

RESEARCH ARTICLE

Predicting subway passenger flows under different traffic conditions

Ximan Ling¹, Zhiren Huang¹, Chengcheng Wang¹, Fan Zhang², Pu Wang^{1*}

1 School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan, China, **2** Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China

* wangpu@csu.edu.cn



Abstract

Passenger flow prediction is important for the operation, management, efficiency, and reliability of urban rail transit (subway) system. Here, we employ the large-scale subway smart-card data of Shenzhen, a major city of China, to predict dynamical passenger flows in the subway network. Four classical predictive models: historical average model, multilayer perceptron neural network model, support vector regression model, and gradient boosted regression trees model, were analyzed. Ordinary and anomalous traffic conditions were identified for each subway station by using the density-based spatial clustering of applications with noise (DBSCAN) algorithm. The prediction accuracy of each predictive model was analyzed under ordinary and anomalous traffic conditions to explore the high-performance condition (ordinary traffic condition or anomalous traffic condition) of different predictive models. In addition, we studied how long in advance that passenger flows can be accurately predicted by each predictive model. Our finding highlights the importance of selecting proper models to improve the accuracy of passenger flow prediction, and that inherent patterns of passenger flows are more prominently influencing the accuracy of prediction.

OPEN ACCESS

Citation: Ling X, Huang Z, Wang C, Zhang F, Wang P (2018) Predicting subway passenger flows under different traffic conditions. PLoS ONE 13(8): e0202707. <https://doi.org/10.1371/journal.pone.0202707>

Editor: Xiaolei Ma, Beihang University, CHINA

Received: May 7, 2018

Accepted: August 7, 2018

Published: August 27, 2018

Copyright: © 2018 Ling et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The minimal data set to replicate this study are uploaded as Supporting Information.

Funding: This work was supported by the National Natural Science Foundation of China (<http://www.nsf.gov.cn>) no. 61473320, and by the National Key Research and Development Program of China (<http://program.most.gov.cn>) no. 2016YFB1200401. These fundings were received by PW. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Introduction

Public transportation plays an indispensable role in modern big cities. Developing public transportation is regarded as the most effective way to solve the ubiquitous traffic congestion problems [1, 2]. The subway is regarded as the backbone of urban public transportation, and is characterized by high speed, convenience, and mass flow features [3–7]. Despite the fact that subway services have been continuously improved in many big cities, the upgraded supply usually cannot meet the even faster growing demands of human mobility, especially in developing countries. Compared with opening new lines or increasing the operating frequency of trains, intelligent operation is a smarter and more cost-efficient way to improve the level of service. This calls for accurate and robust prediction of passenger flows to guide better use of the capacity of subway networks. Despite that some passenger flow prediction models have been proposed, we revisited this important problem from two new perspectives.

First, we analyzed the performance of different predictive models under different passenger flow (traffic) conditions. In general, traffic conditions can be classified into ordinary

Competing interests: The authors have declared that no competing interests exist.

conditions, for example morning commutes in a typical weekday, and anomaly conditions, such as bursts of passenger flow in a specific subway station due to a large commercial or recreational event. Moreover, the variance of travel time under the congestion state is remarkably larger than that under the free-flow state [8]. We identified the traffic conditions of each subway station using the density-based spatial clustering of applications with noise (DBSCAN) algorithm, and explored the high-performance passenger flow models under different traffic conditions.

Second, previous works rarely explored how long in advance that passenger flows can be well predicted by each kind of predictive model. Most models were tested by inputting data collected in a time window to predict passenger flows in the next adjacent time window. However, this type of input data setting is hard to implement in practice because collection of smartcard data usually has a delay. Moreover, for some practical applications, such as preventing a large crowd gathering that may cause a dangerous crowding situation, it is important to predict passenger flows a long time before high-density crowding is realized because it is difficult to evacuate high-density crowds both safely and rapidly.

In the following, we make a brief review of existing traffic prediction models, which can be generally classified into three types: (1) mathematical analytical models; (2) traffic simulation models; and (3) knowledge discovery models.

Early traffic prediction models were mostly based on mathematical analytic approaches. Time series models, which include the auto regression (AR) model, moving average (MA) model, auto regressive moving average (ARMA) model, and autoregressive integrated moving average (ARIMA) model, are typical examples. In 1927, Yule developed the AR model to study the periodicities of Wolfer's sunspot numbers [9]. In this AR model, the curve of the time series was fit by the linear combination of the observed historical values. Walker developed the MA model based on the AR model in 1931 [10]. The MA model used a linear combination of historical random disturbances and prediction errors to obtain the current predictive value. In the same year, Walker proposed the ARMA model, which combined the AR model and MA model. In 1970, Box and Jenkins proposed the ARIMA model [11], which incorporated a differencing process (data values were replaced by the differences of current data values and historical data values) in the ARMA model.

Despite the long history of the time series model, it was first used in transportation studies by Ahmed and Cook in 1979 [12]. They employed the ARIMA model to predict traffic flow in freeways; however, the accuracy of prediction was not satisfying. In the 1980s, Stephanedes and Okutani respectively applied the historical average (HA) model and the Kalman filter model to the urban traffic control system of Minneapolis-St and Nagoya City [13, 14]. Recently, Wang *et al.* [15] developed a general approach for real-time freeway traffic state prediction based on stochastic macroscopic traffic flow modeling and extended Kalman filtering, and Li studied the prediction of traffic flow based on interval type-2 fuzzy sets theory [16]. Given that the HA model is prominently influenced by random disturbance, the Kalman filter model was used to adjust the Kalman gain weight every time, resulting in a heavy computing burden. Time series of traffic states sometimes show obvious periodic variation (quarterly, monthly, weekly, etc.), and thus, the seasonal ARIMA (SARIMA) model was developed to capture periodic variations of traffic states by Williams and Hoel in 2003 [17]. They applied the SARIMA model in the prediction of traffic flow in freeways, and found that it outperformed the HA model. Recently, Schimbinschi *et al.* [18] proposed a novel model named topology-regularized universal vector autoregression (TRU-VAR) for traffic flow prediction, which performs better than ARIMA model. In addition, Xue *et al.* [19] proposed a hybrid model combining the time series model with interactive multiple model (IMM) algorithm to predict the short-term bus passenger demand, it is superior to times series model. Ma *et al.* [20] used a

geographically and temporally weighted regression (GTWR) model to identify the spatiotemporal influence of the built environment on transit ridership.

Traffic simulation models were widely used with the popularization of computers in scientific research. In 2001, Chrobok *et al.* [21] presented an approach based on a micro-simulator to predict traffic flow in the freeway network of North Rhine-Westphalia. In 2010, McCrea *et al.* [22] proposed a novel hybrid approach that combines the advantages of the traffic simulation model and linear system theory. In their model, traffic dynamics was first simulated using a continuum mathematical model to obtain relevant traffic parameters of road segments, and the obtained parameters were used as inputs for the Bayesian model for traffic flow prediction. Under the same requirement of prediction accuracy, the hybrid approach improved the computing efficiency compared to the Bayesian network model.

In recent years, knowledge discovery methods have been used more frequently in traffic prediction. Representative methods include nonparametric regression analysis, artificial neural networks, support vector machines, wavelet analysis, and gradient boosting decision tree [23]. In 1991, Davis and Nihan applied nonparametric regression to predict traffic flow in a freeway; however, the accuracy of prediction was lower than that of the linear time-series method [24]. Twelve years later, Clark applied the method of multivariate nonparametric regression to predict the traffic state of a motorway [25]. The method was simple and easy to implement, requiring only modest data storage, and produced reasonably accurate short-term forecasts of traffic flow and loop occupancies (in the percentage of time a loop is covered by a vehicle).

Artificial neural networks were born in the 1940s, and first introduced in traffic flow prediction by Vythoulkas in 1993 [26]. He employed an artificial neural network to predict the traffic state of a city road network. Two years later, Dougherty summarized the application of neural networks in transportation studies [27]. The transportation research community saw an explosion of interest on neural networks in the 1990s. A variety of neural network models have been proposed to predict traffic conditions. Representative examples include the multilayer perceptron neural network model [28], radial basis function neural network [29, 30], spectral basis artificial neural network [31], time delayed neural network [32], and recurrent neural network [33]. Models combining neural networks with other factors (e.g., time series [34], genetic algorithms [35], fuzzy logic rules [36], empirical mode decomposition [37], etc.) were also studied.

Support vector machines were formally published in 1995 [38], and studies on support vector regression (SVR) began in 1997 [39]. Support vector regression was used for travel-time prediction [40, 41]. Wu *et al.* [40] validated the feasibility of applying support vector regression in travel-time prediction, the mean relative errors for traveling different distances were less than 5% in the test dataset. Vanajakshi *et al.* [41] found the support vector regression performs better than artificial neural network when the training data is less or when there are a lot of variations in the training data. Recently, Jiang *et al.* [42] combined the ensemble empirical mode decomposition with gray support vector machine to predict the short-term passenger flow of high-speed rail (HSR), and the mean absolute percentage errors of the hybrid model is about 6%, which performs better than the SVM model and the ARIMA model.

Wavelet analysis, which was developed in the 1980s, is usually used to decompose a set of original traffic flow signals into signals with different time series to reflect and distinguish the internal variation trend and stochastic disturbance of traffic flows. He *et al.* [43] proposed a method based on wavelet decomposition and reconstruction combined with the time series model for traffic volume prediction. And the processed signals with different characteristics can be combined with the dynamic neural network [44], support vector machines [45], and other methods, to predict traffic flow.

In this study, the smartcard data of more than 6 million subway passengers and geographic information data of the Shenzhen subway network were used. We analyzed four classical

predictive models: the historical average (HA) model, multilayer perceptron (MLP) neural network model, support vector regression (SVR) model, and gradient boosted regression trees (GBRT) model. Different from previous studies, we explored the high-performance models under different traffic conditions, and studied how long in advance that passenger flows could be accurately predicted by each predictive model.

The paper is organized as follows. Section II describes the geographic information data and passenger mobility data used in this study. Section III introduces the passenger flow prediction models and algorithm used to classify passenger flow (traffic) conditions. Section IV analyzes and discusses the passenger flow prediction results of different models, and identifies the high-performance models under different traffic conditions and different model implementation conditions (how long passenger flows are predicted in advance). Section V concludes the results, and discusses future research directions.

Materials and methods

Data

The geographic information systems (GIS) data and smartcard data of Shenzhen subway passengers were both provided by Shenzhen Transportation Authority. Data collection was conducted in 2014; the collection of smartcard data was from October 1, 2014 to December 31, 2014. In 2014, the subway network consisted of 118 subway stations. Stations opened after 2014 were not considered due to lack of smartcard data for the new stations. Once a subway passenger employs his/her smartcard when entering or existing a subway station, the time, card ID, and subway station ID are recorded. In the three-month data collection period, a total of 262 million passenger records were generated. For some days, there was data missing for a few hours or the whole day; therefore, only days with complete records were used in this study (80 days in total).

The three-month observation period was split into 7,680 time windows, with each time window spanning 15 min. Taking the operation period of Shenzhen Metro into consideration, the time period of data collection for each day was from 7:00 a.m. to 10:30 p.m.. Therefore, there are only 62 time windows in each day used for training data and testing data. The time windows from 10:30 p.m. to 7:00 a.m. are not considered because few smartcard data are available during the late-night period. We calculated the number of passengers entering a subway station s during each time window t , in-passenger-flow $N_{in}(s,t)$, and the number of passengers exiting a subway station s during each time window t , out-passenger-flow $N_{out}(s,t)$ (Fig 1A and 1B). Heterogeneous distribution of passenger flows is observed in the studied subway network (Fig 2A and 2B). The in-passenger-flow can be approximated by two different fitting functions for large and small $N_{in}(s,t)$ (gray dashed lines are plotted to guide the eyes):

$$\text{fit1: } P(N_{in}(s,t)) = 0.017 (N_{in}(s,t))^{-0.304} \text{ when } N_{in}(s,t) \leq 150 \text{ persons;}$$

$$\text{fit2: } P(N_{in}(s,t)) = 0.009 \exp(-0.006 N_{in}(s,t)) \text{ when } N_{in}(s,t) > 150 \text{ persons.}$$

The out-passenger-flow also can be approximated by two different fitting functions for large and small $N_{out}(s,t)$ (gray dashed lines are plotted to guide the eyes):

$$\text{fit3: } P(N_{out}(s,t)) = 0.017 (N_{out}(s,t))^{-0.384} \text{ when } N_{out}(s,t) \leq 150 \text{ persons;}$$

$$\text{fit4: } P(N_{out}(s,t)) = 0.005 \exp(-0.004 N_{out}(s,t)) \text{ when } N_{out}(s,t) > 150 \text{ persons.}$$

Roughly 58.47% of in-passenger-flow $N_{in}(s,t)$ and 50% of out-passenger-flows $N_{out}(s,t)$ were smaller than 200 passengers/15 min; for some stations, passenger flows were larger than 1,000 passengers/15 min. In the following sections, measured in-passenger-flow $N_{in}(s,t)$ and out-

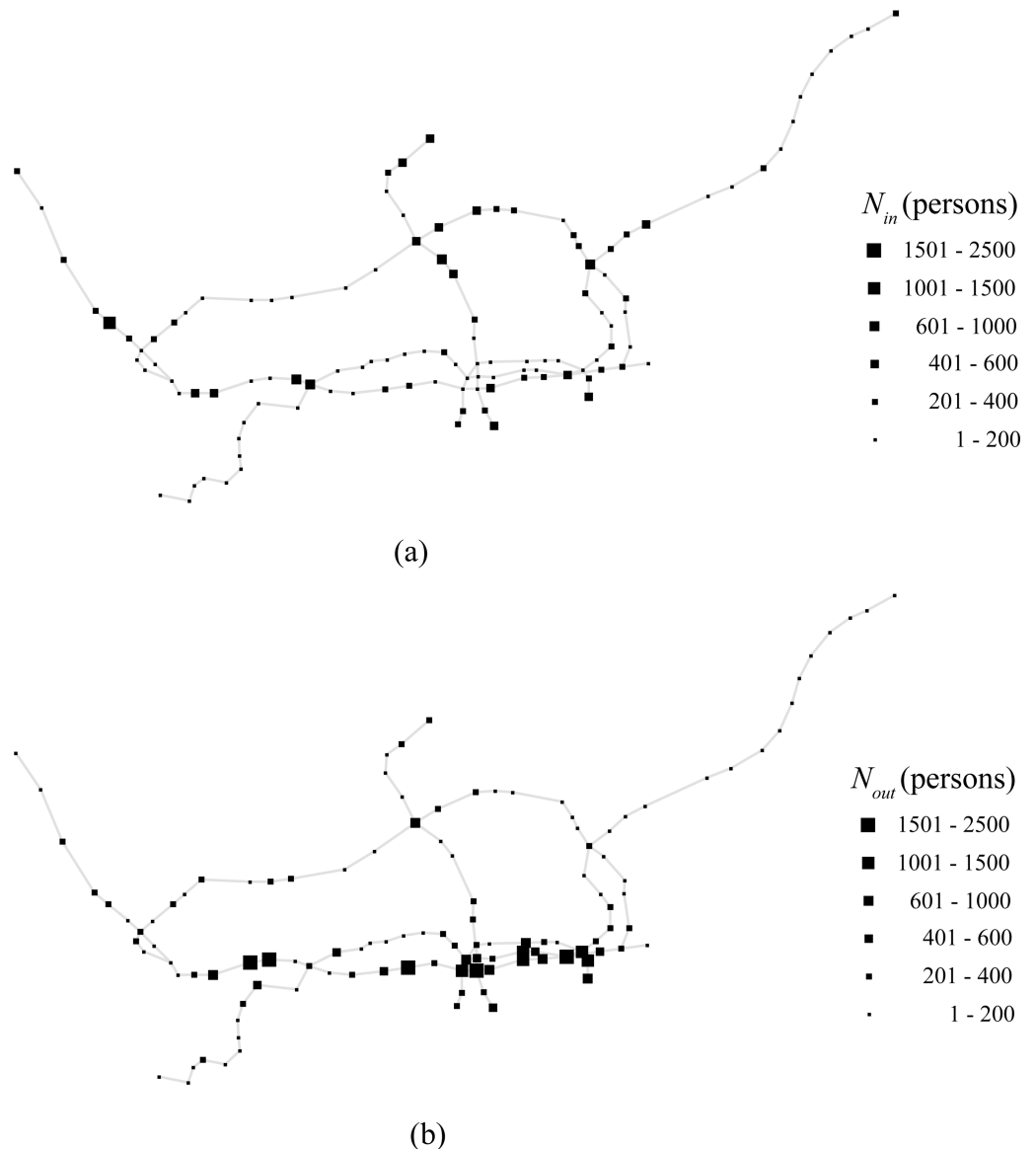


Fig 1. (a), (b) In-passenger-flow $N_{in}(s,t)$ and out-passenger-flow $N_{out}(s,t)$ of each subway station s during the time window 9:00 a.m.–9:15 a.m. of a typical weekday in 2014.

<https://doi.org/10.1371/journal.pone.0202707.g001>

passenger-flow $N_{out}(s,t)$ were used as the ground truth data to train the passenger flow prediction models and validate the predictive results.

The used subway smartcard data were split into two parts. The first part of the data, which recorded the subway passenger trips generated during October and November of 2014, were used as the training dataset. The second part of the data, which recorded the subway passenger trips generated during December of 2014, were used as the testing dataset. Training datasets were denoted by $D = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_n, \mathbf{y}_n)\}$, where $\mathbf{x}_n \in R^d$ represent the input features of the training data, and $\mathbf{y}_n \in R^l$ represent the output results of the training data. The sample size n equals 59 because there were 59 days' smartcard data in the training dataset. Data dimensions d and l represent the number of input and output features used in the models respectively.

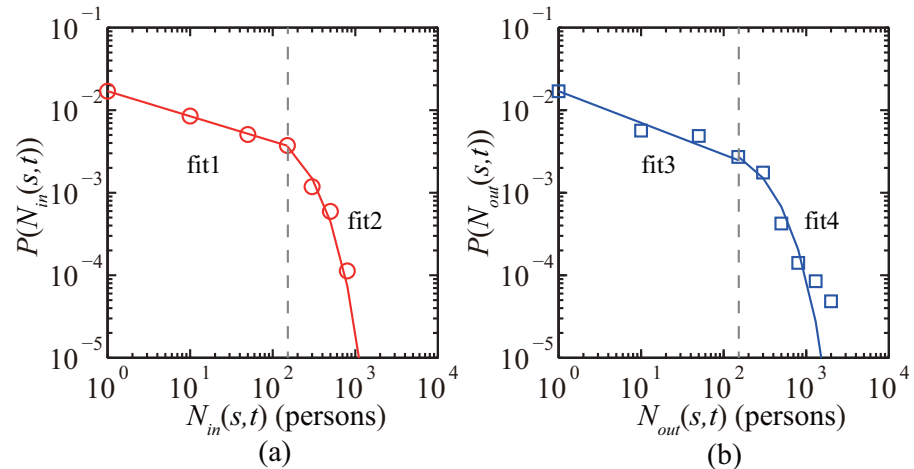


Fig 2. Distributions of passenger-flow show heterogenous patterns in different subway stations. (a) is for in-passenger-flow $N_{in}(s,t)$, (b) is for out-passenger-flow $N_{out}(s,t)$.

<https://doi.org/10.1371/journal.pone.0202707.g002>

The prediction models

When predicting the passenger flow of a subway station s during a time window t_{target} , subway station s is called target station, and time window t_{target} is called target time window. We evaluated the performances of four predictive models under different model implementation conditions; predictions were made in different number (n_{step}) of time windows before t_{target} , and $n_{step} = 1, 2, \dots, 7, 8$ were tested. Here, we briefly introduce the advantageous and disadvantageous features of each of the four predictive models used in this study. The HA model is easy to implement in practice, but performs poorly under unexpected traffic conditions. The multi-layer perceptron (MLP) neural network employed in study is trained using back-propagation. In general, the MLP model works well in capturing complex and nonlinear relations; however, it usually requires a large volume of data and complex training procedures. For the employed SVR model, a linear kernel function was used to predicts passenger flows; however, the selection of best kernel functions is an unsolved problem in this scientific community. Lastly, the GBRT model uses a negative gradient of loss function as an estimate of residuals. In general, the GBRT model also works well in exploring complex and nonlinear relations; however, it cannot train data parallelly.

In the generated HA model, the average in-passenger-flow (or average out-passenger-flow) during the target time window t_{target} of all days in the training dataset were used as the predictive result in the target time window for all days in the testing dataset. Clearly, the HA model was unable to capture the random disturbances of passenger flows, and therefore had the worst prediction accuracy and served as a baseline model for comparison with the other three models. For the MLP model, SVR model, and GBRT model, in-passenger-flows $N_{in}(s,t)$ during time window t of all days in the training dataset were used as inputs, and in-passenger-flows $N_{in}(s,t_{target})$ during the target time window of all days in the training dataset were used as outputs to train the predictive model; t is n_{step} time windows before the target time window t_{target} . In a given day of the testing dataset, the in-passenger-flows $N_{in}(s,t)$ were used as inputs to predict $N_{in}(s,t_{target})$, where t is n_{step} time windows before the target time window t_{target} . Parameter n_{step} determines how long in advance predictions are conducted. Similarly, models were generated to predict $N_{out}(s,t_{target})$. Methods for generating the MLP model, SVR model, and GBRT model are briefly described in the following subsections. Please refer to the literature [46–49] for further details on the generations of these models.

The training dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, $x_n \in R^d, y_n \in R^l$ were used in the MLP model, SVR model, and GBRT model. Parameters d and l represent the dimensions of x and y , respectively. In this paper, parameters $d = 1, l = 1$ are selected because only passenger flows of a station itself are used as the model inputs to predict the passenger flows of the station. Parameter n represents the sample size of D (i.e., 59 days' smartcard data in the training dataset).

Taking the prediction of out-passenger-flows $N_{out}(s, t_{target})$ at the subway station "Window of World" during 9:00 a.m.–9:15 a.m. of December 30 as an example, s denotes the "Window of World" subway station, and t_{target} denotes the target time window 9:00 a.m.–9:15 a.m. When predicting passenger flows at one time window ahead of the target time window ($n_{step} = 1$), historical passenger flows at the station s during the time window $t_{target} - 1$ of all days in the training dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ are used. For the proposed example, x_n represents the out-passenger-flows at the subway station "Window of World" during time window 8:45 a.m.–9:00 a.m. of the n th day in the training dataset, and y_n represents the out-passenger-flows of the "Window of World" station during time window 9:00 a.m.–9:15 a.m. of the n th day in the training dataset.

Multilayer perceptron neural network model. The multilayer perceptron is a forward structure artificial neural network that maps a set of input vectors to a set of output vectors. An MLP consists of multiple layers, including an input layer, one or more hidden layers, and an output layer. Each layer of neurons is interconnected with the next layer of neurons. There is no connection between neurons in the same layer, and there is no cross-layer connection. For both the hidden layer and output layer, neurons have activation functions, whereas on the input layer, neurons only receive the input dataset and do not have activation functions. The learning process in neural networks involves adjusting the connection weights between neurons and the threshold of each functional neuron.

We considered a three-layer MLP network consisting of d input neurons, a hidden layer with q hidden neurons, and an output layer with l output neurons. The threshold of the j th neuron in the output layer is defined as θ_j , and the threshold of the h th neuron in the hidden layer is defined as γ_h . Connection weight v_{ih} represents the weight between the i th neuron in the input layer and the h th neuron in the hidden layer, whereas connection weight w_{hj} represents the weight between the h th neuron in the hidden layer and the j th neuron in the output layer. Therefore, each hidden neuron h firstly computes the net input $\alpha_h = \sum_{i=1}^d v_{ih}x_i$ and generates an output b_h . Each output neuron j uses the outputs of the hidden layer as inputs $\beta_j = \sum_{h=1}^q w_{hj}b_h$.

For a single training sample (x_k, y_k) , $\hat{y}_k = (\hat{y}_1^k, \hat{y}_2^k, \dots, \hat{y}_l^k)$ is the output of the neural network, that is $\hat{y}_j^k = f(\beta_j - \theta_j)$, where $f(\cdot)$ is the activation function, and the rectified linear unit function $f(x) = \max(0, x)$ is used here as the activation function. Therefore, the mean square error of the network is

$$E_k = \frac{1}{2} \sum_{j=1}^l (\hat{y}_j^k - y_j^k)^2. \tag{1}$$

The update of any parameter v is defined as $v \leftarrow v + \Delta v$. The training process of the MLP with backpropagation is as follows.

Step 1: Input the training dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, $x_n \in R^d, y_n \in R^l$ and determine the activation function. In this paper, the number of hidden neurons q was set to 100, the tolerance for stopping criterion is set to 0.0001 (i.e. value of (1) is smaller than 0.0001), and the maximum number of iterations is 200.

Step 2: All connection weights and thresholds in the neural network are initialized randomly in the range of (0, 1).

Step 3: For $(\mathbf{x}_k, \mathbf{y}_k)$, according to current parameters and function $\hat{y}_j^k = f(\beta_j - \theta_j)$, calculate the value of $\hat{\mathbf{y}}_k$. The mean square error of the network is computed as $E_k = \frac{1}{2} \sum_{j=1}^l (\hat{y}_j^k - y_j^k)^2$.

Step 4: Update the connection weights w_{hj} and v_{ih} and the thresholds θ_j and γ_h .

$$w_{hj} \leftarrow w_{hj} + \Delta w_{hj}, \tag{2}$$

$$v_{ih} \leftarrow v_{ih} + \Delta v_{ih}, \tag{3}$$

$$\theta_j \leftarrow \theta_j + \Delta \theta_j, \tag{4}$$

$$\gamma_h \leftarrow \gamma_h + \Delta \gamma_h. \tag{5}$$

The error backpropagation (BP) algorithm based on the gradient descent strategy adjusts the parameters [46, 47].

Step 5: Repeat Steps 1–4 until the value of (1) satisfies the predefined tolerance for stopping criterion.

Support vector regression model. The kernel function Φ is used to map data into a high-dimensional feature space, such that the nonlinear fitting problem in the input space is transformed into a linear fitting problem in the high-dimensional feature space. Common kernel functions include linear kernel, polynomial kernel, gaussian kernel, Laplace kernel, and sigmoid kernel, where the nonlinear mapping function is $k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \cdot \Phi(\mathbf{x}_j)$. The goal of the support vector regression model is to find the partition hyperplane with the maximum margin. The partition hyperplane is represented by $f(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}) + b$, where $\mathbf{w} \in R^d$ is the normal vector, and b is the displacement.

Suppose ε is the error bound between observation value y and predicted value $f(\mathbf{x})$. With $f(\mathbf{x})$ as the center, the epsilon-tube with a width of 2ε is established, and then the problem is formalized [46] as

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \ell_\varepsilon(f(\mathbf{x}_i) - y_i), \tag{6}$$

where C is the penalty coefficient, and

$$\ell_\varepsilon(z) = \begin{cases} 0, & \text{if } |z| \leq \varepsilon; \\ |z| - \varepsilon, & \text{otherwise.} \end{cases} \tag{7}$$

In summary, the support vector regression model can be described as follows.

Step 1: Input training dataset $D = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_n, \mathbf{y}_n)\}, \mathbf{x}_n \in R^d, \mathbf{y}_n \in R^l$ and select a kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$. In this paper, the linear kernel was chosen as the kernel function, the parameter C was set to 1, ε was set to 0.1, and the tolerance for stopping criterion is set to 0.001 (i.e. the value of (6) is smaller than 0.001).

Step 2: Search for the best solution $\bar{\alpha}^{(*)} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \dots, \bar{\alpha}_l, \bar{\alpha}_l^*)$ to

$$\min_{\alpha, \alpha^*} \frac{1}{2} \sum_{i,j=1}^n (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)k(\mathbf{x}_i, \mathbf{x}_j) + \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \mathbf{y}_i \sum_{i=1}^n (\alpha_i^* - \alpha_i) \quad (8)$$

$$\text{s.t.} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0$$

$$0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, 2, \dots, n.$$

Step 3: Calculate parameter b . Select the positive subvector of $\bar{\alpha}$ ($\bar{\alpha}_j > 0$), or the positive subvector of $\bar{\alpha}^*$ ($\bar{\alpha}_j^* > 0$), and calculate parameter

$$b = \mathbf{y}_j - \sum_{i=1}^n (\bar{\alpha}_i^* - \bar{\alpha}_i)k(\mathbf{x}_i, \mathbf{x}_j) + \varepsilon, \quad 0 < \bar{\alpha}_j < C,$$

$$b = \mathbf{y}_j - \sum_{i=1}^n (\bar{\alpha}_i^* - \bar{\alpha}_i)k(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon, \quad 0 < \bar{\alpha}_j^* < C. \quad (9)$$

Step 4: Obtain the model

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i^* - \alpha_i)k(\mathbf{x}, \mathbf{x}_i) + b. \quad (10)$$

Gradient boosted regression trees model (GBRT)

The gradient boosted regression trees model (GBRT) is described as follows.

Step 1: Input the training dataset $D = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_n, \mathbf{y}_n)\}, \mathbf{x}_n \in R^d, \mathbf{y}_n \in R^l$ and initialization function

$$f_0(\mathbf{x}) = \arg \min_c \sum_{i=1}^n L(\mathbf{y}_i, c). \quad (11)$$

The loss function is $L(\mathbf{y}, f(\mathbf{x})) = \frac{1}{2}(\mathbf{y} - f(\mathbf{x}))^2$, where the constant value c minimizes the value of $\sum_{i=1}^n L(\mathbf{y}_i, c)$, namely, c is as close as possible to \mathbf{y}_i . Here, $f_0(\mathbf{x})$ is a tree with only one node.

Step 2: The training dataset is used as input to iteratively build M trees, M was set to 100 in this paper.

(a) For the m th tree, $m = 1, \dots, M$, calculate the negative gradient of the loss function in the current model

$$r_{mi} = -\left[\frac{\partial L(\mathbf{y}_i, f(\mathbf{x}_i))}{\partial f(\mathbf{x}_i)}\right]_{f(\mathbf{x})=f_{m-1}(\mathbf{x})}. \quad (12)$$

Then, use r_{mi} as an estimate of residuals, where $i = 1, \dots, n$, ∂ stands for the derivative, and n is sample size.

(b) Fit a regression tree for r_{mi} to obtain the leaf node regions R_{mj} of tree m , where $j = 1, 2, \dots, J$, and J is the number of leaf nodes, which is not limited in the present study.

(c) For the leaf node region R_{mj} , where $j = 1, 2, \dots, J$, calculating the best fitting value c_{mj} to minimize the loss function $L(\mathbf{y}, f(\mathbf{x}))$.

$$c_{mj} = \arg \min_c \sum_{\mathbf{x}_i \in R_{mj}} L(\mathbf{y}_i, f_{m-1}(\mathbf{x}_i) + c). \quad (13)$$

Then, update

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \sum_{j=1}^J c_{mj} I(\mathbf{x} \in R_{mj}). \quad (14)$$

Step 3: The final prediction model is

$$\tilde{f}(\mathbf{x}) = f_M(\mathbf{x}) = \sum_{m=1}^M \sum_{j=1}^J c_{mj} I(\mathbf{x} \in R_{mj}). \tag{15}$$

Detecting anomalous passenger flow condition

The DBSCAN algorithm was used to identify anomalous passenger flows. We normalized in- or out-passenger-flow of a subway station s during each time window t of a day $N(s,t)$ with the minimum and maximum values of $N(s,t)$ observed in the same time window during the whole data collection period, and take it as the original data set S . In the DBSCAN algorithm, the maximum radius of neighborhood ϵ defines the eps-neighborhood of a data point $i \in S$, denoted by $N_\epsilon(i) = \{j \in S | dist(i,j) \leq \epsilon\}$, and MinPts determines the minimum number of data points within the eps-neighborhood. The Euclidean distance $dist(i,j) = |N(s,t)^j - N(s,t)^i|$ was used to locate the ϵ neighborhood of each data point i , and the typical parameter setting of MinPts = 4 was used. The maximum radius of neighborhood ϵ was set using the fourth distance (4-dist) probability [50]: the distance between a data point and its fourth nearest neighbor is denoted as the 4-dist. The probability distribution of 4-dist was fitted by an exponential function, and the 4-dist value at which the slope of the fitting curve equaled -1 was used as the parameter setting of ϵ .

Passenger flows were classified using the DBSCAN algorithm: passenger flows larger than the maximum flow f_ϵ of the largest cluster were classified into the anomalous passenger flow (traffic) condition. Passenger flows smaller than or equal to f_ϵ were classified into the ordinary traffic condition. We use the out-passenger-flows $N_{out}(s,t)$ at the subway station “Window of World” during the time window 7:00 p.m.–7:15 p.m. as an example (Fig 3B). Here, s denotes the subway station “Window of World”, the target time window of the prediction is $t = 76$. The label of each cluster generated by the DBSCAN algorithm is denoted by the $label(r)$, where $1 \leq r \leq n_c$, n_c is the total number of clusters. During time window t of the i th day, the out-passenger-flows at the studied station s is denoted as $label(N_{out}(s,t)^i)$. When the label $label(N_{out}(s,t)^i)$ is the same with the label of the largest cluster generated $label(r)_{max}$, the threshold passenger flow f_ϵ is determined $f_\epsilon = \max_{label(N_{out}(s,t)^i)=label(r)_{max}} (N_{out}(s,t)^i)$.

In Fig 3A out-passenger-flows $N_{out}(s,t)$ at “Window of World” station of the Shenzhen subway system are illustrated for every 15 min time windows. Using the DBSCAN algorithm, the threshold passenger flow f_ϵ for each 15 min time window was determined. Anomalous growth of passenger flows was observed on December 31, which was caused by the firework show at the plaza of the recreational park at “Window of World” [51]. For all subway stations, anomalous in-passenger-flows were found in 12.2% of time windows, whereas out-passenger-flow were found in 10.3% of time windows.

Results

Predicting dynamical passenger flows

Previous passenger flow prediction models have been seldom analyzed under anomalous traffic conditions, such as abrupt bursts of passenger flows in a particular subway station due to mass commercial or recreational events. Under anomalous traffic conditions, passenger demands may exceed the maximum capability that a subway station can provide; emergent managements are therefore required to protect the safety and order of subway transportation. In addition, under large crowd gatherings, subway service restrictions can be an important way to prevent passengers from flowing into the crowded area, hence avoiding dangerous crowding situations [52]. Therefore, predicting passenger flows under anomalous traffic conditions is even more important than predicting flow under ordinary conditions.

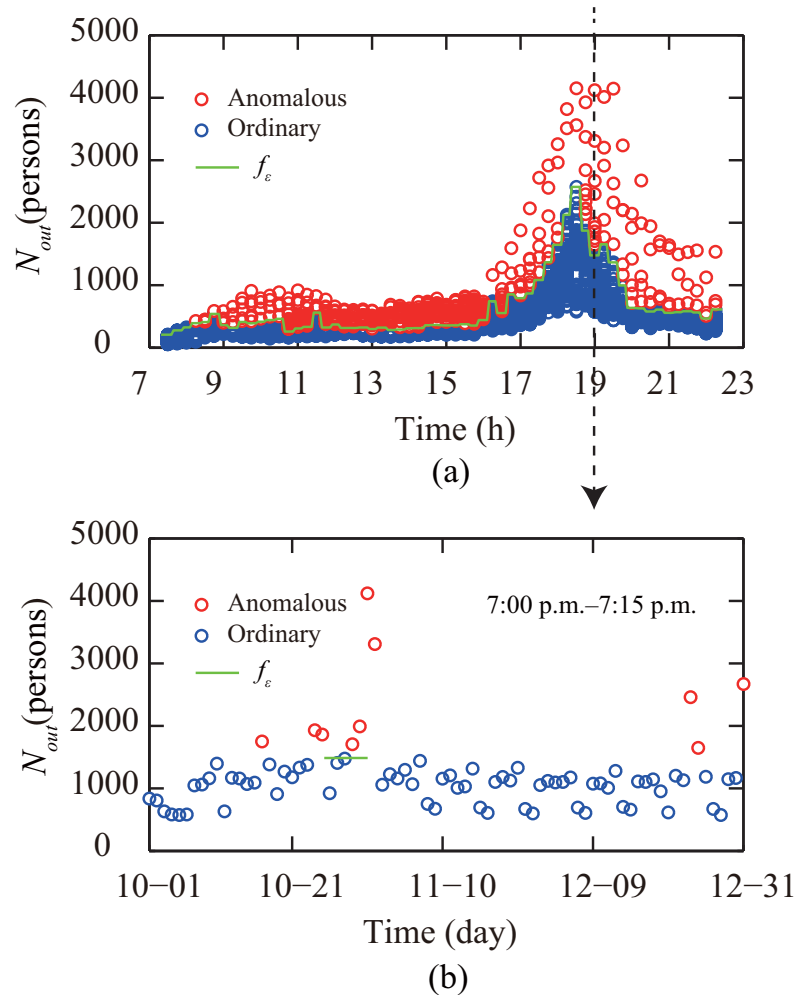


Fig 3. Anomalous passenger flow detection for “Window of World” subway station. (a) Out-passenger-flows of the “Window of World” subway station over the whole observation period. Anomalous flows and ordinary flows are discriminated by the green line f_ϵ . (b) Red circles represent anomalous out-passenger-flow of the station during the time window 7:00 p.m.–7:15 p.m., while blue circles represent passenger flows under ordinary traffic conditions.

<https://doi.org/10.1371/journal.pone.0202707.g003>

Three typical indexes, mean absolute percentage error (MAPE), variance of absolute percentage error (VAPE), and root mean square error (RMSE) were used to evaluate the accuracy of prediction:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%, \quad (16)$$

$$VAPE = Var \left(\frac{|y_i - \hat{y}_i|}{y_i} \right) \times 100\%, \quad (17)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (18)$$

where $y = \{y_1, y_2, \dots, y_n\}$ is the sequence of the observation values, $\hat{y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_1, \dots, \hat{y}_n\}$ is the sequence of the prediction values, and n is the number of observation values.

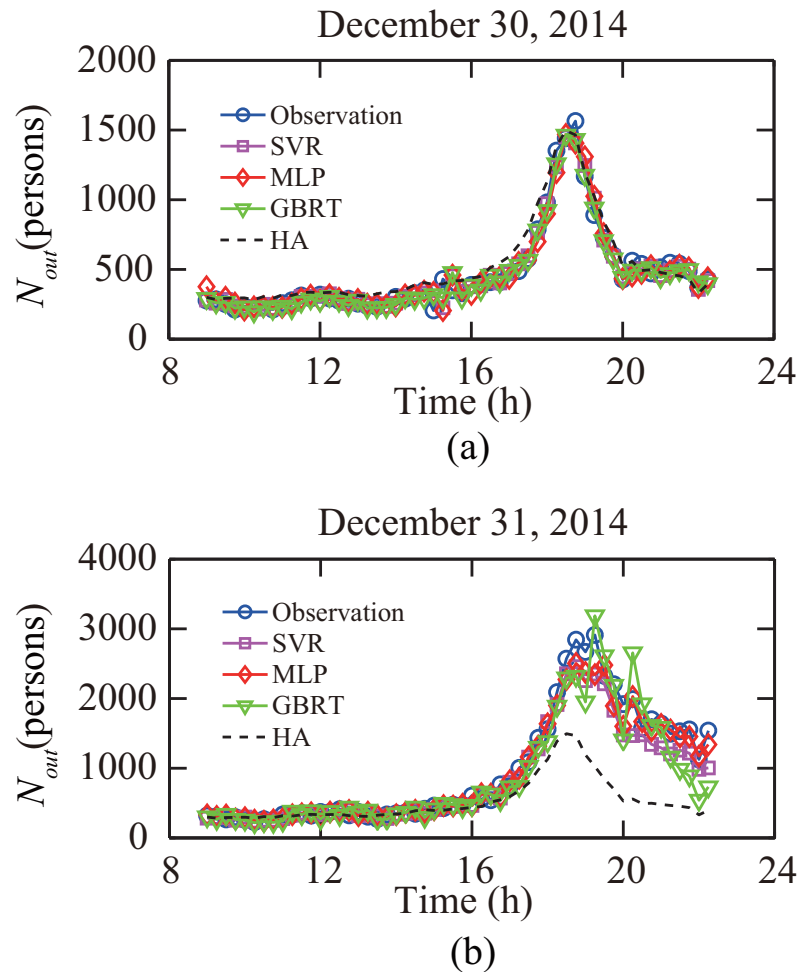


Fig 4. Predictive results versus ground-truth data for subway station “Window of World” when $n_{step} = 1$. (a) Results for a typical weekday. (b) Results for a day when mass events occurred near the station.

<https://doi.org/10.1371/journal.pone.0202707.g004>

Fig 4 shows the prediction results and the ground-truth passenger flow observation at the subway station “Window of World” during a typical weekday (December 30, 2014) and a day when mass events occurred at the plaza near the subway station (December 31, 2014). The performances of four predictive models were similar in the ordinary traffic condition, and all models offered accurate prediction results. However, under the anomalous traffic condition, the HA model failed to capture the trend of abrupt growth of passenger flow as expected, and the prediction of the GBRT model had large fluctuations. Meanwhile, the SVR model and MLP model had a relatively good performance. The results shown in Fig 4 highlight the importance of selecting the proper model under different traffic conditions.

Table 1 shows the RMSE, MAPE, and VAPE values of predictive results of passenger flows at the “Window of World” station based on the SVR, MLP, GBRT and HA model. The prediction time is $n_{step} = 1$ time window ahead of the target time window.

High-performance regions of different predictive models

We analyzed the performances of four predictive models under different numbers of time windows n_{step} that a prediction is made before the target time window n_{target} . When a larger n_{step}

Table 1. The average error of the four models when $n_{step} = 1$.

Method	December 30			December 31		
	RMSE	MAPE (%)	VAPE (%)	RMSE	MAPE (%)	VAPE (%)
SVR	56.22	8.83	1.30	209.10	12.98	0.60
MLP	68.91	11.50	1.39	145.58	10.33	0.44
GBRT	50.37	9.22	0.95	274.75	17.11	1.72
HA	86.72	19.20	2.23	718.70	29.12	6.72

<https://doi.org/10.1371/journal.pone.0202707.t001>

was set, the passenger flow prediction results could be obtained early; meanwhile, the accuracy of prediction decreased as more recent data were not used. Here, we explored the high-performance regions of different predictive models under ordinary and anomalous traffic conditions. Fig 5 shows the predicted passenger flows at the subway station “Window of World” on December 30, 2014. We found that under the ordinary condition, except for the MLP model, the predictive models performed well even when the passenger flow prediction was conducted 2 h before the target time window. The prediction accuracy of the MLP model began to decrease when n_{step} was larger than two time windows, indicating that under the ordinary traffic condition the MLP model only worked well for short-term (less than 30 min) prediction.

Fig 6 shows the predicted passenger flows of the subway station “Window of World” on December 31, 2014. In contrast to the results under the ordinary traffic condition, we found that under anomalous traffic condition, the MLP model performed the best. Given that the HA model is insensitive to the prediction time, the same predictive results were obtained for different numbers of time windows n_{step} that a prediction is made before the target time window, and the HA model could not capture the anomalous traffic condition at all. For all predictive models, the prediction accuracy was not acceptable when the target time window was four time windows (1 h) later than the prediction time. The predicted results of all models had a trend to approach historical average values when $n_{step} \geq 4$.

Table 2 shows the RMSE, MAPE, and VAPE values of predictive results of passenger flows at the “Window of World” station based on the SVR, MLP, GBRT and HA model. The prediction was made 1 to 8 time windows ahead of the target time window, respectively.

We summarized the performance of the four predictive models in Fig 7. Under the ordinary traffic condition, the prediction errors of the SVR model, MLP model, and GBRT model all increased with the increase of the number of time windows n_{step} between the prediction time and target time windows. In particular, the RMSE and MAPE values of the prediction results of the MLP model increased much faster than for the SVR and GBRT models. When prediction was made $n_{step} > 5$ time windows earlier than the target time window, the MLP model had even worse performance than the HA model. The minimum MAPE = 16.9% was generated by the SVR model when $n_{step} = 2$, implying that the most recent data may be not the best data input. Furthermore, the GBRT model had a larger VAPE value than the MLP and SVR models. All the results taken together, the SVR model performed best in the ordinary traffic condition.

When a larger n_{step} was set (the prediction time (t) is earlier than the target time window t_{target}), prediction errors of the SVR model, MLP model, and GBRT model increased faster under the anomalous traffic condition than under the ordinary traffic condition. The RMSE value and MAPE value of the prediction results of the MLP model, SVR model, and GBRT model were similar, but for most n_{step} settings the MLP model was slightly better. The minimum MAPE = 18.0% was generated by the MLP model when $n_{step} = 2$, also implying that the most recent data may be not the best data input. The GBRT model had a larger VAPE value

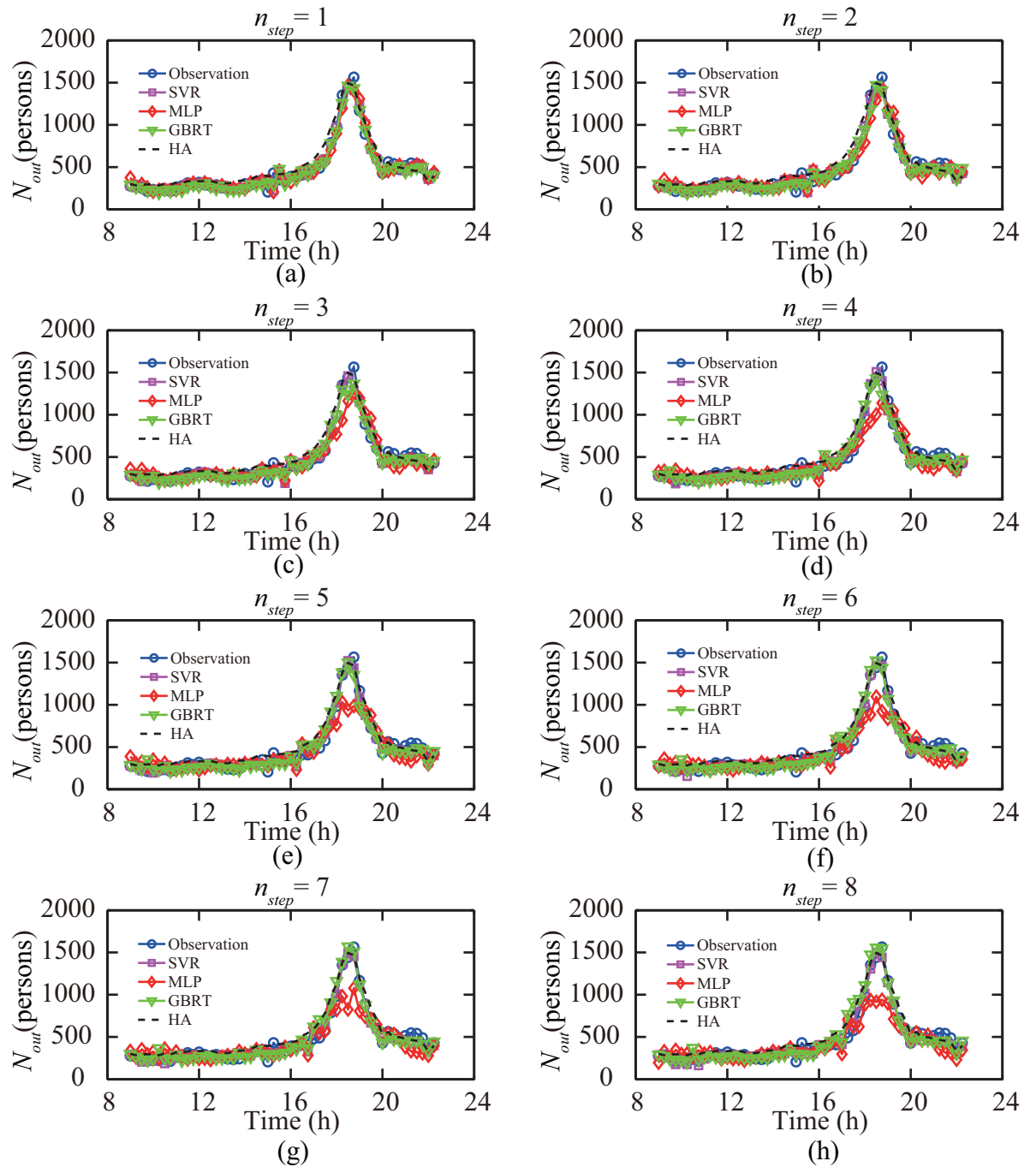


Fig 5. Performance of four models when predicting out-passenger-flow records in subway station “Window of World” on December 30, 2014. (a-h) When the prediction was made 1 to 8 time windows ahead of the target time window, respectively.

<https://doi.org/10.1371/journal.pone.0202707.g005>

when n_{step} was small, which increased slowly with increasing n_{step} ; meanwhile, the VAPE value for the prediction of the MLP model was small when $n_{step} \leq 2$, but had faster growth afterward. Ultimately, the MLP model performed best in the anomalous traffic condition. Table 3 and Table 4 shows the RMSE, MAPE, and VAPE values of the four models under different n_{step} settings for in-passenger-flow and out-passenger-flow predictions.

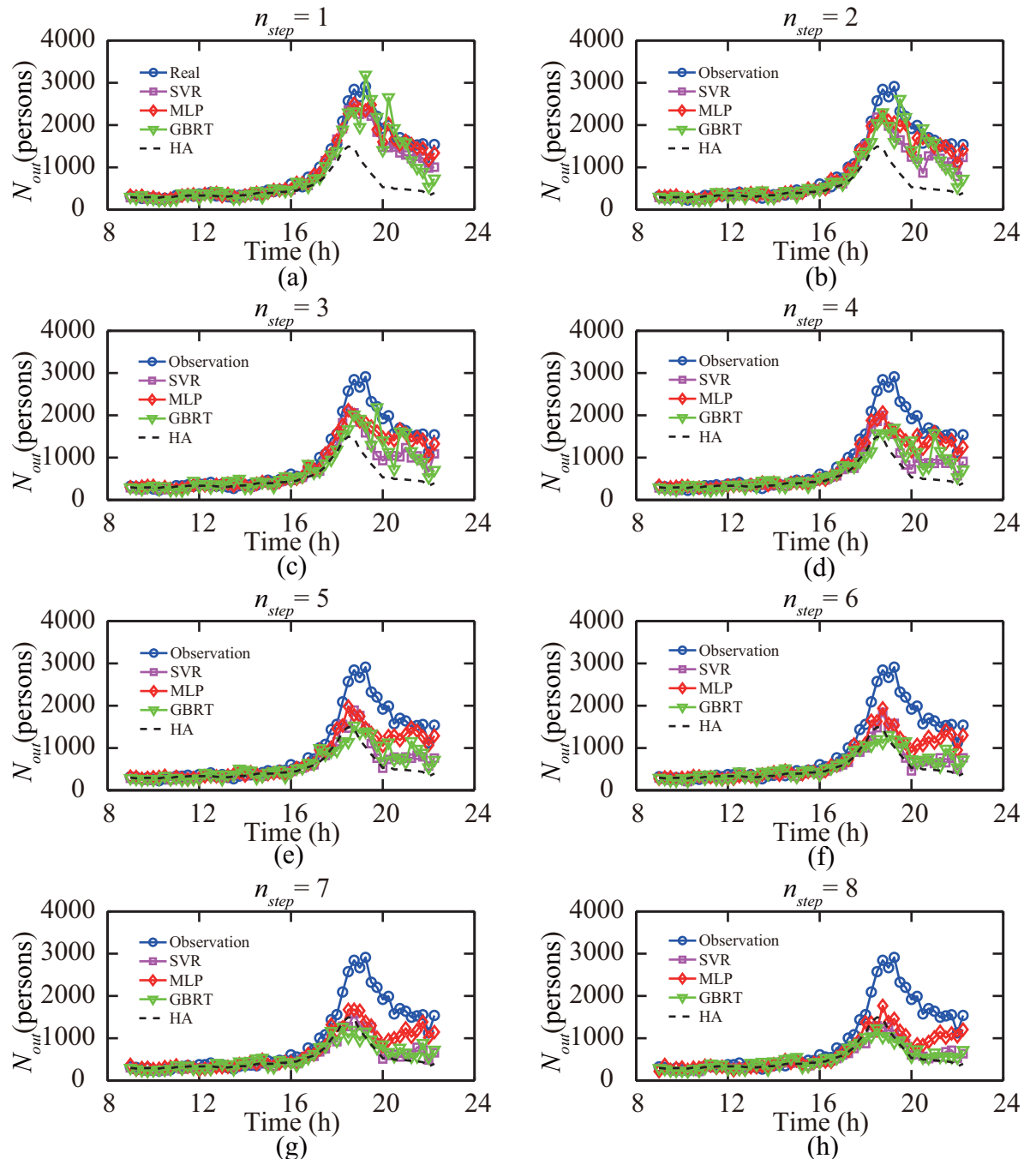


Fig 6. Performance of four models when predicting out-passenger-flow records in subway station “Window of World” on December 31, 2014. (a-h) When the prediction was made 1 to 8 time windows ahead of the target time window, respectively.

<https://doi.org/10.1371/journal.pone.0202707.g006>

Given that all validations were made for all subway stations (118 in total) of the Shenzhen subway network, the average MAPE and RMSE values were enhanced by the majority of low-passenger-flow stations. If we concentrated on subway stations with the top 25% average passenger flows, the minimum MAPE = 11.1% was generated by the SVR model (and GBRT model) when $n_{step} = 2$ for the ordinary traffic condition; meanwhile, the minimum

Table 2. The average error of the four models.

RMSE								
n_{step}	December 30				December 31			
	SVR	MLP	GBRT	HA	SVR	MLP	GBRT	HA
1	56.22	68.91	50.37	86.72	209.10	145.58	274.75	718.70
2	56.68	88.19	58.99	86.72	335.46	216.29	346.33	718.70
3	60.22	110.36	61.43	86.72	434.83	285.31	413.02	718.70
4	57.31	131.77	74.33	86.72	510.15	346.55	507.99	718.70
5	54.90	144.90	71.67	86.72	576.58	403.94	588.75	718.70
6	53.65	151.60	68.89	86.72	612.60	449.57	649.52	718.70
7	55.40	154.21	72.24	86.72	649.76	490.85	689.43	718.70
8	56.16	159.15	85.71	86.72	691.83	536.35	726.10	718.70
MAPE(%)								
n_{step}	December 30				December 31			
	SVR	MLP	GBRT	HA	SVR	MLP	GBRT	HA
1	8.83	11.50	9.22	19.20	12.98	10.33	17.11	29.12
2	10.09	13.50	10.22	19.20	16.80	11.73	19.59	29.12
3	9.98	15.85	10.26	19.20	20.30	14.08	23.90	29.12
4	11.08	18.48	11.96	19.20	23.76	16.53	25.77	29.12
5	10.91	19.14	12.17	19.20	27.06	18.56	28.51	29.12
6	10.72	19.83	12.37	19.20	28.70	21.94	30.08	29.12
7	11.67	20.06	13.08	19.20	28.90	22.80	31.30	29.12
8	11.17	21.03	12.81	19.20	30.15	24.18	33.66	29.12
VAPE(%)								
n_{step}	December 30				December 31			
	SVR	MLP	GBRT	HA	SVR	MLP	GBRT	HA
1	1.30	1.39	0.95	2.23	0.60	0.44	1.72	6.72
2	1.42	1.54	0.90	2.23	1.41	0.91	2.06	6.72
3	1.02	1.29	0.81	2.23	1.91	0.94	2.62	6.72
4	0.97	1.87	0.91	2.23	2.93	1.38	3.21	6.72
5	0.82	1.47	1.30	2.23	3.43	1.75	3.20	6.72
6	0.91	1.87	1.23	2.23	3.93	1.83	4.15	6.72
7	0.89	2.07	1.51	2.23	4.38	1.96	4.42	6.72
8	0.96	2.34	1.95	2.23	5.05	2.69	4.74	6.72

<https://doi.org/10.1371/journal.pone.0202707.t002>

MAPE = 12.3% was generated by the MLP model when $n_{step} = 2$ for the anomalous traffic condition. This result indicates that the inherent pattern of passenger flows at a subway station prominently determines the prediction accuracy. In general, passenger flows of large-flow stations are more predictable than passenger flows of low-flow stations. In addition, for a specific group of subway stations, the best model may be different from the model obtained for all subway stations. In practice, more detailed model selection strategies can be implemented to different subgroups of subway stations. Table 5 and Table 6 describe details of RMSE, MAPE and VAPE values of the four models when $n_{step} = 2$.

Conclusions

An effective and reliable passenger flow prediction model can be beneficial to the management of transportation systems, such as operation planning, revenue planning, and facility

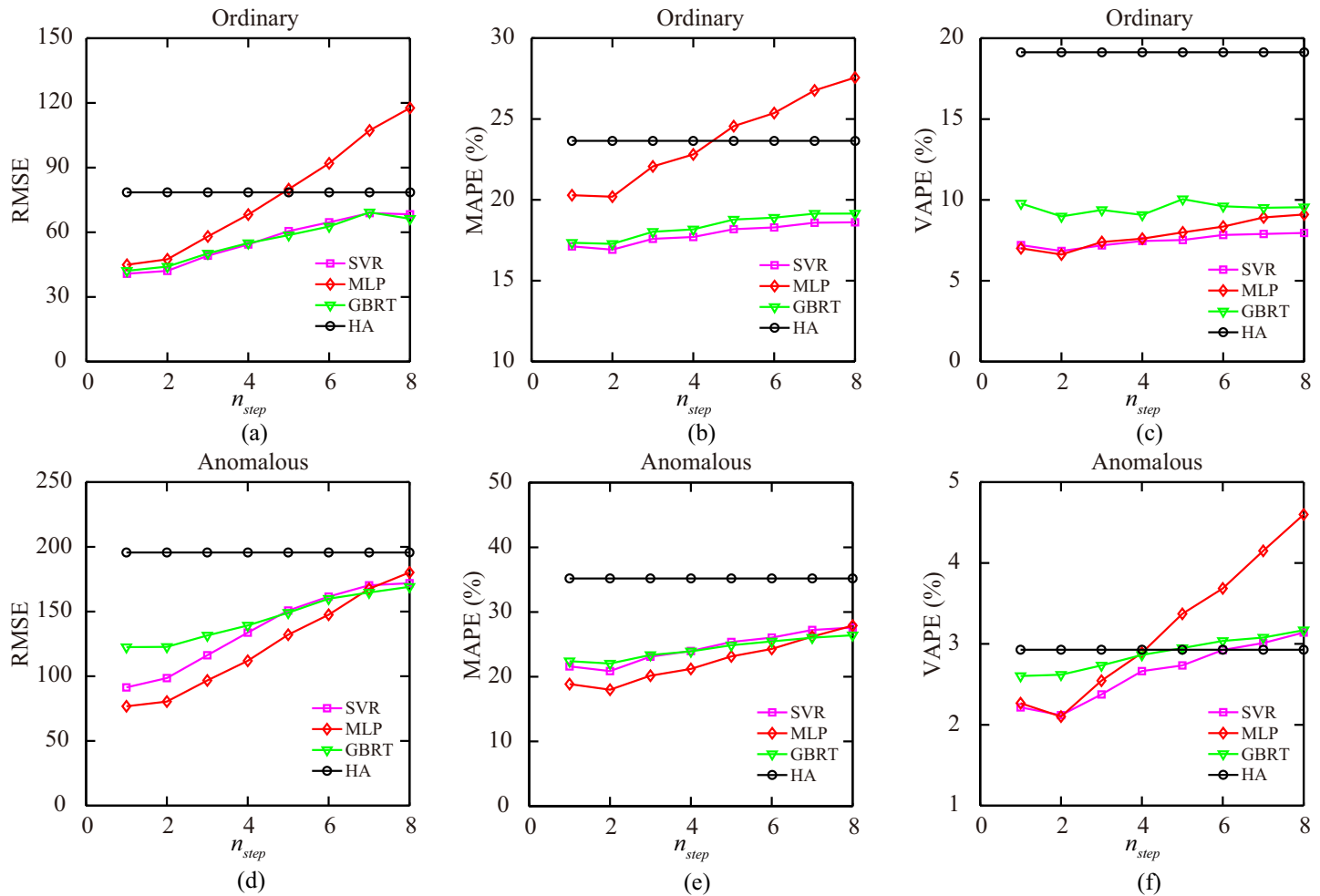


Fig 7. Performance analysis of the predictive models. (a), (b), (c) Performance of the four predictive models under the ordinary out-passenger flow condition for all stations during the whole test period. (d), (e), (f) Same as (a), (b), and (c), but for the results under the anomalous out-passenger flow condition.

<https://doi.org/10.1371/journal.pone.0202707.g007>

improvement. In this paper, we generated four models to predict passenger flow in each station of the Shenzhen subway system. We investigated how long in advance passenger flows could be accurately predicted. Under ordinary traffic condition, acceptable results can be obtained even 2 h in advance, while under anomalous traffic condition, the prediction accuracy of all predictive models was not acceptable when prediction was made 1 h in advance. Li *et al.* [53] compared detrending models and multi-regime models trying to find appropriate traffic prediction models in practices. Our finding highlights the importance of selecting proper models, SVR model and MLP model respectively performed best in ordinary and anomalous traffic conditions. Our finding also highlights that compared with the selection of models, inherent patterns of passenger flows are more prominently influencing the accuracy of prediction. According to the analysis and results of the present study, when passenger flows are relatively stable, SVR prediction model is suggested. When passenger flows show anomalous patterns, the MLP prediction model can achieve more reliable prediction results. In addition, how long the prediction is made in advance of the target time window should also be considered. As the time window ahead of the target time window n_{step} increases, the prediction

Table 3. The average error of all in-passenger-flow of all stations.

RMSE								
n_{step}	Ordinary				Anomalous			
	SVR	MLP	GBRT	HA	SVR	MLP	GBRT	HA
1	32.57	34.27	37.34	87.16	65.71	58.17	86.38	162.09
2	41.15	44.24	43.32	87.16	86.69	73.54	92.37	162.09
3	51.91	56.98	51.11	87.16	98.04	89.02	98.77	162.09
4	64.68	70.85	58.38	87.16	105.96	102.05	106.85	162.09
5	72.32	82.74	64.39	87.16	115.83	115.81	111.66	162.09
6	75.65	92.17	68.65	87.16	127.21	130.67	118.09	162.09
7	78.04	98.19	69.72	87.16	128.26	141.21	124.02	162.09
8	79.88	102.37	70.66	87.16	126.41	148.13	127.95	162.09
MAPE(%)								
n_{step}	Ordinary				Anomalous			
	SVR	MLP	GBRT	HA	SVR	MLP	GBRT	HA
1	15.16	17.05	15.62	24.22	17.12	14.78	18.18	34.21
2	15.75	18.68	16.35	24.22	18.83	16.54	19.73	34.21
3	16.48	20.41	17.05	24.22	20.35	18.14	20.56	34.21
4	17.09	22.01	17.72	24.22	21.79	20.05	21.54	34.21
5	17.50	23.27	18.02	24.22	23.17	22.17	22.39	34.21
6	17.88	24.60	18.46	24.22	24.35	24.22	23.02	34.21
7	18.11	25.53	18.68	24.22	24.67	26.08	23.52	34.21
8	18.32	26.43	18.80	24.22	25.32	28.10	23.91	34.21
VAPE(%)								
n_{step}	Ordinary				Anomalous			
	SVR	MLP	GBRT	HA	SVR	MLP	GBRT	HA
1	5.66	6.70	7.46	13.38	2.13	2.01	2.17	2.27
2	5.87	6.86	7.91	13.38	2.25	2.34	2.40	2.27
3	7.01	8.69	9.24	13.38	2.39	2.48	2.39	2.27
4	8.09	9.98	10.39	13.38	2.65	3.08	2.60	2.27
5	8.58	9.50	10.24	13.38	3.21	4.14	2.77	2.27
6	9.68	11.27	11.48	13.38	3.53	5.28	2.80	2.27
7	9.34	10.61	11.23	13.38	3.42	6.34	2.86	2.27
8	9.37	10.97	10.95	13.38	3.67	8.11	2.93	2.27

<https://doi.org/10.1371/journal.pone.0202707.t003>

Table 4. The average error of all Out -passenger-flow of all stations.

RMSE								
n_{step}	Ordinary				Anomalous			
	SVR	MLP	GBRT	HA	SVR	MLP	GBRT	HA
1	40.76	44.95	42.15	78.53	91.41	76.78	122.53	195.62
2	42.09	47.47	44.09	78.53	98.56	80.46	122.72	195.62
3	49.14	58.06	50.20	78.53	116.20	96.77	131.51	195.62
4	54.38	68.22	54.99	78.53	133.71	111.96	139.23	195.62
5	60.54	80.00	58.68	78.53	150.79	132.11	149.09	195.62
6	64.63	91.99	62.68	78.53	161.55	147.50	159.90	195.62
7	68.90	107.22	69.25	78.53	170.30	167.59	164.74	195.62

(Continued)

Table 4. (Continued)

8	68.33	117.66	66.26	78.53	171.93	180.23	169.14	195.62
MAPE(%)								
	Ordinary				Anomalous			
n_{step}	SVR	MLP	GBRT	HA	SVR	MLP	GBRT	HA
1	17.12	20.29	17.35	23.65	21.60	18.86	22.40	35.19
2	16.92	20.20	17.28	23.65	20.87	17.99	22.03	35.19
3	17.59	22.07	18.02	23.65	23.10	20.16	23.36	35.19
4	17.70	22.81	18.17	23.65	23.97	21.20	23.91	35.19
5	18.19	24.56	18.79	23.65	25.37	23.14	24.87	35.19
6	18.30	25.37	18.90	23.65	26.04	24.29	25.47	35.19
7	18.59	26.77	19.15	23.65	27.22	26.20	26.03	35.19
8	18.61	27.56	19.15	23.65	27.60	27.89	26.40	35.19
VAPE(%)								
	Ordinary				Anomalous			
n_{step}	SVR	MLP	GBRT	HA	SVR	MLP	GBRT	HA
1	7.21	7.01	9.78	19.13	2.21	2.27	2.60	2.93
2	6.84	6.63	8.98	19.13	2.12	2.10	2.62	2.93
3	7.19	7.40	9.38	19.13	2.37	2.55	2.73	2.93
4	7.46	7.60	9.07	19.13	2.66	2.89	2.86	2.93
5	7.52	8.00	10.05	19.13	2.73	3.37	2.95	2.93
6	7.83	8.35	9.61	19.13	2.92	3.69	3.04	2.93
7	7.89	8.92	9.51	19.13	3.01	4.15	3.08	2.93
8	7.96	9.10	9.55	19.13	3.14	4.60	3.17	2.93

<https://doi.org/10.1371/journal.pone.0202707.t004>

Table 5. The average error of all in-passenger-flow ($n_{step} = 2$).

Method	TOP 25%			ALL		
	RMSE	MAPE (%)	VAPE (%)	RMSE	MAPE (%)	VAPE (%)
SVR	82.01	10.51	1.18	50.05	17.29	6.41
MLP	80.73	12.02	1.28	51.41	19.99	6.21
GBRT	87.22	10.90	1.74	56.18	17.72	8.41
HA	174.14	22.61	9.12	95.56	24.73	17.73

<https://doi.org/10.1371/journal.pone.0202707.t005>

Table 6. The average error of all Out -passenger-flow ($n_{step} = 2$).

Method	TOP 25%			ALL		
	RMSE	MAPE (%)	VAPE (%)	RMSE	MAPE (%)	VAPE (%)
SVR	81.19	11.48	1.20	48.32	16.10	5.48
MLP	83.15	13.40	1.48	48.35	18.44	6.36
GBRT	94.29	11.58	1.46	51.09	16.73	7.31
HA	162.82	20.15	4.35	98.25	25.34	12.24

<https://doi.org/10.1371/journal.pone.0202707.t006>

error increases. The prediction errors of the three types of knowledge discovery models gradually approach the prediction error of the simple HA model. Hence, in the condition that n_{step} is large, the simple HA model can be a good option given its low computation cost. We think our results can offer useful information for the management of public transportation, which includes adjusting operating frequency and alleviating passenger congestion.

Finally, we would like to discuss the limitations of this study, and future work. First, the four predictive models were used in their most basic forms and we did not cover all existing models for traffic prediction. Variants of these fundamental models could further improve the accuracy of predictions. Next, better classifications of traffic conditions or subway stations could further improve the prediction accuracy, and are worthy of future work. Finally, transportation information on social media websites is usually prior to the emergence of actual mobility, and therefore incorporating this kind of information with traditional urban transportation data is definitely an interesting future research direction [54, 55].

Supporting information

S1 File. The minimal dataset to replicate this study.
(CSV)

Acknowledgments

The authors thank H. Lu for valuable discussions.

Author Contributions

Conceptualization: Ximan Ling, Pu Wang.

Data curation: Fan Zhang.

Funding acquisition: Pu Wang.

Methodology: Ximan Ling, Zhiren Huang, Fan Zhang, Pu Wang.

Project administration: Pu Wang.

Software: Ximan Ling, Zhiren Huang.

Validation: Ximan Ling.

Visualization: Ximan Ling, Zhiren Huang, Chengcheng Wang.

Writing – original draft: Ximan Ling.

Writing – review & editing: Chengcheng Wang, Fan Zhang, Pu Wang.

References

1. Wang P, Hunter T, Bayen AM, Schechtner K, González MC. Understanding Road Usage Patterns in Urban Areas. *Scientific Reports*. 2012; 2: 1001. <https://doi.org/10.1038/srep01001> PMID: 23259045
2. Li D, Fu B, Wang Y, Lu G, Berezin Y, Stanley HE, et al. Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proceedings of the National Academy of Sciences*. 2015; 112: 669–672. <https://doi.org/10.1073/pnas.1419185112> PMID: 25552558
3. Wang J, Li Y, Liu J, He K, Wang P. Vulnerability analysis and passenger source prediction in urban rail transit networks. *PLoS ONE*. 2013; 8. <https://doi.org/10.1371/journal.pone.0080178> PMID: 24260355
4. He K, Xu Z, Wang P. A hybrid routing model for mitigating congestion in networks. *Physica A: Statistical Mechanics and its Applications*. Elsevier B.V.; 2015; 431: 1–17. <https://doi.org/10.1016/j.physa.2015.02.087>

5. He K, Xu Z, Wang P, Deng L, Tu L. Congestion Avoidance Routing Based on Large-Scale Social Signals. *IEEE Transactions on Intelligent Transportation Systems*. 2016; 17: 2613–2626. <https://doi.org/10.1109/TITS.2015.2498186>
6. Yang X, Chen A, Ning B, Tang T. Measuring route diversity for urban rail transit networks: A case study of the Beijing Metro Network. *IEEE Transactions on Intelligent Transportation Systems*. 2017; 18: 259–268. <https://doi.org/10.1109/TITS.2016.2566801>
7. Zheng Z, Huang Z, Zhang F, Wang P. Understanding coupling dynamics of public transportation networks. *EPJ Data Sci. The Author(s)*; 2018; 7: 23. <https://doi.org/10.1140/epjds/s13688-018-0148-6>
8. Du WB, Zhou XL, Chen Z, Cai KQ, Cao X Bin. Traffic dynamics on coupled spatial networks. *Chaos, Solitons and Fractals*. 2014; 68: 72–77. <https://doi.org/10.1016/j.chaos.2014.07.009>
9. Yule GU. On a method of investigating periodicities in disturbed series, with special reference to Wolf-er's sunspot numbers. *Philosophical Transactions of the Royal Society of London*. 1927; 226: 267–298.
10. Walker G. On Periodicity in Series of Related Terms. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 1931; 131: 518–532. <https://doi.org/10.1098/rspa.1931.0069>
11. Box GEP, Jenkins GM, Reinsel GC. *Time Series Analysis: Forecasting and Control*. *Journal of Time Series Analysis*. 1971; 22: 199–201. <https://doi.org/10.1016/j.ijforecast.2004.02.001>
12. Ahmed MS, Cook AR. Analysis of freeway traffic time-series data by using Box-Jenkins techniques. *Transportation Research Record*. 1979; 116.
13. Edes YJS, Michalopoulos PG, Plum RA. Improved Estimation of Traffic Flow for Real-Time Control. *Transportation Research Record Journal of the Transportation Research Board*. 1981; 28–39.
14. Okutani I, Stephanedes YJ. Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research Part B*. 1984; 18: 1–11. [https://doi.org/10.1016/0191-2615\(84\)90002-X](https://doi.org/10.1016/0191-2615(84)90002-X)
15. Wang Y, Papageorgiou M, Messmer A. Real-time freeway traffic state estimation based on extended Kalman filter: Adaptive capabilities and real data testing. *Transportation Research Part A: Policy and Practice*. Elsevier Ltd; 2008; 42: 1340–1358. <https://doi.org/10.1016/j.tra.2008.06.001>
16. Li R, Jiang C, Zhu F, Chen X. Traffic Flow Data Forecasting Based on Interval Type-2 Fuzzy Sets Theory. *IEEE/CAA Journal of Automatica Sinica*. 2016; 3: 141–148.
17. Williams BM, Hoel L a. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. *Journal of Transportation Engineering*. 2003; 129: 664–672. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2003\)129:6\(664\)](https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664))
18. Schimbinschi F, Moreira-Matias L, Nguyen VX, Bailey J. Topology-regularized universal vector autoregression for traffic forecasting in large urban areas. *Expert Systems with Applications*. 2017; 82: 301–316. <https://doi.org/10.1016/j.eswa.2017.04.015>
19. Xue R, Sun DJ, Chen S. Short-term bus passenger demand prediction based on time series model and interactive multiple model approach. *Discrete Dynamics in Nature and Society*. 2015; 2015: 1. <https://doi.org/10.1155/2015/682390>
20. Ma X, Zhang J, Ding C, Wang Y. A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. *Comput Environ Urban Syst*. Elsevier; 2018; 70: 113–124. <https://doi.org/10.1016/j.compenvurbnsys.2018.03.001>
21. Chrobok R, Wahle J, Schreckenber M. Traffic forecast using simulations of large scale networks. *ITSC 2001 2001 IEEE Intelligent Transportation Systems Proceedings (Cat No01TH8585)*. 2001; 434–439. <https://doi.org/10.1109/ITSC.2001.948696>
22. McCrea J, Moutari S. A hybrid macroscopic-based model for traffic flow in road networks. *European Journal of Operational Research*. Elsevier B.V.; 2010; 207: 676–684. <https://doi.org/10.1016/j.ejor.2010.05.018>
23. Ding C, Wang D, Ma X, Li H. Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees. *Sustain*. 2016; 8. <https://doi.org/10.3390/su8111100>
24. Davis BGA, Member A, Nihan NL. Nonparametric regression and short-term freeway traffic forecasting. *Journal of Transportation Engineering*. 1991; 117: 178–188.
25. Clark S. Traffic Prediction Using Multivariate Nonparametric Regression. *Journal of Transportation Engineering*. 2003; 129: 161–168. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2003\)129:2\(161\)](https://doi.org/10.1061/(ASCE)0733-947X(2003)129:2(161))
26. Vythoukas PC. Alternative Approaches to Short Term Traffic Forecasting for use in Driver Information Systems. *12th International Symposium on Traffic Flow Theory and Transportation*. 1993. p. 22.
27. Dougherty M. A review of neural networks applied to transport. *Science*. 1995; 3: 247–260.
28. Smith BL, Demetsky MJ. Traffic flow forecasting: comparison of modeling approaches. *Journal of Transportation Engineering*. 1997; 123: 261–266. [https://doi.org/10.1061/\(ASCE\)0733-947X\(1997\)123:4\(261\)](https://doi.org/10.1061/(ASCE)0733-947X(1997)123:4(261))

29. Park B, Messer C, Urbanik II T. Short-Term Freeway Traffic Volume Forecasting Using Radial Basis Function Neural Network. *Transportation Research Record: Journal of the Transportation Research Board*. 1998; 1651: 39–47. <https://doi.org/10.3141/1651-06>
30. Li Y, Wang X, Sun S, Ma X, Lu G. Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks. *Transportation Research Part C: Emerging Technologies*. Elsevier Ltd; 2017; 77: 306–328. <https://doi.org/10.1016/j.trc.2017.02.005>
31. Park D, Rilett LR, Han G. Spectral basis neural networks for real-time travel time forecasting. *Journal of Transportation Engineering*. 1999; 125: 515–523.
32. Abdulhai B, Porwal H, Recker W. Short term freeway traffic flow prediction using genetically-optimized time-delay-based neural networks. In: *California Partners for Advanced Transit and Highways (PATH)* [Internet]. 1999 pp. 219–234. Available: <http://escholarship.org/uc/item/4t05p2mp>
33. van Lint J, Hoogendoorn S, van Zuylen H. Freeway Travel Time Prediction with State-Space Neural Networks: Modeling State-Space Dynamics with Recurrent Neural Networks. *Transportation Research Record: Journal of the Transportation Research Board*. 2002; 1811: 30–39. <https://doi.org/10.3141/1811-04>
34. Tseng FM, Yu HC, Tzeng GH. Combining neural network model with seasonal time series ARIMA model. *Technological Forecasting and Social Change*. 2002; 69: 71–87. [https://doi.org/10.1016/S0040-1625\(00\)00113-X](https://doi.org/10.1016/S0040-1625(00)00113-X)
35. Vlahogianni EI, Karlaftis MG, Golias JC. Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. *Transportation Research Part C: Emerging Technologies*. 2005; 13: 211–234. <https://doi.org/10.1016/j.trc.2005.04.007>
36. Zhang Y, Ye Z. Short-term traffic flow forecasting using fuzzy logic system methods. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*. 2008; 12: 102–112. <https://doi.org/10.1080/15472450802262281>
37. Wei Y, Chen MC. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C: Emerging Technologies*. Elsevier Ltd; 2012; 21: 148–162. <https://doi.org/10.1016/j.trc.2011.06.009>
38. Cortes C, Vapnik V. Support Vector Networks. *Machine Learning*. 1995; 20: 273–297. <https://doi.org/10.1007/BF00994018>
39. Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V. Support vector regression machines. *Advances in Neural Information Processing Systems*. 1997; 1: 155–161. <https://doi.org/10.1.1.10.4845>
40. Wu CH, Ho JM, Lee DT. Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems*. 2004; 5: 276–281. <https://doi.org/10.1109/TITS.2004.837813>
41. Vanajakshi L, Rilett LR. Support vector machine technique for the short term prediction of travel time. *IEEE Intelligent Vehicles Symposium, Proceedings*. 2007. pp. 600–605. <https://doi.org/10.1109/IVS.2007.4290181>
42. Jiang X, Zhang L, Chen MX. Short-term forecasting of high-speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China. *Transportation Research Part C: Emerging Technologies*. 2014; 44: 110–127. <https://doi.org/10.1016/j.trc.2014.03.016>
43. He G, Ma S. A study on the short-term prediction of traffic volume based on wavelet analysis. *Proceedings The IEEE 5th International Conference on Intelligent Transportation Systems*. 2002. pp. 731–735. <https://doi.org/10.1109/ITSC.2002.1041309>
44. Jiang X, Adeli H, Asce HM. Dynamic Wavelet Neural Network Model for Traffic Flow Forecasting. *Journal of Transportation Engineering*. 2005; 131: 771–779.
45. Sun Y, Leng B, Guan W. A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. *Neurocomputing*. Elsevier; 2015; 166: 109–121. <https://doi.org/10.1016/j.neucom.2015.03.085>
46. Zhou Z. *Machine learning*. Beijing, China: Tsinghua University Press, 2016, pp. 97–139.
47. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986; 323: 533–536. <https://doi.org/10.1038/323533a0>
48. Friedman JH. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*. 2001; 29: 1189–1232. <https://doi.org/10.1214/aos/1013203451>
49. Li H. *Statistical learning method*. Beijing, China: Tsinghua University Press, 2012, pp. 137–152.
50. Ester M, Kriegel HP, Sander J, Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*. 1996. pp. 226–231. <https://doi.org/10.1.1.1.1980>
51. <http://sjzcsdj.lofter.com/>

52. Huang Z, Wang P, Zhang F, Gao J, Schich M. A mobility network approach to identify and anticipate large crowd gatherings. *Transp Res Part B Methodol*. Elsevier Ltd; 2018; 114: 147–170. <https://doi.org/10.1016/j.trb.2018.05.016>
53. Li Z, Li Y, Li L. A comparison of detrending models and multi-regime models for traffic flow prediction. *IEEE Intelligent Transportation Systems Magazine*. 2014; 6: 34–44. <https://doi.org/10.1109/MITS.2014.2332591>
54. Zheng X, Chen W, Wang P, Shen D, Chen S, Wang X, et al. Big Data for Social Transportation. *IEEE Transactions on Intelligent Transportation Systems*. 2015; 17: 620–630. <https://doi.org/10.1109/TITS.2015.2480157>
55. Lv Y, Chen Y, Zhang X, Duan Y, Li N. Social Media based Transportation Research : the State of the Work and the Networking. *IEEE/CAA Journal of Automatica Sinica*. 2017; 4: 19–26.