A Comparison of Mamdani and Sugeno Fuzzy Inference Systems for Traffic Flow Prediction

Yang Wang
Beijing Key Lab of Traffic Engineering, Beijing University of Technology, Beijing, China
Email: wang_yang@bjut.edu.cn

Yanyan Chen
Beijing Key Lab of Traffic Engineering, Beijing University of Technology, Beijing, China
Email: cdyan@bjut.edu.cn

***Abstract*—Information on the future state of traffic flow provides a solid foundation for the efficient implementation of traffic control and guidance. The prediction approaches based on fuzzy logic theory is of great interests, because the rule-based inference is similar to the way humans process casual relations and fuzzy linguistic variables provide a natural way to deal with uncertainties. This paper presents a comparative study on a set of widely used Mamdani and Sugeno fuzzy inference systems in the application on the short-term prediction for traffic flow based on the historical recordings. To fulfill the comparison, a series of experiments was designed and performed to evaluate prediction performance for each fuzzy inference system in terms of model complexity, execution time, noise resistance, performance consistency, missing data, and multi-step-ahead predictability. Before discussing the primary results, a description on the fuzzy inference systems, evaluation factors and criteria was given. The analyses on the experimental results led to several findings which can be referenced when choosing a FIS for traffic flow prediction based on historical recordings.**

***Index Terms*—Traffic flow prediction, fuzzy inference systems, defuzzification mechanisms, traffic flow time series**

I. Introduction

In intelligent transportation systems (ITS), not only real time traffic flow is of great importance for providing dynamic traffic control and guidance, but also live and accurate traffic flow prediction can help reduce unexpected malfunction and improve efficiency in transportation systems. However, the road traffic system is a complex, open, and time-variant system which often exhibits highly randomness and uncertainty. Such challenge problem has fostered considerable research enthusiasm that has been continuously devoted to this field. As a result, a wide range of prediction algorithms has been developed, such as Kalman filter [1] and its extension [2], support vector machine (SVM) [3], Bayesian networks [4], and hybrid approach [5-8] etc. The prediction approaches based on Kalman filter theory operate recursively on a stream of noisy measurements to predict the future traffic status. Although, this type of approaches do not require to store previous recordings for prediction purpose, a linear dynamic system adopted for to simulate the traffic system is not appropriate. The extended Kalman filter can relax the linear assumption by employing differential equations to model the underlying systems, but the calculation efforts are increased accordingly. One of main advantages of the SVM approaches for traffic flow prediction is their capability to handle the nonlinear problem by means of the kernel transformation from input space to feature space. Furthermore, the prediction accuracy by this type of approaches is generally high even with a small set of training noisy data. The support vectors of a SVM are constructed by solving a quadratic programming problem and this is challenging for a large optimization problem due to the high order matrix calculation. Also, the function after training is not easy to interpret. Bayesian network is a probabilistic reasoning technology and therefore used in the traffic flow prediction to deal with the uncertainty due to the correlation between conditions. In addition, Bayesian network can model casual relationships between variables through learning. However, the main weaknesses for this type of approaches are that the learning result is largely influenced by the *prior* knowledge and Bayesian network has limitation on handling continuous features. The hybrid approaches attempt to overcome the weaknesses of each prediction technique by combining more than one prediction techniques. The prediction accuracy can be improved by the hybrid solution at the cost of extra computation and the improvement made is largely dependent on the fusion method employed.

There are a number of ways to classify these prediction algorithms depending on the grouping factor to be used, such as single road link or correlated road network (sub-network), urban streets or freeways, parametric or non-parametric model, analytical or data-driven approach,

Manuscript received November 1, 2012; revised May 31, 2013; accepted June 3, 2013.

Scientific Research Common Program of Beijing Municipal Commission of Education (KM201010005021); Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry (32004011201201)

Corresponding author: Yang Wang

doi:10.4304/jcp.9.1.12-21

univariate or multivariate method, etc. Nonetheless, the fundamental idea behind these prediction methods is more or less like "yesterday once more". That is, the knowledge extracted from past experiences is used to infer and predict the future status. Thus, it is essential to have good understanding on the past in an attempt to discover the rules governing the evolution.

The evolution of a traffic system may be influenced by many factors, such as weather, road works, and sport events (e.g., Olympic Games) etc. To use these influencing factors, as the latent inputs of the traffic system, for traffic flow prediction, the prediction accuracy largely depends on the quality of the factors supplied with the efforts expended in collecting, analyzing, and processing data. On the other hand, a sequence of historical observations, namely time series, embeds rich information that can be used to infer the future tendency. Therefore, this paper concerns the traffic flow prediction based on historical recordings only.

The modeling techniques based on fuzzy inference systems (FIS) are appealing and have found many successful applications in various fields [9-11]. As compared to the other prediction approaches (such as those mentioned above), the fuzzy approaches have the following features: 1) the rule-based structures are able to capture the dependency between inputs and outputs of a system; 2) the fuzzy linguistic variables provides a natural way to deal with uncertainties; 3) they are capable of modeling nonlinear systems; 4) the singular and linguistic outputs can be easily formed; 5) they are insensitive to random noise. Those unique features of the fuzzy approaches motivate us to investigate their performance in traffic flow prediction further.

However, FISs have seen little application in the traffic flow prediction field over the last decade, apart from a few notable contributions reported in literature. Zhang and Ye proposed a prediction methodology by using fuzzy logic system to fuse the outputs of two methods out of autoregressive integrated moving average, back-propagation neural networks, exponential smoothing method, and Kalman filter, resulting in four different combinations [6]. Similar idea has been adopted in [7], but the two methods mixed by a fuzzy logical model are history mean and artificial neural network models. The paper [8] describes a hybrid methodology that two fuzzy rule-based systems are constructed, one providing the next flow estimation based on the current flow only and the other predicting the one-step-ahead flow based on the current flow at the current location and the upstream location. The paper [12] presents a prediction approach for the short-term traffic flow prediction primarily based on Sugeno fuzzy system (also known as Takagi-Sugeno-Kang, TSK). The initial structure is formed by partitioning the input vector space by the mean shift clustering algorithm and subsequently optimized by eliminating redundant structure using the mean firing technique, and finally the other parameters are determined by particle swarm optimization with the aim to minimize root mean squared error.

The two types of FIS, namely Mamdani [13] and Sugeno FISs [14] are widely accepted and applied to many real-world problems. The predicted traffic flows are frequently used as a significant reference to designing or updating signal timing, route guidance, or variable message signs (VMS) etc deployed in the urban traffic networks. Some of these applications, such as signal timing or route guidance, need the numerical form of prediction result of high accuracy with time constraints, but the others, such as VMS, linguistic form with soft real-time requirement. The output from Mamdani FISs can be easily transformed to linguistic form as the inference result before defuzzification is a fuzzy set [13]. The Sugeno FISs are able to accurately model highly nonlinear systems [15]. Consequently, these two types of FISs are, in theory, suitable for the short-term prediction of traffic flow. Furthermore, although many types of FISs have been proposed, most of them were developed on the basis of Mamdani and Sugeno FISs [15]. Therefore, the comparison made between these two types of FISs in the short-term traffic flow prediction severed as the starting point of our on-going research work and the other FISs will be a focus of our future research.

For Mamdani FISs, the inference for ith rule can be mathematically expressed as follow:

$$R_i : \text{if } x_1 \text{ is } A_{i1} \text{ and} x_2 \text{ is } A_{i2} \text{ and} \ldots x_n \text{ is } A_{in} \text{ then} y_i \text{ is } B_i \tag{1}$$

where x_1 , x_2 , ... , x_n are the input variables and A_{i1}, A_{i2}, …, A_{in} are the fuzzy sets. Unlike the Mamdani FIS, the output from a Sugeno FIS corresponding to ith rule is typically a function of the input vector $\boldsymbol{x}$.

$$R_i : \text{if } x_1 \text{ is } A_{i1} \text{ and} x_2 \text{ is } A_{i2} \text{ and} \ldots x_n \text{ is } A_{in} \text{ then } y_i \text{ is } f_i(\boldsymbol{x}) \tag{2}$$

Sugeno FISs are similar to Mamdani FISs in many aspects, but the consequent parts are quite different. The Sugeno type of FIS uses a mathematical function of the inputs as the rule consequent, instead of fuzzy set employed in Mamdani FISs. The consequent of a rule in a Sugeno FIS is normally a polynomial in the input variables, but it can in theory be any type of functions as long as it can properly present the output within the fuzzy space specified by the antecedent. The Sugeno FIS resulted from a first-order polynomial was originally proposed in [16-17] and is called a first-order Sugeno FIS. Due to the different forms in the rule consequent of Mamdani and Sugeno FISs, the methods used in both types of FIS to get overall crisp output are different too.

In the literature, there a large amount of applications employed either Mamdani or Sugeno FISs, but a few contributions have been made to the comparison between these two types of FIS in terms of their prediction performance. In [18], a comparison of Mamdani and Sugeno FISs was made to evaluate the quality of experience of hapto-audio-visual applications. Although the Sugeno FIS demonstrates more accurate than the Mamdani FIS, the Mamdani FIS displays consistency in their simulations. In addition, there were no noticeable variations between the two types in terms of execution time. A comparative study of Mamdani and Sugeno FISs was reported in [19]. The two types of FIS were compared for a space fault detection application from three aspects: processing time, robustness to noise, and sensitivity analysis of the system's behaviors to changes in input data and the comparison results led to the conclusion that a Sugeno FIS with equivalent functions as Mamdani FIS may improve the overall performance. In [20], a comparison was made between adaptive neuro-fuzzy inference system (ANFIS) and a Mamdani fuzzy inference system for predicting municipal water consumption time series and the comparison indicated that ANFIS is superior to the Mamdani FIS. The paper [21] compared the performances of Mamdani and Sugeno FISs when computing the resonant frequency of rectangular microstrip antennas with thin and thick substrates, and the best result obtained from the experiments was generated by the Sugeno FIS. Hao et al. [22] investigated which type, Mamdani or Sugeno FIS, is more compact for function approximation. They showed that the basic structure of a Sugeno FIS depends on the number of locations of the extrema of the function to be approximated. The comparison results implied that minimal system configurations of Mamdani and Sugeno FISs are comparable. To the best of our knowledge, the literature contains no contributions made on the comparative study on the two types of FIS in traffic flow prediction.

Consequently, this paper is devoted to the comparison on the performances of the two types of FIS in traffic flow prediction based on historical recordings in terms of model complexity, execution time, noise resistance, performance consistency, missing data, multi-step-ahead predictability, through a set of experiments. The next section outlines the basic structures of FISs to be evaluated and briefly describes the algorithms used to determine the rules and parameters, with particular emphasis on the defuzzification methods to be examined. Following that, the evaluation factors and criteria are explained, before presenting the experimental results.

II. MAMDANI AND SUGENO FUZZY INFERENCE SYSTEMS

The evaluation was made on the structure widely adopted for the multi-input single-output (MISO) Mamdani and Sugeno FISs. The Gaussian membership function was used for both types of FIS, but the consequent for each rule in the Sugeno FISs employs linear function, namely, the first-order Sugeno FIS. Both types of FIS use T-norm for conjunction. While "min" is used in the Mamdani FISs for implication, the Sugeno FISs adopts "product" for implication. For the Mamdani FISs, the fuzzy sets resulted from the implication for each rule are aggregated using the maximum before applying the defuzzification. There are five defuzzification methods, namely centroid of area or center of gravity, bisector of area, smallest of maximum, largest of maximum, mean of maximum, which are often used in fuzzy modeling. On the other hand, the defuzzification method equipped in a typical Sugeno FIS is either weighted average or weighted sum. As the defuzzification process is completely different between Mamdani and Sugeno FISs, we compared the FISs with the above stated defuzzifications.

A. Defuzzification Methods

After aggregation, the defuzzification process for the Mamdani type of FIS converts the fuzzy set B ($\mu_B(y)$ is the membership function) in the universe of discourse V into a single crisp value as the final output. There are 5 defuzzification methods frequently adopted in Mamdani type of FIS.

1) Centroid of area (COA)

This defuzzification method returns the output by calculating the centroid of area formed by the aggregated fuzzy sets of the consequents as follow:

$$y_{\mathrm{COA}} = \frac{\int_V y \cdot \mu_B(y)dy}{\int_V \mu_B(y)dy} \tag{3}$$

2) Bisector of area (BOA)

The vertical line corresponding to the output generated by BOA splits the aggregated fuzzy sets into two sub-regions of equal area. This operation can be expressed as follow:

$$\int_{\alpha}^{y_{\mathrm{BOA}}} \mu_B(y)dy = \int_{y_{\mathrm{BOA}}}^{\beta} \mu_B(y)dy \tag{4}$$

where $\alpha = \min\{v \mid v \in V\}$, $\beta = \max\{v \mid v \in V\}$. Note that the value resulted from this method is sometimes coincidently identical to that generated from COA.

3) Smallest of maximum (SOM)

This method generates the crisp output by taking the smallest value that gives the maximum membership degree of the aggregated fuzzy set.

$$y_{\mathrm{SOM}} = \min\{y \mid \mu_B(y) = \max(\mu_B(y))\} \tag{5}$$

4) Largest of maximum (LOM)

Instead of smallest value as SOM, LOM takes the largest value corresponding to the maximum membership degree to yield the final crisp output.

$$y_{\mathrm{LOM}} = \max\{y \mid \mu_B(y) = \max(\mu_B(y))\} \tag{6}$$

5) Mean of maximum (MOM)

In this defuzzification, the mean of maxima is taken as the crisp output.

$$y_{\mathrm{MOM}} = \frac{y_{\mathrm{SOM}} + y_{\mathrm{LOM}}}{2} \tag{7}$$

Note that if the aggregated membership function has a unique maximum degree, rather than a range (i.e., a plateau at the maximum value), the crisp outputs generated by SOM, LOM, and MOM are all identical.

The consequent y_i corresponding to the ith rule (M rules in total) in the Sugeno type of FIS is a function of inputs rather than a fuzzy set in the Mamdani type of FIS. The popular defuzzification methods for the Sugeno type are the following two.

1) Weighted average (WA)

This defuzzification method generates the final output for a Sugeno FIS by averaging the weighted rule outputs.

$$y_{\mathrm{WA}} = \frac{\sum_{i=1}^{M} w_i y_i}{\sum_{i=1}^{M} w_i} \quad (8)$$

2) Weighted sum (WS)

To reduce the computation of WA, the WS method takes only the sum of the weighted rule outputs.

$$y_{\mathrm{WS}} = \sum_{i=1}^{M} w_i y_i \quad (9)$$

B. Rules and Parameters Determination

In this paper, fuzzy c-means (FCM) [23-24] is used to extract rules from a set of training data for the Mamdani FISs and the number of clusters pre-defined is the number of rules. Although the same clustering algorithm is taken for the Sugeno FISs to determine the antecedent, the estimation for the consequent parameters can be formulated as the least square problem and consequently the least estimation method [25] is used to determine the consequent parameters.

III. Evaluation Factors and Criteria

A. Evaluation Factors

To compare the two types of FIS, the following practical aspects have been considered in this paper as they are critical to a success in the traffic flow prediction.

1) Model complexity

While a complex FIS may improve prediction accuracy, computational overhead is often a serious issue for the time-constrained applications due to curse of dimensionality. In this paper, the model complexity is measured by the total number of membership functions for all rules.

2) Execution time

The time required to deliver a solution from a predictor is often critical for traffic control and management or traveler decision support. Prediction made by a FIS is typically consists of two stages, model construction and execution (though the model can be constructed in an off-line manner). However, this paper is not intended to compare various techniques proposed in the literature for model construction. Therefore, the execution time for each FIS under study is concerned in this work and the difference in execution time is largely attributed to the defuzzification mechanism used.

3) Noise resistance

As a sequence of observations on traffic flow is often corrupted by noise, it would be problematic if the future flow is estimated using the model that has overfit to the noisy time series. Thus, this test examines the ability of a FIS to immunize the noise embedded in the time series.

4) Performance consistency

This test measures whether the prediction performance of a FIS is consistent if the traffic situation is changed. A good predictor should have a low fluctuation in response to different traffic situations.

5) Missing data

In practice, there are some recordings missing due to disordered detectors or malfunctioned transformation etc. Although the missing recordings can be made up by many techniques, such as linear or nonlinear interpolation, it is often desired that the future flow can still be estimated without any extra preprocessing in order to minimize computation efforts. In this test, we will examine which FIS is mostly robust to missing data.

6) Multi-steps-ahead predictability

Predicting the first unknown future flow (i.e. one-step-head) is useful in many aspects, for example real time guidance, but additional benefits can be obtained from multi-step-ahead prediction. In this paper, the multi-step-ahead predictability is therefore examined in terms of the prediction accuracy for the increasing prediction horizon.

B. Evaluation Criteria

Three statistical criteria, namely mean square error (MSE) [26], mean absolute percentage error (MAPE) [27], and variance of absolute percentage error (VAPE) [6], were used to assess the prediction quality. The MSE statistic is frequently employed for prediction evaluation, as it indicates a model's ability to predict a value away from the mean. While MAPE calculates the average relative error between the estimated values and actual observed data, VAPE represents the performance stability.

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}(i) - y(i)\right)^2 \quad (10)$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}(i) - y(i)}{y(i)}\right| \times 100\,\% \quad (11)$$

$$\mathrm{VAPE} = \mathrm{var}\left(\frac{\left|\hat{y}(i) - y(i)\right|}{y(i)}\right) \times 100\,\% \quad (12)$$

where $y(i)$ and $\hat{y}(i)$ are ith actual recordings and predicted values and N is the number of data.

IV. Experiments and Results

In order to make a comparison between the Mamdani and Sugeno FISs in traffic flow prediction, a series of experiments was performed in Matlab (version 7.13) and executed on Intel Core Duo processor (2.2 GH) running Windows Vista, based on the traffic flow time series collected and aggregated in the 2 minutes interval for a road in Beijing for the week starting from 20th to 26th November 2006. As the estimation of the consequent parameters for the Sugeno FISs was formulated as a least square fitting problem, there is a potential risk that the model overfits the noisy data. Consequently, a de-noising process was performed using a wavelet-based technique [28] for all experiments, before any evaluation proceeded.

1) Model complexity

The evaluation of complexity was designed by incrementing the number of inputs from 1 to 10 and the number of clusters (i.e., the number of membership function) from 2 to 10 (note that grouping all training data into one cluster is meaningless), respectively. The 10-fold blocked cross-validation [29-30] was used to evaluate each FIS for each combination and 100 independent runs were performed and averaged for each test to obtain the MSE, MAPE and VAPE values. Then, the FISs under test were ranked according to their MSE, MAPE, and VAPE values for each combination and the rank corresponding to the majority over all combinations was assigned to the associated FIS, as shown in Table I.

WA consistently outperformed all the others in terms of MSE, MAPE, and VAPE statistics, while the performance of WS was frequently ranked to be the bottom. Also, it can be seen that the MOM, SOM, LOM, and BOA models are in the second, third, fourth, and fifth places, respectively, in most situations. Therefore, it is evident that WA is able to produce the most accurate prediction with the same complexity as the others.

Table II and Table III summarize the mean MSE values for each method for the different number of inputs and clusters, respectively.

TABLE I.
RANKS OF EACH FIS BASED ON THEIR MSE, MAPE, AND VAPE VALUES FOR TRAINING (DENOTED AS 'T') AND PREDICTING (DENOTED AS 'P') STAGES.

FIS		COA	BOA	SOM	LOM	MOM	WA	WS
MSE	T	6	5	4	3	2	1	7
	P	5	3	6	4	2	1	7
MAPE	T	7	5	3	4	2	1	6
	P	6	5	2	4	3	1	7
VAPE	T	7	6	3	4	2	1	5
	P	6	5	3	4	2	1	7

TABLE II.
MEAN MSE VALUES FOR EACH ALGORITHM FOR THE DIFFERENT NUMBER OF INPUTS ('T' AND 'P' DENOTING TRAINING AND PREDICTING STAGES RESPECTIVELY).

Inputs		1	2	3	4	5	6	7	8	9	10
COA	T	456.4	457.5	460.8	466.9	473.1	480.2	488.8	498.8	508.6	518.8
	P	534.2	538.1	543.4	554.1	564.1	575.7	586.9	600.5	616.4	633.3
BOA	T	315.7	318.0	322.5	328.9	335.2	342.2	351.4	362.5	372.5	383.0
	P	387.1	392.0	397.8	409.8	421.2	433.2	444	457.7	475.3	493.2
SOM	T	188.4	200.0	214.0	227.6	245.0	266.2	288.8	315.1	345.4	376.5
	P	307.6	325.1	345.0	376.8	401.1	430.4	453	490.4	540.0	598.2
LOM	T	181.8	189.4	197.7	211.1	222.7	232.8	246.2	258.6	270.4	283.2
	P	315.9	334.6	353.2	368.4	386.8	407.7	438.3	465.5	494.6	524.7
MOM	T	99.73	101.0	103.3	106.9	111.0	115.8	121.9	128.7	136.3	144.4
	P	169.5	173.5	178.4	185.0	190.2	198.0	207.5	219.4	234.6	251.5
WA	T	2.144	0.272	0.267	0.246	0.237	0.206	0.193	0.193	0.194	0.186
	P	5.018	1.845	2.846	3.966	7.533	8.715	48.48	83.74	83.74	1814
WS	T	19246	7038	3488	2287	1973	2120	2510	3019	3617	4219
	P	18503	6862	3724	2835	2839	3291	3937	4688	5479	6306

The results listed in Table II indicate that the Mamdani type of FIS exhibit different behavior than the Sugeno type of FIS. That is the modeling and prediction errors generated by the Mamdani FISs with different defuzzification mechanisms increase with the number of inputs, but a valley pattern is generally held for the Sugeno FISs. Furthermore, the majority of the MSEs produced by WA （as highlighted in Table II） are lower than the other FISs by approximately 3 orders, indicating that WA model considerably outperformed as compared to the others. However, the poorest performance was generated by the Sugeno FIS with the defuzzification method of weighted sum. The discrepancy of the MSEs, measured for COA, BOA, SOM, LOM, MOM, and WS, between the training and prediction appears relatively stable, but for WA (as highlighted in Table II）, the prediction error increases rapidly when the number of inputs is greater than 6 even though the training error steadily reduces, implying WA overfit to the training data. Finally, near all MSEs measured during the prediction stage are larger than those during training stage.

TABLE III.
MEAN MSE VALUES FOR EACH ALGORITHM FOR THE DIFFERENT NUMBER OF CLUSTERS ('T' AND 'P' DENOTING TRAINING AND PREDICTING STAGES RESPECTIVELY).

Clusters		2	3	4	5	6	7	8	9	10
COA	T	1905	570	344.3	292.1	264.3	246.9	239.2	235.3	231.6
	P	1977	758.2	434.2	384.3	354.4	328.3	317.4	311.2	307.1
BOA	T	1406	403.1	235	197.6	182.3	171.7	167	164.3	161.4
	P	1489	585.3	317.2	279.2	264	241.9	238.3	233.6	231.5
SOM	T	644.6	450.6	318.9	252.4	202.5	165.2	137	120.1	108.9
	P	696.0	744.0	505.0	514.5	461.6	288.8	230	209.9	191.1
LOM	T	818.1	446.6	234.6	158.2	118.4	92.1	75.1	64.6	56.8
	P	1214	706.1	427.6	289.1	243	231.8	214.3	187.6	167.0
MOM	T	338.1	209.4	128.5	96.2	76.1	62.1	52.5	46.6	42.6
	P	457.1	340.2	221	185.3	167.8	130	112.6	100.5	92.1
WA	T	0.441	0.435	0.425	0.416	0.412	0.406	0.4	0.397	0.3917
	P	0.612	0.684	1.041	0.8	1.375	24.15	17.8	7814	120150
WS	T	948.0	2784	3470	3598	4313	5308	6264	7844	10035
	P	1124	3278	4692	5021	5823	6634	7130	8455	10461

From Table III, the mean MSEs for the Mamdani FISs steadily reduce when the number of clusters increases. For WA (as highlighted in Table II）, the modeling error gradually decreases with the number of clusters, but the prediction performance is generally getting worse and worse. Both modeling and prediction errors for WS continuously increase as the number of clusters increments.

From the above discussion, the followings can be summarized: a) to increase the number of membership functions is more efficient to improve the Mamdani FISs performance but when increasing either the number of inputs or membership functions, the prediction accuracy of WA decreases even though modeling error steadily declines; b) WA is capable of delivering the most accurate prediction when the number of inputs and membership functions are chosen properly; c) in general, to increase the number of inputs but membership functions can improve WS performance, but the both training and prediction errors are considerably large.

2) Execution time

To examine the execution time, the same strategy as that in model complexity evaluation was adopted except that the blocked cross-validation was not used in this test. Each FIS was independently executed for 10000 times to obtain the mean execution time. Fig. 1 and Fig. 2 illustrate the typical results obtained for the number of inputs ranging from 1 to 10 when the number of clusters is 4 and for the number of clusters ranging from 2 to 10 when the number of inputs is 4.

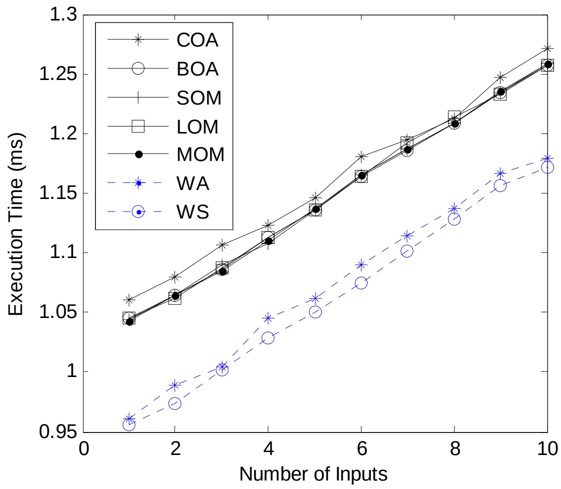


Figure 1. Execution time (ms) for the number of inputs ranging from 1 to 10 when the number of clusters is 4.

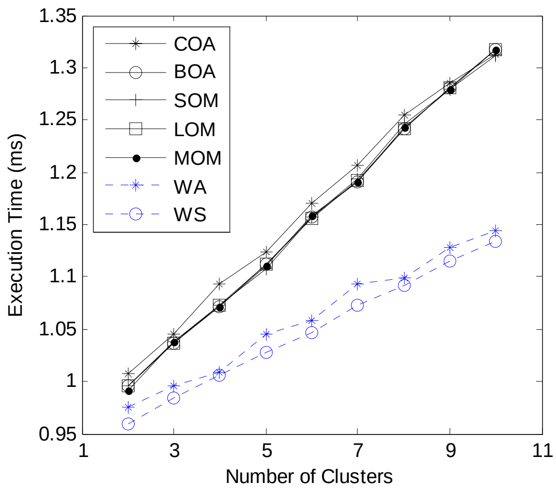


Figure 2. Execution time (ms) for the number of clusters ranging from 2 to 10 when the number of inputs is 4.

From Fig. 1 and Fig. 2, it is evident that COA (shown as solid line with '*' marker) is most computationally expensive while the fastest FIS is WS (shown as dashed line with '○' marker). Although WA (shown as dashed line with '*' marker) requires more time to deliver a solution than WS, but it is faster than all Mamdani FISs under test. Except for COA, the other Mamdani FISs appears to take almost same amount of time during the prediction stage. In addition, the slops for the Mamdani FISs and the Sugeno FISs seems almost same when increasing the number of inputs, but different and the increasing rate is higher for the Mamdani FISs than the Sugeno FISs when increasing the number of clusters (as shown in Fig. 2).

3) Noise resistance

In this set of experiments, the time series of traffic flow was preprocessed to reduce the noise as much as possible and a white noise was subsequently added to the de-noised time series, to generate a set of time series of different signal-to-noise ratio (SNR), starting from 30dB to -10dB with -5dB interval. For each time series, the experiments were independently repeated for 50 times to obtain the averaged performance for all FISs under test. The performance was evaluated for each FIS by examining the MSE of the predicted values from the FIS trained using de-noised time series and noisy time series for each SNR. Fig. 3 and Table IV present the results obtained for different SNRs.

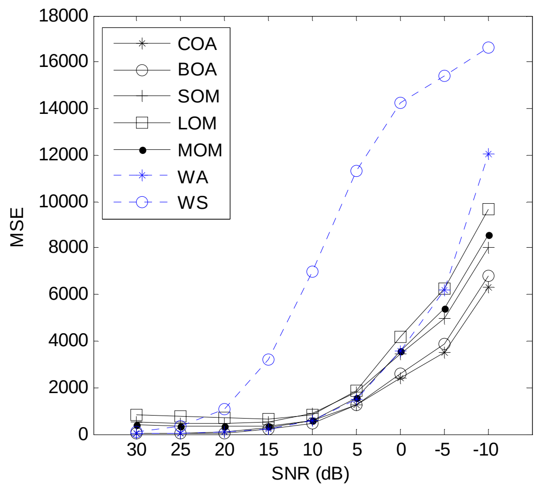


Figure 3. Noise resistance comparison of the FISs (the Mamdani FISs shown as solid lines with markers '*', '○', '+', '□', and '•' for COA, BOA, SOM, LOM, MOM respectively, and the Sugeno FISs denoted by dashed lines with markers '*' and '○' for WA, and WS respectively) for different SNRs.

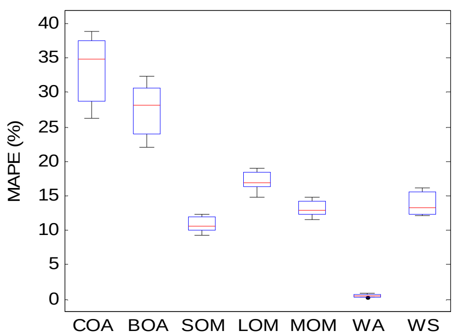
TABLE IV.
MSE VALUES CORRESPONDING TO DIFFERENT SNRS.

SNR (dB)	COA	BOA	SOM	LOM	MOM	WA	WS
30	0.0111	0.0151	0.4904	0.8534	0.4162	0.0190	0.0875
25	0.0369	0.0182	0.4488	0.7679	0.3526	0.0380	0.3181
20	0.1078	0.0606	0.4658	0.6819	0.3181	0.0937	1.0947
15	0.2714	0.1887	0.5464	0.6543	0.3446	0.2319	3.1901
10	0.5573	0.4678	0.8717	0.8502	0.6052	0.5846	7.0153
5	1.2533	1.2342	1.789	1.8471	1.5623	1.5054	11.338
0	2.3945	2.5971	3.4499	4.2061	3.5723	3.5923	14.273
-5	3.5193	3.8954	4.9731	6.2825	5.3721	6.189	15.419
-10	6.3024	6.7862	8.0198	9.6751	8.5917	12.072	16.627

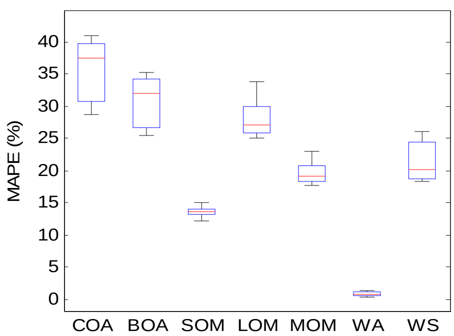
It can be easily identified from Fig. 3 that as the noise level was gradually dominated in the time series, the prediction accuracy generated by WS (shown as dashed line with '○' marker) reduced dramatically and its decline speed is largest among all FISs under test. However, the responses from the other FISs are similar when SNRs are relatively larger, but diverged when SNR is smaller than 5 dB. By a close observation on Fig. 3 and Table IV, COA (shown as solid line with '*' marker and as highlighted in Table IV) is mostly resistant to the noise as its increasing rate in MSE is lowest as compared to the other FISs. Moreover, the performance of WA (shown as dashed line with '*' marker) declined more quickly than all Mamdani FISs when the proportion of noise in the time series became larger and larger. In addition, LOM (shown as solid line with '□' marker) generated the fastest-growing MSE value as compared to the remaining of the Mamdani FISs. Overall, the COA algorithm is mostly insensitive to the noise according to the comparison to the others.

4) Performance consistency

To evaluate the consistency for the FISs under test, the traffic flow data recorded for three different roads (here simply called A, B, and C) over the week, 20th to 26th November 2001, were used and the performance of each FIS was measured using 10-fold blocked cross-validation [28-29]. All parameters were set to be same (3 inputs and 3 rules for all) for the FISs and results were averaged over 100 independent runs. Fig. 4, Fig. 5, and Fig. 6 present the box-plots of MAPEs at training and prediction stages over the week for roads A, B, and C, respectively.

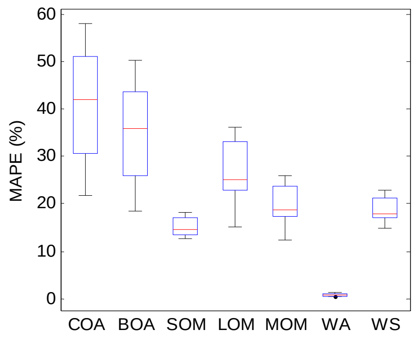


(a) Training performance

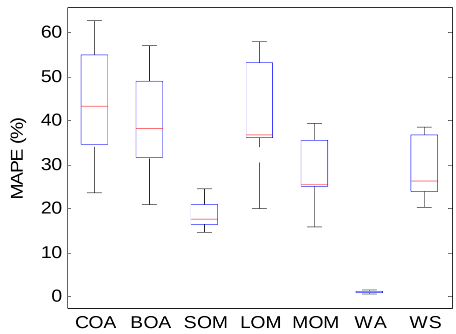


(b) Prediction performance

Figure 4. Box-plots of MAPEs over the week for road A

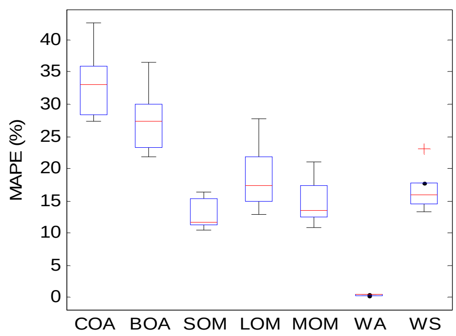


(a) Training performance

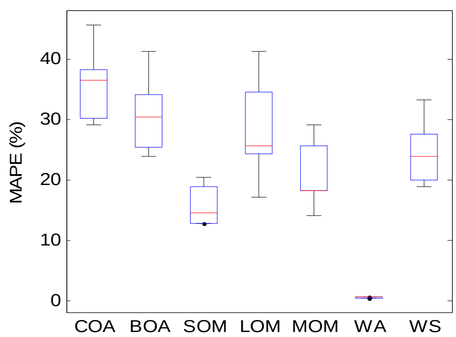


(b) Prediction performance

Figure 5. Box-plots of MAPEs over the week for road B



(a) Training performance



(b) Prediction performance

Figure 6. Box-plots of MAPEs over the week for road C

From Fig. 4, Fig. 5, and Fig. 6, it is evident that WA consistently outperformed as compared to the others as it generated the lowest MAPE values at both training and prediction stages and variations over the week are relatively small. On the other hand, the median MAPEs produced by COA are largest and a relatively large fluctuation over the week can be observed. Among all Mamdani FISs, MOM is the second in performance consistency only to SOM.

5) Missing data

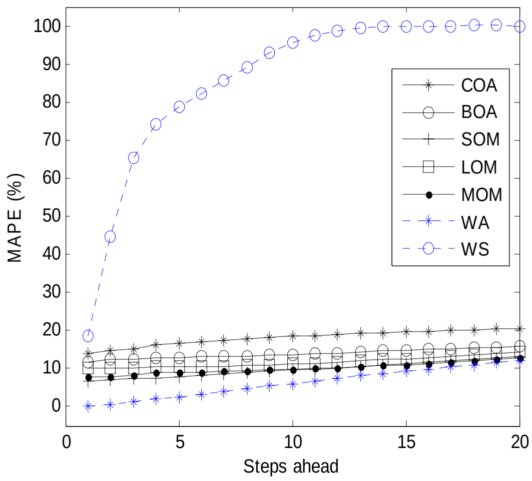
In this paper, the effect of missing data on the prediction performance was examined according to the following procedure: a FIS was constructed using the first half set of the time series that did not have any data missing; the other half set was used as the test data and a predefined number of inputs was replaced by 0 as missing data; the prediction performance for each FIS was evaluated by feeding the set of test data with and without the missing inputs into the FIS; finally, the effect was measured using the index of MAPE by comparing the performance with missing data to that without missing data. In our test, it was assumed that all FISs under test had 5 inputs and therefore we examined the effect with missing inputs from 1 to 5. Also, the other parameters were kept same for all FISs. Each test was performed independently for 100 times and the averaged results are listed in Table V. Except for all inputs fed with 0, LOM (as highlighted in Table V) was influenced much less than the others. In contrast, WA is mostly sensitive to the missing data. One of the interesting facts is that the number of missing inputs did not cause any fluctuation in the prediction performance of SOM. Except for this, the others generally decreased their prediction performances when the number of missing data increased.

TABLE V.
EFFECT OF MISSING DATA ON THE PREDICTION PERFORMANCE MEASURED BY MAPE.

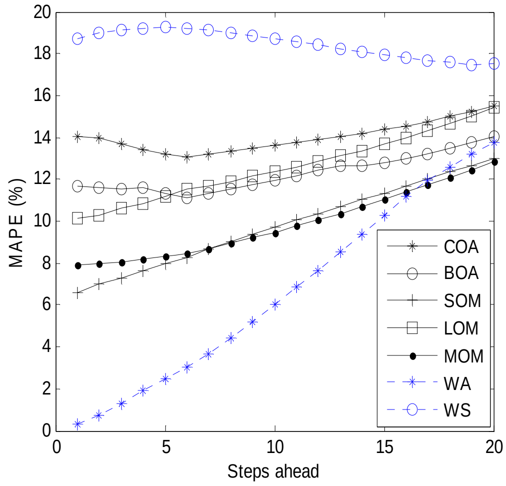
Number of missing data	1	2	3	4	5
COA	241.35	241.79	242.26	245.36	430.38
BOA	311.48	311.88	312.64	317.08	563.95
SOM	504.50	504.50	504. 50	504.50	504. 50
LOM	41.839	43.857	46.855	56.129	617.31
MOM	324.00	324.94	326.64	332.05	660.68
WA (×100)	505.06	618.73	618.15	580.33	451.84
WS	203.12	203.2	201.56	198.97	206.29

6) Multi-step-ahead predictability

When a prediction horizon is higher than 1, the unknown values can be predicted either recursively or directly. The recursive strategy applies the same model of one-step-ahead prediction recursively, using the values estimated as known inputs to predict the next unknown traffic flow. On the other hand, the direct prediction is simple and intuitive as it directly builds a FIS for the specified prediction horizon. In our test, two strategies were evaluated for the prediction horizon ranging from 1 to 20 steps. Again, all parameters (i.e., 3 inputs, 3 rules, and first 300 points and last 300 points used as training and test data) were kept same for the tested FISs and each test was repeated for 100 times. Fig.7 illustrates the prediction performances measured by MAPE for the two strategies.



(a) Recursive prediction



(b) Direct prediction

Figure 7. Multi-step-ahead prediction evaluation for the two strategies: (a) recursive and (b) direct prediction.

As shown in Fig. 7(a), the results obtained by the recursive prediction for the FISs indicate the performance for each FIS generally follows a decreasing pattern with prediction horizon increases. While WA outperformed all the others, WS produced largest errors among all for the test prediction horizon. However, the situation in Fig. 7(b) for the direct prediction is relatively complicated. The traffic flow predicted by WS (shown as dashed line with 'o' markers') are less accurate for a few steps ahead, but its performance gradually improved since the prediction horizon is higher than 5. On the other hand, a reversed pattern is held for COA (shown as solid line with '*' markers') and BOA (shown as solid line with 'o' markers'). The MAPEs produced by the other FISs increase constantly with the prediction horizon being gradually enlarged. WA produced more accurate predictions than the others when the prediction horizon is less than 17 approximately.

To clarify the results presented above, a summary was made by ranking the FISs based on their averaged performance for each evaluation aspect. The overall rank for each FIS is determined by searching the rank dominated over all evaluation aspects. If more than one FIS has been ranked the same place, the place will be assigned to the FIS which has ranked to the place more frequently than the others over all evaluation aspects. Table VI lists the ranks for the FISs. It is clear that WA generally outperformed the others, but WS is the weakest FIS. Also, it should be noticed that although all the tested FISs were ranked, some of them (e.g., SOM, LOM, and MOM) performed almost equally for some evaluation aspects.

TABLE VI.
RANKS OF EACH FIS BASED ON THE EVALUATION RESULTS.

FIS	Model complexity	Execution time	Noise resistance	Performance consistency	Missing data	Multi-step ahead	Overall
COA	6	7	1	7	3	6	6
BOA	5	4	2	6	4	4	4
SOM	3	5	3	2	6	3	3
LOM	4	6	5	5	1	5	5
MOM	2	3	4	3	5	2	2
WA	1	2	6	1	7	1	1
WS	7	1	7	4	2	7	7

V. CONCLUSIONS

This paper presented a comparison between a number of FISs for traffic flow prediction in terms of model complexity, execution time, noise resistance, performance consistency, missing data, and multi-step-ahead predictability. Based on the comparison results, the following main findings can be concluded: 1) as compared to the other tested FISs, WA constantly demonstrated a more accurate prediction, but it is sensitive to noise; 2) a faster processing was realized by the Sugeno FISs mainly due to simplified defuzzification; 3) while LOM is mostly robust to the presence of missing data, WA appears affected severely by missing inputs; 4) WA can deliver a more accurate estimation than the others when predicting the unknown values for more than one-step-ahead in either recursive or direct way. These findings provide an additional reference for choosing a FIS for traffic flow prediction based on historical recordings. For example，based on these findings, one may better to chose WA if the prediction result is used for on-trip routing, when the historical recording is clean and complete or extra preprocessing methods are employed, as the real time requirement of on-trip routing is severer. However, MOM may be the first choice when the information predicted for the next time instant is released via variable messages or radio etc, as these information platforms generally require linguistic forms.

ACKNOWLEDGMENT

This work is partially supported by the Scientific Research Common Program of Beijing Municipal Commission of Education (KM201010005021) and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry (32004011201201). We are grateful for the supports.

REFERENCES

[1] I. Okutani and Y. J. Stephanedes, “Dynamic prediction of traffic volume through Kalman filtering theory,” *Transportation Research Part B*, vol. 18, pp. 1-11, 1984.

[2] D. Gazis, and C. Liu, “Kalman filtering estimation of traffic counts for two network links in Tandem,” *Transportation Research Part B*, vol. 37, pp. 737-745, 2003.

[3] N. Zhang, Y. L. Zhang and H. T. Lu, “Seasonal autoregressive integrated moving average and support vector machine models: prediction of short-term traffic flow on freeways,” *Transportation Research Record*, vol. 2215, 85-92, 2011.

[4] E. Castillo, J. M. Menéndez and S. Sánchez-Cambronero, “Predicting traffic flow using Bayesian networks,” *Transportation Research Part* , vol. 42, pp. 482-509, 2008.

[5] K. Kuhn and A. Nicholson, “Traffic flow forecasting and spatial data aggregation,” *Transportation Research Record*, vol. 2260, pp. 16-23, 2011.

[6] Y. L. Zhang and Z. R. Ye, “Short-term traffic flow forecasting using fuzzy logic system methods,” *Journal of Intelligent Transportation Systems*, vol. 12, pp. 102-112, 2008.

[7] G. J. Shen, *An intelligent hybrid forecasting model for short-term traffic flow*. In Proceedings of the World Congress on Intelligent Control and Automation WCICA, pp. 486-491, Jinan, China, July 2010.

[8] L. Dimitriou, T. Tsekeris and A. Stathopoulos, “Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow,” *Transportation Research Part C*, vol. 16, pp. 554-573, 2008.

[9] G. Y. Zhao, Y. Shen, and Y. L. Wang, “Fuzzy PID position control approach in computer numerical control machine tool,” *Journal of Computers (Finland)*, vol. 8, pp. 622-629, 2013.

[10] Z. Y. Song, X. L. Song, C. Y. Liu, and Y. B. Zhao, “Research on real-time simulation and control of linear 1-stage inverted pendulum,” *Journal of Computers (Finland)*, vol. 8, p 896-903, 2013

[11] J. S. Gao, X. H. Xu, and C. He, “A study on the control methods based on 3-DOF helicopter model,” *Journal of Computers (Finland)*, vol. 7, pp. 2526-2533, 2012.

[12] L. Zhao and F. Y. Wang, *Short-term fuzzy traffic flow prediction using self-organizing TSK-type fuzzy neural network*, In Proceedings of the IEEE International Conference on Vehicular Electronics and Safety, pp. 1-6, Beijing, China, December 2007.

[13] E. H. Mamdani and S. Assilian, “An experiment in linguistic synthesis with a fuzzy logic controller,” *International Journal of Man-Machine Studies*, vol. 7, pp. 1-15, 1975.

[14] M. Sugeno, *Industrial applications of fuzzy control*. New York, NY: Elsevier Science Inc, 1985.
[15] J. R. Jang, C. T. Sun and E. Mizutani, *Neuro-fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. US: Pearson Education, 1996.
[16] M. Sugeno and G. T. Kang, "Structure identification of fuzzy model," *Fuzzy Sets and Systems*, vol. 28, pp. 15-33, 1988.
[17] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Systems Man Cybernet*., vol. 15, pp. 116-132, 1985.
[18] A. Hamam and N. D. Georganas, *A comparison of Mamdani and Sugeno fuzzy inference systems for evaluating the quality of experience of hapto-audio-visual applications*, IEEE International Workshop on Haptic Audio Visual Environments and their Applications, pp. 18-19, Ottawa, Canada, September 2008.
[19] J. J. Jassbi, P. J. A. Serra, R. A. Ribeiro and A. Donati, *A comparison of mandani and sugeno inference systems for a space fault detection application*, World Automation Congress, pp. 1-8, Waikoloa, Hawaii, September 2008.
[20] M. Firat, M. E. Turan and M. A. Yurdusev, "Comparative analysis of fuzzy inference systems for water consumption time series prediction," *Journal of Hydrology*, vol. 374, pp. 235-241, 2009.
[21] K. Guney and N. Sarikaya, "Comparison of Mamdani and Sugeno fuzzy inference system models for resonant frequency calculation of rectangular microstrip antennas," *Progress in Electromagnetics Research B*, vol. 12, pp. 81-104, 2009.
[22] H. Ying, Y. S. Ding, S. K. Li and S. H. Shao, "Comparison of necessary conditions for typical Takagi-Sugeno and Mamdani fuzzy systems as universal approximators," *IEEE Trans. Man, Syst., Cybern.*, vol. 29, pp. 508-514, 1999.
[23] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Norwell, MA: Kluwer Academic Publishers, 1981.
[24] T. Y. Li, Y. Chen, and J. S. Zhang, "Logistics service provider segmentation based on improved FCM clustering for mixed data," *Journal of Computers (Finland)*, vol. 7, pp. 2629-2633, 2012
[25] G. Tsekouras, H. Sarimveis, E. Kavakli, and G. Bafas, "A Hierarchical Fuzzy-clustering Approach to Fuzzy Modeling," *Fuzzy Sets and Systems*, vol. 150, pp. 245-66, 2005.
[26] M. G. Karlaftis, E. I. Vlahogianni and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transportation Research Part C*, vol. 13, pp. 211-34 2005.
[27] Y. Zhang and Y. Liu, "Analysis of peak and non-peak traffic forecasts using combined models," *Journal of Advanced Transportation*, vol. 45, pp. 21-37, 2011.
[28] M. Misiti, Y. Misiti, G. Oppenheim and J.M. Poggi, *Wavelets and their Applications*, Hermes Lavoisier, ISTE Publishing Knowledge, 2007.
[29] C. Bergmeir and J. Benitez, "On the use of cross-validation for time series predictor evaluation," *Information Sciences*, vol. 191, pp. 192-213, 2012.
[30] C. Bergmeir and J. M. Benitez, *Forecaster performance evaluation with cross-validation and variants*. In Proceedings of the 11th International Conference on Intelligent Systems Design and Applications ISDA, pp. 849-854, Córdoba, Spain, November 2011.

Yang Wang was born in Shanxi, China, in 1975. He received his MS degree from The University of York, UK, in 2004, and his PhD degree in intelligent control from Loughborough University, UK, in 2007. He was a Post-doc researcher at University of Portsmouth, UK. Currently, he is a lecturer at Beijing University of Technology, China. His interests are in fuzzy modeling techniques, time series prediction, traffic data processing, and traffic information management.

Yanyan Chen was born in Henan, China, in 1970. She received her PhD degree in civil engineering from Haerbin Institute of Technology, China, in 1997. She was a Post-doc researcher at Harbin Institute of Technology, China. She was a visiting scholar of Imperial College London, UK. Currently, she is a professor at Beijing University of Technology, China. Her research interests include traffic data processing, traffic information management traffic network optimization, and path planning

.