Port Throughput Forecast Based on ARIMA Model-Take Tianjin Port as an Example

Minda Li *

Navigation College, Dalian Maritime University, Dalian 116026, China

* Corresponding Author Email: lmd040827@dlmu.edu.cn

Abstract. Port throughput has long been a crucial metric for assessing the port's capability and even the region's development, so it is very important to forecast the port throughput. This article will use a larger data set than previous papers - 275 sets of data from Tianjin Port from 2001 to 2023 and conduct predictive analysis through the ARIMA model. The parameters of the model were established by ADF, PACF, and ACF, and the data were adjusted. The data can be effectively fitted by the ARIMA model according to the results. and it is also demonstrated that both the original data and the fitted data exhibit annual periodicity. Therefore, the author made further predictions for 12 periods of data and observed that it also exhibited a pattern of initial growth followed by decline, which is consistent with the trends in the shipping industry. The practical significance of this research is to improve port efficiency and achieve accelerated port development.

Keywords: Port throughput; shipping; ARIMA model.

1. Introduction

The shipping industry is an indispensable part of global transportation, and as a hub connecting land and water transportation, ports are particularly important. Port throughput is an important criterion for measuring a port. Port throughput reflects the port's scale and efficiency, its position in the region and waterway, and most importantly, changes in port throughput can directly reflect the port's operating conditions and economic development [1]. Therefore, forecasting port throughput can help port operators grasp port development trends, take effective measures in a timely manner, and better allocate resources. Thereby accelerating the development of the port and promoting each other regional economic development [2].

There are currently many models and algorithms used for prediction. Li proposed an ant colony algorithm to optimize the back propagation neural network (AC-BP) port throughput prediction method, which significantly reduces the mean absolute percentage error (MAPE) [3]. Li introduced the Elman dynamic neural network to predict throughput, which further improved the prediction accuracy compared to BP and RBF neural networks [4]. Li designed a time-varying delay grey system theory model GM(1,1|τi) model and used it to predict the cargo throughput of coastal ports in Fujian Province, improving the dynamic analysis of the system lag effect [5]. Farhan innovatively added seasonal changes when using the seasonal autoregressive integrated model (SARIMA) to predict throughput, making it more consistent with port characteristics [6]. The recent advancements in port throughput prediction, both domestically and internationally, have been summarized by Li, and proposed the research direction of how to predict the throughput of new ports when new ports lack historical data [7].

In addition, the combined use of multiple model algorithms is also a major research direction. Sun combined random forest (RF) with the bidirectional long short-term memory (LSTM) algorithm, by combining the advantages, can enhance the accuracy of predicting Qingdao Port's container throughput [8]. Yu first decomposed the original data through the ensemble empirical mode decomposition (EEMD) algorithm, and then used PSA-optimized parameters (PSO-LSSVR) to predict through intrinsic mode function (IMF) of different frequencies. The first decomposition is integrated, which effectively overcomes the problem of modal aliasing [9]. Li suggested a model for forecasting that incorporates variational mode decomposition (VMD) and gated recurrent units

(GRU), which enhanced the prediction accuracy of inland river ports by improving the decomposition technology and the independence between components [10].

Throughput forecasting also helps address specific time periods, such as the Covid-19 period. The autoregressive integrated moving average (ARIMA) model was employed by Wei to forecast the normal bulk cargo transportation volume of Guangzhou Port in 2020 and compared it with the actual transportation volume affected by the new crown epidemic, calculated the impact index and provided opinions for the subsequent development of the port [11]. Huang proposed a hybrid model configuring S-L2, S-S2, which performs better during the Covid-19 pandemic [12]. To deal with throughput dynamics that are affected by parametric perturbations and external disturbances. Cuong introduced new decision-making strategies that are based on multiple criteria and can be executed with the help of the neural network prediction controller (NNC). Help ports can better respond to abnormal events [13].

Although there are so many prediction algorithms and results, there is still a problem: the data used in the research is not large enough. Unlike other data predictions that can be made based on weeks, days, and seconds, the particularity of port throughput makes its smallest statistical unit is the month, so the insufficient scale of the data may lead to insufficient prediction accuracy. Therefore, this article will use more data and use the ARIMA model to fit the data. Discover predictive accuracy and new possibilities.

2. Methods

2.1. Data Source

Data for this study were taken from the People's Republic of China's Ministry of Transportation's official website. The data contains 275 groups of Tianjin Port throughput and their corresponding months, covering the time span from January 2001 to November 2023. After obtaining the data, the author found that the accuracy of the data has further improved due to the improvement of statistical methods and the development of science and technology. From integers in early years to percentiles in recent years, this may affect subsequent predictions, Therefore, the authors unify the precision integers of the dataset by rounding as shown in Table 1.

Table 1. Data machine accuracy before and after unification (unit: 10,000 tons)

Date	Throughput	Throughput
2019-2	2944.4	2944
2019-3	3871.94	3872
2019-4	4284.496	4284
2019-5	4441.9426	4442
2019-6	4409.1854	4409

2.2. Indicator Selection and Explanation

The development of ports is hindered by numerous factors, such as port facilities, seawater depth, hinterland size, and the connection between land and sea transportation. Despite the difficulty of quantifying these factors numerically, which can be intuitively reflected in the port's throughput. Therefore, the author hopes to make effective predictions by building a time series model based on large amounts of data. Tianjin Port has continued to develop well in the past 20 years and has firmly ranked among the top ten ports in China. The port has a complete infrastructure and is the maritime gateway of Beijing, Tianjin, and Hebei. It has good ability to withstand fluctuations in special events. Therefore, the author takes Tianjin Port as an example.

2.3. ARIMA Model

The ARIMA model combines the AR model's prediction of long-term historical trends with the MA model's ability to handle temporary emergencies and changes, which can better Predict future values based on the properties of a time series. The specific process is shown in Figure 1.

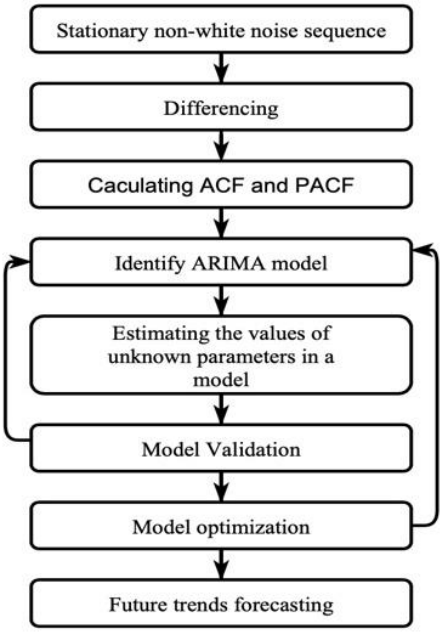

Fig. 1 Process of ARIMA model

In ARIMA (p, d, q), different parameters denote distinct features, " p " denotes the autoregressive order, "d" the difference order, and "q" the moving average order. These three important parameters affect the entire ARIMA model.

The first step is to convert the non-stationary sequence into a stationary sequence through difference operation. The formulas of the first-order and second-order difference operations are as follows:

$$\nabla x_t = x_t - x_{t-1} = x_t - Nx_t = (1-N)x_t \tag{1}$$

$$\nabla^2 x_t = \nabla x_t - \nabla x_{t-1} = (1-N)x_t - (1-N)x_{t-1} = (1-N)^2 x_t \tag{2}$$

The next topic is the p values, which describe the lagged values of the observations used in the model (i.e., the values from the previous p periods). The starting point of the autoregressive model is that the observed value is a linear combination of its previous p values. The general p -order AR model is:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + c + \xi_t \tag{3}$$

Among them, φ is a model parameter, c is a constant, and ξ_t is white noise. The order p of this equation determines the number of observations back in the model.

Note that q stands for "Moving Average". This part describes the lagged value of the error term used in the model (i.e., the value of the previous q periods). The correlation between the current value and past white noise is established by the moving average model. The general q -order MA model is:

$$Y_t = e_t + \mu + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \tag{4}$$

Among them, θ is the model parameter, e_t is the white noise in the current period, and ϵ is the white noise in the past. The order q of this equation determines the amount of white noise backtracking the model.

In summary, the ARIMA model is formed by combining AR and MA and adding the difference operation.

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + c + \xi_t + e_t + +\theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \quad (5)$$

3. Results and Discussion

3.1. Stationary Test

First, to judge whether the data were stable, the author introduced the test method of Augment Dickey fuller (ADF) and assumed that H_0: unit root existed, the sequence was not stable. The result is shown in Table 2. The p=0.096>0.05, fits hypothesis H_0. Thus, the data was processed by first-order difference. At this time, p=0.000<0.05, the null hypothesis was rejected with more than 99% certainty, and the sequence was stable. So, d=1. After that, the author aims to discover the Autoregressive Order p through the autocorrelation function (ACF) and Moving Average Order q through the partial autocorrelation function (PACF).

Table 2. ADF-Test

Differencing Order	t	p	Critical Value		
			1%	5%	10%
0	-2.584	0.096	-3.455	-2.873	-2.573
1	-7.385	0.000	-3.455	-2.873	-2.573

3.2. Model Results

As shown in Figure 2 and 3, neither ACF and PACF has tailing or truncation, so the values of p and q cannot be directly obtained. The authors organized and merged potential p and q values based on actual data, as well as Figures 2 and 3, subsequently comparing the AIC and BIC values of various combinations, the outcomes are displayed in Table 3. Obtained the value of p should be 5 while the value of q should be 7 (Table 3).

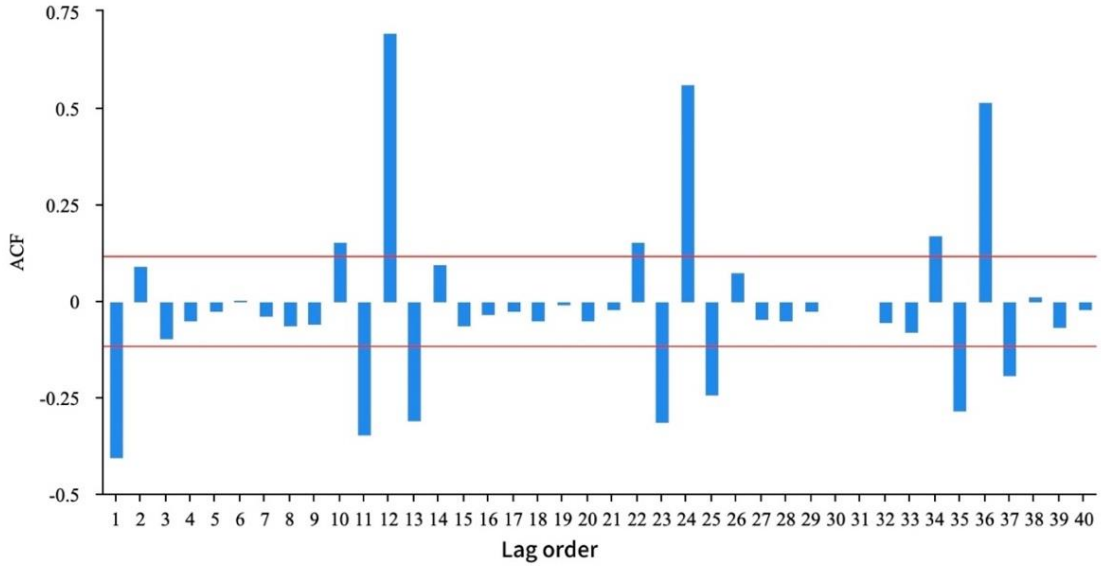

Fig. 2 ACF figure

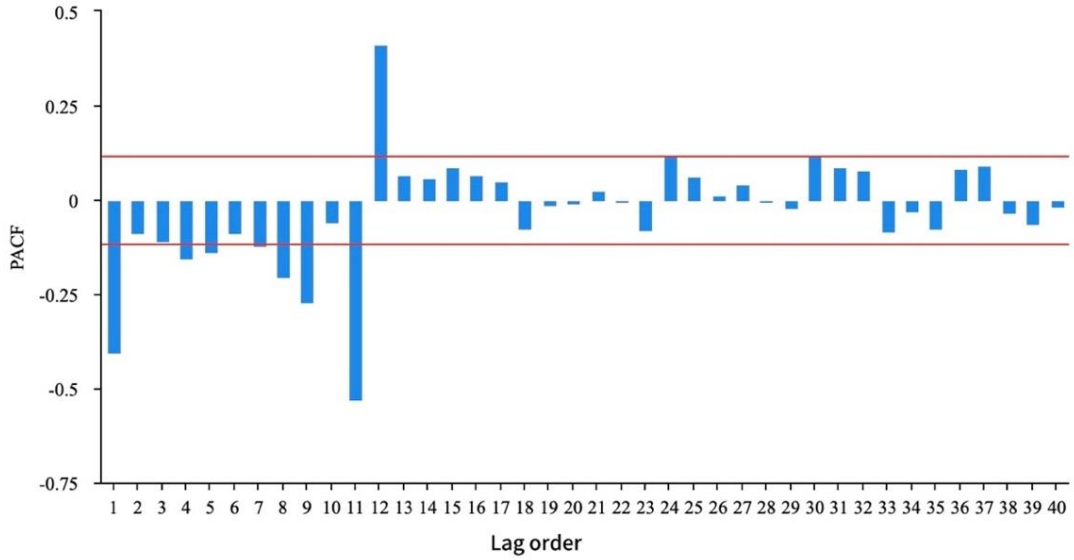

Fig. 3 PACF figure

Table 3. Model Evaluation

ARIMA Model	BIC	AIC
(4,1,6)	4000.31	3956.91
(4,1,7)	4060.41	4013.46
(4,1,8)	3995.61	3944.98
(5,1,6)	4000.26	3953.24
(5,1,7)	3963.53	3909.27
(5,1,8)	3998.46	3947.82
(6,1,6)	3978.32	3927.69
(6,1,7)	3992.42	3938.17
(6,1,8)	3999.69	3941.82

The ARIMA model's formula and specific parameters were obtained after deriving the value of p d q. See Table 4 for details.

$$Y_t = 58.786 - 0.731Y_{t-1} + \cdots - 0.968Y_{t-5} + 0.096\varepsilon_{t-1} + \cdots - 0.046\varepsilon_{t-7} \quad (6)$$

Table 4. ARIMA (5,1,7) Model Parameter

Item	Sign	Coe.	S.E.	Z	P	95%CI
C	c	58.786	19.706	2.983	0.003	20.162 ~ 97.409
AR	α1	-0.731	0.096	-7.599	0.000	-0.920 ~ -0.543
	α2	0.782	0.053	14.841	0.000	0.679 ~ 0.885
	α3	0.787	0.101	7.807	0.000	0.589 ~ 0.985
	α4	-0.696	0.074	-9.391	0.000	-0.842 ~ -0.551
	α5	-0.968	0.046	-21.209	0.000	-1.057 ~ -0.878
	β1	0.096	0.137	0.704	0.481	-0.172 ~ -0.364
	β2	-1.319	0.121	-10.931	0.000	-1.556 ~ -1.083
	β3	-0.414	0.223	-1.854	0.000	-0.852 ~ 0.024
MA	β4	1.204	0.151	7.971	0.000	0.908 ~ 1.500
	β5	0.610	0.134	4.545	0.00	0.347 ~ 0.873
	β6	-0.616	0.093	-6.599	0.000	-0.799 ~ -0.433
	β7	0.046	0.081	-0.567	0.570	-0.206 ~ -0.113

As shown in Figure 4, the predicted value fits the actual value to a high degree, even in the case of cyclical fluctuations, there is a good fit. Therefore, the model can satisfy the forecast of port throughput in the future period.

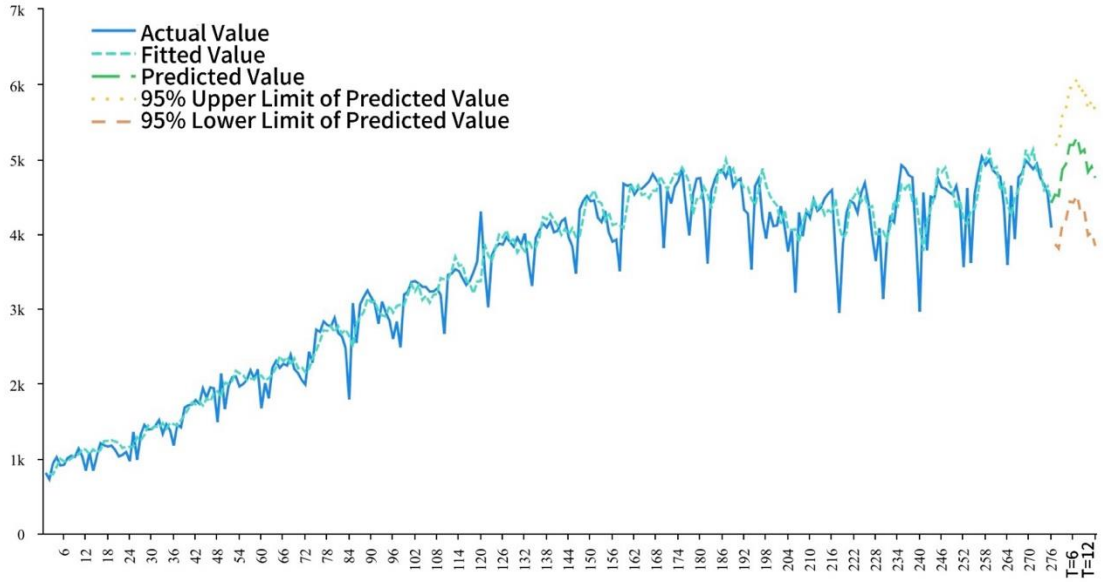

Fig. 4 Data fitting and prediction

The Figure 4 also found that port throughput showed an annual cycle regardless of actual data or forecast data, so the author forecasted 12 periods backward. According to the forecast in Table 5, in the next 12 months, the port throughput of Tianjin Port showed a trend of first rising then declining and reached the maximum value of 52.94 million tons in July. This is in accordance with the actual data and port rules.

Table 5. Prediction results

Prediction	Value	Prediction	Value
T=1	4524	T=7	5294
T=2	4597	T=8	5082
T=3	4859	T=9	5126
T=4	4933	T=10	4825
T=5	5190	T=11	4900
T=6	5187	T=12	4747

4. Conclusion

Through the research of this paper, The ARIMA model has a good performance for large port throughput data Prediction. Also found that port throughput increases and then decreases every 12 months. The above findings can help port decision-makers to know the port throughput in advance, change development strategies in time, and effectively use resources and allocate personnel. During the peak period of the year before the earliest response measures, such as sending more manpower, opening more docks, improve operational efficiency, do a good job of safeguard measures. During the two off-seasons each year, part of the terminal can be closed to avoid wasting resources, and at the same time, additional manpower can be deployed month by month to prepare for the coming peak. Effectively help the port to reduce costs and increase efficiency, sustainable development, but also to drive the economy of the surrounding area.

Hoped that in the future, the actual data and forecast data of major ports around the world can be integrated, so that ports are not limited to their own forecasts but can also refer to ports with similar

geographical locations and similar port sizes to their own ports for comparison and research, which can help with the lack of data in the past. Or find corresponding predictions for newly built ports. Seeking help to promote port and shipping industry development.

References

[1] Munim Z H, Schramm H J. The impacts of port infrastructure and logistics performance on economic growth: the mediating role of seaborne trade. Journal of Shipping and Trade, 2018, 3(1): 1-19.

[2] Santos A M, Salvador R, Dias J C Q, Soares C G. Assessment of port economic impacts on regional economy with a case study on the Port of Lisbon. Maritime Policy & Management, 2018, 45(5): 684-698.

[3] Li C A, Lu X Q, Wu Z Q, Zhang L J. Port throughput prediction based on ant colony algorithm optimizing back propagation neural network. Journal of Metrology, 2020, 41(11): 1398-1403.

[4] Li G R, Zhu Q H. Port cargo throughput prediction based on Elman neural network. Journal of Chongqing Jiaotong University (Natural Science Edition), 2020, 39(6): 8.

[5] Li C, Xie X P. GM(1,1|τi) model with time-varying delay function and its application. Systems Engineering Theory and Practice, 2019, 39(6): 1535-1549.

[6] Farhan J, Ong G P. Forecasting seasonal container throughput at international ports using SARIMA models. Maritime Economics & Logistics, 2018, 20: 131-148.

[7] Li Y, Li T, Zuo Y, Chen C P, Shan Q, Xiao Y, Fan X. A Review of Research on Port Throughput Forecasting. In 2019 IEEE International Conference on Smart Internet of Things (SmartIoT), 2019, 449-453.

[8] Sun X C, Fu Y H. Container throughput prediction based on RF-bidirectional LSTM. Journal of Shanghai Maritime University, 2022, 43(1): 60-65.

[9] Yu T, Sun J Y. Forecasting container throughput of my country's ports based on EEMD and PSO methods. Logistics Technology, 2021, 40(5): 56.

[10] Li X, Zhang D, Chen C, Wan C, Cao J. An Improved VMD-GRU Model for Forecasting the Container Freight Volume of Inland Ports. In 2023 7th International Conference on Transportation Information and Safety (ICTIS), 2023, 615-620.

[11] Wei L, Huang L, Huang H. Research on Forecasting of Port Bulk Freight Volume during the COVID-19 Pandemic Based on Time Series Analysis. In 2022 International Conference on Management Engineering, Software Engineering and Service Sciences (ICMSS), 2022, 132-140.

[12] Huang D, Grifoll M, Sanchez-Espigares J A, Zheng P, Feng H. Hybrid approaches for container traffic forecasting in the context of anomalous events: The case of the Yangtze River Delta region in the COVID-19 pandemic. Transport policy, 2022, 128: 1-12.

[13] Cuong T N, Kim H S, You S S, Nguyen D A. Seaport throughput forecasting and post COVID-19 recovery policy by using effective decision-making strategy: A case study of Vietnam ports. Computers & Industrial Engineering, 2022, 168: 108102.