Research Article

Short-Term Origin-Destination Based Metro Flow Prediction with Probabilistic Model Selection Approach

Xiaoqing Dai (ID),[1] **Lijun Sun,**[1] **and Yanyan Xu** (ID)[2]

[1]*The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai, China*
[2]*Department of Civil and Environmental Engineering, MIT, USA*

Correspondence should be addressed to Yanyan Xu; yanyanxu@mit.edu

Reliable prediction of short-term passenger flow could greatly support metro authorities' decision processes, help passengers to adjust their travel schedule, or, in extreme cases, assist emergency management. The inflow and outflow of the metro station are strongly associated with the travel demand within metro networks. The purpose of this paper is to obtain such prediction. We first collect the origin-destination information from the smart-card data and explore the passenger flow patterns in a metro system. We then propose a data driven framework for short-term metro passenger flow prediction with the ability to utilize both spatial and temporal related information. The approach adopts two forecasts as basic models and then uses a probabilistic model selection method, random forest classification, to combine the two outputs to achieve a better forecast. In the experiments, we compare the proposed model with four other prediction models, i.e., autoregressive-moving-average, neural networks, support vector regression, and averaging ensemble model, as well as the basic models. The results indicate that the proposed approach outperforms the others in most cases. The origin-destination flows extracted from smart-card data can be successfully exploited to describe different metro travel patterns. And the framework proposed here, especially the probabilistic combination method, can improve the performance of short-term transportation prediction.

1. Introduction

Many cities, due to increasing travel demands and ever-extending city coverage, rely more and more on metro systems. With reliable, efficient, and safe service, metro networks are experiencing a sharp rise in ridership [1]. The successful delivery of such a huge amount of passengers requires a high level of operational services where short-term passenger flow prediction plays a key role. Short-term passenger flow prediction is the prerequisite and foundation for the adaptive control of traffic condition in the intelligent transit system. The prediction of passenger information can be extensively applied to the advanced traffic management systems and the advanced travel information systems to help with facility improvement, operation planning, revenue management, and even emergency evacuation.

Short-term traffic prediction and short-term passenger flow prediction are successful applications of short-term transportation prediction in literature [2]. Unlike metro passenger demand forecasting, traffic forecasting has a long research history and has various analytic, statistical, and simulation based models. Parametric models include ARMA, seasonal ARMA, Kalman filtering, etc., while k-nearest neighbors (kNN) approach and spectral analysis are some frequently used nonparametric models [3]. In recent years, various data driven methods and machine learning models have attracted much attention. Many researchers have moved from what can be considered as a classical statistical perspective (the ARMA family of models), to neural and evolutionary computational approaches. This significant leap from analytical to data driven modeling has been marked by an overwhelming increase of computational intelligence and data mining approaches to analyzing the data [4]. Machine learning algorithms like neural network (NN) [5], support vector regression (SVR) [6], and Gaussian Processes [7] have been adopted frequently to perform short-term traffic prediction. Compared to traffic prediction for ground transportation, short-term passenger demand prediction for

transit systems is a relatively new research field. Here are some related works we found. Tsai et al. [8] constructed two types of improved neural network model. The first is multiple temporal units neural network (MTUNN), which deals with distinctive input information via designated connections in the network. The second is parallel ensemble neural network (PENN), which deals with different input information in several individual models. The results show that both MTUNN and PENN outperform conventional multilayer perception neural networks. Wei and Chen [2] developed a hybrid forecasting approach which combines empirical mode decomposition and back-propagation neural networks and found that the approach performs well and stably in forecasting the short-term metro passenger flow. Sun et al. [9] proposed a hybrid model of Wavelet Support Vector Machine (SVM). The method first decomposes the passenger flow data into different high frequency and low frequency series by wavelet and then predicted these series by SVM separately. In the end, the diverse predicted sequences are reconstructed by wavelet. This method is claimed to have the best forecasting performance compared with the state-of-the-art techniques in the year 2015. Li et al. [10] presented a multiscale radial basis function network for forecasting the irregular fluctuation of metro passenger flows. And the prediction performance is said to outperform current prevailing computational intelligence methods for nonregular demand forecasting at least 30 min prior. Silva et al. [11] proposed an approach to analyzing massive transportation systems with goals of quantifying the effects of shocks in the system, such as line and station closures, and to predict traffic volumes. They used past disruptions to predict unseen scenarios, by relying on simple physical assumptions of passenger flow and a system-wide model for origin-destination (OD) movement.

Combining forecast takes advantage of the availability of both multiple information and computing resources for data-intensive forecasting to improve forecasting accuracy [12]. They are wildly used in a postprocessing phase to improve the stability and accuracy of the individual models [12–16]. Research on this methods started from around the sixties and now is available in a wide range from the robust simple average to the far more theoretically complex, such as state-space methods that attempt to model nonstationarity in the combining weights. Among most literatures, it is documented that a combined forecast improves the overall accuracy to a great extent and is often better than the forecast of each component model [17]. A majority of them form a weighted linear combination of the component forecasts. The statistical averaging techniques, e.g., simple average, trimmed mean, Winsorized mean, median, etc., are the most basic ensemble methods, as they do not explicitly determine the combining weights. Many studies found that these fairly simple methods reasonably outperformed a number of more advanced combining schemes [18].

In this work, we propose a framework to select and combine forecasts to achieve better prediction performances. In the first step, two basic models are initially adopted to obtain two forecasts, which are k-nearest neighbors (kNN) and adaptive boosting (Adaboost). Due to the complexity of the travel pattern and flow, these individual prediction models may have different performance when facing different situations. Thus we compared the results based on the fitted regression relationships between forecasting accuracies and features for each basic model. Finally, probabilistic combination is applied to combine the refined results into a final forecasting value. The subsequent validation performed indicates that the proposed model outperforms some widely used forecasting model, i.e., ARMA, NN, SVR, and an averaging ensemble model using the same basic models, in terms of prediction accuracy and stability.

The remainder of the paper is organized as follows. Section 2 presents the dataset used in this study. Section 3 provides an intuitive analysis of the data. Section 4 provides the whole method and a case study of passenger flow prediction using the real world data is presented. Comparative analyses of the prediction performances are provided. Section 5 gives the result and visualizes the prediction performances. Finally, conclusions are drawn and future research directions are indicated in Section 6.

2. Data Description

Metro line 1 in Zhengzhou City, China, is selected for a case study. The city lies in the middle of China and has a very prosperous and promising metro demand for its 9.568 million population. During the period our dataset covered, line 1 was the only metro line in the city with an average daily passenger volume about 300, 000. This dataset reflects a typical growth pattern of a new-built metro system and the exploration of its demand pattern helps the policymaker to understand the passenger volume fluctuation of a medium city.

The datasets were collected from automatic fare collection (AFC) system which covers 667 days, from July 20th, 2014, to May 16th, 2016. AFC records comprise all of the journeys by metro. The useful information in each record for this case consists of a time stamp, a location code, a transaction type, and card IDs. The location code uniquely identifies each of the 19 stations of the system that were active during the period covered by our data. The two transaction types of our interest are generated when a passenger touches the smart-card reader at the entrance ("tap-in" event) or at the exit ("tap-out" event) of a station. Card IDs allow us to pair the tap-in and tap-out records in which way ODs are obtained. We discarded all tap-in records that are not matched to a tap-out and vice versa. Paired OD flows are aggregated by a time interval of 15 minutes. Each day is composed of 96 time intervals. Our analysis covers all weekdays and weekends. Some statistics of instation flow and outstation flow are shown in Table 1, including average 50th and 80th percentile value and standard deviation of 15 minutes' flow for all days and every station. Most stations have balanced in and out station flows. It is noteworthy that Table 1 shows a high volatility of the flows, which also indicates the difficulty of our prediction task.

3. Travel Demand Analysis

3.1. In- and outstation Flow Patterns. We investigate the patterns of the inflow and outflow at each station with the

TABLE 1: Statistics of inflow and outflow (pax/15min).

station	inflow			outflow		
	50%	80%	std	50%	80%	std
1	165	284	146	133	229	112
2	53	94	67	55	89	51
3	115	177	112	126	214	121
4	120	188	120	121	185	108
5	88	136	72	87	138	68
6	94	153	97	96	145	79
7	128	196	105	136	195	100
8	257	426	199	293	478	215
9	249	560	275	277	582	274
10	137	204	99	136	215	103
11	209	315	162	227	337	171
12	137	199	101	146	207	111
13	121	182	98	115	167	102
14	112	212	145	117	206	165
15	105	167	89	99	144	78
16	78	131	83	71	113	77
17	101	159	98	96	148	96
18	173	380	195	214	360	172
19	189	335	161	177	311	153

average counts in each timeslot, as shown in Figure 1. The stations could be split into the following four groups according to their flow patterns. (i) The stations which display a two-peak pattern on both the in- and the outflows on weekdays, e.g., Station 12. The main reason is that these stations are located in the zones which serve as both residence and nonresidence areas. The number of people departing from these zones is comparable with the number of those arriving at these zones during the peak hours. (ii) The stations where the inflow shows only one strong peak during the morning peak hour, while, the outflow shows one strong peak during the evening peak hour, e.g., Station 4. Such stations are located in the zones which mainly serve as residential area. Most of the travelers are commuters in these zones, who usually depart from home to work in the morning peak hour and back to home in the evening. (iii) On the contrary with the second pattern, some other stations have a strong morning peak on the outflow and evening peak on the inflow, e.g., Station 16. These stations are in the nonresidential areas, including the business or industrial area, and provide a great number of jobs. (iv) The stations which display peaks in the noon on both of the inflow and outflow, such as Station 19. Such stations may be located in the popular commercial area, which is the destination of lots of noncommuters. And there are some stations showing mixed characters of those four divisions, like Station 10. They are mostly located in the central district of the city and surrounded by big shopping malls and public utilities like hospitals and theaters. They show mixed characters of types (iii) and (iv).

*3.2. Metro Travel Distance Distribution.* The distribution of metro trip displacements computed from the dataset is shown in Figure 2. A metro trip refers to a travel from the boarding station to alighting station of a passenger, which reflects the movement of an individual within the metro system. The displacement between any OD pair is calculated from station longitudes and latitudes. The probability increases quickly from the beginning and reaches the peak at about 4 ~ 5 km, and another peak is reached at about 9 km. This may result from the pricing structures and travel purposes. With a flat fare of 2 RMB and some time to enter and exit the metro stations, taking metro to complete very short trips is not economical and time efficient. And most metro trips are for the purpose of daily travel, like commuting, which is not likely to have a very long distance, and also the upper bound of the travel distance is restricted by the city scale. We select Weibull distributions to fit the displacements. The likelihood indicated that the Weibull distribution fits the metro trip displacements slightly better, which is different from the result from [19]. The difference may lie in the structure of urban layout.

3.3. Spatial Correlation. The relationship between each two OD pairs in a complex network, especially the correlation among spatially neighboring OD, has always been a research interest. There is a possibility that related OD flows could be used as potential features to improve the prediction. We tried to explore the possibility by finding out the interdependencies between geographically close OD flows. In Figure 3, OD 4–8 (passenger flow from Station 4 to Station 8) is chosen as the target OD pair; neighboring OD pairs include all OD pairs starting from Stations 2–6 and ending at Stations 6–10. Each individual figure shows the correlation between the neighboring OD pairs with the target OD pair. It can be seen that those ODs with correlation higher than 0.5 are OD 3–8, OD 3–9, OD 4–7, OD 4–9, and OD 5–8. Most of these are one station away from the target OD. After conduct the analysis on all OD flows, we choose a spatial window of 1 (one station away) to construct features in prediction. Thus for one target OD flow, 8 neighboring OD flows are used as feature in prediction.

4. Travel Demand Prediction

4.1. Basic Model Formulation. The datasets are split into three parts to train and test the prediction models. The first 80% of them are used as training set. The remaining 20% are used as *testing set* to examine the performance of proposed combined prediction model. For the training set, we randomly disordered the whole sequence (the first 80% of all data) and then divided the disordered sequence by 5:3 to obtain the *Training Set 1* and *Training Set 2*. By doing the random permutation, we excluded the long-term trend from the training dataset. The *Training Set 1* (50% of all data) is used to train the two basic prediction models, Adaboost and kNN; *Training Set 2* (30% of all data) is used to train the probabilistic classifier, RF, here named *Training Set 2*.

It is noteworthy that the weekdays and weekends are treated separately. In the whole computational process of this present method, OD flow in the next timeslot, $OD(S_O, S_D)_{t+1}$, is selected as target. The forecasts of instation and outstation passenger flows can be obtained by aggregating OD flows by
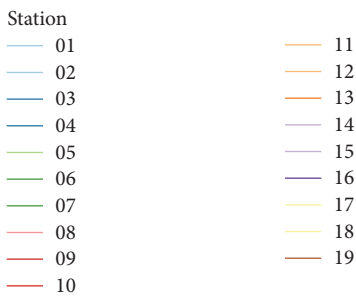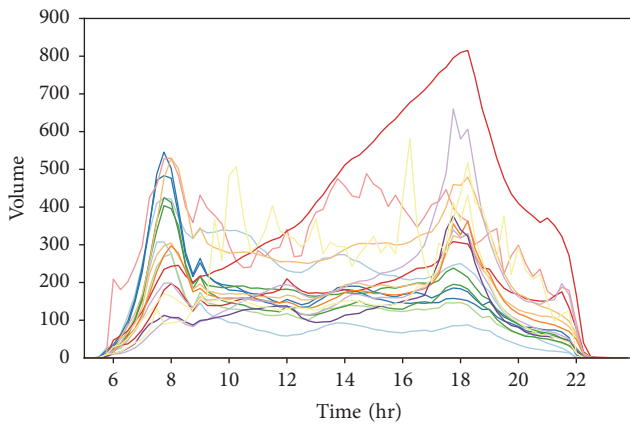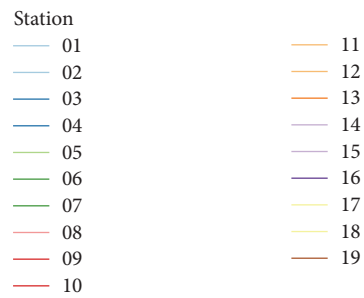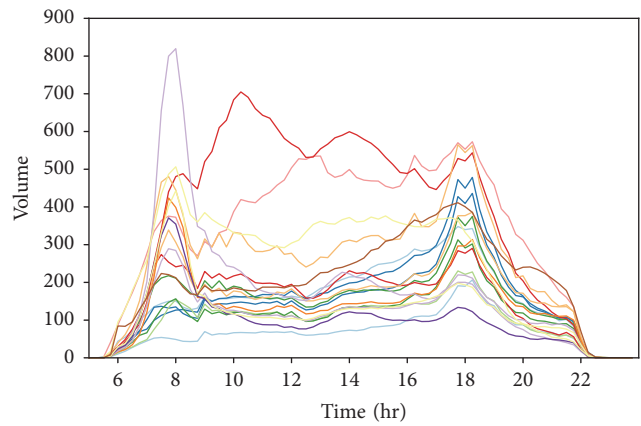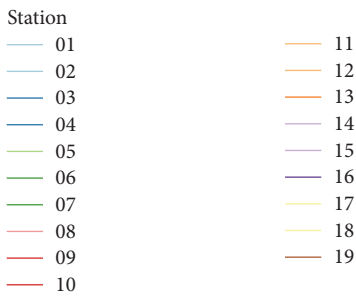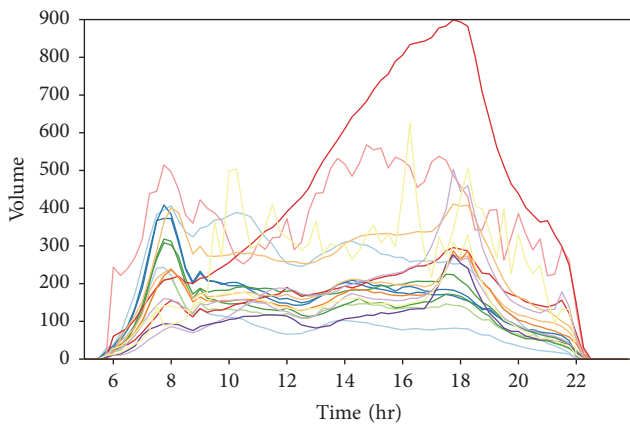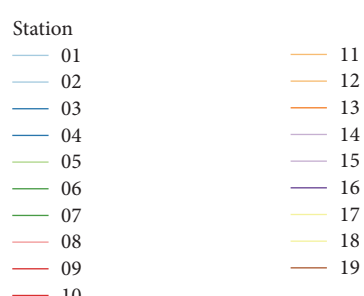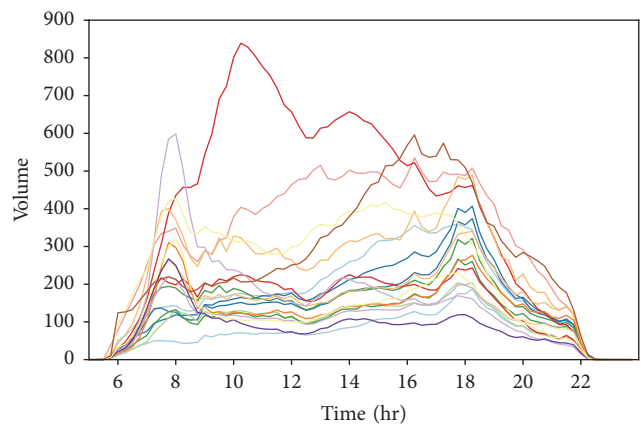
FIGURE 1: The inflow and outflow counts at each station. (a) (b) The average counts of inflow (outflow) every 15-minute timeslot on weekdays. (c) (d) The average counts of inflow (outflow) every 15-minute timeslot on weekends and holidays.

$$1.04 \times \left(\frac{x}{10}\right)^{1.67-1} e^{-\left(\frac{x}{10}\right)^{1.67}}$$
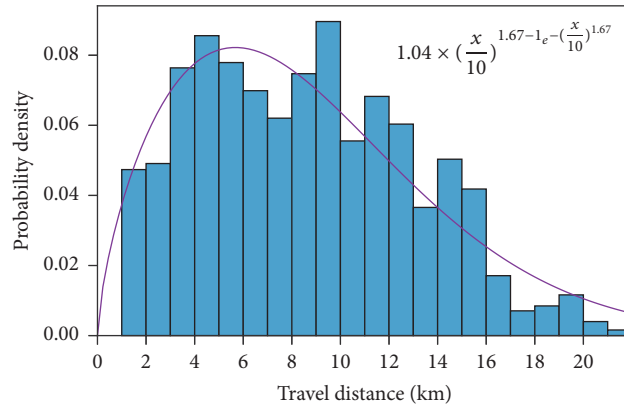
FIGURE 2: Distribution of travel distances approximated by Weibull distribution.
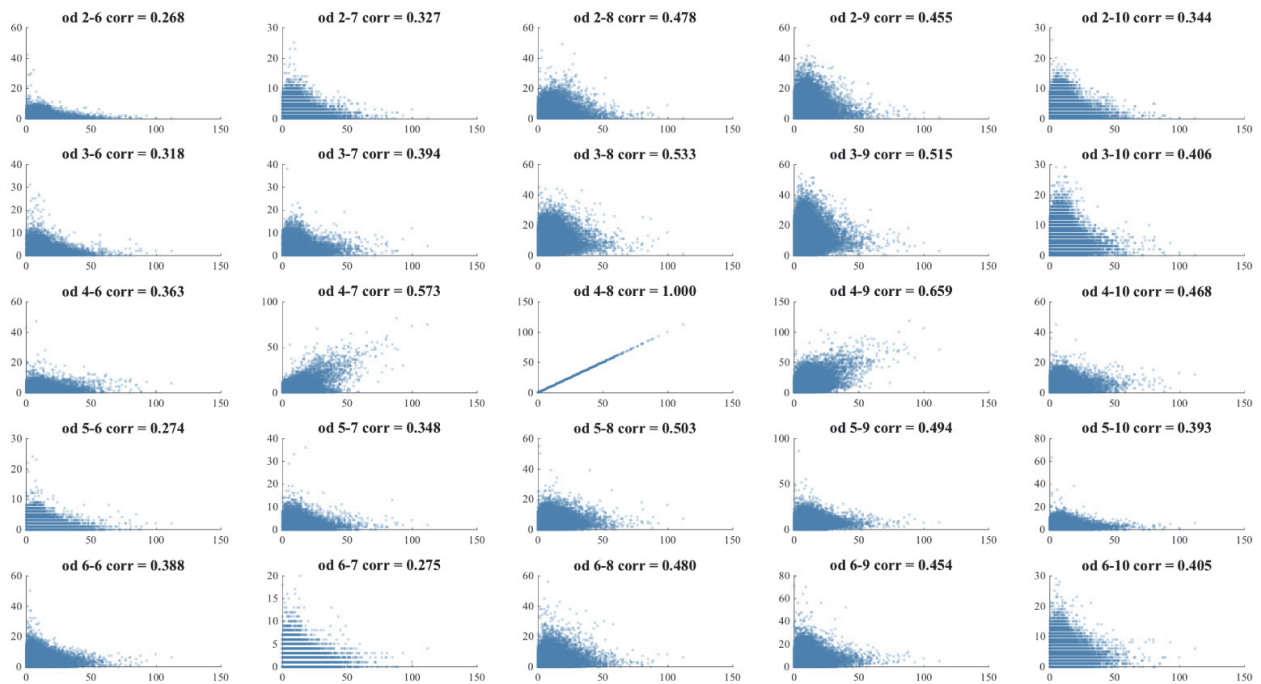
FIGURE 3: The correlation between the target OD flow 4–8 and its neighboring OD flows.

their origin or destination station. Historical values of the targeting and some neighboring OD flows of the target are selected as features, as shown in Figure 4(a).

As we select the spatiotemporal flows as the inputs of the prediction model, the number of features is 45 when the spatial window is $w = 1$(8 neighboring ODs and the target OD) and the time lag is $d = 5$. Such number of features contains redundant information which increases the uncertainty of prediction. Therefore, principal component analysis (PCA) is adopted to decompose the original OD flow into a number of principal components and scores, and it also plays the role of data filtering. In this way, the systematic variations in OD flows are captured in lower dimensions. Consequently, the resulting principal components have a desirable property of being independent of each other, and the weights of principal

components could serve as predictors in prediction process. A detailed description of PCA could be referred to in [20]. In our case, the size of feature matrix $\mathbf{R}$ is $m \times n$. The variable m represents the number of samples; i.e., each row of the data matrix is the estimated OD flow vector in any of the intervals. The variable n represents the number of features. The principal component directions can then be determined by performing Singular Valued Decomposition on the data matrix $\mathbf{R}$

$$\mathbf{R} = \mathbf{U}\Sigma\mathbf{C}^{T} \tag{1}$$

where Σ is a $m \times n$ rectangular-diagonal matrix with positive values called singular values; $\mathbf{U}$ is a $m \times m$ matrix with orthogonal column vectors called the left singular vectors; and $\mathbf{C}$ is a $n \times n$ matrix with orthogonal column vectors called

Independent Variables   **Response**

$$\begin{bmatrix} \mathrm{OD}(S_O, S_D)_{t-d} & \cdots & \mathrm{OD}(S_O, S_D)_t \\ \mathrm{OD}(S_O-1, S_D)_{t-d} & \cdots & \mathrm{OD}(S_O-1, S_D)_t \\ \mathrm{OD}(S_O+1, S_D)_{t-d} & \cdots & \mathrm{OD}(S_O+1, S_D)_t \\ \mathrm{OD}(S_O-1, S_D-1)_{t-d} & \cdots & \mathrm{OD}(S_O-1, S_D-1)_t \\ \mathrm{OD}(S_O, S_D-1)_{t-d} & \cdots & \mathrm{OD}(S_O, S_D-1)_t \\ \mathrm{OD}(S_O+1, S_D-1)_{t-d} & \cdots & \mathrm{OD}(S_O+1, S_D-1)_t \\ \mathrm{OD}(S_O-1, S_D+1)_{t-d} & \cdots & \mathrm{OD}(S_O-1, S_D+1)_t \\ \mathrm{OD}(S_O, S_D+1)_{t-d} & \cdots & \mathrm{OD}(S_O, S_D+1)_t \\ \mathrm{OD}(S_O+1, S_D+1)_{t-d} & \cdots & \mathrm{OD}(S_O+1, S_D+1)_t \end{bmatrix} \qquad \mathrm{OD}(S_O, S_D)_{t+1}$$

$S_O\text{-}1 \quad S_O \quad S_O+1 \quad \cdots \quad S_D\text{-}1 \quad S_D \quad S_D+1$

(a)

Prediction model training phase   **Classifier training phase**

Training Set 1 → PCA → Adaboost → *Training Set 2* → PCA → Prediction of Adaboost → Performance Label

PCA → kNN → *Training Set 2* → PCA → Prediction of kNN →

Performance Label → Random Forest Classifier

Testing Set → PCA → Trained Adaboost / PCA → Trained kNN → Model Selection Probability → **Prediction**

PCA → *Testing Set* ← Random Forest Classifier
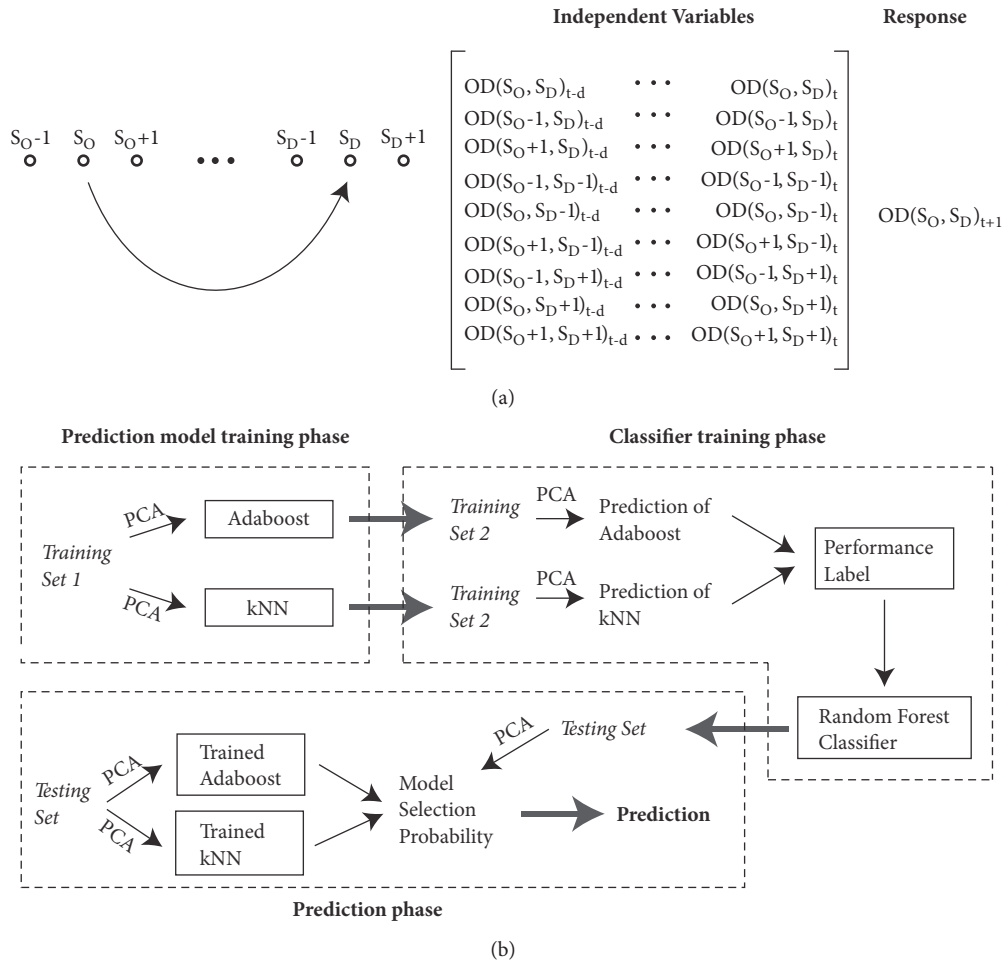
Prediction phase

(b)

FIGURE 4: The framework of the proposed method. (a) The formation of the independent variables and response for OD flow prediction. (b) The pipeline of the proposed probabilistic model selection approach for OD flow prediction.

the right singular vectors. The columns of the matrix **C** are the principal component directions. Alternatively, the columns of C can also be interpreted as the eigenvalues of the matrix $(1/m)\mathbf{R}^T\mathbf{R}$, which is the sample covariance matrix.

Let the individual principal component directions be represented by $\mathbf{c_1}, \mathbf{c_2}, \ldots, \mathbf{c_p}$. Here, $\mathbf{c_1}$ represents the principal component direction with the largest sample variance, $\mathbf{c_2}$ represents the principal component direction with the largest sample variance subject to being orthogonal to $\mathbf{c_1}$, and so on. Assume that only first k directions explain a majority of the variance in the OD flow vector. The first k principal component directions can be represented using a $m\times k$ matrix as

$$\mathbf{C}' = [\mathbf{c_1}, \mathbf{c_2}, \mathbf{c_3}, \ldots, \mathbf{c_k}] \tag{2}$$

Then the $k \times 1$ principal component vector $\mathbf{Z} = [\mathbf{z_1}, \mathbf{z_2}, \mathbf{z_3}, \ldots, \mathbf{z_k}]^T$ of the OD flow vector $\mathbf{r}$ can be written as

$$\mathbf{z} \approx \mathbf{C}'^T\mathbf{r} \tag{3}$$

and the OD flow vector $\mathbf{r}$ can be approximately constructed back as

$$\mathbf{r} \approx \mathbf{C}'\mathbf{z} \tag{4}$$

The extracted principal components are sorted by the percentage of the total variance explained by each principal component, from high to low, and reveal various patterns of OD flow. In this case, we use the first 10 principal components.

The two basic models, Adaboost and kNN, are widely used forecasting approaches. We will not describe them in detail. For Adaboost, we used 500 trees and the maximum depth equals to 4. Interested readers could refer to [21] for an introduction. For kNN, we adopt $k = 12$.

4.2. Probabilistic Model Selection. Two basic forecasts are obtained from previous stage. In the following stage, we combine the prediction results of two individual basic models with a probabilistic model selection approach, random forest classifier, as shown in Figure 4(b). In contrast with the deterministic classification which selects the better one from the individual outputs, the probabilistic classification yields a probability that the output is selected as the better one. This probabilistic mechanism reduces the bias of prediction when the outputs of the individual models are divergent, as one overestimates the traffic flow while another one underestimates the value.

Suppose we have a training set $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots (\mathbf{x}_N, y_N)$, where N is the number of samples in training set. Here in our metro OD flow prediction problem, $\mathbf{x}_i$ is the projected feature vector of the ith independent variables, as shown in Figure 4(a). y_i is the binary label of the ith sample. $y_i = 0$ implies that the prediction of Adaboost is closer to the actual observation than kNN; $y_i = 1$ implies the contrary. In this way, once we applied the trained Adaboost and kNN to the samples in *Training Set 2*, the labels y_i can be collected by comparing the prediction with the ground truth. Consequently, we can adopt the supervised classification method, random forest, to learn the relationship between samples and their labels. Random forest is an ensemble tree model introduced by Leo Breiman [22] consists multiple classification trees. The basic tree models are trained on the data randomly selected from the original training data set, namely, bootstrapping. Usually, the predicted label of a testing sample is the label that appears most frequently in the basic tree models.

For the probabilistic classification, we desire to estimate the probability of each label, that is, $p_{i,k} = P(Y = k | X = \mathbf{x}_i)$. The most simple way is to count the proportion of trees that output class k when observation $\mathbf{x}_i$ is passed to the tree, namely, Voting. In this work, we adopt an out-of-bag method. First, we denote the set of samples used for training the basic trees by $\mathbf{B}_t \in \mathbf{A}$, and the complementary set of $\mathbf{B}_t$ is $\mathbf{B}_t^c \in \mathbf{A}$, a.k.a, out-of-bag method. Here in our work, $\mathbf{A}$ is generated by running the individual predictor on *Training Set 2*. Obviously, for each basic tree model, $\mathbf{B}_t^c$ is not used for training. That is, the terminal nodes are not pure when passing down the samples in $\mathbf{B}_t^c$ to the trained tree. Therefore, $\mathbf{B}_t^c$ gives information about the underlying class probabilities. Suppose that m samples in $\mathbf{B}_t^c$ are assigned to the terminal node i of a basic tree t and these m samples contain k classes; we use the relative frequency of class k as the classification probability of this terminal node. For the terminal node i,

$$p_{k,i} = \frac{\sum_{j=1}^{m} I(j,k)}{m}$$

$$\sum_{k=1}^{K} p_{k,i} = 1 \tag{5}$$

where $I(j,k) = 1$ if the jth sample of m is labeled as class k in $\mathbf{B}_t^c$ and 0 if else. Finally, each terminal node in basic trees is associated with a set of classification probabilities, $p_1, p_2, \dots, p_K$ for a K-class classification problem.

When passing down the testing sample to the trained random forest model, we assign the sample to a terminal node in each basic tree and get a set of classification probabilities. Then we average the probabilities of all basic trees to infer the final classification probabilities. In the context of metro flow prediction, we use two basic prediction models and thus train a binary random forest model to estimate the probability of each prediction model. Once we have applied the probabilistic classification model to the *testing set*, the final prediction can be calculated using the following linear combination:

$$\widehat{\mathbf{C}} = p_0 \widehat{\mathbf{C}_0} + p_1 \widehat{\mathbf{C}_1}$$

$$p_0 + p_1 = 1 \tag{6}$$

where $\mathbf{C_0}$ is the predicted inflow or outflow count of Adaboost and $\mathbf{C_1}$ is that of kNN; p_0 is the probability that Adaboost performs better than kNN on the active sample; p_1 is the probability that kNN performs better than Adaboost.

5. Results Analysis

We implemented several widely used prediction models, i.e., ARMA, neural network (NN) and support vector regression (SVR), using the same dataset to compare predictive performances with the proposed model. It is noteworthy that the proposed combined model is applied to the OD flow first, and then the prediction results are aggregated to get the inflow and outflow, while the three reference methods are directly applied to the aggregated inflow and outflow data, as analyzed in Table 1 and shown in Figure 1, rather than the OD flow. The reason is that we aim to validate not only the proposed combination approach but also the effect of OD flow aggregation. The parameters of p, q, d in ARMA model are calibrated for each time series data (in/outflow of every station), with Bayesian information criterion; For NN, we configure two hidden layers with 5 and 2 nodes, respectively. The model is solved with Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm; for SVR, we choose the radial basis function (RBF) as the kernel, and the two key parameters of RBF are chosen by grid search, with $C = 100$ and $\gamma = 1 \times 10^{-4}$.

Besides, to compare the probabilistic combination with other combining methods, we implement an averaging combination model. Averaging is one of the commonly used model combination method [23]. The model is implemented in the same way with our probabilistic combining model except that the basic models are combined with simple averaging.

To assess the performance of the proposed combined approach, we use two metrics, Mean Absolute Percentage Error (MAPE) and Variance of Absolute Percentage Error (VAPE). MAPE is used to measure the prediction accuracies, and VAPE is used to examine the stability of the proposed and reference models. The formulations of MAPE and VAPE are given as follows:

$$MAPE_{100} = \frac{1}{K} \sum_{k=1}^{K} \frac{|C_k - \widehat{C}_k|}{C_k} \times 100\%, \quad \text{for } C_k \geq 100 \tag{7}$$

$$VAPE_{100} = \text{var}\left(\frac{|\widehat{C}_k - C_k|}{C_k} \right) \times 100\%,$$

$$\text{for } k \in 1, \dots, K \text{ and } \mathbf{C}_k \geq 100 \tag{8}$$

where C_k denotes the actual counts of inflow or outflow at the metro station. Here we only consider the counts larger than 100 as the policymakers are more concerned with large travel flows; the total number of samples is K during the testing phase; $\widehat{C}_k$ is the predicted value produced by the prediction models.

Tables 2, 3, 6, and 7 show the MAPE of the proposed combined model and the referenced models. It can be seen

Table 2: The MAPE of inflow on weekdays.

Station	Combined	ARMA	NN	SVR	Averaging	Adaboost	kNN
1	**13.68**	18.06	14.42	14.27	13.88	13.71	13.83
2	19.06	23.62	19.82	**17.84**	19.26	18.23	20.22
3	**11.69**	20.43	15.82	13.12	11.90	11.84	11.83
4	**12.44**	22.73	17.56	14.7	12.64	12.72	12.47
5	**12.94**	21.56	16.06	14.87	13.13	12.96	13.11
6	11.79	21.84	16.67	13.33	12.00	12.15	**11.70**
7	12.13	23.9	14.13	**11.78**	12.31	12.24	12.28
8	19.99	25.15	22.81	23.89	20.22	20.34	**19.83**
9	**10.23**	13.70	11.79	12.9	10.43	10.33	10.44
10	**11.82**	20.11	14.23	12.72	12.03	11.97	11.87
11	**9.74**	19.31	13.15	10.76	9.96	10.11	9.66
12	11.49	21.01	13.74	12.04	11.69	**11.40**	11.85
13	**11.01**	21.26	14.17	12.16	11.23	11.24	11.04
14	**12.16**	22.34	15.55	13.86	12.38	12.35	12.29
15	**11.33**	19.50	14.85	12.94	11.57	11.65	11.36
16	12.14	21.58	15.88	13.24	12.34	12.62	**11.98**
17	**13.12**	22.54	16.78	14.68	13.33	13.32	13.29
18	20.23	33.79	25.45	29.09	20.54	21.2	**19.94**
19	15.02	17.86	15.13	16.26	15.23	**14.90**	15.33
# of best	11	0	0	2	0	2	4
Average	**13.26**	21.65	16.21	14.97	13.47	13.44	13.40

Table 3: The MAPE of outflow on weekdays.

Station	Combined	ARMA	NN	SVR	Averaging	Adaboost	kNN
1	12.37	16.03	12.91	12.38	12.57	12.31	**11.97**
2	15.66	17.62	14.89	**14.24**	15.84	15.25	13.54
3	**10.27**	16.36	12.21	10.77	10.48	10.56	10.73
4	**11.35**	17.51	14.08	12.85	11.56	11.57	11.96
5	11.78	17.99	13.82	12.54	11.99	11.92	**10.72**
6	**10.63**	22.2	14.77	12.12	10.84	10.96	11.77
7	11.73	17.43	12.97	**11.42**	11.94	11.86	12.11
8	9.85	15.71	10.6	11.64	10.06	**9.97**	14.13
9	**10.44**	14.73	11.35	12.4	10.67	10.70	16.31
10	**11.53**	19.5	13.33	12.18	11.75	11.66	14.79
11	**10.71**	18.62	12.9	11.4	10.92	11.04	13.37
12	**10.47**	21.14	13.1	11.23	10.69	10.76	10.94
13	**10.59**	22.16	15.54	11.3	10.81	11.94	11.76
14	**11.55**	20.18	15.68	13	11.77	11.84	11.90
15	**10.55**	32.62	14.31	11.18	10.76	10.82	11.47
16	12.78	20.56	16.31	14.09	12.98	13.10	**12.38**
17	11.93	22.3	15.62	11.89	12.12	11.98	**11.87**
18	12.40	13.95	**11.48**	12.27	12.59	12.07	12.39
19	12.30	15.37	12.8	13.31	12.50	**12.26**	12.52
# of best	10	0	1	2	0	2	4
Average	**11.51**	18.49	13.61	12.22	11.73	11.66	12.43

TABLE 4: The VAPE of inflow at each station on weekdays.

Station	Combined	ARMA	NN	SVR	Averaging	Adaboost	kNN
1	**1.26**	2.04	3.29	1.39	1.27	1.30	1.56
2	1.39	3.66	1.63	1.41	**1.39**	1.40	1.31
3	**0.73**	2.64	4.03	1.06	0.73	0.77	0.88
4	**0.96**	4.93	1.68	1.35	0.96	1.01	1.19
5	**0.84**	3.84	1.37	1.32	0.85	0.87	0.86
6	**0.90**	3.97	1.05	1.07	0.91	1.01	1.29
7	**0.87**	3.97	1.05	0.91	0.89	0.88	1.21
8	3.15	5.12	4.48	5.19	**3.15**	3.32	4.99
9	**1.60**	1.78	2.03	2.73	1.64	1.56	2.7
10	**0.87**	2.83	2.92	1.07	0.87	0.91	1.62
11	0.69	2.58	0.99	0.77	**0.69**	0.74	1.72
12	0.72	2.79	1.17	0.92	**0.72**	0.73	0.83
13	**0.86**	3.18	1.77	1.08	0.87	0.91	1.17
14	**1.10**	4.89	1.8	1.72	1.11	1.15	1.34
15	**0.82**	2.18	3.15	1.06	0.83	0.86	1.01
16	**0.83**	3.01	1.35	1.06	0.84	0.89	0.89
17	0.98	3.88	1.78	1.62	**0.98**	1.01	1.03
18	**2.84**	8.39	4.82	6.08	2.85	3.03	3.66
19	1.59	2.31	1.58	1.87	1.59	**1.58**	1.94
# of best	13	0	0	0	5	1	0
Average	**1.21**	3.58	2.21	1.77	1.22	1.26	1.64

that among the total 76 (in/outflow at 19 stations on weekdays and weekends) forecasts, the proposed combined model outperforms others in 43 forecasts. By comparing the MAPE of inflow and outflow, we observe that the outflow is generally more predictable than the inflow. That is because the outflow is tied up with the historical inflow, which may be known in the inputs of the prediction models.

The comparison of VAPE is showed in Tables 4, 5, 8, and 9. It can be seen that among the total 76 (in/outflow at 19 stations on weekdays and weekends) forecasts, the proposed combined model outperforms others in 47 forecasts. The improvements are significant compared with ARMA, NN, SVR, and the two basic models. The VAPE of combined model performs best at most of the stations, for inflow or outflow, on weekdays or weekends. However, our probabilistic combined model and averaging combined model yield close VAPEs in many cases. This demonstrates that the prediction ability of combined methods is more stable than other references models. Another interesting finding is that, from the overall view, for the same station, prediction on weekdays is better than weekends and prediction of outflow is better than inflow.

Figures 5 and 6 exhibit the visual comparison of the prediction results with the actual data at selected stations during the operation time of one weekday and weekend, respectively. In contrast with the reference models, our model shows a relatively stable performance both in peak and in small perturbations. Among the four prediction models, ARMA yields more undesired peaks than others. Although the combined model cannot follow the bumpinesses in the actual curve, it gives the most promising prediction among

the four models. However, we observe that the combined model underestimates the evening peaks on weekdays at Stations 4, 12, and 16, as shown in Figures 5(a), 5(c), and 5(d). This shortcoming leaves room for the further improvement. To understand more clearly the ability of predicting peak hours of combined model, we summarized the MAPEs for peak hours in Table 10. Peak hours are selected as 7-8 a.m and 5-6 p.m.

Another aspect of this method we discover during this study is that when applying combination to two basic models, if one individual basic model performs significantly worse than the other, the combination may be not efficient, because the final prediction is a linear combination of the two outputs of the two basic models. In the opposite, another extreme case is that when the two basic models are with very similar characteristics, the combination may also not improve the performance. The reason is that the predicted results of the basic models tend to be same. Therefore, we need to carefully select basic models. The two models should be expert in different scenarios and have overall balanced performance. In this case, the combination will give the best performance.

6. Conclusion

Accurate short-term travel flow prediction in metro systems helps the policymaker to manage the operation of metro more efficiently and also make proactive control possible. With robust and accurate prediction, various control measures could be prepared in advance, for example, volume restriction and timetable reschedule. In this paper, we first analyzed

TABLE 5: The VAPE of outflow at each station on weekdays.

Station	Combined	ARMA	NN	SVR	Averaging	Adaboost	kNN
1	0.90	1.75	1.08	1.17	**0.89**	0.90	1.01
2	0.94	1.42	1.08	1.06	**0.93**	0.94	0.81
3	**0.64**	1.75	0.94	0.74	0.70	0.66	0.84
4	**0.83**	1.99	1.34	1.15	0.88	0.84	1.02
5	0.69	1.74	2.77	0.88	0.70	0.69	**0.63**
6	**0.63**	3.13	1.23	0.83	0.67	0.63	0.90
7	0.72	1.81	0.95	0.81	**0.71**	0.72	1.00
8	**0.84**	1.98	1.2	1.43	0.89	0.84	1.51
9	0.94	1.84	**0.81**	2.37	0.99	0.93	1.99
10	**0.81**	2.37	1.24	0.93	0.86	0.81	1.49
11	**0.80**	2.09	1.22	0.97	0.85	0.81	1.41
12	0.64	2.69	1.12	0.89	0.66	**0.62**	0.81
13	**0.70**	3.58	1.49	0.82	0.74	0.71	0.93
14	0.88	2.81	1.17	1.35	0.87	**0.83**	1.01
15	**0.74**	1.96	1.00	0.75	0.76	0.77	0.85
16	**0.74**	2.88	0.98	1.02	0.78	0.74	0.75
17	**0.86**	3.55	1.39	0.89	0.88	0.88	0.9
18	1.47	1.91	1.75	**1.35**	1.45	1.47	1.77
19	**1.56**	2.14	1.68	1.87	1.58	1.56	1.76
# of best	11	0	1	1	3	2	1
Average	**0.86**	2.29	1.29	1.12	0.88	0.86	1.13

TABLE 6: The MAPE of inflow at each station on weekends.

Station	Combined	ARMA	NN	SVR	Averaging	Adaboost	kNN
1	**14.11**	17.56	15.29	15.54	14.30	14.13	14.28
2	18.08	21.82	18.53	**16.73**	18.27	17.24	19.13
3	11.59	18.88	14.52	13.05	11.76	**11.57**	11.75
4	**12.30**	20.84	14.74	14.48	12.51	12.47	12.35
5	13.29	20.12	14.53	14.96	13.48	**13.13**	13.64
6	**11.71**	20.36	14.81	13.4	11.90	11.97	11.71
7	13.23	23.11	13.63	12.42	13.47	**13.20**	13.57
8	**22.39**	26.23	23.47	28.27	22.61	22.62	22.38
9	**10.49**	13.7	11.44	14.16	10.70	10.60	10.81
10	11.63	19.05	13.46	**12.42**	11.85	11.72	11.75
11	**10.32**	18.8	12.35	11.66	10.55	10.41	10.43
12	11.75	19.88	12.92	11.82	11.95	**11.45**	12.27
13	11.69	20.57	13.76	**12.52**	11.90	11.82	11.78
14	**13.02**	21.49	15.27	14.46	13.24	13.26	13.14
15	**12.41**	18.54	14.76	13.25	12.61	12.54	12.47
16	**12.40**	22.52	16.52	14.2	12.62	12.71	12.39
17	**14.25**	21.78	16.33	15.01	14.46	14.48	14.39
18	**22.23**	33.20	26.41	32.38	22.55	23.25	21.78
19	15.08	16.97	**14.90**	15.66	15.30	14.91	15.55
# of best	11	0	1	3	0	4	0
Average	**13.78**	20.81	15.67	15.60	14.00	13.87	13.98

Table 7: The MAPE of outflow at each station on weekends.

Station	Combined	ARMA	NN	SVR	Averaging	Adaboost	kNN
1	12.64	14.94	**12.20**	12.71	12.84	12.31	12.12
2	16.67	15.77	**14.04**	15.14	16.84	16.21	14.18
3	11.06	14.92	11.96	11.41	11.27	**11.05**	11.04
4	**12.36**	16.41	13.93	13.26	12.57	12.49	12.62
5	12.76	16.26	13.25	12.84	12.96	12.48	**11.15**
6	**11.23**	20.68	13.98	12.24	11.43	11.40	11.92
7	12.37	14.43	13.13	**12.11**	12.58	12.38	12.11
8	**10.16**	15.26	11.41	12.11	10.37	10.29	13.20
9	11.17	13.59	**10.37**	13.30	11.40	11.46	15.25
10	**11.26**	18.38	12.99	12.68	11.47	11.40	14.17
11	**11.37**	17.6	13.41	12.20	11.59	11.49	13.33
12	**11.01**	19.1	13.06	11.53	11.23	11.22	11.53
13	**10.99**	21.15	15.12	11.9	11.22	11.30	11.66
14	**11.95**	19.9	16.15	13.74	12.16	12.12	12.17
15	**10.69**	30.40	13.88	11.47	10.89	10.85	10.88
16	**14.46**	20.45	15.88	15.56	14.67	14.54	13.58
17	11.91	20.84	14.59	12.27	12.12	11.92	**11.62**
18	12.81	13.81	**11.51**	13.61	13.00	12.56	12.59
19	**13.18**	15.96	13.32	17.19	13.37	13.15	13.16
# of best	11	0	4	1	0	1	2
Average	**12.10**	17.94	13.38	13.02	12.31	12.14	12.54

Table 8: The VAPE of inflow at each station on weekends.

Station	Combined	ARMA	NN	SVR	Averaging	Adaboost	kNN
1	**1.31**	2.02	1.44	1.62	1.31	1.35	1.74
2	1.28	2.7	1.89	1.44	1.29	1.28	**1.16**
3	**0.72**	2.29	0.99	1.1	0.72	0.74	0.90
4	**0.84**	3.99	1.19	1.35	0.84	0.87	1.17
5	0.90	3.07	1.47	1.62	0.91	**0.89**	0.90
6	**0.75**	2.86	2.69	1.02	0.76	0.79	1.17
7	0.94	3.57	1.22	**0.93**	0.95	0.93	1.47
8	4.51	6.34	5.64	7.22	**4.50**	4.67	7.78
9	1.85	2.03	2.17	2.87	1.84	**1.81**	3.26
10	**0.83**	2.3	1.13	0.97	0.83	0.85	1.7
11	**0.69**	2.67	3.01	1.03	0.69	0.71	1.84
12	0.76	2.56	1.03	0.89	0.76	**0.75**	0.91
13	**0.84**	2.84	1.13	0.99	0.85	0.86	1.13
14	**1.41**	4.82	2.63	1.91	1.42	1.48	1.78
15	**0.92**	2.09	1.20	1.04	0.92	0.94	1.10
16	**0.91**	2.29	1.52	1.13	0.92	0.97	0.94
17	**1.07**	3.53	1.58	1.47	1.70	1.08	1.12
18	**3.58**	8.66	6.89	7.39	3.60	3.77	5.01
19	1.64	2.08	4.44	1.77	1.64	**1.60**	1.97
# of best	12	0	0	1	1	4	1
Average	**1.35**	3.30	2.28	1.99	1.39	1.39	1.95

TABLE 9: The VAPE of outflow at each station on weekends.

Station	Combined	ARMA	NN	SVR	Averaging	Adaboost	kNN
1	0.99	1.6	1.06	1.22	0.99	**0.94**	1.09
2	**1.00**	1.37	1.06	1.13	1.00	0.98	0.9
3	0.86	1.65	0.92	0.87	**0.85**	0.92	1.06
4	**0.99**	1.92	1.26	1.24	0.99	1.03	1.18
5	0.75	1.54	1.45	0.97	0.75	**0.74**	0.68
6	**0.69**	3.05	1.16	0.90	0.71	0.70	0.88
7	**0.81**	1.74	0.98	0.89	0.81	0.82	0.99
8	**0.87**	1.91	1.18	1.45	0.88	0.88	1.36
9	1.00	1.64	**0.86**	1.74	1.04	1.09	1.95
10	**0.79**	2.3	1.12	1.11	0.80	0.82	1.35
11	**0.84**	2.43	1.29	1.17	0.84	0.88	1.38
12	**0.72**	2.94	3.14	1.16	0.73	0.80	1.03
13	**0.77**	3.2	1.5	0.92	0.79	0.94	1.12
14	**0.83**	2.72	1.68	1.49	0.83	0.85	0.99
15	0.73	6.28	2.12	**0.72**	0.74	0.8	0.84
16	**0.94**	2.58	1.52	1.4	0.94	0.96	0.89
17	0.89	3.17	1.26	1.06	0.89	**0.88**	0.87
18	1.34	1.84	1.36	1.58	1.33	**1.33**	1.57
19	2.13	2.73	2.20	3.86	2.11	**2.07**	2.22
# of best	11	0	1	1	1	5	0
Average	**0.94**	2.45	1.49	1.31	0.95	0.97	1.18

TABLE 10: The MAPE of peak hour on weekdays.

	combine	Avg[1]	Ada[2]	kNN
morning peak outflow	11.78	11.96	12.34	11.69
evening peak outflow	11.02	11.15	11.16	11.28
daytime outflow	11.51	11.73	11.66	12.43
morning peak inflow	10.79	10.98	11.32	10.92
evening peak inflow	14.35	14.38	14.60	14.38
daytime inflow	13.26	13.47	13.44	13.40

[1] Averaging method.
[2] Adaboost method.

the travel pattern of passengers. Then a practical approach is proposed, in which probabilistic model combines two basic models, to improve metro passenger flow forecasting performance. The effective combination benefits from the intermodel diversities, mitigates the risks of using an isolated model, and compensates the drawbacks of the individual models. More importantly, the combination rules could be applied to most individual prediction models. Besides, our model predicts the inflow and outflow at each metro station in an unaggregated way by utilizing OD flows. That is, the OD flows are predicted in a spatiotemporal fashion and then aggregated to get the inflow and outflow at each station. Our assumption is that OD flows enjoy more predictable power for they essentially reflect travel behaviors. Overall, the experiment results indicate that the accuracy and stability of the proposed model outperform the baseline models, ARMA, neural networks, and support vector regression. And the whole framework proposed is generalizable to other complex transportation systems with origin-destination and therefore offers important insights to future research.

In the future research, we would like to improve the usage of spatiotemporal information for the flow prediction of more complex metro networks with a certain number of transfer stations. Another interesting direction is to expand the prediction problem to a multicriteria one. In the past research, assessments of the relative performance of various combinations have generally been made under an accuracy criterion, expressed in terms of MSE, MAE, or MAPE. However, the robustness should attract more attention. Therefore, the variance should also be considered as one of the criteria when developing models. In the end, to build a method to predict the short-term travel demand under disruptions is also in our future plan.

Data Availability

The AFC dataset is not open for public access currently.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.
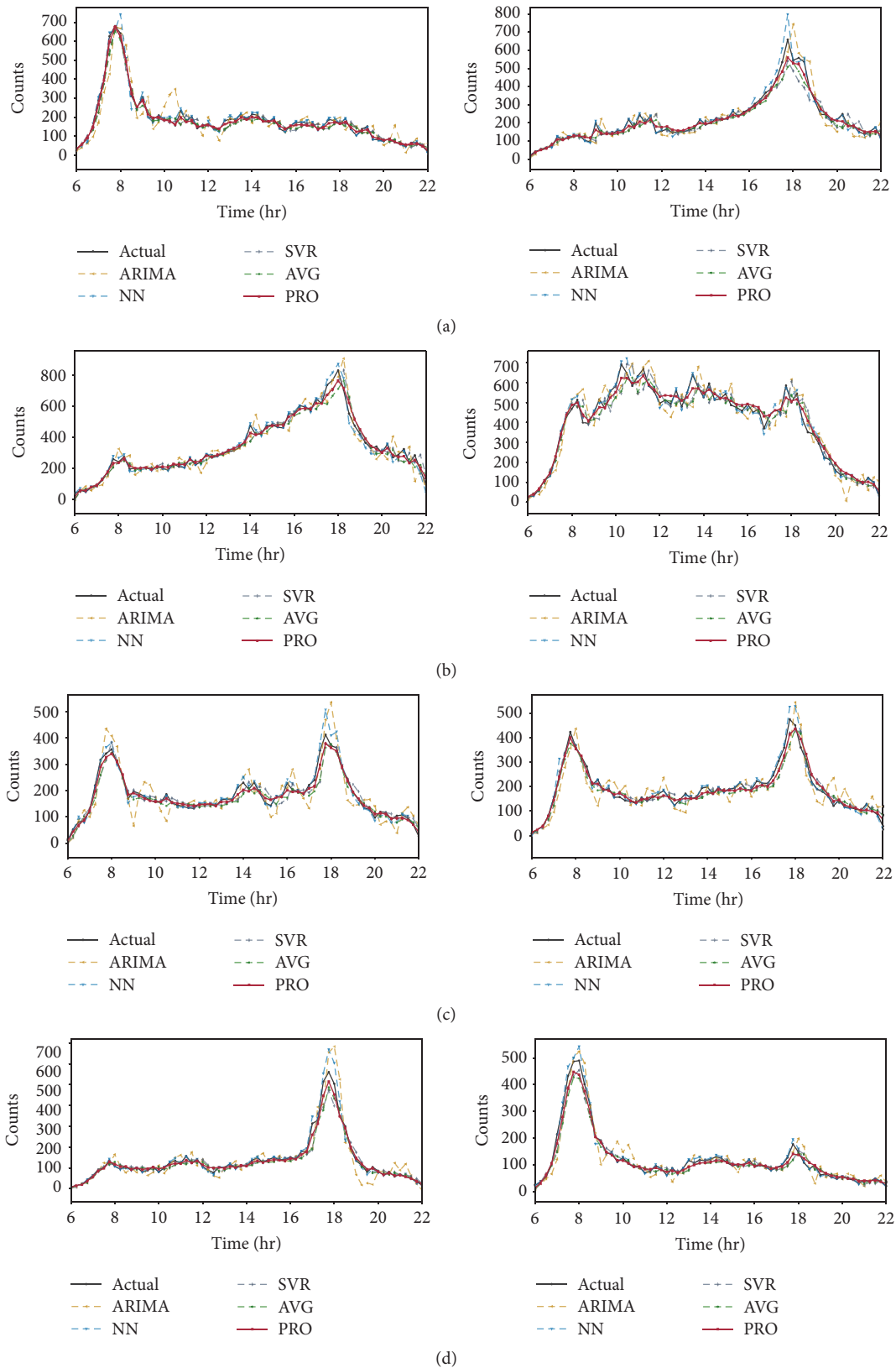
Acknowledgments

FIGURE 5: The comparison of prediction results with referenced models at selected stations during the operation time of one weekday. "AVG" stands for averaging model and "PRO" stands for probabilistic combination. (a), (b), (c), and (d) are the results of Stations 4, 9, 12, and 16, respectively. They are located in different zones. The left column shows the actual and predicted counts of inflow. The right column shows the counts of outflow.
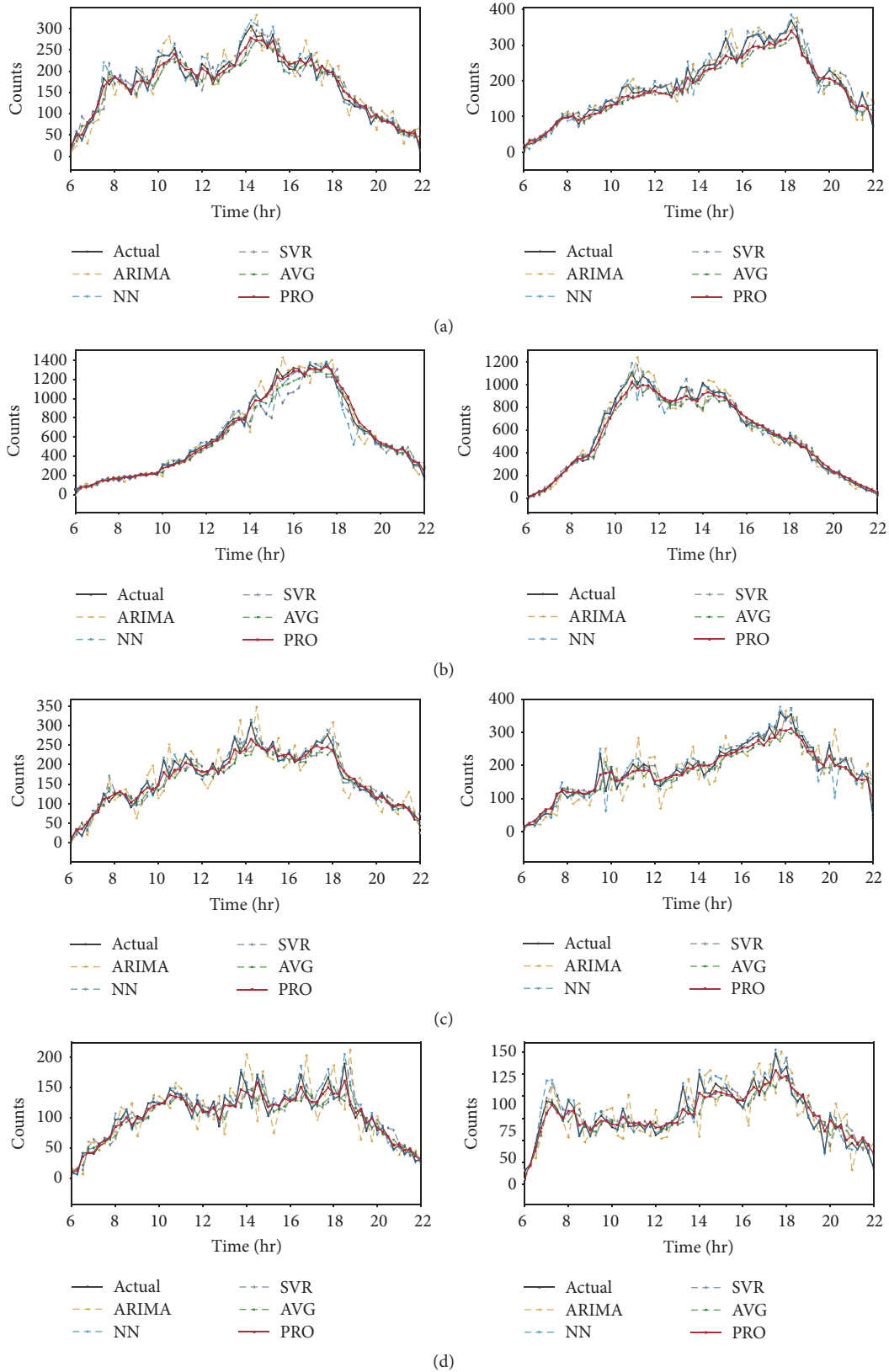
FIGURE 6: The comparison of prediction results with referenced models at selected stations during the operation time of one weekend. "AVG" stands for averaging model and "PRO" stands for probabilistic combination. (a), (b), (c), and (d) are the results of Stations 4, 9, 12, and 16, respectively. They are located in different zones. The left column shows the actual and predicted counts of inflow. The right column shows the counts of outflow.

References

[1] D. Zhang and X. C. Wang, "Transit ridership estimation with network Kriging: A case study of Second Avenue Subway, NYC," *Journal of Transport Geography*, vol. 41, pp. 107–115, 2014.

[2] Y. Wei and M. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 148–162, 2012.

[3] Y. Xu, Q.-J. Kong, R. Klette, and Y. Liu, "Accurate and interpretable bayesian MARS for traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2457–2469, 2014.

[4] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: where we are and where we're going," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 3–19, 2014.

[5] E. Vlahogianni and M. Karlaftis, "Testing and comparing neural network and statistical approaches for predicting transportation time series," *Transportation Research Record*, no. 2399, pp. 9–22, 2013.

[6] Y. Zhang and Y. Xie, "Forecasting of short-term freeway volume with v-support vector machines," *Transportation Research Record*, vol. 2024, no. 1, pp. 92–99, 2007.

[7] Y. Xie, K. Zhao, Y. Sun, and D. Chen, "Gaussian processes for short-term traffic volume forecasting," *Transportation Research Record*, no. 2165, pp. 69–78, 2010.

[8] T.-H. Tsai, C.-K. Lee, and C.-H. Wei, "Neural network based temporal feature models for short-term railway passenger demand forecasting," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3728–3736, 2009.

[9] Y. Sun, B. Leng, and W. Guan, "A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system," *Neurocomputing*, vol. 166, pp. 109–121, 2015.

[10] Y. Li, X. Wang, S. Sun, X. Ma, and G. Lu, "Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 306–328, 2017.

[11] R. Silva, S. M. Kang, and E. M. Airoldi, "Predicting traffic volumes and estimating the effects of shocks in massive transportation systems," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 112, no. 18, pp. 5643–5648, 2015.

[12] L. M. De Menezes, D. W. Bunn, and J. W. Taylor, "Review of guidelines for the use of combined forecasts," *European Journal of Operational Research*, vol. 120, no. 1, pp. 190–204, 2000.

[13] W. Zheng, D. H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: bayesian combined neural network approach," *Journal of Transportation Engineering*, vol. 132, no. 2, pp. 114–121, 2006.

[14] M.-C. Tan, S. C. Wong, J.-M. Xu, Z.-R. Guan, and P. Zhang, "An aggregation approach to short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 60–69, 2009.

[15] A. Stathopoulos, M. G. Karlaftis, and L. Dimitriou, "Fuzzy rule-based system approach to combining traffic count forecasts," *Transportation Research Record*, no. 2183, pp. 120–128, 2010.

[16] H. Qiu, R. Li, and H. Liu, "Integrated model for traffic flow forecasting under rainy conditions," *Journal of Advanced Transportation*, vol. 50, no. 8, pp. 1754–1769, 2016.

[17] R. Adhikari, "A neural network based linear ensemble framework for time series forecasting," *Neurocomputing*, vol. 157, pp. 231–242, 2015.

[18] V. R. R. Jose and R. L. Winkler, "Simple robust averages of forecasts: Some empirical results," *International Journal of Forecasting*, vol. 24, no. 1, pp. 163–169, 2008.

[19] S. Jiang, W. Guan, W. Zhang, X. Chen, and L. Yang, "Human mobility in space from three modes of public transportation," *Physica A: Statistical Mechanics and its Applications*, vol. 483, pp. 227–238, 2017.

[20] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[21] M. Collins, R. E. Schapire, and Y. Singer, "Logistic regression, adaBoost and Bregman distances," *Machine Learning*, vol. 48, no. 1–3, pp. 253–285, 2002.

[22] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[23] Hjort and N. Lid, *Model Selection and Model Averaging*, Cambridge University Press, 2008.