Research on Cargo Volume Prediction of Logistics Networks Based on Time Series Processing and Directed Acyclic Graph Model

Yu Xia *, #, Zihang Wei #

School of Mathematics and Statistics, Central China Normal University, Wuhan, China, 430079

* Corresponding Author Email: xiayu0923@mails.ccnu.edu.cn

#These authors contributed equally.

Abstract. With the rapid development of e-commerce and the logistics industry, accurately predicting the cargo volume at logistics sorting centers to improve the efficiency and service quality of the logistics system has become an important research topic. In response to this, this paper conducts a comprehensive study. First, logistics operation data from the platform are obtained through the competition's official website, followed by data preprocessing, including time series conversion, wavelet transform denoising, white noise detection, differencing and smoothing, and stationarity tests, to transform the cargo volume data of the logistics sorting center into a stationary series. Next, the ARIMA model is applied to forecast the cargo volume for the next 30 days. Furthermore, a directed acyclic graph model of the logistics sorting center is constructed, and historical data is used to calculate node strength. Future transportation route changes are also considered to update node strength and solve the state transition weight matrix. Ultimately, by using the model, the overall prediction accuracy of the cargo volume reached over 80%, with an average predicted cargo volume of 50677.8kg for December, and a total predicted cargo volume of 1520334kg for the entire month of December. This paper holds significant importance in the study of transportation and cargo volume forecasting within logistics networks.

Keywords: Wavelet Transform, Directed Acyclic Graph, Autoregressive Integrated Moving Average (ARIMA) Model.

1. Introduction

With the rapid development of e-commerce and the logistics industry, logistics sorting centers play a crucial role in modern logistics networks. The cargo handling capacity of these centers directly affects the overall efficiency and service quality of the logistics system. However, as the cargo volume in logistics sorting centers is influenced by various factors, such as seasonal fluctuations, holiday promotions, and changes in market demand, accurately predicting cargo volumes has become a significant research topic [1].

In this study, This paper employed the wavelet denoising algorithm, differencing and smoothing algorithms, and a distributed consensus algorithm based on directed acyclic graphs (DAG). The wavelet denoising algorithm, introduced by Donoho, is widely applied in signal processing [2-5]. Drawing on the principles of wavelet denoising, we adapted this method from processing signal systems to discrete data, which helps ensure stability for solving this issue. The distributed consensus algorithm based on DAG has proven to be a powerful tool, significantly aiding the coordination and prediction of the logistics DAG model we established later. This algorithm also plays a significant role in work scheduling and job assignments.

In this paper, we first preprocess the data by sorting and removing missing values, followed by wavelet denoising [6], white noise detection, and differencing and smoothing to ensure data stability. Next, we apply the ARIMA model [7-8] to forecast the cargo volume and obtain prediction values. Subsequently, we construct the DAG model and adjacency matrix for the logistics sorting center, calculate and update node strength, and solve the state transition matrix. Finally, by combining the time series prediction with the state transition matrix, we predict the future cargo volume. It lays an important foundation for finding the supply balance of cargo volume and the actual transportation planning of goods for real-world logistics network companies.

2. Data Source and Preprocessing

These data for this study are the historical records of a logistics company in an actual city selected for the competition. They come from the website www.mathorcup.org.

2.1. Data Preprocessing

First, consider the time series sorting operation on the dataset to convert irregular data into time series data. The unit of time for Annex 1 is the date and the unit of time for Annex 2 is the hour.

Due to the missing values in the dataset, the dataset is treated as an outlier as follows, and the original time series dataset is set as $T=\{(t_1,\tau_1),(t_2,\tau_2),\cdots,(t_n,\tau_n)\}$, where the time point is denoted t_i and the observed value τ_i is represented at the time point t_i.

If an observation τ_i is missing (usually represented by NaN, None, null, or other special markers), we remove the corresponding value from the dataset. The dataset after the filtering operation is

$$T'=\{(t_i,\tau_i)\in T\,|\,\tau_i \text{ is not a missing value}\} \tag{1}$$

For example, the trend of the daily cargo volume in the four months after the pretreatment and the daily and hourly cargo volume in November are shown in Figure 1.

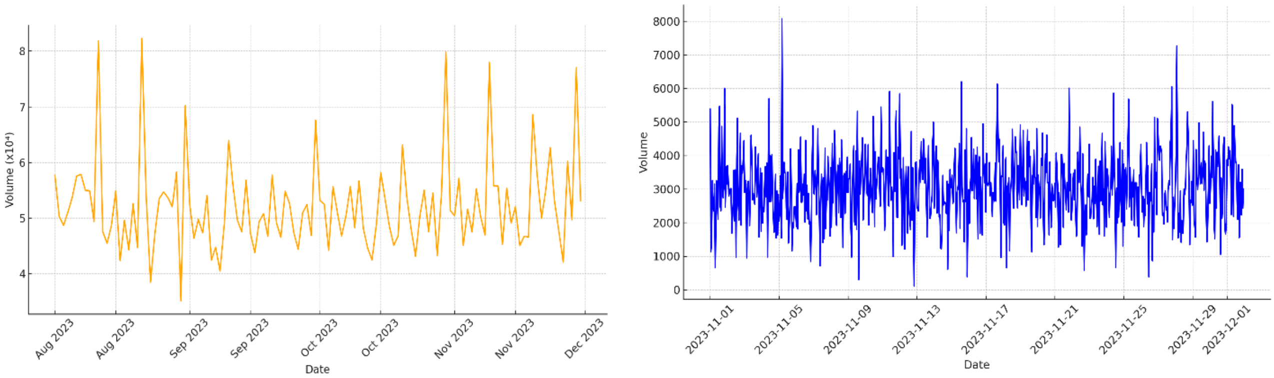

Figure 1. Daily Volume Trend for Center SC1

From Figure 1, it can be observed that the cargo volume of the Center SC1 shows a changing trend from August to November and throughout the entire month of November in the future. The monthly trend is generally higher during the summer holiday period and in November, which can be approximately attributed to the demand influenced by summer holidays and the Double 11 festival. The cargo volume in November shows volatility, aligning with market regulation patterns.

2.2. Noise reduction processing and data standardization based on wavelet transform

Due to the multiple influences of time series datasets, there is a high signal-to-noise ratio, so it is necessary to consider denoising, smoothing, and normalizing the datasets.

In the case of time-frequency localization, the wavelet transform is more widely applied than the traditional Fourier transform. Due to the artificial controllability of the transportation of goods in the logistics center, the time series dataset of this topic contains many frequency components, so the wavelet transform is considered to denoise the data.

Based on the data characteristics, the following discrete wavelet transform is considered.

$$\int_{-\infty}^{\infty}\psi(t)\,\mathrm{d}t=0 \quad \Rightarrow \quad \psi_{a,b}(t)=\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{2}$$

where the wavelet function is denoted $\psi(t)$, $\{\psi_{a,b}(t)\}$ is a family of functions transformed by translation and scaling of $\psi(t)$, which a,b are all parameters in the process of translation and scaling, $a,b \in \mathbb{R}$ and $a > 0$.

For a and b, discretization is performed to complete the discretization process, which yields

$$\begin{cases} \psi_{m,n}(t) = a_0^{-\frac{m}{2}} \psi(a_0^{-m} t - k b_0) \qquad m,n,k \in \mathbb{Z} \\ m = \log_\alpha a' \\ |b'| = k|b| \end{cases} \tag{3}$$

where a', b' denotes the coordinates of a,b after discretization. Therefore, the discrete wavelet transform in the dataset is

$$Wf(m,n) = \int_{-\infty}^{\infty} f(t) \psi_{m,n}(t) \mathrm{d}t \tag{4}$$

$$Wf(m,n) = a_0^{-\frac{m}{2}} \int_{-\infty}^{\infty} f(t) \psi_{m,n}(a_0^{-m} t - k b_0) \mathrm{d}t \tag{5}$$

where $f(t)$ denotes the inner product of the discrete wavelet transform.

After applying discrete wavelet denoising to the aforementioned time series dataset, the resulting relatively stable dataset is

$$\begin{cases} \Upsilon_1 = \{(t_{1,1}, \tau_{1,1}), (t_{1,2}, \tau_{1,2}), \cdots\} \\ \Upsilon_2 = \{(t_{2,1}, \tau_{2,1}), (t_{2,2}, \tau_{2,2}), \cdots\} \\ \cdots \\ \Upsilon_N = \{(t_{N,1}, \tau_{N,1}), (t_{N,2}, \tau_{N,2}), \cdots\} \\ \Upsilon'_N = \{(t'_{N,1}, \tau'_{N,1}), (t'_{N,2}, \tau'_{N,2}), \cdots\} \end{cases} \tag{6}$$

For the logistics sorting center SC1, the trend of the data after wavelet transform compared to the original cargo volume data is shown in Figure 2.

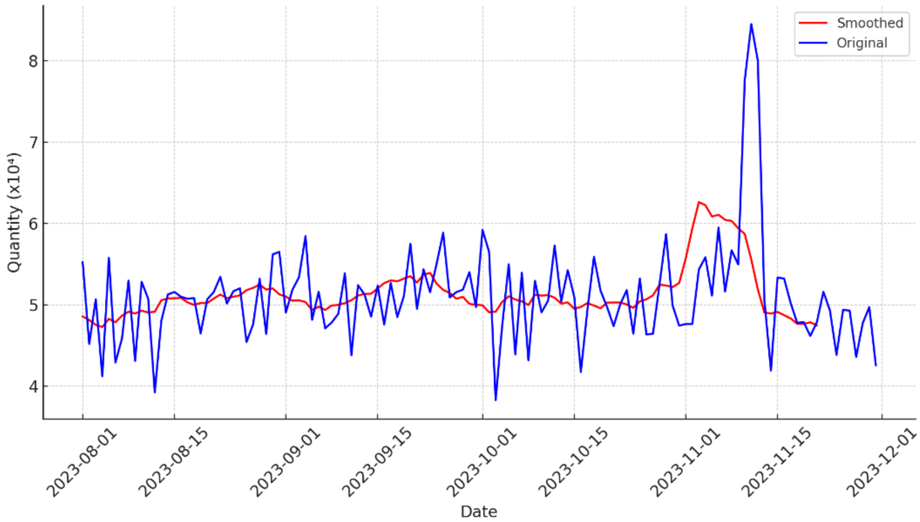

Figure 2. Comparison between the Wavelet-Transformed SC1 Data and the Original Data

It is evident from the dataset that the cargo volume in mid-November is relatively high, which can be approximately attributed to the influence of the "Double Eleven" shopping festival. Therefore, for the logistics sorting center SC1, the volume of sorted cargo around November 11 is higher than on other dates.

3. ARIMA for Logistics Volume Prediction

3.1. Dataset Smoothing and Stabilization Processing

To stabilize the data, we consider applying differencing to the processed dataset, thereby transforming the non-stationary and non-smooth data into stationary and smooth data. Here, we use the smooths function to smooth the processed dataset and obtain a trend chart showing the smoother changes in the data [9].

To avoid the occurrence of non-stationarity in the time series dataset, we consider further processing the data using the differencing method. For the dataset sequence (t_i, τ_i), assuming the original time series is $\{\widetilde{\tau_i}\}$, where t represents the time point, the first-order differenced sequence $\{\Delta\widetilde{\tau_t}\}$ is defined as

$$\Delta\widetilde{\tau_t} = \widetilde{\tau_t} - \widetilde{\tau_{t-1}} \tag{7}$$

Here, $\Delta\widetilde{\tau_t}$ represents the first-order difference at the time point t, $\widetilde{\tau_t}$ is the original data value at the time point t, and $\widetilde{\tau_{t-1}}$ is the data value at the previous time point $t-1$. Additionally, the second-order difference is defined as

$$\Delta^2\widetilde{\tau_t} = \Delta\widetilde{\tau_t} - \Delta\widetilde{\tau_{t-1}} = \widetilde{\tau_t} - 2\widetilde{\tau_{t-1}} + \widetilde{\tau_{t-2}} \tag{8}$$

Taking logistics sorting centers SC1, SC10, SC12, SC14 and SC15 from Attachment 1 as examples, the trend of the cargo volume in the different datasets over time is shown in Figure 3.

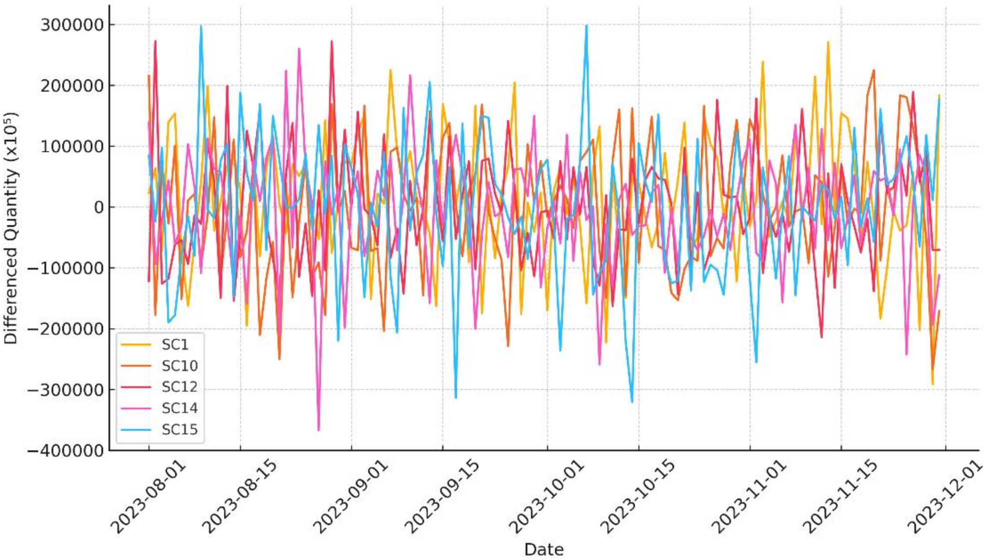

Figure 3. Different Quantity Data for Multiple Centers

From Figure 3, it can be seen that the cargo volume of centers SC1, SC10, SC12, SC14, and SC15 from August to November 2023 shows fluctuating changes. This aligns with market regulation trends and provides an effective foundation for the subsequent forecasting.

3.2. Analysis of the ARIMA Time Series Forecasting Model

Based on the processed datasets above, we first conduct stationarity and white noise tests to determine whether the datasets meet the conditions for time series forecasting [10].

$$\Delta\widetilde{\tau}_t = \alpha + \beta t + \gamma\widetilde{\tau}_t + \sum_{i=1}^{p}\delta_i \Delta\widetilde{\tau}_t + \varepsilon \tag{9}$$

Conducting the white noise test, the corresponding Box-Pierce Q statistic and Ljung-Box Q statistic satisfy the following expressions[11]..

$$Q_{BP}(m) = n\sum_{k=1}^{m}\widehat{\rho_k^2} \tag{10}$$

$$Q_{LB}(m) = n(n+2)\sum_{k=1}^{m}\frac{\widehat{\rho_k^2}}{n-k} \tag{11}$$

Here, $\widehat{\rho_k}$ represents the estimated value of the autocorrelation coefficient at a lag of k periods. After stationarity testing and white noise processing, the time series dataset has characteristics that allow for accurate forecasting. Therefore, we consider the following ARIMA model.

In this model, AR represents the autoregressive component and p becomes the autoregressive coefficient, which depends only on its p historical values. The corresponding expression is

$$X_t = \phi_1 X_t + \phi_2 X_{t-2} + \cdots \phi_p X_{t-p} + \varepsilon_t \tag{12}$$

For the stepwise correlation and processing between sequential data, the differencing order is defined. Based on the autoregressive model and differencing order calculations, the following moving average model is defined to represent the time series dataset.

$$X_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \tag{13}$$

Thus, this time series forecasting model can be integrated into the following form [12].

$$\begin{cases} X_t = \sum_{i=1}^{p}\phi_i X_{t-i} + \varepsilon_t \\ X_t = \Delta X_t = \Delta^2 X_t = \Delta^3 X_t = \cdots \Delta^p X_t \\ X_t = \mu + \sum_{j=1}^{q}\theta_q\varepsilon_{t-q} + \varepsilon_t \end{cases} \tag{14}$$

The predicted changes in cargo volume at logistics sorting center SC1 from Attachment 1 over the next 30 days and hourly changes for the next 30 days are shown in Figure 4.

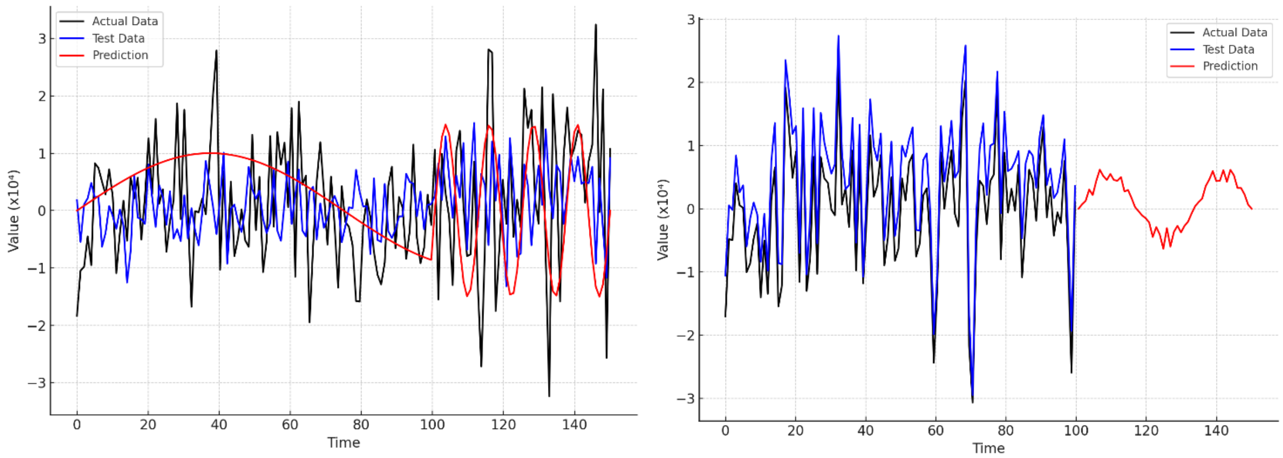

Figure 4. Comparison of ARIMA Time Series Prediction Results

Figure 4 shows that the prediction (represented by the red line) closely follows the trends in both the actual and test data, highlighting the accuracy of the time series forecasting model. The application of wavelet transform denoising effectively minimized noise, resulting in smoother and more stable data. This improvement in data quality significantly enhanced the model's predictive performance. The combination of this preprocessing technique with ARIMA modeling ensures reliable and accurate estimates of future cargo volumes, particularly in scenarios where the logistics data is complex and subject to significant fluctuations.

4. Logistics Network Modeling Based on Directed Acyclic Graph

4.1. Construction of the Directed Acyclic Graph Model for Logistics Sorting Centers

The dataset provided in Attachment 3 contains both the originating and destination logistics sorting centers, with specific constraints on whether each pair of centers is connected. Additionally, the cargo volume corresponding to each connecting line segment is also provided [13-14].

Therefore, we consider constructing a directed acyclic graph (DAG) that encompasses all logistics sorting centers, routes, and cargo volumes, satisfying the following conditions.

$$\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{W}) \tag{15}$$

$$c_{ij} = \begin{cases} W(e_{ij}) & e_{ij} = (v_i, v_j) \in \mathrm{E} \\ 0 & e_{ij} = (v_i, v_j) \notin \mathrm{E} \end{cases} \tag{16}$$

Here, n represents the number of nodes, and the number of sorting centers in the logistics network. In this case, $n = 57$ $W(e_{ij})$ represents the weight of the edge e_{ij}, which corresponds to the cargo volume on each transportation route between the sorting centers.

Finally, the directed acyclic graph constructed using MATLAB is shown in Figure 5.

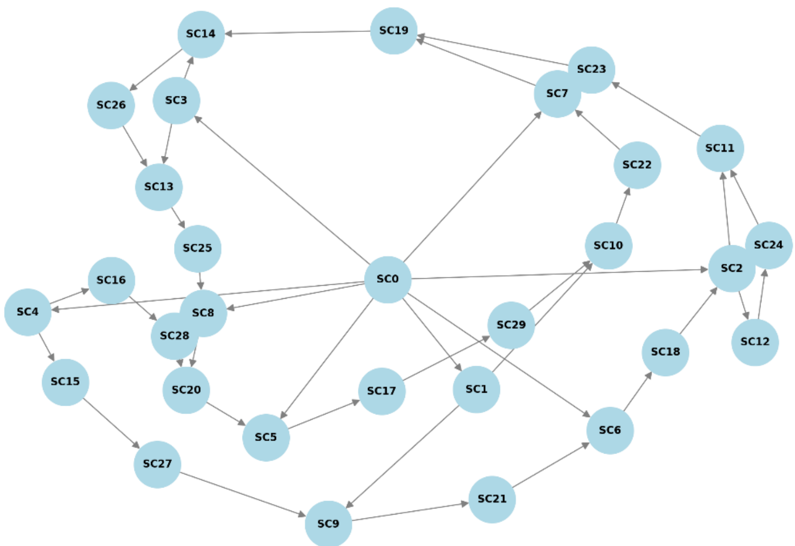

Figure 5. Directed Acyclic Graph of the Logistics Sorting Center

From Figure 5, we can see a directed acyclic graph with the logistics sorting center SCi as the node. This graph reflects the sequence and directional relationships of transportation between the centers.

4.2. Node Strength Calculation Based on Edge Weights

For any node in the logistics network, $v \in \mathrm{V}$ representing each sorting center. On one hand, the weighted in-degree $W_1(v_i)$ of this node is represented by the sum of the weights of all edges pointing to the node v (the total cargo volume transported to this sorting center from other centers), reflecting the center's ability to receive goods. Thus, we can get

$$W_1(v_i) = \sum_{v_i \in \mathrm{V}, e_{ji} \in \mathrm{E}} W(e_{ji}) \tag{17}$$

On the other hand, the weighted out-degree $W_2(v_i)$ of the node v is represented by the sum of the weights of all edges originating from this node (the total cargo volume transported from this sorting center to other centers), reflecting the center's ability to send goods. Thus, we can obtain

$$W_2(v_i) = \sum_{v_i \in \mathrm{V}, e_{ij} \in \mathrm{E}} W(e_{ij}) \tag{18}$$

Let the node strength of node v_i S_i be, representing the importance and influence of the sorting center v_i within the logistics network. This satisfies the following conditions

$$S_i = W_1(v_i) + W_2(v_i) \tag{19}$$

At this point, each obtained node strength S_i corresponds to the previous node strength value from the historical data in the dataset corresponding to the website.

Due to the complexity of the logistics network, this paper no longer considers the changes in cargo volume on the transportation routes between sorting centers caused by route changes between centers, i.e., the weights of the directed edges.

When an edge $e_{kl} = (v_k, v_l) \in \mathrm{E}$ is disconnected, meaning that the logistics sorting center v_k no longer transports goods to the logistics sorting center v_l, the weighted in-degree and weighted out-degree satisfy the following conditions.

$$\begin{cases} W_1'(v_l) = W_1(v_l) - W(e_{kl}) \\ W_2'(v_k) = W_2(v_k) - W(e_{kl}) \end{cases} \tag{20}$$

Thus, the corresponding new node strength satisfies the following conditions[15].

$$S_k' = W_1(v_k) + W_2'(v_k) = W_1(v_k) + W_2(v_k) - W(e_{kl}) = S_k - W(e_{kl}) \tag{21}$$

$$S_l' = W_1(v_l) + W_2'(v_l) = W_1(v_l) + W_2(v_l) - W(e_{kl}) = S_k - W(e_{kl}) \tag{22}$$

When an edge $e_{kl} = (v_k, v_l) \in \mathrm{E}$ is connected, meaning the logistics sorting center v_k begins transporting goods to the logistics sorting center v_l, the weighted in-degree, weighted out-degree, and new node strength satisfy the following conditions.

$$\begin{cases} W_1'(v_l) = W_1(v_l) + W(e_{kl}) \\ W_2'(v_k) = W_2(v_k) + W(e_{kl}) \\ S_k' = W_1(v_k) + W_2'(v_k) = W_1(v_k) + W_2(v_k) + W(e_{kl}) = S_k + W(e_{kl}) \\ S_l' = W_1(v_l) + W_2'(v_l) = W_1(v_l) + W_2(v_l) + W(e_{kl}) = S_k + W(e_{kl}) \end{cases} \tag{23}$$

Using the logistics sorting center SC60 as the central node, its situation is illustrated in Figure 6.

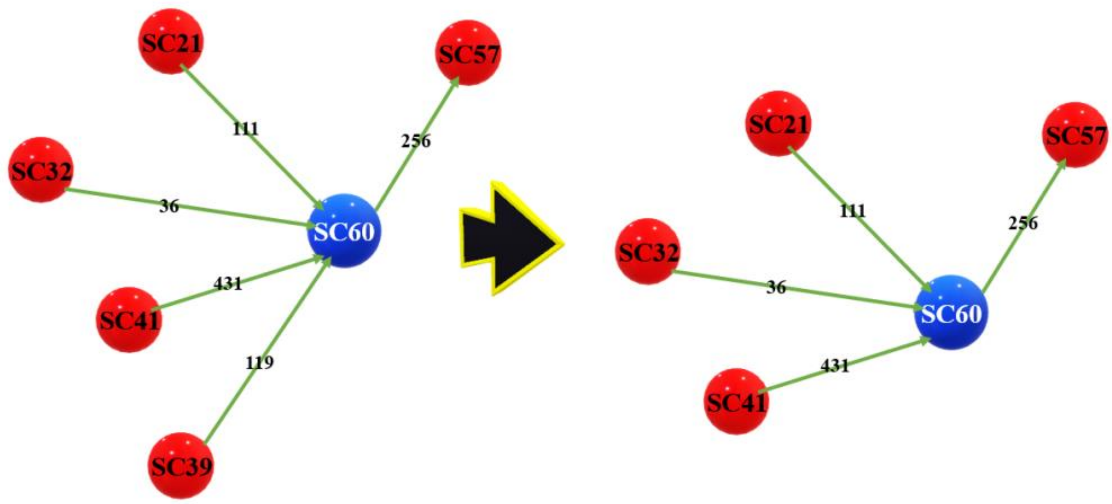

Figure 6. Node Transfer Diagram of Logistics Sorting Center SC60

Figure 6 shows the state transition changes of center SC60. After the calculation, it was found that the transferred logistics sorting center should not include SC39, reflecting the update in node strength.

4.3. Cargo Volume Prediction Model Based on the Weighted State Transition Matrix

The adjustment of the cargo volume transported between sorting centers is represented by the ratio of the adjusted node strength to the original node strength for each transportation route. A state transition matrix H is established, which satisfies the following conditions.

$$\begin{cases} H = (h_{ij})_{n\times n} & n = 57 \\ h_{ij} = \begin{cases} \frac{S_i'}{S_i} & i = j \\ 0 & i \neq j \end{cases} & i,j \in \mathbb{Z} \end{cases} \tag{24}$$

Each row of the matrix represents a vector that characterizes the change in weight values of the corresponding logistics sorting center's features before and after route adjustments. This can be used to weight the prediction data, thereby obtaining the sorting volume of the logistics center after route adjustments.

Let X be the matrix representing the predicted cargo volumes for each of the 57 logistics sorting centers over the next 30 days before route adjustments, and Y be the matrix representing the cargo volumes for each of the 57 sorting centers over the next 30 days after route adjustments, which needs to be determined. According to the principle of matrix multiplication, the state transition equation for the daily cargo volume matrix is as follows.

$$\sum_{i=1}^{57}\sum_{j=1}^{30} y_{ij} = \sum_{i=1}^{57}\sum_{j=1}^{57}\sum_{k=1}^{57}\sum_{l=1}^{30} h_{ij} x_{kl} \tag{25}$$

Similarly, the corresponding cargo transfer matrix for each hour over the next 30 days satisfies

$$\sum_{i=1}^{57}\sum_{j=1}^{720} y'_{ij} = \sum_{i=1}^{57}\sum_{j=1}^{57}\sum_{k=1}^{57}\sum_{l=1}^{720} h'_{ij} x'_{kl} \tag{26}$$

Using the cargo volume prediction model from Problem 1 and the dataset from Attachment 1, the future daily cargo volumes for the next 30 days for the 57 sorting centers are predicted. By multiplying X_1 with the state transition matrix H_{ij}, we obtain the matrix Y_1, which represents the adjusted daily cargo volumes for the next 30 days.

Similarly, using the dataset from Attachment 2 ("Hourly Cargo Volume for Each Sorting Center Over the Past 30 Days"), we predict the hourly cargo volumes for the next 30 days for the 57 sorting centers, resulting in the matrix X_2. After multiplying it by the state transition matrix, we obtain the matrix Y_2, representing the adjusted hourly cargo volumes for the next 30 days.

Finally, using the logistics sorting center SC1 as an example, the obtained data is shown in the Figure 7[16].

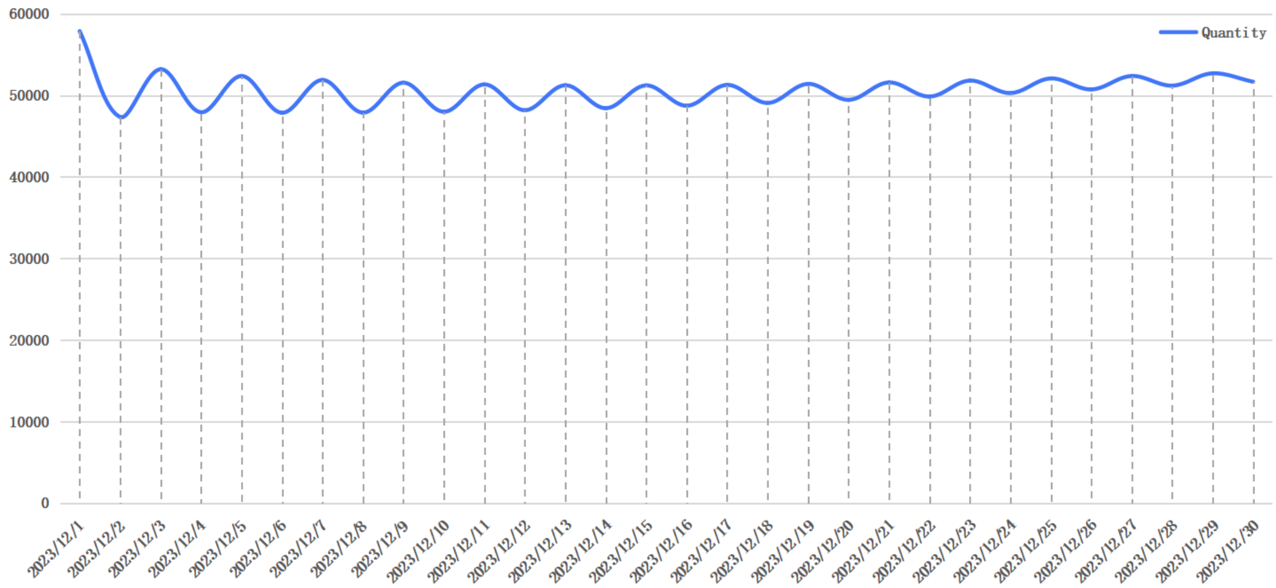

Figure 7. Logistics Sorting Center SC1 Cargo Volume Prediction Results for December

Based on Figure 7, the cargo volume forecast for the logistics sorting center SC1 shows that its distribution aligns with the fluctuating trend. The data's variation is consistent with stable fluctuations. Under this distribution, cargo transportation can effectively regulate the market, managing both surpluses and shortages, facilitating the operation of the logistics sorting center.

5. Conclusion

This study addresses the logistics network problem by first preprocessing the given data, including sorting, removing missing values, and applying wavelet transform for denoising. The ARIMA model is then used for time series forecasting, predicting daily and hourly cargo volumes for the next 30 days for each sorting center. Next, a directed acyclic graph model is constructed based on the complex relationships between sorting centers, incorporating a distributed consensus algorithm and updating the state transition matrix. Combined with the time series forecasting model, future cargo volumes are predicted.

For example, the predicted cargo volume for SC1 on December 1, 2023, is 57868 kg, and on December 30, is 51661 kg. This data falls between the maximum and minimum values of the early cargo volume data for center SC1, fully reflecting the stable fluctuations in the data. The logistics sorting center SCi can use the predicted data from the above model to effectively regulate surpluses and shortages, achieving smooth control and development of production and transportation.

This study presents a research approach and framework that combines time series forecasting with a network graph model, which can be widely applied in supply chain management. The use of wavelet transformation for denoising, the accurate prediction of the ARIMA model, and the introduction of a distributed consensus algorithm based on directed acyclic graphs, along with the calculation of node strength and updating of the state transition matrix, resulting in more accurate cargo volume predictions. When compared with actual scenarios, the application proves highly effective, demonstrating the feasibility of the overall method and significantly contributing to logistics networks and production management.

References

[1] Gurnak V ,Volynets L ,Khalatska I .Intellectualization of logistic supply chains on the basis of forecasting volumes of cargo transportation [J].MATEC Web of Conferences,2019,29404013.

[2] Chavan P P , Singh M .Intelligent Image Compression Model on the Basis of Wavelet Transform and Optimized Fuzzy C-Means-Based Vector Quantisation [J].Journal of Information & Knowledge Management,2024.

[3] WANG Hongjuan , SHANG Yingying. Optimization of speech signal denoising algorithm based on wavelet transform [J].Electroacoustic Technology,2024,48 (05):67-69.

[4] Naieni H K ,Sepandi M ,Eshrati B , et al.Comparative performance of hybrid model based on discrete wavelet transform and ARIMA models in prediction incidence of COVID-19 [J].Heliyon,2024,10 (13):e33848-e33848.

[5] Wang Taiyue,Dai Yanqing. Digital image denoising algorithm based on wavelet transform [J].Journal of Hubei University of Technology,2022,38 (05):25-30+55.

[6] LI Weisong,XU Weijie,ZHANG Tao. Improvement of threshold denoising algorithm based on wavelet transform [J].Computer Simulation,2021,38 (06):348-351+356.

[7] Weng Y ,Wang Z .ARIMA time series based logistics route cargo volume forecasting research[J].Academic Journal of Computing & Information Science,2023,6 (8).

[8] Bhambu A ,Gao R ,Suganthan N P .Recurrent ensemble random vector functional link neural network for financial time series forecasting [J].Applied Soft Computing,2024,161111759.

[9] Artemios-Anargyros S ,Evangelos S ,Vassilios A .Data augmentation for univariate time series forecasting with neural networks [J].Pattern Recognition,2023,134.

[10] WEI Lingxiang, DONG Jianjun, CHEN Zhilong, et al. Time series prediction model of urban freight volume based on correlation vector machine [J].Journal of Wuhan University of Technology (Transportation Science and Engineering),2021,45 (01):88-92.

[11] JIANG Jian,HE Shiwei. Prediction of the whole society's freight turnover based on improved time series method [J].Railway Freight,2007, (06):6-8+45.

[12] PAN Zhengtong. China Storage and Transportation,2022, (01):85-86.

[13] HUANG Xiaoge, WU Yuhang, YIN Hongbo, et al. Personalized federated mutual distillation learning method based on directed acyclic graph blockchain in Internet of Vehicles [J].Journal of Electronics & Information Technology,2024,46 (07):2821-2830.

[14] YU Kefan, ZHU Qing, DING Yulin, et al. Dynamic Arrangement Method of Landslide-Dammed Lake-Flood Disaster Chain Model Based on Directed Acyclic Graph [J].Surveying and Mapping,2023,46 (05):211-216.

[15] YANG Jiahao.Research on distributed consensus algorithm based on directed acyclic graph[D].North China University Technology, 2022.

[16] Li Hailang. Research on demand forecasting and functional area layout planning of Xuzhou railway logistics center [D].China University of Mining and Technology, 2023.