Forecasting Cargo Throughput with Modified Seasonal ARIMA Models

MING-HUNG SHU
National Kaohsiung Uni. of Applied Sciences
Dept. of Ind. Engineering & Management
415, Chien Kung Rd., Kaohsiung 80778
Taiwan, R.O.C
workman@cc.kuas.edu.tw

THANH-LAM NGUYEN*
Lac Hong University
10, Huynh Van Nghe, Bien Hoa, Dong Nai
Vietnam
National Kaohsiung Uni. of Applied Sciences
Grad. Inst. of Mechanical & Precision Engineering
415, Chien Kung Rd., Kaohsiung 80778
Taiwan, R.O.C

* *Corresponding author:* green4rest.vn@gmail.com

BI-MIN HSU
Cheng Shiu University
Dept. of Industrial Engineering & Management
840, Chengcing Rd., Kaohsiung 83347, Taiwan
bmhsu@csu.edu.tw

CHUNWEI LU
Shu-Zen Junior College of Medicine & Management
Department of Applied Japanese
452, Huanqiu Rd., Kaohsiung 82144, Taiwan
wini@hotmail.co.jp

JUI-CHAN HUANG
National Kaohsiung Uni. of Applied Sciences
Dept. of Ind. Engineering & Management
415, Chien Kung Rd., Kaohsiung 80778, Taiwan
y7891225@yahoo.com.tw

Abstract: The importing-and-exporting goods normally requires organizational cooperation among different international sectors ranging from purchasing, manufacturing, transporting, inventory, to distribution centers. For having their smooth coordination in the international trade, accurately forecasting the volume of imported-exported goods has become the core issue. This paper aims to obtain efficient forecasting models for estimating the cargo throughput at sea ports. We comprehensively compare the predicting performance of traditional autoregressive integrated moving average (ARIMA) models, Grey models, and their residual Fourier-modified models. The forecasting accuracy of the conventional models combined with the Fourier modification in residual is discovered to be significantly boosted. Two empirical studies of cargo and container throughputs collected from the Hong Kong and Kaohsiung ports were investigated, where two Fourier-modified seasonal ARIMA models, $FSARIMA(4,1,4)(1,1,0)_4$ and $FSARIMA(4,1,4)(0,1,0)_{12}$, are strongly suggested due to their superior forecasting powers. Finally, both models are further employed to provide the valid forecasts of two-port cargo and container throughputs in 2013.

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 throughput used in international trade, several forecasting models have been proposed as reviewed by Peng and Chu [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]. Traditional regression is used to identify causal relationship between dependent and in-dependent variables. This method performs nicely if the data series are stationary and no special pattern or trend is displayed. If this assumption is violated, the strength of the relationship is likely to be inflated which will result in an illusion of causal relationship. In practice, there are several circumstances where there is a certain causal relationship between unrelated variables, called spurious regression. Recently, the regression approach has been less preferred in favor of a heuristic approach which considers a combination of historical data and several microeconomic and macroeconomic variables.

Maritime economy has been considered extremely complex. In order to get a full understanding of its impacts, it is critical to identify its major variables. There are many different determinants of the cargo throughput at a certain port, such as the world economy, seaborne commodity trade, average haul, transport cost, political events, economic and social shocks. It is obviously that several variables can't be easily accessed; and there are still many other determinants that haven't been identified. Due to these limitations, in order to forecast the cargo throughput, we propose using Grey model, 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 [7,8]. Grey forecasting models have recently popularly used in time-series forecasting due to their simplicity and ability to characterize an unknown system with few data points.

Recorded data of cargo/ container throughput at a certain port is actually a time series. Thus, Autoregressive Integrated Moving Average (ARIMA) model is suggested because it is well-known for its special ability in dealing with time series. Unlike other forecasting models, ARIMA can efficiently work with unknown underlying economic model or structural relationships of the data set which are assumed that past values of the series plus previous error terms contain information for the purposes of forecasting. ARIMA forecasting model is advantageous because it only requires data of the interested variable be a time series. Moreover, ARIMA model usually outperforms more sophisticated structural models in providing robust short-term forecasts [9,10]; it is therefore preferably used in practice. As a consequence, the ARIMA forecasting technique is also selected and presented in this paper. The following two subsections respectively review the two traditional models.

1.1 Grey Model

Basically, Grey theory has been considered as a good literature to deal with several practical circumstances where the collection of data for analysis may have not only certain difficulties in obtaining expected number of observations but also problems of poor information which is said to be "partial known, partial unknown" [7,8]. A part of its advances and several applications was reviews by Liu [8]. A Grey dynamic model, also called Grey Model (GM), the core of the theory, offers a good approach to execute the short-term forecasting operation with no strict hypothesis for the distribution of the original data series [11]. The general GM model comes in 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. For $d = 1$ and $v = 1$, we have the basic model of Grey model, $GM(1, 1)$, a first-order differential model with one input variable. $GM(1, 1)$ has been successfully applied to many different areas, such as geology, oil and natural gas distribution, medicine, hydrology, agriculture, waste flow allocation planning, image compression, shipboard fire-detection system, hazard and alert time determination for an aircraft flying in windshear field, multi-radar low-altitude little target tracking system, processing of measuring data in reverse engineering [8]. Recently, there have been many other applications of $GM(1, 1)$ in forecasting power system [13], port throughput [14], global integrated circuit industry [15], logistics demand [16], crude oil production and consumption in China [17], inventory in global supply chain [18], the rigidity change of a linear motion guide [19], accounting earnings [20], and tourism demand [11,21,23].

$GM(1, 1)$ is obtained based on the following procedure.

Step 1:

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

$$x^{(0)} = \{x_1^{(0)}, x_2^{(0)}, \ldots, x_k^{(0)}, \ldots, x_n^{(0)}\}. \quad (1)$$

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

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

$$x^{(1)} = \{x_1^{(1)}, x_2^{(1)}, \ldots, x_k^{(1)}, \ldots, x_n^{(1)}\}. \quad (2)$$

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

Step 3:

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

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

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:

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

where:

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

$$Y = \begin{bmatrix} x_2^{(0)} \\ x_3^{(0)} \\ \vdots \\ x_n^{(0)} \end{bmatrix}$$

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

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

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

$$\hat{x}^{(0)} = \{\hat{x}_1^{(0)}, \hat{x}_2^{(0)}, \cdots, \hat{x}_k^{(0)}, \cdots, \hat{x}_n^{(0)}\} \quad (6)$$

where:

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

1.2 ARIMA Model

There are several practical phenomenons whose data are presented in time series with seasonal characteristic. A seasonal time series is defined as a series with a regular pattern of changes that repeats over S time-periods, i.e., the average values at some particular times within the seasonal intervals are usually significantly different from those at other times. Thus, a seasonal time series is usually a non-stationary series which should be made stationary by using either differencing or logging techniques before ARIMA models are used to do the forecasting for the series.

1.2.1 Non-seasonal ARIMA model

The non-seasonal ARIMA model usually has the form of $ARIMA(p,d,q)$, where:

- p is the number of lags of the differenced series appeared in the forecasting equation, called autoregressive parameter,
- d is the difference levels to make a time series stationary, called integrated parameter, and
- q is the number of the lags of the forecast errors, called moving-average parameter. "Auto-Regressive" term refers to the lags of the differenced series appeared in the forecasting equation and "Moving Average" term refers to the lags of the forecast errors. This "Integrated" term refers to the difference levels to make a time series stationary.

1.2.2 Seasonal ARIMA model

The variation of a time series is usually affected by several different factors, including seasonality. Seasonality may make several non-stationary time series significantly vary. And, due to the environmental influence, such as periodic trend, the variations induced by seasonal factor sometimes dominate the variations of the original series. A seasonal time series is usually a non-stationary time series that follows some kind of seasonal periodic trend and can be made stationary by seasonal differencing which is defined as a difference between one value and another one with lag that is a multiple of S.

Seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model with the form of $SARIMA(p,d,q)(P,D,Q)_S$, where:

- p, d, q are the parameters in non-seasonal ARIMA model as mentioned above.
- P is the number of seasonal Autoregressive order,
- D is the number of seasonal differencing,
- Q is the number of seasonal Moving Average order, and
- S is the time span of repeating seasonal pattern.

An ARIMA/SARIMA model is obtained based on the following procedure with three fundamental steps [22], including:

Step 1: Identifying the possible models

- Examine the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) graphs to identify non-seasonal terms.

  Before identifying possible ARIMA models for a time series, it is critical to make sure that the series is stationary. If it is not, it must be transformed by either differencing or logging to become stationary. For auto-regression (AR) or auto-regression moving average (ARMA) models, it is mandatory that the modulus of the roots of the AR polynomial be greater than unity, and, for the moving average (MA) part to be invertible, it is also crucial that the roots of the MA polynomial lie outside the unit circle [9].

  By differencing the seasonal time series to make it stationary, we can easily determine the difference order of differencing required rendering the series stationary before identifying an appropriate ARMA form to model the stationary series. Traditionally, Box-Jenkins procedure is frequently used, which is a quasi-formal approach with model identification relying on subjective assessment of plots of auto-correlation function (ACF) and partial auto-correlation function (PACF) of the series [9]. A time series is considered stationary if the lag values of the ACF cut off or die down fairly quickly. If the series is not stationary, it should be differenced gradually until it is considered stationary. Then, the d value in the model is obtained. If ACF graph cut off after lags q fairly quickly and PACF graph cut off after lags p fairly quickly, $ARMA(p, q)$ is achieved. $ARIMA(p, d, q)$ is accordingly identified.
- Examine the patterns across lags that are multiples of S to identify seasonal terms. Judge the ACF and PACF at the seasonal lags in the same way.

Step 2: Fitting the model.

In this step, the parameters of the model are estimated. Nowadays, with the advancement of the science and technology, the parameter estimation is usually done with the assistance of computational software, such as STATA, Eviews, SPSS, etc.

Step 3: Testing the model for adequacy.

This step formally assesses each of the time series models, and involves a rigorous evaluation of the analytical tests for each of the competing models. Because different models may wisely perform similarly, their alternative formulations should be retained for further assessment at the stage of checking forecasting power of the models.

There are several analytical methods available for testing the models. Among them, there are two popularly used, including plotting the residuals of the estimated model to detect either any outliers that may affect parameter estimates or any possible auto-correlation or heteroscedasticity problems; and, plotting the ACF and PACF of the residuals to check the model adequacy. The residuals from the model must have normal distribution and be white-noise (also known as random). This test can be done with one of the following ways:

- Testing the normal distribution of the residuals by considering the normal probability plot and testing the white-noise of the residuals by considering its ACF and PACF graphs where individual residual autocorrelation should be small and its value is within $\pm 2/\sqrt{n}$ from the central point of zero.
- Ljung-Box Q statistic [12]:

$$Q_m = n(n+2)\sum_{k=1}^{m} \frac{e_k^2}{n-k} \qquad (7)$$

  where: e_k is the residual autocorrelation at lag k; n is the number of residuals; and, m is the number of time lags includes in the test.

  The model is considered adequate only if the p-value associated with the Ljung-Box Q Statistic is higher than a given significance.

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

1.3 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 [11, 13-15, 21]. Hence, this methodology is proposed in this paper. The procedure to obtain the modified model is as the following.

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

$$\varepsilon^{(0)} = \left\{\varepsilon_2^{(0)}, \varepsilon_3^{(0)}, \cdots, \varepsilon_n^{(0)}\right\} \quad (8)$$

where: $\varepsilon_k^{(0)} = x_k^{(0)} - f_k^{(0)} \left(k = \overline{2,n}\right)$ and $f_k^{(0)}$ is the forecasted value at the k^{th} entry.

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

$$\varepsilon_k^{(0)} = \frac{a_0}{2} + \sum_{i=1}^{F} \left[a_i \cos\left(\frac{2\pi ik}{n-1}\right) + b_i \sin\left(\frac{2\pi ik}{n-1}\right)\right] \quad (9)$$

where: $F = [(n-1)/2 - 1]$ is called the minimum deployment frequency of Fourier series [21] and only take integer number [13-15]. And therefore, the residual series is rewritten as:

$$\varepsilon^{(0)} = P \cdot C \quad (10)$$

where:

$$P = \left(\left[\frac{1}{2}\right]_{(n-1)\times 1} \quad P_1 \quad \cdots \quad P_k \quad \cdots \quad P_F \right)$$

$$P_k = \begin{pmatrix} \cos\left(\frac{2\pi\times 2\times k}{n-1}\right) & \sin\left(\frac{2\pi\times 2\times k}{n-1}\right) \\ \cos\left(\frac{2\pi\times 3\times k}{n-1}\right) & \sin\left(\frac{2\pi\times 3\times k}{n-1}\right) \\ \vdots & \vdots \\ \cos\left(\frac{2\pi\times n\times k}{n-1}\right) & \sin\left(\frac{2\pi\times n\times k}{n-1}\right) \end{pmatrix}$$

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

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:

$$C = (P^T P)^{-1} P^T \left[\varepsilon^{(0)}\right]^T \quad (11)$$

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

$$\hat{\varepsilon}_k^{(0)} = \frac{1}{2}a_0 + \sum_{i=1}^{F} \left[a_i \cos\left(\frac{2\pi i}{n-1}k\right) + b_i \sin\left(\frac{2\pi i}{n-1}k\right)\right] \quad (12)$$

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

$$\tilde{x}^{(0)} = \{\tilde{x}_1^{(0)}, \tilde{x}_2^{(0)}, \ldots, \tilde{x}_k^{(0)}, \ldots, \tilde{x}_n^{(0)}\} \quad (13)$$

where

$$\begin{cases} \tilde{x}_1^{(0)} = f_1^{(0)} \\ \tilde{x}_k^{(0)} = f_k^{(0)} + \hat{\varepsilon}_k^{(0)} \quad (k = \overline{2,n}) \end{cases}$$

Once a forecasted series is obtained, it should be compared with the original series in order to evaluate the accuracy of the forecasting model. The absolute residual error of the k^{th} entry and its relative error, respectively denoted by ε_k and $\wp_k$, are defined as [14,25]

- Residual error: $\varepsilon_k = x_k^{(0)} - f_k^{(0)}$, $(k = 1, \ldots, n)$.
- Relative error:

$$\wp_k = \frac{|\varepsilon_k|}{x_k^{(0)}}$$

Practically, in order to better evaluate the accuracy of a forecasting model, more advanced indexes are usually taken into consideration. Among them, the following four indexes are frequently used.

- The mean absolute percentage error (MAPE) [7, 9,11,13,14,21,24,26,28-31]:

$$MAPE = \frac{1}{n}\sum_{k=1}^{n} \wp_k.$$

- The post-error ratio C [16,24,32]:

$$C = \frac{S_2}{S_1},$$

where:

$$S_1 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left[x_k^{(0)} - \frac{1}{n}\sum_{k=1}^{n} x_k^{(0)}\right]^2}$$

$$S_2 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left[\varepsilon_k - \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k\right]^2}$$

The ratio C, in fact, is the ratio between the standard deviation of the original series and the standard deviation of the forecasting error. The smaller the C value, the higher accuracy the model has since smaller C value results from a larger S_1 and/or a smaller S_2.

- The small error probability P [16,24,32]:

$$P = p\left\{\left(|\varepsilon_k - \frac{1}{n}\sum_{k=1}^{n}\varepsilon_k|/S_1\right) < 0.6745\right\}.$$

The P value indicates a probability of the ratio of the difference between the residual values of data points and the average residual value with the standard deviation of the original series smaller than 0.6745 [16,24]. Thus, the higher the P value, the higher accuracy the model has.

- The forecasting accuracy ρ [16,24]: $\rho = 1 - MAPE$.

Based on the above indexes, there are four grades of accuracy stated in Table 1.

Table 1: Four grades of forecasting accuracy

Grade level	$MAPE$	C	P	ρ
Very good	< 0.01	< 0.35	> 0.95	> 0.95
Good	< 0.05	< 0.50	> 0.80	> 0.90
Qualified	< 0.10	< 0.65	> 0.70	> 0.85
Unqualified	≥ 0.10	≥ 0.65	≤ 0.70	≤ 0.85

2 Empirical Study

In this section, we employ the both traditional and modified models of the Grey model and ARIMA model to forecast the cargo throughput at Hong Kong port and Kaohsiung Port which are two busy ports in the maps of international sea ports in Asia. The two ports are selected as typical examples in illustrating the procedure to obtain good statistical models to forecast cargo throughput. Due to the difference in their locations, they may have different models to forecast their cargo flows.

2.1 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 [33]. The data from Q1-1997 - Q3-2012, plotted in Fig.1, 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.

2.1.1 ARIMA Model

From Fig.1, it can be concluded that seasonality exists in the series of tourism demand. Therefore, only seasonal ARIMA model is considered in this section. By observing the auto-correlation function graph (ACF) and partial auto-correlation function graph (PACF) in Fig.2, we can conclude that the original series is non-stationary. However, at one degree of both non-seasonal and seasonal difference, the series becomes stationary as shown in Fig.3 and $SARIMA(4,1,4)(1,1,0)_4$ (HK-1) is found appropriate. The ACF and PACF of its residual series in Fig.4 and its histogram in Fig.5 indicate that $SARIMA(4,1,4)(1,1,0)_4$ (HK-1) is random-walk (white-noise). Hence, it is selected. Its Fourier modified model is $FSARIMA(4,1,4)(1,1,0)_4$ (HK-2).

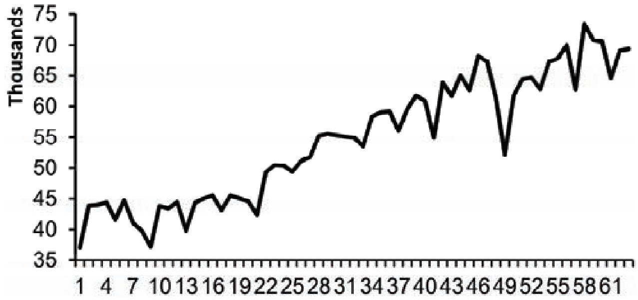
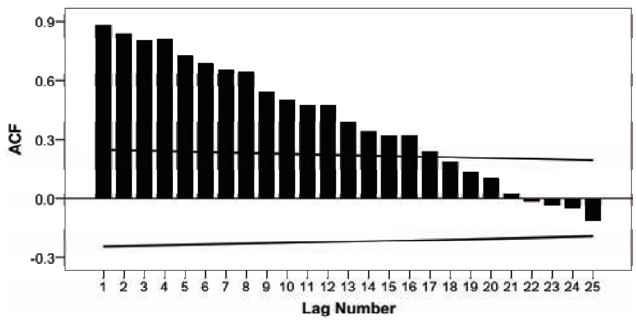
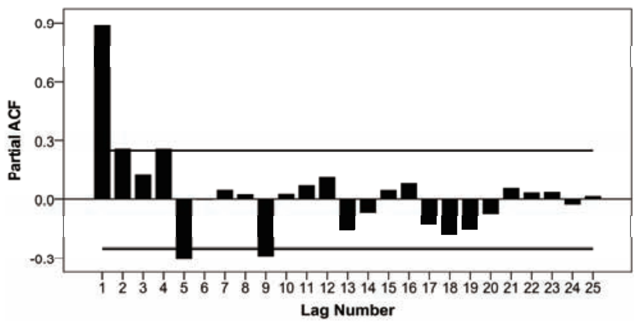


Figure 1: Quarterly throughput at Hong Kong port

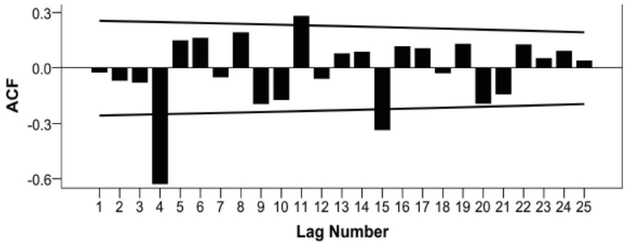


a) Auto-correlation function graph

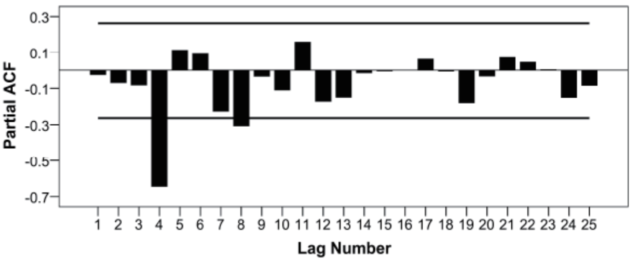


b) Auto-correlation function graph

Figure 2: ACF and PACF of original series



a) ACF of differenced series



b) ACF of differenced series

Figure 3: ACF and PACF at one degree of both non-seasonal and seasonal difference

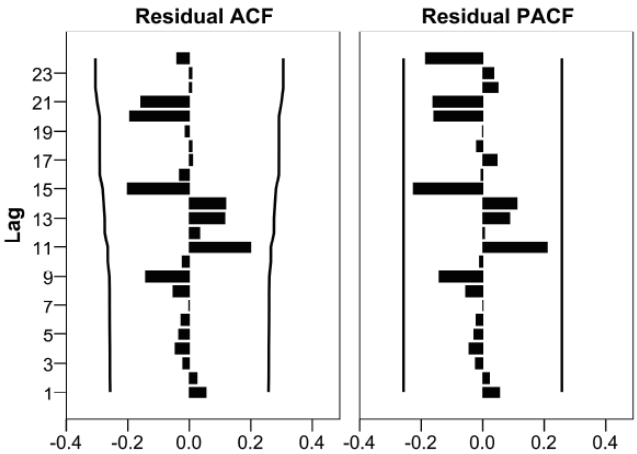


Figure 4: Noise residuals ACF and PACF

2.1.2 Grey Model

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

$$\hat{x}_k^{(1)} = 4293068.99e^{0.009425(k-1)} - 4256019.99,$$

$\left(k = \overline{1,n}\right)$. 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.

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.

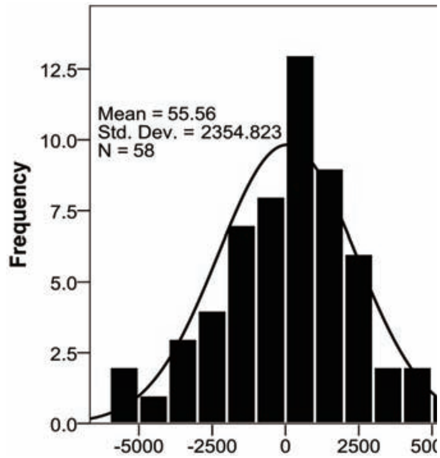


Figure 5: Histogram of Noise residuals

Table 2: Model accuracy (Hong Kong port)

Model	HK-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
Power	Good	Very good	Good	Good

Table 3: Checking forecasting power HK-2

Quarter	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

The forecasted values across in the investigated period follow closely to the actual observations as plotted in Fig. 6, indicating the good forecasting of the proposed model.

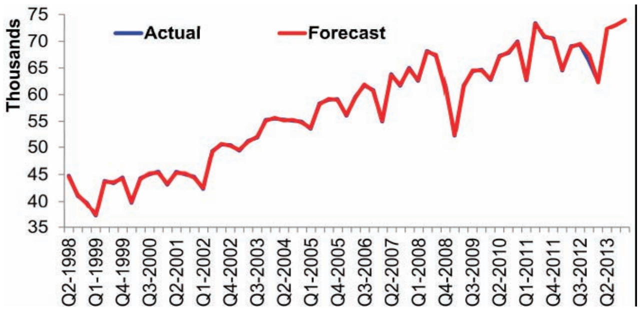


Figure 6: Actual observations versus Forecast

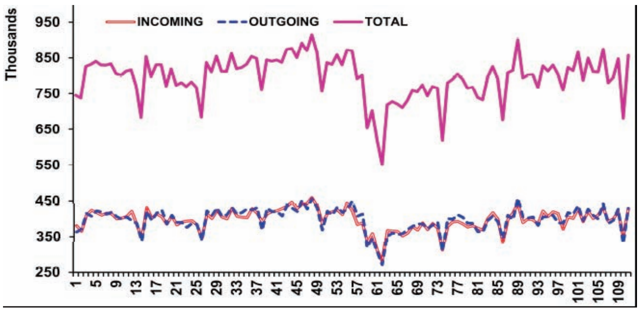


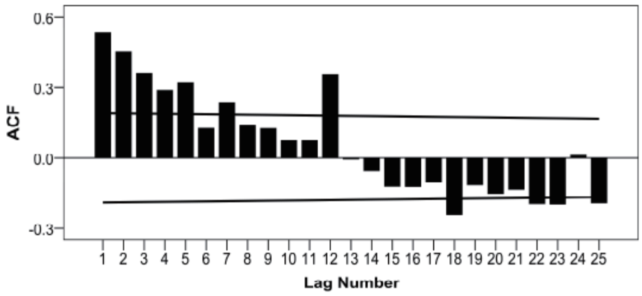
Figure 7: Monthly throughput at Kaohsiung port

2.2 Container throughput at Kaohsiung port

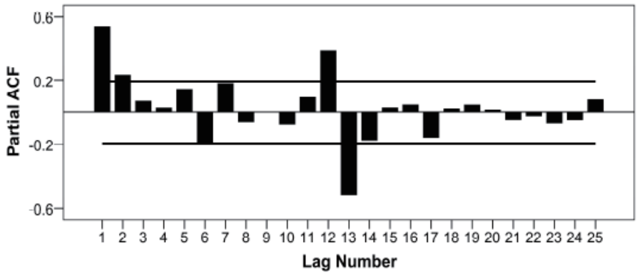
By using the same approach, we now consider the second case studies at Kaohsiung port which is the largest container port in Taiwan and an important port in the international transportation. Historical data of the container throughput including incoming and outgoing container throughput at Kaohsiung 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) [34]. There are totally 111 observations recorded which are plotted as in Fig.7. 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.

2.2.1 ARIMA Model

The original time series has seasonality characteristic and therefore, only seasonal ARIMA model is considered in this section. The ACF graph and PACF graph



a) Auto-correlation function graph



b) Partial auto-correlation function graph

Figure 8: ACF and PACF of original series

in Fig.8 show that the original series of the container throughput at Kaohsiung port is non-stationary. However, at one degree of both non-seasonal and seasonal difference, the series becomes stationary as plotted in Fig.9.

From Fig.9, four possible SARIMA models are identified, 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 possible 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).

2.2.2 Grey Model

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

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

$\left(k=\overline{1,n}\right)$. 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.

Based on Table 5, both Fourier modified models (KA-2 and KA-4) are considered very good in term of forecasting power. In order to further evaluate the

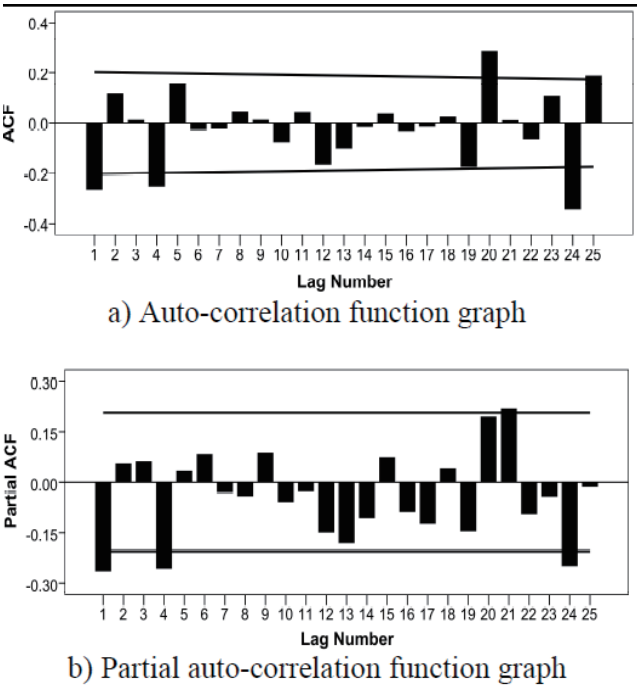


Figure 9: ACF and PACF of differenced series

Table 5: Model accuracy (Kaohsiung port)

Model	KA-1	KA-2
$MAPE$	0.04038	0.00343
C	0.62089	0.04260
P	0.74490	1.00000
ρ	0.95962	0.99657
Power	Qualified	Very good

Model	KA-3	KA-4
$MAPE$	0.05780	0.00553
C	0.9898	0.08059
P	0.62963	1.00000
ρ	0.94220	0.99447
Power	Un-qualified	Very good

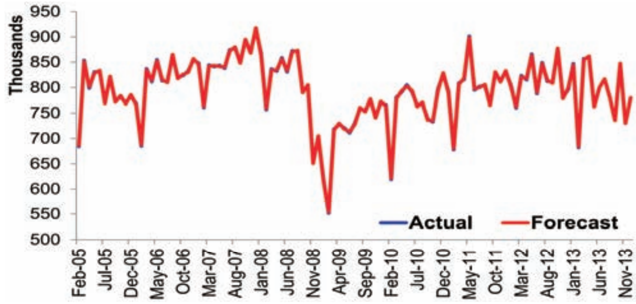


Figure 10: Actual observations versus Forecast

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 illustrated 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 7.

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		

Fig.10 further proves the accuracy of the values obtained $FSARIMA(4,1,4)(0,1,0)_{12}$ model compared to the actual container throughput at Kaohsiung port.

3 Conclusion

Due to the difficulties in assessing relevant statistical data of cargo throughput, the use of traditional regression possibly leads to spurious results. Therefore, in this paper, we proposed alternative approaches with Grey and ARIMA models. In order to enhance the accuracy of the two conventional models, their residual series are modified with Fourier series.

As typical examples, two empirical studies of the cargo throughput at Hong Kong port and Kaohsiung port were investigated. It was found that despite of the hard assessment of relevant data about its determinants, ARIMA models modified with Fourier series outperform those of Grey models. Particularly, the quarterly throughput at Hong Kong port can be effectively forecasted with $FSARIMA(4,1,4)(1,1,0)_4$ which resulted in a very low mean absolute percentage error (MAPE) of 0.01057 in comparison of the forecasted values and actual observations in the fourth quarter of 2012 and the first quarter of 2013; and, similarly, the monthly container throughput at Kaohsiung port can be forecasted with $FSARIMA(4,1,4)(0,1,0)_{12}$ with the MAPE of 0.00344 in forecasting the throughput in the first three months of 2013. The two ports have different forecasting models due to the difference in their location which affects the cargo flows.

From these two examples, the Fourier residual modification has been further proved to improve the accuracy level of the traditional ARIMA model. With the usage of seasonal ARIMA models, we can obviously conclude the existence of seasonality in the time series of cargo throughput. More importantly, the low MAPE values of the two obtained models demonstrate their applicability in practice to provide accurate forecasts. 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.

Acknowledgements: The authors are grateful to the anonymous referees who made some useful comments and suggestions for the improvement of the paper. This work was partially supported by the National Science Council of Taiwan under grants NSC 101-2628-E-151-002-MY3 and NSC 101-2410-H-230-021-MY2.

References:

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

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

[3] K. Pawlas and D. Zall, 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* , 20112.

[4] C. L. Jain, Benchmarking forecasting models, *J. Bus. Forecasting* 26, 2007, pp. 15-18 & 35.

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

[6] C. C. Chou, C.W. Chu and G. S. Liang, The accuracy of extrapolation (time series) methods: Results of a forecasting competition A modified regression model for forecasting the volumes of Taiwan's import containers, *Math. Comput. Modell.* 47, 2008, pp. 797-807.

[7] C. C. Hsu and C. Y. Chen, Applications of improved grey prediction model for power demand forecasting, *Energy Convers. Manage.* 44, 2003, pp. 2241-2249.

[8] F. Liu, Advances in Grey system theory and its applications, *Proc. 2007 IEEE Int. Conf. Grey Syst. Intell. Serv.*, 2007.

[9] A. Meyler, G. Kenny and T. Quinn, Forecasting Irish inflation using ARIMA models, *Technical paper 3/RT/98*, 1998.

[10] S. Abdullah, M. D. Ibrahim, A. Zaharim and Z.M. Nopiah, Statistical Analysis of a Nonstationary Fatigue Data Using the ARIMA Approach, *WSEAS Trans. Math.* 7, No.2, 2008, pp.59-66.

[11] M. L. Kan, Y. B. Lee and W. C. Chen, 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.

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

[13] M. Askari and A. Fetanat, Long-term load forecasting in power system: Grey system prediction-based models, *J. Appl. Sci.* 11, 2011, pp. 3034-3038.

[14] Z. Guo, X. Song and J. Ye, A Verhulst model on time series error corrected for port throughput forecasting, *J. East. Asia Soc. Transp. Stud.* 6, 2005, pp. 881-891.

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

[16] J. Hua and Q. Liang, Application research of the Grey forecast in the Logistics demand forecast, *First Int. Workshop Edu. Technol. Comput. Sci.* 3, 2009, pp. 361-363.

[17] H. Ma and Z. Zhang, Grey prediction with Markov-Chain for Crude oil production and consumption in China, *Adv. Intell. Soft Comput.* 56, 2009, pp. 551-561.

[18] S. J. Wang, W. L. Wang, C. T. Huang and S. C. Chen, Improving inventory effectiveness in RFID-enabled global supply chain with Grey forecasting model, *J. Strategic Inf. Syst.* 20, 2011, pp. 307–322.

[19] Y. F. Hsiao, Y. S. Tarng and K. Y Kung, Comparison of the grey theory with neural networks in the rigidity prediction of linear motion guide, *WSEAS Trans. Appl. Theor. Mech.* 4, No.1, 2009, pp. 31–41.

[20] Z. L. Chen, C. C. Chiu and C. J. Lai, Grey group model forecasting of quarterly accounting earnings, *WSEAS Trans. Inf. Sci. Appl.* 6, No.8, 2009, pp. 1269–1278.

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

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

[23] L. C. Hsu and C. H. Wang, Apply multivariate forecasting model to tourism industry, *Tourism Int. Interdiscip. J.* 56, No.2, 2008, pp. 159–172.

[24] T. L. Nguyen, M. H. Shu, Y. F. Huang and B. M. Hsu, Accurate forecasting models in predicting the inbound tourism demand in Vietnam, *J. Stat. Manage. Syst.* 16, No.1, 2013, pp. 25–43.

[25] P. Liang, H. Zhou and J. Zheng, Forecast research of droplet size based on Grey theory, *IFIP Adv. Inf. Commun. Technol.* 258, 2008, pp. 653–658.

[26] L. C. Hsu and C. H. Wang, The development and testing of a modified Diffusion model for predicting tourism demand, *Int. J. Manage.* 25, 2008, pp. 439–445.

[27] P. Gonzlez, and P. Moral, An analysis of the international tourism demand in Spain, *Int. J. Forecasting* 11, 1995, pp. 233–251.

[28] T. L. Nguyen, M. H. Shu and B. M. Hsu, Forecasting international tourism demand- An empirical case in Taiwan, *Asian J. Empirical Res,* 3, No.6, 2013, pp. 711–724.

[29] Y. W. Chang and M. Y. Liao, A seasonal ARIMA model of tourism forecasting: The case of Taiwan, *Asia Pac. J. Tourism Res.* 15, 2010, pp. 215–221.

[30] M. Memmedli and O. Ozdemir, An application of fuzzy time series to improve ISE forecasting, *WSEAS Trans. Math.* 9, No.1, 2010, pp. 12–21.

[31] D. Aydin and M. Mammadov, A comparative study of hybrid, neural networks and nonparametric regression models in time series prediction, *WSEAS Trans. Math.* 8, No.10, 2009, pp. 593–603.

[32] R. C. Tsaur and T. C. Kuo, The adaptive fuzzy time series model with an application to Taiwan's tourism demand, *Expert Syst. Appl.* 38, 2011, pp. 9164–9171.

[33] 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].

[34] 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].