Forecasting the Cargo Throughput for Haiphong Port in Vietnam

Hai Dang Bui * · Hwa-Young Kim**

ABSTRACT

Port throughput forecasting is fundamental in port optimization. A reliable prediction model is essential for the terminal operators to make decisions on planning and renovation of building structure and other port facilities. By monitoring the changes in seasonal patterns and business cycles in months or quarters, the predicted values help port managers in decision making and planning in the context of small and unexpected changes. In this paper, the authors reviewed a various of commonly used forecasting methods applied for the time-series data in the short-term. By applying a set of monthly data of Haiphong port from January 2003 to February 2019 to these models and evaluating forecast accuracy by root mean squared error (RMSE), we found that the Winters exponential smoothing method appears to be the best model for forecasting total cargo throughput with trend and seasonal variations. The empirical results could be used as a reliable scientific source for the port managers and the departments to make short-term plans for upgrading facilities and setting up effective loading and unloading plans, and then contribute to avoiding congestion and reducing unnecessary waste.

Keywords: Forecast, Cargo throughput, Trend and seasonal components

* Mokpo Maritime University, 91 Haeyangdaehak-ro, Mokpo, Jeollanamdo province, Korea
** Mokpo Maritime University, 91 Haeyangdaehak-ro, Mokpo, Jeollanamdo province, Korea

1. Introduction

With the rapid development in economic globalization, ports are becoming increasingly vital in the operation of international trade activities (Notteboom, 2016), particularly in the 21st century, owing to the significant growth in container shipping. With such a vast volume of cargo transported, the need for a sufficiently accurate cargo throughput forecast is not surprising since it can significantly influence the port development strategy, investments in infrastructure, daily operations management. According to Hyndman and Athanasopoulos (2018), forecasting is the process of making statements about events that actual outcomes have not yet been observed, but forecasting will estimate of the values at certain specific times in the future, which can help people and organizations planning and making decisions. Accurate port throughput prediction can not only avoid repetitive construction, but also improve the port resource utilization efficiency. However, if the prediction of throughput provides poor accuracy, port authorities will make unsuitable decisions, which may lead to significant financial losses (Xie, 2017). Thus, it is essential to implement accurate port throughput forecasting in port transportation system nowadays.

The forecasts for port operations are usually conducted in the long-term, but short-term forecasts are also important. In monitoring the changes in seasonal patterns and business cycles, the fact has been proved that short-term forecasts often yield better results than long-term forecasts (Franses and Dijk, 2005). As the prediction period is shorter, fewer unexpected factors may arise, short-term forecasts would be more accurate than long-term forecasts. Furthermore, forecasts on a short-term basis are necessary for the control and scheduling of a port system, and for the terminal operator in decision making and planning in the context of small and unexpected changes (Peng and Chu, 2009). However, the role of forecast in the seaport system is not paid much attention in Vietnam, mainly comes from weak statistical work. Currently, forecast is mainly conducted by the Government in the long term, e.g. the government introduced a master plan related to the development of seaport systems by 2030. However, regarding each port, the forecast task remains plenty of limitations, especially in the short term. Operational plans are mainly based on past experiences or sentiment, without forecast results and scientific bases, which leads to inaccurate decisions.

The aim of this study is to adopt monthly cargo throughput datasets of Haiphong port to forecast the values until the end of 2020. Future values of time series are assumed to be based solely on past values. Therefore, historical data are analyzed in an attempt to identify a pattern, and a similar historical pattern is assumed to continue in the future. The authors apply monthly data to the several suitable models for forecasting in the short-term, then conclude, among the applied models, which model is capable of generating the most accurate prediction of cargo throughput useful for Haiphong port authority. In this study, the data was collected from the planning and statistics department of Haiphong Port and analyzed with the support of statistical software such as Eviews 10 and Microsoft Excel 2016.

2. Literature review

Forecasting models may be classified according to different factors. In order to ground the problem presented in this paper, the literature review section focus on two classification factors: time horizon and method approach.

Forecast may be dealt with at different time-horizons according to the type of planning problem. Long and medium-term forecasts provide key inputs for port infrastructure planning (terminal capacity, berth utilization, and expansion decisions), in-port services and operational plans. Kuroda and Takebayashi (2005) forecasted long-term demand of container throughput in Indonesia in order to evaluate the investment required to expand the capacity of the Indonesian ports for the following 15 years. Schulze and Prinz (2009) focused on medium-term forecast models (SARIMA and Holt–Winters methods) to forecast the container transshipment in German ports in 3 different economic regions namely Asia, Europe and North America.

Short-term forecasts provide key inputs for monthly operation management decisions. These include port operation scheduling, congestion delay occurrences, maintenance planning, revenue management, among others. Chou, Chu and Liang (2008) built a short-term forecasting model for the number of monthly volumes of import container at ports in Taiwan. In an attempt to optimize cargo handling procedure, Peng and Chu (2009) made forecast for container throughput in three ports of Taiwan by month.

Regarding forecasting methods, several approaches have been considered to predict cargo throughput at ports. Among the most commonly used approaches two are worth highlighting: causal econometric and time series. The econometric approach seeks to identify relationships between demand-related factors (e.g. total cargo throughput) and social, economic, and service-related factors. The time series approach relies on historical data series to generate forecasts making use of the correlation between present and past observations.

Regarding time series approaches, Grubb and Mason (2001) applied the Holt-Winters method by considering monthly time series to demonstrate forecasting performance for long lead times. Another forecasting application was developed by Peng and Chu (2009). They found out the best forecasting model of container throughput from six difference methods, namely: classical decomposition, the trigonometric model, the seasonal dummy variables, the grey forecast, the hybrid grey forecast and SARIMA. The result shows that the classical decomposition has the best performance to predict container throughput in Keelung port and Taichung port. While in the prediction of container throughput in Kaohsiung Port, SARIMA and the classical decomposition perform better than other methods.

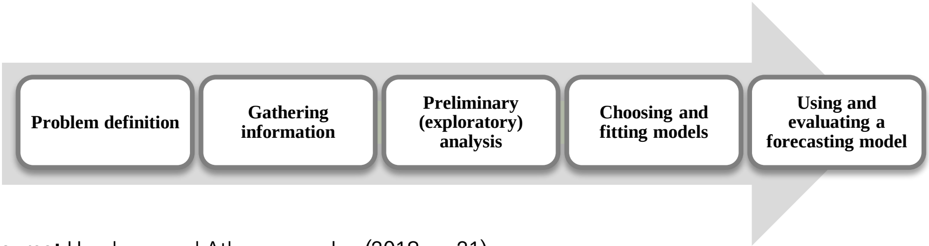
Regarding causal and econometric modelling, an application presented by Chou, Chu and Liang (2008) used leading indicators such as macroeconomic variables as dependent variables for regression models to predict short-term volumes of import containers in Taiwan. After comparing the accurate prediction of modified regression and traditional regression the predictive performance of modified regression had less error than traditional regression. Gosasang, Chandraprakaikul and Kiattisin (2011) built a linear regression model regarding dependent variables: GDP, exchange rate, population, interest rate, inflation rate and fuel price. The model was applied to predict monthly container throughput at Bangkok port.

In the current study, total cargo throughput at Haiphong port, selected because it is ranked as one of three largest port complexes of Vietnam and has the largest capacity in the northern region, was inputted into some commonly used models for forecasting purposes. Even though there are several studies on forecasting port cargo throughput, up to now there has been no practical study conducted at Haiphong port. Furthermore, this study concerns a comparison of different methods to choose the best suited model to forecast regarding the simplicity and rapidity in decision making.

3. Methodology

According to Hyndman and Athanasopoulos (2018), forecasting task usually involves five basic steps as can be seen in Figure 1. In the beginning, defining the problem carefully requires an understanding of the way the forecasts will be used, who requires the forecasts, and how the forecasting function fits within the organization requiring the forecasts. A forecaster needs to spend time talking to everyone who will be involved in collecting data, maintaining databases, and using the forecasts for future planning. Next, it is necessary to gather enough statistical data from the past, then analyze it to understand the nature and identify the pattern of datasets. After that, we choose and validate models. The best model to use depends on the availability and pattern of historical data, and the accuracy of forecasts measured through parameters. Finally, once a model has been selected and its parameters estimated, the model is used to make forecasts, then the results should be clearly presented to people who may use.

Figure 1. The basic steps in a forecasting task



Source: Hyndman and Athanasopoulos (2018, p. 21)

Based on the length of time, it is divided into long-term forecast, medium-term forecast and short-term forecast. Long-term forecasts are used in strategic planning with a forecast period of up to 15 years. Such decisions must take account of market opportunities, environmental factors and internal resources. Medium-term forecasts are needed to determine future resource requirements, in order to purchase raw materials, hire personnel, or buy machinery and equipment. Short-term forecasts are needed for the scheduling of personnel, production and transportation. The choice depends on the purpose of forecasters, what data are available and the predictability of the quantity to be forecast. In this paper, six commonly used forecasting models for the time series in the short-term are presented.

3.1. Naive model

The naïve forecast models assume that the nearest periods are the best estimates for the future. They can be represented as follow:

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1}) \text{ or } \hat{Y}_{t+1} = Y_t . \frac{Y_t}{Y_{t-1}} \quad (1)$$

$$\hat{Y}_{t+1} = Y_{t-11} \quad (2)$$

$$\hat{Y}_{t+1} = Y_{t-11} + \frac{(Y_t - Y_{t-12})}{12} \quad (3)$$

Where $\hat{Y}_{t+1}$ is the forecast value in period t+1, which is estimated based on the actual value in the past. (1) is applied for time-series data with trend component while (2) is used if the monthly data is adopted or the data contains seasonal component. (3) is adjusted naïve method, which is applied for the data has both trend and seasonal components.

3.2. Moving averages model

The moving averages model uses some of the closest observations as forecast values, which is shown by the following formula:

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + \cdots + Y_{t-11}}{k} \quad (4)$$

Thus, the moving average for period t is the mean value of k the closest observations. In a moving average, the weights of each observation are equal to 1/k. The moving averages forecasting model is suitable for stationary series.

3.3. Exponential smoothing model

The exponential smoothing method is based on averaging all the past values of the data series as exponentially decreasing weight. There are three common methods, which is presented as follow:

3.3.1. Simple exponential smoothing model

$$\hat{Y}_{t+1} = \alpha.\ Yt + (1 - \alpha)\ \hat{Y}_t \quad (5)$$

Where: $\hat{Y}_{t+1}$: the forecast value in period t+1, α: exponential coefficient, Y_t: the actual value in period t, $\hat{Y}_t$: the forecast value in period t.

Thus, the simple exponential smoothing model assumes that the new forecast value is an average between the actual value and the predicted value in period t. The exponential coefficient value determines the degree of influence of the current observation on the predicted value.

3.3.2. Holt's exponential smoothing model

Holt's exponential smoothing model is presented in the three following equations:

Estimate current average value:

$$L_t = \alpha Y_t + (1 - \alpha).\ (L_{t-1} + T_{t-1}) \quad (6)$$

Estimate trend value:

$$T_t = \beta.\ (L_t - L_{t-1}) + (1- \beta)\ T_{t-1} \quad (7)$$

Forecast the future value in period p:

$$\hat{Y}_{t+P} = L_t + p.\ T_t \quad (8)$$

Where: L_t: the estimated current average value, α: the exponential coefficient of L_t ($0<\alpha<1$), Y_t: the actual value in period t, β: the exponential coefficient of T_t ($0< \beta<1$), T_t: the estimated trend value, p: the forecast period, $\hat{Y}_{t+P}$: the forecast value in period p

In short, Holt's exponential smoothing model is suitable for data series with trend component.

3.3.3. Winter's exponential smoothing model

Winter's exponential smoothing model is presented in the three following equations:

Estimate current average value:

$$L_t = \alpha\frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + T_{t-1}) \qquad (9)$$

Estimate trend value:

$$T_t = \beta. (L_t - L_{t-1}) + (1-\beta)\, T_{t-1} \qquad (10)$$

Estimate seasonal value

$$S_t = \gamma\frac{Y_t}{L_t} + (1-\gamma)S_{t-s} \qquad (11)$$

Forecast the future value in period p:

$$\hat{Y}_{t+P} = (L_t + pT_t).\, S_{t-s+p} \qquad (12)$$

Where: L_t: the estimated current average value, α: the exponential coefficient of L_t ($0<\alpha<1$), Y_t: the actual value in period t, β: the exponential coefficient of T_t ($0< \beta<1$), T_t: the estimated trend value, γ: the exponential coefficient of S_t, S_t: the estimated seasonal value, p: the forecast period, $\hat{Y}_{t+P}$: the forecast value in period p

To sum up, Winter's exponential smoothing model is applied for data series with trend and seasonal components.

3.4. Regression-based trend model

Trend is the up and down movement of data for a long time. This movement can be described by a straight line (linear trend) or a curve (nonlinear trend). The trend can be modeled by an appropriate regression function between the forecast variable (variable Y) and time (variable t). This regression function is then used to generate future forecast values. The commonly used model is presented as follow:

$$Y_t = \beta_1 + \beta_2.t + U_t \qquad (13)$$

3.5. Time-series decomposition model

In this model, the time series is decomposed into four separate components: trend, cyclical, seasonal and irregular factors. Of these four components, forecasting models can only analyze the trend and seasonal changes. The cyclical component requires a data series of at least 30 years while unusual fluctuations cannot be predicted. Therefore, time-series decomposition model mainly refers to trend and seasonal components in order to understand how these components relate to the original data series. There are two types of models: the multiplicative model and the additive model. In this paper, we take the multiplicative approach and state the time series as:

$$Y_t = TR_t \, . \, SN_t \, . \, CL_t \, . \, IR_t \tag{14}$$

Where Y_t is the observed value of the time series in time period t, TR_t is the trend component in time period t, SNt is the seasonal component in time period t, CLt is the cyclical component in time period t, IRt is the irregular component in time period t. In this method, the concept is eliminating the seasonal component first, then applied the aforementioned models to the after-adjusted data series.

Table 1. A guide to selecting an appropriate forecasting method

Forecasting method	Data pattern	Forecast horizon
Naive	Stationary	Very short
Moving averages	Stationary	Very short
Simple exponential smoothing	Stationary	Short
Holt's exponential smoothing	Linear trend	Short to medium
Winter's exponential smoothing	Trend and seasonality	Short to medium
Regression-based trend	Linear and nonlinear trend	Short to medium
Time-series decomposition	Trend, seasonal and cyclical patterns	Short, medium and long

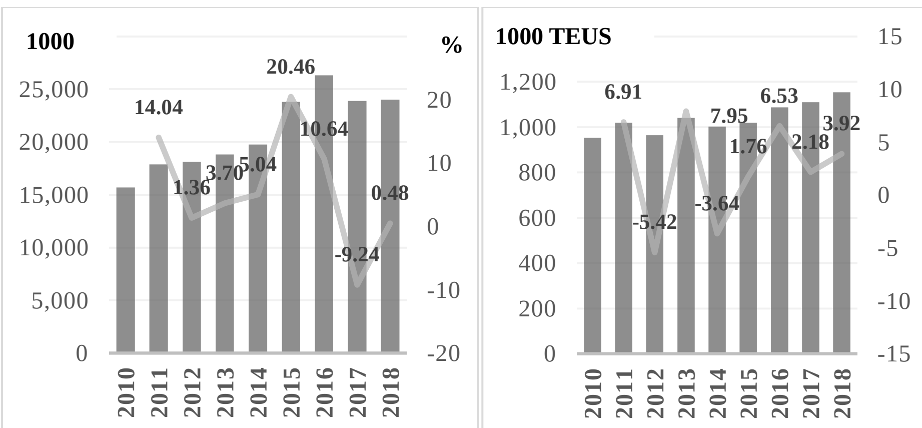
Source: Hanke and Wichern (2005)

4. Empirical study

4.1. Haiphong port overview

Vietnam's seaport system consists of six groups with 44 seaports in total. Among these, group 1 and group 5 account for largest percentage regarding cargo throughput with 35% and 62% respectively. Belonging to group 1, Haiphong port established in early 1874 has a long-lasting history. It is ranked as one of three largest port complexes of Vietnam, along with Danang and Ho Chi Minh city. Besides, Haiphong port is considered as the leading seaport in the North of Vietnam. In recent years, many new container port terminals have been developed in the downstream area, which accommodates larger container vessels and reduces the need for transshipments.

Figure 2. Total cargo throughput of Haiphong port from 2010 to 2018



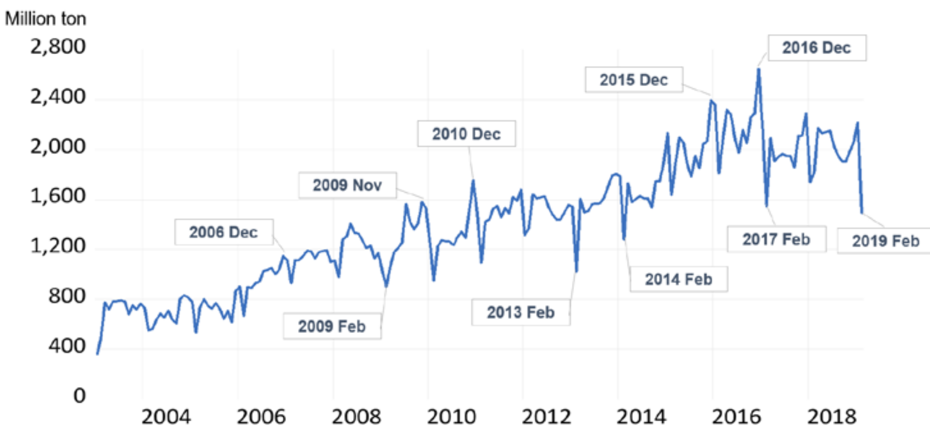
Source: Planning and Statistics Department – Haiphong Port

From 2010 to 2018, there is an upward trend in total cargo throughput of Haiphong port. Only in 2017 and 2018 did the number decrease coming from the restructuring and merging of company in 2017. The highest number can be seen in 2016 with over 26 million tons of cargo throughput. In which, container accounts for over 75% of total throughput, followed by other categories. Over the period from 2010 to 2018, the container throughput has increased slightly from 900 thousand TEUS to nearly 1.2 million TEUS. Moreover, the figure is forecasted to be considerably greater in the next following years when Lach Huyen port terminal is fully operating.

4.2. Data description

The monthly data of Haiphong Port's total throughput used in this study are from the planning and statistics department of Haiphong port, covering the period from January 2003 to February 2019, as illustrated by Fig. 3. We considered data from January 2003 to December 2016 as training set (in-sample) and the observations from January 2017 to February 2019 as test set (out-of-sample). In order to save space, the original data are not listed here.

Figure 3. Total cargo throughput of Haiphong Port by month from 2003 to 2019 February



Source: Planning and Statistics Department – Haiphong Port

As shown in Figure 3, the time-series data of the total cargo throughput is a series with trend and seasonal components. From January 2003 to February 2019, the total cargo throughput witnessed a continuous fluctuation. Although the volume of throughput did not increase steadily every year, the overall trend during the studied period was an upward trend. In addition, the data series also contains seasonal component. Specifically, the total volume of throughput usually plummets in the first months of the year (January, February, March) and often has a significant increase in the last months of the year (November, December).

This can be explained by the end of the year, the demand for transporting goods for the Chinese New Year holidays increased sharply. However, in the first months of the year, this demand tends to decrease, especially in the months after the holiday, when the productions are not busy as before, leading to a decrease in the demand for maritime transportation.

According to Hanke and Wichern (2005), appropriate forecasting methods applied for time - series data with trend and seasonal components are presented as follow:

- Adjusted naïve method (trend & seasonality)
- Winter's exponential smoothing method
- Time-series decomposition method (after eliminating the seasonal component, we can apply one of these methods for adjusted time-series data: Adjusted naïve method (trend), Holt's exponential smoothing method, Regression-based trend method)

4.3. Validate model

Since there is no universally accepted measure of accuracy or forecast error that can be applied to every forecasting situation, several criteria are normally used to give a comprehensive assessment of forecasting models. A forecast error is the difference between an observed value and its forecast. In this study, we measure accuracy or error of the forecast models by using the root mean squared error (RMSE) criteria, which defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum e_t^2}{n}} = \sqrt{\frac{\sum (Y_t - \hat{Y}_t)^2}{n}} \qquad (15)$$

Where Y_t and $\hat{Y}_t$ are the actual and the predicted values of the time series in period t, respectively. Obviously, RMSE is positive in value and the smaller the RMSE value obtained is the better the performance of the forecasting method.

With the support of EViews 10 software, we evaluate the aforementioned models to conclude which model gives the smallest error (the difference between the actual value and the forecast value in test dataset is smallest). The comparative results of forecasting accuracy of the five methods applying the test dataset of total cargo throughput of Haiphong port from January 2017 to February 2019 are presented in Table 1:

Table 2. Performance of methods of forecasting total cargo throughput of Haiphong port

No.	Forecasting method	Forecasting error
1	Adjusted naïve method (trend & seasonality)	RMSE = 236,844.389
2	Winter's exponential smoothing method	RMSE = 101,143.700
3	Adjusted naïve method (trend)	RMSE = 219,279.801
4	Holt's exponential smoothing method	RMSE = 111,180.000
5	Regression-based trend method	RMSE = 231,410.510

Table 2 shows that the Winter's exponential smoothing is clearly the best forecasting model since it has the lowest value of RMSE. The Holt's exponential smoothing appears to be the second best model for forecast accuracy. On the other hand, adjusted naïve method is found to be the worst method for predicting total throughput of Haiphong port.

4.4. Forecast results

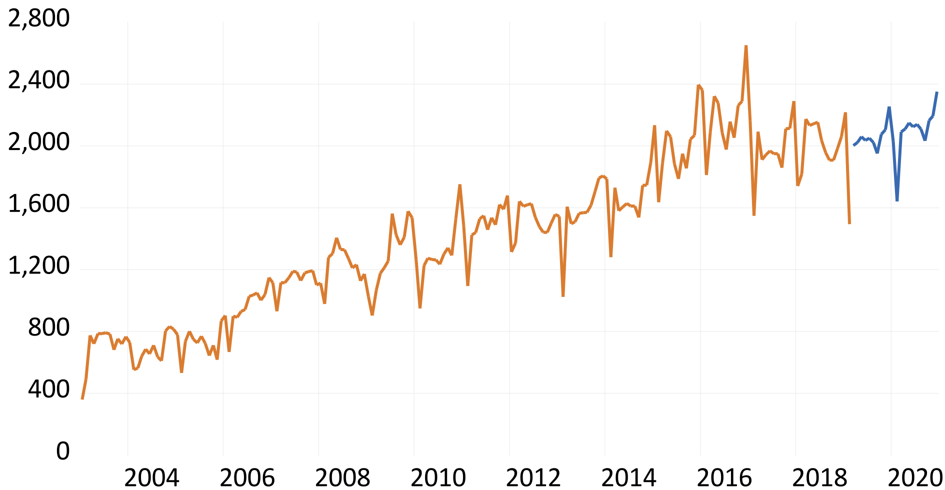
With the support of EViews 10 software, we forecast the volume of total cargo throughput of Haiphong port until 2020 by applying Winter's exponential smoothing method. The results are presented in Table 2.

Table 3. Forecast results of cargo throughput by month

No.	Month	Forecast results (tons)	No.	Month	Forecast results (tons)
1	2019 Mar	2,003,027	12	2020 Feb	1,641,276
2	2019 Apr	2,021,922	13	2020 Mar	2,090,604
3	2019 May	2,059,060	14	2020 Apr	2,110,004
4	2019 Jun	2,035,747	15	2020 May	2,148,435
5	2019 Jul	2,049,976	16	2020 Jun	2,123,792
6	2019 Aug	2,022,088	17	2020 Jul	2,138,318
7	2019 Sep	1,950,252	18	2020 Aug	2,108,916
8	2019 Oct	2,075,011	19	2020 Sep	2,033,697
9	2019 Nov	2,104,561	20	2020 Oct	2,163,478
10	2019 Dec	2,254,781	21	2020 Nov	2,193,971
11	2020 Jan	2,028,310	22	2020 Dec	2,350,235

To give a clearer view of the comparison, we plot the actual and the predicted values generated by the best method (Winter's exponential smoothing method) in Figure 4.

Figure 4. Forecast results of cargo throughput by month (thousand tons)



From the forecast results, the cargo throughput of Haiphong port, in general, is predicted to follow the upward and seasonal pattern as in the past. To specify, comparing the forecast results of the coming months in 2020 to those in the previous period, the overview can be seen as an increasing trend. On the other hand, it is easy to notice that seasonal component in the coming months continue to appear in the series of data on the throughput of goods. The total volume of goods throughput continued to decline sharply in the first months of the year (January, February, March) and significantly recovered in the last months of the year (November and December).

5. Conclusion

Port throughput forecasting is fundamental in port optimization. Accurate cargo throughput forecasting, especially in the short-term, not only facilitates the future development trend of ports, but also help to shorten transport time, reduce trade costs, and manage the port transportation system effectively. By monitoring the changes in seasonal patterns and business cycles in months or quarters, the predicted values help port managers in decision making and planning in the context of small and unexpected changes.

In this paper, we review some commonly used forecasting methods applied for the time-series data in the short-term. The monthly data of Haiphong port's total throughput covering the period from January 2003 to February 2019 was used in this study, which was a series with trend and seasonal components. The available dataset was separated into two portions, training data (from January 2003 to December 2016) and test data (from January 2017 to February 2019). After measuring the accuracy of the models through forecast errors (RMSE), the most suitable model chosen to forecast total cargo throughput by month for Haiphong port until 2020 was the Winters exponential smoothing method.

In the future, the cargo throughput of Haiphong port is predicted to follow the upward and seasonal pattern as in the past. Noticeably, the total volume of goods throughput continued to decline sharply in the first months of the year (January, February, March) and significantly recovered in the last months of the year (November and December). These forecast results can be used as a reliable scientific source for the port managers and the departments to make short-term plans for upgrading facilities and setting up effective loading and unloading plans, contribute to avoiding congestion and reducing unnecessary waste.

The study suggests that the port authority should realize the important role of forecast to the development of port in general and short-term forecast in particular. Besides, in behalf of the simplicity and rapidity of the chosen method, the port should continue collecting and updating data for the upcoming periods. If there are discrepancies between the forecast results and the actual results, there should be timely adjustments to the forecasting model and the forecasting method. Due to the lack of scientific papers focusing on forecasting in Haiphong port, we believe that this is one of the most important contributions of this study. However, in the future, other more sophisticated methods are necessary in order to lower the forecast errors.

References

Chou, C., Chu, C., Liang, G. (2008). A modified regression model for forecasting the volumes of Taiwan's import containers. *Mathematical and Computer Modelling*.

Gosasang, V., Chandraprakaikul, W., & Kiattisin, S. (2011). A Comparison of Traditional and Neural Networks Forecasting Techniques for Container Throughput at Bangkok Port. *The Asian Journal of Shipping and Logistics*.

Grubb, H., Mason, A. (2001). Long lead-time forecasting of UK air passengers by Holt-Winters methods with damped trend. *International Journal of Forecasting*.

Hyndman, R.J., Athanasopoulos, G. (2018). *Forecasting: principles and practice, 2nd edition.* OTexts: Melbourne, Australia.

Kuroda, I.S., Takebayashi, M. (2005). Forecasting the demand of container throughput in Indonesia. *Memoirs of Constructon Engineering Research Institute*.

John E.Hanke, Dean W.Wichern. (2005). *Business Forecasting, 8th Edition, Chapter 2&3*.

Notteboom, T. (2016). The adaptive capacity of container ports in an era of mega vessels: The case of upstream seaports Antwerp and Hamburg. *Journal of Transport Geography, volume 54*, 295-309.

Franses, P.H., Van Dijk, D. (2005). The forecasting performance of various models for seasonality and nonlinearity for quarterly industrial production. *International Journal of Forecasting, 21*, 87-102.

Peng, W.Y., Chu, C.W. (2009). A comparison of univariate methods for forecasting container throughput volumes. *Mathematical and Computer Modelling*.

Schulze, P.M., Prinz, A. (2009). Forecasting container transshipment in Germany. *Applied Economics*.

Peng, W.Y., Chu, C.W. (2009). A comparison of univariate methods for forecasting container throughput volumes. *Mathematical and Computer Modelling, volume 50, isuses 7-8*, 1045-1057.

Xie, G. (2017). Data characteristic analysis and model selection for container throughput forecasting within a decomposition-ensemble methodology. *Transportation Research Part E: Logistics and Transportation Review, volume 108*, 160 - 178.