Optimization research on logistics network cargo volume prediction based on ARMA model

Zhicheng Wang [1, *], Kairen Xue [2], Junlei Xu [2]

[1] International Engineering College, Xi'an University of Technology, Xi'an, China, 710054

[2] School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, China, 710054

* Corresponding Author Email: OnlyOswald@outlook.com

Abstract. With the rapid development of power grid logistics, predicting the cargo volume of logistics network sorting centers and reasonable personnel scheduling play a crucial role in their development. This article studies the problem of cargo volume prediction and personnel scheduling in logistics network sorting centers. A time series prediction model, logistics graph and network analysis model, and 01 integer programming model are established to obtain a more reasonable cargo volume prediction and personnel scheduling plan. To predict the daily and hourly cargo volume of 57 sorting centers for the next 30 days, an ARMA time series prediction model was established based on time series data. Firstly, the original sequence was preprocessed, and then the ARMA model was used to predict and analyze the daily and hourly cargo volume of each sorting center. Finally, a goodness of fit test was conducted on the model, and it was found that the model passed the test. When there are changes in the transportation routes between sorting centers in the next 30 days, this article establishes a graph and network model to analyze the transportation routes, analyze the changed routes, and finally dynamically adjust the predicted daily and hourly cargo volume for the next 30 days based on the built model. After calculation, the final dynamic adjustment result is obtained.

Keywords: Time Series, Integer Programming, Dynamic Adjustment, Personnel Scheduling.

1. Introduction

Create a fast and convenient consumption and transportation environment for e-commerce. Online shopping has gradually become popular among consumers, indicating that building a more intelligent logistics network is particularly important. Under various online preferential policies such as holidays and Double Eleven, the number of logistics packages will increase sharply at a specific time. Therefore, it is very important to allocate personnel reasonably for scheduling and predict the cargo volume of logistics network sorting centers. [1] In the past, companies often relied on traditional time series prediction methods, but in practical applications, these methods often exhibited limited prediction accuracy. This article analyzes and compares historical data of logistics centers, and applies the ARMA model to establish a demand forecasting model. [2] Firstly, the original data is visualized and analyzed, and it is found that the original sequence is not stationary. The periodicity and trend are eliminated through first-order difference processing. Secondly, this article uses Daniel's test to test the difference data and determine which one is more stable between the original sequence and the first-order difference sequence. Then, pattern recognition is used to determine which model is more suitable for solving the problem. After establishing the model, the model is bounded and the freight volume of each sorting center in the next 30 days is predicted. Finally, the model used in this article is tested to determine its goodness of fit. [3]

2. Research on stationary time series prediction

2.1. Stability test

In Spearman rank correlation test, it is not necessary to consider the real data of the time series, and the quantity is only 4 or more [4]. The Daniel test is a statistical analysis method based on the Spearman rank correlation coefficient. The general idea of this method is to calculate T from the

Spearman rank correlation coefficient Q of the time series $X_{\theta,t}$ at the significance level, and construct a statistical measure $(t, R_{\theta,t}) t = 1,2,\cdots,n$. If $|T| \le t_{\frac{\alpha}{2}}(n-2)$ is used, H0 is accepted, indicating that sequence $X_{\theta,t}$ is a stationary sequence.

$$q_s = 1 - \frac{6}{n^2(n-1)} \sum_{i=1}^{n} \left(t - R_{\theta,t}\right)^2 \tag{1}$$

$$T = \frac{q_s\sqrt{n-2}}{\sqrt{1-q_s^2}} \tag{2}$$

Among them, $R_{\theta,t} = \{max(X_{\theta,t}) - min(X_{\theta,t})\}$ performs a stationarity test on the data after the first-order difference, and for a significance level of$\alpha = 0.05$, the calculation results are shown in Table 1 and Table 2:

Table 1. Results of stationarity test for the first-order differential sequence of sorting center SC1

/	q_s	$\|T\|$	$t_{\frac{\alpha}{2}}(n-2)$	Stationarity
Original sequence$X_{\theta,t}$	0.037	3.096	1.96	No, it's not
First-order differential sequence$X^{(1)}{}_{\theta,t}$	-0.015	1.219	1.96	yes

Table 2. Results of stationarity test for first-order differential sequences in sorting centers

/	q_s	$\|T\|$	$t_{\frac{\alpha}{2}}(n-2)$	Stationarity
Original sequence$X_{\theta,t}$	-0.369	10.642	1.96	No, it's not
First-order differential sequence$X^{(1)}{}_{\theta,t}$	0.019	0.497	1.96	yes

Therefore, the first order difference data should be used to establish the model.

2.2. Model recognition

The second-order characteristics of the$ARMA(p,q)$ process, AR (p) process, and MA(q) process are shown in Table 3. This article can identify the model by calculating the tailing and truncation of the autocorrelation function (ACF) and partial autocorrelation function (PACF).

Table 3. Properties of sequences

model	$ARMA(p,q)$	$MA(q)$	$AR(p)$
Autocorrelation function(ACF)	Trailing	$Q - order\ truncation$	Trailing
Partial autocorrelation function($PACF$)	Trailing	Trailing	$Q - order\ truncation$

The following are the calculation formulas for the autocorrelation function (ACF, Equation (1)) and the partial autocorrelation function (PACF, Equation (2)) [5]:

$$\gamma_k = \frac{\sum_{t=1}^{n-k}\left(X_t^{(1)} - \bar{X}\right)\left(X_{t+k}^{(1)} - \bar{X}\right)}{\sum_{t=1}^{n}\left(X_{t-1}^{(1)} - \bar{X}\right)^2} \tag{3}$$

$$\begin{cases} \gamma_1 \quad , \quad k = 1 \\ \dfrac{\gamma_k - \sum_{j=1}^{k-1}\varphi_{k-1,j}\cdot\gamma_{k-j}}{1 - \sum_{j=1}^{k-1}\varphi_{k-1,j}\cdot\gamma_k} \quad , \quad k = 2,3,\cdots \\ \varphi_{kj} = \varphi_{k-1,j} - \varphi_{kk}\varphi_{k-1,k-j}, j = 1,2,\cdots,\mathrm{k}-1 \end{cases} \tag{4}$$

Using Python to calculate the library functions in the class library, the results are as follows:

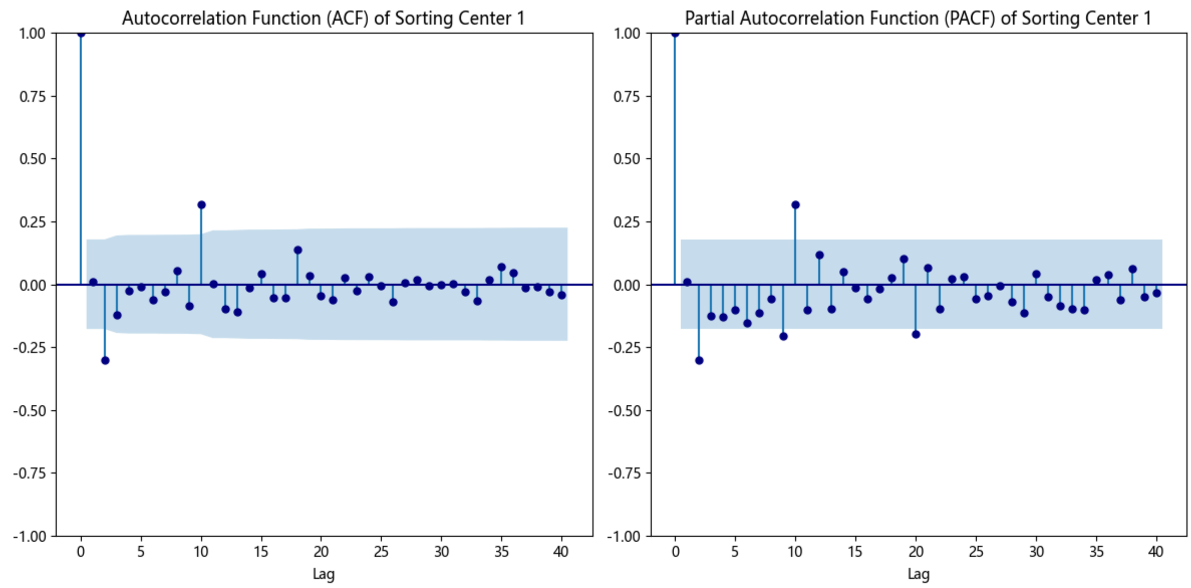

Figure 1. Autocorrelation coefficient and partial autocorrelation coefficient of SC_1 first-order differential sequence

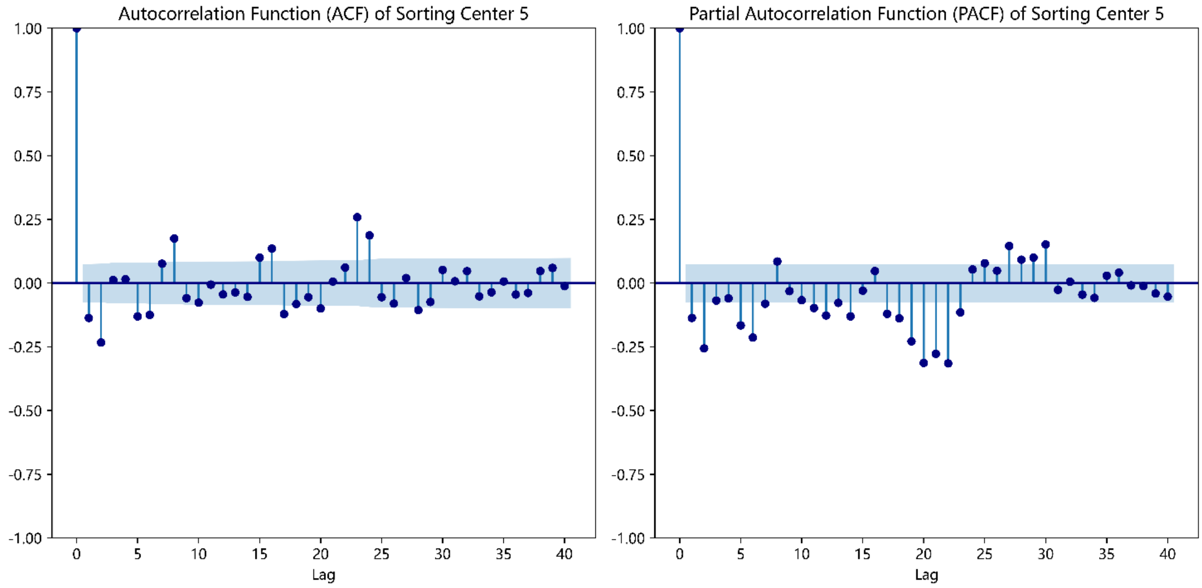

Figure 2. Autocorrelation coefficient and partial autocorrelation coefficient of SC_5 first-order differential sequence

From figure. 1 and 2, the autocorrelation coefficient of the first-order difference sequence has a significant tail effect, that is, as time increases, the corresponding coefficient value gradually approaches 0. The tail of the partial autocorrelation coefficient graph oscillates steadily around zero, without obvious truncation or tailing [6]. Therefore, the $ARMA(p,q)$ model is used for prediction.

2.3. Order determination of the model

In the previous step, the model recognition has been determined to use $ARMA(p,q)$, where $p,q \neq 0$ is used for predictive analysis. The key to the next problem is to determine the order [7], which is the determined value. The penalized AIC criterion contains more parameters. Here, the penalized AIC ordering criterion is used to uniformly determine the values of p and q in model $ARMA(p,q)$ for all sorting center $SC_\theta(\theta = 1,2,\cdots,57)$ [8]. The specific form of the function is:

$$AIC(\mathrm{S}) = 2\ln\hat{\sigma} + \frac{2S}{N} \tag{5}$$

Among them, S is the total number of unknown parameters in the model, $\hat{\sigma}^2$is the estimated value of standard deviation, N is the total number of samples, and $S = p + q$.

p, q. The determination of q is the minimum value point p, q of AIC (S), denoted as $p, q \in [0,3]$. Tables 4 and 5 below show the calculated values of AIC and BIC:

Table 4. Values calculated from first-order difference sequences over the past four months

p	q	AIC	BIC
0	0	130902.8	130909.6
0	1	128492.2	128505.9
0	2	128293.1	128313.6
0	3	127986.3	128013.7
1	0	129281.5	129295.2
1	1	128204.1	128224.7

Table 5. Calculated p and q values for hourly first-order difference sequences over the past 30 days

p	q	AIC	BIC
0	0	542100.7	542109.1
0	1	515559.4	515576.2
0	2	515064.9	515090.1
0	3	514790.2	514823.9
1	0	531326.2	531343.1
1	1	515134.7	515159.9

This article calculates the first order difference sequence from the past four months, and when $p = 2, q = 1$, the corresponding AIC value is the smallest. Therefore, $ARMA(2, 1)$ is used to predict the cargo volume of each sorting center in the next 30 days. Based on the first-order differential sequence of the past 30 days per hour, it was calculated that at$p = 3, q = 2$, the corresponding AIC value was the smallest. Therefore, $ARMA(3, 2)$ was used to predict the hourly cargo volume of each sorting center for the next 30 days.[9]

2.4. parameter estimation

Let time series$X_{\theta,t}^{(1)}(t = 1,2,\cdots, n_1)$ be the measured data after first-order difference, and then use $ARMA(2, 1)$and$ARMA(3,2)$ to fit the model, as shown in equations (6) and (7). Calculate the parameter estimates σ^2 for the autoregressive coefficient vectors $\varphi = (\varphi_1, \varphi_2)^T$ and $\varphi = (\varphi_1, \varphi_2, \varphi_3)^T$ moving average coefficient vectors $w = (w_1)^T$ and $w = (w_1, w_2)^T$, as well as the random sequence variance $\hat{\varphi}, \hat{w}, \hat{\sigma}^2$.

$$\begin{cases} X_{\theta,t} - \varphi_1 X_{\theta,t-1}^{(1)} - \varphi_2 X_{\theta,t-2}^{(1)} = Z_{\theta,t} - w_1 Z_{\theta,t-1} \\ \{Z_{\theta,t}\} \sim WN(0, \sigma^2) \end{cases} \tag{6}$$

$$\begin{cases} X_{\theta,t} - \varphi_1 X_{\theta,t-1}^{(1)} - \varphi_2 X_{\theta,t-2}^{(1)} - \varphi_3 X_{\theta,t-3}^{(1)} = Z_{\theta,t} - w_1 Z_{\theta,t-1} - w_2 Z_{\theta,t-2} \\ \{Z_{\theta,t}\} \sim WN(0, \sigma^2) \end{cases} \tag{7}$$

First, calculate the Yule Walker estimates $\hat{\varphi} = (\hat{\varphi}_1, \hat{\varphi}_2)^T$ and $\hat{\varphi} = (\hat{\varphi}_1, \hat{\varphi}_2, \hat{\varphi}_3)^T$ for the autoregressive coefficient vector, and then calculate the estimates$\hat{w} = (\hat{w}_1)^T$and $\hat{w} = (\hat{w}_1, \hat{w}_2)^T$for the moving average coefficient vector and $\hat{\sigma}^2$ for the random sequence variance, respectively. The calculated$ARMA(2, 1)$ model result for sorting center SC_1 is:

$$X_{1,t}^{(1)} = 0.658 X_{1,t-1}^{(1)} - 0.278 X_{1,t-2}^{(1)} + 0.760 Z_{1,t-1} \tag{8}$$

$$X_{1,t} = X_{1,t}^{(1)} + C_{1,t-1} \tag{9}$$

For this, only $ARMA(3,2\ )$of the sorting center SC_1 is written. The model result is:

$$X_{1,t}^{(1)} = -0.040X_{1,t-1}^{(1)} - 0.999X_{1,t-2}^{(1)} - 0.043X_{1,t-3}^{(1)} + 0.007Z_{1,t-1} - 0.988Z_{1,t-1} \tag{10}$$

$$X_{1,t} = X_{1,t}^{(1)} + C_{1,t-1} \tag{11}$$

Among them, $X_{1,t-1}^{(1)}$, $X_{1,t-2}^{(1)}$represents the data after the first order difference of $t-1$, $t-2$ and period, $X_{1,t}$ represents the predicted value of period, $\{Z_{1,t-1}\}\sim WN(0,\sigma^2)$, $C_{1,t-1}$ represent the historical data of the $t-1$ period of sorting center SC_1.

2.5. Model validation

After establishing a model based on a stationary time series, it is necessary to conduct model validation. This article uses goodness of fit test, which is a statistical significance test for constructing χ^2 statistics, as shown in formula (12).

$$\chi^2 = \frac{(f_0-f_e)^2}{f_e} \tag{12}$$

Where f_0 is the observed frequency, and f_e is the expected frequency.

This article re predicts the last 10 days of the original data, and performs a goodness of fit test between the obtained data and the observed values. Given $\alpha = 0.1$, through calculation, $\chi^2 = 234.396$ is obtained. Looking up the table, it can be seen that $\chi^2_{0.1}(10-1)$ =21666. Therefore, it is not in the rejection domain, therefore H_1 is rejected and H_0 is accepted, indicating that there is no significant difference between the observed value and the predicted value, indicating that the fit is considered good and the model passes the test.

2.6. Prediction results

The following are partial solution results for the next 30 days (SC_1~SC_{16}),as shown in fig.3.

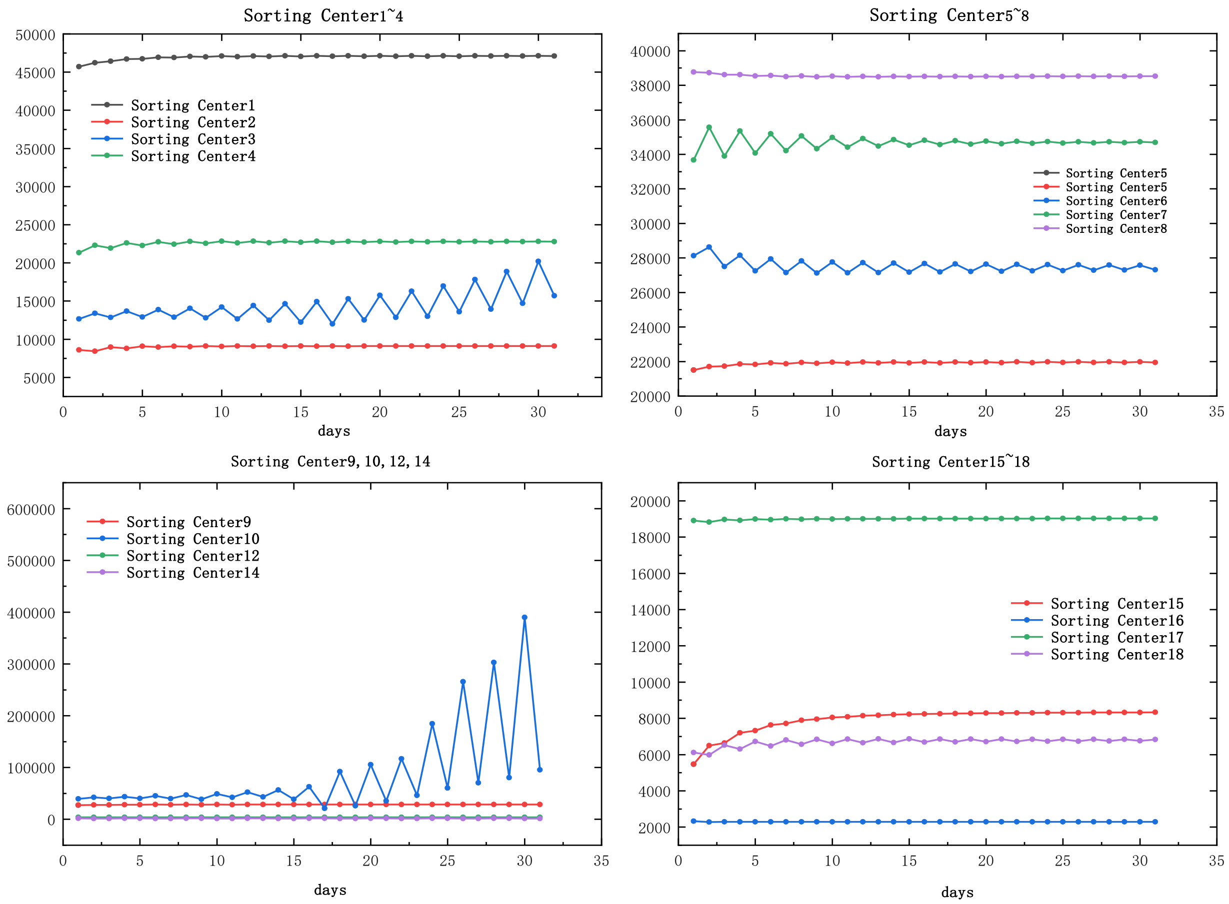

Figure 3. Partial daily forecast results for the next 30 days

Partial solution results per hour for the next 30 days:

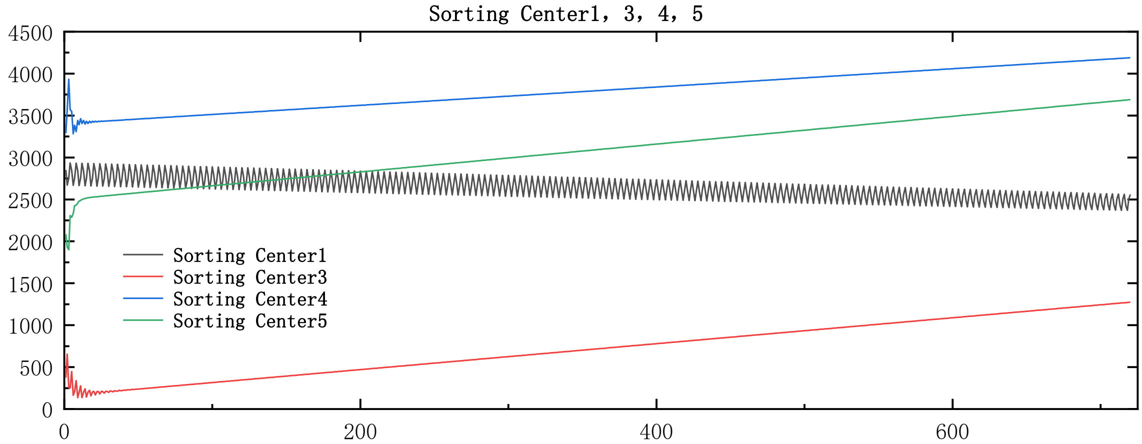

Figure 4. Partial hourly forecast results for the next 30 days

From fig. 4, the cargo volume of each sorting center in the next 30 days changes relatively steadily, which is consistent with the characteristics of time series prediction. Thus, the model establishment and solution are all completed.[10]

3. Conclusion

This article has 57 sorting centers. Given the daily cargo flow of each sorting center for the past 4 months and its hourly or hourly cargo flow for 30 days, a prediction model is established to predict and analyze the cargo flow of these 57 sorting centers for the next 30 days. Time series analysis can be well used to solve dynamic data sets that are interrelated with each other. It can also accurately analyze and predict data that changes over time. The integer programming model is an extension of linear programming, which is characterized by the fact that the set decision variables can only be rounded to integer values. This model is often used for planning problems such as resource allocation and personnel allocation, and can more accurately consider actual situations. Graph theory and network planning models can effectively handle various highly complex network problems, including different types of nodes, connections, and constraints. This enables the model to be applicable to various practical situations and meet different optimization objectives.

References

[1] Ma Tingwei. Research on order picking strategy of logistics center [D]. Beijing University of Posts and Telecommunications, 2015.

[2] Tian Tang, (2021) Supply Chain Demand Forecasting Model based on ARIMA and BP Neural Network and its comparative analysis Ying Yong Shu Xue Jin Zhang, 10 (06), 2041 - 2049 https: //doi.org/10.12677/aam. 2021. 106214.

[3] Liŭ, D. (2024). Enterprise Digital Retail Business data analysis and Forecasting based on Time series analysis. Advances in Economics, Management and Political Sciences, 77 (1), 206 – 212. https: //doi.org/10.54254/2754-1169/77/20241678 (arma).

[4] Plante, J. (2017). Rank correlation under categorical confounding. Journal of Statistical Distributions and Applications, 4 (1). https: //doi.org/10.1186/s40488 - 017 - 0076 - 1.

[5] Alitasb, G. K. (2024). Assessment of fractional and integer order models of induction motor using MATLAB/Simulink. Modelling and Simulation in Engineering, 2024, 1 – 11. https: //doi.org/10.1155/2024/2739649.

[6] Zhang Zhaochi. A Study on Predicting the Number of College Entrance Examinations Based on ARMA Model: A Case Study of H City [J]. Heilongjiang Science, 2023,14 (05): 92 - 95.

[7] LIANG G, WILKES D M, CADZOW J A. ARMA modelorder estimation based on the eigenvalues of the covariance matrix [J]. IEEE Transactions on Signal Processing, 1993, 41 (10): 3003 - 3009.

[8] Tang Minggang, Qi Min, Wang Dajun, etc Application of ARMA model in accuracy analysis of radar dynamic measurement [J]. Modern Radar, 2019, 41 (05): 77 - 81.

[9] ZHENG Zhuo, CHEN Ao, LI Yueju. An improved ARIMA prediction model for ship systems [J]. Ship and Ocean Engineering, 2023, 39 (06): 29 - 34+55.

[10] Wang, Y. (2023). Overview of logistics demand forecasting methods. Frontiers in Business, Economics and Management, 9 (2), 251 – 255. https: //doi.org/10.54097/fbem.v9i2. 9293.