two test locations in urban areas are selected to examine the transferability of the proposed fusion based prediction framework. Two different types of traffic variables, traffic flow and travel time are used. Moreover, the accuracy of the proposed framework is tested under both normal and incident conditions to examine framework robustness.

Traffic flow data used in this paper are from Inductive Loop Detectors (ILDs), part of SCOOT (Robertson and Bretherton, 1991) traffic control system in Central London. Traffic flow data collected from Cromwell Road and Marylebone Road were selected. Cromwell Road is a major urban road located in the Royal Borough of Kensington and Chelsea. A map of the Cromwell Road and the location of ILDs is shown in Fig. 3(a). The Marylebone Road is a famous thoroughfare in the centre of London from Euston Road at Regent's Park to the A40 Westway at Paddington. This corridor has three lanes of traffic in both directions. A map of the Marylebone corridor is shown in Fig. 3(b).

Link travel time data from London was obtained from the London Congestion Analysis Project (LCAP) (TfL, 2010). LCAP is operated by Transport for London (TfL) to estimate and store link travel time data in London based on ANPR camera data. Link 1309 on the A40 in London with a length of 5.63 km was selected for this application. The topology of this road link in the LCAP system is shown in Fig. 3(c). The direction of travel is from West to East.

In this paper, the proposed fusion-based prediction model was assessed under four test cases using both traffic flow and travel time data as described below:

Case 1: 5-min traffic flow data from Cromwell Road (Fig. 3(a)) is used for this test case. In this case, due to the accessibility of traffic flow data in Central London, traffic flow data from 05:00 to 22:00 was extracted. Traffic data are divided into training dataset and testing dataset. The training dataset was collected from 1st November 2013 to 16th December 2013. The testing dataset was collected from 17th December 2013 to 23rd December 2013. Data from weekends are excluded since traffic profiles are different during weekends.

Case 2: 15-min traffic flow data from the Marylebone corridor is used in Case 2. The training dataset was from April, May and June 2008. Only traffic data under normal conditions were used for training. A severe traffic incident happened on the testing day, 20th June 2008, a Friday. The incident happened around 17:30 and was cleared around 20:00. The location of this incident was near the intersection of Mac Farren Place and Marylebone Road (point A in Fig. 3(b)).

Case 3: Link travel time data from A40 in London (Fig. 3(c)) under normal traffic conditions without incidents or accidents was selected. Travel time data for a period of three months between January and March is divided into two datasets. Training data is from 3rd January 2011 to 17th March 2011; while testing data is from 17th March 2011 to 24th March 2011. Data from weekends are excluded from training and test datasets.

Case 4: Link travel time data from A40 in London during the incident is used for this test case. In order to evaluate the accuracy of the proposed model when traffic data was disrupted, travel time prediction during incidents is tested in Case 4. Training data collected during periods of normal traffic are from weekdays during October, November and December 2010; testing data is from 21st December 2010. One lane was blocked on A40 on the testing day because of a broken down vehicle. This incident occurred around 18:00 and was cleared around 18:40. The location of this incident is shown at point A in Fig. 3(c).

#### 5. Experimental results

Traffic data were predicted using individual machine learning methods, NN, SVR and RF, and with the fusion-based structure using the averaged fusion method, the weighted fusion method and the kNN fusion based method. All the methods were tested using both traffic flow and travel time data under normal and incident conditions. Figs. 4–6 and 7 are examples of time series plots of prediction results using different methods in Case 1, Case 2 Case 3 and Case 4, respectively. The results of comparisons are shown in Table 2 for four test cases with and without fusion-based structures. The mean values of MPE, MAPE and RMSE of three individual methods also shown in Table 2 are used as a baseline in order to examine the accuracy improvement of fusion based methods.

In test Case 1 – short-term traffic flow prediction under normal traffic conditions, the mean value of MAPE using three individual predictors is 6.42%. The MAPE values using three fusion based methods of averaged, weighted and kNN fusion based methods are 6.41%, 6.40% and 6.28% respectively. There is no significant accuracy improvement using Wilcoxon signed-rank test for these fusion methods in traffic flow prediction under normal traffic conditions.

In test Case 2 – short-term traffic flow prediction under abnormal conditions, the values of MPE, MAPE and RMSE under incident conditions were calculated based on the period starting 30 min before the occurrence of incidents and finishing around 30 min after the clearance of incidents. The MAPE values using simple averaged and weighted fusion based methods are 37.31% and 37.19%. The kNN fusion based method gives the most accurate results, 19.45% of MPE, 36.10% of MAPE and 303.46 veh/h of RMSE. Compared with the baseline of MAPE, the improvement of kNN fusion based method is 8% under incident traffic conditions.

In test Case 3 – travel time prediction under normal conditions, the improvements of weighted and kNN fusion based methods using MAPE metrics are 8.98% and 9.77% respectively, compared with the baseline. Similarly, the value of RMSE is reduced from 45.19 s to 39.02 s of the weighted fusion method and 36.39 s of the kNN fusion based method.

In test Case 4, the abnormal event happened around 17:57 and cleared around 18:40. It is clear that, under incident conditions, travel time prediction accuracy in terms of MAPE and RMSE increases when the fusion based framework is applied, regardless of the fusion methods used. For example, the mean value of MAPE by individual methods is 10.50%, while MAPE values of 8.85%, 8.47% and 8.30% are yielded when averaged fusion-based method, weighed fusion-based method and kNN fusion-based method are used.

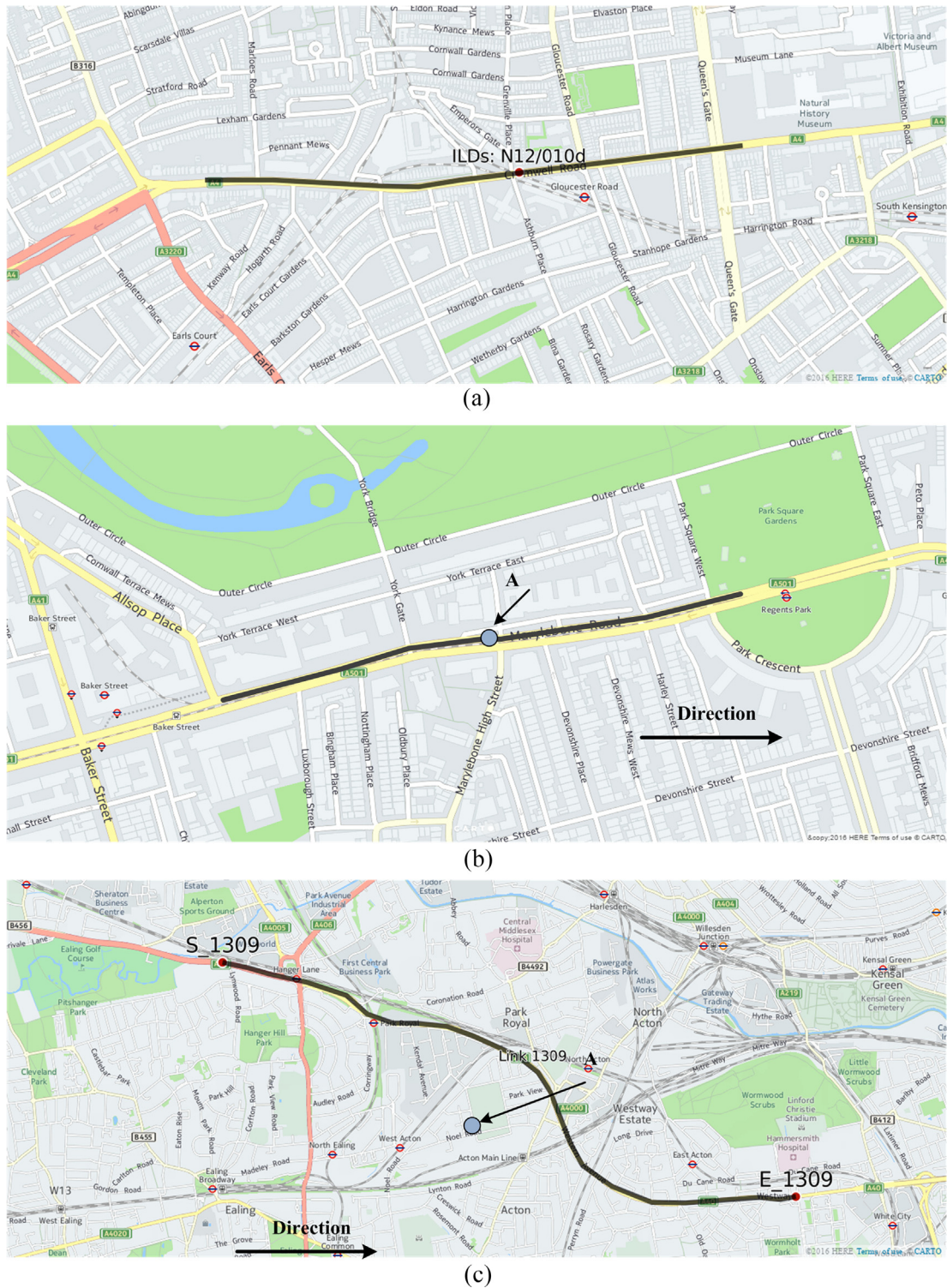


Fig. 3. Selected links for the test cases (Map Source: CARTO): (a) Cromwell Road corridor, (b) Marylebone Road corridor and (c) A40 corridor.

Moreover, the kNN fusion based method is more accurate in short-term travel time prediction during incidents. For example, the improvement in MAPE value is 21.85% using kNN fusion based method compared with the individual methods in Case 4. Similarly, the RMSE metric improves from 199.12 s of the baseline to 153.75 s using the kNN fusion based method, a 22.79% improvement.

From the viewpoint of fusion methods, the results of Table 1 indicate that simple fusion methods such as averaged method and

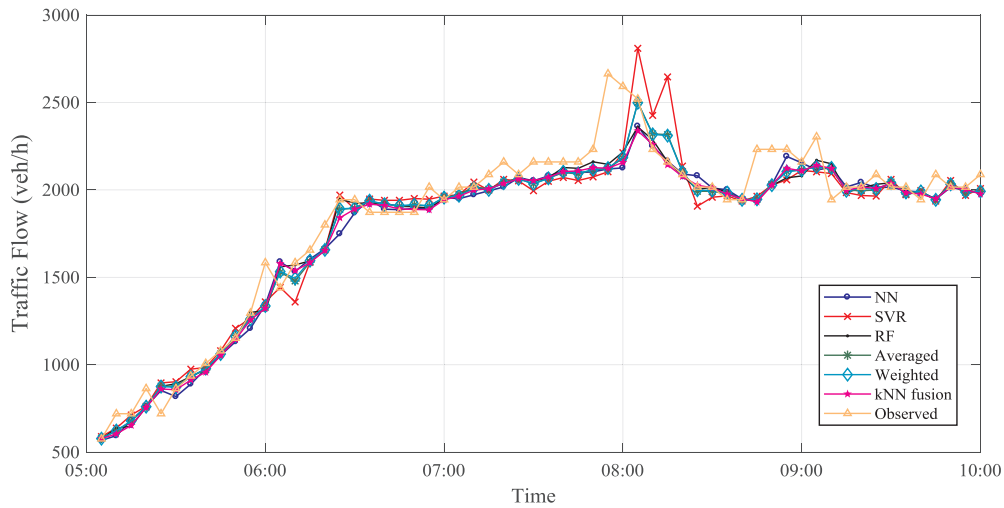


Fig. 4. Example of time series plot of predicted and observed traffic data in Case 1.

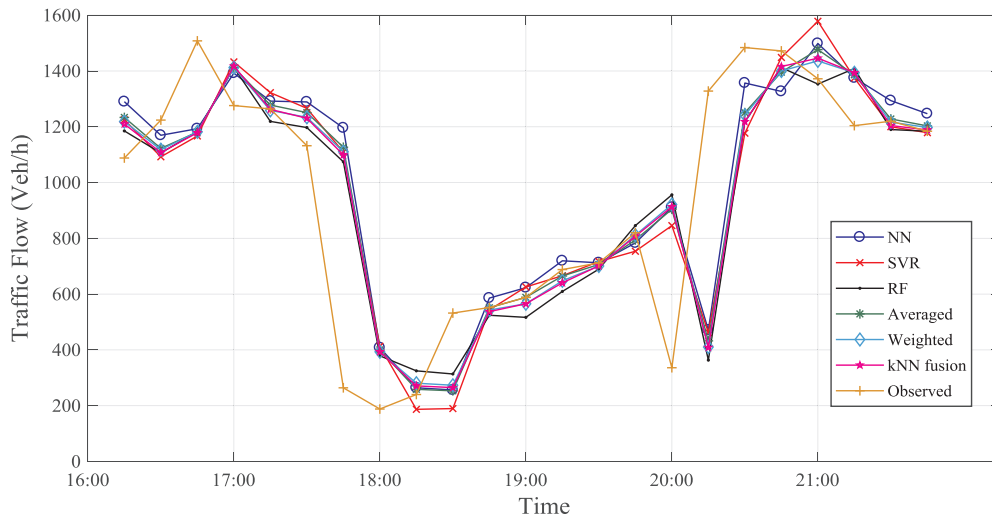


Fig. 5. Time series plot of predicted and observed traffic data in Case 2.

weighted methods only offer moderate improvements to the overall prediction accuracy. By contrast, the kNN fusion based method improved the final accuracy in four different test cases, especially Case 2 and Case 4 – incident traffic conditions. For example, in Case 4, the MAPE value is reduced to 8.30% from 9.82% which is the best result using the individual method, a 15.5% increase in accuracy. In Case 2 and Case 4 – forecasting during incident traffic conditions, the prediction results in terms of the absolute percentage errors (APE) cannot be analysed using Wilcoxon signed-rank test because the incident periods are not long enough to create sufficient samples. When the absolute predictive percentage errors of NN and kNN-based fusion method using the whole testing day data including incident periods were analysed as the samples, the results show that the predictive errors of kNN based fusion method are significantly smaller than those of NN when the significance level was set to 0.025 for hypothesis testing.

In this paper, there are two types of input data, traffic flow and travel time. In short-term traffic flow prediction, the average improvement of kNN fusion based method using the MAPE metric is 6.43% under both normal and incident conditions. In short-term travel time prediction, the corresponding improvement is 15.81%.

## 6. Conclusions

Traffic states are affected by traffic incidents and accidents, which cause uncertainties for short-term traffic prediction models using conventional machine learning tools. This paper presents a framework for the fusion of multiple predictors to reduce these uncertainties and improve the performance of short-term traffic prediction. A series of experiments using both traffic flow and travel time data from Central London under different traffic conditions were undertaken to evaluate the accuracy of different individual predictors and different fusion methods. The results of these experiments indicate that there are significant differences in the

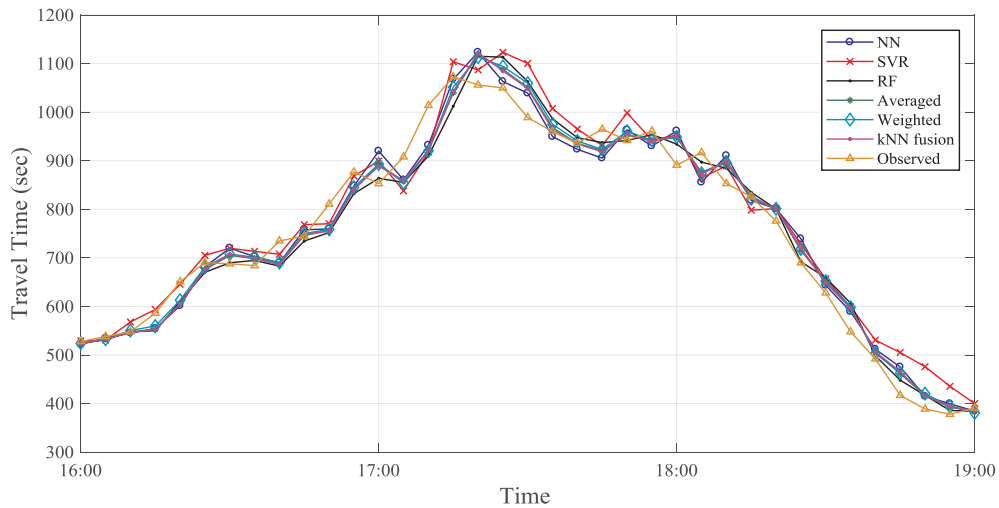


Fig. 6. Example of time series plot of predicted and observed traffic data in Case 3.

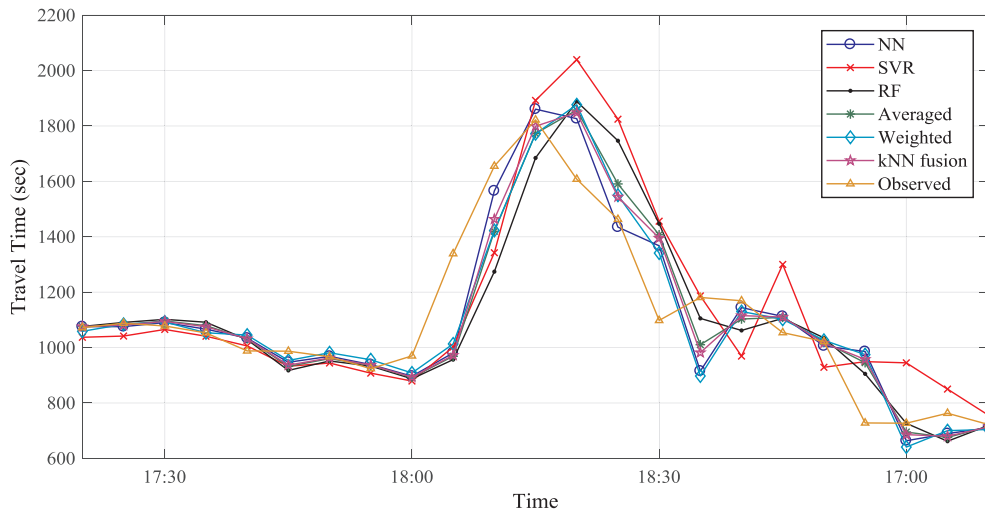


Fig. 7. Time series plot of predicted and observed traffic data in Case 4.

**Table 2**  
Prediction accuracy of different methods.

|          | Case 1: Traffic flow prediction<br>under normal conditions |             |                 | Case 2: Traffic flow prediction<br>during incident period |              |                 | Case 3: Travel time prediction<br>under normal conditions |             |               | Case 4: Travel time prediction<br>during incident period |             |               |
|----------|------------------------------------------------------------|-------------|-----------------|-----------------------------------------------------------|--------------|-----------------|-----------------------------------------------------------|-------------|---------------|----------------------------------------------------------|-------------|---------------|
|          | MPE<br>(%)                                                 | MAPE<br>(%) | RMSE<br>(veh/h) | MPE (%)                                                   | MAPE<br>(%)  | RMSE<br>(veh/h) | MPE (%)                                                   | MAPE (%)    | RMSE<br>(sec) | MPE (%)                                                  | MAPE<br>(%) | RMSE (sec)    |
| NN       | 1.31                                                       | 6.42        | 49.71           | 24.89                                                     | 39.61        | 318.76          | 0.76                                                      | 4.72        | 44.61         | −1.39                                                    | 9.82        | 189.78        |
| SVR      | 1.22                                                       | 6.43        | 50.83           | 22.44                                                     | 39.76        | 316.85          | 1.57                                                      | 5.89        | 50.34         | <b>0.23</b>                                              | 11.06       | 211.50        |
| RF       | 1.38                                                       | 6.40        | 50.03           | 21.34                                                     | 38.25        | 319.50          | <b>0.42</b>                                               | 4.75        | 40.63         | −0.82                                                    | 10.62       | 196.07        |
| Baseline | 1.30                                                       | 6.42        | 50.19           | 22.89                                                     | 39.21        | 318.37          | 0.92                                                      | 5.12        | 45.19         | −0.66                                                    | 10.50       | 199.12        |
| Averaged | 1.31                                                       | 6.41        | 49.50           | 21.69                                                     | 37.31        | 315.03          | −1.01                                                     | 5.11        | 43.48         | −1.07                                                    | 8.85        | 161.01        |
| Weighted | 1.30                                                       | 6.40        | 49.50           | 21.31                                                     | 37.19        | 314.12          | −0.80                                                     | 4.66        | 39.02         | −1.35                                                    | 8.47        | 156.73        |
| kNN      | <b>0.80</b>                                                | <b>6.28</b> | <b>47.68</b>    | <b>19.45</b>                                              | <b>36.10</b> | <b>303.46</b>   | 0.61                                                      | <b>4.62</b> | <b>36.39</b>  | −1.17                                                    | <b>8.30</b> | <b>153.75</b> |

The bold numbers are the smallest absolute values in each column.

prediction accuracy of different fusion strategies used. The results show that while the accuracy of the averaged and weighted fusion methods is comparable to the accuracy of individual methods, the kNN fusion based method can achieve significantly superior results, especially during disrupted traffic conditions. Overall, the results support the hypothesis that suitably configured fusion based

methods that exploit the complementary strengths of different predictors can indeed improve final prediction results. They also indicate, however, that the way in which the fusion is performed is critical. However, the proposed fusion framework requires calibrating multiple prediction models in parallel. This increases the computational complexity compared with the standalone prediction methods used within the framework. Thus, there is a trade-off to be had between prediction accuracy and computational efficiency of the fusion method.

Future work will aim to refine further our understanding of the most effective fusion methods. More machine learning methods will be used as individual predictors within the fusion framework. In addition, the prediction performance of methods discussed in the literature needs to be compared with those of the proposed framework under disrupted conditions.

## Acknowledgements

The authors would like to thank Transport for London (TfL) for providing the traffic data used in this paper. This work was undertaken as a part of the CoSy Academic Engagement Programme [ITC 11402] supported by TfL; oneTRANSPORT Project supported by Innovate UK [Award Number 53472-404172]; and was partially supported by the UK Engineering and Physical Sciences Research Council [Award Number EP/F005156/1 and EP/I038837/1].

## References

- Abdulhai, B., Porwal, H., Recker, W., 1999. Short-term freeway traffic flow prediction using genetically optimized time-delay-based neural networks. In: *Proceedings of the 87th Annual Meeting of the Transportation Research Board Washington D.C., USA*.
- Bonissone, P.P., Xue, F., Subbu, R., 2011. Fast meta-models for local fusion of multiple predictive models. *Appl. Soft Computing* 11 (2), 1529–1539.
- Box, G.E.P., Jenkins, G.M., 1970. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Castro-Neto, M., Jeong, Y.S., Jeong, M.K., Han, L.D., 2009. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Syst. Appl.* 36 (3), 6164–6173.
- Chan, P.K., Stolfo, S.J., 1997. On the accuracy of meta-learning for scalable data mining. *J. Intell. Inform. Syst.* 8 (1), 5–28.
- Cherkassky, V., Ma, Y., 2004. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Netw.* 17 (1), 113–126.
- Clark, S., 2003. Traffic prediction using multivariate nonparametric regression. *J. Transport. Eng. – ASCE* 129 (2), 161–168.
- Clemen, R.T., 1989. Combining forecasts: a review and annotated bibliography. *Int. J. Forecast.* 5 (4), 559–583.
- Dia, H., 2001. An object-oriented neural network approach to short-term traffic forecasting. *Eur. J. Oper. Res.* 131 (2), 253–261.
- Djuric, N., Radosavljevic, V., Coric, V., Vucetic, S., 2011. Travel speed forecasting by means of continuous conditional random fields. *Transport. Res. Record: J. Transport. Res. Board* 2263, 131–139.
- Guo, F., Krishnan, R., Polak, J., 2012. Short-term traffic prediction under normal and incident conditions using singular spectrum analysis and the k-nearest neighbour method. In: *Proceedings of the 17th International Conference on Road Transport Information and Control (RTIC)*, London, UK.
- Guo, F., Krishnan, R., Polak, J., 2017. The influence of alternative data smoothing prediction techniques on the performance of a two-stage short-term urban travel time prediction framework. *J. Intell. Transport. Syst.* 21 (3), 214–226.
- Guo, F., Krishnan, R., Polak, J.W., 2014. A novel three-stage framework for short-term travel time prediction under normal and abnormal traffic conditions. In: *Proceedings of the 93rd Annual Meeting of the Transportation Research Board, Washington D.C., USA*.
- Guo, F., Polak, J., Krishnan, R., 2010. Comparison of modelling approaches for short term traffic prediction under normal and abnormal conditions. In: *Proceedings of the 13th International IEEE Annual Conference on Intelligent Transportation Systems, Madeira Island, Portugal*.
- Habtemichael, F.G., Cetin, M., 2016. Short-term traffic flow rate forecasting based on identifying similar traffic patterns. *Transport. Res. Part C: Emerg. Technol.* 66, 61–78.
- Hamed, M.M., Almasaeid, H.R., Said, Z.M.B., 1995. Short-term prediction of traffic volume in urban arterials. *J. Transport. Eng.-ASCE* 121 (3), 249–254.
- Hastie, T., Tibshirani, R., Friedman, J.H., 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Verlag, New York.
- Ishak, S., Alecsandru, C., 2004. Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings. *J. Transp. Eng.* 130 (4), 452–465.
- Ishida, T., Kinoshita, K., 2008. Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics* 24 (11), 1344–1348.
- Krishnan, R., Polak, J.W., 2008. Short-term travel time prediction: an overview of methods and recurring themes. In: *Proceedings of the Transportation Planning and Implementation Methodologies for Developing Countries Conference (TPMDC 2008)*, Mumbai, India.
- Leshem, G., Ritov, Y., 2007. Traffic flow prediction using adaboost algorithm with random forests as a weak learner. *World Academy of Science. Engineering and Technology. Citeseer, Bangkok, Thailand*, pp. 193–198.
- Loh, G.H., Henry, D.S., 2002. Predicting conditional branches with fusion-based hybrid predictors. In: *Proceedings in the 2002 International Conference on Parallel Architectures and Compilation Techniques*.
- Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.Y., 2015. Traffic flow prediction with big data: a deep learning approach. *IEEE Trans. Intell. Transp. Syst.* 16 (2), 865–873.
- Makridakis, S., 1989. Why combining works? *Int. J. Forecast.* 5 (4), 601–603.
- Mitchell, T.M., 1997. *Machine Learning*. McGraw-Hill, New York.
- Mizianty, M.J., Kurgan, L., 2009. Meta prediction of protein crystallization propensity. *Biochem. Biophys. Res. Commun.* 390 (1), 10–15.
- Park, D., Rilett, L.R., 1999. Forecasting freeway link travel times with a multilayer feedforward neural network. *Comput.-Aided Civ. Infrastruct. Eng.* 14 (5), 357–367.
- Polson, N.G., Sokolov, V.O., 2017. Deep learning for short-term traffic flow prediction. *Transport. Res. Part C: Emerg. Technol.* 79, 1–17.
- Qiu, H., Li, R., Liu, H., 2016. Integrated model for traffic flow forecasting under rainy conditions. *J. Adv. Transport.* 50 (8), 1754–1769.
- Robertson, D.L., Bretherton, R.D., 1991. Optimizing networks of traffic signals in real time-the SCOOT method. *IEEE Trans. Veh. Technol.* 40 (1), 11–15.
- Smith, B.L., Demetsky, M.J., 1997. Traffic flow forecasting: comparison of modeling approaches. *J. Transport. Eng. - ASCE* 123 (4), 261–266.
- Smith, B.L., Williams, B.M., Keith Oswald, R., 2002. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transport. Res. Part C: Emerg. Technol.* 10 (4), 303–321.
- Szeto, W.Y., Ghosh, B., Basu, B., O'Mahony, M., 2009. Multivariate traffic forecasting technique using Cell Transmission Model and SARIMA model. *J. Transp. Eng.* 135 (9), 658–667.
- Tan, M.C., Wong, S.C., Xu, J.M., Guan, Z.R., Zhang, P., 2009. An aggregation approach to short-term traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* 10 (1), 60–69.
- Tao, Y., Yang, F., Qiu, Z.J., Ran, B., 2005. Travel time prediction in the presence of traffic incidents using different types of neural networks. In: *Proceedings of the Transportation Research Board 85th Annual Meeting, Washington D.C., USA*.
- Tselentis, D.I., Vlahogianni, E.I., Karlaftis, M.G., 2015. Improving short-term traffic forecasts: to combine models or not to combine? *IET Intell. Transp. Syst. Instit. Eng. Technol.* 193–201.
- Van Arem, B., Kirby, H.R., Van Der Vlist, M.J.M., Whittaker, J.C., 1997. Recent advances and applications in the field of short-term traffic forecasting. *Int. J. Forecast.* 13 (1), 1–12.

- Van Der Voort, M., Dougherty, M., Watson, S., 1996. Combining kohonen maps with arima time series models to forecast traffic flow. *Transport. Res. Part C: Emerg. Technol.* 4 (5), 307–318.
- van Hinsbergen, C.P.I., van Lint, J.W.C., van Zuylen, H.J., 2009. Bayesian committee of neural networks to predict travel times with confidence intervals. *Transport. Res. Part C: Emerg. Technol.* 17 (5), 498–509.
- Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. Springer Verlag, New York.
- Verikas, A., Gelzinis, A., Bacauskiene, M., 2011. Mining data with random forests: a survey and results of new tests. *Pattern Recogn.* 44 (2), 330–349.
- Vlahogianni, E.I., 2015. Optimization of traffic forecasting: Intelligent surrogate modeling. *Transport. Res. Part C: Emerg. Technol.* 55(Supplement C), 14–23.
- Vlahogianni, E.I., Golias, J.C., Karlaftis, M.G., 2004. Short-term traffic forecasting: overview of objectives and methods. *Transp. Rev.* 24 (5), 533–557.
- Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2014. Short-term traffic forecasting: Where we are and where we're going. *Transport. Res. Part C: Emerg. Technol.* 43, Part 1, 3–19.
- Williams, B.M., Hoel, L.A., 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results. *J. Transport. Eng.-ASCE*. 129 (6), 664–672.
- Wu, C.H., Ho, J.M., Lee, D.T., 2004. Travel-time prediction with support vector regression. *IEEE Trans. Intell. Transp. Syst.* 5 (4), 276–281.
- Zheng, W., Lee, D.H., Shi, Q., 2006. Short-term freeway traffic flow prediction: Bayesian combined neural network approach. *J. Transp. Eng.* 132 (2), 114–121.