Application of Grey forecasting model to estimate the Cargo throughput in Vietnam seaports

Van Thanh Phan[1], Thi Ngoc Lien Ho[1] , Vinh Dang[1], Thi Tieu Tien Nguyen[1], Thi Thuy Duong Tran[1]

[1.]Vietnam – Korea University of Information and Communication Technology, the University of Danang

Email: pvthanh@vku.udn.vn; lienhtn.22el@vku.udn.vn, dvinh@vku.udn.vn, tienntt.21it@vku.udn.vn, duongttt.23itb@vku.udn.vn,

Abstract: This study employs the GM (1,1) and FGM (1,1) models to forecast cargo throughput at seaports in Vietnam and compares their forecasting performance to determine the more accurate model. The forecasting value from 2025 to 2030 was conducted based on the annual datasets from 2015 to 2023 collected from the General Statistics Office of Vietnam. The findings indicate that both models are applicable in this context. However, the FGM (1,1) model demonstrates superior accuracy with the forecasting precision achieved 99.91%. As a result, the FGM (1,1) model is recommended for future cargo throughput forecasting at Vietnamese seaports. This result is also a scientific basis for planners, investors, and the government to have the basis to strengthen the inspection and operation activities and to make appropriate policies for the construction of seaport infrastructure in Vietnam. In addition, the forecast will help port managers to plan and make appropriate plans for seaports.

Keywords: the cargo throughput, GM (1,1), FGM (1,1), seaport, Vietnam.

1. Introduction

Seaports play a crucial role in promoting the development of maritime economic and the national economy. They are regarded as gateways for the export of goods and serve as the foundation for the transition from maritime transport to rail, road and inland waterway transport. Moreover, seaports are driving force that creating markets and serve as key links connecting economies between countries, stimulating market development. Especially, seaports help attract investors, traders and manufacturers to expand business activities. Export activities, foreign trade and other logistics service are facilitated due to the development of port system. The development of seaports act as catalysts in formation of industrial zones, export processing zones, financial and commercial centers along with maritime service hubs which are formed and developed around the seaports.

The volume of cargo throughput in seaports is an indicator used to assess the efficiency of seaports operations and the development policies set by the government. The index also reflects the the level of growth in international trade activities. In Vietnam, the volume of cargo throughput in seaport reached approximately 346,539 million tons in the first five months of 2024, representing a 17% increase compared to the same period in 2023. Specifically, the export cargo volume is estimated to be 106,873 million tons increasing by 27%. Domestic cargo is estimated to reach 154,283 million tons up to 11% and transshipment cargo reached around 1,531 million tons [2]. This index is also used by the General Statistics Office and the General Department of Customs for reporting purposes. Furthermore in recent years, this index has been utilized by economists and researcher for future forecasting. Bui et al. (2019) used traditional models such as the Naive adjustment model, SARIMA seasonal forecasting model and multiple regression model to predict the cargo throughput at Hai Phong seaport based on data from January 2003 to February 2019. The results indicated that the SARIMA forecasting model was the most suitable in this case [4]. Pham (2020) also used a multiple regression model to forecast the cargo throughput at the Hai Phong seaport. The results showed that the multiple regression model was appropriate for this case [5]. In the world, this index has also been studied by many research. For example, Shu et al. (2013) used the SARIMA seasonal forecasting model and GM (1,1) model to forecast the cargo throughput at Hong Kong and Kaohsiung seaports, utilizing datasets from Hong Kong Port (1997 – 2013) and Kaohsiung Port (2004 – 2013). The study concluded that both models produced suitable results [6]. Chen et al. (2021) employed the DFA – ARIMAZ (Dynamic Factor Analysis – ARIMA modelling) method, comparing it with the forecasting results from ARIMA and HW models to identify the most appropriate method for forecasting cargo throughput at seaports. The study concluded that the DFA – ARIMAZ method provided better results, though seasonal variations had not been adequately captured. Therefore, the authors plan to continue their research and use the SARIMAXINT – ANN forecasting method in the future to improve forecast accuracy [7]. Wang and Phan (2014) utilized the grey system theory, specifically the GM (1,1) model and an improved GM (1,1) model using Fourier series to forecast the cargo volume at the Kaohsiung international trade seaport (Taiwan) using data from 2003 to 2012. The study showed that the improved model yielded better forecasting results than the traditional Grey model with a forecasting accuracy of nearly 100% [8].

The Grey system theory was proposed by Professor Julong Deng from China in 1982 [9]. This method is considered novel and focuses on solving problems involving unclear and uncertain information. Currently, this theory is widely used by scholars around the world and has achieved promising results in various fields including energy [10], environment [11], tourism [12], finance [13], and economics [14],…

Recognizing the importance of forecasting cargo throughput at seaports and the strengths of grey forecasting models, this study employs two specific grey models GM (1,1) and FGM (1,1) with the following objectives: (1) to examine whether these two models are suitable for forecasting cargo throughput at seaports in Vietnam; (2) to determine which of the two models provides the most accurate and appropriate results; (3) to provide forecast outcomes that can serve as a valuable foundation for policymakers and enterprises in formulating strategic plans, zoning policies and making informed decisions amend at promoting stable and sustainable maritime economic development.

2. Research methods

2.1 GM (1,1) model

The GM (1,1) is based on GM (n, m) where n is the order of Grey difference equation and m is the number of variables. Among the family of Grey forecasting model most of the previous researchers have focused on GM (1, 1) model in their prediction. GM (1,1) model ensure a fine agree between simplicity and accuracy of the results. The GM (1,1) model is described by Wang and Phan [6] through the following six steps:

Step 1: Assume that $x^{(0)}$ denotes a non-negative sequence of raw data as

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots x^{(0)}(n)), n \geq 4 \quad (1)$$

Where n is the sample size of data

Step 2: Accumulating Generation Operator (AGO) is used to smooth the randomness of primitive sequence. The AGO converting the original sequence into a monotonically increasing sequence. A new sequence $x^{(1)}$ is generated by one- time AGO as:

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \ldots x^{(1)}(n)), n \geq 4 \quad (2)$$

Where, $x^{(1)}(1) = x^{(0)}(1)$ and $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 2,3,\ldots,n.$ (3)

The mean generating sequence $z^{(1)}$ of $x^{(1)}$ is defined as:

$$z^{(1)} = (z^{(1)}(1), (z^{(1)}(2), \ldots, (z^{(1)}(n)) \quad (4)$$

Where $z^{(1)}(k)$ is the mean value of adjacent data, i.e

$$z^{(1)}(k) = 0.5(x^{(1)}((k + (k-1)), k = 2,3,\ldots,n. \quad (5)$$

Step 3: The GM (1, 1) model can be constructed by establishing a first order differential equation for $x^{(1)}(k)$ as: $\frac{dx^{(1)}(k)}{d(k)} + ax^{(1)}(k) = b$ (6)

The solution in the Eq. (6) and also known as time response function is given by:

$$\hat{x}^{(1)}(k) = \left[x^{(0)}(1) - \frac{b}{a}\right]e^{-a(k-1)} + \frac{b}{a} \quad (7)$$

Step 4: To estimate the parameters a and b of GM (1,1) model we use the Ordinary Least Squares (OLS) method:

$$[a,b]^T = (B^T B)^{-1} B^T Y \quad (8)$$

In that

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \\ x^{(0)}(n) \end{bmatrix}; \; B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \ldots\ldots\ldots & . \\ \ldots\ldots\ldots & . \\ -z^{(1)}(n) & 1 \end{bmatrix} \quad (10)$$

Where: Y is called data series, B is called data matrix, and $[a,b]^T$ is called parameter series.

Step 5: Establish the forecasting equation to compute the predicted values of the GM (1,1) model.

$$\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right]e^{-ak} + \frac{b}{a};\ k = 1,2,3\ldots n, n+1,\ldots \tag{11}$$

Step 6: Calculate the forecast values of the GM (1,1) model by using inverse AGO *(IAGO)*:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k);\ k = 1,2,\ldots.. \tag{12}$$

Where: $\hat{x}^{(0)}(1) = x^{(0)}(1)$

Or $\hat{x}^{(0)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right]e^{-ak}(1-e^{a})$ (13)

2.2 The FGM (1,1) model

The FGM (1,1) model was mentioned in the research of Wang and Phan in 2014 [8]. The main ideas in this paper is that based on the advantages of the Fourier series in modifying the noise of residual error of the GM (1,1). In addition of 6 calculation steps as in the GM model (1,1), we have to calculate the following additional steps:

First: We have a residual series named ε is defined as:

$$\varepsilon = (\varepsilon(k)),\ k = 2,3,\ldots m \tag{14}$$

Where $\varepsilon(k) = x(k) - v(k)$, $k = 2,3,\ldots m$

$x_{(k)}$ is the original series of actual value and $v_{(k)}$ is the predicted series (obtained from GM (1,1).

Second: the residual sequence of GM (1,1) can be modify by the Fourier series approximately expressed as:

$$\hat{\varepsilon}(k) = \frac{1}{2}a_{(0)} + \sum_{i=1}^{Z}\left[a_i \cos\left(\frac{2\pi i}{m-1}(k)\right) + b_i \sin\left(\frac{2\pi i}{m-1}(k)\right)\right],\ k = 1,2,3,.,\ m \tag{15}$$

Where $Z = (\frac{m-1}{2}) - 1$ called the minimum deployment frequency of Fourier series [8] and only take integer number, therefore, the residual series is rewritten as:

$$\varepsilon = P.C \tag{16}$$

Where

$$P = \begin{bmatrix} \frac{1}{2} & \cos\left(\frac{2\pi\times 1}{m-1}\times 2\right) \sin\left(\frac{2\pi\times 1}{m-1}\times 2\right) & \ldots & \cos\left(\frac{2\pi\times Z}{m-1}\times 2\right) \sin\left(\frac{2\pi\times Z}{m-1}\times 2\right) \\ \frac{1}{2} & \cos\left(\frac{2\pi\times 1}{m-1}\times 3\right) \sin\left(\frac{2\pi\times 1}{m-1}\times 3\right) & \ldots & \cos\left(\frac{2\pi\times Z}{m-1}\times 3\right) \sin\left(\frac{2\pi\times Z}{m-1}\times 3\right) \\ \ldots & \ldots & \ldots & \ldots \\ \frac{1}{2} & \cos\left(\frac{2\pi\times 1}{m-1}\times m\right) \sin\left(\frac{2\pi\times 1}{m-1}\times m\right) & \ldots. & \cos\left(\frac{2\pi\times Z}{m-1}\times m\right) \sin\left(\frac{2\pi\times Z}{m-1}\times m\right) \end{bmatrix} \tag{17}$$

And $C = [a_0, a_1, b_1, a_2, b_2, \ldots, a_Z, b_Z]$ (18)

Third: the parameter a0, a1, b1, a2, b2… a_Z, b_Z are obtained by using the ordinary least squares method (OLS) which results in the equation of:

$$C = \left(P^T P\right)^{-1} P^T \varepsilon^T \quad (19)$$

Once the parameters are calculated, the modified residual series is then achieved based on the following expression:

$$\hat{\varepsilon}(k) = \frac{1}{2} a_{(0)} + \sum_{i=1}^{Z} \left[a_i \cos\left(\frac{2\pi i}{m-1}(k)\right) + b_i \sin\left(\frac{2\pi i}{m-1}(k)\right)\right] \quad (20)$$

Fourth: From the predicted series v and $\hat{\varepsilon}$, the Fourier modified series $\hat{v}$ of series v is determined by:

$$\hat{v} = \left(\hat{v}_1, \hat{v}_2, \hat{v}_3, \ldots, \hat{v}_k, \ldots, \hat{v}_m\right) \quad (21)$$

Where

$$\hat{v} = \begin{cases} \hat{v}_1 = v_1 \\ \hat{v}_k = v_k + \hat{\varepsilon}_k \quad (k = 2,3,\ldots, m) \end{cases} \quad (22)$$

2.3. Evaluative precision of forecasting models

Currently, there are various indexes for measuring the accuracy of forecasting models, such as mean absolute error (MAE), mean squared error (MSE), mean absolute percentage error (MAPE), and root mean squared error (RMSE). However, the mean absolute percentage error (MAPE) is more commonly used by researchers. Therefore, this index has been selected to used in this study and is calculated using the following formula:

$$\text{MAPE}=\frac{1}{n}\sum_{k=1}^{n}\left|\frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)}\right| x\, 100\% \quad (23)$$

Where $x^{(0)}(k)$: the actual value in time period k

$\hat{x}^{(0)}(k)$: The forecast value in time period k

And the grade of MAPE are divided into four grades [8]:

Table 1: The grade of MAPE

MAPE	**≤ 1%**	**1%-5%**	**5%-10%**	**>10%**
Grade level	Excellent	Good	Qualified	Unqualified

3. Empirical results

3.1 Data collection

The cargo throughput at Vietnamese seaports from 2015 to 2023 was collected from the website of the General Statistics Office. The entire datasets is visualized in Figure 1, with the unit of measurement expressed in million tons [1].

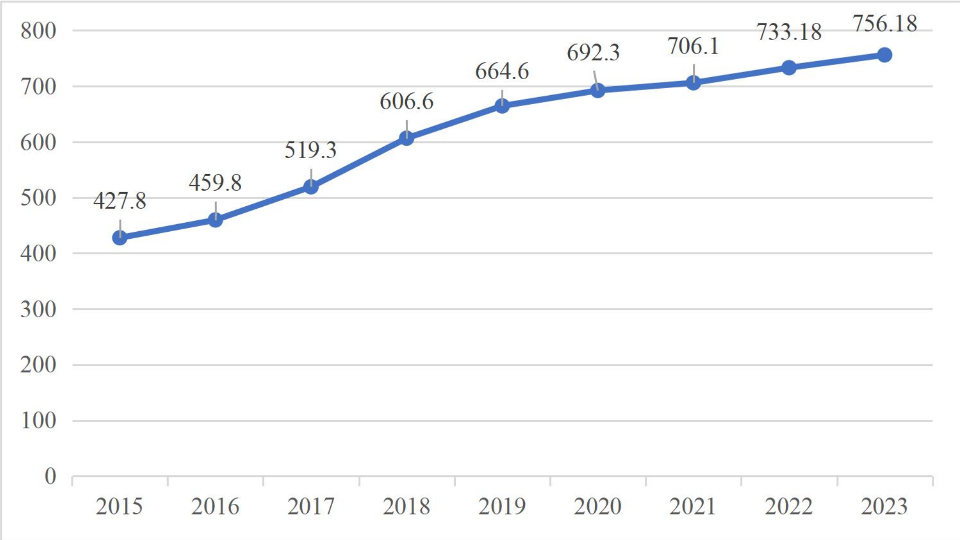

Figure 1: Cargo throughput at Vietnamese seaports from 2015 to 2023

Figure 1 illustrates the upward trend in cargo throughput at Vietnamese seaports over the years. In 2023, the volume of total cargo throughput reached 756,18 million tons, marking an increase of over 328.38 million tons compared to 2015. The slowest growth was observed between 2020 and 2021, with an increase of only 13.8 million tons, primarily due to the impact of the COVID-19 pandemic. During this period, international trade activities were disrupted, and regional as well as global supply chains experienced significant breakdowns, which negatively affected the flow of goods globally, including in Vietnam. The period between 2017 and 2018 witnessed the most substantial growth, with cargo throughput rising by 87.3 million tons.

3.2 Tools and function used

To calculate and simulate the GM (1,1) and FGM(1,1) models, this study utilizes Microsoft Excel, a widely used software developed by Microsoft Corporation. Excel is commonly adopted due to its user-friendly interface and a wide range of built-in functions that support computational processes. To estimate the parameter values of the two aforementioned models, the study applies basic mathematical operations along with two essential functions: the matrix multiplication function "MMULT (matrix1, matrix2)" and the matrix inversion function "MINVERSE (matrix)". These operations are fundamental for deriving the parameter estimates within the models. Following the completion of the calculations in Microsoft Excel, the forecasting results of the two models are presented in the subsequent section.

3.3 Forecasting result of Cargo throughput by GM (1,1) model

By integrating the datasets on cargo throughput at seaports during the period 2015–2023 with the algorithm of the GM (1,1) model, this study identified two key parameters, *a* and *b*. Through the computational process, the estimated values were determined as $a = -0.06186$ and $b = 470.30133$. Accordingly, the response function for forecasting cargo throughput at Vietnamese seaports is formulated as follows:

$$\hat{x}^{(1)}k = 8030.951963 * e^{-0.06186*(k-1)} - 7603.151963$$

The forecasting results and corresponding errors of the model are presented in detail in Table 2:

Table 2: Forecast values and percentage error of the GM (1,1) model

Year	Actual value	GM (1,1) model	error (%)
2015	427.8	427.8	-
2016	459.8	512.45	11.45
2017	519.3	545.15	4.98
2018	606.6	579.93	4.40
2019	664.6	616.94	7.17
2020	692.3	656.31	5.20
2021	706.1	698.18	1.12
2022	733.18	742.73	1.30
2023	756.18	790.13	4.49
Percentage Error (MAPE)			5.01
Accuracy = (100-MAPE) (%)			94.99%
Evaluation			Qualified

The results presented in Table 2 indicate that the GM (1,1) model is able to accepted for forecasting cargo throughput at Vietnamese seaports, with an accuracy belong the range of 90% to 95%. This results only meets the criteria for an "qualified" level of evaluation. Moreover, the error value of each year during the period time from 2015 to 2023 range from 1.12% to 11.45%.

3.4 Forecasting result of Cargo throughput by FGM (1,1) model

By combining the datasets collected from 2015 to 2023 with the algorithm of the FGM (1,1) forecasting model presented in Section 2.2, the study determined the parameter values as. Based on these values, the FGM (1,1) behavior model was developed to forecast cargo throughput at Vietnamese seaports as follows. The calculated results are presented in Table 3:

Table 3: Forecast values and percentage error of the FGM(1,1) model

Year	Actual value	FGM (1,1) model	error (%)
2015	427.8	427.8	-
2016	459.8	459.22	0.13

2017	519.3	519.88	0.11
2018	606.6	606.02	0.10
2019	664.6	665.18	0.09
2020	692.3	691.72	0.08
2021	706.1	706.68	0.08
2022	733.18	732.60	0.08
2023	756.18	756.76	0.08
MAPE			0.09
Accuracy = (100-MAPE) (%)			**99.91%**
Evaluation			**Excellent**

The results presented in Table 3 show that the accuracy of the FGM(1,1) model is 99.91%, which is 4.92% lower than the accuracy of the GM (1,1) forecasting model. Furthermore, the MAPE coefficient for the values from 2015 to 2023 ranges from 0.08% to 0.13%. Based on these results, this study demonstrates that the performance of the FGM (1,1) forecasting model is also satisfactory for this case. In addition, figure 2 emphasize that the curve of FGM(1, 1) extremely closed with actual data than the curve of GM (1,1) model.

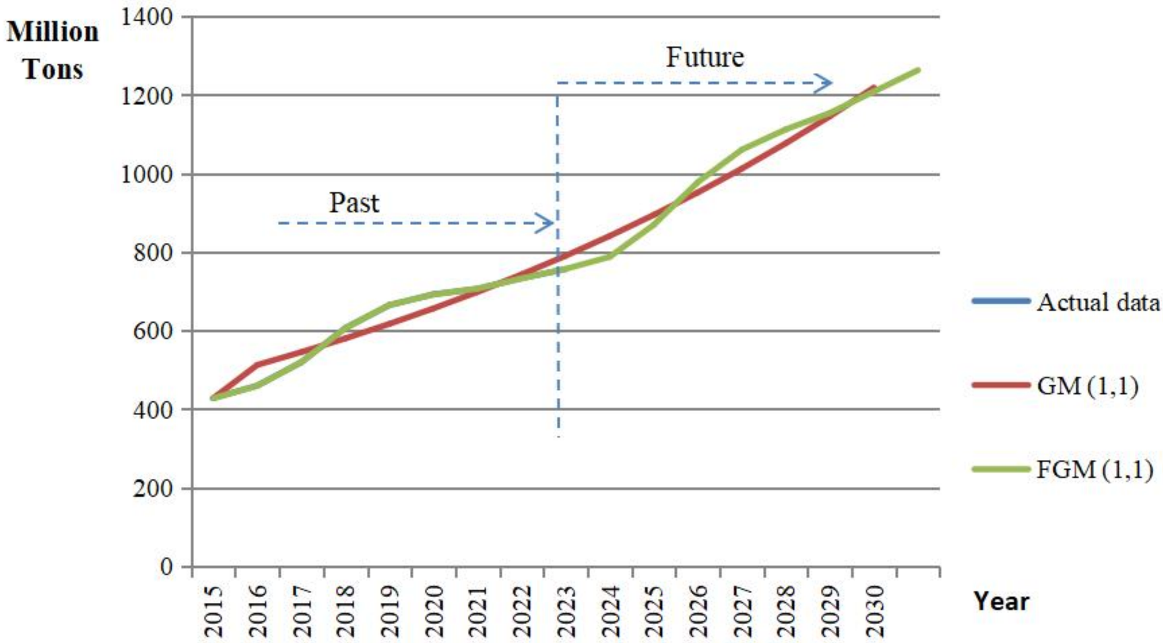

Figure 2: Cargo throughput at Vietnamese seaports from 2015 to 2030

3.5 Cargo throughput forecasting at seaports in Vietnam for the period 2025 – 2030

Through the comparison of the accuracy between the two forecasting models, this study suggests using the GM (1,1) model for forecasting cargo throughput at Vietnamese seaports for the period 2025 – 2030, as the accuracy of the GM (1,1) model is superior to that of the FGM (1,1) model. The forecasting results are presented in Table 4.

Table 4: Cargo throughput at Vietnamese seaports from 2025 to 2030

Year	Volume of Cargo throughput (Unit: million ton)
2024	787.31
2025	868.92
2026	977.32
2027	1,060.18
2028	1,111.91
2029	1,153.70
2030	1,208.13

The results in Table 4 shows that the volume of cargo throughput at Vietnamese seaports is expected to continue its strong growth in the coming years. Specifically, the forecast of cargo throughput at Vietnamese seaports will reach nearly 870 million tons in 2025 (an increase of more than 14.8% compared to 2023) and is projected to exceed 1,208 million tons in 2030. This will serve as an important data source for regulatory authorities to guide and plan the development of seaport activities in the near future.

4. Conclusions

Based on the simulation results, the study has identified the best forecasting model for cargo throughput forecasting at Vietnamese seaports for the period 2025–2030, which is the FGM(1,1) model with an accuracy rate of 99.91%. This result serves as a basis for the government to assess Vietnam's international trade activities and plan for the upgrading of the seaport system to meet the demands of the national economy. For businesses, this provides a foundation for developing business strategies and optimizing operations. Although the accuracy rate is quite high, there are still objective and subjective factors beyond the control of the model, such as global economic fluctuations, political and social factors, or technological advancements in the industry. Therefore, additional studies are needed to further clarify these factors and to apply the forecasting model more flexibly in real-world contexts. In the future, the authors propose studying cargo throughput at individual clusters of ports or specific ports to better analyze the regional characteristics of the North, Central, and South regions. This would support the government and businesses in making more accurate decisions regarding resource allocation, investment, and infrastructure development for seaports. To optimize this task, the team will continue to apply other forecasting models to test their accuracy and effectiveness, thus selecting the best forecasting model for each phase and specific conditions.

References

[1] Website of General Statistics Office of Vietnam. (2024). *Data on cargo throughput at Vietnamese seaports from 2015 to 2023*. Available online: https://www.gso.gov.vn/tin-tuc-thong-ke/2023/03/van-tai-bien-tiep-tuc-khoi-sac-trong-2-thang-dau-nam-2023/

[2] Website of Ministry of Transport of Vietnam. (2024). *Data on cargo throughput at Vietnamese seaports in 2024*. Available online: https://mt.gov.vn/vn/tin-tuc/92417/hang-qua-cang-bien-tang-truong-hai-con-so.aspx

[3 Website of Government of Vietnam *Approval of the adjustment of the master plan for the development of the Vietnam seaport system to 2020, with a vision to 2030* Available online: https://chinhphu.vn/?pageid=27160&docid=174573&tagid=2&type=1

[4] Bui, H. D., & Kim, H. Y. (2019). Forecasting the Cargo Throughput for Haiphong Port in Vietnam. International Journal of Maritime Affairs and Fisheries, 11(2), 1-15.

[5] Pham Thi Thu Hang (2020) *Forecasting cargo throughput at Hai Phong seaport for 2025 and 2030. Maritime Technology Journal, 1*(6)

[6] Shu, M. H., Huang, Y. F., & Nguyen, T. L. (2013). Forecasting models for the cargo throughput at Hong Kong Port and Kaohsiung Port. Recent Researches in Applied Economics and Management. WSEAS Press, Chania, Crete Island, Greece, 507.

[7] Hou, J., Chen, Y., & Li, T. (2014, December). The forecast of port cargo throughput based on combination forecasting model. In 2014 Seventh International Symposium on Computational Intelligence and Design (vol. 1, pp. 585-588). IEEE.

[8] Chia-Nan, Wang., & Van-Thanh, Phan. (2014). An improvement the accuracy of grey forecasting model for cargo throughput in international commercial ports of Kaohsiung. WSEAS transactions on business and economics, 11, 322-327.

[9] Liu S., Jeffrey F & Yang Y. (2012). A brief introduction to Grey systems theory. *Grey Systems: Theory and Application*, no. 2, pp. 89-104.

[10] Lin, C. C., He, R. X., & Liu, W. Y. (2018). Considering multiple factors to forecast CO2 emissions: A hybrid multivariate grey forecasting and genetic programming approach. Energies, 11(12), 3432.

[11] Huiping Wang and Zhun Zhang (2022). Forecasting CO_2 emissions using a novel grey Bernoulli model: A case of Shanxi province in China. *International journal of Environment research and public healthy.*

[12] Yi – Chung Hu & Peng Jiang (2020). Fuzzified grey prediction models using neural networks for tourism demand forecasting. *Computational and Applied mathematics.*

[13] Sheng, Y., Zhang, J., Tan, W., Wu, J., Lin, H., Sun, G., & Guo, P. (2021). Application of grey model and neural network in financial revenue forecast. Computers, Materials & Continua, 69(3), 4043-4059.

[14] Xuemei, L., Cao, Y., Wang, J., Dang, Y., & Kedong, Y. (2019). A summary of grey forecasting and relational models and its applications in marine economics and management. *Marine economics and management*, 2(2), 87-113.

[15] Nguyen, N. T., Phan, V. T., Nguyen, V. Đ., Le, T. H., & Pham, T. V. (2022). Forecasting the coffee consumption demand in Vietnam based on grey forecasting model. Vietnam Journal of Computer Science, 9(03), 245-259.