Traffic & Transportation

Estimating Vehicle Turn-In Rate of Expressway Rest Areas via ETC Gantry Data – An ADPC-GMM Approach

Yubin ZHENG[1], Cheng CHENG[2], Yong ZHANG[3], Lingyi WANG[4], Qixuan LI[5], Hailin ZHANG[6]

Original Scientific Paper
Submitted: 25 Dec 2023
Accepted: 11 Apr 2024

[1] beijingzyb@qq.com, Tongji University, Key Laboratory of Road and Traffic Engineering of the Ministry of Education
[2] Corresponding author, michael_cheng1989@126.com, Tongji University, Key Laboratory of Road and Traffic Engineering of the Ministry of Education
[3] 21155818@qq.com, Hebei Province Expressway Jingxiong Management Centre
[4] 1028886841@qq.com, Hebei Province Expressway Jingxiong Management Centre
[5] 517868920@qq.com, Hebei Province Expressway Jingxiong Management Centre
[6] 591274158@qq.com, Jiaoke Transport Consultants LTD

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT

Vehicle turn-in rate is a critical and widely adopted input for expressway rest area design and operation. With the implementation of expressway ETC gantries, the ERA turn-in rate can be further estimated by measuring the travel speed distribution via ETC gantry data. This paper proposed an adaptive density peak clustering Gaussian mixture model (ADPC-GMM) for ERA turn-in rate estimation. The ADPC algorithm is applied to generate the GMM's inputs accommodating to the traffic characteristic of ERA expressway segments and GMM would further provide the turn-in rate estimation results. To validate the model precision, the turn-in rate data of four selected ERAs in Sichuan, China, as well as the ETC gantry data of their corresponding expressway sections are obtained. According to the estimation results, the MAE and RMSE are 0.0228 and 0.0267 for the passenger car scenario and 0.0264 and 0.0356 for the commercial truck scenario, respectively. These results are also at the lowest level compared with the results acquired from ordinary GMM, K-Means and DBSCAN algorithms. The proposed method has good applicability for vehicle turn-in rate estimation and can be deployed at different ERAs, especially those ERAs without traffic monitoring.

KEYWORDS

expressway rest area; turn-in rate; ETC gantry; adaptive density peak clustering; Gaussian mixture model.

1. INTRODUCTION

An expressway rest area (ERA) is a specially constructed place deployed along the expressway to provide rest and recreational spaces, as well as travel-related services for travellers [1]. It performs a critical role in easing travel fatigue, increasing expressway travel quality, and at the same time creating service-related incomes for its operation and maintenance [2]. To meet travellers' parking and service requirements and in avoidance of overinvestment, the information on ERA travel volumes, or the percentage of expressway travel volume entering ERA, commonly known as turn-in rate, is basic and critical in both the ERA design and operation process [3].

Generally, ERA's vehicle volume can be obtained via manual or automatic vehicle counting techniques. When adopting the turn-in rate for ERA design and analysis, the traffic volume of the expressway section is further needed and can be easily collected via roadside surveillance facilities. Al Kaisy et al. investigated the entering traffic volumes, characteristics and trends at rest areas in the state of Montana, and found that a majority of the ERA turn-in rates were distributed between 8.4 to 12.3% [2, 4]. Kay et al. studied the factors that impact the ERA travel preference via regression models by investigating the passenger car and commercial truck turn-in rate of 47 ERA in Michigan, USA [5]. Griffin and Yan conducted a six-month study on ERAs in Louisiana and found that the mainline vehicle turn-in rate within the state ranges from 2% to 20% [6]. Similarly, Shen et al. collected the ERA's travel characteristics at Henan, China and proposed prediction models for measuring ERA's turn-in rate [7, 8]. Sun et al. proposed an ERA travel demand prediction model via XGBoost and an improved particle swarm optimisation model with over 3 months of observation data [9]. Zhao et al. proposed an ERA short-term traffic flow prediction model via a seasonal-trend decomposition procedure based on LOESS and with an optimal model selection process [10]. The proposed model provides further demand information for ERA operation. Although the vehicle counting method has been widely adopted worldwide, it is facilitated with additional surveillance facilities or manpower and leads to the increase of ERA operation costs.

Recently, with the implementation of roadside surveillance techniques, especially the expressway ETC gantry, enormous travel data, including vehicle plate numbers, vehicle types, temporal-spatial location, etc. can be captured with high precision [11, 12]. Extensive studies have been proposed on measuring travel patterns via ETC gantry data and some of the research attempted to adopt it into ERA turn-in rates estimation. Liao et al. implemented the XGBoost method for turn-in rate prediction considering temporal traffic volume and traffic composition features. The RMSE and MAE achieved 0.033 and 0.026 [13]. Cai et al. proposed an ERA access vehicle recognition method by considering the vehicle travel speed of three consecutive road sections, and spatiotemporal and traffic condition features [14]. Although the precisions of these two turn-in rate estimation methods are high, the traffic data of the upstream and downstream sections need to be further collected. Moreover, the current models assume that the number of upstream and downstream sections is unique and would not be accommodated to the conditions with the upstream merging or downstream diversion sections. Therefore, the proposed models are only applicable to ERAs with favourable data collection conditions.

Since the turn-in vehicles would stay at ERA for some time, their travel speed would have an obvious decrease compared to the through vehicles, leading to a multi-peak travel speed distribution pattern at the expressway section where ERA was built. The distribution information can be further implemented for ERA turn-in rate estimation. Currently, clustering methods including k-means [15], DBSCAN [16] and Gaussian mixture model (GMM) [17] have been widely adopted for measuring travel characteristics. However, the preset parameters of these methods are hard to determine due to the varying features of section length, speed limit and number of vehicle clusters at ERA's road section. These differences cause difficulty for ERA turn-in rate estimation applications on a large scale. To fill in this gap, this paper proposes a modified GMM model for estimating ERA turn-in rates by adopting an adaptive density peak clustering algorithm (ADPC). The ADPC would be able to adjust the preset parameters for GMM adaptively, thus increasing its generalisation capability for ERA turn-in rate estimation.

The remainder of this paper is organised as follows. Section 2 presents the framework for ERA turn-in rate estimation. Section 3 introduces the details of the proposed methodology, especially the process of the ADPC-GMM algorithm. A case study with four different ERAs is adopted to verify model precision and is proposed in Section 4. Finally, the conclusion, limitations and future research directions are further discussed in Section 5.

2. ERA TURN-IN RATE ESTIMATION FRAMEWORK

The ERA turn-in rate estimation framework via ETC gantry data is presented in *Figure 1* with three major steps: data collection, data processing and ERA turn-in rate estimation. In the data collection step, the data from the upstream and downstream ETC gantry of the ERA's expressway section would be collected with vehicle information and ETC gantry mileage information. With the ETC gantry data, the vehicle travel time and travel speed at the ERA's road section are estimated according to the unique vehicle plate number. After outlier elimination, the travel speed data would be implemented in the ERA turn-in estimation process. In the final step, the ADPC algorithm would be first adopted to achieve the optimal number of clusters and the initial input for the GMM. The final clustering results would be generated from the GMM model. According to the mean travel speed threshold for selecting ERA turn-in vehicles, the ERA turn-in vehicle cluster would be finally identified and the turn-in rate can be calculated.

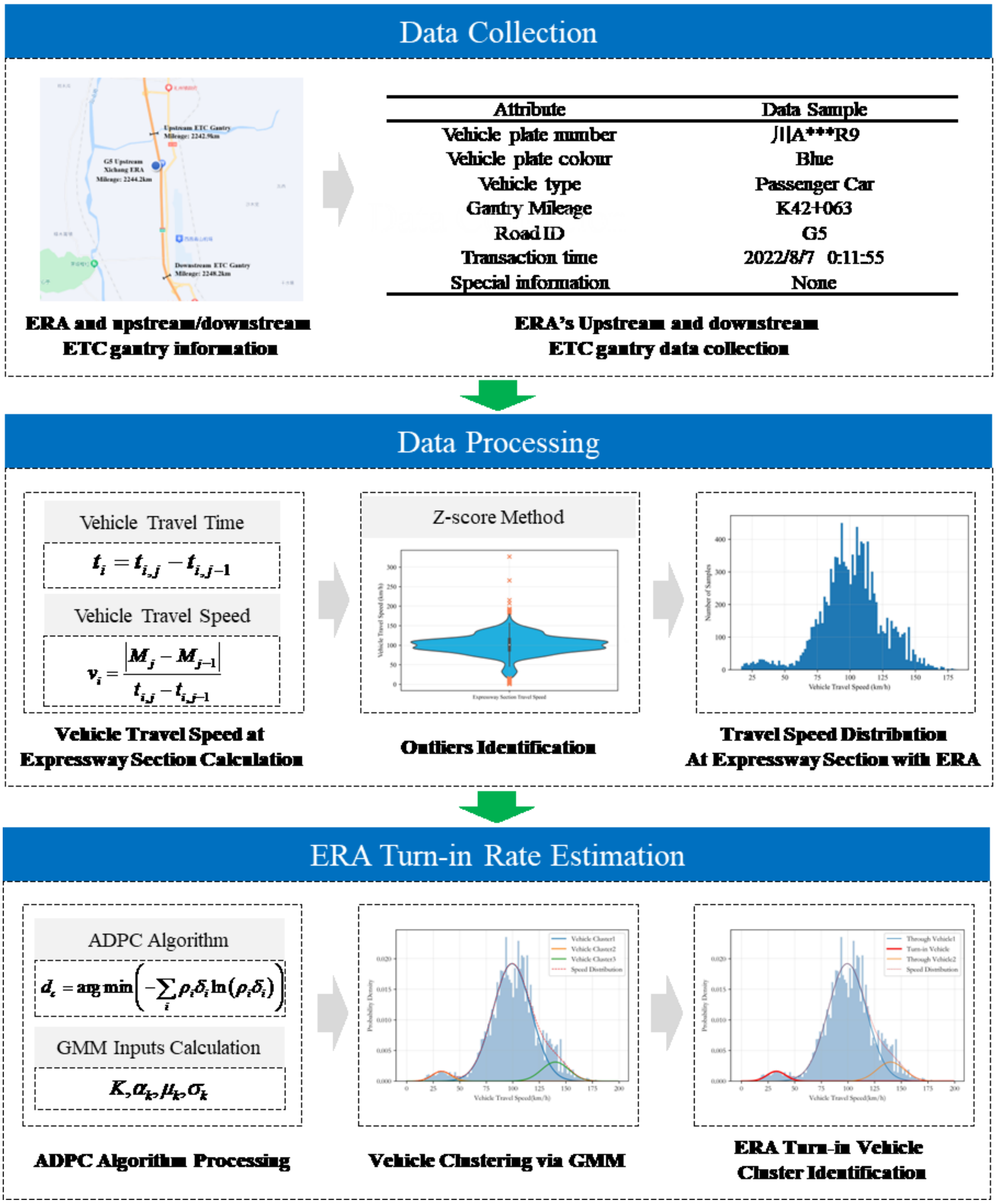


Figure 1 – The framework for ERA turn-in rate estimation via ETC gantry data

3. METHODOLOGY

3.1 Data collection

ETC gantries are the roadside facilities deployed to support section-based expressway tolling. It integrates a variety of sensing techniques including radio frequency identification, license plate recognition and vehicle type identification. Therefore, both the information on vehicles with ETC on-board units and vehicles using manual passing cards can be captured. To measure the ERA's vehicle turn-in rate, the ETC gantry data containing vehicle plate number, vehicle type, gantry mileage, road ID, transaction time, etc. at both the upstream and downstream ETC gantry closest to ERA should be collected. We have to mention that some of the ETC gantries may be deployed before the upstream merging area or after the downstream diversion area of the expressway section with ERA. For these special but common scenarios, the closest ETC gantries at each expressway prior to the merging area or constructed later than the diversion area are needed for analysis. A proper sample of the ETC gantry data is presented in *Table 1*.

Table 1 – A data sample of the ETC gantry

Attribute	Data sample
Vehicle plate number	川 A***R9
Vehicle plate colour	Blue
Vehicle type	Passenger car
Gantry mileage	K42+063
Road ID	G5
Transaction time	2022/8/7 0:11:55
Special information	None

3.2 Data processing

Vehicle travel speed calculation

To calculate the section travel speed of passing vehicles, the data from the upstream ETC gantry would be first extracted according to the transaction time and would be matched to the specific data with the same vehicle plate number, colour and with the later but closest transaction time generated from the downstream ETC gantry. The time difference between the matched two pieces of data would be the travel time of the specific vehicle plate number:

$$t_i = t_{i,j} - t_{i,j-1} \tag{1}$$

where t_i is the travel time of the i^{th} vehicle at the expressway section with ERA. $t_{i,j}$ is the transaction time of i^{th} vehicle at the j^{th} ETC gantry.

With vehicle section travel time and the mileage information of the upstream and downstream ETC gantry, the travel speed can be calculated via *Equation 2*, where M_j is the mileage of the j^{th} ETC gantry.

$$v_i = \frac{|M_j - M_{j-1}|}{t_{i,j} - t_{i,j-1}} \tag{2}$$

Outliers identification

Due to the existence of equipment failure, wireless signal interference, misrecognition and other issues, duplication, anomality and data missing may occur during the travel transaction process. The calculated vehicle travel speed would be further inspected for further estimation. The data inspection rules are designed as follows:

1) To avoid data duplication, for each vehicle passing the expressway section with ERA, the difference in transaction time at ERA's upstream ETC gantry should be greater than 15 minutes. Otherwise, the travel speed data with the latter transaction time should be removed from the dataset.
2) To avoid data anomality, the travel speed outliers would be identified via the Z-score method [18]. The data with travel speeds ranging from $\overline{v} - 3\sigma_v$ to $\overline{v} + 3\sigma_v$ would be reserved, where $\overline{v}$ is the average travel speed of the selected expressway section and σ_v is the travel speed standard deviation.

3.3 ERA turn-in rate estimation

Gaussian mixture model

The GMM is able to separate the data that come from several groups by calibrating each group's distribution parameter and data proportion via the Gaussian distribution model. Due to its model stability, it has been widely adopted in studies related to travel time or travel speed pattern analysis [19, 20]. In this study, the travel speed distribution of an expressway section with ERA can be calibrated via the following function:

$$P(x) = \sum_k \alpha_k N(x|\mu_k, \sigma_k) \tag{3}$$

$$0 \le \alpha_k \le 1 \tag{4}$$

$$\sum_k \alpha_k = 1 \tag{5}$$

$$N(x|\mu_k,\sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} exp\left(-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right) \tag{6}$$

where $P(x)$ is the probability of vehicle with travel speed equaling to x. α_k is the proportion of kth group vehicle. $N(x|\mu_k,\sigma_k)$ is the travel speed Gaussian distribution function of the kth group vehicle with mean travel speed and its standard deviation equaling to μ_k and σ_k, respectively.

The optimal calibration results of α_k, μ_k, and σ_k would be achieved once the log-likelihood function of *Equation 3* reaches to the maximum, as shown in *Equation 7* and the results can be acquired via the expectation-maximisation algorithm with pre-set number of clusters *K* and the initial value of α_k, μ_k, and σ_k according to Dempster et al. [21]

$$LL(\mu_k,\sigma_k,\alpha_k) = \sum_x log\left(\sum_k \alpha_k P_k(x|\mu_k,\sigma_k)\right) \tag{7}$$

Adaptive density peak clustering algorithm

Although the GMM can be directly implemented for vehicle clustering, the model precision and efficiency are sensitive to the initial inputs of the parameters. The number of pre-determined parameters is also related to the input of the number of groups, which indicated that the input data structure would not be stable when implemented in ERA's expressway section with different travel speed distribution features [22].To reduce the number of preset parameters and improve GMM's model universality for ERA turn-in rate estimation, the adaptive density peak clustering algorithm is adopted to determine the optimal number of vehicle groups and estimate the initial value of α_k, μ_k, and σ_k.

The initial density peak clustering (DPC) method was proposed by Rodriguez and Laio [23]. The method assumes that the clustering centre of each group is surrounded by data and the region closer to the clustering centre is with higher sample density. At the same time, the centre of each group is located relatively far from other group centres. To capture the features, Rodriguez and Laio defined two key indicators, namely local density ρ_i and minimum distance to another point with higher density δ_i, and for a specific sample x_i, ρ_i and δ_i can be estimated via *Equation 8–11*, where d_c is the preset cutoff distance. The clustering centre would be selected according to the product of ρ_i and $\delta_i$in descending order. Once a clustering centre is selected, the samples within the range of d_c from the clustering centre will be extracted. The clustering centre selection process would continue for the rest of the samples once all the samples are stamped with a cluster. The initial α_k, μ_k, and σ_k for GMM can be calculated according to the grouping results from the density peak clustering algorithm.

$$\rho_i = \sum_{i\neq j} f(d_{ij} - d_c) \tag{8}$$

$$d_{ij} = \|x_i - x_j\| \tag{9}$$

$$f(x) = \begin{cases} 1, x < 0 \\ 0, x \geq 0 \end{cases} \tag{10}$$

$$\delta_i = \begin{cases} max(d_{ij}), \forall\rho_i > \rho_j \\ min(d_{ij}), \forall\rho_j > \rho_i \end{cases} \tag{11}$$

Compared with the ordinary GMM, the DPC algorithm uses cutoff distance d_c for vehicle clustering instead of a number of vehicle groups and each group's travel speed distribution parameters, which improves clustering simplicity [24]. We further adopted the information entropy to determine the optimal d_c. The optimal d_c would achieve the minimum entropy of the clustering result [25]. Since the centre of a cluster would with both greater ρ_i and δ_i, we define the information of sample *i* being a cluster centre as the product of ρ_i and δ_i. The optimal d_c would achieve minimum information entropy as shown in *Equation 12*. By iterating d_c automatically and selecting the result with minimum entropy, the DPC method would be adaptive to various data features.

$$d_c \in argmin\left(-\sum_i \rho_i \delta_i \ln(\rho_i \delta_i)\right) \tag{12}$$

By implementing the APDC method, the travel speed data would be automatically classified into different groups and the vehicle proportion, mean travel speed and travel speed variance of each group (α_k, μ_k and σ_k) would be set as the initial value of GMM inputs for further clustering.

Turn-in vehicle cluster identification

Although the proposed ADPC-GMM is capable of figuring out vehicle clusters via travel speed distribution of expressway section with ERA, the threshold for determining the specific vehicle cluster as an ERA turn-in vehicle is still needed. In this paper, the average travel speed of passenger car and commercial truck clusters for accessing ERA are set to be less than 80% and 70% of the speed limit, respectively, and the proportion summation of clusters of specific vehicle types following the determined rule is regarded as the ERA turn-in rate.

Meanwhile, as for expressways, freight vehicle volume is relatively low, and thus it may be difficult for the ADPC-GMM model to split the freight travel speed data into multiple clusters, due to the sample number limitation. For this scenario, the travel speed data are sorted and for the two consecutive samples with maximum speed difference, the average speed of the two samples is taken as the threshold, and the samples with travel speed lower than the threshold are regarded as ERA turn-in vehicles.

For the ERA's expressway section with multiple upstream or downstream ETC gantries, the ERA turn-in rate would be the weighted proportion of selected clusters estimated from travel speed data of different ETC gantry pairs:

$$R = \frac{\sum_p Q_p p_{t,p}}{\sum_p Q_p} \tag{13}$$

where R is the ERA turn-in rate, Q_p is the vehicle volume estimated from the p^{th} ETC gantry pair and the $p_{t,p}$ is the proportion of the vehicle cluster identified as ERA turn-in vehicle group of the p^{th} ETC gantry pair.

3.4 Model validation

To validate the effectiveness of the proposed RF model, mean absolute error (MAE) and root mean square error (RMSE) are adopted. The calculation procedures are presented in *Equation 14* and *Equation 15*, where n is the number of ERA turn-in rate samples, R_i and R_i' are the ERA turn-in rate via video surveillance and estimated turn-in rate via ETC gantry data, respectively.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(R_i - R_i')^2}{n}} \tag{14}$$

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^{n} |R_i - R_i'| \tag{15}$$

4. CASE STUDY AND RESULT ANALYSIS

4.1 Case information

To test the accuracy of the proposed ADPC-GMM ERA turn-in rate estimation method, four different rest areas, namely, G5 upstream Xichang ERA, G93 upstream Pujiang ERA, S1 upstream Anzhou ERA and G75 upstream Nanbu ERA at Sichuan Province, China, are selected for model validation. The video surveillance data of the four ERAs access ramp on 7 August 2022, as well as the upstream and downstream ETC gantries data at the expressway section with ERAs, were collected from the Sichuan Highway Management Centre. The location of ERAs and their upstream and downstream ETC gantries, as well as the expressway section speed limits, are shown in *Figure 2*.

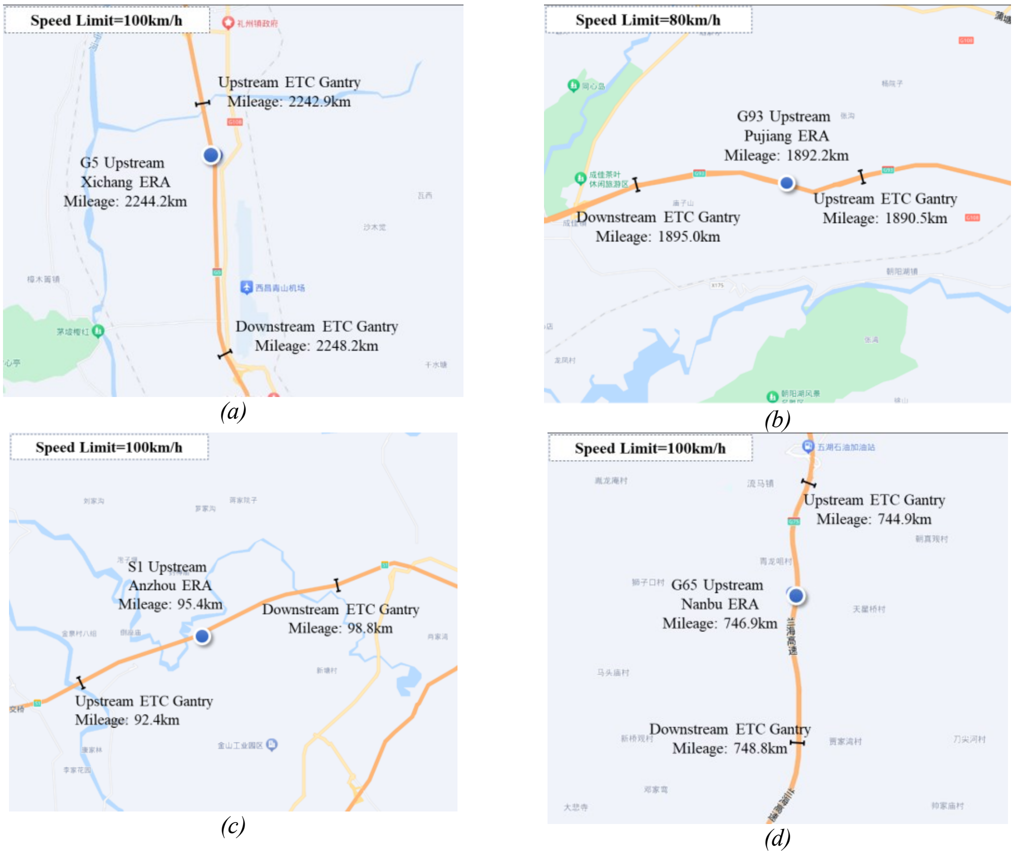


Figure 2 – The location information of selected ERAs and the upstream and ETC gantries: a) G5 upstream Xichang ERA; b) G93 upstream Pujiang ERA; c) S1 upstream Anzhou ERA; d) G75 upstream Nanbu ERA

The number of ERAs' turn-in passenger cars and trucks each hour from 8:00 to 16:00 are manually counted from surveillance video and the turn-in rate would be calculated using the hourly road section passing volume calculated from ETC gantries data. The hourly travel volumes of each ERA's section are presented in *Table 2*. At the same time, the ADPC-GMM model is implemented to estimate turn-in rates using hourly vehicle travel speed acquired from ETC gantry data. The cutoff distances d_c are set to be ranged from 20 to 50 and the iteration interval is determined as 5 for model usage. The threshold for the expectation-maximisation algorithm for GMM result generation is set as 10^{-5}.

Table 2 – Traffic volume of the selected ERAs

Time	Hourly travel volume (vehicle per hour)							
	G5 upstream Xichang ERA		G93 upstream Pujiang ERA		S1 upstream Anzhou ERA		G75 upstream Nanbu ERA	
	Passenger car	Truck	Passenger car	Truck	Passenger car	Truck	Passenger car	Truck
8:00-9:00	213	125	588	59	238	126	375	46
9:00-10:00	292	126	1012	55	394	96	442	56
10:00-11:00	354	101	1145	79	468	143	497	70
11:00-12:00	500	94	1151	84	423	140	399	76
12:00-13:00	524	88	851	70	317	153	292	60
13:00-14:00	640	65	670	66	389	138	385	51
14:00-15:00	798	102	706	65	438	110	536	66
15:00-16:00	785	98	669	58	466	136	507	49
16:00-17:00	794	96	602	66	482	105	428	60

4.2 Turn-in rate result and analysis

To validate the clustering capability of the ADPC-GMM model, the vehicle clustering results of both passenger cars and commercial trucks of the selected ERAs at 10:00 on 7 August 2022, are presented in *Figure 3* and *Figure 4*. It can be clearly shown that the optimal number of vehicle clusters, as well as the turn-in vehicle cluster, can be automatically identified. The separated clusters' distributions are also able to describe the whole distribution of the travel speed sample of the expressway section with ERA, especially for the passenger cars estimation results.

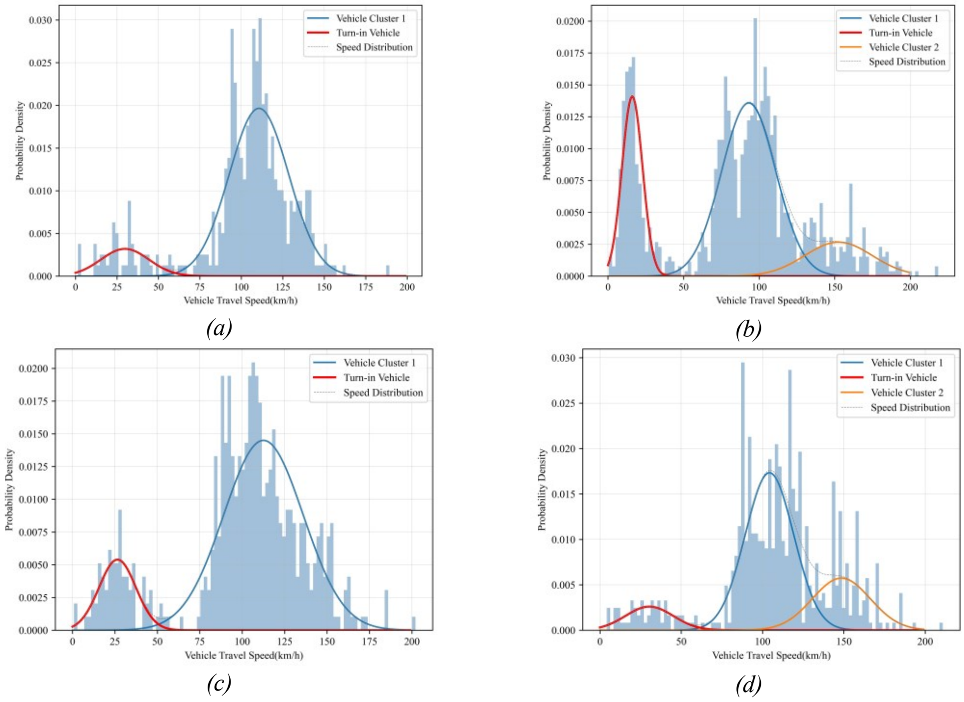


Figure 3 – The vehicle cluster results of passenger cars at selected ERAs: a) the passenger car clusters at G5 upstream Xichang ERA; b) the passenger car clusters at G5 upstream Pujiang ERA; c) the passenger car clusters at S1 upstream Anzhou ERA; d) the passenger car clusters at G75 upstream Nanbu ERA

The turn-in trucks can be extracted from the ADPC-GMM algorithm via travel time acquired from the ETC gantry data of the ERA expressway section according to *Figure 4*. However, since the truck volumes at each ERA expressway section are relatively small, the misclassified truck would have a higher influence on turn-in rate estimation accuracy. The model error may be higher than the passenger car estimation results.

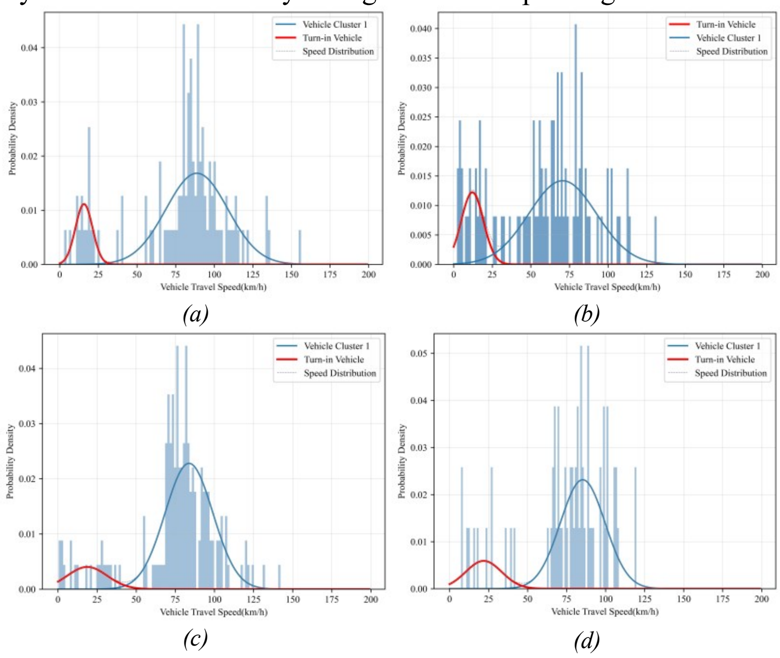
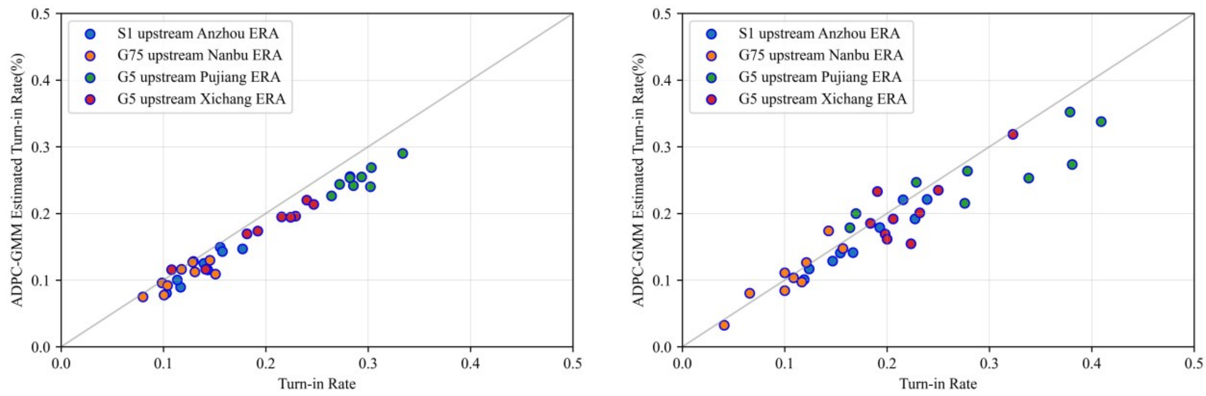


Figure 4 – The vehicle cluster results of trucks at selected ERAs: a) the truck clusters at G5 upstream Xichang ERA; b) the truck clusters at G5 upstream Pujiang ERA; c) the truck clusters at S1 upstream Anzhou ERA; d) the truck clusters at G75 upstream Nanbu ERA

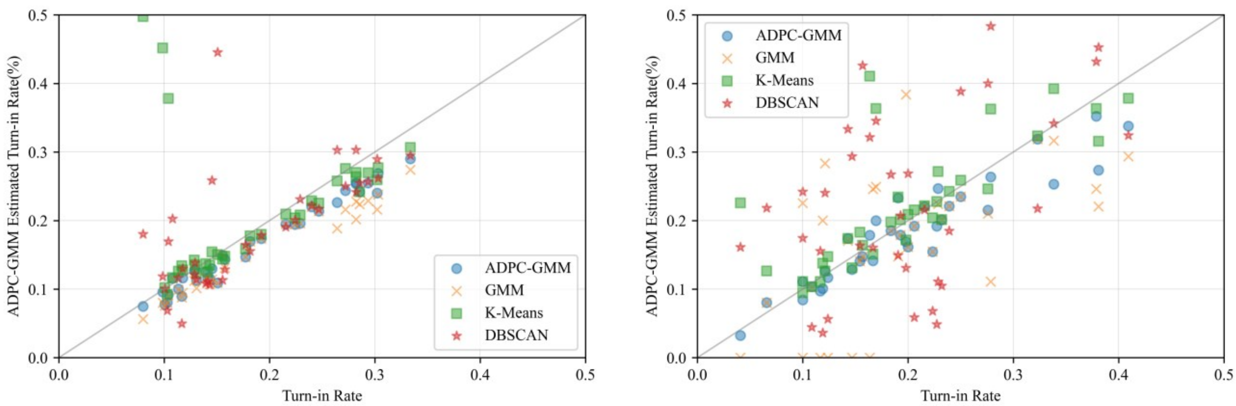
The estimated passenger car and commercial truck turn-in rates of selected ERAs are presented in *Figure 5*. The MAE of the passenger car and commercial truck turn-in rates are 0.0228 and 0.0264, respectively, which indicates the accuracy of the proposed ADPC-GMM model. It can also be seen that the scatters of the actual turn-in rates and the estimated turn-in rate of all the selected ERAs are distributed near the contour line in *Figure 5*. However, the ADPC-GMM model tends to underestimate the turn-in rates since most of the estimated results are less than the actual values, which is especially obvious in the case of passenger car estimation. This may be due to the fact that the proposed method only uses the vehicle travel speed of the ERA's expressway segment and lacks information on speed variation patterns of consecutive sections. Therefore it failed to deal with the vehicles with high travel speeds at the upstream and downstream sections, and with relatively short parking duration at ERA on the analysed section, which caused the underestimation. In terms of the dispersion of estimation results, the variability of passenger car turn-in rates is much lower, with an RMSE of 0.0267, while the estimations of commercial truck turn-in rates are more discrete, with an RMSE of 0.0356.



(a) *(b)*

Figure 5 – The turn-in rate estimation results via ADPC-GMM model: a) estimated passenger car turn-in rate via ADPC-GMM model; b) estimated commercial truck turn-in rate via ADPC-GMM model

To further validate the accuracy of the proposed model, three different cluster algorithms, namely GMM, k-means and DBSCAN, are also employed for ERA turn-in rate estimation. The number of clusters is defined as two for these algorithms. The input average travel speeds for the two separated clusters in GMM are 50 km/h and 80 km/h with sample percentage and standard deviation equaling 0.5 and 15 km/h, respectively. In the DBSCAN algorithm, the epsilon is set as 5 km. The model comparison results of passenger car and commercial truck turn-in rates, as well as the calculated MAE and RMSE, are presented in *Figure 6* and *Table 3*, respectively.



(a) *(b)*

Figure 6 – Turn-in rate model comparison: a) passenger car turn-in rate estimation comparison; b) commercial truck turn-in rate estimation comparison

With the implementation of the ADPC algorithm and the proposed ERA turn-in rate criteria for small sample scenarios, the turn-in rate estimation results have achieved high improvement compared with the ordinary GMM. The model results have the lowest MAE and RMSE in both passenger car and commercial truck conditions and over 50% accuracy increase compared with GMM. The proposed method also figures out the problem of GMM estimation with small data samples. Along with the ADPC-GMM model, the k-means method also acquires relatively accurate estimation in a majority of cases, however, this method has its shortage in achieving reasonable estimation under specific travel speed datasets. As for the DBSCAN algorithm, the distribution of the estimation errors is highly dispersed and is not reasonable for turn-in rate estimation.

Table 3 – Model comparison result analysis

	Passenger cars		**Commercial trucks**	
	MAE	**RMSE**	**MAE**	**RMSE**
ADPC-GMM	0.0228	0.0267	0.0264	0.0356
GMM	0.0337	0.0400	0.0744	0.1005
K-means	0.0419	0.1033	0.0379	0.0672
DBSCAN	0.0400	0.0643	0.1016	0.1193

5. CONCLUSION AND FUTURE RESEARCH

ERAs are the necessary places deployed along expressways for easing travellers' travel fatigue and increasing expressway travel quality. To design and operate ERAs to satisfy travel demand and avoid overinvestment, the ERA turn-in rate is the basic and critical input for ERA design and operation. Currently, the ERA's turn-in rates can be calculated by obtaining turn-in demand and expressway travel volume with manual or automatic vehicle counting techniques. Recently, with the implementation of expressway ETC gantries, enormous travel data can be captured with high precision and can be further implemented for ERA turn-in rate estimation, especially for those ERA uncovered with video surveillance.

For those vehicles that drive into ERAs, their travel speed at the expressway section with ERA would be lower than the through vehicles. Therefore the ERA turn-in rate estimation problem can be regarded as a data clustering problem. Although the ordinary clustering algorithm, such as GMM or k-means, can be directly implemented for turn-in rate estimation, the pre-set parameters of these methods cannot be decided adaptive to the various features of vehicle travel time distribution at ERA's road sections. In this paper, an ADPC-GMM method is proposed for turn-in rate estimation adaptively via the ETC gantry data. The ADPC algorithm is implemented to obtain the initial inputs for GMM, which improves GMM's generalisation capability for various vehicle travel speed distribution patterns of ERA's expressway segments. To validate the model's accuracy, four different ERA's turn-in rates are obtained and the turn-in rates are estimated by ADPC-GMM and other three ordinary clustering methods with the usage of ETC gantry data. Among the tested four methods, the proposed method achieved the minimum MAE and RMSE in both passenger car and commercial truck scenarios which indicated its model precision.

Similar to other studies, this research has its limitations. Due to the difficulty in obtaining both video surveillance data and ETC gantry data with a longer time span, the number of samples for model validation is relatively small. The proposed method would need further validation in the future. The proposed model only considers the travel speed distribution of the ERA's expressway segment and fails to capture the speed-varying pattern among consecutive segments. The model accuracy would be further improved once the vehicle travel speed at both prior and latter sections are obtained. Moreover, other innovative algorithms providing turn-in rate estimations with accuracy and efficiency would offer great help in ERA design and operations and would be the direction for further research.

ACKNOWLEDGEMENTS

This work was sponsored by Key Technologies Research of New Generation Collaborative Control Smart Highway via Data and All Media Integration (JX-202002) and the Science and Technology Program of Zhejiang Provincial Department of Transportation (2023001).

REFERENCES

[1] Ding W, Wang Y, Chu P, et al. Safety-oriented planning of expressway truck service areas based on driver demand. *Frontiers in Public Health*. 2022;19:976495. DOI: 10.3389/fpubh.2022.976495.

[2] Al-Kaisy A, et al. Traffic use of rest areas on rural highways. *Transportation Research Record Journal of the Transportation Research Board*. 2011; 2255:146–155. DOI: 10.3141/2255-16.

[3] Ministry of Land. *Infrastructure and Transportation. Expressway Design Handbook*, Ministry of Land, Seoul, Korea, 2009.

[4] Al-Kaisy A, et al. Practical guidelines for estimation of rest area use on rural interstates and arterial highways. *Transportation Research Record Journal of the Transportation Research Board*. 2012;2303:117–124. DOI: 10.3141/2303-13.

[5] Kay J, Gates T, Savolainen P, et al. Empirical models of demand levels and turn-in rates at roadside rest areas. *Transportation Research Record Journal of the Transportation Research Board.* 2014;2430:200–206. DOI: 10.3141/2430-21.

[6] Griffin D, Yan X. *Water use, waste generation, and traffic counts at interstate rest areas in Louisiana*. Department of Civil Engineering - Louisiana Tech University, June 2003.

[7] Shen X, Wang L, Liu H, et al. Estimation of the percentage of mainline traffic entering rest area based on BP Neural network. *Journal of Applied Sciences*, 2013;13:2632–2638. DOI: 10.3923/JAS.2013.2632.2638.

[8] Shen X, Zhang F, Lv H, et al. Prediction of entering percentage into expressway service areas based on wavelet neural networks and genetic algorithms. *IEEE Access*, 2019;2019:2913177. DOI: 10.1109/ACCESS.2019.2913177.

[9] Sun C, Lyu H, Yang R, et al. Traffic volume prediction for highway service areas based on XGBoost model with improved particle swarm optimization. *Journal of Beijing Jiaotong University*. 2021;45:74–83.

[10] Zhao J, Yu Z, Yang X, et al. Short term traffic flow prediction of expressway service area based on STL-OMS. *Physica A: Statistical Mechanics and its Applications*. 2022;595:126937. DOI: 10.1016/j.physa.2022.126937.

[11] Guo F, Zou F, Luo S, et al. Positioning method of expressway ETC gantry by multi-source traffic data. *IET Intelligent Transportation Systems*. 2022;00:1–15. DOI: 10.1049/itr2.12280.

[12] Qi W, Bin R, Fu C. A novel filtering method of travel-time outliers extracted from large-scale traffic checkpoint data. *Journal of Transportation Engineering, Part A: Systems*. 2024;150:04023133. DOI: 10.1061/JTEPBS.TEENG-8082.

[13] Liao L, Li Z, Lai S, et al. An expressway traffic congestion measurement under the influence of service areas. *Plos One*. 2023;2023:0279966.1-0279966.22. DOI: 10.1371/journal.pone.0279966.

[14] Cai Q, Yi D, Zou F et al. Recognition of vehicles entering expressway service areas and estimation of dwell time using ETC data. *Entropy*. 2022;24:1208. DOI: 10.3390/e24091208.

[15] Chiabaut N, Faitout R. Traffic congestion and travel time prediction based on historical congestion maps and identification of consensual days. *Transportation Research Part C: Emerging Technologies*. 2021;124:102920. DOI: 10.1016/j.trc.2020.102920.

[16] Liu F, Bi W, Hao W, et al. An improved fuzzy trajectory clustering method for exploring urban travel patterns. *Journal of Advanced Transportation*. 2021;2021:6651718. DOI: 10.1155/2021/6651718.

[17] Kim S, Chung J. Exploration on origin–destination-based travel time variability: Insights from Seoul metropolitan area. *Journal of Transport Geography*. 2018;70:104-113. DOI: 10.1016/j.jtrangeo.2018.05.021.

[18] Smiti A. A critical overview of outlier detection methods. *Computer Science Review*. 2020;38:100306. DOI: 10.1016/j.cosrev.2020.100306.

[19] Zheng K, Yao E, Zhang J, et al. Traffic flow estimation on the expressway network using toll ticket data. *IET Intelligent Transportation Systems*. 2019;13:886-895. DOI: 10.1049/iet-its.2018.5375.

[20] Yang S, Wu Y. Mixture models for fitting freeway travel time distributions and measuring travel time reliability. *Transportation Research Record Journal of the Transportation Research Board*. 2016;2594:95–106. DOI: 0.3141/2594-13.

[21] Dempster A, Laird N, Donald R. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1977;39:1–38. DOI: 10.1111/j.2517-6161.1977.tb01600.x.

[22] Figueiredo M, Jain A. Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2022;24(3):381–396, DOI: 10.1109/34.990138.

[23] Rodriguez A, Laio A. Clustering by fast search and find of density peaks. *Science*. 2014;344(6191):1492-1496. DOI: 10.1126/science.1242072.

[24] Wang Y, Qian J, Hassan M, et al. Density peak clustering algorithms: A review on the decade 2014–2023. *Expert Systems with Applications*. 2024;238:121860. DOI: 10.1016/j.eswa.2023.121860.

[25] Weng Y, Zhang N, Yang X. Improved density peak clustering based on information entropy for ancient character images. *IEEE Access*. 2019;7:81691–81700. DOI: 10.1109/ACCESS.2019.2923694.

郑育彬，成诚（通讯作者），张勇，王凌一，李奇轩，张海林

基于 ETC 门架数据的服务区驶入率估计：一种 ADPC-GMM 方法

摘要

车辆驶入率是高速公路服务区设计和运营中广泛应用的关键输入。随着高速公路 ETC 门架的应用，服务区驶入率可以进一步通过 ETC 门架数据进行估计。本文提出了一种自适应密度峰值聚类高斯混合模型（ADPC-GMM）用于估算高速公路服务区驶入率，其中，ADPC 算法用于估计高速公路服务区所在路段的特征参数，特征参数作为 GMM 模型的输入参数进行驶入率估计。为验证模型精确性，论文获取了中国四川四条高速公路服务区的驶入率数据及 ETC 门架数据。验证结果显示，乘用车驶入率估计结果的 MAE 和 RMSE 分别为 0.0228 和 0.0267，货运车辆驶入率估计结果 MAE 和 RMSE 分别为 0.0264 和 0.0356。与传统 GMM、K-Means 和 DBSCAN 算法相比，模型估计误差最小，表明所提出的方法在估算服务区驶入率方面具有良好的适用性。论文提出的方法具备普适性，对于未布设有服务区驶入车辆监控的服务区将有更大帮助。

关键词

高速公路服务区；驶入率；ETC 门架；自适应密度峰值聚类；混合高斯模型