Forecasting Models for the Cargo Throughput at Hong Kong Port and Kaohsiung Port

MING-HUNG SHU[1], YING-FANG HUANG[2], THANH-LAM NGUYEN[3,*]
[1] Department of Industrial Engineering and Management
[2,3] Graduate Institute of Mechanical and Precision Engineering
National Kaohsiung University of Applied Sciences
415, Chien Kung Road, Kaohsiung 80778
TAIWAN, R.O.C.
(* Corresponding author: green4rest.vn@gmail.com)

Abstract: - The import and export of goods normally require the involvement of many different sectors from purchasing, manufacturing, transporting, inventory, distribution, etc. In order to have proper plans, accurately forecast the volume of imported-exported goods is the core issue. By comparing the performance of autoregressive integrated moving average (ARIMA) model, Grey model, and their joint Fourier modified models, this paper aims at obtaining an efficient model to forecast the cargo throughput at a sea port. The accuracy of the conventional models is found to be significantly boosted with the Fourier modification. In the empirical case studies of cargo throughput at Hong Kong port and container throughput at Kaohsiung port, modified seasonal ARIMA models, respectively $FSARIMA(4,1,4)(1,1,0)_4$ and $FSARIMA(4,1,4)(0,1,0)_{12}$, are strongly suggested due to their satisfactorily high forecasting power. We further employ these models to provide the forecast in 2013 at the two ports in order to assist the control and scheduling of the two port systems and for the terminal operators in decision making and planning.

Key-Words: - ARIMA model, Grey Model, Cargo throughput, Container throughput, Fourier modification, Hong Kong port, Kaohsiung port

1 Introduction

International trade, one of the most critical factors to develop the national economies, includes two basic activities: Import and Export. The import and export of goods normally require the involvement of many different sectors from purchasing, manufacturing, transporting, inventory, distribution, etc., especially the engagement of the customs authorities in both the country of export and the country of import. Therefore, it is of great significance to have appropriate planning from macro to micro levels so that the trade flow among nations becomes affluent. In order to have proper plans, accurately forecast the volume of imported-exported goods is the core issue [1-3].

In predicting the container throughputs used in international trade, several forecasting models have been proposed as in the review by [1]. Time series characteristic of the throughput has received special attention of many scholars [3-5] and regression methods have been commonly used [1]. However, conventional regression methods disregard the non-stationary relationship among the volume of containers and the macro-economic variables, possibly resulting in the spurious regression forecasting models [6]. One of well-known model dealing with a time series is Autoregressive Integrated Moving Average (ARIMA) model [7, 8]. Therefore, ARIMA model is taken into consideration in this paper.

Furthermore, due to the limitations in accessing the relevant data as well as the determinants of the cargo throughput, we also consider the Grey model, which can effectively deal with the which has been widely employed in different areas due to its ability to deal with the problems of uncertainty with few data points and/or "partial known, partial unknown" information [9, 10].

In order to improve the accuracy of the forecasting models, in this study, a new approach to minimize the errors obtained from the conventional models is suggested by modifying the residual series with Fourier series. Those models are then compared based on certain evaluation indexes. In this study, two empirical studies of cargo throughput at Hong Kong port and Kaohsiung port are conducted to find out appropriate forecasting

ISBN: 978-960-474-323-0

models for the control and scheduling of port system and the terminal operations.

2 Literature Review

2.1 ARIMA Model

ARIMA model is usually used to forecast a time series which can be made stationary by differencing or logging. A time series may have non-seasonal or seasonal characteristics which have respectively the form of *ARIMA(p,d,q)* and *SARIMA(p,d,q)(P,D,Q)$_S$*, where p, d, q, P, D, Q, and S are the model parameters as reviewed by Nguyen et al. (2013). There are three basic steps in the overall procedures to obtain an ARIMA or SARIMA model [11, 12].

2.2 Grey Model

Grey theory offers a new approach to deal mainly with the problems of uncertainty with few data points and/or poor information which is said to be "partial known, partial unknown" [13, 14]. The core of the theory is the grey dynamics model which is usually called Grey model (GM). The Grey model is used to execute the short-term forecasting operation with no strict hypothesis for the distribution of the original data series [15]. The general GM model has the form of *GM(d,v),* where d is the rank of differential equation and v is the number of variables appeared in the equation. The basic model of Grey model is GM(1,1), a first-order differential model with one input variable which has been successfully applied in many different researches. It is obtained based on the following procedure.

Step 1: Suppose an original series with n entries is $x^{(0)}$:

$$x^{(0)} = \left\{ x_1^{(0)}, x_2^{(0)}, \ldots, x_k^{(0)}, \ldots x_n^{(0)} \right\} \tag{1}$$

where $x_k^{(0)}$ is the value at time k $\left(k = \overline{1,n}\right)$.

Step 2: From the original series $x^{(0)}$, a new series $x^{(1)}$ can be generated by one time accumulated generating operation (1-AGO), which is

$$x^{(1)} = \left\{ x_1^{(1)}, x_2^{(1)}, \ldots, x_k^{(1)}, \ldots x_n^{(1)} \right\} \tag{2}$$

where $x_k^{(1)} = \sum_{j=1}^{k} x_j^{(0)}$

Step 3: A first-order differential equation with one variable is expressed as:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{3}$$

where a is called developing coefficient and b is called grey input coefficient. These two coefficients can be determined by the least square method as the following:

$$\left[a, b\right]^T = \left(B^T B\right)^{-1} B^T Y \tag{4}$$

where

$$B = \begin{bmatrix} -\left(x_1^{(1)} + x_2^{(1)}\right)/2 & 1 \\ -\left(x_2^{(1)} + x_3^{(1)}\right)/2 & 1 \\ \cdots & \cdots \\ -\left(x_{n-1}^{(1)} + x_n^{(1)}\right)/2 & 1 \end{bmatrix} \qquad Y = \begin{bmatrix} x_2^{(0)} \\ x_3^{(0)} \\ \cdots \\ x_n^{(0)} \end{bmatrix}$$

Therefore, the forecasting equation for GM(1,1) is expressed as:

$$\hat{x}_k^{(1)} = \left[x_1^{(0)} - \frac{b}{a} \right] e^{-a(k-1)} + \frac{b}{a} \qquad \left(k = \overline{1,n}\right) \tag{5}$$

Based on the operation of one time inverse accumulated generating operation (1-IAGO), the predicted series $\hat{x}^{(0)}$ can be obtained as following:

$$\hat{x}^{(0)} = \left\{ \hat{x}_1^{(0)}, \hat{x}_2^{(0)}, \ldots, \hat{x}_k^{(0)}, \ldots, \hat{x}_n^{(0)} \right\} \tag{6}$$

where

$$\begin{cases} \hat{x}_1^{(0)} = \hat{x}_1^{(1)} \\ \hat{x}_k^{(0)} = \hat{x}_k^{(1)} - \hat{x}_{k-1}^{(1)} & \left(k = \overline{2,n}\right) \end{cases}$$

2.2 Fourier Residual Modification

The model accuracy of Grey models has been significantly improved after the residual series obtained from the model is modified with Fourier series [14-18]. Hence, this methodology is proposed The procedure to obtain the modified model is as the following.

Based on the predicted series $\hat{x}^{(0)}$ obtained from one of the traditional models, its residual series named $\varepsilon^{(0)}$ is defined as:

$$\varepsilon^{(0)} = \left\{ \varepsilon_2^{(0)}, \varepsilon_3^{(0)}, \ldots, \varepsilon_k^{(0)}, \ldots, \varepsilon_n^{(0)} \right\} \tag{2}$$

where $\varepsilon_k^{(0)} = x_k^{(0)} - \hat{x}_k^{(0)}$ $\quad \left(k = \overline{2,n}\right)$

Expressed in Fourier series, $\varepsilon_k^{(0)}$ is rewritten as:

$$\varepsilon_k^{(0)} = \frac{1}{2} a_0 + \sum_{i=1}^{F} \left[a_i \cos\left(\frac{2\pi i}{n-1} k\right) + b_i \sin\left(\frac{2\pi i}{n-1} k\right) \right] \tag{3}$$

where $F = \left[(n-1)/2 - 1\right]$ called the minimum deployment frequency of Fourier series [18] and only take integer number [14, 15, 17]. And therefore, the residual series is rewritten as:

$$\varepsilon^{(0)} = P.C \tag{4}$$

where

$$P = \begin{bmatrix} \frac{1}{2} & \cos\left(\frac{2\pi\times 1}{n-1}\times 2\right) & \sin\left(\frac{2\pi\times 1}{n-1}\times 2\right) & \cdots & \cos\left(\frac{2\pi\times F}{n-1}\times 2\right) & \sin\left(\frac{2\pi\times F}{n-1}\times 2\right) \\ \frac{1}{2} & \cos\left(\frac{2\pi\times 1}{n-1}\times 3\right) & \sin\left(\frac{2\pi\times 1}{n-1}\times 3\right) & \cdots & \cos\left(\frac{2\pi\times F}{n-1}\times 3\right) & \sin\left(\frac{2\pi\times F}{n-1}\times 3\right) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{2} & \cos\left(\frac{2\pi\times 1}{n-1}\times n\right) & \sin\left(\frac{2\pi\times 1}{n-1}\times n\right) & \cdots & \cos\left(\frac{2\pi\times F}{n-1}\times n\right) & \sin\left(\frac{2\pi\times F}{n-1}\times n\right) \end{bmatrix}$$

$$C = \left[a_0, a_1, b_1, a_2, b_2, \ldots, a_F, b_F\right]^T$$

The parameters $a_0, a_1, b_1, a_2, b_2, \ldots, a_F, b_F$ are obtained by using the ordinary least squares method (OLS) which results in the equation of:

$$C = \left(P^T P\right)^{-1} P^T \left[\varepsilon^{(0)}\right]^T \quad (5)$$

Once the parameters are calculated, the predicted series residual $\hat{\varepsilon}^{(0)}$ is then easily achieved based on the following expression:

$$\hat{\varepsilon}_k^{(0)} = \frac{1}{2}a_0 + \sum_{i=1}^{F}\left[a_i \cos\left(\frac{2\pi i}{n-1}k\right) + b_i \sin\left(\frac{2\pi i}{n-1}k\right)\right] \quad (6)$$

Therefore, based the predicted series $\hat{x}^{(0)}$, the predicted series $\tilde{x}^{(0)}$ of the modified model is determined by:

$$\tilde{x}^{(0)} = \left\{\tilde{x}_1^{(0)}, \tilde{x}_2^{(0)}, \ldots, \tilde{x}_k^{(0)}, \ldots, \tilde{x}_n^{(0)}\right\} \quad (7)$$

where

$$\begin{cases} \tilde{x}_1^{(0)} = \hat{x}_1^{(0)} \\ \tilde{x}_k^{(0)} = \hat{x}_k^{(0)} + \hat{\varepsilon}_k^{(0)} & \left(k = \overline{2,n}\right) \end{cases}$$

In order to evaluate the accuracy of the forecasting model, four basic indexes, including mean absolute percentage error (MAPE), post-error ratio (C), small error probability (P), and forecasting accuracy (ρ), have been usually used as reviewed by Nguyen et al. [12]. Based on the combination of these four indexes, the forecasting accuracy is classified into four grades as stated in Table 1.

Table 1: Four grades of forecasting accuracy

Grade level	*MAPE*	*C*	*P*	ρ
I (Very good)	< 0.01	< 0.35	> 0.95	> 0.95
II (Good)	< 0.05	< 0.50	> 0.80	> 0.90
III (Qualified)	< 0.10	< 0.65	> 0.70	> 0.85
IV (Unqualified)	≥ 0.10	≥ 0.65	≤ 0.70	≤ 0.85

3 Empirical Study

3.1 Cargo throughput at Hong Kong port

Historical data of the cargo throughput at Hong Kong port from the first quarter (Q1) of 1997 – Q1 of 2013 are obtained from the quarterly statistical data published by the Census and Statistics Department of Hong Kong [19]. The data from Q1-1997 – Q3-2012 are used to build SARIMA and Grey models whose residual series are then modified with Fourier series to become modified models with higher accuracy. Data from Q4-2012 – Q1-2013 are used to check the forecast power of the modified models before one of them is employed to forecast the throughput in the other three quarters of 2013.

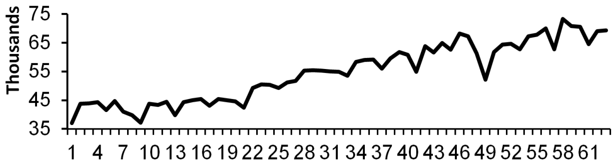

Figure 1. Quarterly throughput at Hong Kong port

3.1.1 ARIMA Model

From Figure 1, it can be concluded that seasonality exists in the series of tourism demand. Therefore, only seasonal ARIMA model is considered in this section. At one degree of both non-seasonal and seasonal difference, the series becomes stationary and *SARIMA(4,1,4)(1,1,0)*$_4$ (HK-1) is found appropriate. Its Fourier modified model is *FSARIMA(4,1,4)(1,1,0)*$_4$ (HK-2).

3.1.2 Grey Model

From the data, GM(1,1) (HK-3) is obtained as:

$$\hat{x}_k^{(1)} = 4293068.99e^{0.009425(k-1)} - 4256019.99 \quad \left(k = \overline{1,n}\right)$$

Its Fourier modified model is named as FGM(1,1) (HK-4).

The performance of these four models is briefly summarized as in Table 2.

Table 2. Model accuracy (Hong Kong port)

Model	KH-1	HK-2	HK-3	HK-4
MAPE	0.03392	0.00265	0.04746	0.02085
C	0.24045	0.01473	0.32703	0.11123
P	1.00000	1.00000	0.98412	1.00000
ρ	0.96608	0.99735	0.95254	0.97915
Forecasting power	Good	Very good	Good	Good

Based on the evaluation indexes in Table 2, among the four models, *FSARIMA(4,1,4)(1,1,0)*$_4$ (HK-2) outperforms others and it is therefore selected. In order to further evaluate the forecasting power of this model, we now compare the forecast values (unit: 1,000 tons) in Q4-2012 and Q1-2013 with the actual observations in the same period, which results in Table 3.

Table 3. Checking forecasting power HK-2

Month	Actual	Forecast	APE
Q4-2012	66,288	67,475	0.01791
Q1-2013	62,505	62,303	0.00323
Mean absolute percentage error			0.01057

With the low MAPE value of 0.01057, *FSARIMA(4,1,4)(1,1,0)*$_4$ is considered powerful to be employed to forecast the cargo throughput at Hong Kong port in the next three quarters of 2013 as shown in Table 4.

Table 4. Forecast throughput at Hong Kong port

Quarter	Forecast (1,000 tons)
Q2-2013	72,342
Q3-2013	73,027
Q4-2013	73,968

3.2 Container throughput at Kaohsiung port

By using the same approach, we now consider the second case studies at Kaohsiung port which is the largest container port in Taiwan and an important port in the international transportation. Historical data of the container throughput including incoming and outgoing container throughput at Kaohsiung port from January 2004 – March 2013 are obtained from the monthly statistical data published on the website of Ministry of Transportation and Communication R.O.C (MOTC) [20]. There are totally 111 observations recorded which are plotted as in Figure 2. It is well noted that there is a close relationship between the incoming and outgoing container throughput in the time investigated. And, therefore, in this paper, we only consider the total throughput for brevity. However, only data from January 2004 – December 2012 are used to build ARIMA and Grey models whose residual series are then modified with Fourier series because data from January 2013 – March 2013 are used to check the forecast power of the modified models before one of them is employed to forecast the container throughput in the other months in 2013.

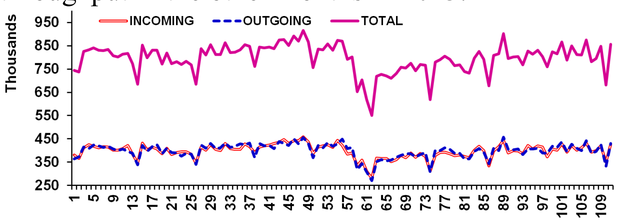

Figure 2. Monthly throughput at Kaohsiung port

3.2.1 ARIMA Model

The original time series has seasonality characteristic and therefore, only seasonal ARIMA model is considered in this section. At one degree of both non-seasonal and seasonal difference, the series becomes stationary and there are four possible SARIMA models found, including:

SARIMA(1,1,1)(0,1,0)$_{12}$ *SARIMA(1,1,4)(0,1,0)*$_{12}$
SARIMA(4,1,1)(0,1,0)$_{12}$ *SARIMA(4,1,4)(0,1,0)*$_{12}$

Among the four models, *SARIMA(4,1,4)(0,1,0)*$_{12}$ (KA-1) was found to be superior in term of lowest MAPE. Therefore, it is selected for further improvement with Fourier series, which results in *FSARIMA(4,1,4)(0,1,0)*$_{12}$ (KA-2).

3.2.2 Grey Model

From the data, GM(1,1) (KA-3) is obtained as:

$$\hat{x}_k^{(1)} = -2833807980e^{-0.00028573(k-1)} + 2834552950 \quad \left(k = \overline{1,n}\right)$$

Its Fourier modified model is named as FGM(1,1) (KA-4).

The performance of these four models is briefly summarized as in Table 5.

Table 5. Model accuracy (Kaohsiung port)

Model	KA-1	KA-2	KA-3	KA-4
MAPE	0.04038	0.00343	0.05780	0.00553
C	0.62089	0.04260	0.98980	0.08059
P	0.74490	1.00000	0.62963	1.00000
ρ	0.95962	0.99657	0.94220	0.99447
Forecasting power	Qualified	Very good	Un-qualified	Very good

Based on Table 5, both Fourier modified models (KA-2 and KA-4) are considered as very good in term of forecasting power. In order to further evaluate the forecasting power of these two models, we now compare the forecast values (unit: TEUs) in the first quarter of 2013 with the actual observations in the same period as in Table 6.

Table 6. Checking forecasting power

Month	Actual	KA-2	KA-4
Jan. 2013	847,324	844,612	763,724
Feb. 2013	681,878	684,590	768,829
Mar. 2013	857,143	854,431	770,823
MAPE		0.00344	0.10896

With the low MAPE value of 0.00344, *FSARIMA(4,1,4)(0,1,0)*$_{12}$ (KA-2) is finally selected to be employed to forecast the container throughput (unit: TEUs) at Kaohsiung port in the next three quarters of 2013 as shown in Table 4.

Table 7. Forecast throughput at Kaohsiung port

Month	Forecast	Month	Forecast
Apr. 2013	862,094	Sep. 2013	735,468
May. 2013	762,114	Oct. 2013	847,740
Jun. 2013	799,701	Nov. 2013	729,843
Jul. 2013	816,934	Dec. 2013	779,959
Aug. 2013	780,505		

4 Conclusion

The accuracy level of a traditional ARIMA model can be significantly improved with the procedure called Fourier residual modification. Particularly, in

ISBN: 978-960-474-323-0

the case of the cargo throughput at Hong Kong port and Kaohsiung port, despite of the hard assessment of relevant data about its determinants, the quarterly throughput at Hong Kong port can still be forecasted effectively with $FSARIMA(4,1,4)(1,1,0)_4$; and, the monthly container throughput at Kaohsiung port can be forecasted with $FSARIMA(4,1,4)(0,1,0)_{12}$. It can be also concluded that the cargo throughput has seasonal characteristic. Precise forecasting results obtained from these two models are essential for the control and scheduling of the two port systems and for the terminal operators in decision making and planning.

References:

[1] Peng, W.Y., and Chu, C.W., A comparison of univariate methods for forecasting container throughput volumes, *Mathematical and Computer Modelling*, Vol. 50, 2009, pp. 1045-1057.

[2] Chen, S.H., and Chen, J.N., Forecasting container throughputs at ports using genetic programming, *Expert Systems with Applications*, Vol. 37, 2010, pp.2054-2058.

[3] Pawlas, K., and Zall, D., Analysis of forecasting methods and applications of system dynamics and genetic programming: Case studies on country throughput, *Master's thesis of the School of Computing at Blekinge Institute of Technology, Sweden*, 2012.

[4] Jain, C.L., Benchmarking forecasting models, *Journal of Business Forecasting*, Vol. 26, 2007, pp. 15-18 & 35.

[5] Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R., The accuracy of extrapolation (time series) methods: Results of a forecasting competition, *Journal of Forecasting*, Vol. 1, 1982, pp. 111-153.

[6] Chou, C.C., Chu, C.W., and Liang, G.S., A modified regression model for forecasting the volumes of Taiwan's import containers, *Mathematical and Computer Modelling*, Vol. 47, 2008, pp. 797-807.

[7] Chou, C.C., Chu, C.W., and Liang, G.S., Competitiveness analysis of major ports in Eastern Asia, *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 5, 2003, pp. 682-697.

[8] Gosasang, V., Chandraprakaikul, W., and Kiattisin, S., An application of neural networks for forecasting container throughput at Bangkok port, *Proceedings of the World Congress on Engineering 2010*, Vol. 1, 2010, London, UK.

[9] Hsu, C.C., and Chen, C.Y., Applications of improved grey prediction model for power demand forecasting, *Energy Conversion and Management*, Vol. 44, 2003, p.2241-2249.

[10] Liu, S.F., Advances in Grey system theory and its applications, *Proceedings of 2007 IEEE International Conference on Grey Systems and Intelligent Services*, 2007.

[11] Hanke, J.E. and Wichern, D.W., *Business Forecasting,* 8th edition, Pearson, Prentice Hall, New Jersey, 2005.

[12] Nguyen, T.L., Shu, M.H., Huang, Y.F., and Hsu, B.M., Accurate forecasting models in predicting the inbound tourism demand in Vietnam, *Journal of Statistics and Management Systems*, Vol. 16, 2013, pp. 25-43.

[13] Ljung, G.M., and Box, G.E.P., On a measure of lack of fit in time series models, *Biometrika*, 65, 1978, 297-303.

[14] Hsu, L.C., Applying the grey prediction model to the global integrated circuit industry, *Technological Forecasting & Social Change*, Vol. 70, 2003, pp. 563-574.

[15] Guo, Z., Song, X. and Ye, J., A Verhulst model on time series error corrected for port throughput forecasting, *Journal of the Eastern Asia society for Transportation studies*, Vol. 6, 2005, pp. 881-891.

[16] Kan, M.L., Lee, Y.B. and Chen, W.C., Apply grey prediction in the number of Tourist, in *The fourth international conference on Genetic and Evolutionary computing* in Shenzhen, China, 2010, pp. 481-484.

[17] Askari, M., and Fetanat, A., Long-term load forecasting in power system: Grey system prediction-based models, *Journal of Applied Sciences*, Vol. 11, 2011, pp. 3034-3038.

[18] Huang, Y.L. and Lee, Y.H, Accurately forecasting model for the Stochastic Volatility data in tourism demand, *Modern Economy*, Vol. 2, 2011, pp. 823-829.

[19] Census and Statistics Department of Hong Kong, Table 085: Port cargo throughput, [Online], *http://www.censtatd.gov.hk/hkstat/sub/sp130.jsp?tableID=085&ID=0&productType=8* [Access: May 15, 2013].

[20] Ministry of Transportation and Communication R.O.C., Table 5-14: Container throughput by international commercial ports in Taiwan area, [Online], *http://www.motc.gov.tw/en/home.jsp?id=609&parentpath=0,154* [Access: May 15, 2013].