Time series forecasting of domestic shipping market: comparison of SARIMAX, ANN-based models and SARIMAX-ANN hybrid model

Cemile Solak Fiskin*

Department of Maritime Business Administration,
Ordu University,
Ordu, Turkey
Email: cemilesolak@odu.edu.tr
*Corresponding author

Ozgu Turgut and Sjur Westgaard

Department of Industrial Economics and Technology Management,
Norwegian University of Science and Technology,
Trondheim, Norway
Email: ozgu.turgut@ntnu.no
Email: sjur.westgaard@ntnu.no

A. Güldem Cerit

Maritime Faculty,
Dokuz Eylul University,
İzmir, Turkey
Email: gcerit@deu.edu.tr

Abstract: Seaborne transport forecasting has attracted substantial interest over the years because of providing a useful policy tool for decision-makers. Although various forecasting methods have been widely studied, there is still broad debate on accurate forecasting models and preprocessing. The current paper aims to point out these issues, as well as to establish the forecasting model of the domestic cargo volumes using SARIMAX, MLP, LSTM and NARX and SARIMAX-ANN hybrid models. Based on the domestic cargo volumes of Turkey, findings suggest that SARIMA-MLP models can be considered as an appropriate alternative, at least for time series forecasting of shipping. Pre-processed data provides a significant improvement over those obtained with unpreprocessed data, with the accuracy of the models found to be significantly boosted with the Fourier term of decomposition. The results indicate that SARIMAX-MLP, with a mean absolute percentage error (MAPE) of 4.81, outperforms the closest models of SARIMAX, with a MAPE of 6.14 and LSTM with Fourier decomposition with a MAPE of 6.52. Findings have implications for shipping policymakers to plan infrastructure development, and useful for shipowners in accurately formulating shipping demand.

Keywords: time series forecasting; shipping; artificial neural network; ARIMA; machine learning; hybrid model.

Reference to this paper should be made as follows: Fiskin, C.S., Turgut, O., Westgaard, S. and Cerit, A.G. (2022) 'Time series forecasting of domestic shipping market: comparison of SARIMAX, ANN-based models and SARIMAX-ANN hybrid model', *Int. J. Shipping and Transport Logistics*, Vol. 14, No. 3, pp.193–221.

Biographical notes: Cemile Solak Fiskin has BSc, MSc and PhD degrees from the Maritime Business Administration of the Maritime Faculty, Dokuz Eylul University in Izmir, Turkey. She is a Research Assistant at the Maritime Business Administration Department at Ordu University, Turkey. She focuses on machine learning, trade forecasting and shipping.

Ozgu Turgut received her BSc from the Industrial Engineering Department at Bogazici University in Istanbul, Turkey. She entered this department by ranking within the top 0.01% on the national college entrance exam. She then received her MSc at the Systems Engineering Department of Yeditepe University and PhD from the Industrial and Systems Engineering Department at Wayne State University in Michigan. Prior to starting her position at the Norwegian University of Science and Technology, she worked as a Postdoc Associate at the MIT's Center for Transportation and Logistics for one year and at LLamasoft Inc. for three years as a research scientist.

Sjur Westgaard obtained his MSc and PhD in Industrial Economics from the Norwegian University of Science and Technology and MSc in Finance from the Norwegian School of Business and Economics. He is now a Professor at the Norwegian University of Science and Technology, and also has an adjunct position at the Norwegian University of Life Sciences. His teaching involves managerial economics, financial accounting, corporate finance, derivatives and real options, empirical finance and financial risk management. He has been a project manager for several research projects involving power companies and the Norwegian Research Council. He is also a co-editor of the *Journal of Energy Markets* and has been a co-editor of the *Journal of Banking and Finance*.

A. Güldem Cerit received her BSc from the Engineering School of the Middle East Technical University in Ankara, Turkey. She worked in the private industry for nine years at various engineering and marketing positions, all at managerial levels. She completed the MBA program and PhD in the marketing discipline of the Business Program at the Dokuz Eylul University. She joined the Dokuz Eylul University School of Maritime Business and Management in 1993 as an Assistant Professor. She has served as the Director/Dean of the school between 1997–2017. She received her Associate Professor and Professor degrees in the maritime transport business discipline.

1 Introduction

The need for accurate forecasting stems from the perishable nature of the services in the shipping industry. Forecasting the future is crucial for shipping industry stakeholders to make investment decisions on ordering a ship, establishing the chartering type, planning the future, gaining more profit and calculating the risk [Stopford, (2009), p.697]. Such forecasts are necessary in particular for Turkey, due to the potential to modal shift from road to sea. A total of 95% of domestic shipping in Turkey is done through road transport

(Çelikkaya, 2012). Turkey attaches an increasing importance on domestic shipping due to road transport excess and as a requirement of Turkey's EU accession process (Özen, 2007). Domestic shipping is a sustainable alternative to road transport. Forecasting domestic shipping in Turkey is essential to plan for the aging of Turkish flagged coasters, to develop incentives to Turkish maritime transport and to generate prospects for the future of Turkey's domestic shipping. A failure to forecast the demand of domestic shipping market may lead to discrepancies between demand and supply, errors in ship and port investments and dwell time in ports and congestion, and consequently the loss of stakeholders. In order to overcome these problems, it is necessary to apply accurate forecasting models.

When the demand for shippers' commodities increases, shipowners are forced to increase their fleet size. If a shipowner decides to invest in new ships, this increased capacity creates incremental changes in the entire supply chain. In order to overcome this capacity problem, which is based on the derived and discontinuous demand characteristics of domestic shipping, shipowners need to obtain good market forecasts. Nevertheless, there is no consensus on the best forecasting models for shipping. With this motivation, our paper benchmarks forecasting models for the players involved in domestic shipping in a developing country case. In this respect, developing countries are unique in the sense that they generally do not have continuous data on related variables for a convenient model, and insufficient data restricts the use of various techniques [Akal, (2004), p.568].

Based on the above-mentioned reasons, shipping markets have been forecast for years, with different forecasting models having been developed. Studies have generally focused on specific ports or regions, and mainly include containerised cargo in order to increase the accuracy and to cope with data collection limitations. This study also attempts to establish the model of forecasting domestic cargo volumes. Therefore, it focuses on a wider geographic area, and includes diversified types of cargo.

In terms of the methods used, SARIMAX, MLP, LSTM and NARX and SARIMAX-ANN hybrid models are developed in order to compare the models. Several researchers have emphasised the seasonal characteristics of the cargo throughput [e.g., Chen and Chen, 2010; Farhan and Ong, 2016; Shu et al., (2013), p.193]. The SARIMA model is selected, in that data has the seasonal time series characteristics. It also has the capability to increase the explanatory power of the model with the help of exogenous variables. The research is extended with a neural network and hybrid technique that use the methodology developed by Zhang (2003). The selected hybrid model consists of both time series and a soft computing method. Furthermore, the LSTM and NARX models are also included in order to improve the benchmarking portfolio. To the best of our knowledge, none of these methods have previously been used in maritime. In addition, data pre-processing is considered while comparing the methods used to increase accuracy.

This paper proceeds as follows: Section 2 briefly reviews the relevant literature. Section 3 proposes the theoretical framework of the used time series methods, whereas Section 4 presents the empirical study and research findings. Finally, Section 5 concludes by discussing the study's contributions and implications for future research.

2 Literature review

Research about forecasting and modelling shipping markets has a long history, although it has attracted a particularly wide amount of attention in the last few decades. The earliest studies go back to Tinbergen (1959), Beenstock (1985), Beenstock and Vergottis (1993) and Charemza and Gronicki (1981), which are also generally shown as theoretical foundation studies. Several articles on maritime forecasting have been published after 2000, with the majority of these studies indicating empirical research and a focus on quantitative models that use different forecasting methods such as econometric modelling, time series modelling and soft computing techniques. Reviewed studies have generally concentrated on a specific region in order to increase their accuracy, and to cope with data collection limitations.

Seaborne trade forecasting has especially guided issues about planning, design, supervision, maritime safety [Feng et al., (2011), p.446], economic cooperation development, construction and renovation [Li et al., (2015), p.243] and traffic control [Lv et al., (2016), p.1]. Forecasting seaborne trade is a complex nonlinear dynamic process, though a variety of models have been developed to address this issue [Li et al., (2015), pp.243–244]. Winston (1981) predicted market demand for ocean container services with a multinomial probit model; De Gooijer and Klein (1989) forecast the maritime steel traffic flow with ARIMA and VARMA at the Port of Antwerp; Klein (1996) forecast the maritime traffic flow with SARIMA at the Port of Antwerp; Veenstra and Haralambides (2001) forecast the trade flow through the USA, North Africa, Europe, the Middle East and the Far East, Babcock and Lu (2002) forecast inland waterway traffic on the Mississippi River; Feng et al. (2011) estimated traffic flow with SVM on the Yangtze River and Li et al. (2015) forecast the traffic flow at Tianjin with ARIMA and hybrid soft computing methods. Moreover, Mostafa (2004) forecast the maritime traffic flow at the Suez Canal with ARIMA and ANN, and found that ANN forecast performance more accurately than univariate autoregressive integrated moving average (ARIMA). Goulielmos and Kaselimi (2011) used nonlinear models to forecast transshipment containers at the Port of Piraeus, and emphasised the importance of forecasting shipping demand. Lastly, Pang and Gebka (2017) forecast container throughput using aggregate or terminal-specific data, and found that SARIMA produced the worst forecasts among the VECM, ASHM and MSHW.

Several studies have attempted to formulate accurate models for the shipping trade forecast. For instance, econometrics-based models, such as simple regression (e.g., Lam et al., 2004; Jugović et al., 2011), multinominal logit and probit (e.g., Winston, 1981; Lee et al., 2017) vector autoregression models (e.g., Veenstra and Haralambides, 2001), time series techniques such as ARMA, ARIMA, SARIMA, SARIMANT (e.g., De Gooijer and Klein, 1989; Klein, 1996; Mostafa, 2004; Schulze and Printz, 2009), soft computing techniques such as neural networks, least squares support vector machines (Liu et al., 2014; Xie et al., 2013), as well as hybrid models such as projection pursuit regression and genetic programming, hybrid robust v-support vector regression and chaotic simulated annealing particle, swarm the optimisation and multivariable adaptive regression splines model (Huang et al., 2015; Geng et al., 2015), the two-state Markov model chain and Monte Carlo simulations (Grifoll, 2019). Modelling the estimation and forecasting the shipping demand has evolved with increasingly complicated models. Although few studies indicated that hybrid models generally develop more accurate results (e.g., Huang et al., 2015; Geng et al., 2015), there is no consensus on better modelling and forecasting shipping demand.

The study by Xie et al. (2013) is closely related with current work: they forecast container throughput at the Shanghai Port and Shenzhen Port, and compared LSSVM, SVM, SARIMA and hybrid approaches (SARIMA-LSSVR, SD-LSSVR and CD-LSSVR). Du et al. (2019, p.10) mentioned that hybrid models could increase the accuracy of forecasts with combining the advantages of two or more models. Xie et al. (2013) concluded that the hybrid approaches performed better than the single forecasting models. They showed the seasonal and nonlinear nature of container throughput series as the primary reason for better forecasting performance. Chan et al. (2019) benchmarked the time series forecasting methods, which include MA, MARS, ARIMA, the grey model, ANN and SVM, and found that SVM outperforms than the benchmarked models. The authors also suggested that further research is needed to include more input data to increase the accuracy of the forecast. The GMDH-based hybrid model has also been constructed by Mo et al. (2018), with this model presenting a more superior accuracy than other single and hybrid models. In addition, there are some recent time series studies, such as Li et al. (2017), which modelled a ship's motion with the NARX network based on three learning strategies. Yang and Mehmed (2019) also used NARNET and NARXNET to analyse the effect of FFA on freight index forecasting. Moeeni et al. (2017) found that ANFIS-ANN had the best accuracy among ANN, ANFIS and SARIMA-ANN, while SARIMA-ANN performed better than ANN and SARIMAX. Moeeni and Bonakdari (2017) proposed hybrid SARIMA-ANN forecasts as better than single methods.

In this study in particular, SARIMA, MLP, LSTM, and NARX are used for benchmarking. To the best of our knowledge, LSTM and NARX are used for the first time in the forecasting of shipping studies. Moreover, this study considers all types of cargoes loaded at Turkish domestic ports, whereas Xie et al. (2013) and similar port forecasting studies only focus on specific type of cargoes and specific ports. As suggested by Chan et al. (2019), more input data is also considered while developing models.

There are some papers on shipping forecasting literature that mention the pre-processing process of the data: Klein (1996) uses univariate time series models with data transformations and intervention models to forecast the volumes of 22 maritime traffic flows in the Port of Antwerp. The models obtained after preprocessing produced forecasts demonstrate a substantial improvement over those obtained with unadjusted data. Shu et al. (2013) forecast with the SARIMA model, the grey model and their joint Fourier modified models (with residuals). They also indicated that the accuracy of the conventional models is found to be significantly boosted with the Fourier modification. Recently, Moeeni et al. (2017) showed that the preprocessing of data could improve the forecasting accuracy by reducing the problematic effects of time series components.

As a result, SARIMAX-ANN and SARIMAX models provide accurate forecasts for the domestic shipping industry. While developing neural networks models, preprocessing should be considered, with a Fourier adjustment enhancing the forecasting ability. The significance of the bootstrap tests are also pointed out in the paper.

Abbreviations

ACF	Autocorrelation
ADP	Absolute deviation percent
AIC	Akaike information criterion
ANFIS	Adaptive neuro fuzzy interference systems
ANN	Artificial neural network
API under MIT	Application programmer interface under Massachusetts Institute of Technology
ARIMA	Autoregressive integrated moving average
ASHM	Additive seasonal holt-winters
CD-LSSVR	Classical decomposition-least squares support vector regression
EU	European union
GMDH	The group method of data handling
LSSVR	Least squares support vector regression
LSTM	Long short-term memory
MA	Moving average
MAD	Mean absolute deviation
MAPE	Mean absolute percentage error
MARS	Multivariate adaptive regression splines
MLP	Multilayer perceptron
MSE	Mean squared error
MSHW	Multiplicative seasonal holt-winters
NARX	Nonlinear autoregressive network with exogenous inputs
PACF	Partial autocorrelation
RMSE	Root mean square error
RNN	Recurrent neural network
SARIMA	Seasonal autoregressive integrated moving average
SARIMAX	Seasonal autoregressive integrated moving average with exogenous variables
SEP	Standard error prediction percent
SD-LSSVR	Seasonal decomposition-least squares support vector regression
STL	Seasonal trend with loess
SVM	Support vector machines
VARMA	Vector autoregressive moving average
VECM	Vector error correction model

3 Time series forecasting models

3.1 SARIMAX

ARIMA models were introduced in the 1990s, and popularised by the George Box and Gwilym Jenkins in the 1970s [Chase, (2013), p.203]. ARIMA models have been widely used in maritime research (e.g., De Gooijer and Klein, 1989; Babcock and Lu, 2002; Mostafa, 2004) as in other research areas. A seasonal ARIMA model includes the additional seasonal terms in the ARIMA model, which is called SARIMA. The SARIMA (p, d, q) (P, D, Q)s model has six parameters: autoregressive parameters (p), number of differencing passes (d), and moving average parameters (q); seasonal parameters contain seasonal autoregressive (P), seasonal differencing (D), and seasonal moving average parameters (Q); s is for the length of the seasonal period [Zhang et al., (2018), p.121]. If we want to increase forecasting performance with exogenous variables, then the model is called SARIMAX. SARIMAX is generally mathematically expressed as follows [Vagropoulos et al., (2016), p.2]:

$$\varphi_p(B)\Phi_P\left(B^s\right)\nabla^d\nabla_s^D y_t = \beta_k x'_{k,t} + \theta_q(B)\Theta_Q\left(B^S\right)\varepsilon_t$$

y_t	series at time period
$\varphi_p(B)$	the regular AR polynomial of order p
$\Phi_P(B^s)$	the seasonal AR polynomial of order P
$\theta_q(B)$	is the regular MA polynomial of order q
$\Theta_Q(B^s)$	is the seasonal MA polynomial of order Q
∇^d	the seasonal differentiating operator
∇_s^D	eliminate the non-seasonal and seasonal non-stationarity, respectively
B	the backshift operator
s	seasonal period
ε_t	errors at time period
$x_{k,t}$	the vector including the k^{th} explanatory input variables at time t
β^k	the coefficient value of the k^{th} exogenous input variable.

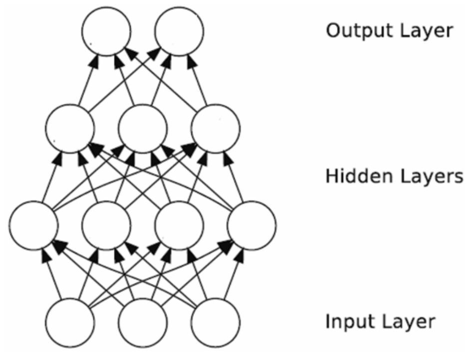
3.2 ANN-based models

ANN-based models are known as a useful tool for time series forecasting that can model the nonlinear structure [Yolcu et al., (2013), p.1340]. Various ANN structures and training algorithms have been used over time [Ihle, (2016), p.260]. There is also an extensive literature on using ANN-based models for maritime related forecasting (e.g., Mostafa, 2004; Santos et al., 2014; Eslami et al., 2016).

3.2.1 MLP

Multilayer perceptron (MLP) is the most recognised and implemented ANN structure [Zhang et al., (1998), p.37]. A MLP illustrated in Figure 1 consists of an input layer and output layer directly linked with the intermediate hidden layer [Palit and Popovic, (2006), p.84].

Figure 1 A MLP



Source: Graves (2012, p.16)

MLP can be expressed in the following equation:

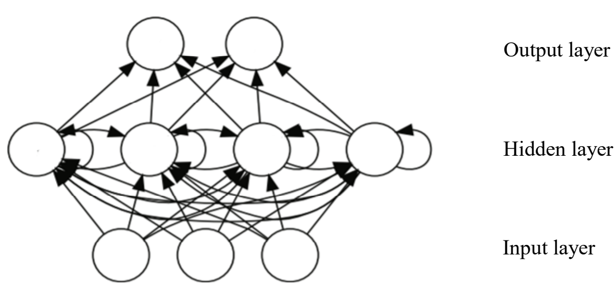
$$output = f\left(\sum_{i}^{Inputs}\left(x_i \cdot w_i + b_i\right)\right)$$

x is the input of the neuron, w is the weight on each connection to the neuron, b is the bias and $f(\ldots)$ is the activation function [Gensler et al., (2016), p.2].

3.2.2 LSTM

MLP does not contain any cycles and output only based on the current input [Witten et al., (2016), p.241]. RNN structures incorporate this dependence [Gulli and Pal, (2017), p.175], and as shown in the Figure 2 [Graves, (2012), p.22], can be mapped from the entire history of previous inputs to each output. RNN is also designed to process a sequential or time varying pattern [Medsker and Jain, (2001), p.1].

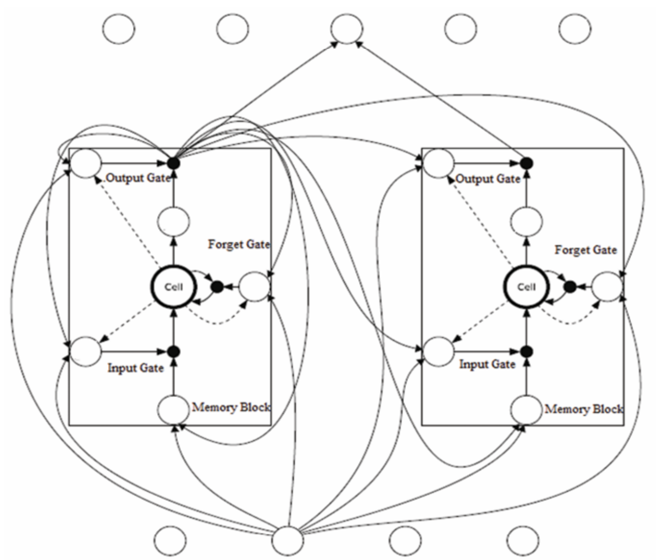
Figure 2 A recurrent neural network (RNN)



Source: Graves (2012, p.22)

However, a standard RNN has a vanishing gradient problem that resulted in the exponential degenerations of the input as it cycles around the network's recurrent connections. LSTM networks solve that problem [Graves, (2012), p.37], specifying the concept of gates and memory cells in each hidden layer [Zheng et al., (2017), p.3]. LSTM is a redesign of RNN architecture around special 'memory cell' units [Graves, (2012), p.1]. An LSTM network is illustrated in Figure 3. The input, the output and the forget gates are in each memory block and provide similar operations, such as writing, reading and resetting for the cell [Graves, (2012), p.38].

Figure 3 An LSTM network



Source: Graves (2012, p.40)

3.2.3 NARX

The NARX model is a recurrent dynamic network based on the linear ARX model. The NARX neural network model is a combination of a multilayer feed forward neural network (MLP), a RNN and a time delay neural network [Mitrea et al., (2009), p.67]. The defining equation for the NARX model is shown as follows:

$$Y_t = f\left(y_{t-1}, y_{t-2}, \ldots\ldots, y_{t-n}, x_{t-1}, x_{t-2}, x_{t-n}\right)$$

Y_t is the series at time period (output variable), *f*(..) is the nonlinear feedforward neural network and x_t is the explanatory input variable.

3.3 The hybrid methodology

Zhang (2003) proposed that a time series is to be composed of a linear autocorrelation structure and a nonlinear component as shown in the following equation, where L_t denotes the linear component and N_t denotes the nonlinear component:

$$y_t = L_t + N_t$$

The hybrid methodology consists of two steps. First, ARIMA models are used to analyse the linear component. Next, a neural network model is used to model the residuals from ARIMA models.

First, the linear model has to be estimated; after that, the residuals of the linear model contain the only nonlinear relationship. The residual at time *t* from the linear model is expressed as found with the following equation, where $\hat{L}_t$ is the forecast value for time t from the estimated relationship:

$$e_t = y_t - \hat{L}_t$$

With n input nodes, the ANN model for the residuals are modelled as where f is a nonlinear function determined by the neural network and ε_t is the random error:

$$\hat{N} = f\left(e_{t-1}, e_{t-2}, e_{t-3}, \ldots\ldots, e_{t-n}\right) + \varepsilon_t$$

Therefore, the correct model identification is critical, and the combined forecast will be:

$$\hat{y}_t = \hat{L}_t + \hat{N}_t$$

4 Time series forecasting models

In this section, the overall process of the applied methodology is presented. Moreover, the data and pre-processing process used to perform forecasting are described, respectively.

The overall process used while applying SARIMAX is presented as a pseudo code in Table A1. After the normality and stationarity tests, logarithmic transformations are applied in order to convert the percentage change. The non-stationarity and seasonality affect found above is handled with the SARIMAX, so any differencing process is not needed. The dataset is split into two as the train and test dataset. First, the SARIMA model with all variables (SARIMAX model) is estimated. In order to obtain the optimal hyper-parameters for the SARIMAX 'auto_arima' from pyramid (Pmdarima, 2019) is used, which is an API under an MIT license based on selecting the model that minimises the Akaike information criterion (AIC). The function selects the differencing terms with a test of stationarity (such as an augmented Dickey-Fuller test), and determines seasonality (such as the Canova-Hansen test) for seasonal models (Pmdarima Documentation, 2019).

While applying ANN-based models, the pseudo code presented in Table A2 is used. In order to execute MLP and LSTM, the Keras deep learning package is used (Chollet, 2015). The NARX model is formed with MATLAB's neural network toolbox (MATLAB, R2018b). There are some discussion about the necessity of preprocessing while using an artificial neural network modelling in a time series [Zhang and Qi, (2005), p.501]. Nelson et al. (1999, p.359) proposed that deseasonalised data could produce better forecasts than those which were not deseasonalised. Zhang and Qi (2005, p.501) examined SARIMA models and neural networks, and found that seasonal or trend variations are not effectively detected by the neural networks with the unpreprocessed data. Due to the ongoing debate, examinations in this study are preferred to be based on both the raw data and preprocessed data. After normality and stationarity tests, the STL and Fourier decomposition of the dependent variable is executed. The first differencing is then applied to the variables. In order to normalise the data, a min-max scaler is applied. Lastly, forecasts are collected with the use of MLP and LSTM models.

In light of the above-mentioned theoretical model, a hybrid model for the specified sample is developed in this study. After modelling the SARIMAX for the linear relationship, the MLP is modelled in order to reveal any nonlinear relationship. According to Moeeni and Bonakdari (2017), this nonlinear relationship is highly dependent on the number of neurons. As a result, there is no accurate method to find the most suitable input combination. Because of the limited data, lags have not been considered, and input trials are limited to four lags. For this study, the following combinations are identified, where Q is the:

SARIMAX-ANN-1: $Qt-1, Qt-4$

SARIMAX-ANN-2: $Qt-1, Qt-2$

SARIMAX-ANN-3: $Qt-1, Qt-2, Qt-3$

SARIMAX-ANN-4: $Qt-1, Qt-2, Qt-3, Qt-4$

SARIMAX-ANN-5: $Qt-1, Qt-3$

SARIMAX-ANN-6: $Qt-3, Qt-4$

SARIMAX-ANN-7: $Qt-2, Qt-4$

4.1 Dataset

Turkey has an important role in providing a link among the European Union, the Middle East, the Caucasus, the Mediterranean, the Aegean and the Black Seas [Mueller, (2007), p.6]. Therefore, Turkey has a significant potential to be a logistics hub in the region, and needs substantial investments in order to benefit from this geographical opportunity.

The domestic shipping market of Turkey was 29 million tons at 2003. Since then, there has been a significant increase of approximately 107%, with the total domestic shipping market raised up to 60 million tons by 2018, in which the share of the total handling volume of Turkey was 13% (MTI, 2019). This amount is planned to increase up to 18% until 2023 [SBB, (2019), p.127]. When the domestic shipping market is analysed by cargo volume, liquid cargoes are the most handled cargo type at 21 million tons. Moreover, 15 million tons of dry bulk cargo and 12 million tons of general cargo are handled. Of the total domestic shipping volume, 10 million tons are containerised cargo and 5 million tons are ro-ro (MTI, 2019). Although the opportunities of the market as usage of economies of scale, risk reduction, fuel prices excluded from the special consumption tax and high road freight rates, the market has some threats that weaken the market demand. A decrease in the number of Turkish shipowners, an inadequate demand for domestic shipping, too many intermediaries and operations are becoming considerable threats to the market on the horizon. With the new transportation networks and sustainable trends in the region, the domestic shipping of Turkey needs an elaborate evaluation.

Following the relative literature research that consists of forecasting studies realised in the maritime shipping literature, the variables in Table 1 are determined as possible factors to forecast domestic shipping volume loaded at the ports of Turkey.

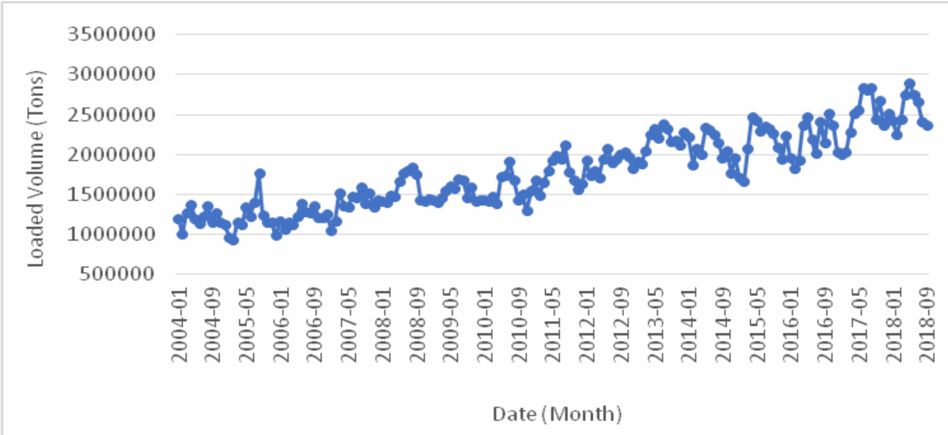
Table 1 Definitions and sources of variables

Variables	*Definitions*	*Data period (MM.YYYY)*	*Sources*
Loaded*	Domestic shipping volume loaded at the s of Turkey (tons)	01.2004–09.2018	MTI**
Exports (demand variable)	Foreign trade by months, 2003–2018 (exports values (thousand US$/FOB)	01.2004–09.2018	TUIK***
Imports (demand variable)	Foreign trade by months, 2003–2018 (imports values (thousand US$/FOB)	01.2004–09.2018	TUIK***
IPI (demand variable)	Industrial production index (2015 = 100), seasonally and calendar adjusted indices	01.2004–09.2018	TUIK***
Crude oil (supply variable)	Crude oil Brent FOB UK ports	01.2004–09.2018	UNCTADSTAT
BDI (supply variable)	Baltic exchange: Baltic dry index	01.2004–01.2009	Investing 2019
BDI (supply variable)	Baltic exchange: Baltic dry index	01.2009–09.2018	Thomson Reuters Eikon

Notes: *Dependent variable.
**Ministry of Transport and Infrastructure of Turkey.
***Turkish Statistical Institute.

The collected data is publicly available, and compromises 177 months. The sample is split into two as the train and test for SARIMAX model [159 train data and 18 test data]. For ANN-based models, the sample is split into three as train, validation and test data [141 train data, 18 validation data and 18 test data]. The domestic shipping volume loaded at the ports of Turkey is visualised in Figure 4. Data have a strong seasonality, as the increase in the summer months and the decrease in the winter months can be observed. A steady increase in the total volume can also be remarked upon, which might be related to the cement loads in summer being higher in general. Cement is one of the most loaded commodities at the domestic ports in Turkey. Diesel oil load is the reason for the spike in August 2005, which is contrary to the experienced increase in diesel fuel prices in July 2005. Seasonality and volatility become more visible towards the end of the sample, especially after 2014. This may be explained by an accelerated decline in the economic development of Turkey after 2013, which may be a result of increased volatility in the demand for the shipping of these commodities. After the sample of the collected data, data characteristics such as stationarity, normality and seasonality were investigated.

Figure 4 Domestic shipping cargoes loaded at the ports of Turkey (see online version for colours)



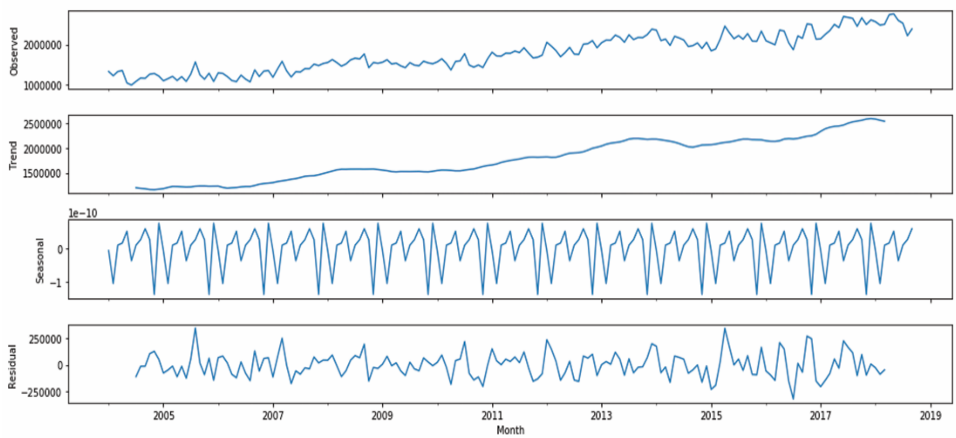
4.2 Dataset

Seasonality is a repeated characteristic of the time series that occurs at a fixed time interval [Xie et al., (2017), p.163]. Therefore, seasonal dummy variables can be used to detect seasonality in the time series [Hylleberg et al., (1993), p.321]. Deterministic seasonal dummies are used to test the seasonality with EViews software, with the results of the analysis indicating an obvious seasonality for the used sample data. After the detection of seasonality, STL decomposition and Fourier decomposition methods are used for the preprocessing of the data. STL decomposition is presented by Cleveland et al. (1990, p.3) in order to decompose the time series into trend, seasonality and reminder components. STL decomposition provides an estimate of the trend and extracts a changing seasonal component with iterative Loess smoothing [Dokumentov and Hyndman, (2015), p.3]. The STL seasonal decomposition of domestic loaded shipping cargoes split the series into seasonal, trend, and residual components as presented in Figure 5. It is observed that the time series of domestic cargo load has an upward trend, showing a gradually increasing trend and certainly has a periodicity. The remaining signal illustrates the residual. The STL decomposition procedure is implemented with the 'STL decompose' package of Python that is under an MIT license (STL Decompose, 2019).

Fourier terms decomposition as a trigonometric representation of seasonal components based on the Fourier series [De Livera et al., (2011), p.1516]. In this series, the seasonal pattern follows a simple sine and cosine wave [Birn and Birn, (2002), p.500]. Fourier terms decomposition as being implemented with a forecast package of RStudio Version 1.1.463 (Hyndman et al., 2018).

The descriptive statistics of the variables are given in Table A3. All variables are platykurtic, which have a flat top near the mean, and shorter, thinner tails that refer to a low probability for very extreme values. Loaded, BDI and crude oil variables have a positive skewness that show that occurring increases are more probable than decreases. Export and import variables have a negative skewness that refer to occurring declines, which are more than probable than occurring increases at the data.

Figure 5 STL decomposition of the domestic shipping cargoes loaded at the ports of Turkey (see online version for colours)



A statistical association between the loaded endogenous variable and used exogenous variables are searched for with a cross correlations matrix, which is shown in Table A4. As a result, the highest correlation existed between loaded and export and import variables. In addition to that, the BDI variables are highly correlated with the loaded amount. Export and import are the strongly correlated endogenous variables. The loaded data have a severe auto-correlation, and presented that a loaded series is serially correlated for all 12 lags. A prominent autocorrelation is observed, especially in lag 1 and 11.

In this study, stationarity is tested with a Dickey-Fuller and Philips Perron unit root test. The results are presented in Table 2, and reveal that all variables apart from the import variable have stationarity at their first differences.

Table 2 Unit root test results of the variables

Variable	*Level (ADF)*	*Level (P-P)*	*First difference (ADF)*	*First difference (P-P)*
Loaded	0.117444 (11)	–2.0472 (14)	–8.0811 (10)***	–22.2403 (14)***
BDI	–1.935572 (5)	–2.3601 (14)	–6.7586 (4)***	–9.5062 (14)***
Crude oil	–2.672187 (1)	–2.2464 (14)	–8.5000 (0)***	–7.9731 (14)***
IPI	–0.247671 (1)	–0.4011 (14)	–17.6818 (0)***	–17.4042 (14)***
Exports	–1.68583 (13)	–2.684 (14)*	–3.7866 (12)***	–29.0695 (14)***
Imports	–2.56416 (13)*	–2.998 (14)**	–3.5277 (14)***	–20.7374 (14)***

Notes: *, ** and *** represent statistical significance at the 10%, 5%, and 1% level, respectively. Trend and intercept are added according to statistical significance in the model.

Various time series methods rely on the assumption that data were sampled from a normal distribution [Machiwal and Jha, (2012), p.32]. Normality tests resulted in the sample data being normally distributed (>99%).

5 Results

In this study, the predictive capabilities of the conventional time series forecasting model SARIMA, the machine learning models MLP, LSTM and NARX, as well the hybrid model SARIMAX-MLP, are compared. To evaluate the accuracy of the models, the model performance was tested by calculating the MAPE, RMSE, MSE, MAD, SEP and ADP of the testing dataset. RMSE, MAD, MSE and MAPE are the popular measures for forecasting accuracy [Shim, (2000), p.133]. RMSE depicts the error distribution [Chai and Draxler, (2014), p.1248], whereas MAPE provides a framework to judge models by percentage [Klimberg et al., (2010), p.140]. MAPE is evaluated based on the following scale: a MAPE ≤ 10% means a high prediction accuracy, a 10% ≤ MAPE ≤ 20% means good prediction, a 20% ≤ MAPE ≤ 50% means a reasonable prediction, while a MAPE ≥ 50% means inaccurate forecasting (Lewis, 1982). MSE and MAD measures the amount of error, while MSE measures the amount of dispersion of the errors [Klimberg et al., (2010), p.139]. SEP and ADP are referred to by Zaji et al. (2018, p.721) as useful when comparing different variables. SEP and ADP are scale-dependent and non-dimensional.

Firstly, the SARIMA model with all variables (SARIMAX model) is estimated. In order to find out the best model, a variable selection is applied with a random forest, eliminating with the highest probability and stepwise regression with forward and backward elimination. However, these methods did not provide the best accuracy. Therefore, the models listed in Table 3 are attempted to be based on the accuracy methods of RMSE and MAPE. The model generated from the dataset is SARIMA (2, 1, 1) × $(0, 1, 1)_{12}$. In terms of RMSE, the best model is composed of IPI and the export variables, while the MAPE indicates that the best model is composed of crude oil, IPI and export variables. When two of the accuracy methods are compared, RMSE points out more relative forecasting models.

Table 3 Comparison of the SARIMAX forecast according to the accuracy methods

Model SARIMAX (2, 1, 1) × $(0, 1, 1)_{12}$		
Variables	*RMSE*	*MAPE*
BDI, crude oil, IPI, export, import	233,767	7.36
Crude oil, IPI, export, import	225,966	7.07
Crude oil, export, import	227,854	7.37
Crude oil, IPI, export	199,101	6.14
Export	196,923	6.5
Crude oil, IPI	206,216	6.15
Crude oil, export	197,708	6.19
Crude oil	208,955	6.24
IPI, export	*188,939*	6.14
BDI, crude oil, IPI, export	199,138	*6.06*
BDI, IPI, export	189,956	6.1
BDI, IPI	200,085	6.59
BDI, export	193,632	6.32

A logarithmic transformation of the series is used to change the units of the data to percentages in order to provide an equality of the characteristics to the data. The non-stationarity and seasonality affect found above is handled with the 'auto_arima' of the Phyton library, so any differencing process is not needed. The estimated significant parameters are illustrated in Table 4. The equation used is presented as follows:

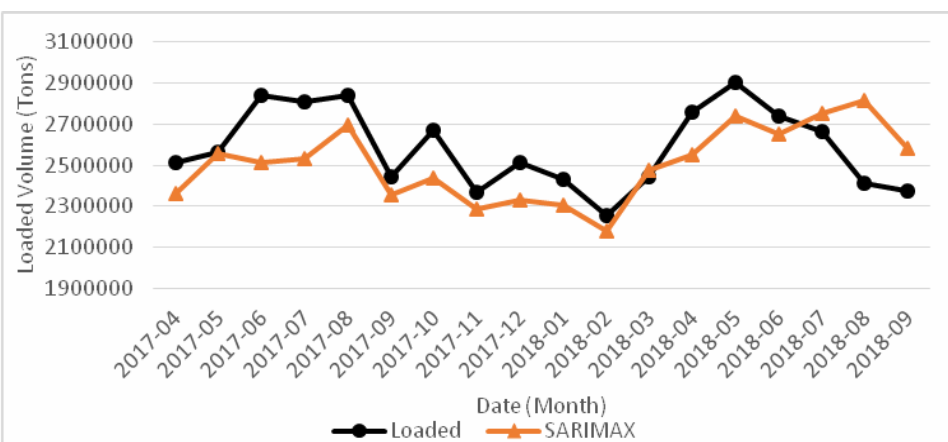
$$Y_t = 0.0003 + 0.1273\, X_1 + 0.1920\, X_2 + 0.2597\, Y_{t-1} \\ -0.1886\, Y_{t-2} + \varepsilon_t - 0.7610\varepsilon_t - 0.9028\varepsilon_{t-12}$$

The sample forecast of the domestic shipping cargo load obtained from the SARIMAX model is presented in Figure 6. As shown in Figure 6, out of sample forecasts of domestic shipping loads can be produced with the SARIMAX model, with the model being capable of producing effective forecasts.

Table 4 Model estimation of the SARIMAX model

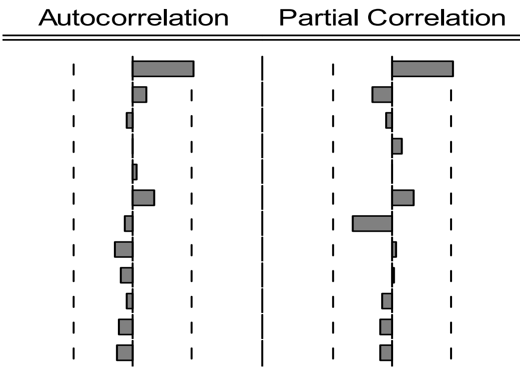
Variable	*Coefficient (standard errors)*	*Z-value*	*P > |z|*
C	0.0003 (0.001)	0.554	0.579
IPI (X_1)	0.1273 (0.231)	0.550	0.582
Export (X_2)	0.1920 (0.082)	2.345	0.019
ar.L1	0.2597 (0.116)	2.244	0.025
ar.L2	–0.1886 (0.100)	–1.878	0.060
ma.L1	–0.7610 (0.101)	–7.531	0.000
ma.S.L12	–0.9028 (0.171)	–5.267	0.000
sigma2	0.0061 (0.001)	5.036	0.000

Figure 6 Results of SARIMAX forecasting for domestic shipping cargoes loaded at the ports of Turkey (see online version for colours)



After the model is estimated, in order to check the adequacy of the model, the ACF and PACF of the residuals were examined. The Ljung-Box Q test is adopted to diagnose whether the estimated residuals are white noise series or not. As seen in Figure 7, the result of the test indicates that the residuals of the estimated model are white noise, and do not have a serial correlation. Thus, it is deduced that the developed SARIMAX model was established correctly.

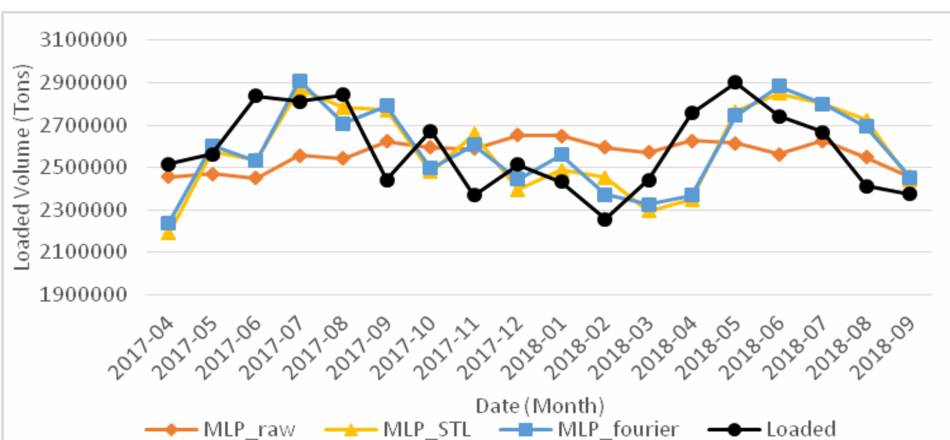
Figure 7 ACF and PACF of residuals



	AC	PAC	Q-Stat	Prob
1	0.489	0.489	5.0591	0.024
2	0.123	-0.153	5.3976	0.067
3	-0.037	-0.042	5.4302	0.143
4	0.007	0.087	5.4316	0.246
5	0.045	0.007	5.4880	0.359
6	0.173	0.179	6.3873	0.381
7	-0.058	-0.309	6.4955	0.483
8	-0.137	0.046	7.1754	0.518
9	-0.096	0.024	7.5471	0.580
10	-0.042	-0.070	7.6249	0.665
11	-0.100	-0.087	8.1393	0.701
12	-0.115	-0.096	8.9267	0.709

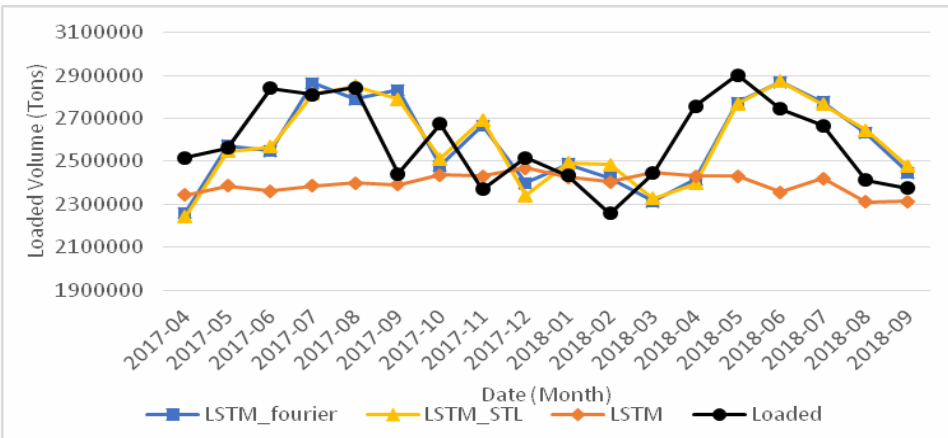
The results of MLP models are presented in Figure 8. MLP architecture is utilised as five input variables, three hidden layers with five, five and three neurons terminating at an output layer with tanh activation. Hidden layers utilised tanh activation functions, and the network epoch was 120. Fourier terms the decomposed MLP model and MLP with the raw data model forecast similarly in this scenario. However, it is observed that a MLP with the Fourier terms of decomposition can capture the fluctuations, while MLP with a raw value model stayed around a limited line. It can be observed that the Fourier terms of decomposition achieved the highest accuracy between the MLP models. The STL decomposing the MLP model results were quite spurious, with a larger MAPE and RMSE result.

Figure 8 Results of MLP forecasting models for domestic shipping cargoes loaded at the ports of Turkey (see online version for colours)



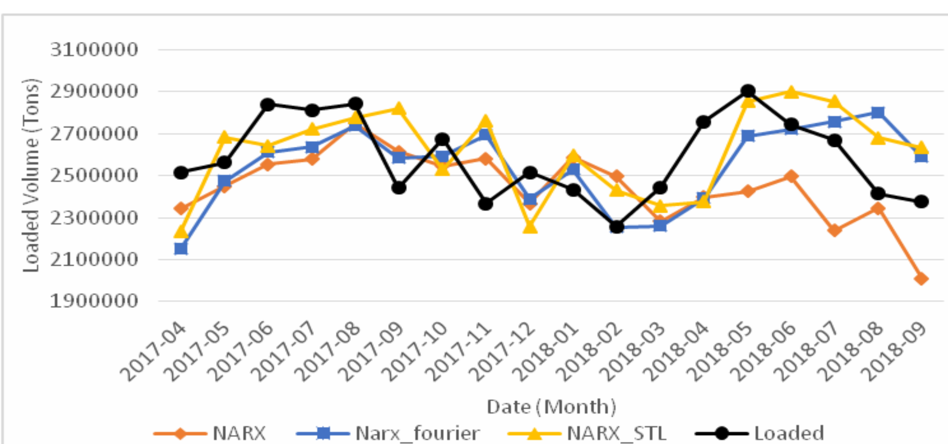
The results of the LSTM models are illustrated in Figure 9. LSTM architecture is utilised as five input variables, one hidden layer with 240 neurons terminating at an output layer with a tanh activation. Hidden layers utilised tanh activation functions, and the network epoch was 50. LSTM with seasonal adjustment achieved the performance closest to the SARIMAX model. Fourier terms decomposing the LSTM model performed better than the MLP models, the NARX model and other LSTM models.

Figure 9 Results of LSTM forecasting models for domestic shipping cargoes loaded at the ports of Turkey (see online version for colours)



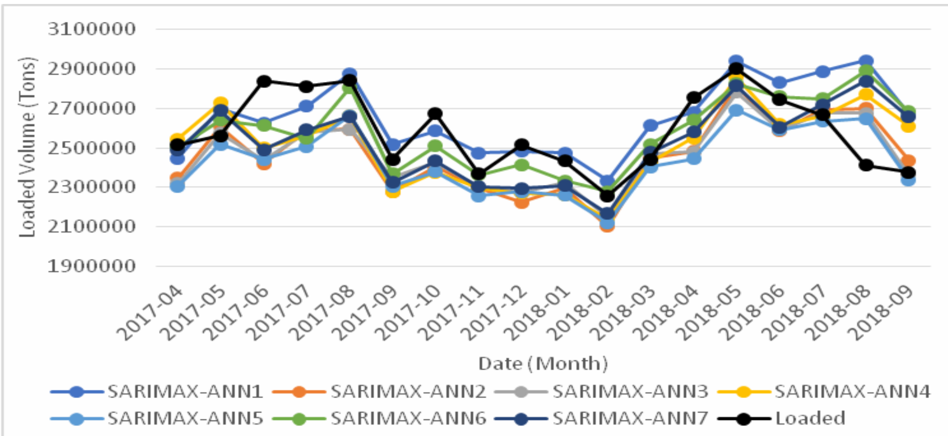
The results of the NARX models are presented in Figure 10. In this study, preparets (a MATLAB function that prepares data for network simulation or training) fixed the data in the correct format for our NARX, and there were ten hidden neurons. The training function is the Levenberg-Marquardt backpropagation. NARX models performed worse than other benchmarked forecasting models. The Fourier terms of decomposition consistently achieved the highest accuracy between the NARX models.

Figure 10 Results of the NARX forecasting model for domestic shipping cargoes loaded at the ports of Turkey (see online version for colours)



Forecasted combinations of the hybrid SARIMAX-ANN models are presented in Figure 11. The hybrid models, excluding the SARIMAX-ANN5 model, achieved more precise forecasts than the ANN-based models. The SARIMAX-ANN6 model (input nodes Qt – 3, Qt – 4 and residual value) had the lowest out of sample error. In order to check the adequacy of the model, the ACF and PACF of the residuals were examined. A Ljung-Box Q test was adopted, with the result of the test showing that the residuals of the estimated model are white noise, and do not have serial correlation.

Figure 11 Comparison of the hybrid SARIMAX-ANN model forecasts (see online version for colours)



This study presents the comparison of the forecasting results obtained from the individual models and hybrid models. Figure 12 shows the best models according to the out-of-sample errors from each method. The hybrid SARIMAX-ANN model outperformed all the other benchmarked methods. SARIMAX-ANN finds a 4.81 MAPE value, so it is concluded that the model forecasts a high prediction accuracy.

Figure 12 Benchmark of the forecasting models of domestic shipping loaded at the ports of Turkey (see online version for colours)

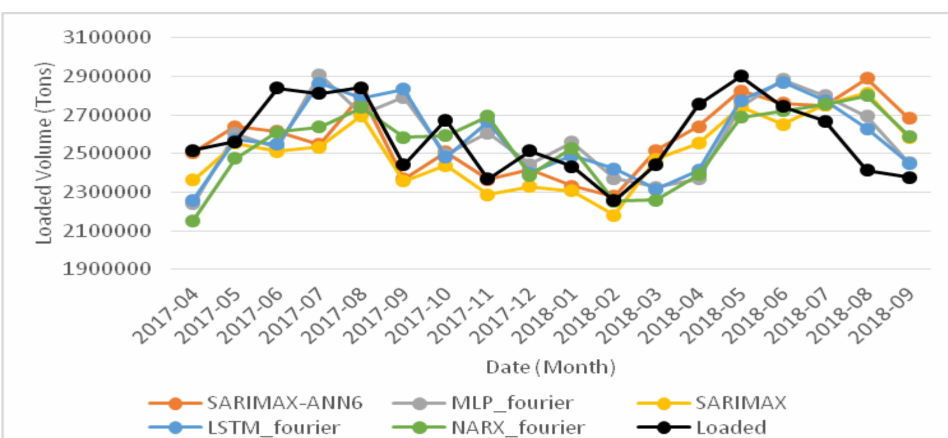


Table 5 Accuracy comparisons of the forecasting models of domestic shipping loaded at the ports of Turkey

	RMSE	*MAPE*	*MSE*	*MAD*	*SEP*	*ADP*
MLP						
Raw values with normalisation	205,447	6.95	42,208,636,950.72	180,290.50	7.94	6.98
First difference and STL decomposition	214,234	7.04	45,896,538,091.09	180,272.07	8.28	6.98
First difference and Fourier decomposition	204,782	6.94	41,935,720,606.78	179,320.29	7.91	6.95
LSTM						
Raw values with normalisation	269,594	7.89	72,680,942,764.05	213,796.833	10.42	8.28
First difference and STL decomposition	202,722	6.64	41,096,210,474.50	168,821.546	7.83	6.54
First difference and Fourier decomposition	198,428	6.52	39,348,978,005.47	166,926.525	6.47	7.66
NARX						
Raw values with normalisation	252,368	8.66	63,689,812,705.88	225,471.032	9.75	8.74
First difference and STL decomposition	228,631	8.02	52,272,342,736.79	202,956.580	8.83	7.86
First difference and Fourier decomposition	211,749	6.93	44,838,000,147.24	178,022.415	8.18	6.90
SARIMAX $(2, 1, 1) \times (0, 1, 1)_{12}$						
Logarithmic transformation	188,939	6.14	35,697,995,690.72	159,995.611	7.30	3.17
SARIMAX-ANN						
Logarithmic transformation	172,431	4.81	29,732,500,548.62	123,682.022	6.66	4.79

Table 5 also shows that the SARIMAX-ANN model outperformed all the other methods with the smallest forecasting errors. In this study, the performance of the NARX method was considered in inaccurate forecasting, with the largest RMSE and MAPE. The SARIMAX method forecast the domestic shipping load well with that horizon. Therefore, preprocessed data with a Fourier series also produce reasonable forecasts. The SARIMAX and SARIMAX-ANN hybrid are two models with a seasonal adjustment that performed best in this scenario due to the fact that the sample has a strong seasonal pattern and a clear upper trend. Moreover, the SARIMAX-ANN model can capture both the linear and nonlinear nature of the data, and uses the advantages of both models. SARIMAX accounts for the seasonality nature of the data, and reflects the effects of lags. However, artificial neural network-based models fail to model the effects of lag and seasonality properly, despite the preprocessed data.

Rating the results of Table 5 shows that SARIMAX-ANN, SARIMAX and LSTM_Fourier are the superior models. According to the RMSE, MAPE, MSE and the MAD SARIMAX-ANN method is the best performing model for forecasting the domestic shipping market of Turkey. The SARIMAX model is also a superior model, which has the closest accuracy measures to the SARIMAX-ANN. LSTM_Fourier and LSTM_stl are the following models. In terms of SEP, the best models are LSTM_Fourier, SARIMAX-ANN and SARIMAX, relatively speaking. SARIMAX, SARIMAX-ANN and LSTM_STL are the best performing models according to the ADP. LSTM_stl, MLP_Fourier, MLP_raw and MLP_stl are very close in terms of their performances. LSTM_raw and NARX_raw resulted in poor performances. This result confirms Bonakdari et al. (2019), who state that individual soft computing models did not present a better accuracy than the linear models that soft computing models need to be hybridised.

The main reason for the accuracy improvement of the hybrid method is the ability of combining both the linear and nonlinear parts of the time series after decomposing seasonal components with SARIMAX. Hence, finding accurate lag combinations that have a significant impact on the time series is another reason for this improvement. The superior performance of SARIMAX over other single methods is due to being adequate in determining the seasonal terms of time series. Moreover, LSTM Fourier has a good accuracy among the other developed soft computing models. This result can be linked with the cyclical seasonal effects of observations and memory blocks in the LSTM networks.

To discuss the developed models in detail, t-tests are conducted to examine the significances of the model performances. In this paper, t-test results are based on the comparison of the MAPE of the residuals. The p-value for each t-test is presented in Table 6. The tests results have a significance level of 95%; therefore, p-values < 0.05 indicate models that do vary significantly. Except for the SARIMAX and Hybrid model, statistically significant differences cannot be concluded between the developed models. Furthermore, a Diebold-Mariano test is adopted in order to evaluate the statistical significance of the performances of developed models in depth. A DM test aims to test the null hypothesis of equal accuracy of the two competing models. In this paper, the null hypothesis of the DM test cannot be rejected for developed models. It is revealed that the developed models' forecasting accuracy are similar. The DM test and paired t-test are implemented via the R software package.

Table 6 T-test and DM test results

	MLP	*SARIMAX*	*LSTM_Fourier*	*NARX_Fourier*	*Hybrid*
MLP		0.51742 (0.6115)	0.63187 (0.5359)	–0.30737 (0.7623)	–0.7744 (0.4493)
SARIMAX	*0.7321* *(0.474)*		–26,541 (0.7939)	–0.80961 (0.4293)	–0.8894 (0.3862)
LSTM_Fourier	*1.0466* *(0.3099)*	*0.29532* *(0.7713)*		–0.48472 (0.6341)	–0.5579 (0.5841)
NARX_Fourier	*0.016785* *(0.9868)*	*0.73693* *(0.4712)*	*0.38401* *(0.7057)*		–1.1547 (0.2642)
Hybrid	*–1.4886* *(0.1549)*	*–2.032* *(0.05809)*	*–1.0638* *(0.3023)*	*–1.6672* *(0.1138)*	

Note: *italic fonts are the results of the t-test.

6 Conclusions and discussion

Almost all decision-makers in the shipping industry, i.e., shipowners, investment bankers, governments, shipbuilders, etc., need to predict cargo volumes in order to make accurate strategic planning. In this study, the predictive capabilities of the conventional time series forecasting model SARIMAX, and the machine learning models MLP, LSTM and NARX and the hybrid model SARIMAX-ANN, are compared.

When we consider the pre-processing of the data, the literature proposes that the pre-processed data result in the best forecasts. The results of the present paper also supported the studies by Klein (1996) and Shu et al. (2013) in this regard. The adjusted data provide a substantial improvement over those obtained with unadjusted data, and the accuracy of the conventional models is found to be significantly boosted with the Fourier modification.

In general, it can be theoretically demonstrated that hybrid models can obtain as good and even better results than one of the individual models [Khashei and Hajirahimi, (2018), p.2639]. In shipping markets, a similar remark is made by Xie et al. (2013), with the conclusion that among the benchmarked models the hybrid models are the best performing models, while among the single models SARIMA is the best performing model. Xiao et al. (2016) also showed that SARIMA outperformed benchmarked single models, such as a generalised regression neural network, a wavelet neural network and a feedforward neural network. Chan et al. (2019) also found that machine learning approaches may not necessarily present a better performance than the traditional regression-based models. In our study, it is also found that SARIMAX-ANN and SARIMAX can produce more accurate results than the compared models. However, it should be noted that the forecasting performances of these models are found to be statistically insignificant. The predictive abilities of the developed models are the same, and they did not show any better performance than one another statistically. To sum up, it is recommended that bootstrap tests should be applied in order to reveal the accuracy of the developed models.

The primary motivation behind this study is to support decision-makers of the industry in terms of the analysis and forecasting of the domestic shipping at the ports of Turkey. The methodological benchmarking of the candidate forecasting techniques, as well as the establishment of a clear-cut forecasting process, is claimed as the technical contribution of this study, with the findings providing a consistent result with the literature regarding the performances of the SARIMAX and hybrid models. The results of the current study demonstrate that the SARIMAX-ANN models can be considered as an appropriate alternative, at least for shipping time series forecasting.

This research has several limitations, and some of them can be considered for future research. First, Turkey's domestic shipping data with only 177 observations is used in this study; however, data from a wider and varied geographic sample should be utilised in future research. STL decomposition and the Fourier decomposition applied in this study, and decomposing with other methods and comparing their performances, may be attempted in the future. Five exogenous variables are included in the developed models. Therefore, different input combinations are also possible, and can be developed with evolutionary algorithms or selected with other classification algorithms. Varying input combinations of the hybrid model could also have been selected with computing algorithms. Furthermore, developing soft computing methods with an optimisation grid and different learning algorithms can contribute to the forecasting ability of shipping markets. Future research can include a benchmark of the models that is hybridised by only soft computing methods, and also hybridised by both the time series and soft computing methods. Last but not least, evolutionary algorithms, and GMDH networks with multi-variables can be considered as well.

References

Akal, M. (2004) 'Forecasting Turkey's tourism revenues by ARMAX model', *Tourism Management*, Vol. 25, No. 5, pp.565–580.

Babcock, M.W. and Lu, X. (2002) 'Forecasting inland waterway grain traffic', *Transportation Research Part E: Logistics and Transportation Review*, Vol. 38, No. 1, pp.65–74.

Beenstock, M. (1985) 'A theory of ship prices', *Maritime Policy and Management*, Vol. 12, No. 3, pp.215–225.

Beenstock, M. and Vergottis, A. (1993) *Econometric Modelling of World Shipping*, Springer Science & Business Media, London.

Birn, R.J. and Birn, R. (2002) *The International Handbook of Market Research Techniques*, Kogan Page Publishers, London.

Bonakdari, H., Moeeni, H., Ebtehaj, I., Zeynoddin, M., Mahoammadian, A. and Gharabaghi, B. (2019) 'New insights into soil temperature time series modeling: linear or nonlinear?', *Theoretical and Applied Climatology*, Vol. 135, Nos. 3–4, pp.1157–1177.

Çelikkaya, A. (2012) 'Türkiye'de deniz taşımacılığına sağlanan vergi teşvikleri üzerine bir inceleme', *Maliye Dergisi*, January–June, Vol. 162, pp.73–102.

Chai, T. and Draxler, R.R. (2014) 'Root mean square error (rmse) or mean absolute error (mae)? – arguments against avoiding RMSE in the literature', *Geoscientific Model Development*, Vol. 7, No. 3, pp.1247–1250.

Chan, H.K., Xu, S. and Qi, X. (2019) 'A comparison of time series methods for forecasting container throughput', *International Journal of Logistics Research and Applications*, Vol. 22, No. 3, pp.294–303.

Charemza, W. and Gronicki, M. (1981) 'An econometric model of world shipping and shipbuilding', *Maritime Policy and Management*, Vol. 8, No. 1, pp.21–30.

Chase, C.W. (2013) *Demand-Driven Forecasting: a Structured Approach to Forecasting*, John Wiley & Sons, New Jersey.

Chen, S.H. and Chen, J.N. (2010) 'Forecasting container throughputs at ports using genetic programming', *Expert Systems with Applications*, Vol. 37, No. 3, pp.2054–2058.

Chollet, F. (2015) *Keras* [online] https://keras.io/ (accessed 15 May 2018).

Cleveland, R.B., Cleveland, W.S., McRae, J.E. and Terpenning, I. (1990) 'STL: a seasonal-trend decomposition', *Journal of Official Statistics*, Vol. 6, No. 1, pp.3–73.

De Gooijer, J.G. and Klein, A. (1989) 'Forecasting the Antwerp maritime steel traffic flow: a case study', *Journal of Forecasting*, Vol. 8, No. 4, pp.381–398.

De Livera, A.M., Hyndman, R.J. and Snyder, R.D. (2011) 'Forecasting time series with complex seasonal patterns using exponential smoothing', *Journal of the American Statistical Association*, Vol. 106, No. 496, pp.1513–1527.

Dokumentov, A. and Hyndman, R.J. (2015) *STR: a Seasonal-Trend Decomposition Procedure based on Regression (No. 13/15)*, Working Paper, Department of Econometrics and Business Statistics, Monash University [online] https://robjhyndman.com/papers/wp13-15.pdf (accessed 8 April 2018).

Du, P., Wang, J., Yang, W. and Niu, T. (2019) 'Container throughput forecasting using a novel hybrid learning method with error correction strategy', *Knowledge-Based Systems*, Vol. 182, No. 104853, pp.1–12.

Eslami, P., Jung, K., Lee, D. and Tjolleng, A. (2016) 'Predicting tanker freight rates using parsimonious variables and a hybrid artificial neural network with an adaptive genetic algorithm', *Maritime Economics & Logistics*, Vol. 19, No. 3, pp.538–550.

Farhan, J. and Ong, G.P. (2016) 'Forecasting seasonal container throughput at international ports using SARIMA models', *Maritime Economics & Logistics*, Vol. 20, No. 1, pp.131–148.

Feng, H., Kong, F. and Xiao, Y. (2011) 'Vessel traffic flow forecasting model study based on support vector machine', in *International Conference on Electronic Commerce, Web Application, and Communication*, Springer, Berlin, Heidelberg, April.

Geng, J., Li, M.W., Dong, Z.H. and Liao, Y.S. (2015) 'Port throughput forecasting by MARS-RSVR with chaotic simulated annealing particle swarm optimization algorithm', *Neurocomputing*, January, Vol. 147, pp.239–250.

Gensler, A., Henze, J., Sick, B. and Raabe, N. (2016) 'Deep Learning for solar power forecasting – an approach using AutoEncoder and LSTM neural networks', *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Budapest, October.

Goulielmos, A.M. and Kaselimi, E. (2011) 'A non-linear forecasting of container traffic: the case-study of the Port of Piraeus, 1973–2008', *International Journal of Shipping and Transport Logistics*, Vol. 3, No. 1, pp.72–99.

Graves, A. (2012) *Supervised Sequence Labelling with Recurrent Neural Networks*, pp.15–35, Springer, Berlin, Heidelberg.

Grifoll, M. (2019) 'A statistical forecasting model applied to container throughput in a multiport gateway system: the Barcelona-Tarragona-Valencia case', *International Journal of Shipping and Transport Logistics*, Vol. 11, No. 4, pp.316–333.

Gulli, A. and Pal, S. (2017) *Deep Learning with Keras*, Packt Publishing Ltd., Birmingham.

Huang, A., Lai, K., Li, Y. and Wang, S. (2015) 'Forecasting container throughput of Qingdao port with a hybrid model', *Journal of Systems Science and Complexity*, Vol. 28, No. 1, pp.105–121.

Hylleberg, S., Jørgensen, C. and Sørensen, N.K. (1993) 'Seasonality in macroeconomic time series', *Empirical Economics*, Vol. 18, No. 2, pp.321–335.

Hyndman, R.J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M. and Yasmeen, F. (2018) *Forecast: Forecasting Functions for Time Series and Linear Models. R Package Version 8.9* [online] http://pkg.robjhyndman.com/forecast (accessed 10 December 2019).

Ihle, N. (2016) 'Short-term load forecasting methods for maritime container terminals', *Proceedings of the ICCBR 2016 Workshops*, Atlanta.

Investing (2019) *Baltic Dry Historical Data* [online] https://www.investing.com/indices/baltic-dry-historical-data (accessed 20 January 2018).

Jugović, A., Hess, S. and Poletan Jugović, T. (2011) 'Traffic demand forecasting for port services', *PROMET-Traffic & Transportation*, Vol. 23, No. 1, pp.59–69.

Khashei, M. and Hajirahimi, Z. (2018) 'A comparative study of series arima/mlp hybrid models for stock price forecasting', *Communications in Statistics-Simulation and Computation*, Vol. 48, No. 9, pp.2526–2640.

Klein, A. (1996) 'Forecasting the Antwerp maritime traffic flows using transformations and intervention models', *Journal of Forecasting*, Vol. 15, No. 5, pp.395–412.

Klimberg, R.K., Sillup, G.P., Boyle, K.J. and Tavva, V. (2010) 'Forecasting performance measures – what are their practical meaning?', in *Advances in Business and Management Forecasting*, pp.137–147, Emerald Group Publishing Limited, Bingley, UK.

Lam, W.H., Ng, P.L., Seabrooke, W. and Hui, E.C. (2004) 'Forecasts and reliability analysis of port cargo throughput in Hong Kong', *Journal of Urban Planning and Development*, Vol. 130, No. 3, pp.133–144.

Lee, S.Y., Lim, H. and Kim, H.J. (2017) 'Forecasting container port volume: implications for dredging', *Maritime Economics & Logistics*, Vol. 19, No. 2, pp.296–314.

Lewis, C.D. (1982) *Industrial and Business Forecasting Methods*, Butterworths Publishing, London.

Li, G., Kawan, B., Wang, H. and Zhang, H. (2017) 'Neural-network-based modelling and analysis for time series prediction of ship motion', *Ship Technology Research*, Vol. 64, No. 1, pp.30–39.

Li, M., Han, D. and Wang, W. (2015) 'Vessel traffic flow forecasting by RSVR with chaotic cloud simulated annealing genetic algorithm and KPCA', *Neurocomputing*, June, Vol. 157, pp.243–255.

Liu, S., Tian, L. and Huang, Y. (2014) 'A comparative study on prediction of throughput in coal ports among three models', *International Journal of Machine Learning and Cybernetics*, Vol. 5, No. 1, pp.125–133.

Lv, P., Zhuang, Y. and Yang, K. (2016) 'Prediction of ship traffic flow based on BP neural network and Markov model', *ICTTE 2016 in MATEC Web of Conferences*, Vol. 81, pp.1–5.

Machiwal, D. and Jha, M.K. (2012) *Hydrologic Time Series Analysis: Theory and Practice*, Springer Science & Business Media, New York.

Medsker, L. and Jain, L. (2001) *Recurrent Neural Networks: Design and Applications*, CRC Press, Boca Raton.

Mitrea, C.A., Lee, C.K.M. and Wu, Z. (2009) 'A comparison between neural networks and traditional forecasting methods: a case study', *International Journal of Engineering Business Management*, Vol. 1, No. 2, pp.67–72.

Mo, L., Xie, L., Jiang, X., Teng, G., Xu, L. and Xiao, J. (2018) 'GMDH-based hybrid model for container throughput forecasting: selective combination forecasting in nonlinear subseries', *Applied Soft Computing*, Vol. 62, pp.478–490.

Moeeni, H. and Bonakdari, H. (2017) 'Forecasting monthly inflow with extreme seasonal variation using the hybrid SARIMA-ANN model', *Stochastic Environmental Research and Risk Assessment*, Vol. 1, No. 8, pp.1997–2010.

Moeeni, H., Bonakdari, H. and Ebtehaj, I. (2017) 'Integrated SARIMA with neuro-fuzzy systems and neural networks for monthly inflow prediction', *Water Resources Management*, Vol. 31, No. 7, pp.2141–2156.

Mostafa, M.M. (2004) 'Forecasting the Suez Canal traffic: a neural network analysis', *Maritime Policy & Management*, Vol. 31, No. 2, pp.139–156.

MTI (2019) *Maritime Statistics* [online] https://atlantis.udhb.gov.tr/istatistik/istatistik_yuk.aspx (accessed 25 November 2019).

Mueller, R. (2007) *Technical Assistance to Transportation Infrastructure Needs Assessment for Turkey*, Final Report, Joint Venture, TINA, Turkey (accessed 6 January 2019).

Nelson, M., Hill, T., Remus, W. and O'Connor, M. (1999) 'Time series forecasting using neural networks: should the data be deseasonalized first?', *Journal of Forecasting*, Vol. 18, No. 5, pp.359–367.

Özen, C. (2007) *Avrupa Birliği'ne giriş sürecinde Avrupa denizcilik politikaları ve Türkiye'nin uyumu*, Unpublished Master's thesis, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Turkey.

Palit, A.K. and Popovic, D. (2006) *Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications*, Springer Science & Business Media, London.

Pang, G. and Gebka, B. (2017) 'Forecasting container throughput using aggregate or terminal-specific data? The case of Tanjung Priok Port, Indonesia', *International Journal of Production Research*, Vol. 55, No. 9, pp.2454–2469.

Pmdarima (2019) [online] https://github.com/tgsmith61591/pmdarima (accessed 8 April 2019).

Pmdarima Documentation (2019) [online] https://www.alkaline-ml.com/pmdarima/about.html?highlight=aic (accessed 8 April 2019).

Santos, A.A., Junkes, L.N. and Pires Jr., F.C. (2014) 'Forecasting period charter rates of VLCC tankers through neural networks: a comparison of alternative approaches', *Maritime Economics & Logistics*, Vol. 16, No. 1, pp.72–91.

SBB (2019) *11th Development Plan of Turkey* [online] http://www.sbb.gov.tr/wp-content/uploads/2019/07/OnbirinciKalkinmaPlani.pdf (accessed 25 November 2019).

Schulze, P.M. and Prinz, A. (2009) 'Forecasting container transshipment in Germany', *Applied Economics*, Vol. 41, No. 22, pp.2809–2815.

Shim, J.K. (2000) *Strategic Business Forecasting: the Complete Guide to Forecasting Real World Company Performance*, CRC Press, Boca Raton.

Shu, M.H., Huang, Y.F. and Nguyen, T.L. (2013) 'Forecasting models for the Cargo throughput at Hong Kong Port and Kaohsiung Port', *Recent Research in Applied Economics and Management*, WSEAS Press, Crete Island, Greece.

STL Decompose (2019) [online] https://pypi.org/project/stldecompose/ (accessed 9 April 2019).

Stopford, M. (2009) *Maritime Economics*, 3e, Routledge.

Tinbergen, J. (1959) 'Tonnage and freight', in Klaassen, L.H., Koyck, L.M. and Witteveen, H.J. (Eds.): *Jan Tingergen Selected Papers*, North-Holland Publishing Company, Amsterdam, Netherlands.

Vagropoulos, S.I., Chouliaras, G.I., Kardakos, E.G., Simoglou, C.K. and Bakirtzis, A.G. (2016) 'Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting', in *2016 IEEE International Energy Conference (ENERGYCON)*. April.

Veenstra, A.W. and Haralambides, H.E. (2001) 'Multivariate autoregressive models for forecasting seaborne trade flows', *Transportation Research Part E: Logistics and Transportation Review*, Vol. 37, No. 4, pp.311–319.

Winston, C. (1981) 'A multinomial probit prediction of the demand for domestic ocean container service', *Journal of Transport Economics and Policy*, September, Vol. 15, No. 3, ,pp.243–252.

Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J. (2016) *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann.

Xiao, Y., Wang, S., Liu, J.J., Xiao, J. and Hu, Y. (2016) 'Throughput estimation-based port development and management policies analysis', *Maritime Policy & Management*, Vol. 43, No. 1, pp.84–97.

Xie, G., Wang, S., Zhao, Y. and Lai, K.K. (2013) 'Hybrid approaches based on LSSVR model for container throughput forecasting: a comparative study', *Applied Soft Computing*, Vol. 13, No. 5, pp.2232–2241.

Xie, G., Zhang, N. and Wang, S. (2017) 'Data characteristic analysis and model selection for container throughput forecasting within a decomposition-ensemble methodology', *Transportation Research Part E: Logistics and Transportation Review*, December, Vol. 108, pp.160–178.

Yang, Z. and Mehmed, E.E. (2019) 'Artificial neural networks in freight rate forecasting', *Maritime Economics & Logistics*, Vol. 21, No. 3, pp.390–414.

Yolcu, U., Egrioglu, E. and Aladag, C.H. (2013) 'A new linear & nonlinear artificial neural network model for time series forecasting', *Decision Support Systems*, Vol. 54, No. 3, pp.1340–1347.

Zaji, A.H., Bonakdari, H. and Gharabaghi, B. (2018) 'Reservoir water level forecasting using group method of data handling', *Acta Geophysica*, Vol. 66, No. 4, pp.717–730.

Zhang, G., Patuwo, B.E. and Hu, M.Y. (1998) 'Forecasting with artificial neural networks: the state of the art', *International Journal of Forecasting*, Vol. 14, No. 1, pp.35–62.

Zhang, G.P. (2003) 'Time series forecasting using a hybrid ARIMA and neural network model', *Neurocomputing*, January, Vol. 50, pp.159–175.

Zhang, G.P. and Qi, M. (2005) 'Neural network forecasting for seasonal and trend time series', *European Journal of Operational Research*, Vol. 160, No. 2, pp.501–514.

Zhang, Q., Yan, L. and He, J. (2018) 'Time series analysis of correlativity between pulmonary tuberculosis and seasonal meteorological factors based on theory of human-environmental inter relation', *Journal of Traditional Chinese Medical Sciences*, Vol. 5, No. 2, pp.119–127.

Zheng, J., Xu, C., Zhang, Z. and Li, X. (2017) 'Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network', in *2017 51st Annual Conference on Information Sciences and Systems (CISS)*, March.

Appendix

Table A1 Pseudo code of SARIMAX model

Pseudo code 1	*SARIMAX model fitting algorithm*
	Start
1	Load dataset
2	Normality and stationarity tests
3	Log transformations
4	X = independent variables
5	Y = dependent variables
6	datasets = [train, test]
7	Variable selection
8	Execute SARIMAX with the auto_arima
9	model = auto_arima(trainy, trainx) with least-AIC in models
10	model.fit(trainy,trainx)
11	model.predict(exogenous=testx,n_periods=18)
12	Invert log transform
	End

Table A2 Pseudo code of ANN-based models

Pseudo code 2	*ANN-based models fitting algorithm*
	Start
1	Load dataset
2	Normality and stationarity tests
3	STL or Fourier decomposition
4	First differencing
5	MinMaxScaler(copy=True, feature_range=(-1,1))
6	sc.fit_transform(X) "X = independent variables"
7	sc.fit_transform(X) "Y = dependent variables"
8	datasets = [train, validation, test]
9	If Execute MLP
	model.add(Dense(10, input_shape=(n_cols,))
	model.add(Activation("tanh"))
	model.add(Dense(5))
	model.add(Activation("tanh"))
	model.add(Dense(3))
	model.add(Activation("tanh"))
	model.add(Dense(3))
	model.add(Activation("tanh"))
	model.add(Dense(1, activation="tanh"))
	model.compile(optimizer="RMSprop", loss="mse", metrics=["mape"])
	model.fit(trainy,trainx)
	model.predict(testx,n_periods=18)
9:	If Execute LSTM
	Reshape input
	model = Sequential()
	model.add(LSTM(5, input_shape=(X_train.shape[1], X_train.shape[2]), activation="tanh"))
	model.add(Dense(240, activation="tanh"))
	model.add(Dense(1, activation="tanh"))
	RMSprop= optimizers.RMSprop(lr=0.9, rho=0.9, epsilon=None, decay=0.0)
	model.compile(loss="mse", optimizer="RMSprop", metrics=["mape"])
	history = model.fit(X_train, Y_train, epochs=50, batch_size=2, validation_data=(X_test, Y_test), verbose=0, shuffle=False)
	model.predict(X_test)
	Invert reshape
10:	Invert scaling
11:	End

Table A3 Descriptive statistics of variables used for forecasting

Variable	*Mean*	*Standard deviation*	*Min*	*Max*	*Skewness*	*Kurtosis*
Loaded	1,787,193	472,663.1	929,856	2,902,909	0.288378	–0.83340
BDI	2,569.791	2,307.55	317	11,440	1.75013	2.86610
Crude oil	74.95927	26.7692	30.8	133.87	0.329643	–1.12446
Export	10,399,018	2,782,979	3,664,503	15,554,395	–0.505538	–0.8949
Import	16,030,611	4,374,194	6,139,442	23,245,301	–0.402714	–0.95412

Table A4 Correlation coefficients of the used variables

	Loaded	*BDI*	*Crude oil*	*Export*	*Import*
Loaded	1.0000	0.6307	0.9253	0.9882	0.9834
BDI		1.0000	0.7129	0.6508	0.6578
Crude			1.0000	0.9517	0.9654
Export				1.0000	0.9949
Import					1.0000