Application of Quantitative models, MNLR & ANN in Short term forecasting of Ship data

***P. Oliver jayaprakash[1] and K.Gunasekaran[2]**

[1] Ph.D student, Transportation Engineering Division, Civil Engineering Department,Anna University Chennai,Tamil Nadu,600 025,India.

[2] Associate Professor, Transportation Engineering Division, Civil engineering Department,Anna University Chennai, Tamil Nadu,600 025,India.

Abstract

Data Forecasting is a crucial task performed by the port authorities to plan their daily operations and future resource requirements. The short term forecasting is an effective tool for estimating the resource requirements for service time of ships of similar tonnage and Cargo. Forecasting ship service time for required operations is conventionally done using the standard algorithms and assumptions. The regular forecasting methods were decomposition, smoothing, and Box-Jenkins procedures. The focus on forecasting is to minimise the inherent errors and to make the inevitable errors as smallest as possible. An attempt has been made to perform short term forecasting using MNLR & ANN utilising the time series port data and comparison of results is presented.

Keywords: *Forcasting, Bulk cargo ships, service time, artificial neural network technique, Multivariate nonlinear regression model, Quantitative model, Performance accuracy.*

1. Introduction

Time series data is a historical data collected over time for a particular item of interest. It has five major components: average, trend, seasonal influence, cyclical movements, and random error. There are numerous tools and methods for preprocessing, such as sampling, selecting a representative subset of a large population; transformation, manipulating raw data to produce a single input; denoising, organizing data for more efficient access; and feature extraction, pulling out specified data that is significant in some context. In a customer (ships) oriented port management, data pre-processing is a component which extracts valid sets of data showcasing the shippers' behaviour in ports called *user transactions*, which provide more useful information about user's transactions at the port. Data mining can assist researchers by speeding up their data analyzing process.

1.1 Past literature

Daniel Penai [1] compared the structure of three time series models for estimating future growth. Vedat Yorucu [2] evaluated four forecasting methods namely, Actual static, Double Exponential Smoothing, Holt Winters (HW) and Auto Regressive Moving Averages (ARMA) to forecast the accuracy of tourist arrival data to North Cyprus and Malta. Amaury Lendasse et.al [3] proposed a generic non-linear approach for time series forecasting. George Duncan et.al [4] adopted the Bayesian pooling approach to draw information from analogous time series data to model and forecast future datasets. Haipeng Shen and Jianhua Z. Huang [5] considered forecasting the latent rate profiles of inhomogeneous Poisson processes.

Siem Jan Koopman et.al [6] explored a periodic analysis in the context of unobserved components; Prajakta S. Kalekar [7] concentrated on the analysis of seasonal time series data using Holt-Winters Exponential Smoothing methods namely multiplicative and the additive seasonal models. Michael J Dueker [8] presents a new QualVAR model incorporating information from qualitative and discrete variables in vector auto regressions. Roselina Sallehuddin [9] developed a better alternative hybrid model by combining a linear ARIMA and Grey Relational Artificial Neural Network (GRANN) for forecasting the time series data. The model was validated by comparing outputs with ARIMA, Multiple Regression (MR), GRANN, several hybrid models (MARMA, MR ANN, ARIMA ANN) and ANN trained using Levenberg Marquardt algorithm. Sabine Garbarino et.al [10] defined and reviewed the case of combining qualitative and quantitative approaches to impact evaluation. Fang Chen [11] proposed a new Markov chain Monte Carlo (MCMC) algorithm, which enables a wide range of complex statistical models for fitting a variety of linear and nonlinear, single-level and multilevel Bayesian models. Alysha M De Livera [12] emphasized that the predictability of an event depends on the factors that contribute to its occurrence, and unexplained variability involved. Forecasting situations might vary with time horizons, data patterns and the predictability of the quantity to be forecast. In this paper, the application of the various models to forecast service time of Ships carrying Bulk Cargo is analysed using MNLR & ANN. And comparison of results is presented.

2. Data collection:

The ship data related to port performance was collected from the Tuticorin port, South india [13] &[14], located in proximity to the world maritime route. The study port has 10 berths, out of which one is a dedicated berth for containers.The berths were handling various classified commodities such as Bulk, Break-bulk, Loose bulk and Coal. The Fig.1 shows the typical arrangement of berths in the study port. The bulk cargo ships related data was collected for five years (2005-2009) for analysis and an attempt has been made to forecast the future ship service time data using the various methods. The Table 1 shows the descriptive statistics of the Bulk cargo ship data collected and Fig.3 depicts the Log plot of the ship data.

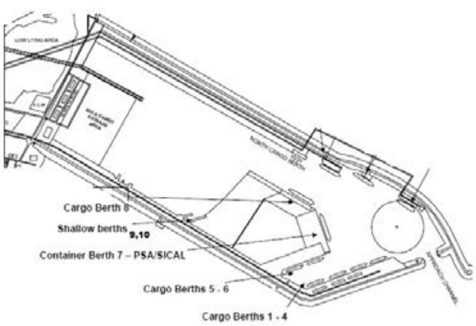

Fig.1 Location of berths in study port (Ref:13)

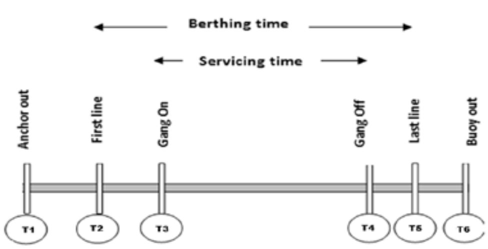

Fig 2 Concept of Ship service time

The ship service time is measured as the time lapse between the ship anchoring time at port and time of departure. The Fig.2 illustratres the concept behind the ship service time.

Table 1 Descriptive statistics of Bulkcargo Ship data

Statistics	Value
Count	1556
Min.	0.21
Max.	11.77
Mean	2.51
Median	2.38
S. D	1.70

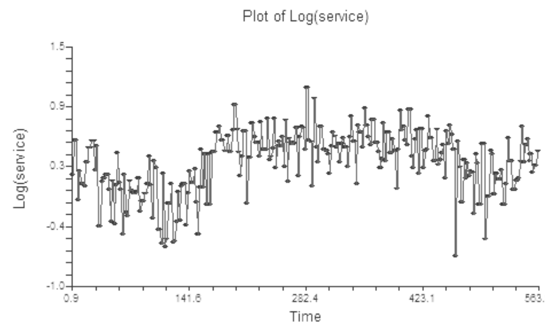

Fig.3 Log plot Bulkcargo ship data

3. METHODS OF FORECASTING

3.1 Quantitative Methods

The quantitative methods are widely adopted for short range time series based data forecasting in practice; however, the accuracy of such prediction would be less compared to the casual relationship models. The present research analyses the pattern and trend of the Ship data and to forecast the future values using the various algorithms such as simple average(SA) , moving average(MA), weighing moving average (WMA), single exponential smoothing (SES), single exponential smoothing with trend (SEST), Double exponential smoothing(DES), Double exponential smoothing with trend (DEST), Adaptive exponential smoothing (AES), Linear regression with time (LR), Holt winters additive algorithm (HWA), Holt winters multiplicative algorithm (HMA) were used to forecast the future ship service time from the time series data using a commercial software is presented in Table 2. And the observed and forecasted ship service time values were presented in Fig.4. The outcomes of forecasting differ with the method of forecasting. Also the trend of future values varied with each method because of the algorithm implemented and trend of historical time series data on ship service time.

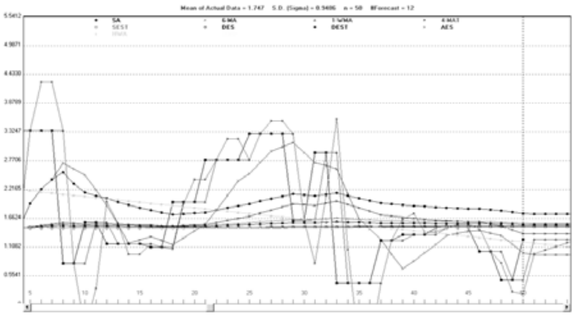

Fig.4 The observed and forecasted values of Ship service time by various Quantitative models

Table 2 Validation of Quantitative forecasting Models

Models	Performance validation of Quantitative models					
	MAD	**MAPE**	**Track Signal**	**CFE**	**MSE**	**R^2**
SA	.87	84.3	-10.8	-9.38	.97	.09
MA	.33	32.6	-6.7	-.22	.51	1
WMA	.80	89.4	-6.32	-5.03	.97	.35
MAT	.33	32.6	-0.67	-.22	.51	1
SES	.74	56.6	17.9	13.3	.97	.08
SEST	.74	56.6	17.9	13.3	.97	.08
DES	.74	56.6	17.9	13.3	.97	.08
DEST	.74	56.6	17.9	13.3	.97	.08
AES	.74	56.6	17.9	13.3	.97	.08
LR	.73	61.0	0.00	0.00	.78	.11
HWA	.74	56.6	17.9	13.3	.97	.08
HWM	.74	56.6	17.9	13.3	.97	.08

3.2 Casual Relationship Models

Since, the quantitative methods could not forecast the future service time effectively; the casual relationship methods such as Multivariate Non-Linear Regression (MNLR) and Artificial Neural Network technique (ANN) were attempted to forecast the bulk cargo ships' service time. The casual relationship model relates the independent variables which had a significant correlation with the dependant variable.

The independent variables chosen for the analysis were, Tonnage of cargo carried by the ship, loading rate of cargo, Number of manpower (gangs) employed and berth utilization rate in percentage; A multivariate correlation analysis was performed, to estimate the Pearson correlation coefficients among dependant and independent variables. The MNLR built to forecast the Bulk cargo ship service time is presented in Eq (1).

Nonlinear regression Model for future ship service time (Y)

$$Y = \left[\frac{\begin{array}{c} 0.68 + 1.8e^{-9}T^2 - 1.09e^{-18}T^4 - 14310.1 * \frac{1}{L^2} + 0.11 * T^2 * \frac{1}{L^2} \\ -3.22e^{7} * \frac{1}{L^4} + 1.18 * \frac{1}{G^2} + \\ 1.9e^{-8}T^2 * \frac{1}{G^2} - 43682 * \frac{1}{L^2} * \frac{1}{G^2} - 1.31 * \frac{1}{G^4} + 26.4 * \frac{1}{B^2} \\ -1.94e^{-7}T^2 * \frac{1}{B^2} + 6726597 * \frac{1}{L^2} * \frac{1}{B^2} - 491 * \frac{1}{G^2} * \frac{1}{B^2} - 2181 * \frac{1}{B^4} \end{array}}{\begin{array}{c} 1 + 8.1e^{-3} * T^2 * \frac{1}{L^2} + 1.14e^{-9} * T^2 * \frac{1}{G^2} \\ -56159 * \frac{1}{L^2} * \frac{1}{G^2} - 7.5e^{-8} * T^2 * \frac{1}{B^2} \\ +1.36e^{7} * \frac{1}{L^2} * \frac{1}{B^2} - 217 * \frac{1}{G^2} * \frac{1}{B^2} \end{array}} \right] \quad (1)$$

3.3 Artificial Neural Network Model

Artificial Neural Network is an emulation of biological neural system which could learn and calibrate itself. It was developed with a systematic step-by-step procedure to optimize a criterion, the learning rule. The input data and output training was fundamental for these networks to get a optimized output. The neural network was good at studying patterns among the input data and learns. The prediction accuracy increases with the number of learning cycles and iterations. The capabilities of MATLAB's neural networks module was utilized to build an ANN model. The inputs were given as batch files and the script programming was used to run neural network model iterations.

Fig.5.0 shows the hidden layer structure of the neural network. The validation of neural network model was done by estimating the NASH coefficient (NSE) and Sum of squares of errors (SSE values), by comparing the observed values and predicted values of ship's service time. Table 3 shows the number of data sets used for training, testing and production. The observed values and the ANN forecasted values of ship service time were presented in Fig.6.

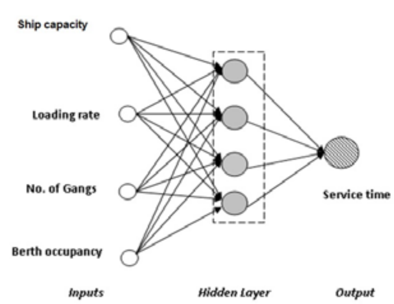

Fig.5. Hidden structure of proposed artificial neural network

4. Validation of ANN models

4.1 Nash–Sutcliffe Efficiency Index (NSE)

Recognizing the limitations of the correlation coefficient, Nash and Sutcliffe, proposed an alternative Goodness-of-fit index, referred to as the efficiency index presented in Eq (2). In statistical terms, it was a quantity that reflects the total variation of the observed values about the mean.

$$NSE_j = 1 - \sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2 / \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 \qquad (2)$$

4.2 Sum of squares of error (SSE)

The SSE measure the total deviation of the response values from the fit to the response values and the expression is presented in Eq (3).

$$SSE = \sum_{i=1}^{n} w_i \left(Y_i - \hat{Y}_i\right)^2 \qquad (3)$$

Where,
$Y_i^{\wedge}$ and *Yi were* predicted and measured values of the criterion dependent variable *Y*;
$\overline{Y}$ Was the mean of measured values of *Y*; and
n. was the sample size.

Table 3 Datasets used in ANN model

Category	Training	Testing	Production
Number of data sets	623	466	467
Percentage	39.9	29.9	30.1
Neurons used	22		

5. Results and Discussion

The Fig.6 shows plot of the observed and forecasted values of Bulk cargo ship service time by MNLR and ANN models. The validation of Multivariate nonlinear regression model was done using the measures such as Cumulative forecast error(CFE),Mean absolute deviation (MAD),Mean square error (MSE), Mean absolute percent error (MAPE) and R^2 values. The outputs of the quantitative models presented in fig.4 shows the mismatch between the trend of the ship data and forecasted data.

Also, the validation parameters such as, MAD error (0.33 to 0.87), MAP error (32.6 to 89.4), Track signal (-10.8 to 17.9),CFE (-9.38 to13),MSE (0.51 to 0.97), R^2 Value (0.08 to 1) were estimated. Also few quantitative models, such as MAT,LR, forecasted the future values of ship service time with minimum accuracy. The validation results of these models were presented in Table 4.

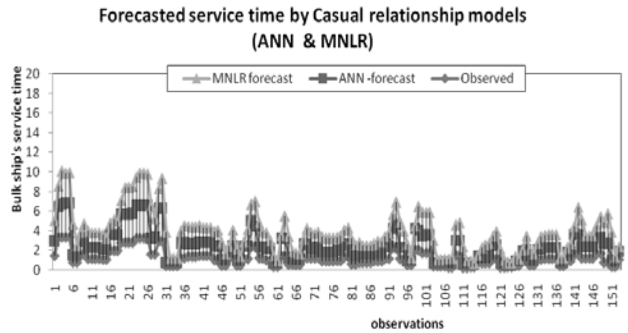

Fig.6 Forecasted ship service time values by MNLR & ANN

Table 4 Performance of ANN & MNLR models

Model	Criteria	Values
ANN	NSE	0.98
	SSE	0.0004
	R^2	0.90
MNLR	RMS	0.534
	CV	0.21
	R^2	0.90

6. Conclusions

The ship service time forecasting was carried out with the different quantitative models utilizing the historical time series data. Also, an attempt was made to forecast the ship data selecting a set of independent variables using Multivariate non-linear regression) and Artificial Neural Network technique (ANN). The MNLR model and ANN model (R^2, 0.9 and NSE, 0.98) outperforms the conventional time series quantitative methods. Hence, the casual relationship models such as nonlinear regression and ANN techniques would be best suited to obtain reliable and accurate forecasts of Service time of Ships carrying bulk cargo. However, Quantitative models could be used for crude predictions with historical time series data**.**

References

[1] Daniel Pena, “Forecasting growth with time series models”, Journal of Forecasting, vol.14, No.2, 1995,97-105.
[2] Vedat Yorucu, “The Analysis of Forecasting Performance by Using Time Series Data for two Mediterranean Islands, Review of Social, Economic and Business Studies, Vol.2, 2003, 175-196.
[3] Amaury Lendasse, Michel Verleysen, Eric de Bodt,Marie Cottrell, Philippe Grégoire,“ Forecasting Time-Series by Kohonen Classification”, ESANN'1998 proceedings - European Symposium

on Artificial Neural Networks Bruges (Belgium), D-Facto public., ISBN 2-9600049-8-1,1998, 221-226.
[4] George Duncan Wilpen Gorr,Janusz Szczypula H. John Heinz,“Forecasting Analogous Time Series”,School of Public Policy and Management, Carnegie Mellon University, Pittsburgh, PA 15213-3890,2001.
[5] Haipeng Shen, Jianhua Z. Huang,“Analysis of call centre arrival data using singular value decomposition”,Applied Stochastic Models in Business and Industry,Vol.21,No.3,2005,251-263.
[6] Siem Jan Koopman , Marius Ooms, “Forecasting daily time series using periodic unobserved components time series models”, Computational Statistics & Data Analysis,Volume 51, Issue 2,2006,885-903.
[7] Kalekar , Prajakta S. “Time series Forecasting using Holt-Winters Exponential Smoothing”, Kanwal Rekhi School of Information Technology, 2004.
[8] Michael J. Dueker & Katrin Wesche, “Forecasting macro variables with a Qual VAR business cycle turning point index," Federal Reserve Bank of St. Louis, Working Papers 2001-019, 2005.
[9] Roselina Sallehuddin, Siti Mariyam Hj. Shamsuddin, Siti Zaiton Mohd. Hashim and Ajith Abraham, , “Forecasting time series data using hybrid grey Relational artificial neural network and Auto regressive integrated Moving average model”,Neural Network World,vol.6,2007, 573-605.
[10] Sabine Garbarino and Jeremy Holland, “Quantitative and Qualitative Methods in Impact evaluation and Measuring Results”, UK (DFID), Emerging Issues Research Service of the Governance and Social Development Resource Centre (GSDRC). ISBN 0 7044 2756 7 / 9780704427563, 2009.
[11] Fang Chen, “Bayesian Modeling Using the MCMC Procedure”, SAS Global Forum 2009, Paper 257-2009.
[12] Alysha M De Livera, Rob J Hyndman and Ralph D Snyder, “Forecasting time series with complex seasonal patterns using exponential smoothing”, Working Paper 15/09, ISSN 1440-771X,2010.
[13] Tuticorin port trust website; http://www.tuticorinporttrust.org

[14] Ports of India website; http://www.portsofindia.nic.in

P.Oliver jayaprakash is at present Assistant professor in Civil Engineering Department, Mepco Schlenk Engineering College, Sivakasi,Tamilnadu. and currently pursuing his PhD. programme at Anna University, Chennai. His research interest includes Application of Soft computing in Transportation planning and management.

Dr.K.Gunasekaran is currently Associate professor in Transportation Engineering Division of Anna University, Chennai.Tamilnadu. His research interest includes Road safety, GIS & GPS applications in Traffic analysis and Management. He published several research articles in various journals.