Chinese Studies, 2018, 7, 12-20
http://www.scirp.org/journal/chnstd
ISSN Online: 2168-541X
ISSN Print: 2168-5428

Prediction of Container Throughput in China

Zhao Jin[1], **Yaozong Ding**[2]

[1]South China University of Technology, Guangzhou, China
[2]Beijing Jiaotong University, Beijing, China
Email: 397042520@qq.com

How to cite this paper: Jin, Z., & Ding, Y. Z. (2018). Prediction of Container Throughput in China. *Chinese Studies, 7,* 12-20.
https://doi.org/10.4236/chnstd.2018.71002

Received: November 23, 2017
Accepted: January 28, 2018
Published: February 1, 2018

Abstract

In this paper, we will use the time series analysis method to predict the throughput of the research, through the establishment of time series SARIMA model, using the January-February 2017 domestic container throughput statistics of the main container port forecast from March 2017 to December 2020 container throughput data and analyze the trend of the throughput of major container ports in China in the coming year according to the forecast results. At the same time, we will use R software to analyze the time series of the throughput of China's major container ports. By decomposing the time series and correcting the seasonal differences of the data, a seasonally revised time series chart is obtained. Based on the above analysis, this paper predicts that by 2020 China's major container port throughput will be at the level of 250 million TEU-270 million TEU.

Keywords

Container Port, Throughput, Time Series Analysis, SARIMA Model

1. Introduction

In the context of world economic integration, international trade is increasingly frequent. As the most widely used means of transport in international trade, container shipping not only promotes the mechanization and automation of transportation production, but also greatly improves the efficiency of international transportation and reduces the transportation cost. Chinese foreign trade activities are very frequent. But in recent years, the changes in the international and domestic transportation environment have caused the balance of supply and demand of container transport to change. On the one hand, with the economic restructuring in our country, the labor-intensive industries are gradually shrinking. The transition of foreign trade products to "technology-based" commodities has led to changes in the demand for container transport. On the other hand, since the 21st century, the construction of large-scale ports in our country has greatly increased the handling capacity of container ports in our country. The supply of container transport has also undergone great changes. Forecasting the throughput of container ports has increasingly become a topic of concern to the industry.

The research of this paper is based on the actual situation of the port, combines the achievements of the relevant scholars in this field and makes innovations and improvements, and uses the time series analysis method to analyze the data of container port transportation dynamically. And we combine the latest transport market situation to analyze the container port transport, which has some guidance on the production of first-line transport.

2. Literature Review

In the actual statistics, the throughput of the container port is a time series of random fluctuations influenced by various factors such as economy and politics. Therefore, the forecast of container port throughput has some difficulties and uncertainties in the research. Many scholars put forward many prediction models and methods in the prediction of container port throughput. The most common prediction methods are curve fitting method, regression analysis method, gray prediction method, nonlinear trend analysis method and neural network method. Jian Ye (2005) introduced four classic series models: time series model, gray series model, regression model and artificial neural network model. On the basis of cluster analysis, the main coastal ports of China are divided into three different growth ports according to the growth characteristics of their respective container traffic volume. The growth of these three types of ports were ordinary growth, accelerated growth and volatility growth, different growth ports in the prediction of container throughput forecast model used is also different. The study predicts the throughput of China's major coastal container ports and summarizes the forecasting methods that should be used for container ports of different growth types. Xing Xu and Xijun Shi (2002) introduced the BP learning and generalization advantages of BP artificial neural network model in forecasting container port throughput in Shenzhen Port. Xidong Zhai (2006) relied on the transportation background of Dalian Port and predicted the container throughput of Dalian Port by GM (1, 1) residual correction prediction model and completing the validation of the model. Xinhua Ma (2010) based on the original data of the container throughput of China's major ports in his research, conducted a prediction study using the spatial state model. In Xiaomeng Zhu's study (2014), the relatively novel time series formed by the complementary advantages of time series analysis and causal analysis. In this model, GDP replaces the time factor in the traditional time series forecasting method as a new explanatory variable. Based on the container throughput data of Wuhan Port in recent ten years, Yamei Peng (2016) used gray forecasting model and regression analysis model to construct a combined model to forecast the container throughput of Wuhan Port in the coming period.

In this paper, the seasonal variation of the time series of container port throughputs is obtained by using the Autoregressive Moving Average Model (SARIMA) and the model established in R software. And we forecast the throughput of China's major container ports in the next 4 years in the future. Based on the analysis of the impact of China's economic development and the throughput of foreign trade on the throughput of container ports, we corrected the results predicted by the model.

3. Model

1) Model description

ARIMA, the differential autoregressive moving average model, consists of three parts. One is autoregression (AR part). The current value of a time series can be expressed as a linear combination of lagged P-period observations; The second is the whole (I (d) part), d is the number of difference required for the time series to become a stationary sequence; Third, the moving average (MA part), that the current value of the time series can be expressed as a linear combination of q-order residuals. The expression of the model is:

$$W_t = C + \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q}$$

W_t : variables after the d-order difference; ε: moving average;

p: autoregressive terms; q: moving average term.

2) Data sources and analysis

The data in this paper come from the data of China Transportation Database. Since China joined the WTO in 2001, we take the monthly data of the container throughput of the major ports in the country from January 2001 to February 2017 as the research object. The data are shown in Table 1.

The throughput data of container ports in China are analyzed, and the form of test is judged from the time series chart, the results shown in Figure 1. It can be seen from the figure that the throughput fluctuates relatively large. Generally, from 2001 to 2017, it has been an upward trend and has some cyclical changes. In February, throughput has reached the lowest throughout the year. In December monthly throughput reached its maximum. Therefore, the time series of throughput consists of three parts: the trend part, the seasonal part and the irregular part.

In this paper, the time series is decomposed into three parts according to the results shown in Figure 2. Figure 2 is divided into four parts, from top to bottom: the original time series and the estimated trend, the season Graphs and irregular graphs.

As can be seen from Figure 2, the estimated trend has steadily risen year by year since 2001 and declines around 2009, but then steadily increased. Figure 2 estimated the seasonal part of the image is stable, indicating that the container port throughput there is a certain seasonal pattern, consistent with the previous analysis.

Table 1. Monthly domestic container port throughput.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2001	188.00	170.60	212.80	212.00	205.40	217.30	221.70	237.30	248.40	246.20	235.70	341.35
2002	247.60	206.10	264.40	282.70	285.60	298.80	304.10	328.00	337.50	319.30	335.70	429.60
2003	342.22	273.50	368.30	372.41	376.31	387.24	393.87	433.13	427.12	438.42	439.30	558.23
2004	421.93	384.98	458.75	481.79	486.86	504.25	506.00	532.91	549.77	550.45	542.98	690.97
2005	573.31	444.05	560.81	607.17	611.67	608.76	637.03	661.55	665.95	662.54	660.73	796.22
2006	688.57	539.89	704.18	736.49	740.39	754.26	782.30	800.83	833.25	827.14	820.87	1006.68
2007	849.29	763.35	803.57	914.21	912.00	949.21	943.87	980.31	1008.06	985.90	989.29	1203.08
2008	1044.06	839.98	1026.67	1050.99	1066.87	1099.57	1090.40	1094.48	1074.55	1071.08	1041.43	1221.84
2009	899.13	706.02	942.24	935.87	1020.49	1003.30	1056.41	1088.87	1107.10	1078.11	1085.23	1129.78
2010	1135.53	947.81	1109.75	1162.85	1243.92	1226.30	1272.42	1312.95	1273.17	1222.32	1309.26	1315.07
2011	1350.47	995.02	1277.93	1327.62	1383.14	1376.16	1441.20	1445.78	1430.23	1386.43	1397.27	1457.04
2012	1363.43	1193.46	1408.47	1439.43	1512.28	1499.85	1510.47	1511.64	1587.65	1508.22	1527.95	1598.28
2013	1549.99	1217.96	1529.03	1573.71	1613.47	1604.09	1649.50	1638.52	1626.12	1594.87	1638.64	1674.24
2014	1656.49	1277.11	1632.11	1675.38	1681.50	1696.19	1704.64	1756.38	1725.95	1737.74	1755.81	1777.93
2015	1784.05	1432.45	1674.14	1762.47	1793.49	1780.36	1792.72	1789.88	1776.54	1744.46	1788.84	1789.68
2016	1808.23	1427.28	1760.20	1816.55	1838.34	1846.39	1881.09	1911.93	1857.41	1847.10	1879.58	1834.95
2017	1858.99	1537.96										

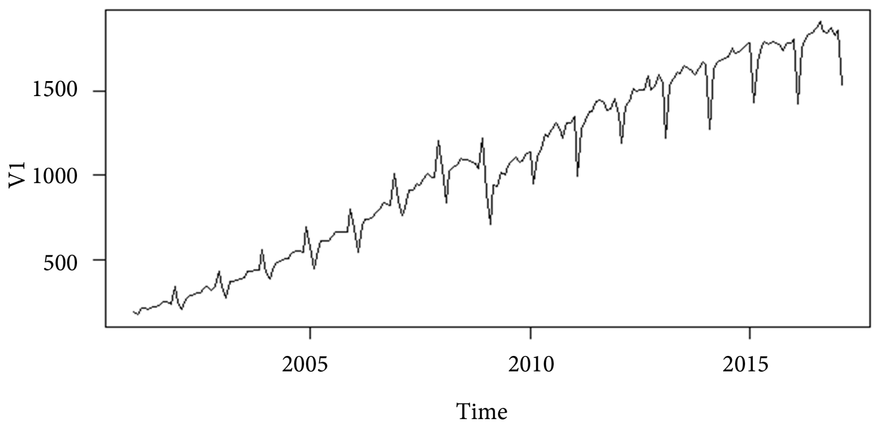

Figure 1. Sequentially.

3) Stationarity test

Unit root test of container throughput data, due to the upward trend in the time series, the unit root test needs to include both intercept and trend items. As shown in **Figure 3**, the ADF test of the throughput time series shows that the P value ratio is much larger than the default p value (0.05), and we can see that the null hypothesis can not be rejected. The sequence is not stable. Therefore, a difference of the throughput sequence is made to obtain a Dttl sequence and an ADF test on the Dttl sequence. As shown in **Figure 4**, a P value of 0.01 is smaller

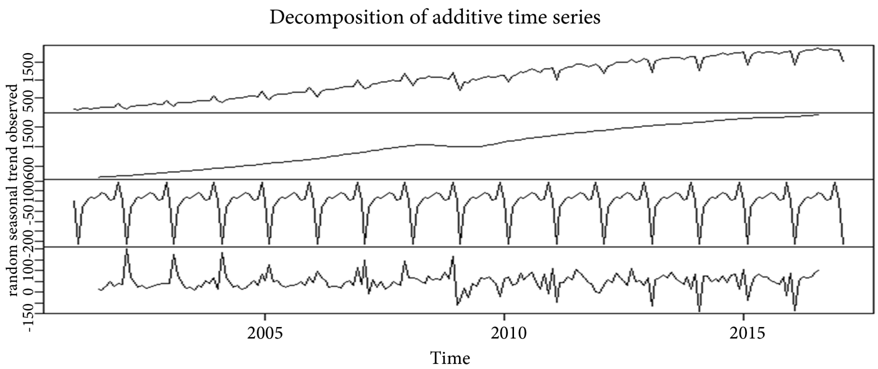

Figure 2. Time series decomposition.

Augmented Dickey-Fuller Test
Test Results:
PARAMETER:
Lag Order:11
STATISTIC:
Dickey-Fuller: 7.855
P VALUE: 0.99
p-value greater than printed p-value

Figure 3. ADT test of throughput ttl time series.

Augmented Dickey-Fuller Test
Test Results:
PARAMETER:
Lag Order:11
STATISTIC:
Dickey-Fuller: -9.2901
P VALUE: 0.01
p-value smaller than printed p-value

Figure 4. D (throughput) ADF test.

than a default P value (0.05), so that the original hypothesis can be rejected and the Dttl sequence obtained is stable.

4) Model estimation and parameter setting

Through the correlation analysis of Dttl (throughput), it is found that in the multiple of 12, as shown in Figure 5, Figure 6, the autocorrelation coefficient is not significant, so the sequence is seasonal and seasonal difference is needed. Therefore, this article selects seasonal ARIMA model or SARIMA model.

After making a seasonal difference adjustment on Dttl, we denote the auto-correlation diagrams and the autocorrelation diagrams of the sequences

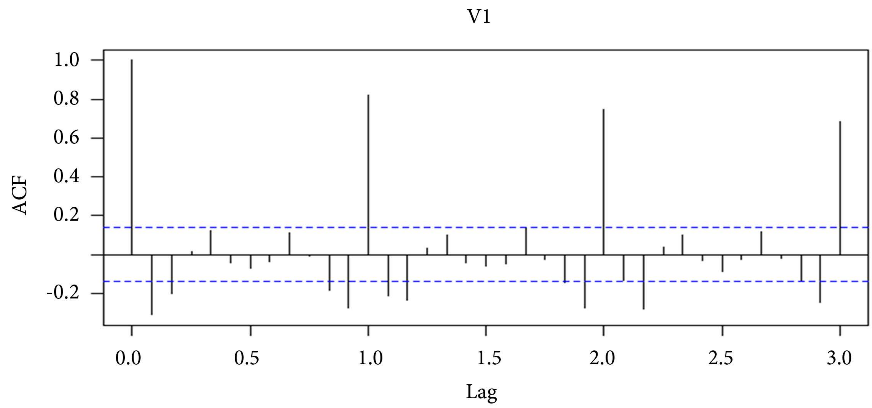

Figure 5. Dttl's autocorrelation function.

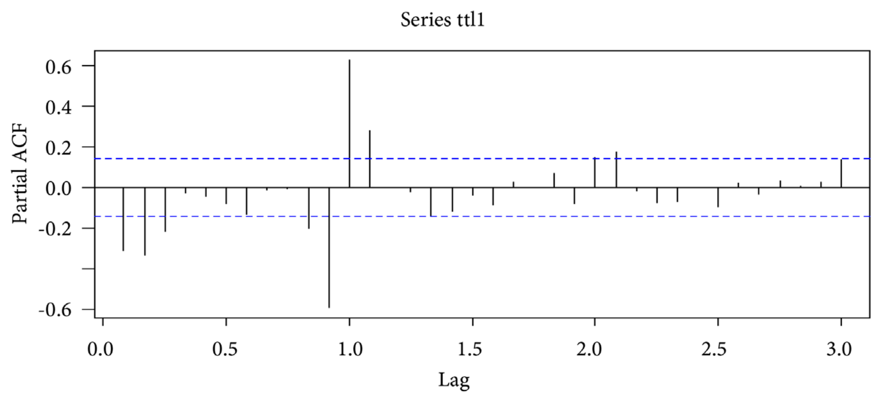

Figure 6. Partial autocorrelation function of Dttl.

SDttl and SDttl, as shown in **Figure 7**, **Figure 8**, we can see that the seasonal autocorrelation gradually disappears after the first order.

According to the autocorrelation and partial autocorrelation of sequences, we found that SARIMA, (2, 1, 2) (2, 1, 1) 12, AIC is the smallest, which is selected according to the AIC principle. Model parameters are shown in **Figure 9**.

5) Residual analysis

Figure 10 shows the residual analysis. It can be seen that the residuals have no autocorrelation and according to Ljung-Box, all P values are greater than 0.5. Therefore, the model is established.

6) Analysis of forecast results

Using the obtained SARIMA model to predict the throughput of container ports nationwide and get the forecast conclusion. As shown in **Table 2**, it can be seen from the results that throughput will continue to increase in the future, decreasing in February of each year and increasing in December.

4. Conclusion

Although China's GDP growth will stabilize at 6.5% during the "13th Five-Year Plan" period, economic and industrial institutions will face a transformation and the throughput growth of major ports will slow down further. However, the

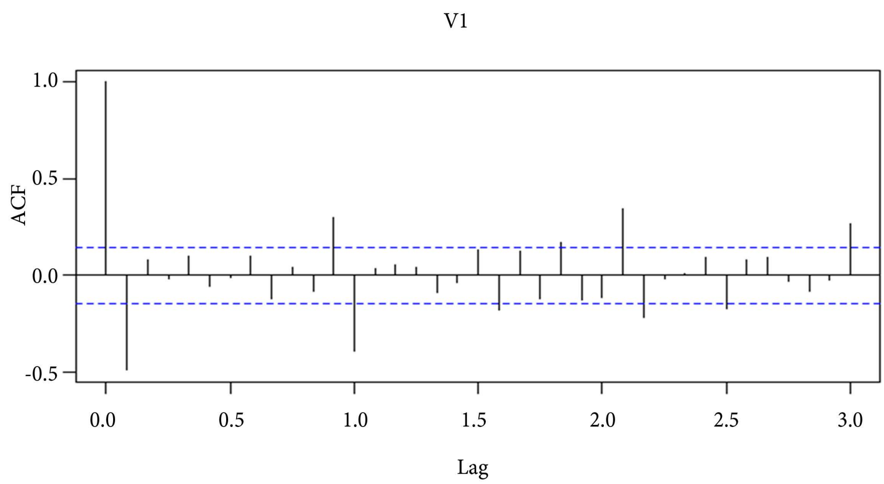

Figure 7. SADttl's autocorrelation function.

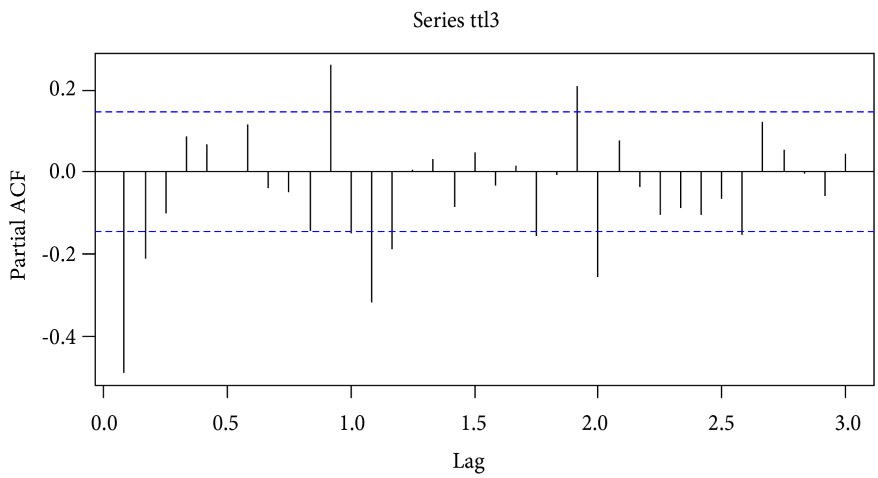

Figure 8. Partial autocorrelation function of SADttl.

ARIMA(x = ttl, order = c(2, 1, 2), seasonal = list(order = c(2, 1, 1), period = 12))

Coefficients:

	ar1	ar2	ma1	ma2	sar1	sar2	sma1
	-0.3095	-0.1349	-0.3113	0.1020	-0.9293	-0.4045	0.4148
s.e.	0.9178	0.2933	0.9221	0.3215	0.2720	0.1124	0.3015

sigma^2 estimated as 1570: log likelihood = -925.2, aic = 1866.4

Figure 9. Model parameters.

restructuring of China's port structure is tending to be rationalized and the proportion of container shipping will further increase. With the further release of the favorable strategy of "One Belt, One Road" and other major strategies, the throughput of China's major container ports will also maintain a relatively steady growth. Based on the above analysis, this paper predicts that by 2020 China's major container port throughput will be at the level of 250 million

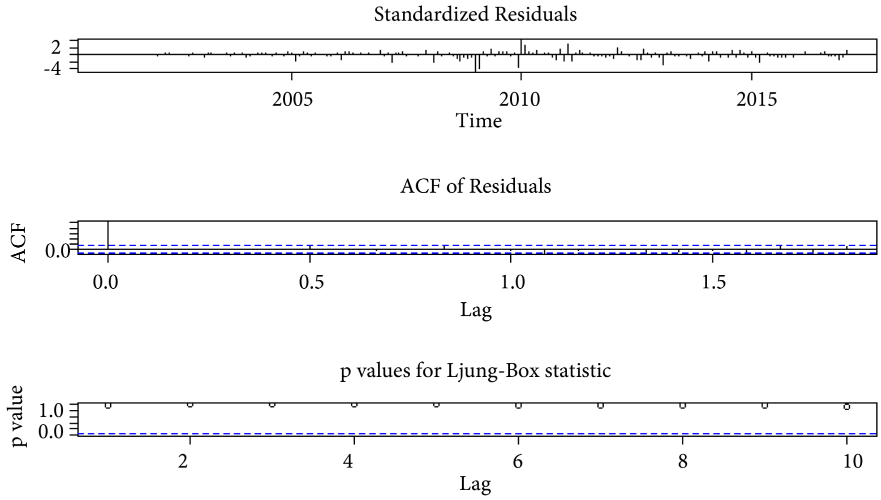

Figure 10. SADttl Residual Analysis.

Table 2. Predicted results.

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Mar-17	1805.988	1754.660	1857.317	1727.488	1884.489
Apr-17	1876.891	1822.622	1931.161	1793.893	1959.890
May-17	1903.077	1842.421	1963.733	1810.312	1995.842
Jun-17	1903.645	1836.591	1970.700	1801.095	2006.196
Jul-17	1927.827	1854.810	2000.844	1816.157	2039.497
Aug-17	1947.261	1868.709	2025.813	1827.126	2067.369
Sep-17	1914.621	1830.894	1998.347	1786.572	2042.669
Oct-17	1897.140	1808.540	1985.740	1761.638	2032.642
Nov-17	1932.018	1838.799	2025.238	1789.452	2074.585
Dec-17	1918.465	1820.844	2016.085	1769.167	2067.762
Jan-18	1933.937	1832.106	2035.769	1778.200	2089.675
Feb-18	1586.992	1481.117	1692.866	1425.070	1748.913
Mar-18	1878.028	1759.900	1996.156	1697.367	2058.689
Apr-18	1944.030	1820.238	2067.822	1754.706	2133.354
May-18	1969.121	1938.610	2099.632	1769.521	2168.720
Jun-18	1970.818	1933.710	2107.926	1761.129	2180.507
Jul-18	1996.685	1853.241	2140.129	1777.306	2216.064
Aug-18	2017.961	1868.441	2167.481	1789.290	2246.633
Sep-18	1981.780	1826.420	2137.140	1744.177	2219.383
Oct-18	1965.460	1804.471	2126.448	1719.248	2211.671
Nov-18	1999.950	1833.522	2166.377	1745.421	2254.479
Dec-18	1981.368	1809.675	2153.062	1718.785	2243.951
Jan-19	1998.227	1821.424	2175.030	1727.830	2268.624

Continued

Feb-19	1655.474	1473.705	1837.243	1377.482	1933.466
Mar-19	1942.788	1747.961	2137.615	1644.826	2240.750
Apr-19	2009.583	1807.991	2211.175	1701.275	2317.891
May-19	2034.851	1825.373	2244.329	1714.482	2355.220
Jun-19	2036.366	1819.073	2253.659	1704.045	2368.687
Jul-19	2061.960	1837.079	2286.840	1718.035	2405.885
Aug-19	2082.938	1850.710	2315.167	1727.775	2438.101
Sep-19	2047.330	1807.977	2286.683	1681.271	2413.389
Oct-19	2030.822	1784.550	2277.093	1654.181	2407.462
Nov-19	2065.375	1812.373	2318.376	1678.442	2452.307
Dec-19	2047.606	1788.050	2307.163	1650.648	2444.564
Jan-20	2064.241	1798.290	2330.192	1657.505	2470.978
Feb-20	1720.810	1448.616	1933.004	1304.525	2137.095
Mar-20	2008.726	1723.521	2293.931	1572.542	2444.910
Apr-20	2075.393	1782.354	2368.432	1627.228	2523.557
May-20	2100.632	1798.801	2402.463	1639.021	2562.242
Jun-20	2102.176	1791.617	2412.736	1627.217	2577.136
Jul-20	2127.814	1808.727	2446.902	1639.812	2615.816

Point	Point	Point	Point	Point	Point
Aug-20	2148.841	1821.440	2476.242	1648.124	2649.557
Sep-20	2113.140	1777.630	2448.650	1600.022	2626.258
Oct-20	2096.662	1753.234	2440.090	1571.434	2621.890
Nov-20	2131.205	1780.037	2482.373	1594.141	2668.269
Dec-20	2113.305	1754.565	2472.045	1564.659	2661.951

TEU-270 million TEU.

References

Jian, Y. (2005). *Study on Port Container Throughput Prediction Based on Cluster Analysis*. Dalian University of Technology.

Ma, X. H. (2010). *Research on the Balance between Supply and Demand of Container Port Throughput in China*. Dalian Maritime University.

Xu, X., & Shi, X. J. (2002). Real-Time Forecasting Model of Shenzhen Port Based on BP Neural Network. *Journal of Hohai University (Natural Science).*

Zhai, X. D. (2006). *Study on Container Container Throughput Prediction Model*. Dalian University of Technology.

Zhu, X. M. (2014). *Study on the Correlation Model of Container Throughput and GDP in Coastal Ports*. Dalian Maritime University.