Ain Shams Engineering Journal 11 (2020) 535–551

Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

Electrical Engineering

Traffic congestion prediction based on Hidden Markov Models and contrast measure☆

John F. Zaki, Amr Ali-Eldin, Sherif E. Hussein, Sabry F. Saraya, Fayez F. Areed

Computer and Control Systems Engineering, Faculty of Engineering, Mansoura University, Mansoura, Egypt

ARTICLE INFO

Article history:
Received 8 July 2019
Revised 23 August 2019
Accepted 9 October 2019
Available online 7 November 2019

Keywords:
Hidden Markov models
Traffic congestion prediction
Freeway traffic
Contrast measure
Correlation analysis
Empirical evaluation

ABSTRACT

Traffic congestion is an important socio-economic problem that swelled in the last few decades. It affects the social mobility of people, length of trips, quality of life, and the economy of countries. As a major problem in most countries, it has been tackled by governments, universities, and advanced research using intelligent transportation systems (ITS) to solve the problem or at least ease its adverse effects. Hidden Markov Models (HMM) represent one of the methods that are suitable for congestion prediction. In this paper, a new model, based on Hidden Markov Model and Contrast, is proposed to define the traffic states during peak hours in two dimensional space (2D). The proposed model uses mean speed and contrast to capture the variability in traffic patterns. Empirical evaluation shows that the proposed approach has improved prediction error in comparison to HMM related work and neuro-fuzzy approaches.

1. Introduction

In today's life, everybody commutes for various intents such as business reasons, personal purposes or socializing. In most cases, people travel at the same time of the day which produces congestion on the roads. For instance, going to work during the weekdays. Hence, an increasingly large number of people are planning an additional time for their trips "just in case". They attempt to use whatever way possible to avoid congestion and guarantee arrival to the required destinations on time. Therefore, methods of traffic forecasting are of utmost importance. That is, ability to provide people with an informed decision regarding their trips is invaluable to both traffic officers and travelers alike. This forecasting would not have been possible without the use of Intelligent Transportation Systems (ITS) technologies. One of such technologies is the modeling and forecasting of traffic systems based on real time data. An accurate prediction can help decision makers to improve the transportation service and reduce associated cost. There is a myriad of methods that can be used for traffic forecasting where recent research streams can be grouped into deterministic, non-deterministic, and statistical methods shown in the literature review section.

In this research, Hidden Markov Model approach is used for predicting traffic state during peak periods on the UK motorways. This work is based on general HMM theory used in [31,46,27,52] to predict traffic states. Besides, it extends and generalizes the work conducted by Qi and Ishak [32] in using contrast in the context of traffic. Contrast is a second order statistical measure that *is used to make traffic states distinguishable* from each other. Its physical meaning represents the local variation or disturbances in traffic conditions. It is unlike variance; it does not only show the distribution of the data but also shows its orientation. Thus, such a measure is important to enhance the throughput of the transportation system, and improve officers' and travelers' decision making.

The paper starts with a review of the recent research work followed by the introduction of the real-time dataset used in the empirical evaluation. Next, it discusses the HMM-C proposed model and its architecture. Sequentially, the analysis of the experiments, the results and discussions, and the concluding remarks are presented.

☆ Peer review under responsibility of Ain Shams University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.asej.2019.10.006
2090-4479/

Nomenclature

λ	model initial parameters
λ^*	model optimal parameters
μ_{speed}	mean of speed
π	initial probability
a_{ij}	transition probability from state $s_{T-1} = s_i$ to state $s_T = s_j$
A_{NxN}	transition probability matrix
B_{NxM}	emission probability matrix
c_1, c_2	intensity values
M	number of observations
N	number of traffic states
O	set of observations
$P(S_T\|S_{T-1})$	probability of state S_T given state S_{T-1}.
S	collection of random Variables
s_i	a traffic State i
ANN	artificial Neural Network
ARIMA	AutoRegressive Integrated Moving Average
Con_{speed}	contrast of speed
GLCM	gray level co-occurrence matrix
HMM	Hidden Markov model
HMM-C	Hidden Markov model with contrast
ITS	Intelligent Transportation System
KF	Kalman Filter
PCA	Principal Component Analysis

2. Literature review

Traffic prediction has been a hot topic for few decades. Different challenges have been reviewed in Vlahogianni et al. [45,42]. Additionally, researchers have exerted much effort over the years exploring traffic prediction using a multitude of methods. Among the methods are deterministic mathematical methods such as Kalman Filter (KF) and Traffic Time Series. Non-deterministic methods include Artificial Neural Networks (ANN), Fuzzy Logic, hybrid Network-Fuzzy algorithm, and Fuzzy Clustering. Statistical methods include Hidden Markov Models, Regime Switching, and Nonlinear Dynamics. In this section, deterministic, non-deterministic, and statistical research efforts are reviewed as follows:

2.1. Deterministic methods

Okutani and Stephanedes [30] were the first to use Kalman Filtering (KF) method for short-term traffic volume prediction. The prediction performance was enhanced by using data from various links to reach a prediction error of 9%. In another research, Xia and Chen [49] developed a multi-step look ahead short-term corridor travel time prediction model. An embedded adaptive KF was used to adjust the prediction error from real-time data. An approach for real-time traffic flow variables estimation was introduced by Wang et al. [47] based on stochastic macroscopic traffic flow modeling and extended KF. Their algorithm had three main advantages among which was the adaptation to external conditions.

Another deterministic short-term traffic prediction method is based on time series models which started in the 70s by Ahmed and Cook [3]. They used Box-Jenkins to predict the freeway time series of volume and occupancy. Time series then became popular in the late 90s, Hamed et al. [13] developed a Box-Jenkins autoregressive integrated moving average (ARIMA) model to predict future traffic volumes on urban arterials. Recently, Cheng et al. [11] studied the spatio-temporal autocorrelation structure of London's road networks. They proved that forecasting dynamic and local structures is more suitable than global parameters. A different research by Liu et al. [25] integrated Chaotic Time Series Analysis (CTSA) and Support Vector Machine (SVM), and introduced Dynamic Time Wrapping (DTW) and the "Dynamic-K" strategy.

2.2. Non-deterministic methods

As for the non-deterministic models, Zaki et al. [54] developed a time-aware traffic congestion prediction model based on neuro-fuzzy and hidden Markov models. They focused on prediction of traffic patterns during different times of the day and different days of the week. Their best algorithm reduced the error to only 6%. While Chen et al. [10] used neural networks with memory to predict traffic congestion in Beijing, China. The short-long term algorithm used to analyze complex nonlinear variation of speed with promising prediction accuracy. Li et al. [24] reviewed the four major problems facing traffic research including traffic prediction. They focused on Auto-Regressive Moving Average (ARMA) and Feed-Forward Neural Network (FFNN) for traffic Prediction. Another approach by Yin et al. [53] implemented fuzzy-neural model (FNM) to predict the traffic flow in an urban network. They classified the input data into a number of clusters using a fuzzy approach, and specified the input-output relationship using a neural network approach. Shankar et al. [35] used traffic flow information captured from a traffic camera to evaluate congestion using three different fuzzy techniques. They used Adaptive Network-Fuzzy Inference System (ANFIS) model to predict the traffic condition. Abdulhai et al. [2] presented short-term traffic flow prediction based on a combination of both Artificial Neural Network (ANN) and Genetic Algorithm (GA). They explored both spatial and temporal effects on the prediction error. Vlahogianni [41] proposed an ANN prediction scheme for pattern-based evolution of traffic which enhances the prediction accuracy in traffic flow regimes compared to time-series techniques. Further, Abdi et al. [1] presented an approach using improved neuro-fuzzy techniques for short term traffic flow prediction.

Tang et al. [36] used fuzzy C-means with genetic algorithm for imputation of missing data. A matrix structure was used to interpret weekly similarities while Gastaldi et al. [12] used C-means for estimating annual average daily traffic from one week data. The approach used fuzzy sets to define the boundaries of road groups and ANN to assign a road segment to one or more groups. In a different research, Azimi and Zhang [5] used K-means, fuzzy C-means, and CLARA (clustering large applications) to classify freeway traffic flow. Zhu et al. [56] used fuzzy C-means to improve the level of intersection traffic flow with application on West Beijing Road and Xi Kang Road in Nanjing.

2.3. Statistical probabilistic methods

Bouyahia et al. [8] used Markov random field and Markov decision process as two stage method for traffic prediction. Their solution aimed at self-organizing traffic control system. Cetin and Comert [9] used (ARIMA) as the forecasting model to account for regime changes and mixed it with Expectation Maximization (EM) and Cumulative Sum (CUSUM) to detect shifts in the mean level. The percent improvements in the error measures ranged from 44% to 76%. In a recent paper, Kamarianakis et al. [19] accounted for nonlinear traffic dynamics and space-time dependencies by non-overlapping linear regimes with temporal thresh-

olds. They used Least Absolute Shrinkage and Selection Operator (LASSO) to perform model selection and coefcient estimation simultaneously. Recently, Ma et al. [28] implemented regime-switching to avoid unknown structural changes in the time-series used for short-term traffic prediction. The algorithm was robust and effective in modeling the complex multivariate stochastic nature of traffic.

In a different study by Ishak and Qi [17], short term speed prediction is discussed using three different stochastic methods to explore the stochastic nature of traffic during peak periods. They compared the performance of their algorithms using Root Mean Square Error (RMSE). Xia et al. [50] tackled the problem of online traffic state identification using online clustering procedure and optimizing data points through joint utilization of the Bayesian Criterion. In another study, Yang et al. [51] proposed an online adaptive model and extended it into a more flexible state space model. The model was evaluated with data on Interstate 405 near Irvine, California. The results were within the 95% accuracy.

Qi and Ishak [32] introduced contrast for traffic prediction for the first time. They used it for 5 min short term highway congestion prediction. Although the research by Ishak and Qi introduced the contrast as a traffic prediction measure, there are still some unresolved issues such as providing the theoretical generalization of the contrast concept and properties. For instance, their work bypassed whether the contrast value range changes with the change of the dataset. They skipped addressing the contrast suitability for aggregated data or longer sample periods. Moreover, they did not experiment if contrast still holds as a traffic prediction indicator with longer prediction horizons.

The novelty introduced in this work addresses the previous shortcomings and broaden the perspective towards the generalization of contrast use for traffic prediction with HMM. Contrast is extended to 90 min short-term prediction horizon with different range of contrast values at a sampling interval of 15 min. Thus, it is shown that contrast is a good measure that still holds for aggregated data. Additionally, a more practical set of 9 traffic states is defined compared to Qi and Ishak [32] 30 traffic states. They have clearly indicated that a number of the 30 states rarely appear in practical traffic patterns if any. Finally, the model is proven to be robust due to the right choice of training data.

Based on the literature review introduced here, one can see that there are different methods proposed to tackle the prediction problem. These methods belong to the deterministic, non-deterministic and statistical models. The deterministic mathematical models such as time series usually provide good results for linear stable traffic conditions while they fall back under disturbances, unstable or complex stochastic conditions. This has given the space for the evolutionary non-deterministic methods such as ANN or Fuzzy sets to take place. They are very successful in classification and prediction due to their embedded learning process, ability to deal with vagueness, and their interdisciplinary application. They also provide a cost effective solution when the mathematical solution is very costly. Statistical models such as HMM which are based on Bayesian inference, although dependent on the quality of sample data, they are arguably better in predicting nonlinear stochastic processes where randomness is present. Nonetheless, one has to highlight that the accuracy versus simplicity battle between statistical and non-deterministic models remains and the need for a consistent prediction measure that is intuitive enough to help decision makers is still required.

3. The dataset

This research is based on data collected from the Highways in England. They are commonly referred to as the motorways and are divided into six regions that are roughly based on the regions of England. Those are South West, London & South East, East, Midlands, North West and North East. The network is composed of 4400 miles of major motorways in England and accounts for only 2% of all England's roads [14].

The Highway Agency has made public "Network Journey Time and Traffic Flow Data". The data is a year on year statistics from 2009 to date. It is publicly available in monthly ".CSV" files [39]. Each monthly file contains roughly 7 million records of data for 2499 junctions on the motorways. The data resolution is 15 min for all the junctions resulting in 96 readings per junction per day (2976 readings per junction per month). The content of a monthly data file is similar to the sample below in Table 1 and the fields are explained in Table 2.

4. The proposed HMM-contrast model (HMM-C)

In this paper, HMM will be utilized for traffic prediction because its structure matches traffic behaviour. In HMM, only the observations can be seen and a probabilistic inference is to be made regarding states of the system which are unknown (hidden) traffic states. Rabiner [34] declared HMM as a doubly stochastic process where the stochastic underlying process (in this study the traffic state) is only observed through a sequence of observation which is another stochastic process (i.e. traffic speed). The dynamics of the stochastic hidden process known as the traffic state are captured into HMM State Transition Matrix and the dynamics of the observation are captured into HMM Emission Matrix.

4.1. Hidden Markov model

Since traffic is complex in nature, the following simple example matches traffic nature to the HMM definition and structure. Assuming a vehicle is traveling at a speed of 40 km/h. This speed can be observed at a free-flow traffic condition, medium flow condition, and a breakdown or recovery from traffic congestion. That is, the speed as a point measurement does not show the traffic state which is hidden. Therefore, traffic states must be defined over a time interval using statistics. The use of such statistics enables the variation and the trend of traffic parameters to be observed. In other words, the traffic conditions are unknown which comply to the HMM hidden states and can be captured into a traffic state transition matrix. The speed is known which conforms to HMM observation and can be captured into a traffic emission matrix. Consequently traffic can be modeled using HMM. The following steps describe the application of the HMM model to the traffic problem in hand:

1. Define the model states and observation based on data characteristics and analysis.
   Markov chain is a mathematical system of a random process which is a collection of random variables $S = \{S_1, S_2, S_3, \ldots, S_N\}$, where N is a finite or infinite countable number of traffic states [15], that undergoes random transitions over time $t \in T$ from one traffic state, say s_i, to another traffic state, say s_j, between a certain number of possible traffic states where s_t is the state of the system at time t [23]. If the time t is an interval, the process is a continuous Markov process, while it is a discrete Markov process if $t \subset Z$ [34] where Z is the integers. The above model is an observable Markov Model where the output at each time instant is a set of states corresponding to a physical observable event [34]. For an HMM, the state of the system is unknown and only an event associated with that state is observed. This event is a probabilistic function of the unknown state. Defining the set of distinct observation as

Table 1
Sample Traffic Data.

LinkRef	Date	TimePeriod	AverageJT	AverageSpeed	LinkLength	Flow
AL1260	01/12/2014	0	114	82.11	2.6	10.75
AL1260	01/12/2014	1	113.71	82.31	2.6	10.25
AL1260	01/12/2014	2	127.19	73.59	2.6	8.5
AL1260	01/12/2014	3	110.75	84.51	2.6	6.5
AL1260	01/12/2014	4	124.7	75.06	2.6	5.5
AL1260	01/12/2014	5	124.31	75.3	2.6	5
AL1260	01/12/2014	6	123.76	75.63	2.6	5.5
AL1260	01/12/2014	7	121.26	77.19	2.6	5.5
AL1260	01/12/2014	8	115.38	81.12	2.6	5.75
AL1260	01/12/2014	9	120.23	77.85	2.6	6.5
AL1260	01/12/2014	10	132.81	70.48	2.6	7.5

Table 2
Meaning of Column Headers in Data Files [14].

Variable name	Variable description
LinkRef	A unique alphanumeric link id representing a junction to junction link
Date	Date of travel
TimePeriod	One of 96 15-min intervals in the day (0–95 where 0 indicates 00:00 to 00:15)
AverageJT	The average journey time to travel across the LinkRef in seconds
AverageSpeed	The average speed (km/h) of vehicles entering the link within a given 15-min time period
Link Length	The length of the link (km)
Flow	An average of the observed flow for the link, time period and day type

$O = \{O_1, O_2, O_3, \ldots, O_M\}$ where M is the number of observations. This means, given the Markov model is in a particular hidden traffic state, the connections between the hidden traffic states and the observation represent the probability of generating a particular observation or event.

2. Define the model initial parameters λ (initial transition probability, state transition matrix, and emission matrix) by making an initial stochastic estimation.
The system has the "memoryless" property where the future traffic state conditional probability is independent of all the past traffic states given the present state $P(S_t|S_{t-1}, S_{t-2}, S_{t-3}, \ldots S_1) = P(S_t|S_{t-1})$. The $P(S_t|S_{t-1})$ is called the traffic state transition probability $A_{N\times N}$ where a_{ij} from state $S_{t-1} = s_i$ to state $S_t = s_j$.

$$a_{ij} = P(S_t = s_j|S_{t-1} = s_i), 1 \leqslant i,j \leqslant N \tag{1}$$

which obeys the following standard probability constraints:

$$a_{ij} \geqslant 0 \tag{2}$$

$$\sum_{j=1}^{N} a_{ij} = 1 \tag{3}$$

$$\pi = P(S_t) = s_1 \tag{4}$$

The probability matrix denoted as π is a special type of the traffic state transition probability which represents the initial probability. In addition, the probability of generating a particular observation given a particular hidden traffic state is held in the emission matrix $B_{N\times M}$ where b_{jk} is:

$$b_{jk} = P(O_k \ at \ t|S_t = s_j) 1 \leqslant j \leqslant N, \ 1 \leqslant k \leqslant M, \ t = 1,2,3,\ldots \tag{5}$$

For convenience, the HMM model is usually referred to by the notion $\lambda = \{A, B, \pi\}$ where A is the traffic state transition matrix, B is the emission matrix, and π is the initial probability.

3. Training the HMMs to find the optimal $\lambda^* = \{A^*, B^*, \pi^*\}$ by iterative estimation till convergence using maximum likelihood and expectation maximization.
In any HMM construction, there are three questions that must be answered:
   - What is the probability of generating an observation sequence? That is, given an observation sequence $O = \{O_1, O_2, O_3, \ldots O_T\}$ and initial model parameters $\lambda = \{A, B, \pi\}$ what is the $P(O|\lambda)$?
   - How to find the initial model parameters? That is, given an observation sequence $O = \{O_1, O_2, O_3, \ldots O_T\}$ what are the initial optimal model parameters $\lambda^* = \{A^*, B^*, \pi^*\}$ that maximizes the probability $P(O|\lambda)$?
   - Given a set of observation sequence $O = \{O_1, O_2, O_3, \ldots O_T\}$ and the HMM parameters $\lambda = \{A, B, \pi\}$, what is the optimal traffic state transition sequence?

The answers to these questions are available in [34]. These answers will provide the algorithm for solving a standard HMM problem and will help applying HMM to traffic.

4. Predict the hidden traffic states using Viterbi dynamic programming algorithm.
The focus now is shifted to finding the optimal traffic state sequence associated with a given observation sequence. That is, predicting the optimal traffic state sequence. As defined by Rabiner [34] there could be more than one optimal criterion. There could be an optimal criterion where the *individual states* are most likely. Another criterion where the optimal solution could be optimizing for more than one state; a state sequence. The Viterbi algorithm [40] was implemented to find the optimal traffic state sequence $S = \{S_1, S_2, \ldots, S_T\}$ for a given observation sequence $O = \{O_1, O_2, \ldots, O_T\}$.

The interested reader can have more insight into the theory behind HMM from [34,40,6].

4.2. Contrast definition and mapping

In this study, statistics are required to find the hidden traffic states. Such statistics are defined over a rolling time window composed of a number of observations. The rolling time window rolls forward one step at a time where it drops the oldest observation and captures one new observation. The statistics are calculated over the time window for every forward move. This way, the traffic trend and changes are captured over a short period of time. Mean speed, standard deviation and contrast (defined in the next paragraph) are among the statistics that could be used. Mean speed μ_{speed} is selected to capture the central tendency of the observation while the contrast CON_{speed} and standard deviation σ_{speed} are used to measure the variation in traffic state. The same statistics are calculated for the density of traffic $\mu_{density}$, $CON_{density}$,and $\sigma_{density}$. Combining the central tendency and the variation variables together

provide better confidence in the traffic state classification and prediction [16]. A graph and sample example is shown in Fig. 1.

While mean and standard deviation are well defined variables, contrast in the context of traffic is not. Contrast is a second order statistical measure. Its physical meaning is attributed to the degree of local variation of traffic states which indicates the severity of disturbances in a sequence of observations. It makes traffic states distinguishable from each other. It exploits the underlying relations among the different states of traffic. High contrast values indicate severe local disturbances which may be attributed to abrupt changes in demand or capacity, heavy weaving, or merging maneuvers. Low contrast values are attributed to change under stable conditions such as stable free flow or stable congestion condition. Additional elaboration around the properties and advantages of using contrast are demonstrated under the results and discussion section.

Contrast was first developed in the field of image processing as a measure of the intensity between a pixel and its neighbour [37]. It shows the local variation of a gray level in the image where both the contrast and its reverse are identical. Contrast is calculated from the second order histogram, the gray level co-occurrence matrix (GLCM) [32]. The GLCM is created by finding how often a data point with gray level (intensity) value c_1 occurs in relative distance d to a data point with intensity value c_2. In an image, this distance could represent a pixel and its consecutive pixel if $d = 1$ or two pixels apart if $d = 2$. Moreover, an image is a 2-D matrix. Therefore, the orientation of the pixel ϕ is taken into consideration. For instance, horizontal orientation between a pixel and its immediate right side discussion pixel will have a $\phi = 0$ or immediately above will have a $\phi = 90$ degrees.

$$CON = \sum_{c_1,c_2} |c_1 - c_2|^2 a_{c_1c_2} \tag{6}$$

where

$$a_{c_1c_2} = \frac{number\ of\ pairs\ at\ distance\ d\ with\ gray\ level\ (c_1, c_2)}{total\ number\ of\ possible\ pairs} \tag{7}$$

In traffic context, CON defined in Eq. 6 is redefined according Eq. (8) where it is possible to analyze the data evolution with increasing or decreasing trend [33] by incorporating the positive and negative signs of GLCM. The terms c_1 and c_2 in traffic context represent two different speed vectors. The term $a_{c_1c_2}$ is defined according to Eq. 7 and accumulates the total occurrence of each speed pair combination. Since traffic data points are measured over time, the orientation will be quantified in the horizontal direction with $\phi = 0$ and only consecutive data pairs with distance $d = 1$ will be considered. Thus CON will not only provide information regarding the distribution of data but also regarding the relative location and orientation of the data.

$$CON = \sum_{c_1,c_2} |c_1 - c_2|(c_1 - c_2) a_{c_1c_2} \tag{8}$$

Assuming a sequence of 10 traffic speeds shown in Fig. 1 [33] $O = [10\ 10\ 7\ 10\ 8\ 9\ 10\ 10\ 8\ 8]$, the output element $a_{10,10}$ contains the value 0.22 because there are two instants in the dataset with adjacent data points having the value of 10 and 10 respectively while the total number of pairs is 9 as shown in Fig. 1 where every two consecutive numbers in the sequence is one pair.

The contrast following Eq. (8) is calculated as

$$\begin{aligned} CON &= (9-8)^2 \times 0.11 + (10-9)^2 \times 0.11 + (10-7)^2 \times 0.11 \\ &\quad -(10-7)^2 \times 0.11 - (10-8)^2 \times 0.22 \\ &= -0.67 \end{aligned}$$

Contrast is then used in conjunction with speed to predict traffic states. Both the speed and contrast are used together to construct a 2D space. Such a 2D space maps different traffic states from different clusters which is discussed in Section 5.3.1 to Section 5.3.5. The mapping provides an insight into the transformation of a traffic observation (speed) into intelligence regarding the traffic hidden states (free flow, congestion, . . .etc). Thus, utilizing the contrast to predict the traffic states. By definition, contrast represents the variation in traffic states. Following Eq. (8), if there is no or little variation between a traffic state and its adjacent state, the resulting contrast is zero or small values respectively, given $a_{ij} < 1$ according to Eq. 7. In opposition, severe changes in the traffic states correspond to large contrast values.

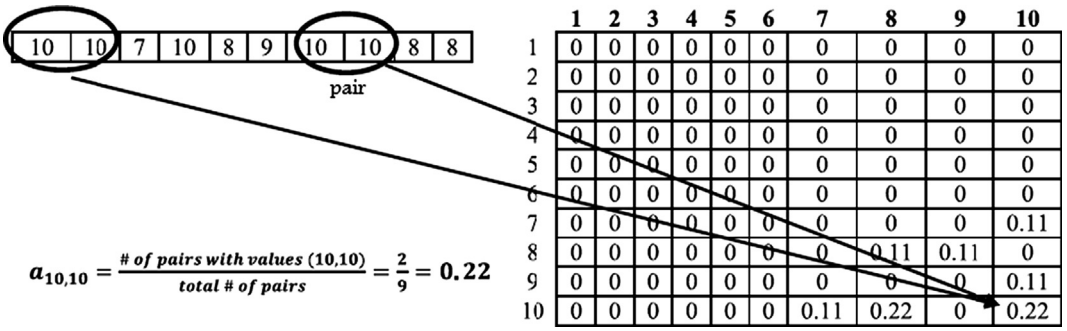
5. HMM-C Model High Level Architecture

Fig. 2 is called HMM-C owed to HMM and Contrast. It shows a diagram that depicts the architectural components of the HMM-C methodology presented in Section 4. These components are discussed in the following sections with the specified output of each component as shown Fig. 2. The first component (Data Preprocessing) summarizes the data operations while the component (Traffic Variable Definition) selects the right traffic variables. The third component (Traffic Pattern Classification) determines the number of traffic clusters to assist in finding the traffic states of the system. Components (HMM Training and Testing) depict the actual methodology discussed in Section 4.

5.1. Data preprocessing

5.1.1. Data preparation

When the data was initially received, SQL queries under parallel Perl scripts were applied to sort, group, and select junctions to serve the purpose of this research. Since this research is using a statistical model to predict traffic congestion, it is largely dependent on the quality of data and the traffic profile this data represents. For example, a profile of a certain junction does not have any sizable congestion pattern or another junction profile that only contains slow speeds; both could bias the study. Therefore, smart



	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0.11
8	0	0	0	0	0	0	0	0.11	0.11	0
9	0	0	0	0	0	0	0	0	0	0.11
10	0	0	0	0	0	0	0.11	0.22	0	0.22

Fig. 1. GLCM for a speed sequence [32].

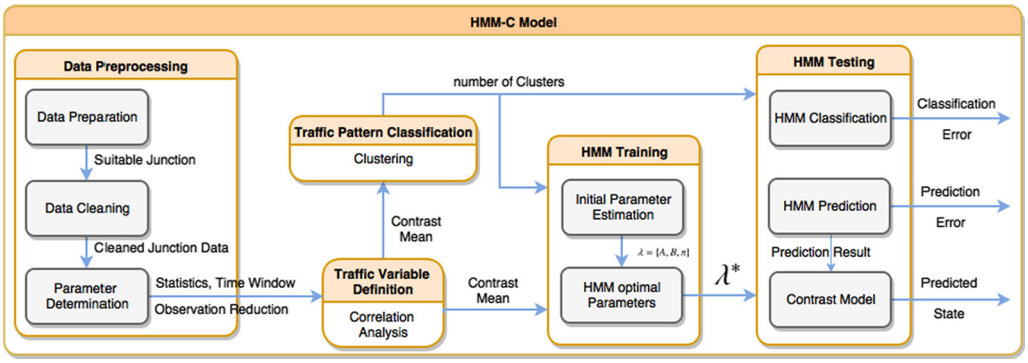


Fig. 2. HMM-C Model High Level Architecture.

routines were developed to qualify the junctions based on the data profile available to suit the study at hand. Each junction data was plotted against the theoretical relationship defined by Kerner [21]. The choice of the junction is based on the condition that the junction traffic profile covers all traffic conditions as defined by Kerner [21].

This is apparent from Fig. 3 showing flow-density curve and speed-flow curve that coincide with the theoretical curves defined by Kerner [21] and covers all different traffic conditions. One year data for a junction called "AL1260" representing the A453 between A50 and A42 is chosen for this research. Additionally, similar amount of data for another junction "AL2701" between A45 and A46 is used for verification of the algorithm. A standardization of the units is applied to allow for the calculation of additional variables and solid analysis.

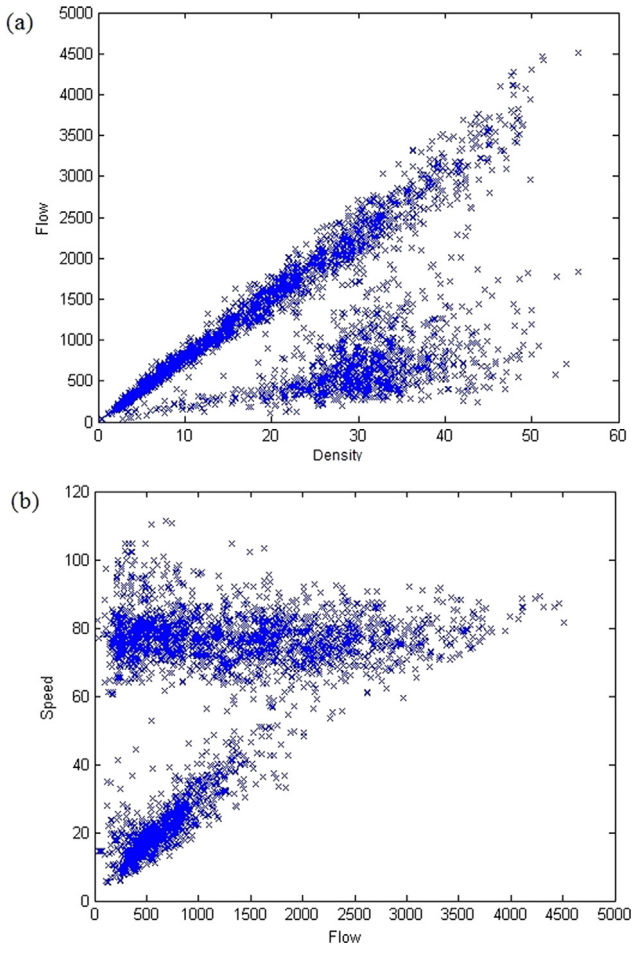


Fig. 3. Flow-Density and Speed-Flow Relationships for AL1260.

The application of HMM in traffic is based on an assumption that the traffic states are stationary over time. As it stands, traffic is nonlinear and non-stationary time series. It exhibits various deterministic and nonlinear processes such as trends, abrupt changes, seasonality, and oscillating behaviour. Different methods, such as differencing, are used in attempt to "stationarize" the time series. Other methods use Recurrence Plots (RP) and Recurrence Qualitative Analysis (RQA) in order to be able to analyse and predict traffic behaviour. Such methods are beyond the scope of this work. One widely adopted approach in the literature, although does not guarantee stationary characteristics, is the segmentation of non-stationary time series [4,7] and searching for stationarity in shorter time periods as discussed by Vlahogianni et al. [44]. That is, dividing the traffic into shorter periods, say peak and non-peak periods. For argument sake, five periods a day; early off-peak period, morning peak, noon off-peak, evening peak, late evening off-peak. The bounds of those periods depend on the experience and analysis. It is assumed that the traffic system is stationary during those periods. Thus, the transition of states is stochastic over time. Generally, assumptions are not investigated a priori, they are validated later based on the prediction success [44]. In this research, HMM is applied to traffic morning peak period on the freeway since peak periods experience more breakdowns and congestion. The morning peak period is taken from 5:00 a.m. to 10:00 a.m. The data points for these periods are selected to cover weekdays and exclude both weekends and public holidays. This reduces the pattern variation and increases confidence level in the data since both weekends and public holidays have different commuting patterns than that of weekdays.

5.1.2. Data cleaning

After the data is selected, it is cleaned from outliers and other errors. The data is originally collected using loop detectors which may randomly malfunction resulting in missing, duplicate or incorrect data. Investigating the data resulted in no missing or duplicated records. This is because the data has been aggregated by the UK Department of Transport [38]. That is, it has been preprocessed before. Therefore, the data has no empty or duplicated records which ease out the problem of missing or duplicated data. It may have also been processed for outliers. However, this study uses the mean speed which is not robust to outliers. Taking for instance, the speed on the UK motorways is 70 miles/h which is roughly 112 km/h. Therefore, any data point that is more than 120 km/h is truncated to 120 giving a margin (8 km/h or 5 miles/h) over the maximum motorway speed. Such a margin is important considering in reality people may increase speed for

overtaking, have faulty or not calibrated speedometers, or simply drive faster than the speed limit for a period of time or at a certain stretch of the road. On the other hand, a single point of extremely low value may or may not be an outlier. If traffic appears to be in free flow state where speed is greater than 60 km/h, a single point of extremely low value less than 10 km/h is considered an outlier. The outlier is then replaced by the average of the two points around it. However, this point is not an outlier if traffic is in stop and go condition.

5.1.3. Parameter determination

- Time Window and Statistics Calculation: Subsequent to the data cleaning, statistics are calculated. A rolling time window of 90 min is used to define the statistics for short-term traffic prediction. Since the data resolution is 15 min and the period under consideration is 5 h, the 90 min window is good enough to capture traffic trend while keeping details about the change of traffic states during the 5 h peak period. Each 90 min window contain 6 points of 15 min measurements generating 16 speed vectors within the 5 h peak period. It rolls forward one step at a time dropping the oldest observation and capturing a new one. Thus enabling the statistics calculations over the time window for every forward move while keeping information regarding the change of states. Variables containing trend and state are mean (μ), standard deviation (σ), and second order statistics contrast.
- HMM Observation Reduction: The speed limit on the UK motorways is 70 miles/h which is roughly 112 km/h while during the data cleaning process, a threshold of 120 km/h was applied to the data. This threshold floors any extreme points as explained in Section 5.1.1. In HMM terms, the speed range is $0 < v \leqslant 120$ km/h means that there are 120 different observations. This results in a large emission matrix leading to excessive computational requirements. To reduce the complexity, the speed is split into equally spaced ranges of 5 km/h reducing the size and complexity of the calculations. Thus reducing the 120 different observations into 24 different observations as shown in Table 3. For instance, any speed between 30 and 35 km/h is represented by number 7 in the calculation as shown in Table 3.

5.2. Traffic variable definition

On the grounds that it is not necessary to use all the statistical variables defined earlier to determine the traffic state, Principal Component Analysis (PCA) [18] is used to define the most relevant variables to determine the traffic condition owing to the fact that some variables may be correlated. When the PCA is applied, it results in components arranged in a descending order of variance. That is, the first component contains the largest variability of the dataset.

As shown in Table 4, it is clear that the first 2 principal components account for nearly 95% of the variation in the dataset while the first 3 principal components account for almost 98% of the variation. The remaining 3 components contribute to only 2% of the variation. Thus the first 3 components provide good representation of the dataset. By further expanding the analysis, it is clear from Table 5 that μ_{speed}, CON_{speed}, and $\mu_{density}$ have high loading corresponding to the first 3 components PC1, PC2, and PC3. This means that those 3 variables store most information regarding the dataset than the remaining variables. Hence, they are selected for further analysis.

To ensure that the statistical data used in HMM is reliable, correlation analysis is conducted on the grounds that correlated variables may affect the accuracy of the model. The Pearson correlation analysis in Table 6 shows the correlation among each pair of variables under consideration. Since the PCA analysis conducted above resulted in the interest of μ_{speed}, CON_{speed}, and $\mu_{density}$ variables only, it can be seen that only $\mu_{density}$ has a correlation > 0.5 with μ_{speed} while the other variables have < 0.5 correlation between each pair. With such a result in mind, one of the two variables $\mu_{density}$ or μ_{speed} can be used with CON_{speed} for traffic condition prediction. Herein, the μ_{speed} has been chosen as the observation symbol for HMMs. The correlation analysis discussed above is shown in the second component (Traffic Variable Definition) in Fig. 2.

Table 4
Eigen values.

PC	percentage EV	Commulative EV
PC1	**0.779**	**0.779**
PC2	**0.164**	**0.944**
PC3	**0.036**	**0.9805**
PC4	0.019	0.9998
PC5	0.0001	1
PC6	0	1

Table 5
Eigen vectors.

	PC1	PC2	PC3	PC4	PC5	PC6
μ_{speed}	0.02	**-0.8731**	**-0.4656**	0.1257	-0.0636	-0.0231
CON_{speed}	**-0.9995**	**-0.0293**	0.0110	0.0006	0.0003	-0.0001
σ_{speed}	-0.0001	-0.0009	0.0020	0.0023	-0.0037	-0.0005
$\mu_{density}$	0.0212	**-0.4554**	**0.7149**	-0.3241	0.4121	-0.0782
$CON_{density}$	-0.0004	-0.0374	-0.0653	-0.8082	-0.5786	-0.0786
$\sigma_{density}$	0.0021	-0.0610	0.0465	-0.0809	-0.0232	0.9934

Table 6
Pearson correlation analysis.

	μ_{speed}	CON_{speed}	σ_{speed}	$\mu_{density}$	$CON_{density}$	$\sigma_{density}$
μ_{speed}	1.00	0.00	0.16	0.57	0.18	0.56
CON_{speed}	0.00	1.00	-0.17	-0.16	0.05	-0.06
σ_{speed}	0.16	-0.17	1.00	0.59	-0.08	0.33
$\mu_{density}$	**0.57**	-0.16	0.59	1.00	0.78	0.64
$CON_{density}$	0.18	0.05	-0.08	0.78	1.00	0.27
$\sigma_{density}$	0.56	-0.06	0.33	0.64	0.27	1.00

Table 3
Observation reduction: speed ranges.

HMM Traffic Symbol	Speed Range (km/h)	HMM Traffic Symbol	Speed Range (km/h)	HMM Traffic Symbol	Speed Range (km/h)
1	$0 < v \leqslant 5$	9	$40 < v \leqslant 45$	17	$80 < v \leqslant 85$
2	$5 < v \leqslant 10$	10	$45 < v \leqslant 50$	18	$85 < v \leqslant 90$
3	$10 < v \leqslant 15$	11	$50 < v \leqslant 55$	19	$90 < v \leqslant 95$
4	$15 < v \leqslant 20$	12	$55 < v \leqslant 60$	20	$95 < v \leqslant 100$
5	$20 < v \leqslant 25$	13	$60 < v \leqslant 65$	21	$100 < v \leqslant 105$
6	$25 < v \leqslant 30$	14	$65 < v \leqslant 70$	22	$105 < v \leqslant 110$
7	$30 < v \leqslant 35$	15	$70 < v \leqslant 75$	23	$110 < v \leqslant 115$
8	$35 < v \leqslant 40$	16	$75 < v \leqslant 80$	24	$115 < v \leqslant 120$

5.3. Traffic pattern classification

Following to the selection of variables to be used, K-means clustering analysis is performed using Ward method and applied using Matlab® to group similar observations and to show the evolution of different states. It partitions the observations based on their distance from each other into a number of clusters. Each cluster has a centroid where the observations within this cluster are close to each other as possible and far from data points in other clusters as possible. The main question in clustering analysis is: *what is the optimal number of clusters to be used?* To answer such a question, within clusters sum-of-squares is employed to find the optimal number of clusters and the result is shown in Fig. 4(a). Sum-of-Squares method shows local minimas at 3 and 5 indicating the optimal number of clusters. This result is verified using Silhouette internal evaluation. In the verification, {3,4,5 and 6} clusters are tested iteratively. The test result is shown in the Silhouette plot Fig. 4(b). The bigger the silhouette value, the more points that better belong to this cluster than any other cluster. If the silhouette value has a negative sign, it means there are some points classified into one cluster while they probably belong to another cluster. From Fig. 4(b), one can argue that 3 clusters may be the optimal numbers of clusters to be used due to lack of negative values. However, 5 clusters show larger positive Silhouette values across all

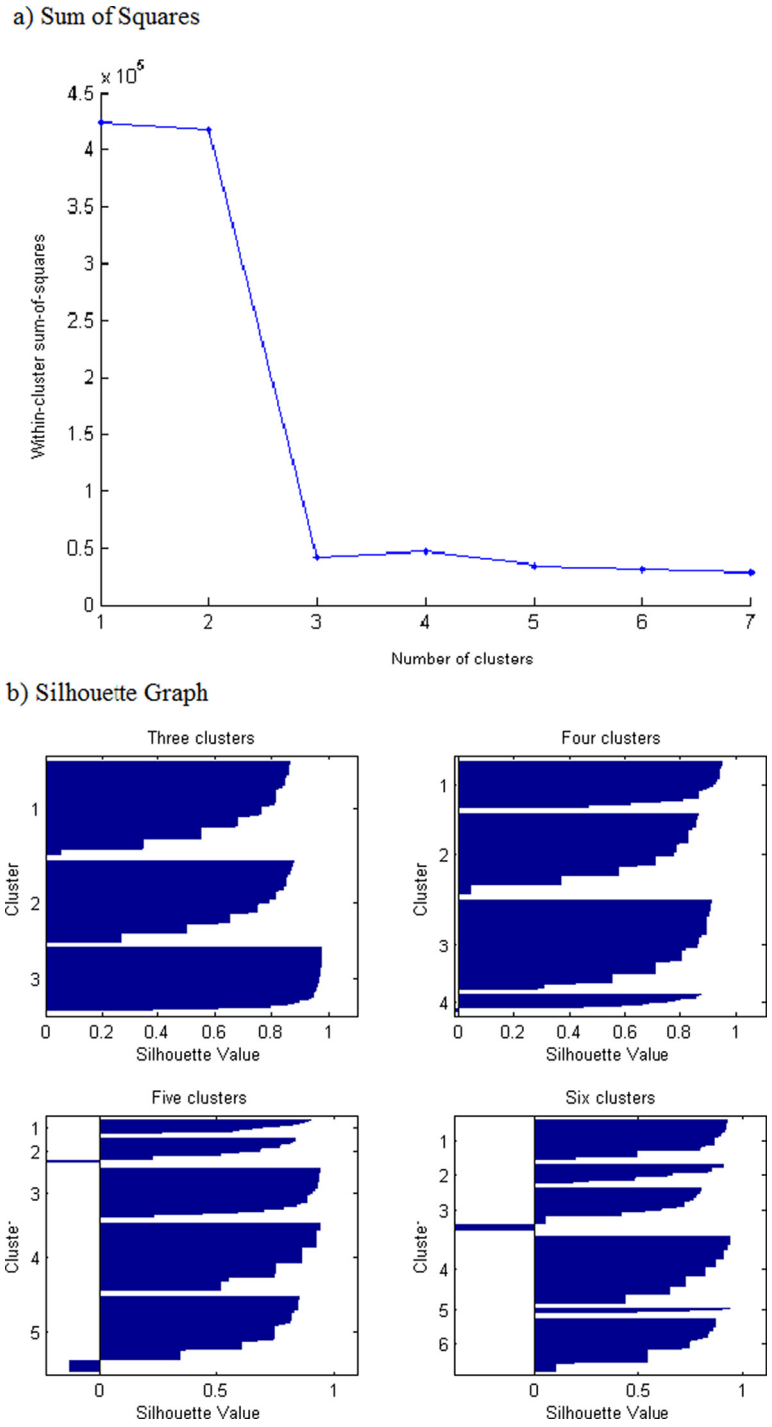


Fig. 4. Evaluating the Right Number of Clusters.

clusters than those of 3 clusters. Comparing the result from sum-of-squares and Silhouette evaluation to the gold standard, which is human judgement in this case, Fig. 5 is required. It shows the difference between 3 and 5 clusters for the same data. That is, clustering into 5 clusters provide a better explanation of traffic evolution than 3 clusters. Since each cluster contains a number of traffic state evolution over time and collecting the arguments from sum-of-squares result in Fig. 4(a), Silhouette result in Fig. 4(b), and human judgement in Fig. 5, one can see that 5 clusters is a better clustering candidate. The output of the component (Traffic Pattern Classification) in Fig. 2 is the number of clusters used for HMM construction. Moreover, the explanation of the 5 clusters is discussed as follows:

5.3.1. Cluster 1 - from free flow to congestion

Traffic speed shown in Fig. 5 (b - Cluster 1) represents cluster 1 data. The cluster illustrates a traffic condition starting at free flow with speed between 60 and 80 km/h. Then the speed drops dramatically to 20 km/h indicating a deterioration of traffic condition from free flow to congestion. This happens when the traffic on the highway is less than the road capacity (free-flow), then entering a peak period the traffic demand increases till it reaches the capacity of the highway causing a breakdown (congestion). This is apparent by the rapid drop of speed.

5.3.2. Cluster 2 - free flow

Traffic speed illustrated in Fig. 5 (b - Cluster 2) shows a traffic condition of free flow. The traffic enters a period at speed around 80 km/h and continues at the same speed during the full period. This means the traffic is stable under free flow condition. Vehicles flowing smoothly through the link with no congestion. Such a situation happens when the demand capacity is far less than the highway capacity with no apparent incidents or congestion.

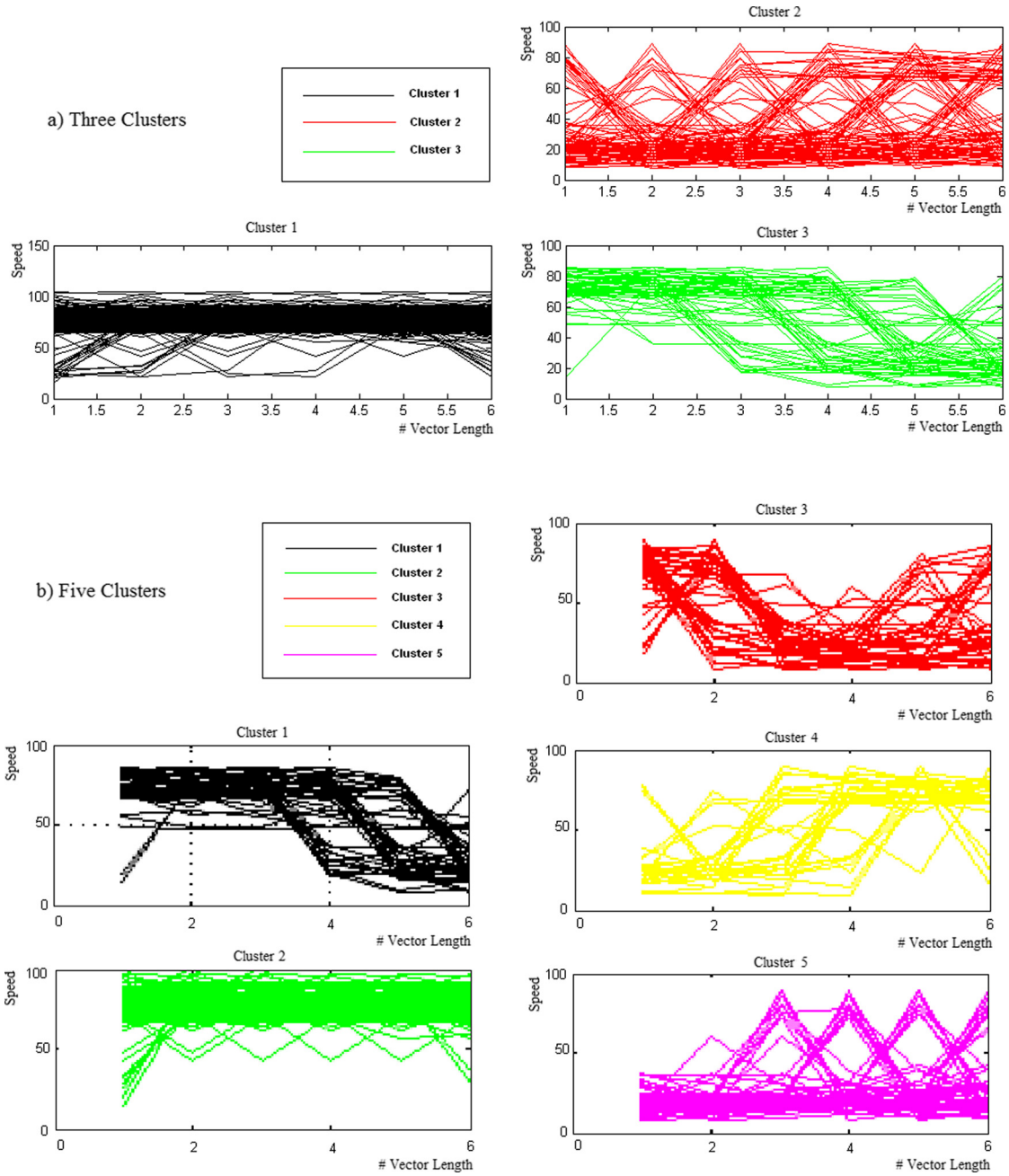


Fig. 5. Three VS Five Clusters.

5.3.3. Cluster 3 from free flow to congestion and back

The peak period traffic speed profile presented in Fig. 5 (b - Cluster 3) shows traffic starting at free flow speed of 60 km/h. Then the speed deteriorates till it reaches 20 km/h. It stays there for a period of time, 30–60 min according to Fig. 5 (b - Cluster 3), then it picks up again. It recovers to a speed of 40 km/h lower than the free flow speed. There is a lot of speed fluctuation around 20 km/h indicating an unstable traffic condition. This unstable condition could be due to driving behaviour such as a stop-and-go. This does not indicate any stable traffic condition. Cluster 3 is a typical peak period condition where the traffic starts off at a free flow speed, then it deteriorates to congestion speed, later it recovers towards free flow. Hence, there is a slow flow level despite the congestion and could be named as a moderate congestion condition or medium traffic flow.

5.3.4. Cluster 4 - from congestion to free flow

The traffic speed in Fig. 5 (b - Cluster 4) shows congestion and a trend for recovery. This means the traffic during the peak period exceeds the highway capacity. Nevertheless, there is a slow flow with low service level till it recovers. Since the peak period under consideration is 5 h while the trend shown represents 90 min (6 measurements spanned by 15 min each), this period of congestion to recovery may fall at the end of the 5 h peak period. The speed after recovery is the free flow speed around 70 km/h.

5.3.5. Cluster 5 - congestion

The last cluster shown in Fig. 5 (b - Cluster 5) demonstrates congestion traffic state. The peak period starts at 20 km/h with some speed vectors fluctuation every 15 min. The average speed during the peak period is around 25 km/h. This means that congestion lasts for a long time of 90 min window represented in this graph. The low level of service indicates heavy demand that exceeds the highway capacity or an incident that took time to clear.

Comparing the clustering results with those in the literature, Lumpe and Vo [26] clustered Melbourne, Australia traffic pattern into two and three clusters. Their research proved the meaningful cluster decomposition is between 2 and 5 clusters. Their two and three clusters did not show sufficient details regarding traffic conditions. Qi and Ishak [32] clustered traffic into 5 clusters ranging from free flow to congestion. Zhang et al. [55] clustered traffic pattern into 4 different clusters of Heavy Jam, Light Jam, Jam, and smooth. They used K-means, C-means, and grey relational analysis for clustering of traffic data.

The clustering analysis shown in this section represents 5 different traffic patterns. These patterns cover almost all possible traffic states. It is important to note that traffic evolution does not follow one pattern and there is no one rule to capture the variations or the different trends. It is not known when a congestion will happen, how long it will last, or when it will recover which emphasises the fact that non-deterministic probabilistic models such as HMM are suitable for the problem of traffic prediction.

5.4. HMM training

The successive step to clustering is the HMMs construction and training. In the construction, there must be a number of HMMs equal to the number of clusters; one for each cluster. The dataset is split into 70% and 30% for both training and validation of HMM respectively. To predict the traffic state, three time periods are required in HMM construction as shown in Fig. 6. The first period is the length of the prediction horizon. That is, how far in the future the sequence of traffic states (hidden states) is predicted. Since the statistics were defined earlier over 90 min horizon, this horizon is used to predict the traffic state. The second period is how often the prediction is updated. This is based on the 15 min data resolution. The prediction can be updated once new data becomes available. Thus, it cannot be updated in time less than 15 min. In fact it can be multiples of 15 min. The third period is the actual data resolution of 15 min which represents the state transition. This is summarized in component (HMM Training) in Fig. 2.

5.5. HMM testing

After the training of the HMMs, they are utilized to find the optimal traffic state (hidden state) sequence for an observed sequence of traffic speeds. A concept similar to Model Predictive Control (MPC) receding horizon [29] is used to define the prediction horizon of HMM. Knowing the current speed at the beginning of a 90 min, the traffic state at the coming 90 min is unknown. Building a sequence of previous observations, a horizon (sequence) of traffic speeds is formed. Each speed point of the horizon corresponds to a hidden traffic state. One can find the optimal traffic state sequence using HMMs for the sequence of the available speeds. Since each cluster may contain more than one state as discussed in Section 6.1. The prediction of states is achieved through finding the optimal cluster first and the HMMs find the optimal state sequence within the optimal cluster. This optimal sequence

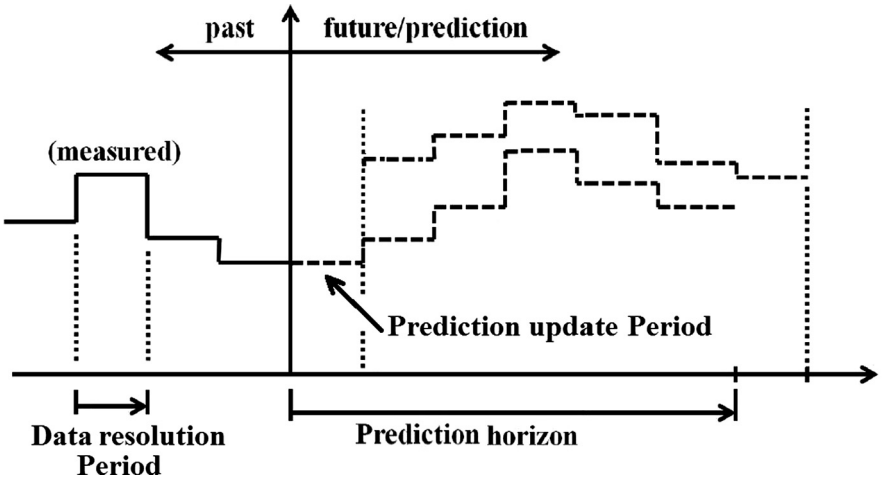


Fig. 6. Time Periods used for prediction.

is found using Viterbi Algorithm [40] and implemented using Kevin Murphy toolbox in Matlab. The prediction operates in a way that the optimal cluster is found based on the log-likelihood of the sequence. The last point of the optimal state sequence is the predicted traffic condition for the coming (90 min) window associated with the last traffic speed observation of the traffic speed sequence. The algorithm finds the maximum likelihood of the next state through the transition matrix by means of induction. This is based on the emission probability of each state and the corresponding cluster. The algorithm is repeated according to the rolling window of 15 min and the coming 90 min status is updated every 15 min. The component (HMM testing) in Fig. 2 shows the prediction method as described in Section 4.

6. Results and discussion

Herein, the speed and contrast are discussed to predict traffic states. Both the speed and contrast are used together to construct a 2D space representing different traffic states. Fig. 7 (from a to e) represents 2D mapping of the speed and contrast relationship from different clusters discussed in Section 5.3.1 to Section 5.3.5 and shown in Fig. 7 (f). The mapping provides an insight into the transformation of a traffic observation (speed) into intelligence regarding the traffic hidden states (free flow, congestion, . . .etc). Thus, utilizing the contrast to predict the traffic states. By definition, contrast represents the variation in traffic states. Following Eq. (8), if there is no or little variation between a traffic state and its adjacent state, the resulting contrast is zero or small values respectively, given $a_{ij} < 1$ according to Eq. 7.

Generally speaking, such a condition would appear in Fig. 7 (from a to e) represented with the zero contrast line for zero variation. The area bounded between the left and right red lines represents little variation. The exact bounds of such an area are shown in Table 7 later in this section and based on the researcher experience, analysis, and the characteristics of the dataset. The areas beyond the red lines on either side have larger contrast values and hence, they represent large variation in the state of traffic. The magnitude of the contrast whether negative or positive is affected by the change in traffic behaviour. Whenever, there is drop in speed, the contrast will have a negative value indicating a change to worse traffic condition. Similarly, an increase in traffic speed would have positive contrast value indicating a better change of traffic condition. The detailed explanation of the suggested mapping is discussed below for various clusters.

6.1. Clusters 2D speed-contrast mapping

Investigating cluster 1 in Fig. 7 (f), the speed is mostly at free flow then it drops to congestion speed near the end of the curve as described in Section 5.3.1. When people are driving at free flow speed, they tend to stay at the same speed. Thus, no or little variation of traffic state. If something happens, they start slowing down whether slowly or quickly depending on the incident ahead of them. Mapping this traffic behaviour to Fig. 7(a), the contrast discussed in the previous paragraph can be examined. The section of Cluster 1 with free flow speed results in zero or small contrast values apparent in the area between the two vertical red lines and concentrated at free flow speed between 60 and 90 km/h (state 8). Thus, indicating no or little variation of the traffic state at free-flow.

Later, the speed drops near to 20 km/h. A group of circles (points) in Fig. 7(a) in the bottom left corner of the graph (state 1) to the left of the negative vertical red line showing large negative contrast values confirming the speed drop in cluster 1 in Fig. 7 (f). This drop means a severe change in traffic condition from free flow to congestion, hence large negative contrast values. The magnitude of the contrast depends on the size of the variation. Therefore, some points in the bottom left corner between the zero contrast line and negative vertical red line (state 2), indicating small variation of traffic condition at low speed. That is, once the speed dropped in Cluster 1 Fig. 7 (f), it stayed there or dropped a little further generating small negative contrast values.

Additionally, there are points in Fig. 7(a) showing small positive contrast values (state 2) which are due to the congestion waves or failed attempts of traffic to pick-up. Finally in this cluster, there are few points in the top right corner to the right of the red line with large contrast values (state 9). These points are mapping of the few lines in the cluster shown in Fig. 7 (f) which were at very low speed and moved up to free flow indicating a change from congestion to free flow traffic condition at the beginning of the cluster points.

Similarly, clusters 2–5 mapping can be explained. However, to avoid repetition and for space limitation, only some interesting findings will be mentioned. For instance, traffic state in the bottom right corner (state 3) in the speed-contrast representation of all the clusters is empty. This is because it is physically impossible to have large positive traffic variation and still have low speed. Top left corner of the five clusters also does not have any points (state 7). Although this is physically possible, it is extremely difficult. For that corner to have any points, it is only possible if someone is driving at an extreme speed say 200 km/h and then drops his speed to the average free flow speed. Additionally, cluster 3 has a wider spread of points across the contrast graph. This is a result of the cluster speed data multiple stop and go. Moreover, it has more contrast points between speeds 30 and 60 km/h than any other cluster owing to the fact that it has more speed curves settling at medium speed than any other cluster.

Since the traffic states in relation to contrast values have been explained in general terms, the range which define "small value" contrast or "large value" must be set. This is shown by investigating Fig. 7(a–e). By extrapolating the information from the graphs and linking cluster speed data to 2D map in Fig. 7 (f), one could see some boundaries of speed and contrast. It is evident that free flow speed is above 60 km/h, medium flow speed is between (30 and 60) km/h while congestion speed is up to 30 km/h which is also confirmed by clustering in Section 5. Meanwhile, contrast ranges from −900 to 1100 spanning a range of 2000 numbers. Therefore, it is safe to roughly split the contrast into 3 equal ranges as shown in Table 7 for the interpretation of results. Instead, the speed is discretized into 3 ranges. The μ_{speed} and *CON* ranges are based upon the traffic characteristics evident from Fig. 7 (a to e). Having split the contrast into 3 ranges and speed into 3 ranges, the combination of different speed and contrast results in 9 different traffic states as shown in Table 7 and marked with numbers from 1 to 9 in Fig. 7 (a to e). It is important to distinguish between the 9 states defined here and the five clusters defined earlier. Each cluster may have more than one traffic state.

6.2. Results validation by additional junctions

To validate the method, an additional junction "AL2701" is used. Two periods from the junction are tested. A morning peak period from 5:00am to 10:00am and a late evening period from 7:00 pm to 12:00am. The classification and prediction of "AL2701" traffic states are conducted using the HMMs trained with the data from the junction "AL1260". Fig. 8(e) shows the clustering of "AL2701" morning peak traffic pattern. The remainder of the sub-figures in Fig. 8 shows the 2D mapping traffic states in the contrast-speed domain. It can be seen from Fig. 8(e) that there are 4 clusters similar to those presented in Sections 5.3.1–5.3.5 except for cluster 5.3.3 does not have any data. This is due to the

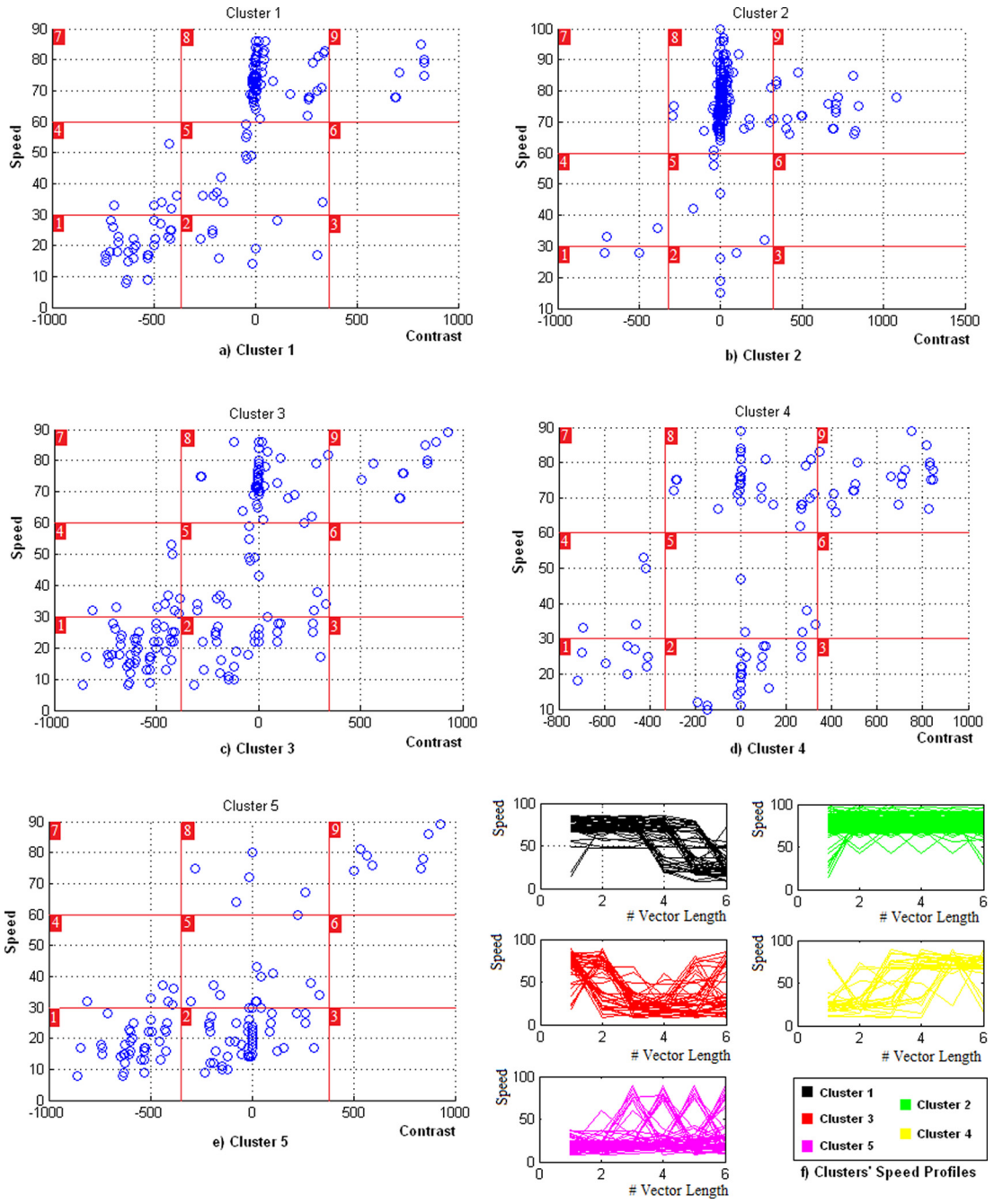


Fig. 7. Scatter Plot of Speed and Contrast.

Table 7
Traffic states.

Contrast	Speed		
	⩽ 30	30–60	⩾ 60
⩽ −330	1	4	7
−330 to 330	2	5	8
⩾ 330	3	6	9

smooth transition of speed that appears in most of the clusters of "AL2701". Thus, according to the smooth nature of the traffic profile of this junction, the 90 min (6 sampling points) is not long enough to capture a pattern going from free flow to congestion and back. Focusing on the available four clusters, it can be seen that all of them have very smooth speed transitions unlike junction "AL1260". This smooth transition according to contrast definition results in low contrast values $-330 < CON < 330$.

For instance, in cluster 1 the speed drops smoothly from around 60 km/h at the beginning of Fig. 8(e) to 40 km/h near the middle of the graph and ending with a speed near to 20 km/h. This smooth transition means the traffic state did not have an abrupt change or hit congestion all of a sudden. It rather swiftly moves to medium traffic flow and then to congestion. This can be seen from the mapping Fig. 8(a) that some points have contrast values $-330 < CON$and speed between 30 and 60 km/h (state 5) as defined in Table 7 then speed drops less than 30 km/h which also correspond to low contrast values (state 2). Some positive contrast values $CON < 330$ are also seen in the 2D mapping correspond to small positive changes of few speed vectors in Fig. 8(e). Similarly,

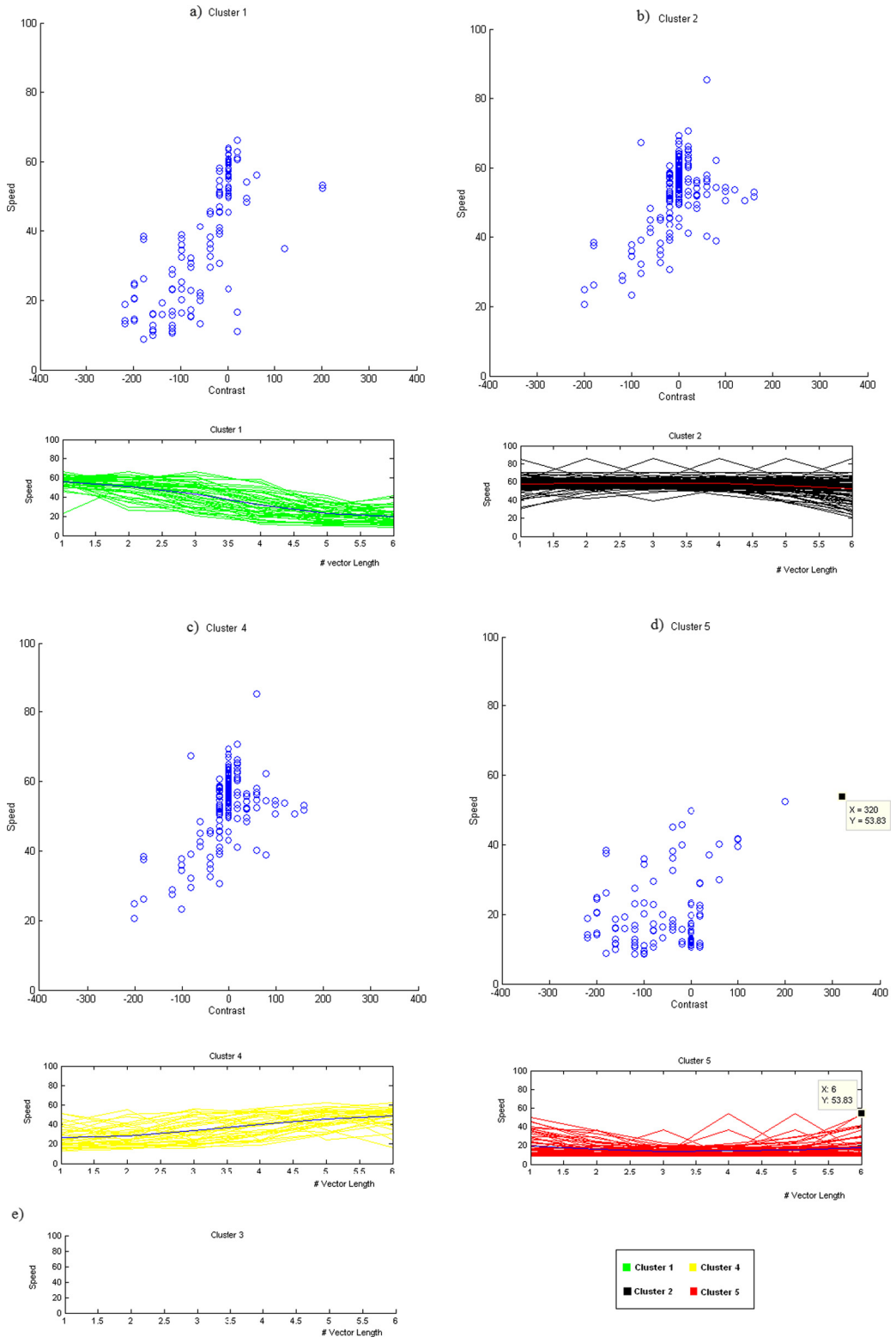


Fig. 8. Junction AL2701 Morning Peak Hour.

cluster 4 Fig. 8(e) have smooth transition of speed in the opposite direction to cluster 1. The mapping shown in Fig. 8(c) shows positive contrast < 330 (state 5 and state 8) owing to the increment in traffic speed. Again, the negative contrast in Fig. 8(c) due to small negative change in the speed trajectory of some vectors. The same could be argued on cluster 2 and cluster 5. Additionally, in cluster 5, one point has a contrast value of 320 which approaches the 330 boundary as shown in the data-tip Fig. 8(d) due to a speed change

from nearly 30 km/h to 53 km/h (still $<$ 330 though almost going from congestion upper bound of 30 km/h to free flow speed). Hence, relatively large contrast value within the stable bounds.

Fig. 9(a) presents the clustering of "AL2701" late evening off peak traffic pattern. Three clusters are shown as compared to 5 clusters in Sections 5.3.1–5.3.5. Since traffic during off-peak hours have less density, it is mostly flowing at free flow speed. Thus, most of traffic shown in Fig. 9(a) is concentrated in cluster 1 while the 2 other clusters show little activity. The explanation of the clusters 2D mapping is easily followed from previous discussions. Nevertheless, few additional comments appear in the off-peak clusters. For instance, in cluster 1, there are 2 points on the speed-contrast Fig. 9(b) with high contrast values above 330. Those result from the two lines shown in the cluster Fig. 9(a) that go from 20 and 40 km/h up to 80 and 60 km/h respectively. While cluster 2 Fig. 9(c) has one point with a large negative contrast value <-330 due to a large drop near the end of the curve in Fig. 9(a). Similarly in Cluster 3, one large positive contrast value which is due to the large fast increase in the speed of one vector.

6.3. Important findings

Having tested additional peak and off-peak periods, it is important to know that HMM fitted to one junction is able to accurately classify and predict data from different junctions provided that the following condition is true. *"The junction used for training must have a profile that covers most if not all the realistic available traffic states. That is, free-flow, medium flow, and congestion with their up and down transitions as shown in Fig. 3.* "This can be achieved by ensuring the speed-flow and flow-density relationships, of the training junction coincides with the theoretical graphs defined by Kerner [21] and shown for the training junction in Fig. 3. Thus covering the majority of traffic states. It is also interesting to note that an adaptive model which is capable of tailoring the HMMs online according to the junction could produce better results. Though care must be exercised regarding the complexity of such an approach.

In a nutshell, the change of contrast corresponds to the variation of traffic speed and hence the change in traffic state. Small increase or decrease of speed corresponds to contrast increase or decrease within the range -330 to 330. This range relates to a relatively stable traffic condition whether free flow, medium traffic, or congestion. Any large drop of speed would reflect with a large negative change in contrast values (negatively $\leqslant -330$). This means going into congestion. While any large speed increase would reflect with a large positive change in contrast values ($\geqslant$ 330). This large positive or negative change conform with unstable traffic flow such as rapid breakdown or recovery. One could argue, that both the speed and contrast could be split further into smaller ranges, to tightly represent various congestion states, which is outside the scope of this work. It is also worth noting that the choice of these ranges is based on the researcher experience and the analysis which is apparent from the 2D mapping of clusters. Not to forget an important finding, that each dataset has a different range of contrast.

The states defined in Table 7 represent various traffic conditions resulted from using the combination of both speed and contrast ranges. A total of 9 traffic states is shown in Fig. 10 and in Table 7. Fig. 10 (b, c, d) represent special cuts of Fig. 10(a). They show traffic states 1, 2, 5, 8 and 9 as indicated with the respective numbers on the graph and marked with black rectangles. Some of states are known traffic conditions such as the state where the contrast is between -330 and 330 while speed is $\leqslant$ 30. This presents a stable traffic under congestion conditions (state 2) Fig. 10(b). Similarly, the same contrast range at high speed above 60 km/h correspond to stable traffic under free flow conditions (state 8) as shown in 10(d). The same contrast range at speeds between 30 and 60 km/h conform to traffic flow at medium speed conforming to (state 5). Whereas, contrast range $\geqslant$ 330 with a speed $\leqslant$ 30 km/h conform to a traffic state in the table which is physically impossible as explained earlier (state 3). As for the same contrast range $\geqslant$ 330 but with speed between 30 to 60 does not resemble a well known traffic condition or a traffic condition that appear very regularly in real time traffic profiles. It is a medium speed with large contrast value where it is rare to have a driving conditions at medium speed but also with great variation of speed that results in large contrast values (state 6). It is a medium condition which is reached through picking up from congestion or diverting away from free flow. Hence, the difficulty of resulting in large contrast values. Thus rarely happens in reality. Meanwhile traffic state 1 and 9 resemble traffic breakdown and recovery respectively.

6.4. Comparison with related work

HMM used in this work predicts the state of traffic. With the aid of contrast, the predicted states can be explained. To assess the method developed here, it is compared to a simple Network-Fuzzy approach [35]. Shankar developed three approaches for traffic congestion prediction. Their best approach, based on simple ANFIS model, is replicated on the dataset in hand. The results are shown in Table 9. It represents the accuracy of the prediction with respect to observed values. The error is measured based on the Mean Absolute Error (MAE) according to the following Eq. (9):

$$Error = \frac{\sum |\text{SystemOutput} - \text{ActualValues}|}{\text{TotalDataPoints}} \tag{9}$$

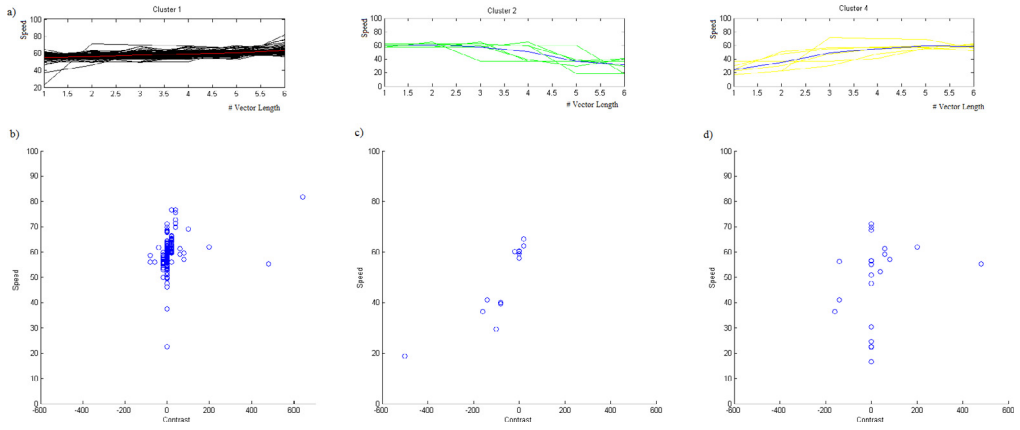


Fig. 9. Junction AL2701 Late Evening Off Peak.

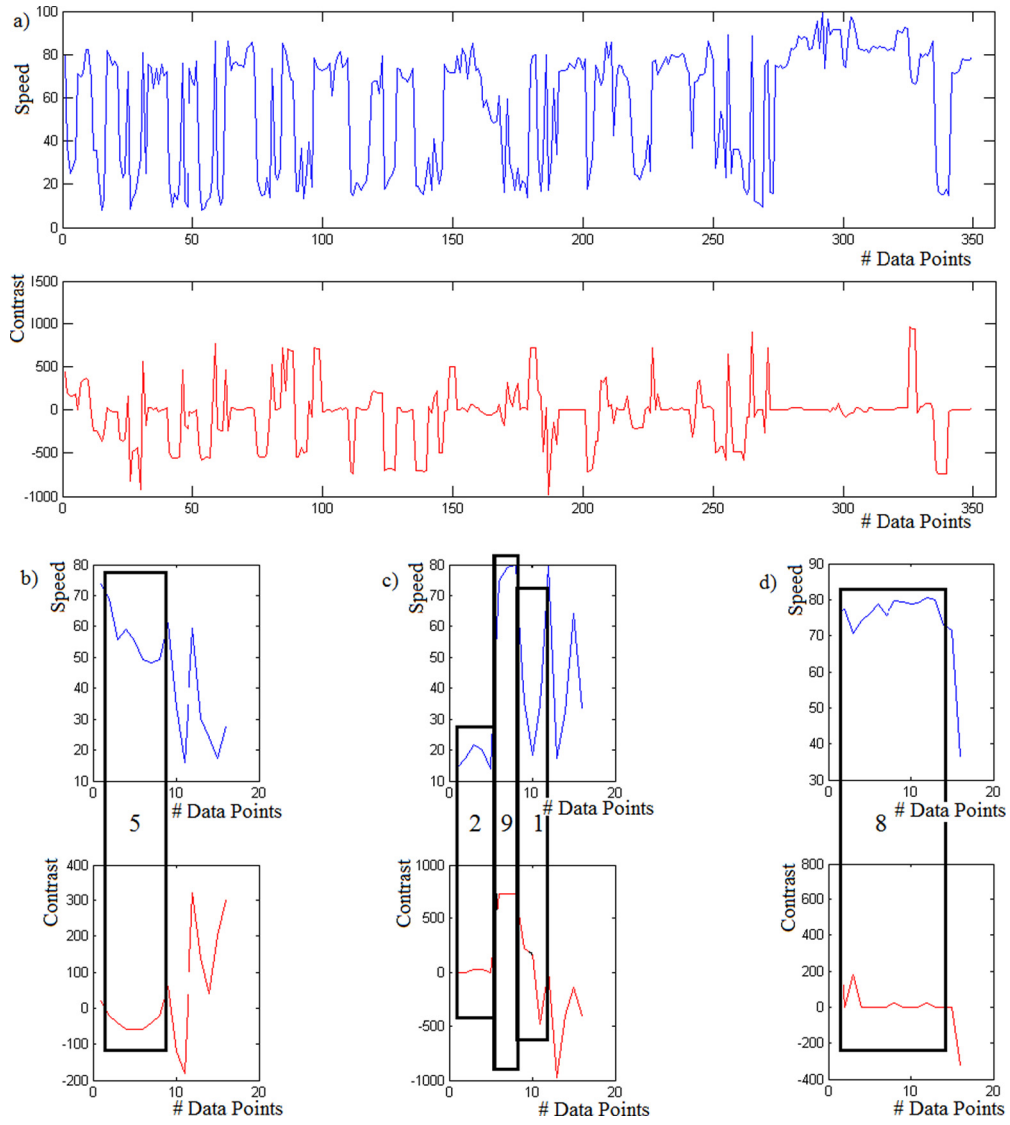


Fig. 10. Traffic states resulting from Speed and Contrast Relation.

The classification error of HMM refers to the ability of HMM to classify a test sequence into its respective cluster. That is the number of miss/hits by the total number of test vectors provides the error/accuracy respectively. The ground truth of the correct cluster is obtained through the clustering algorithm while the system output depends on the Log-Likelihood of the test vector as obtained by the HMMs. For prediction, the ground truth refers to observed values while the system output is the predicted value.

Fig. 11(a) shows the predicted versus observed values while (b) shows the residuals of the prediction process. In order to test the goodness of model fit, different tests are discussed in [43,20] such as Augmented Dickey-Fuller, PhillipsPerron, Breusch-Godfrey and Ljung-Box. In this work, Ljung-Box test is applied to the residuals. The null hypothesis is defined as H_0: the model does not exhibit lack of fit while H_α: exhibits lack of fit where α is the significance level of 5%.

$$Q = n(n+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{n-k} \tag{10}$$

where n is the sample size, $\hat{\rho}_k$ is the autocorrelation at lag k, m is the number of lags being tested, and Q is the test statistics. Table 8 shows the results of Ljung-Box test for the first 5 lags with h degrees of freedom where the $P-value$ suggests that the null hypothesis holds and the model exhibits good fit. More lags have been tested for verification. Sample autocorrelation and partial autocorrelation functions are shown in Fig. 12.

From such discussions, it can be seen that the proposed approach outperforms related work with a prediction accuracy of 91.2% as shown in Table 9.

The novelty introduced by this research showed that contrast range of values differs for different datasets. The contrast concept has been extended from 5 min prediction horizon as in [32] to 90 min horizon with good prediction results. It can also be concluded that contrast values depend heavily on the dataset resolution, and the prediction horizon. The method developed in this work is not only for data collected on small intervals as in [32] but also for aggregated data at longer intervals and longer prediction horizons. Future work includes verifying the concept

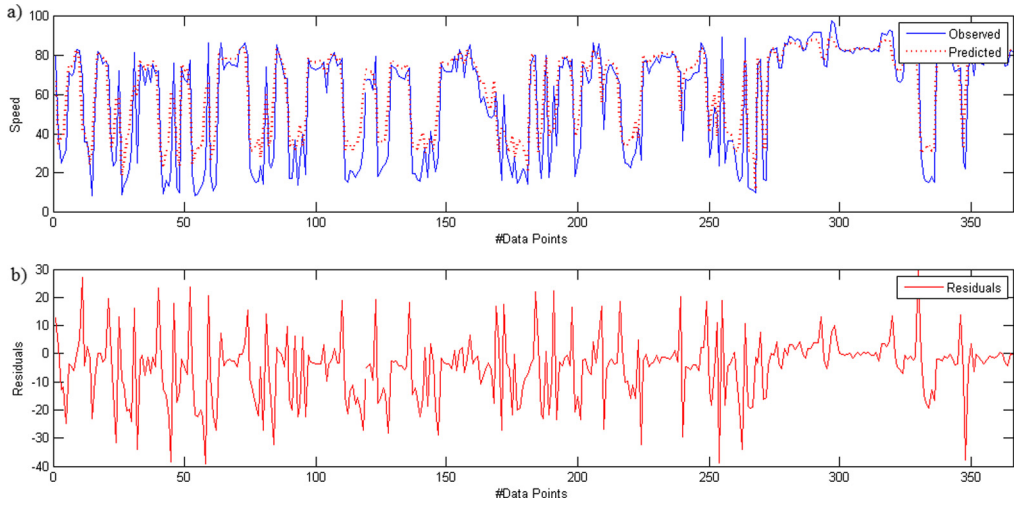


Fig. 11. Observed VS Predicted values.

Table 8
Ljung-Box goodness of fit results.

Lags	1	2	3	4	5
h	0	0	0	0	0
P − value	0.3975	0.4914	0.3733	0.4701	0.2704
Q	0.7158	1.4211	3.1214	3.5514	6.3866

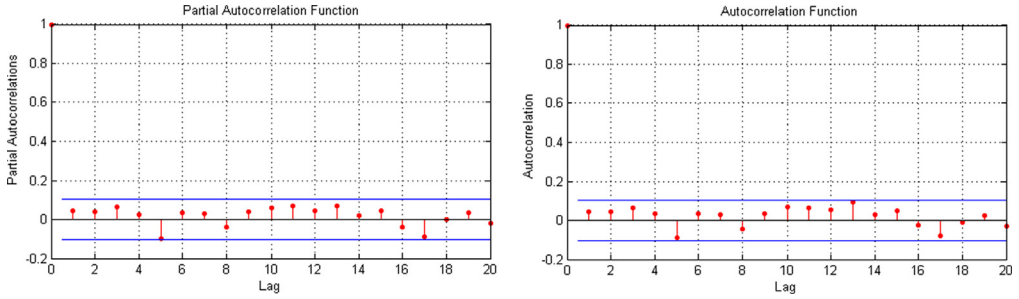


Fig. 12. Residuals Autocorrelation.

Table 9
Error results.

Method	Error %
ANFIS Prediction Error (Actual of [35])	11.77%
ANFIS Prediction Error (Replication of Shankar in this work)	13%
HMM Prediction Error [32]	10%
HMM Prediction Error (This work)	8.8%
Seasonal ARIMA [22]	10%
Random Walk [48]	10.4%

of contrast on additional datasets, studying the spatiotemporal contrast-speed relationship, adaptively adjusting HMM parameters as well as different methods for stationarizing the time series.

7. Conclusion

This paper is discussing the use of HMM for short-term traffic prediction. The contribution of this work can be summarized as follows: (1) Proposing new HMM model (HMM-C) which combines Hidden Markov Model (HMM) and Contrast measure for traffic state prediction. (2) Predicting traffic condition using different traffic horizons from 5, 15 up to 90 min. (3) Contrast principal can be used for predicting traffic conditions. Since, it has been applied in other research on a dataset from USA and in this research on a dataset from UK, it was able to determine the states. However, contrast range of values differ depending on the dataset even for similar traffic conditions.

To summarize, one can make inference and generalization regarding contrast, its properties, and its use with different datasets. It can be utilized on datasets each having a different sampling interval. It is valid for different prediction horizons ranging from few minutes up to one and half hour (90 min). Though, care is required when partitioning the contrast range into smaller periods representing the different states of traffic since it does not have a single range of values for all datasets. Instead, a different range of values exists for every dataset. Thus, it is fairly safe to restrictively generalize contrast properties and its use as an indicator for traffic state prediction. The dynamics of traffic are captured into the HMM state transition probability matrix and the observations of speed are captured into the emission probability matrix. The trend of traffic and variations are captured through the statistics. The variables of the dataset are analysed for correlation using PCA and Pearson Correlation test. Consequently, the right choice of variables is made for construction of HMM. Clustering analysis is exploited to serve the HMM construction. Following, the training

of HMM is performed using expectation maximization. Later, the trained HMM is used to predict different traffic states using the validation dataset. There are 9 different traffic states defined according to the ranges of speed and contrast. This is a more practical set of states compared to the 30 traffic states defined earlier where a number of the states do not appear in real traffic pattern. For a sequence of traffic speed observations, HMM estimates the most likely sequence of traffic states using Viterbi Algorithm path backtracking.

The purpose of the proposed HMM model is to assist traffic departments in their short-term decision making. Whether to send traffic officers to a particular location, use variable speed signs upstream to divert traffic into alternative roads, or increase the number of recovery vehicles in a particular stretch of road at certain times. It also supports traffic management and improvement strategies on the longer term such as ramp metering. Such forward looking approaches can only be applied if prediction is available.

Acknowledgement

The authors would like to thank Prof. Sherif Ishak and Dr. Yan Qi for their patience and prompt responses to questions regarding their work. Also, Jay Symonds from the UK government for providing the dataset. They would also like to thank the anonymous reviewers for their insights and valuable comments.

References

[1] Abdi J, Moshiri B, Abdulhai B, Sedigh AK. Forecasting of short-term traffic-flow based on improved neurofuzzy models via emotional temporal difference learning algorithm. Eng Appl Artif Intell 2012;25(5):1022–42. ISSN 0952–1976. http://www.sciencedirect.com/science/article/pii/S0952197611001643.
[2] Abdulhai B, Porwal H, Recker W. Short-term traffic flow prediction using neuro-genetic algorithms. ITS J, Copyright Taylor & Francis 2002;7(3–41).
[3] Ahmed M, Cook A. Analysis of freeway traffic time-series data by using box-jenkins techniques. Transp Res Rec 1979;722:1–9.
[4] Aue A, Horvth L, HuÅkov M. Segmenting mean-nonstationary time series via trending regressions. J Econ 2012;168(2):367–81.
[5] Azimi M, Zhang Y. Categorizing freeway flow conditions by using clustering methods. Transport Res Rec J Transport Res Board 2010;2173(1):105–14.
[6] Baum LLE, Sell GR. Growth functions for transformations on manifolds. Pac J Math 1968;27(2):211–27.
[7] Bocharov A, Thiesson B. Segmentation of nonstationary time series with geometric clustering. Advances in intelligent systems and computing, vol. 204. Springer; 2013.
[8] Bouyahia Z, Haddad H, Jabeur N, Yasar A. A two-stage road traffic congestion prediction and resource dispatching toward a self-organizing traffic control system. Pers Ubiquit Comput 2019;1(1–12).
[9] Cetin M, Comert G. Short-term traffic flow prediction with regime switching models. Transport Res Rec J Transport Res Board 1965;3:2006.
[10] Chen M, Yu G, Chen P, Wang Y. Traffic congestion prediction based on long-short term memory neural network models. In: 17th COTA International conference of transportation professionals.
[11] Cheng T, Haworth J, Wang J. Spatio-temporal autocorrelation of road network data. J Geogr Syst 2011;14(4):1–25.
[12] Gastaldi M, Gecchele G, Rossi R. Estimation of annual average daily traffic from one-week traffic counts. a combined ann-fuzzy approach. Transport Res Part C: Emerg Technol 2014;47:86–99.
[13] Hamed MM, Al-Masaeid HR, Said ZMB. Short-term prediction of traffic volume in urban arterials. J Transport Eng 1995;121(3):249–54.
[14] Highways England Government Website. <http://www.highways.gov.uk/>. <http://www.highways.gov.uk/>. Page Visited 15-May-2015.
[15] Hoel P, Port S, Stone C. Introduction to stochastic processes., volume (ISBN 0-88133-267-4). Waveland Press, Inc.; 1987.
[16] Ishak S, Alecsandru C. Analysis of freeway traffic incident conditions by using second-order spatiotemporal traffic performance measures. Transport Res Rec, J Transport Res Board 1925;20–28:2005.
[17] Ishak S, Qi Y. Stochastic approach for short-term freeway traffic prediction during peak periods. IEEE Trans Intell Transp Syst 2013;14(2):660–72.
[18] Jolliffe I. Principal component analysis. Springer; 2002.
[19] Kamarianakis Y, Shen W, Wynter L. Real-time road traffic forecasting using regime-switching space-time models and adaptive lasso. Appl Stoch Models Bus Indust, John Wiley & Sons Ltd 2012;28:297–315.
[20] Karlaftis M, Vlahogianni E. Memory properties and fractional integration in transportation time-series. Transport Res Part C: Emerg Technol 2009;17:444–53.
[21] Kerner BS. The physics of traffic. Springer; 2004.
[22] Kumar SV, Vanajakshi L. Short-term traffic flow prediction using seasonal arima model with limited input data. Eur Transp Res Rev 2015.
[23] Levin DA, Peres Y, Wilmer EL. Markov chains and mixing times. American Mathematical Society; 2008.
[24] Li L, Su X, Zhang Y, Lin Y, Li Z. Trend modeling for traffic time series analysis: an integrated study. IEEE Trans Intell Transp Syst 2015;16(6):3430–9.
[25] Liu X, Fang X, Qin Z, Ye C, MX. A short-term forecasting algorithm for network traffic based on chaos theory and svm. J Network Syst Manage 2011;19(4):427–47.
[26] Lumpe M, Vo QB. Finding the k in k-means clustering: a comparative analysis approach. AI 2015: Advances in Artificial Intelligence, Lecture Notes in Computer Science, Springer 2015;9457:356–364.
[27] Lwin HT, Naing TT. Road traffic congestion estimation system using gps data. In: Proceedings of seventh TheIIER international conference, Singapore.
[28] Ma T, Zhou Z, Abdulhai B. Nonlinear multivariate time-space threshold vector error correction model for short term traffic state prediction. Transp Res Part B 2015;76:27–47.
[29] Maciejowski JM. Predictive control with constraints. Prentice Hall, volume 978-0201398236; September 2000.
[30] Okutani I, Stephanedes YJ. Dynamic prediction of traffic volume through kalman filtering theory. Transport Res Part B Methodol Feb 1984;18(1):1–11.
[31] Porikli F, Li X. Traffic congestion estimation using hmm models without vehicle tracking. In: Intelligent vehicles symposium. IEEE; 2004. p. 188–93.
[32] Qi Y, Ishak S. A hidden Markov model for short term prediction of traffic conditions on freeways. ELSEVIER - Transport Res Part C 2014;43:95–111 URL www.elsevier.com/locate/trc.
[33] Qi Y, Ishak S. Probabilistic models for short term traffic conditions prediction. PhD thesis, Louisiana State University The Department of Civil and Environment Engineering; 2014b.
[34] Rabiner LR. A tutorial on hidden markov models and selected aapplication in speech recognition. IEEE 1989;77(2):257–87.
[35] Shankar H, Raju PLN, Rao KRM. Multi model criteria for the estimation of road traffic congestion from traffic flow information based on fuzzy logic. J Transport Technol 2012;2:50–62.
[36] Tang J, Zhang G, Wang Y, Wang H, Liu F. A hybrid approach to integrate fuzzy c-means based imputation method with genetic algorithm for missing traffic volume data estimation. Transport Res Part C: Emerg Technol 2015;51:29–40.
[37] Theodoridis S, Koutroumbas K. Pattern recognition. Academic Press, volume 0-12-686140-4; 1999.
[38] Traffic England government website. <http://www.trafficengland.com>. <http://www.trafficengland.com>. Page Visited 15-May-2015.
[39] UK Government. <http://data.gov.uk/dataset/dft-eng-srn-routes-journey-times>. <http://data.gov.uk/dataset/dft-eng-srn-routes-journey-times>.
[40] Viterbi AJ. The viterbi algorithm. Proc IEEE 1973;61:268–78.
[41] Vlahogianni E. Enhancing predictions in signalized arterials with information on short-term traffic flow dynamics. J Intell Transport Syst Technol Plan Oper 2009;13(2):73–84.
[42] Vlahogianni E, Golias J, Karlaftis M. Short-term traffic forecasting: overview of objectives and methods. Transp Rev 2004;24(5):533–57.
[43] Vlahogianni E, Karlaftis M. Testing and comparing neural network and statistical approaches for predicting transportation time series. Transport Res Rec: J Transport Res Board 2013;2399(02).
[44] Vlahogianni E, Karlaftis M, Golias J. Statistical methods for detecting nonlinearity and non-stationarity in univariate short-term time-series of traffic volumec. Transport Res Part C: Emerg Technol 2006;14:351–67.
[45] Vlahogianni E, Karlaftis M, Golias J. Short-term traffic forecasting: where we are and where we–re going. Transport Res Part C, 2014;43(Part 1):3–19.
[46] Wang X, Peng L, Chi T, Li M, Yao X, Shao J. A hidden markov model for urban-scale traffic estimation using floating car data. PLoS ONE 2015;10(12).
[47] Wang Y, Papageorgiou M, Messmer A. Real-time freeway traffic state estimation based on extended kalman filter: adaptive capabilities and real data testing. Transport Res Part A: Policy Pract 2008;42(10):1340–58.
[48] Williams BM, Hoel LA. Modeling and forecasting vehicular traffic flow as a seasonal arima process: theoretical basis and empirical results. ASCE J Transport Eng 2003;129(6–6):664–72.
[49] Xia J, Chen M. Dynamic freeway corridor travel time prediction using single inductive loop detector data. In: Transportation research board 88th annual meeting, Washington DC.
[50] Xia J, Huang W, Guo J. A clustering approach to online freeway traffic state identification using its data. KSCE J Civil Eng 2012;16(3):426–32.
[51] Yang F, Yin Z, Liu H, Ran B. Online recursive algorithm for short-term traffic prediction. Transport Res Rec J Transport Res Board Jan 2004;1879(1):1–8.
[52] Ye N, qin Wang Z, Malekian R, Lin Q, chuan Wang R. A method for driving route predictions based on hidden markov model. Math Probl Eng; 2015.
[53] Yin H, Wong S, Xu J, Wong C. Urban traffic flow prediction using a fuzzy-neural approach. Transport Res Part C Emerg Technol 2002;10:85–98.
[54] Zaki JF, Ali-Eldin AM, Hussein SE, Saraya SF, Areed FF. Time aware hybrid hidden markov models for traffic congestion prediction. Int J Electr Eng Inform 2019;11(1).
[55] Zhang Y, Ye N, Wang R, Malekian R. A method for traffic congestion clustering judgment based on grey relational analysis. Int J Geo-Inform, ISPRS 2016;5(71).
[56] Zhu Y, Gao N, Wang J, Liu C. Study on traffic flow patterns identification of single intersection intelligent signal control. Proc Eng 2016;137:452–60.