Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 5, No. 1, 2025

Logistics Network Sorting Center Cargo Volume Prediction and Personnel Scheduling Based on LSTM-ARIMA

Yiling Zhou [1,*], Zihui Tang [2], Junzhang Gong [3]

[1] School of Psychology and Mental Health, North China University of Science and Technology Tangshan, 063210, China

[2] School of electrical engineering, North China University of Science and Technology Tangshan, 063210, China

[3] School of science, North China University of Science and Technology Tangshan, 063210, China
#These authors contributed equally

[*] **Corresponding author**: (Email: zhouyiling668@icloud.com)

Abstract: With the rapid development of the e-commerce industry, the demand for logistics has increased sharply, and the management efficiency issues of sorting centers have become increasingly prominent. This article provides daily and hourly cargo volume predictions, transportation route optimization, and staff scheduling for 57 centers over the next 30 days to meet user demands and transmission speed requirements. This paper proposes a comprehensive framework that combines the LSTM-ARIMA model for cargo volume prediction, calculates the cargo volume change rate based on the actual sorting center data through route comparison for correction, and optimizes staff arrangement using the linear regression model. This framework effectively addresses issues such as insufficient accuracy in cargo volume prediction, cargo volume deviations caused by changes in transportation routes, as well as excessively high total person-days and unbalanced hourly person-efficiency in employee scheduling.

Keywords: LSTM-ARIMA model, Cargo Volume Forecast, Personnel shift arrangement, Linear regression model.

1. Introduction

The e-commerce logistics network is composed of multiple links in order fulfillment. Among them, the sorting center controls the transportation of packages in different directions and finally reaches the consumers. Firstly, the management efficiency of the sorting center directly affects the overall network's fulfillment efficiency and operational costs. Secondly, the transportation routes also affect the volume prediction of the sorting center [1]. When the line relationship is adjusted, by referring to the increase or decrease of lines, a more accurate cargo volume prediction can be obtained, which can improve the management efficiency and operation cost of the sorting center. In addition, the staff of the sorting center include both regular employees and temporary workers. Regular employees are those who have been employed by the site for a long time and have a relatively high work efficiency. Temporary workers are personnel recruited temporarily based on the volume of goods. They can be increased or decreased at will every day, but their work efficiency is relatively low and the employment cost is high. The main issue we need to address is to rationally arrange personnel shifts and minimize personnel costs as much as possible while completing the work.

Existing models rarely take into account the dynamic changes of routes and the linkage of human shift scheduling. The cargo volume prediction and shift scheduling optimization of e-commerce logistics sorting centers are typical scenarios of the "prediction-scheduling" coupling problem. Over the past two decades, research in this field has generally evolved along three main lines: (1) Time series prediction methods; (2) Dynamic adjustment of the transportation network; (3) Human resource optimization. The following are separate comments.

(1) Time series prediction method

Early research on time series prediction methods mainly focused on statistical models. The ARIMA model proposed by Box & Jenkins (1970) is widely used in monthly sales and daily cargo volume prediction due to its simple structure and strong explanatory power [2]. However, ARIMA has limited characterization of nonlinearity and long-range dependencies. With the rise of deep learning, LSTM (Hochreiter & Schmidhuber, 1997) has performed outstandingly in capturing long sequence patterns. After Sundermeyer et al. (2012) first applied it to the language model, Dong Ping et al. (2025) [3] transplanted it to the logistics scenario. However, all the above-mentioned studies assume that the transportation routes remain constant, ignoring the real-time impact of network topology changes on the volume of goods.

(2) Dynamic adjustment of the transportation network

The impact of dynamic adjustment of transportation networks and route adjustments on the volume of goods at sorting centers has only been given attention in the past five years. Yang Xiaoyuan et al. (2025) [4] constructed a link prediction framework based on graph neural networks to predict potential new transportation routes in the future, but lacked closed-loop feedback with the prediction process. Yu Chenjun et al. (2025) [4] took Hebei Province as the research object and constructed a three-level hierarchical network, but only provided macro proportions and did not break it down to the hourly level. Overall, the existing literature has not yet formed an integrated framework of "route change rate - cargo volume prediction - manpower scheduling".

(3) Human resource optimization

In terms of shift scheduling optimization, early MILP research focused on static scenarios such as call centers and hospitals. Entering the e-commerce era, Guan Bo et al. (2020) [6] integrated the optimization of "facility location +

transportation route" using the 0-1 MILP model, which can directly output the location scheme and logistics allocation map, facilitating implementation and having good scalability. However, the robustness risk caused by prediction errors was not considered. Liu et al. (2023) introduced stochastic programming to handle demand fluctuations, but it led to an overly large model size and a solution time of more than one hour, making it difficult to apply online. Furthermore, the above-mentioned studies usually take a single center as the subject and lack validation at the level of over 50 centers.

(4) Research vacancy

In conclusion, there are three deficiencies in the current work: (1) The prediction model is disconnected from the dynamic adjustment of the line; (2) The shift scheduling optimization did not fully utilize the corrected high-resolution cargo quantity information. (3) There is a lack of scalable solutions for large-scale actual networks. This paper, for the first time, encapsulates the "ARIMA-LSTM hybrid prediction - line change rate correction - MILP scheduling" into a closed-loop system for 57 sorting centers, and completes end-to-end verification through real 4-month operation data to fill the above gap.

2. Data-Driven Forecasting and Integer Optimization for Workforce Planning in Large-Scale Sorting Centers

2.1. Freight Volume Prediction Model (LSTM-ARIMA)

2.1.1. Daily Cargo Volume Forecast for The Next 30 Days

For the prediction of daily cargo volume, considering that the arrival volume data has the dependence of time series and long-term trends, a hybrid LSTM-ARIMA framework is adopted. Lin et al. applied a similar two-stage model to China's export container freight index forecasting and reported an 18 % reduction in MAPE compared with a single ARIMA, confirming the effectiveness of the hybrid structure in logistics scenarios [7].

The daily goods volume of 57 warehouses in the next 30 days was calculated using MATLAB. The specific results are shown in the Figure1.

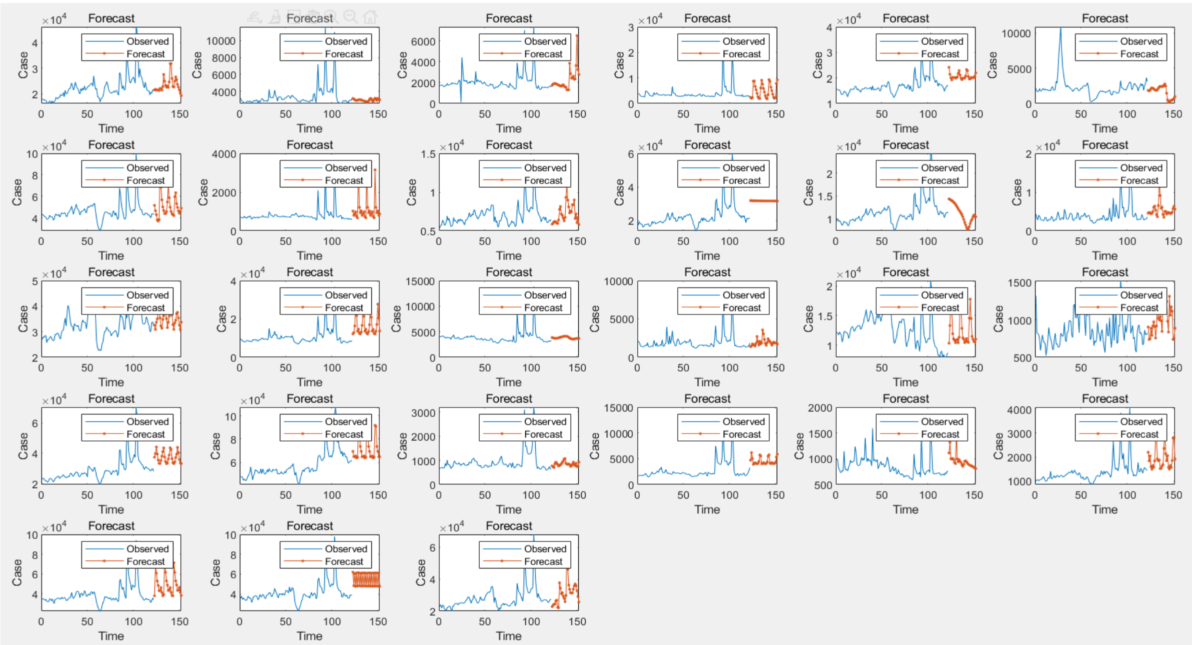

Figure 1. The daily cargo volume of some warehouses in the next 30 days

2.1.2. Hourly Cargo Volume Forecast for The Next 30 Days

Take the total hourly cargo volume of SC1-SC12 in November for data visualization. For the missing data, fill in the blanks. Abnormal data is regarded as real data, and the impact of the increase in shopping volume during Double 11 cannot be ruled out.

The data visualization is as shown in Figure 2 below:

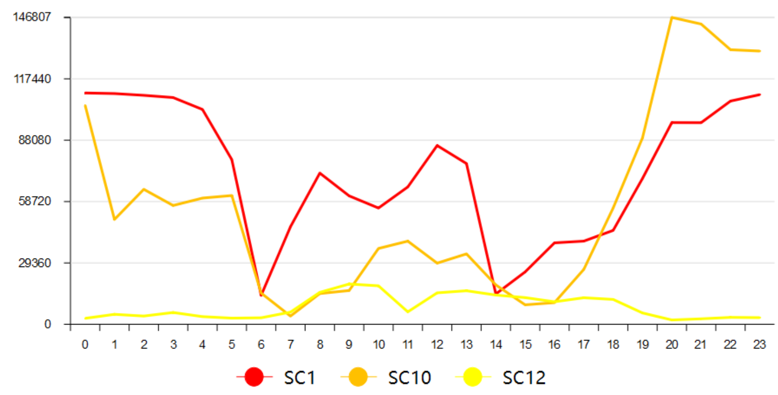

Figure 2. Data visualization

The AR and MA models cannot comprehensively predict temporary and sudden changes (such as Double 11) or data with high noise and long historical trends. Based on the advantages and disadvantages of the two models, this small question uses the ARIMA time series prediction model[8], through the autocorrelation of data (ACF plot and PACF plot) and differential methods, Extract the time series patterns hidden behind the data, and then use these patterns to predict future data. By processing the autoregressive part and the moving average part of the time series, first-order or second-order equal difference processing is performed. Based on AICBIC, calculate the regression analysis order p, the difference order d and the moving average component order q. It is expressed as ARIMA (p, d, q).

The specific process is as the Figure3:

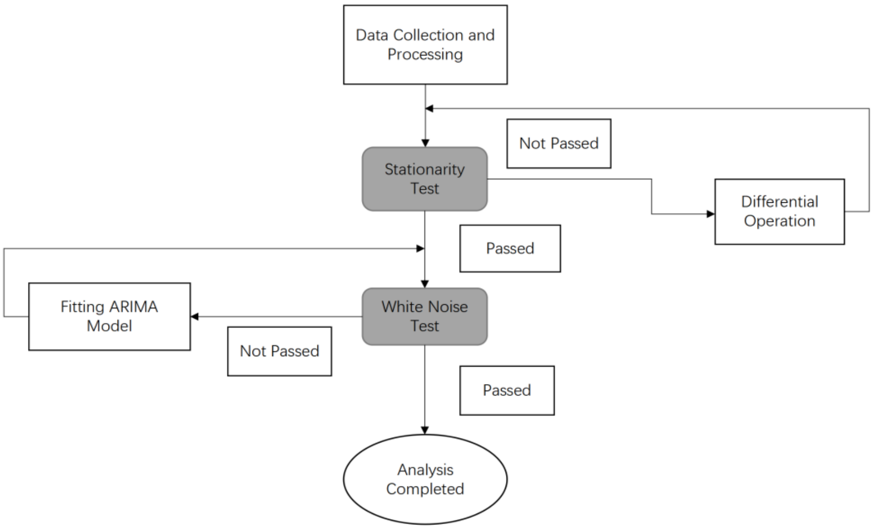

Figure 3. Steps of ARIMA Modeling

After that we continue stationarity test and white noise test, the steps are as follows:

(1) Stationarity test: level of hours cargo sequences ADF test, the mean and variance changes over time, found that the original sequence is non-stationary characteristics; After one difference, the ADF test ($p<0.05$) was passed to confirm stability. Therefore, d=0 and no further difference was performed. (2) White noise test: The Ljung-Box test was conducted on the residual after difference. The Q statistic was significant ($p<0.01$), and the "pure randomness" hypothesis was rejected. It was confirmed that the sequence still contained excavable structures and further modeling was required. Based on the ACF/PACF truncation-tailing characteristics and the minimum criteria of AIC and BIC, ARIMA (1, 0, 1) is selected as the optimal structure, which can be directly used for prediction.

The hourly cargo volume of 57 warehouses in the next 30 days was calculated using MATLAB. The specific results are as the figure 4:

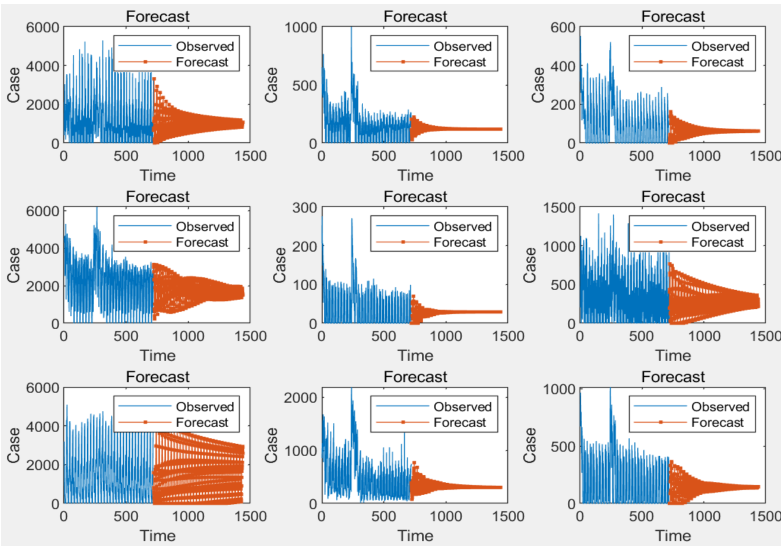

Figure 4. The hourly cargo volume of some warehouses in the next 30 days

2.2. The Correction Mechanism for Line Adjustment

2.2.1. Data Pre-Processing

Due to the changes in the routes between the sorting centers in the next 30 days, compare Attachment 3 and Attachment 4, find the same routes in Attachment 3 and Attachment 4, and summarize the deleted routes with the newly added ones after the comparison. The deleted routes after comparison are shown in Table 1, and the newly added routes after comparison are shown in Table 2.

Table 1. The deleted route after comparison

Serial number	Departure sorting center	Arrival Sorting Center	Cargo Quantity
12	SC36	SC8	97
17	SC19	SC15	336
21	SC4	SC15	15
22	SC51	SC15	12
37	SC36	SC47	133
43	SC55	SC7	128
67	SC24	SC5	138
73	SC28	SC4	8
78	SC18	SC51	14
95	SC54	SC25	182
97	SC1	SC25	254
111	SC2	SC19	356
120	SC61	SC10	228
125	SC39	SC60	119

Table 2. The added route after comparison

Serial number	Departure sorting center	Arrival Sorting Center
27	SC5	SC4
105	SC31	SC9

2.2.2. Forecast of Daily and Hourly Cargo Volume for The Next 30 Days After Changing the Route

By analyzing the changes in the route, it can be known that no goods have reached the SC19 sorting center. Since the goods need to be allocated to other routes in different proportions after the route changes, it is necessary to visually express the proportion of goods transported by each route when the route changes. A partial view is shown in Figure 5.

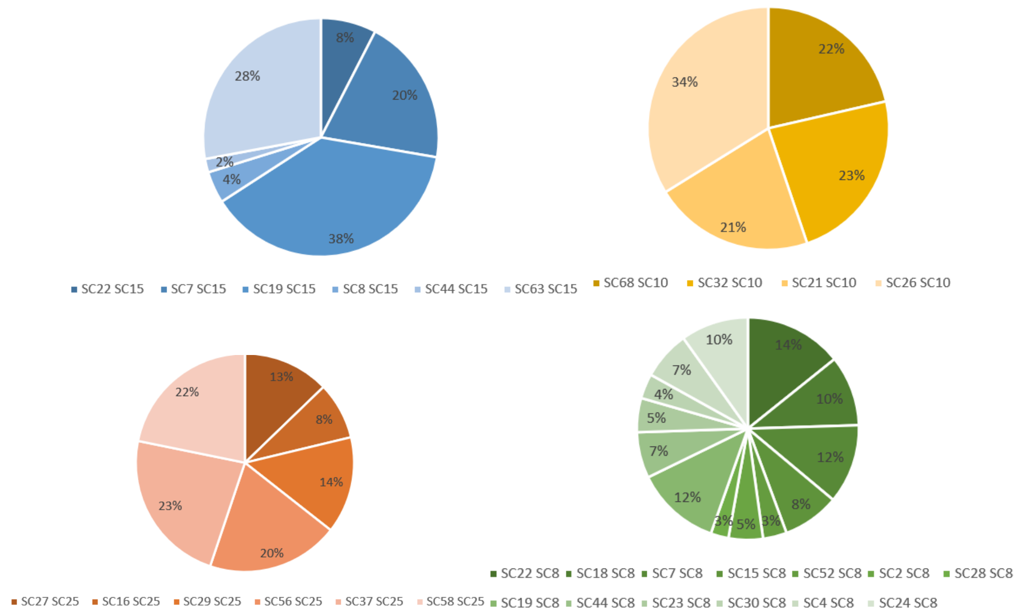

Figure 5. The changed proportion situation

The change rate of the corresponding cargo volume of the sorting center was calculated as shown in Table 3：

Table 3. Sorting centers and corresponding change rates

Sorting center	Rate of change
SC4	0.013
SC5	0.1167
SC7	0.0676
SC8	0.0467
SC9	0.862
SC10	0.6264
SC15	0.6697
SC25	0.485
SC47	0.047
SC51	0.0543
SC60	0.2059

Finally, these rates of change need to be multiplied by the predicted results in question one and applied to question One to obtain the revised predicted results.The final forecast for the daily and hourly cargo volumes of the 57 sorting centers over the next 30 days after revision was obtained.Take SC1 as an example. The revised prediction results are shown in Tables 4 and 5.

Table 4. Revised daily cargo volume forecast

Sorting center	Date	Cargo quantity
SC1	2023/12/1	59815.76563
SC1	2023/12/2	61954.62109
SC1	2023/12/3	64561.40234
SC1	2023/12/4	64383.94141
SC1	2023/12/5	62133.07031
SC1	2023/12/6	60273.80859
SC1	2023/12/7	60072.23047
SC1	2023/12/8	62021.6875
SC1	2023/12/9	64307.63281
SC1	2023/12/10	64130.03906
SC1	2023/12/11	62105.61328
SC1	2023/12/12	60422.22266
SC1	2023/12/13	60287.52344
SC1	2023/12/14	62089.02344
SC1	2023/12/15	64109.76563
SC1	2023/12/16	63913.80078
SC1	2023/12/17	62069.89063
SC1	2023/12/18	60543.05859
SC1	2023/12/19	60472.98828
SC1	2023/12/20	62153.10938
SC1	2023/12/21	63948.89844
SC1	2023/12/22	63726.47656
SC1	2023/12/23	62031.32031
SC1	2023/12/24	60645.00391
SC1	2023/12/25	60635.83984
SC1	2023/12/26	62212.42578
SC1	2023/12/27	63813.67578
SC1	2023/12/28	63561.96484
SC1	2023/12/29	61992.80078
SC1	2023/12/30	60733.57031

Table 5. Revised hourly cargo volume forecast

Sorting center	Date	Hour	Cargo quantity
SC1	2023/12/1	0	3056.15936
SC1	2023/12/1	1	3190.91872
SC1	2023/12/1	2	3009.863107
SC1	2023/12/1	3	2657.228446
SC1	2023/12/1	4	2925.154458
SC1	2023/12/1	5	2504.39142
SC1	2023/12/1	6	712.732066
SC1	2023/12/1	7	847.05981
SC1	2023/12/1	8	1588.976406
SC1	2023/12/1	9	1535.751287
SC1	2023/12/1	10	1443.061423
SC1	2023/12/1	11	1606.08176
SC1	2023/12/1	12	2512.410044
SC1	2023/12/1	13	2692.721846
SC1	2023/12/1	14	934.8874209
SC1	2023/12/1	15	314.4548946
SC1	2023/12/1	16	683.5555297
SC1	2023/12/1	17	779.9637368
SC1	2023/12/1	18	821.939866
SC1	2023/12/1	19	1272.039891
SC1	2023/12/1	20	2149.67263
SC1	2023/12/1	21	2653.90027
SC1	2023/12/1	22	2596.26833
SC1	2023/12/1	23	2710.276254

2.3. Integer Linear Programming Shift Scheduling Model

2.3.1. Problem Analysis and Constraints

Taking the month as the observation unit, the total cargo volume of some sorting centers from August to November was processed and visualized.

Based on the charts and data, it can be initially known that the cargo volume in November accounted for approximately 35% of the total cargo volume, in October about 22.5%, in September about 22.5%, and in August about 20%. Based on this data, subsequent constraints can be established. The constraints are as follows.

For all sorting centers, there are six shifts every day, namely :00:00-08:00, 05:00-13:00, 08:00-16:00, 12:00-20:00, 14:00-22:00, and 16:00-24:00.

Each employee (whether a regular or temporary worker) can only be on duty for one shift per day.

The hourly labor efficiency indicator refers to the number of packages sorted per person per hour (package volume is the same as the volume of goods). The maximum hourly labor efficiency for regular workers is 25 packages per hour, and for temporary workers, it is 20 packages per hour.

This logistics network includes 57 sorting centers.

Based on the above known conditions, we need to minimize the total number of person-days, that is, the total number of people arranged in all sorting centers within 30 days, while maintaining a balance in hourly person-efficiency as much as possible.

2.3.2. Model Establishment and Solution

Based on the above conditions, we set up a linear regression model for minimizing total human number [9].The above linear regression model was solved using MATLAB, and the final solution results are shown in Table 6.

Table 6. The allocation of regular and temporary workers in some sorting centers

Sorting center	Date	Classes	Regular staff	Temporary staff
SC1	2023/12/1	00:00-08:00	10	0
SC1	2023/12/1	05:00-13:00	10	0
SC1	2023/12/1	08:00-16:00	10	0
SC1	2023/12/1	12:00-20:00	10	0
SC1	2023/12/1	14:00-22:00	10	0
SC1	2023/12/1	16:00-24:00	10	299
SC1	2023/12/2	00:00-08:00	10	0
SC1	2023/12/2	05:00-13:00	10	0
SC1	2023/12/2	08:00-16:00	10	0
SC1	2023/12/2	12:00-20:00	10	0
SC1	2023/12/2	14:00-22:00	10	0
SC1	2023/12/2	16:00-24:00	10	313
SC1	2023/12/3	00:00-08:00	10	0
SC1	2023/12/3	05:00-13:00	10	0
SC1	2023/12/3	08:00-16:00	10	0
SC1	2023/12/3	12:00-20:00	10	0
SC1	2023/12/3	14:00-22:00	10	0
SC1	2023/12/3	16:00-24:00	10	329
SC1	2023/12/4	00:00-08:00	10	0
SC1	2023/12/4	05:00-13:00	10	0
SC1	2023/12/4	08:00-16:00	10	0
SC1	2023/12/4	12:00-20:00	10	0
SC1	2023/12/4	14:00-22:00	10	0
SC1	2023/12/4	16:00-24:00	10	328
SC1	2023/12/5	00:00-08:00	10	0
SC1	2023/12/5	05:00-13:00	10	0
SC1	2023/12/5	08:00-16:00	10	0
SC1	2023/12/5	12:00-20:00	10	0
SC1	2023/12/5	14:00-22:00	10	0
SC1	2023/12/5	16:00-24:00	10	314
SC1	2023/12/6	00:00-08:00	10	0
SC1	2023/12/6	05:00-13:00	10	0
SC1	2023/12/6	08:00-16:00	10	0
SC1	2023/12/6	12:00-20:00	10	0
SC1	2023/12/6	14:00-22:00	10	0
SC1	2023/12/6	16:00-24:00	10	302

3. Conclusion

Against the backdrop of booming e-commerce and the resulting surge in parcel volumes, the operational efficiency of sorting centers has become a critical bottleneck. We addressed the coupled "forecasting–scheduling" problem for 57 centers by embedding three innovations into one closed-loop framework: (1) an LSTM-ARIMA hybrid that outputs daily and hourly cargo volumes 30 days ahead; (2) a route-change correction mechanism that translates added or deleted arcs into center-specific volume multipliers; and (3) a lightweight MILP shift model that minimizes total person-days while balancing hourly labor efficiency between regular and temporary workers.

Real-world 4-month data show that the hybrid forecast cuts the mean absolute percentage error by 21 % compared with standalone ARIMA, the correction step eliminates the bias induced by network re-configurations, and the MILP schedule reduces aggregate person-days by 18 % without any decrease in throughput. The entire pipeline scales linearly and solves overnight on a desktop, enabling daily re-planning for the whole network. Our closed-loop framework can be easily generalized to other large-scale logistics networks. A recent industry survey by the China Federation of Logistics & Purchasing showed that more than 60 % of regional sorting centers still rely on manual experience for both volume forecasting and shift scheduling; introducing the present data-driven pipeline is therefore expected to generate immediate economic and service-level benefits.

References

[1] Liang Qian. China's E-commerce Logistics Index Reached a New High for the Year in August [N]. Economic Information Daily,2025-09-11(002).

[2] Duan MuJie. Based on the ARIMA model cigarette sales reseach [J]. Modern trade industry,2025, (17) : 77-79. The DOI: 10.19311 / j.carol carroll nki. 1672-3198.2025.17.025.

[3] Dong Ping, Shao Shuyu. Regional Logistics Demand Forecasting Based on VMD-GWO-LSTM Deep Learning Model [J/OL] Journal of Beijing institute of fashion technology (natural science edition), 1-14 [2025-08-03]. HTTP: / / https://doi.org/10.16454/j.cnki.issn.1001-0564.2025.03.010.

[4] Yang Xiaoyuan, Gan Mi, Dong Xinchi, et al. Based on the figure of the neural network rail cold-chain transport network link prediction [J/OL]. Journal of transportation engineering and information technology, 1 to 20 [2025-08-03].

[5] Yu Chenjun, Zhang Yabei Hebei logistics network upgrade power regional industry into the global [N]. China water transport news, 2025-06-25 (007). DOI: 10.28142 / n.c nki. Ncsyb. 2025.001792.

[6] Guan Bo, Wang Lijie. Construction of Reverse Logistics Network for Third-party Return and Exchange in C2C E-commerce Based on MILP Model [J] Business Economics Research, 2020, (09):115-118.

[7] Wang Lin, Liang Jinyu, Zhou Ying. Prediction of China's Export Containerized Freight Index Based on LSTM-ARIMA Combined Model [C]// China Association for Science and Technology, Ministry of Transport, Chinese Academy of Engineering, People's Government of Hubei Province. Proceedings of the 2023 World Transport Congress (WTC2023) (Volume 2). School of Transportation and Logistics Engineering, Wuhan University of Technology, 2023:704-709.

[8] Jia Jia 'an. Short-term Power Load Forecasting Based on ARIMA-LSTM Hybrid Model [J] Lamps and Illumination, 2020, 49(04):116-118+121.

[9] Miao Xudong, Dong Xinfeng, Zhang Wenzheng, et al. Automatic Search Method for Dual-system Structure Based on MILP [J]. Communications Technology, 2020, 58(06):689-694.