Research Article

Demand Analysis and Distribution of Single-Trip Ticket Cards for Urban Rail Transit

Zi-jia Wang [1,2] **Jing-qi Li** [1] **Jiang-yue Wu** [3] **and Zhi-gang Yang**[4]

[1]*Department of Civil Engineering, Beijing Jiaotong University, Beijing 100044, China*
[2]*Beijing Engineering and Technology Research Center of Rail Transit Line Safety and Disaster Prevention, No. 3 Shangyuancun, Haidian District, Beijing 100044, China*
[3]*Department of Urban Planning and Design, The University of Hong Kong, Hong Kong 999077, China*
[4]*The Traffic Planning Department, Beijing Municipal Institute of City Planning & Design (BICP), No. 60 Nanlishi Street, Beijing 100045, China*

Correspondence should be addressed to Jiang-yue Wu; wujiangyue@hotmail.com

In the current urban rail transit systems, nearly 15% of passengers are noncommuter travelers who use single-trip ticket cards (ticket cards). Accordingly, the effective management of ticket cards is of great importance. This article suggests a time series model for use in predicting ticket card storage based on the characteristics of ticket cards collected by an automatic fare collection (AFC) system. The distribution cycle, station types, and distribution volume of each station are also determined. Then, drawing on small package transportation feasibility theory, an unbalanced distribution model between production and demand (unbalanced distribution model), as well as a hybrid distribution model of loading and unloading (hybrid distribution model), is established. Application of these models to the Beijing Subway system is used to verify the efficiency and feasibility of such a hybrid distribution model. The analysis and results offer insights into usage patterns of urban rail transit ticket cards, providing solid evidence for a relative decision-making process.

1. Introduction

One notable barrier to effective management of ticket cards is the geographical imbalance of ticket card storage among subway system's stations. After ticket cards are sold by an AFC system, they are collected by an automatic gate machine (AGM) as passengers exit the station. However, uncertain passenger flows inevitably result in geographical imbalances of ticket cards. Furthermore, ticket storage capacity management in each station is based on the assumption that "more is better." Flexible distribution of ticket cards based on the specific circumstances of each station has thus far rarely been investigated. However, the present approach to managing ticket cards is not sufficient to effectively respond to growing traffic demand. In response to these issues, this paper proposes a set of positive management methods for forecasting and dynamically distributing ticket cards among stations so as to reduce total storage demand for ticket cards, improve ticket card turnover rate, and ensure better travel experiences for passengers.

Precise short-term predictions of passenger flows are the basis for determining distribution cycle, distribution volume, and other basic statistics of ticket cards. Many existing studies have explored short-term passenger flow prediction. After Box and Jenkins [1] proposed a systematic method for identifying, analyzing, and predicting time series, a number of other researchers improved their time series model [2, 3], applying it more broadly [4]. In recent years, the emergence of big data has done much to improve the stability and applicability of the time series model. As the data collected become complicated, many data mining methods are used as a complement operator to the classic ARIMA model, processing unstable smart-card data. The hybrid model is also used to predict short-term passenger flows. Common

applications include the ARIMA-Kalman filter model [5], ARIMA-ANN model [6], ARIMA-SVM model [7], ARIMA-GA model [8], and EMD-BPN [9] model. Thus, the time series model has been proved to be an effective approach with which to predict short-term passenger flows in an urban transit system. This paper uses the classical time series model for prediction, based on AFC records, to establish a storage prediction model.

Second, although ticket card distribution is seldom addressed in the literature, distribution itself has been studied extensively, especially in relation to surface transportation for time and cost savings. Many logistics models have been introduced to solve problems relating to delivery of goods. However, little research has been conducted on dynamic distribution in a subway setting primarily because single-trip ticket cards are rarely recycled and because the potential cost savings associated with distribution models have been underestimated by decision makers and subway companies alike. To tackle the distribution problem in a subway environment, a logistics model is needed that reflects specific constraints corresponding to the actual situation of each station. This paper mainly introduces two distribution models, the unbalanced distribution model and the hybrid distribution model, which are based on two logistic models: the unbalanced production-marketing problem and the vehicle routing problem with simultaneous pickup and delivery (VRPSDP). The hybrid distribution model offers more cost savings than the unbalanced distribution model by considering both "pickup" and "delivery" demands simultaneously, thus saving time and manpower. Within the context of card collection and distribution, "pickup" refers to the collection of surplus cards at a station and "delivery" to the transporting of surplus cards to stations having a card deficit. Essentially, it can be considered an extension of VRPSDP.

Clarke and Wright [10] proposed the traveling salesman problem (TSP), a special case of the VRPSDP. Min [11] first abstracted the VRPSDP model from the real-world problem of delivering books among libraries and solved it using the genetic algorithm. Since then, the VRPSDP model has been gradually improved to apply to different practices, including by adding time windows [12], allowing multiple vehicles [13], and reservations for repeated service [14]. Beyond improvements to constraints, its objective is becoming increasingly generalized from just distance [14] and time [12] to generalized cost savings, such as through manpower arrangements for accomplishing distribution [13]. At present, methods of solving the VRPSDP include mainly stereotype algorithms [15] and heuristic algorithms. Instead of providing the best solution, heuristic algorithms are more adaptive to big data demands—for example, tabu research [16], the tourism partition heuristic algorithm [17], the genetic algorithm [18], simulated annealing [19], and ant colony systems [20]. To improve algorithmic efficiency and avoid single method limitations, a hybrid algorithm is widely accepted as tabu research with guided local research [21], ant colony systems with a tabu research operator [22], and a combination of the metaheuristic algorithm [23], and the like, and has become a focus of mainstream research. The

genetic algorithm is particularly powerful when used in the VRPSDP logistics model. It has been used to tackle certain types of the vehicle routing problem in recent decades, often in combination with other algorithms to improve its effectiveness. However, existing studies have mostly concentrated on applications to road traffic problems, with few having focused on distribution balance. This paper proposes a distribution model based on the urban rail transit network and the characteristics of ticket card distribution.

This paper is organized as follows: Section 2 proposes a storage prediction model based on a time series model. Section 3 develops a reasonable distribution model on the basis of the forecast data. Section 4 offers a feasible distribution plan using data for the Beijing Subway from December 2016.

2. Storage Prediction Model of Single-Trip Ticket Cards

A time series model is used to predict ticket card data. For this step, ticket card flow volume, distribution volume, and storage settings are considered, which are defined in the current management standards [24] as follows:

(1) Flow volume:

$$F = S - R, \tag{1}$$

where F is the flow volume of ticket cards, S is the total volume of ticket cards sold, and R is the recycled number of ticket cards.

(2) Distribution volume:

The distribution volume of ticket cards is the number of ticket cards to be deployed from stations that have a ticket card surplus to those that have a deficit.

(3) Storage capability settings:

$$C_{\max} = a * \overline{S} * 120\%, \tag{2}$$

$$C_{s} = a * \overline{S}, \tag{3}$$

$$C_{\min} = a * \overline{S} * 80\%, \tag{4}$$

where $C_{\max}$ is the maximum storage capability, C_{s} is the safe storage capability, $C_{\min}$ is the minimum storage capability, a is a coefficient that equals 5 for public holidays but equals 4, otherwise, and $\overline{S}$ is the average daily sales volume of the last distribution cycle.

2.1. ARIMA Model. ARIMA(p, d, q) is the most general form of the classical time series model. The classical ARIMA model can be applied only to stable time series, so the test ADF is used to judge whether a time series is stable. Then, parameters can be estimated, and the t-test and q-test are used to ensure that the ARIMA model is error-tolerant compared with the actual times series. Several possible ARIMA models may be proposed in the foregoing steps, after which the AIC test [25] is used to select the best ARIMA

model. Using the inverse difference operator, final prediction results are acquired. The general process of establishing the ARIMA (p, d, q) model [26–28] is shown in Figure 1.

According to the general process of the ARIMA (p, d, q) model, this paper uses STATA [29] to make a short-term prediction concerning the sales volume and flow volume of each station. Data cover the period from December 19, 2016, to December 25, 2016.

2.2. Prediction Model.

Depending on ticket card characteristics, the minimum storage capacity of each station should satisfy the sales volume of the first day in this distribution cycle as well as in the next. The associated mathematical expressions are given in equations (5) and (6):

$$C_{\min} > S_1, \tag{5}$$

$$C_{\min} - \sum_{i=1}^{N} F_i > S_{N+1}, \tag{6}$$

where S_i is the sales volume or the prediction of sales volume, $i = 1, 2, \ldots, N + 1$, F_i is the flow volume or prediction of flow volume, $i = 1, 2, \ldots, N$, and N is the prediction cycle.

The ARIMA model is applied to each station to estimate its sales volume for the first day in the next distribution cycle. Using equations (5) and (6), final minimum storage capability is obtained. The number of ticket cards in one ticket box is added to the final minimum storage capability of 2,000 pieces to address uncertainty about passenger flow. Under the current regulations (equations (2) and (4)), the relative ratio between the maximum storage capability and the minimum storage capability is fixed, so the maximum storage capability can be determined.

3. Balanced Distribution Model of Single-Trip Ticket Cards

The subway itself is believed to be an effective carrier for distributing ticket cards. A large number of existing studies have demonstrated the rationality and feasibility of using the subway to transfer goods during nonpeak hours. Liu [30] and Shen [31] studied the feasibility of using the Beijing Subway system as a nighttime logistics circulation system in urban areas. Some studies [32, 33] have also explored the feasibility of using the Tyne rail transit for cargo transport in the United Kingdom. The benefits of using rail transit as a carrier include improving the utilization rate of the subway in low-peak hours and avoidance of technical problems such as alterations to subway coaches and subway tracks. Ticket cards are smaller and easier to transport than other heavy goods, confirming the feasibility of distributing ticket cards using the subway system itself.

In distribution models, station-to-station distribution is often applied to actual situations, especially distribution between an AFC clearing center (ACC) and subway stations. Furthermore, in practice, distribution follows a "more is better" approach, so subway companies generally neglect full
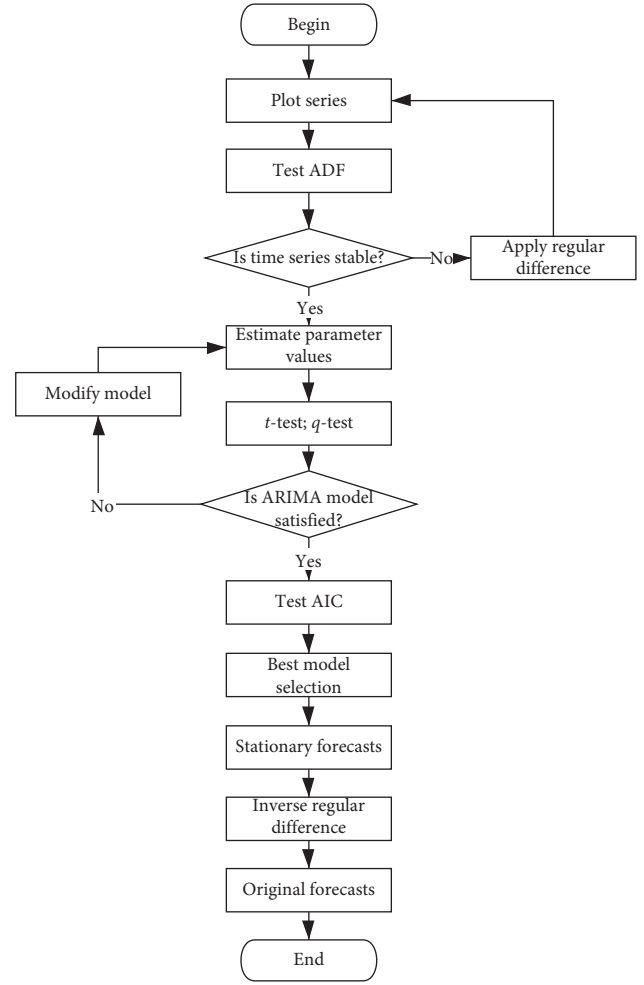

Figure 1: Classical time series model flowchart.

utilization of ticket cards and manpower. An unbalanced distribution model is established as a simulation of the actual distribution, and a hybrid distribution model is proposed as an effective alternative. In this part, some necessary parameters and concepts used in distribution models are defined, after which two distribution models are introduced. Next, the reasons for proposing these two models, with a particular focus on total distance and management cost, are illustrated in detail.

3.1. Distribution Volume.

Assuming that the current storage of a station is c, the number of ticket cards to be allocated is

$$\Delta c = C_s - c, \tag{7}$$

where Δc is the number of ticket cards to be distributed, C_s is the safe storage capability, and c is the current storage.

For a station x,

$$\begin{cases} \Delta c > 0, & \text{station } x \text{ is a loss station,} \\ \Delta c < 0, & \text{station } x \text{ is a surplus station.} \end{cases} \tag{8}$$

Whether the station participates in the distribution depends on distribution volume and actual sales volume.

Different types of stations should have different standards. For a loss station i,

$$\begin{cases} |\Delta c| > S_l, & \text{station } i \text{ participates in the deployment,} \\ |\Delta c| < S_l, & \text{otherwise,} \end{cases} \tag{9}$$

where S_l is the loss number threshold of ticket cards.

Similarly, for a surplus station j,

$$\begin{cases} |\Delta c| > S_s, & \text{station } j \text{ participates in the deployment,} \\ |\Delta c| < S_s, & \text{otherwise,} \end{cases} \tag{10}$$

where S_s is the surplus number threshold of ticket cards.

3.2. Distribution Model Unbalanced between Production and Demand.

In practice, total distribution between surplus stations and loss stations is not equal, according to Beijing AFC data. The ACC conducts the distribution of the subway system and provides or recycles ticket cards. Suppose that the total surplus is greater than the total deficit—specifically that m stations have a surplus of ticket cards and n stations a deficit. Because more ticket cards need to be reused, ACC is taken as a station with a deficit of ticket cards so that the total number of loss stations is $n + 1$. Based on the unbalanced logistics model, some constraints in the context of the subway system are added, with the ultimate mathematical expressions of the unbalanced distribution model given in equations (11) to (14):

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n+1} D_{ij} x_{ij}, \tag{11}$$

$$\text{s. t} \sum_{j=1}^{n+1} x_{ij} = s_i, \quad i = 1, 2, \ldots, m, \tag{12}$$

$$\sum_{i=1}^{m} x_{ij} = l_j, \quad j = 1, 2, \ldots, n+1, \tag{13}$$

$$\sum_{i=1}^{m} s_i = \sum_{j=1}^{n+1} l_j, \tag{14}$$

where x_{ij} is the distribution volume between surplus station i and loss station j ($x_{ij} \geq 0$), D_{ij} is the distance between surplus station i and loss station j, s_i is the total distribution volume of surplus station i ($s_i > 0$), l_j is the total distribution volume of loss station j ($l_j > 0$), m is the number of surplus stations, and n is the number of loss stations.

3.3. Hybrid Distribution Model of Loading and Unloading.

According to the logistical model that is the basis of the unbalanced distribution model, the foregoing distribution model determines optimal station-to-station distributions. According to the Ticket Distribution Management Regulations [34], at least one ticket management crew and one security guard are required for each station-to-station distribution. Accordingly, this method will cause significant wastage of human resources. Furthermore, distribution itself is arranged at night or other nonpeak hours when in response to reduced workload, fewer staff are present. As a result, human resources should be more fully utilized. Another disadvantage of unbalanced distribution is that no distribution tasks are involved in the backhaul of station-to-station distribution, meaning that the objective value, the total distance, can still be reduced by a change in distribution methods.

Based on the two preceding reasons, the VRPSDP is used to establish a hybrid distribution model. Assuming that n stations participate in distribution, the total surplus exceeds the total loss. ACC is equivalent to a loss station, and its station number is 0. Thus, the hybrid distribution model can be expressed as

$$\min \sum_{i=0}^{n} \sum_{j=0}^{n} D_{ij} x_{ij}, \tag{15}$$

$$\text{s.t} \sum_{i=0}^{n} x_{ij} = 1, \quad i \neq j, \tag{16}$$

$$\sum_{j=0}^{n} x_{ij} = 1, \quad i \neq j, \tag{17}$$

$$z_{ij} \geq 0, \quad i \neq j, \tag{18}$$

$$\sum_{j=1}^{n} z_{0j} = \sum_{i=1}^{n} (l_i - s_i), \tag{19}$$

$$\text{or} \sum_{j=1}^{n} z_{0j} = \sum_{i=1}^{n} (s_i - l_i), \tag{20}$$

$$\sum_{i=0}^{n} z_{ij} + (s_i - l_i) \sum_{i=0}^{n} x_{ij} = \sum_{i=0}^{n} z_{ji+1}, \tag{21}$$

$$\begin{aligned} l_i > 0, \\ \text{or } s_i > 0, \end{aligned} \tag{22}$$

$$x_{ij} = \begin{cases} 1, & \text{there is deployment between station } i \text{ and station } j \\ 0, & \text{otherwise} \end{cases} \quad (i \neq j), \tag{23}$$

where D_{ij} is the distance between station i and station j, $i \neq j$, x_{ij} is a parameter, $i \neq j$, z_{ij} is the number of ticket cards involved in distribution from station i to station j, l_i is the number of ticket cards to bring to station i, where if $l_i > 0$, then station i is a loss station, and s_i is the number of ticket cards to bring out for station i, where if $s_i > 0$, then station i is a surplus station.

Objective function (15) minimizes the total distance. Constraints (16) and (17) mean that each station can be served only once; constraint (18) ensures that the number of ticket cards during the distribution is positive; and

TABLE 1: Prediction data of Wangfujing Station.

Date	Dec. 23, 2016	Dec. 24, 2016	Dec. 25, 2016	Dec. 26, 2016
Flow volume(pieces)	1,242	1,618	1,767	
Sales volume (pieces)	9,593	14,421	12,690	14,224

constraint (19) indicates that when starting from ACC, the number of ticket cards equals the sum of the distribution number in all loss stations minus the sum of the distribution number in all surplus stations. Constraint (20) represents the opposite situation. Only one of these two constraints can be met in the actual circumstances. Constraint (21) ensures that the demand of ticket cards for each station could be met; constraint (22) requires that the distribution volume be positive.

4. Case Study

To better investigate the performance of the preceding models, the Beijing Subway is taken as a real-world example. In the Beijing Subway system, single-trip ticket cards account for 15% of all total ticket cards in the AFC system. The data used in the case study cover the period from December 23, 2016, to December 26, 2016. We took the initial number of ticket cards for all stations to be baseline of safe storage capability on the beginning of December 23. Section 4.1 analyzes the advantages and feasibility of the storage prediction model and then determines the distribution cycle of ticket cards; Section 4.2 solves, compares, and analyzes the two distribution models in a MATLAB environment.

4.1. Distribution Cycle. Taking Wangfujing Station as an example, we used the prediction model to forecast the associated statistics, given in Table 1.

Wangfujing Station is a loss station, with an average sales volume of 12,732 pieces. Under the current regulations, the storage for the next distribution cycle should be 40,743–61,114 pieces. However, a storage predicting model predicts that storage will be 20,851–31,277 pieces or 49% less.

Thus, the storage prediction model can significantly reduce total storage of ticket cards while improving ticket card circulation frequency and utilization rate. However, because the model has some limitations that prevent it from guaranteeing an opportune response to unexpected fluctuations in passenger flow, storage settings should be raised commensurately in actual application. Even with such adjustments, the Wangfujing example indicates that the storage predicting model can still save ticket card resources and attendant costs of management.

Determination of the distribution cycle is intended to stabilize the number of stations involved and the total distribution volume of ticket cards within a reasonable range. According to relevant regulations [24] and studies [35] of the Beijing Subway, S_l is set at 2,000 and S_s at 1,000. The number of stations to be deployed after day's end is shown in Figures 2 and 3.

Figure 2 shows a steep increase on the fifth day in both number of loss stations and number of surplus stations,
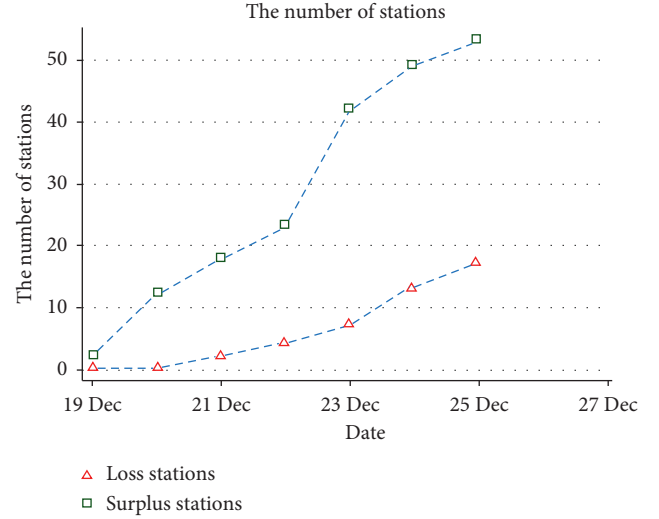
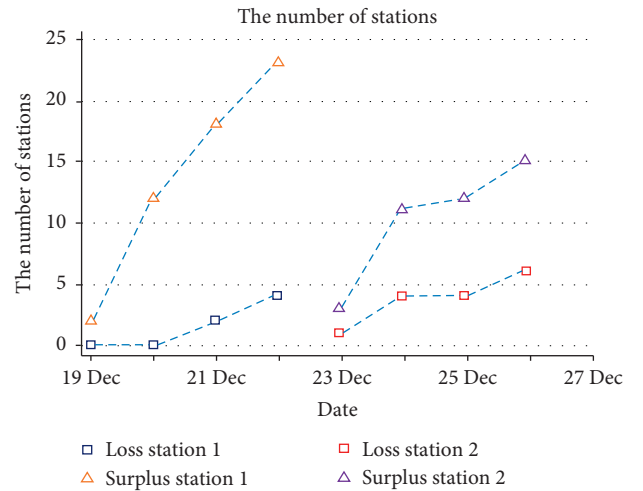

FIGURE 2: The number of stations to be deployed 1.

FIGURE 3: The number of stations to be deployed 2.

indicating that most stations need distribution after 4 days' operation. If setting the distribution cycle to a week, total distribution volume will increase to 55,129, and the total number of stations to be deployed to about 70, or fourth of total stations—a difficult task to accomplish in a single distribution. Accordingly, 4 days is a more suitable threshold than a week. For a 4-day distribution cycle, as Figure 3 shows, the total number of stations that needs to participate in distribution is approximately 25, of which about 5 are loss stations. Not only the number of distributed stations but also total distribution volume varies within an acceptable range, with 19,780 ticket cards needing deployment in the first 4 days and 21,249 in the next 4. Accordingly, the distribution
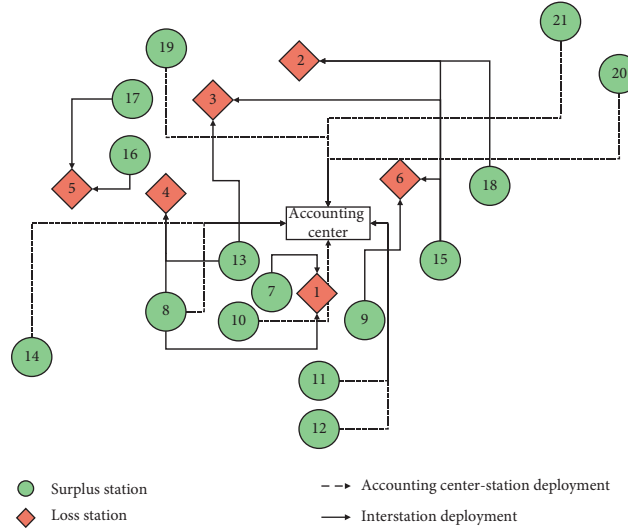
Figure 4: Schematic of the unbalanced distribution model.

Table 2: Partial simulation results.

Number	Total iterations	Shortest path	Number of best iterations	Total distance (km)
1	100	0–14–8–1–7–10–12–11–9–6–18–3–2–0; 0–21–20–15–13–4–5–16–17–0; 0–19–0	70	215.36
2	150	0–21–20–15–6–10–9–13–4–5–16–17–0; 0–14–8–1–7–12–11–18–3–2–0; 0–19–0	66	235.75
3	150	0–8–7–9–10–12–11–13–4–6–18–3–2–0; 0–21–20–15–1–14–5–16–17–0; 0–19–0	106	227.4
4	200	0–14–8–11–12–10–7–1–9–6–18–2–0; 0–21–20–15–13–4–3–5–16–17–0; 0–19–0	132	220.05
5	200	0–14–8–7–1–9–11–12–10–6–18–2–0; 0–21–20–15–13–4–3–5–16–17–0; 0–19–0	121	218.88
6	200	0–21–20–15–6–18–3–2–0; 0–14–8–1–11–12–9–10–7–13–4–5–16–17–0; 0–19–0	92	215.35

cycle for nonholidays is determined to be 4 days. Because holidays in China generally last 3 days, the distribution cycle for holidays is set to 3 days.

4.2. Distribution Plan. According to the foregoing analysis, 21 stations will need distribution on December 26, 2016: 6 loss stations and 15 surplus stations. ACC participates in the distribution as well.

In two distribution models, the Dijkstra algorithm [36] is used to solve for key parameter D_{ij}, the shortest distance between two stations. ACC is a certain location at the center of all stations, so D_{0j} or D_{i0} is simplified as the mean of all distances between ACC and each station: 17 km. Results show that total distance in the unbalanced distribution model is 550.5 km, with 21 station-to-station distributions. Figure 4 shows a schematic of the distribution.

A few points should be explained concerning the hybrid distribution model:

(1) When calculating fitness, two constraints were added to improve distribution reasonableness. The first is intended to limit the degree of dispersion among stations. If the cumulative distance reaches 100 km (a value determined through multiple experiments), the crew should return to the ACC and restart distribution with a new branch line. The second is that the remaining number of ticket cards must meet the demand of the next station in the branch line after serving one station. If it does not, the crew should return to the ACC and restart with a new branch line.

(2) The population size is set at 400 and the number of iterations at 200 (values determined through
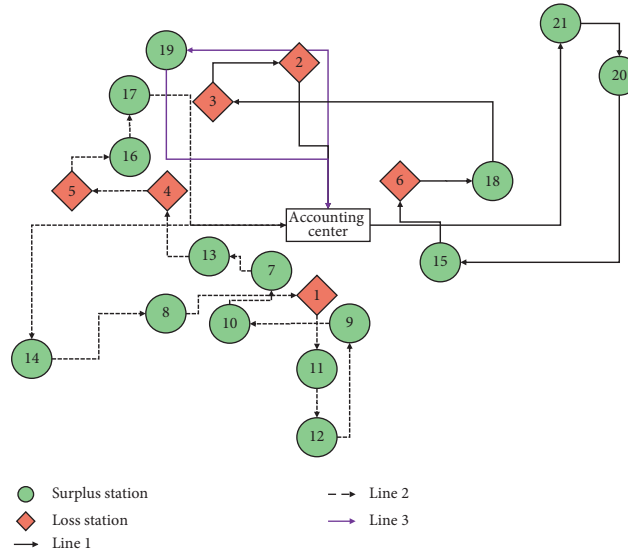
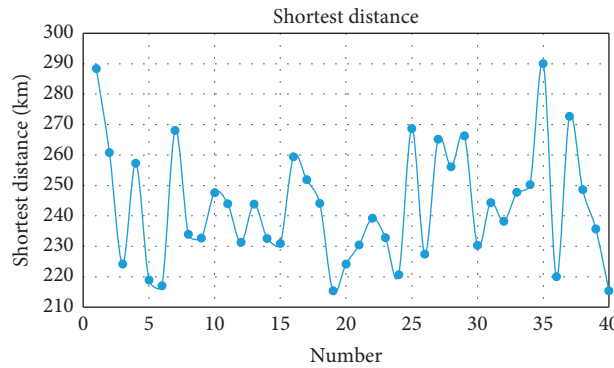FIGURE 5: Schematic of the hybrid distribution model.

FIGURE 6: Schematic of the optimal experimental convergence result.

multiple experiments), with a probability of chromosome crossing PC (term of genetic algorithm) of 0.85 and a probability of variation PM of 0.01.

To obtain better optimization results, several simulation experiments were carried out. Table 2 shows the results of partial experiments.

The best experimental result is for 215.35 km, with three branch lines. Figure 5 shows a schematic of distribution.

Figure 6 shows the convergence result of the optimal experiment.

Clearly, the final convergence result is inconsistent with the optimal solution, which appears in the early stage. Considering the change of path, it can be reasonably inferred that the final convergence result is generated from the optimal solution, which is itself generated by the cross-over operation—indicating that room remains for improvement of the algorithm.

To solve this problem, a simulated annealing operator [37] is added to enhance the global search ability and prevent the optimal chromosome from occupying the entire population too quickly. We set the initial temperature t_0 to 90 and the attenuation coefficient α to 0.95. Because simulated annealing reduces convergence speed, the number of iterations is doubled to 400. Table 3 gives partial experimental results.

Figure 7 shows a significantly better convergence result than can be achieved using a single algorithm, albeit a slight difference in the minimum distance. In addition, the total number of lines is not limited to three, indicating that the diversity of itineraries is also increased. These results thus demonstrate the greater strength of the algorithm's global search ability.

In conclusion, the hybrid distribution model significantly outperforms the unbalanced distribution model. Not only is the total distance clearly reduced but distribution efficiency is also greatly increased. In addition, the unbalanced distribution plan involves 21 station-to-station distributions, requiring at least 21 groups of staff, whereas in the hybrid distribution model, three groups of personnel can be assigned to the independent lines. Freed from any time limit, one group of personnel would be sufficient, greatly improving the utilization rate of human resources. The genetic

TABLE 3: Partial simulation results.

| Number	Total iterations	Shortest path	Number of best iterations	Total distance (km)
1	400	0–19–17–16–5–6–9–10–11–12–1–7–0; 0–21–20–15–18–2–3–13–4–8–0; 0–14–0	304	250.35
2	400	0–20–21–18–6–5–16–3–2–0; 0–13–4–11–12–10–7–1–9–15–8–14–17–0; 0–19–0	321	246.05
3	400	0–21–20–4–13–12–11–5–16–17–0; 0–14–10–7–8–1–9–15–6–18–3–2–0; 0–19–0	292	232.7
4	400	0–19–17–16–5–14–0; 0–21–0; 0–4–0; 0–13–12–11–10–7–8–1–9–15–6–18–3–2–0; 0–20–0	383	235.7
5	400	0–10–9–11–12–6–15–8–7–1–13–4–5–0; 0–19–17–16–3–18–2–0;0–21–20–0; 0–14–0	311	253.18
6	400	0–21–20–15–6–18–3–2–0; 0–14–8–1–7–10–9–11–12–13–4–5–16–17–0; 0–19–0	361	213.5
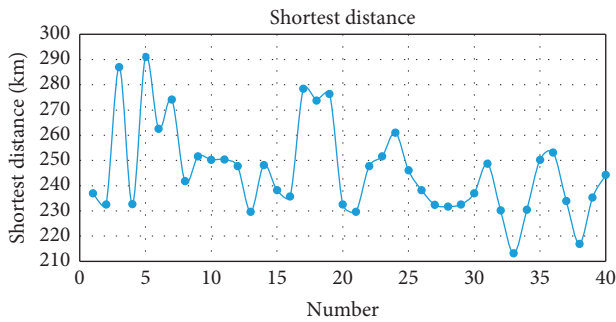
FIGURE 7: Schematic of the optimal experimental convergence result (with the simulated annealing operator).

algorithm used in the hybrid distribution model is highly adaptable to various practical problems. Accordingly, after a comprehensive comparison, the hybrid distribution model offers better simulation effect, much room for improvement, and a more convincing practical application.

5. Conclusions

This paper, being based on AFC data, mainly studies forecast and distribution problems related to ticket cards. The current distribution follows the principle of "more is better," having the problem of low recycled rate and ineffective manpower arrangements. To propose a feasible and efficient distribution model, a forecast model of sales volume and a determination of recycled threshold are demanded. With the prediction of loss stations, surplus stations, and the total distribution volume of ticket cards, a hybrid distribution model is introduced and compared with the current distribution model. Its primary conclusions are as follows:

(1) A quantitative method extended from times series prediction for determination of storage settings is established that can improve empirical determinations. Based on a case study, the storage predicting model can reduce total storage 40%, on average, from that seen under the current regulations.

(2) By employing the prediction model, distribution value, loss stations, and surplus stations are determined. The distribution cycle is set as 4 days for nonholidays and 3 days, otherwise, on the basis of ticket card data.

(3) A new distribution model is established using the theory of reverse logistics. Experimental results show that searching behavior can be enriched, with convergence and diversity stressed and well balanced with multiple operators. Compared with the current distribution method, simulations demonstrate the effectiveness and feasibility of the proposed distribution method.

(4) However, the model does have certain limitations. First, its lack of a buffer against emergencies is an inevitable shortcoming: in practice, such a buffer should be provided. Second, determination of the distribution cycle was made based on short-term data; further comprehensive analysis based on long-term AFC data is expected in the future. Third, a quantitative expression of manpower can be included in the objective of the distribution model in the future research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco, CA, USA, 2015.

[2] R. F. Engle and C. W. J. Granger, "Co-integration and error correction: representation, estimation, and testing, Econometrica," *Journal of the Econometric Society*, vol. 55, no. 2, pp. 251–276, 1987.

[3] Van Der, M. Voort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transportation Research Part C: Emerging Technologies*, vol. 4, no. 5, pp. 307–318, 1996.

[4] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003.

[5] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 50–64, 2014.

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

[7] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6164–6173, 2009.

[8] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 211–234, 2005.

[9] Y. Wei and M.-C. 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.

[10] G. Clarke and J. W. Wright, "Scheduling of vehicles from a central depot to a number of delivery points," *Operations Research*, vol. 12, no. 4, pp. 568–581, 1964.

[11] H. Min, "The multiple vehicle routing problem with simultaneous delivery and pick-up points," *Transportation Research Part A: General*, vol. 23, no. 5, pp. 377–386, 1989.

[12] R. Bent and P. Van Hentenryck, "A two-stage hybrid local search for the vehicle routing problem with time windows," *Transportation Science*, vol. 38, no. 4, pp. 515–530, 2004.

[13] J. Wang, Y. Zhou, Y. Wang, J. Zhang, C. L. P. Chen, and Z. Zheng, "Multiobjective vehicle routing problems with simultaneous delivery and pickup and time windows: formulation, instances, and algorithms," *IEEE Transactions on Cybernetics*, vol. 46, no. 3, pp. 582–594, 2016.

[14] S. Mitra, "A parallel clustering technique for the vehicle routing problem with split deliveries and pickups," *Journal of the Operational Research Society*, vol. 59, no. 11, pp. 1532–1546, 2008.

[15] J. Dethloff, "Vehicle routing and reverse logistics: the vehicle routing problem with simultaneous delivery and pick-up," *OR-Spektrum*, vol. 23, no. 1, pp. 79–96, 2001.

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

[17] G. Mosheiov, "Vehicle routing with pick-up and delivery: tour-partitioning heuristics," *Computers & Industrial Engineering*, vol. 34, no. 3, pp. 669–684, 1998.

[18] G. Jeon, H. R. Leep, and J. Y. Shim, "A vehicle routing problem solved by using a hybrid genetic algorithm," *Computers & Industrial Engineering*, vol. 53, no. 4, pp. 680–692, 2007.

[19] G. Martinovic, I. Aleksi, and A. Baumgartner, "Single-commodity vehicle routing problem with pickup and delivery service," *Mathematical Problems in Engineering*, vol. 2008, Article ID 697981, 17 pages, 2008.

[20] Y. Gajpal and P. Abad, "An ant colony system (ACS) for vehicle routing problem with simultaneous delivery and pickup," *Computers & Operations Research*, vol. 36, no. 12, pp. 3215–3223, 2009.

[21] J.-F. Chen and T.-H. Wu, "Vehicle routing problem with simultaneous deliveries and pickups," *Journal of the Operational Research Society*, vol. 57, no. 5, pp. 579–587, 2006.

[22] E. E. Zachariadis, C. D. Tarantilis, and C. T. Kiranoudis, "A hybrid metaheuristic algorithm for the vehicle routing problem with simultaneous delivery and pick-up service," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1070–1081, 2009.

[23] A. S. Tasan and M. Gen, "A genetic algorithm based approach to vehicle routing problem with simultaneous pick-up and deliveries," *Computers & Industrial Engineering*, vol. 62, no. 3, pp. 755–761, 2012.

[24] Beijing Subway co. LTD., *Management of ticket cards manual*, Beijing Subway co. LTD., Beijing, China, 2014, https://www.bjsubway.com/in Chinese.

[25] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proceedings of the Second International Symposium on Information Theory*, Budapest: Akademiai Kiado, Tsahkadsor, Armenia, pp. 267–281, September 1973.

[26] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366a, pp. 427–431, 1979.

[27] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[28] A. Pankratz, *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases*, Vol. 224, John Wiley & Sons, Hoboken, NJ, USA, 2009.

[29] Tiberlake, http://www.stata-uk.com/manuals.

[30] C. X. Liu, *Beijing Subway at Night and Off-Peak Hours for Urban Logistics System*, School of Economics, Beijing Supplies University, Beijing, China, 2011, in Chinese.

[31] W. S. Shen, "Reverie Beijing low carbon life and the development of subway freight function," *Beijing Observation*, vol. 1, pp. 13-14, 2010, in Chinese.

[32] A. Dampier and M. Marinov, "A study of the feasibility and potential implementation of metro-based freight transportation in Newcastle upon Tyne," *Urban Rail Transit*, vol. 1, no. 3, pp. 164–182, 2015.

[33] J. Kelly and M. Marinov, "Innovative interior designs for urban freight distribution using light rail systems," *Urban Rail Transit*, vol. 3, no. 4, pp. 238–254, 2017.

[34] Beijing Subway co. LTD., *Ticket Distribution Management Regulation*, Beijing Subway co. LTD., Beijing, China, 2014, in Chinesehttps://www.bjsubway.com/.

[35] L. U. Kai, M. H. Bao, L. U. Fang et al., "Urban rail transit in China: progress report and analysis (2008–2015)," *Urban Rail Transit*, vol. 2, no. 3-4, pp. 93–105, 2016.

[36] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[37] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.