Container-Throughput Forecasting in the Port of Pointe-Noire Using ARIMA Model

Dimitri D.H. Owassa^{1*}, Jingbo Yin¹

School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiaotong University, China

OPENOACCESS

Accepted September 09,2021 Published September 19,2021

*Corresponding Author: Dimitri D.H. Owassa

DOI:<u>https://doi.org/10.5281/z</u> enodo.5516224 Pages: 01-18

Funding: None

Distributed under Creative Commons CC BY 4.0

Copyright: © The Author(s)

How to cite this article (APA): Owassa D.D.H., Yin J. (2021). Container-Throughput Forecasting in the Port of Pointe-Noire Using ARIMA Model. *North American Academic Research, 4*(9), 01-18. doi: <u>https://doi.org/10.5281/zenodo.</u> 5516224

Conflicts of Interest There are no conflicts to declare.

ABSTRACT

Because the Port of Pointe-Noire (POPN) is the economic lung of the Republic of Congo, its rapid development will help boost the national revenues. The port's container terminal is a key element in the management of the port. Thus, an accurate estimation of the future container traffic is essential to the optimization of its performance. In this study, an ARIMA model was developed to forecast the port's container throughput. Monthly data from 2011 to 2020 were collected, giving a total of 120 observations. The model performance was evaluated using three accuracy measuring indexes. The forecasted values fit well the actual values, demonstrating that the model performed quite well. Finally, the container throughput for the year 2021 was forecasted, and the results showed a slight increase in the container traffic compared to the year 2020, from 928,000 to 957,000 TEUs.

Keywords: PORT OF POINTE-NOIRE, CONTAINER THROUGHPUT, ARIMA MODEL

1 Introduction

The Port of Pointe-Noire (POPN) is situated in the city of Pointe-Noire (See Figure 1), the economic capital of the Republic of Congo. The POPN is the only seaport within the country. It is one of the most modern ports in the Gulf of Guinea and a major stopover in the maritime service

in Central Africa. It is a significant access point to the political capital Brazzaville, and plays a significant role as a transshipment port for the ports of Matadi (Democratic Republic of Congo), Luanda, Cabinda, and Soyo (Angola) (1). The POPN offers particularly attractive navigation and mooring facilities compared to other ports in the sub-region. Its 84-hectare basin opens onto the bay via a 250-meter-wide entrance pass, extended by a 2,500-meter access channel which is dredged to -16 meters. The POPN has infrastructure and equipment that make it one of the largest port platforms on the continent, capable of receiving and handling all types of ships. It is mainly composed of a 32-hectare container terminal with a 4.5-hectare container loading and unloading area linked to a railway line; an eight-hectare wood park; six quay gantries cranes; 18 yard gantries cranes; four Gottwald-type mobile cranes; one cement silo, one wheat silo; a quay length of around 1,800 meters; some refrigeration installations; a half a dozen high-power tugs (2,000 to 5,000 hp); motorboats; and service gear(2,3). North American Academic Research, 4(9) [September 2021 | https://doi.org/10.5281/zenodo.5516224 Monthly Journal by TWASP, USA | 1



Figure 1. Location of the Port of Pointe-Noire, with the Congo-Ocean Railway network

Port-throughput forecasting has become crucial for port operations, as it helps to prevent uncertainties and avoid possible logistical issues in ports. Therefore, forecasting container throughput in the POPN is an important task, as this will help greatly in the decision-making process regarding the port management, task planning, and development and improvement of port facilities for better global infrastructure while also avoiding congestion in the future (4). Talluri and Van Ryzin (2004) stated that the volume forecast in the port warehouse can provide a helpful clue, leading to more efficient strategies and better decisions (5). If the forecast data considered turns out to be inaccurate, significant financial losses may occur as a result (6). Therefore, accurate port-throughput forecasting should be developed to provide a solid basis for guidance on how to regulate the port's development and curtail wasting of resources.

A great amount of research related to the forecasting of container throughput in ports around the world has been conducted, but few studies have examined the POPN in the Republic of Congo and its container terminal. Thus, the objective of this study is to provide the local government with a trustworthy forecasting model for the throughput of the POPN container terminal, one of the lungs of the national economy, in order to create an effective port policy that will lead to better port facility handling while simultaneously maximizing profits. In doing so, the study will contribute to the rise of the national economy and help the government to go further in the national development program. To do so, the container-throughput data over ten years (2011–2020) were collected and observed, and then an ARIMA model that could fit the time series well was developed and used to forecast the container throughput for the year 2021.

The rest of this paper is structured as follows: section 2 is the literature review, where the research background is examined. Section 3 presents the data. Section 4 presents the methodology used to conduct this study. Section 5 presents the results. Section 6 summarizes and concludes the study.

2 Literature Review

Research on forecasting has played a considerable role in the scientific world for decades, and, with advances in technology and increased demand, an enormous number of forecasting models have been developed in many fields with the objective of continually improving performance and delivering more accurate results. The issue of forecasting port container traffic has become a hot topic among researchers in the maritime transportation field. Forecasting models vary from casual but quite accurate time series models, such as ARMA, ARIMA, SARIIMA, and exponential smoothing models, to more advanced models using machine learning, etc. Models combining two or more models have also been developed.

Dragan and Kramberger (2004) proposed using the Box–Jenkins auto-regressive integrated moving average (ARIMA) time series model for forecasting container throughput in the Port of Koper in Slovenia (7). They found that the accuracy of the ARIMA forecasting model was high. The model showed good performance and achieved a pretty good forecast outcome. Chen et al. (2006) used the Pearl Curve Model, the G-M (1,1), and the Exponential Smoothing model to propose a minimum-variance based combination forecasting model for predicting cargo throughput. The results of their research imply that a single model has lower precision than the combination model. However, they indicated that because of the unceasing changes in the shipping industry, which bring some uncertainty, the model they used in their research may not always be the most accurate (8). Chen (2013) adopted the Triple Exponential Smoothing-Grey combined model to forecast the container searail intermodal transport volume of Qingdao Port. The results showed that the combined model improved the accuracy of the initial forecast (9). Wu and Liu (2015), in their research on container sea-rail transport volume in the port of Ningbo, used a combination model of the Grey model and the Radial Basis Function (RBF) neural network to forecast container sea-rail transport volume progression for the next two years based on data collected over a six-year period. As a result, with the combination model, they obtained the expected outcome that combining the Grey model and the RBF neural network was more accurate than using them separately as single models (10).

Seabrooke et al. (2003) selected the regression analysis method to forecast the cargo throughput of Hong Kong Port, then compared it with the local forecast considered in the port management policy decisions. The results showed that their forecast was not congruent with the official forecast, which was pessimistic, and they predicted that the cargo throughput should still increase over the next few years, although at a moderate rate. Through their research based on a prediction of the container volume of the Port of Abidjan in West Africa using diverse models (11), Gamassa and Chen (2017) found that the most accurate model was the Double Exponential Smoothing model, although the combination model also gave some good results (12). This study shows that, even if the combined method produces good results, it does not always give the best results. We can conclude from these studies that single time series forecasting models are also widely used for forecasting container throughput and can provide very good results. Sometimes, combining models can be a good idea, but using single models may also be enough to obtain accurate results.

Fang and Lahdelma (2016) proposed two methods for forecasting district heat systems in Espoo in Finland: North American Academic Research, 4(9) | September 2021 | <u>https://doi.org/10.5281/zenodo.5516224</u> Monthly Journal by TWASP, USA | 3 the linear regression model and the seasonal auto-regressive integrated moving average (SARIMA) model. The results showed that the proposed linear regression model (T168h) was more accurate compared to the SARIMA model (13). Georgakarakos et al. (2006) used an Auto-Regressive Integrated Moving Average (ARIMA) model, artificial neural networks (ANNs), and a Bayesian dynamic model to forecast annual loliginid and ommastrephid landings registered at the principal fishing ports in the Northern Aegean Sea (1984–1999). The results showed that the selected models were all quite efficient, although the Bayesian model did show better performance than the others (14). Maia and De Carvalho (2010) used the Holt's Exponential Smoothing and neural network models to predict interval-valued time series, and then used a hybrid technique combining both methods. They used real interval-valued stock market time series to prove the practicability of their techniques via simulation analysis and applications. Their results suggested that their models can be efficient for interval-valued time series prediction (15).

The results of the above studies demonstrate that more than one model can actually perform well for the same data set. The very existence of this many studies related to forecasting demonstrates its importance in any field and reveals that forecasting work is necessary for the development of many fields, as well as illustrating that a multitude of such studies are now being conducted worldwide using numerous different methods. This paper aims to provide a good forecasting model for container traffic in the POPN, as not enough studies of the POPN have been conducted, despite the importance of this structure in the sub-region.

3 Data

The monthly data from the year 2011 to the year 2020 were collected directly from the Commercial Department of the Port of Pointe-Noire. With a total of 120 observations, the container monthly throughput time series is plotted in Figure 2-a while Figure 2-b plots the annual throughput and the growth rate. Table 1 shows the yearly empirical data and the growth rate for each year.





Measure: TEU (Twenty-foot Equivalent Unit)



Figure 2-b. Port of Pointe-Noire Container Annual Throughput and Growth Rate, 2011–2020 Measure: TEU (Twenty-foot Equivalent Unit)

Figure 2-a shows that from 2014 to 2016, the traffic had decreased slightly; then from 2017, it started to increase again. This can be explained by the collapse in oil prices during the year 2014, leading the country to a recession, as the country's economy, due to the lack of diversification, relies mainly on petroleum exports (16). It is also relevant to note that the World Bank reported that between 2014 and 2016, the Republic of Congo's economic complexity index (ECI) deteriorated significantly, from -0.924 to -1.49. Figure 2-b shows the instability of the growth rate over this decade. It increased from 2011 to 2012, then constantly decreased from 2012 to 2015, then increased till 2019 before decreasing again in 2020. This instability is due to many factors, such as the lack of port efficiency, the volatility of the prices of raw materials, or, in the case of the year 2020, the impact of the COVID-19 pandemic.

4 Methodology

To conduct this study, a Box-Jenkins ARIMA model was designed.

4.1 Introduction to the model

The model was first introduced by George Box and Gwilym Jenkins (17), both British statisticians, for forecasting time series. The acronym ARIMA stands for Auto-Regressive (AR) Integrated (I) Moving Average (MA). The model aims to describe autocorrelations in data, and the past values are used to obtain the forecast values. Each component of the ARIMA model is specified by a parameter, so we have ARIMA (p,d,q), where p is the parameter for AR, representing the lag order (number of lag observations); d is the parameter for I,

representing the degree of differencing (the number of times actual series observations are differenced in order to make the series stationary); and q is the parameter for MA, representing the order of the moving average model (number of lagged prediction errors in the forecast equation). Basically, the ARIMA model is used as a non-stationary model; this explains the inclusion of I. When the series is already stationary at the beginning, there is no use in differencing, then d = 0, which gives us ARIMA (p, 0, q), which is equal to ARMA (p, q).

The ARMA (p, q) model with a forecast value of y at time t is given by Eq.(1):

$$\hat{y}_t = c + \varphi_1 y_{t-1} + \dots + \varphi_k y_{t-k} + \alpha_t - \theta_1 \alpha_{t-1} - \dots - \theta_q \alpha_{t-q} \tag{1}$$

where \hat{y}_t is the forecast value at period t, y_{t-1} is the forecast value of the previous period (t-1), c is the constant, φ_p is the coefficient of AR at lag p to be estimated, θ_p is the coefficient of MA at lag p to be estimated and $\alpha_{t-p} = y_{t-p} - \hat{y}_{t-p}$ is the forecast error at time t-p (α 's are a series of residuals, white noise).

Please note that the MA terms in this model are written with a negative sign, in accordance with Box and Jenkins' original work.

The ARIMA (p, d, q) model is obtained when the standard ARMA (p, q) model is combined with a differential process (18). It can be expressed as follows in Eq.(2):

$$\Delta^d \hat{y}_t = c + \varphi_1 \Delta^d y_{t-1} + \dots + \varphi_k \Delta^d y_{t-k} + \alpha_t - \theta_1 \alpha_{t-1} - \dots - \theta_q \alpha_{t-q}$$
(2)

where Δ is the differencing operator and $\Delta^d \hat{y}_t$ is the new series, differenced d times.

A simpler form of the ARIMA structure using the Box–Jenkins backshift operator can be written as follows:

$$\varphi_p(B)(1-B)^d \hat{y}_t = \theta_q(B)\alpha_t \tag{3}$$

where $(1 - B) = \Delta$ (backshift operator).

4.2 Model construction and validation

The ARIMA (p, d, q) model consists of three phases:

- Phase 1: Identification of the model
- Phase 2: Estimation and diagnostic checking of the model

- Phase 3: Forecasting

To evaluate the forecasting accuracy, we considered three indexes: MAPE (Mean Absolute Percentage Error), RMSE (Root Mean Square Error), and MAE/MAD (Mean Absolute Error/Mean Absolute Deviation). They are expressed as follows:

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|}{n} \times 100 \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (A_t - F_t)^2}{n}}$$
(5)

$$MAD/MAE = \frac{\sum_{t=1}^{n} |A_t - F_t|}{n}$$
(6)

where A_t is the actual value at time *t*; F_t is the forecast value of period (*t*), and *n* is the number of observations. We used the EViews economic software to construct the ARIMA (*p*, *d*, *q*) model. The time series is named "THROUGHPUT."

5 Results

5.1 Identification of the model

The stationarity of the time series is first checked. From Figure 2, we can suspect at first sight of the graph pattern that the data are non-stationary, in which case the use of the ARIMA model may be efficient. To confirm this suspicion, the correlogram (Figure 3) and a standard unit root test named the Augmented Dickey–Fuller (ADF) test (Figure 4) are run. As monthly data over 10 years are used, up to 24 lags are selected to conduct the tests.

The plot of the correlogram is shown in Figure 3:

	Correlogram of THROUGHPUT					
Date: 07/05/21 Time Sample: 2011M01 20 Included observation	e: 20:16 020M12 s: 120					
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.898	0.898	99,164	0.000
		2	0.843	0.193	187.43	0.000
	· •	3	0.802	0.096	267.98	0.000
· ·	() () () () () () () () () ()	4	0.759	0.010	340.61	0.000
•	() () ()	5	0.720	0.018	406.65	0.000
	1 1 1	6	0.678	-0.027	465.60	0.000
		7	0.671	0.167	523.86	0.000
	'⊑''	8	0.633	-0.082	576.28	0.000
· •		9	0.639	0.202	630.16	0.000
	' '	10	0.616	-0.082	680.71	0.000
	'[]'	11	0.580	-0.075	725.96	0.000
	' '	12	0.542	-0.101	765.80	0.000
	1 1	13	0.505	-0.005	800.71	0.000
	ויםי	14	0.490	0.051	833.83	0.000
· -	· 🗖 '	15	0.434	-0.126	860.07	0.000
	' '	16	0.384	-0.135	880.86	0.000
	ומי	17	0.340	-0.031	897.25	0.000
· • •	וםי	18	0.300	-0.053	910.14	0.000
· •	ון ו	19	0.270	0.027	920.73	0.000
· 🗖	1 1	20	0.236	-0.008	928.90	0.000
· •		21	0.215	0.019	935.77	0.000
· • 🖻	וייםי	22	0.193	0.046	941.35	0.000
· • •		23	0.183	0.018	946.42	0.000
· 🟳 ·	'뎍'	24	0.153	-0.093	949.98	0.000

Figure 3. Correlogram of THROUGHPUT at level (d = 0)

Figure 3 shows that lags in the autocorrelation function (ACF) are significant (outside the 95% confidence interval) and gradually decreasing, but that most of the lags are non-significant at the partial autocorrelation function (PACF). These ACF and PACF patterns usually indicate that the series is non-stationary. To obtain a more accurate identification of the stationarity, the ADF test is run.

Table 1. ADF	Unit Root Test on	THROUGHPUT
--------------	-------------------	------------

Augmented Dickey-Fuller Unit Root Test on THROUGHPUT						
Null Hypothesis: THROUGHPUT has a unit root Exogenous: Constant, Linear Trend Lag Length: 1 (Automatic - based on SIC, maxlag=12)						
t-Statistic Prob.*						
Augmented Dickey-Fuller test statistic -3.073085 0.1176 Test critical values: 1% level -4.037668 5% level -3.448348 10% level -3.149326						
*MacKinnon (1996) one-s	sided p-values.					
Augmented Dickey-Fuller Test Equation Dependent Variable: D(THROUGHPUT) Method: Least Squares Date: 06/16/21 Time: 20:12 Sample (adjusted): 2011M03 2020M12 Included observations: 118 after adjustments						
Variable	Coefficient	Std. Error	t-Statistic	Prob.		
THROUGHPUT(-1) D(THROUGHPUT(-1)) C @TREND("2011M01")	-0.197878 -0.223766 7301.857 63.94001	0.064391 0.091036 2361.313 26.19668	-3.073085 -2.457996 3.092287 2.440767	0.0027 0.0155 0.0025 0.0162		

Table 1 shows that the value of the *t*-statistic of the ADF test is higher than those of the test critical values (TCV) (1% level, 5% level, and 10% level), and the probabilities of the constant *c* and *@trend* are significant (< 0.05). These results indicate that the series is non-stationary and needs to be differenced in order to become

stationary.

Once the first differencing has been applied (d = 1), the same tests are run again to check the stationarity of the new differenced series, which is named "DTHROUGHPUT."

The plot of the correlogram of DTHROUGHPUT is shown in Figure 4, below:

Correlogram of D(THROUGHPUT)						
Date: 07/05/21 Time Sample (adjusted): 2 Included observation Autocorrelation	e: 20:19 :011M02 2020M12 s: 119 after adjustme Partial Correlation	nts	AC	PAC	Q-Stat	Prob
		1 1 2 3 4 5 6 7 7 8 9 10 11 12 13 14 15 16 17 18 19 20 20 20 20 20 20 20 20 20 20 20 20 20	-0.322 -0.035 0.003 -0.196 0.218 -0.226 0.093 0.090 -0.066 0.013 -0.084 0.192 -0.019 -0.032 0.047 -0.116 0.032 0.047 -0.116 0.032 -0.057 -0.013 -0.003	-0.322 -0.154 -0.063 -0.028 -0.242 0.071 -0.207 -0.036 0.064 -0.025 -0.046 -0.079 0.075 0.151 0.031 0.060 -0.059 -0.042 -0.042 -0.023 -0.080 -0.016	12.617 12.765 12.772 12.838 12.839 17.748 23.838 30.460 31.595 32.673 33.254 33.276 34.224 39.293 39.343 39.343 39.343 39.488 39.798 41.725 41.876 42.368 42.369	0.000 0.002 0.005 0.012 0.025 0.007 0.001 0.000 0.000 0.000 0.000 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.002 0.002 0.004 0.006
; P ;	ן יקי	23	0.009	-0.049	43.083	0.010

Figure 4. Correlogram of DTHROUGHPUT (d_=_1)

Figure 4 shows that both ACF and PACF have similar patterns, and most of the coefficients are non-significant (within the 95% CI); this generally indicates that the series is stationary. To confirm the stationarity of the new series, the ADF test is conducted:

Augment	ed Dickey-Ful	ler Unit Root	Test on D(TH	ROUGHPUT)		
Null Hypothesis: D(THROUGHPUT) has a unit root Exogenous: Constant, Linear Trend Lag Length: 0 (Automatic - based on SIC, maxlag=12)						
t-Statistic Prob.*						
Augmented Dickey-Fuller Test critical values:	r test statistic 1% level 5% level 10% level		-14.97215 -4.037668 -3.448348 -3.149326	0.0000		
*MacKinnon (1996) one-sided p-values. Augmented Dickey-Fuller Test Equation Dependent Variable: D(THROUGHPUT,2) Method: Least Squares Date: 06(16(21Time: 20:13						
Sample (adjusted): 2011M03 2020M12 Included observations: 118 after adjustments						
Variable	Coefficient	Std. Error	t-Statistic	Prob.		
D(THROUGHPUT(-1)) C @TREND("2011M01")	-1.322082 731.7907 -3.256219	0.088303 1038.658 14.94741	-14.97215 0.704554 -0.217845	0.0000 0.4825 0.8279		

From Table 2, it appears that the *t*-statistic value of the ADF test is greater than those of the TCV and that the *p*-values of the constant *c* and *@trend* are non-significant (> 0.05). The results of the ADF test show that the series became stationary after the first differencing.

Once the time series has become stationary, we proceeded by identifying the significant lags at the ACF to determine q and the PACF to determine p. Figure 5 shows that for ACF, lag 1, lag 6, lag 7, lag 8, and lag 14 are significant and for PACF, lag 1, lag 6, and lag 8 are significant. This means that q can be 1, 6, 7, or 8 and p can be 1, 6, or 8. Table 3 shows the possible combinations of p and q to form an ARIMA model we can get from Figure 5:

q	1	6	7	8	14
p					
1	(1,1)	(1,6)	(1,7)	(1,8)	(1,14)
6	(6,1)	(6,6)	(6,7)	(6,8)	(6,14)
8	(8,1)	(8,6)	(8,7)	(8,8)	(8,14)

Table 3. Possible models combinations of *p* and *q*

To select the best model from the combinations of models above, they are compared using these criteria: smallest volatility (SIGMASQ), smallest Akaike Info Criterion (ACI), smallest Hannan–Quinn Information criterion (HQIC), smallest Schwarz Criterion (SC), most significant coefficients, and the highest R-squared and Adjusted R-squared. Moreover, it is important to notice that in time series modeling, the well-known principle of parsimony should also be taken into consideration. This suggests the use of as few parameters as possible; the simpler the model, the better it is, at least when it produces good results. Over-parameterized models are to be avoided when possible. This reasoning added to the model comparison leads us to select p = 1 and q = 1, which gives us an ARIMA (1, 1, 1), as this model gave the best results during the comparison.

5.2 Estimation and diagnostic checking of the model

The ARIMA (1, 1, 1) is estimated using the Least Squares (LS) method, and the results are as follows:



Figure 5. Equation estimation of the model

Table 4. ARIMA (1,1,1) equation estimation results

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C AR(1) MA(1) SIGMASQ	384.8434 0.383044 -0.765302 28163186	201.9105 0.146638 0.106204 3825320.	1.906010 2.612178 -7.205991 7.362309	0.0591 0.0102 0.0000 0.0000
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood F-statistic Prob(F-statistic)	0.146695 0.124435 5398.405 3.35E+09 -1189.661 6.590056 0.000378	Mean depend S.D. depende Akaike info cr Schwarz crite Hannan-Quin Durbin-Watso	lent var ent var iterion rion in criter. on stat	391.0504 5769.276 20.06154 20.15495 20.09947 2.058709

To make sure that the model is good, the correlogram of residuals (CR) and the correlogram of residuals squared (CRS), also known as the Ljung–Box statistical test, are performed.

	Correlogram of Residuals					
Date: 07/05/21 Time: 23:40 Sample (adjusted): 2011M02 2020M12 Q-statistic probabilities adjusted for 2 ARMA terms						
Autocorrelation	Partial Correlation	AC PAC Q-Stat Prob				
		$ \begin{array}{cccccccccccccccccccccccccccccccccccc$				
1 þ 1	ן ים י	24 0.040 -0.058 23.161 0.393				

Figure 6. Correlogram of residuals

Date: 07/05/21 Time: 23:40 Sample (adjusted): 2011M02 2020M12 Included observations: 119 after adjustments Autocorrelation Partial Correlation	A	c	PAC	Q-Stat	Prob
	0. 0. 0. 0. 0. 0. 0. 0. 0. 0.	105 038 013 026 219 026 219 035 035 035 035 035 035 040 059 028 106 028 106 028 106 028 040 102 064 0055	0.105 0.269 -0.095 -0.055 0.084 -0.020 0.079 0.029 0.179 -0.089 0.036 0.087 -0.013 0.014 -0.024 0.085 -0.081 0.185 0.072 -0.069 0.049 0.013 -0.048	1.3457 10.817 10.995 11.016 11.175 11.204 12.635 18.902 19.067 20.962 21.127 21.361 21.832 21.941 23.825 24.442 32.233 32.468 33.972 34.560	0.246 0.004 0.012 0.026 0.082 0.084 0.125 0.026 0.039 0.034 0.046 0.082 0.082 0.093 0.093 0.109 0.093 0.109 0.093 0.108 0.021 0.028 0.026 0.031 0.026

Figure 7. Correlogram of residuals squared

Figure 6 shows that the CR is not totally flat; we can see that for the ACF, lag 14 is still significant, meaning that some information at lag 14 needs to be captured, and from Figure 7, we see that a lot of p-values for the CRS are significant (< 0.05). These results indicate that the model still has problems that need to be fixed.

To do so, we adjust the model by adding ma(14) when estimating the equation as follows:

Equation spe Dep and	ecification endent variable followed by list of regressors including ARMA PDL terms, OR an explicit equation like Y=c(1)+c(2)*X.
d(throughp	ut) c ar(1) ma(1) ma(14)
Estimation se	sttings
Method: LS	- Least Squares (NLS and ARMA)

Figure 8. Equation estimation of the adjusted model

Then the CR and the CRS for the adjusted model are performed:

Co	orrelogram of	fRe	siduals			
Date: 07/05/21 Time: 23:41 Sample (adjusted): 2011M02 Q-statistic probabilities adjus	2020M12 sted for 3 ARN	1A te	rms			
Autocorrelation Partial	Correlation		AC	PAC	Q-Stat	Prob
		1 1 3 3 4 5 6 7 8 9 10 11 12 13 14 15 17 18 20 21 22 22 23	-0.004 0.016 0.016 -0.037 -0.085 -0.162 0.097 -0.169 0.075 0.106 0.033 -0.015 0.033 -0.015 0.033 -0.067 -0.062 -0.020 -0.026 0.024 0.024 0.024	-0.004 0.016 0.016 -0.037 -0.086 0.164 0.100 -0.168 0.078 0.010 -0.053 0.013 -0.006 0.003 -0.004 0.020 -0.101 -0.064 0.020 -0.101 -0.061 0.019 0.019 0.096	0.0024 0.0339 0.0638 0.2371 1.1479 4.5112 5.7313 9.4506 10.179 11.666 11.789 11.938 11.938 12.180 12.801 13.047 13.103 15.599 15.778 16.337 16.438 16.527 18.059	0.626 0.563 0.211 0.220 0.117 0.112 0.161 0.217 0.350 0.350 0.384 0.444 0.518 0.409 0.469 0.469 0.562 0.622 0.622

Figure 9. Correlogram of residuals of the adjusted model

Figure 9 shows that the CR is now totally flat; the residuals are white noise, and all the information has been well captured. The model should be good. To confirm that the adjusted model is good, the Ljung–Box statistical test is also performed:

Correlogram of Residuals Squared								
Date: 07/05/21 Time Sample (adjusted): 2 Included observation Autocorrelation	e: 23:41 011M02 2020M12 s: 119 after adjustme Partial Correlation	nts	AC	PAC	Q-Stat	Prob		
		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 6 17 12 20 21 22 23	0.054 0.117 -0.090 0.092 0.102 -0.099 0.111 -0.017 0.159 -0.065 0.108 -0.052 0.129 -0.052 0.127 0.127 0.127 0.127 0.053 0.039 -0.002 -0.053 0.039 -0.002 -0.010	0.054 0.114 -0.104 0.092 0.118 -0.150 0.129 0.011 0.088 -0.047 0.100 -0.051 -0.084 0.165 -0.038 0.024 -0.047 0.079 0.054 -0.008 0.008 0.008 0.018 -0.091	0.3516 2.0255 3.0417 4.1023 5.4245 6.6611 8.2408 8.2776 11.604 12.160 13.721 13.862 14.227 16.506 16.658 18.558 20.836 23.135 23.533 23.760 23.760 23.774 23.798	0.553 0.363 0.385 0.322 0.363 0.353 0.312 0.407 0.237 0.275 0.249 0.358 0.358 0.340 0.292 0.234 0.234 0.234 0.215 0.253 0.359 0.359 0.359 0.3415		
· 🗗 ·	ן ימי	24	-0.089	-0.027	25.001	0.406		

Figure 10. Correlogram of residuals squared of the adjusted model

We can see from Figure 10 that all the *p*-values are now non-significant (> 5%), which indicates good results. This proves that the adjusted model is good and can be used for forecasting.

North American Academic Research, 4(9) | September 2021 | https://doi.org/10.5281/zenodo.5516224 Monthly Journal by TWASP, USA | 13

5.3 Forecasting

The adjusted ARIMA model is used to forecast the data set.



Figure 11. Forecast results for the adjusted ARIMA Model

Table 5 shows the results of the accuracy indexes for the adjusted ARIMA model.

Evaluation statistics						
Model	RMSE	MAE	MAPE			
Adjusted ARIMA	5198.920	4030.039	7.713			
model						

Table 5. Accuracy measures for the adjusted ARIMA model

It can be seen from Figure 11 that the forecasted values fit quite well the actual values of the original time series. This shows that the adjusted ARIMA model generated good forecast results. Table 5 shows a fairly low value of MAPE, which generally indicates that the model did well. Seeing high values of RMSE and MAE should not make us doubt the accuracy of the model, as both have the same units as the dependent variable, which are large numbers in our data set.

These results prove that the model is good and can be used to forecast the container throughput for the year 2021. Figure 12 plots the graph of the monthly forecasted throughput of the year 2021 whereas Table 6 presents its annual total throughput.



Figure 12. Forecasted throughput for the year 2021

Table 6. Forecasting the Port of Pointe-Noire's throughput for the year 2021 using the adjusted ARIMA

model	
Period	Throughput (TEU
Total throughput of 2021	957,500

Figure 12 shows how the throughput is likely to increase in 2021. Table 6 indicates that the total annual throughput of the POPN is expected to reach 957,000 TEUs in 2021, with a slight growth rate of +3.17% from 2020, where the annual throughput reached 928,124 TEUs. This weak growth rate, which has been observed since 2020, can be explained by the minor impact of the COVID-19 pandemic on maritime shipping in the Central Africa sub-region.

6 Conclusion

In this study, an adjusted ARIMA model is designed to predict the container traffic of the POPN in the Republic of Congo. The port is briefly presented, and the model is introduced, built, tested and then used to forecast the traffic for the year 2021 on the basis of the monthly historical data from 2011 to 2020 (120 observations). The adjusted ARIMA model fit the actual data quite well, showing that it can be used to forecast future values of container throughput.

By the end of the year 2021, the port container throughput is expected to reach 957,000 TEUs, with a low growth rate of only +3.17% due to the COVID-19 pandemic. Seeing the general trend of the forecast and knowing that the export/import demand is likely to increase the more the pandemic is stemmed over time, solutions must be found in order to avoid any logistical issues in the future.

As there is a lack of research papers related to the container traffic forecasting in Central Africa main ports in general and in the Republic of Congo in particular, this paper represents an important contribution toward filling this gap. Furthermore, the findings of this study, by providing a trustworthy model, may help the local authorities to get prepared and develop new strategies and management policies to improve the sustainability of the port infrastructure.

This study produced good results. However, the scope of this paper only covered the container traffic forecast based on historical data. Further studies may consider more inputs, such the evolution of the Gross Domestic Product (GDP) through the years.

References

- 1. Schweikert A. Republic of the Congo, Port of Pointe Noire [Internet]. The Logistics Capacity Assessment;

 2017.
 Available

 https://dlca.logcluster.org/display/public/DLCA/2.1.1+Republic+of+the+Congo+Port+of+Pointe+Noire
- Congo Terminal, Bolloré Ports. Congo Terminal receptionne deux nouveaux portiques de parc [Internet]. Bolloré Transport & Logistics; 2021. Available from: https://congoterminal.net//reception deux nouveaux rtg.html
- Bhalat S. Port de Pointe-Noire, Porte Océane De L'Afrique Centrale 1939-2019: Historique, développement et perspectives à l'horizon 2035. Port Autonome De Pointe Noire; 2019.
- Chen S-H, Chen J-N. Forecasting container throughputs at ports using genetic programming. Expert Systems with Applications [Internet]. 2010 Mar 15;37(3):2054–8. Available from: http://www.sciencedirect.com/science/article/pii/S095741740900640X
- Talluri K, Van Ryzin G. The Theory and Practice of Revenue Management. Springer US; 2004. 713 p. (International Series in Operations Research & Management Science; vol. 68).
- Xie G, Zhang N, Wang S. Data characteristic analysis and model selection for container throughput forecasting within a decomposition-ensemble methodology. Transportation Research Part E: Logistics and Transportation Review [Internet]. 2017 December;108:160–78. Available from: http://www.sciencedirect.com/science/article/pii/S1366554517300455
- 7. Dragan D, Kramberger T. Forecasting the Container Throughput in the Port of Koper using Time Series ARIMA model. In Celje, Slovenia; 2014. Available from: https://www.researchgate.net/publication/333974184_Forecasting_the_Container_Throughput_in_the_P ort_of_Koper_using_Time_Series_ARIMA_model/link/5d10bc7f458515c11cf308d3/download

- Chen Z, Chen Y, Li T. Port Cargo Throughput Forecasting Based On Combination Model. In: 2016 Joint International Information Technology, Mechanical and Electronic Engineering Conference [Internet]. Atlantis Press; 2016. Available from: https://doi.org/10.2991/jimec-16.2016.25
- Chen J. Research on Container Sea-Rail Intermodal Transport in Qingdao Port Based on Volume Forecast and Benefit Analysis [Internet]. Ocean University Of China; 2013. Available from: https://wenku.baidu.com/view/9b606f5f2f60ddccda38a0a8.html
- Wu H, Liu G. Container Sea-Rail Transport Volume Forecasting of Ningbo Port Based on Combination Forecasting Model. In: International Conference on Advances in Energy, Environment and Chemical Engineering [Internet]. Atlantis Press; 2015. Available from: https://doi.org/10.2991/aeece-15.2015.91
- Gamassa P, Chen Y. Application of Several Models for the Forecasting of the Container Throughput of the Abidjan Port in Ivory Coast. International Journal of Engineering Research in Africa. 2017 Jan 1;28:157–68.
- Seabrooke W, Hui ECM, Lam WHK, Wong GKC. Forecasting cargo growth and regional role of the port of Hong Kong. Cities [Internet]. 2003 Feb 1;20(1):51–64. Available from: http://www.sciencedirect.com/science/article/pii/S0264275102000975
- Fang T, Lahdelma R. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. Applied Energy [Internet]. 2016 Oct 1;179:544–52. Available from: http://www.sciencedirect.com/science/article/pii/S0306261916309217
- Georgakarakos S, Koutsoubas D, Valavanis V. Time series analysis and forecasting techniques applied on loliginid and ommastrephid landings in Greek waters. Fisheries Research [Internet]. 2006 Apr 1;78(1):55– 71. Available from: http://www.sciencedirect.com/science/article/pii/S0165783605003292
- Maia ALS, de Carvalho F de AT. Holt's exponential smoothing and neural network models for forecasting interval-valued time series. International Journal of Forecasting [Internet]. 2011 Jul 1;27(3):740–59. Available from: http://www.sciencedirect.com/science/article/pii/S0169207010000506
- Husted TF, Arieff A. The Republic of Congo (Congo-Brazzaville) [Internet]. Library of Congress. Congressional Research Service; 2019 Sep. Report No.: CRS In Focus, IF11301. Available from: https://www.hsdl.org/?view&did=829483
- Box GEP, Jenkins GMJ. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day; 1970.

 Box GEP, Jenkins GMJ, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control. 5th Edition. New-York: John Wiley and Sons; 2015. 712 p.



© 2021 by the authors. Author/authors are fully responsible for the text, figure, data in above pages. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/)

