

# Prediction of Marine Traffic Density Using Different Time Series Model From AIS data of Port Klang and Straits of Malacca

Akim Ramin<sup>a</sup>, Masnawi Mustaffa<sup>b</sup>, Shaharudin Ahmad<sup>a</sup>

In the study of ocean engineering, marine traffic is referring to the study of the pattern of the density of ships within the particular boundaries at certain periods. The Port Klang and Straits of Malacca are known for one of the heaviest traffics in Malaysia and the world. The study of traffic within this area is important, because it enables ships to avoid traffic congestion that might happen. Thus, this study is mainly aimed at predicting or forecasting the density of the ships using the route through this waterway by using quantitative methods which are time-series models and the associative models from the Automatic Identification System (AIS) data. The moving averages, weight moving average, and exponential smoothing for the time series model and associative model have used multiple regression. The results show an exponential smoothing alpha 0.8 and give the lowest MAPE as 20.701%, thereby making this method to be the best in forecasting the future traffic density among the method categories.

## KEY WORDS

- ~ AIS
- ~ Traffic density
- ~ Port Klang
- ~ Malaysia
- ~ Straits of Malacca
- ~ Navigation
- ~ Time series model

a. Universiti Teknologi MARA, Faculty of Mechanical Engineering  
e-mail: akim.ramin@gmail.com

b. Universiti Teknologi MARA, Faculty of Applied Sciences  
e-mail: masnawimustaffa@gmail.com

doi: 10.7225/toms.v09.n02.006

This work is licensed under 

## 1. INTRODUCTION

A high traffic route is an alarming situation for any vessel using it as it is bound to have higher possibilities of accidents, causing financial loss, injury, or casualties, and damage to the environment. In this respect the shipping traffic at Strait of Malacca and Port Klang waterways is one of the highest traffic in Malaysia and the world for marine routes. The average statistic of Port Klang keeps increasing as, in 2013, the cargo and vessel statistics are 200,278,901 and 16,703 respectively (Mustaffa et al., 2019).

The AIS technology is mainly aimed at giving information in studying this field. In a brief explanation, the Automatic Identification System (AIS) technology is a system that enables a vessel to obtain about the encountered vessels, such as their position (Zaman et al., 2015), course, speed, and other parameters automatically by very high frequency (VHF) radio transmission (Kundakci et al., 2018). Using AIS (Wu et al., 2017; Kang et al., 2018) will improve and remedy the traffic congestion on open water as it gives required data for traffic analysis (Xiao et al., 2015; Zhang et al., 2019; Fiorini et al., 2016).

In studying the pattern of data, it always takes into consideration the affecting and responding variables and the season or weather that may affect the traffic congestion. Previous studies show that weather during storms or disasters affects the traffic or behaviour of ships in manoeuvres of vessels (Gao et al., 2017). Thus a seasonal pattern technique has been used in the analysis.

Several methods have been used for this study and the effectiveness of each method is measured by the percentage of error. It presents the graphs to show the pattern of the forecast and the actual situation. This is expected to improve the

effectiveness of predicting the overall density in the study area. Furthermore, the prediction or forecast may be useful for vessels in planning to use a route to avoid traffic congestion.

The scope of this research is the forecasting using time series and associative models. These methods are based on a set of data from July 2016 to December 2017. These data have been collected from our laboratory, using our own AIS receiver. These data do not represent real data of the whole Port Klang and Straits of Malacca. We are not using any data from related authorities in Malaysia. Rather, we are using our own data to study the pattern and measure the effectiveness of the methods.

### 1.1. Straits of Malacca

At the centre of one of the busiest shipping lanes that connect East and West, the Straits of Malacca play a vital role for the shipping industry (Cheng et al., 2019). The Straits of Malacca are located on the east coast of Indonesia's Sumatra Island and the west coast of the Malaysia Peninsula, and extend to the Straits of Singapore at its southeast end. Marine navigational hazards will emerge when using this straits due to the increasing amount of traffic, in addition to a geographically narrow straits (Mustaffa et al., 2019).



**Figure 1.** Traffic density using Straits of Malacca and Port Klang, Malaysia (Mustaffa et al., 2019).

## 2. QUANTITATIVE FORECASTING

Forecasting is a science of predicting future events as it gives the good application in economic forecasting, technology forecasts and also demands forecasts which are a projection of

a company's sales for each period in the planning horizon (Gao et al., 2017; Mustafa et al., 2019). Furthermore, the main aim of time series modeling is to carefully collect and rigorously study the past observations of a time series to develop an appropriate model which describes the inherent structure of the series. This model is then used to generate future values for the series, time series forecasting can thus be termed as the act of predicting the future by understanding the past (Ratnadip et al., 2013).

A forecasting system should follow seven basic steps(Heizer, 2016):

1. Determine the use of the forecast
2. Select the items to be forecast
3. Determine the time needed to make a forecast
4. Select the forecasting models
5. Gather the data needed to make the forecast
6. Make the forecast
7. Validate and implement the results

The quantitative forecasting and qualitative forecasts are a general approach to forecast categorized into time-series models and the associative model.

### 2.1. Time-series Model

This type of model predicts the assumption that the future is a function of the past. By using this, the data will analyse the previous past data to make a forecast. In this study, the approach used is Moving Average, Weighted Moving Average and the Exponential Smoothing.

### 2.2. Associate Model

This model is like the linear regression. It incorporates the variables or factors that might influence the quality being forecast. The variables that give the effect of the forecast is the seasonal pattern and this study involves this variable. The Multiple-Regression approach is used in this study for this category.

## 3. METHODOLOGY

### 3.1. AIS Data

All collected data was encrypted, analyzed daily and stored in the database. This dataset contains all traffic in the area of study.

The scope of this study is obtaining the general equation or the pattern of the data distribution of traffic density. The dataset used contains the statistical data from July 2016 until December 2017.

The data are a collection of 499 day dataset, containing the total of ships. We categorized the data by the four seasonality

quarter of the year. Each season represents months from January to March, April to June, July to September, and October to December, denoted as 1 to 4 respectively.

### 3.2. Moving Average (MA)

A moving average forecast uses many historical actual data values to generate a forecast. A 4-month moving average is found by simply summing the demand during the past 4 months and dividing it by 4. The following is the formula for calculating the moving average:

$$MA = \frac{\sum \text{Num.of ships in previous } n \text{ period}}{n} \quad (1)$$

Where  $n$  is the number of periods in the moving average.

### 3.3. Weighted Moving Average (WMA)

The trend of the pattern can be more emphasis on recent value when considering the weights. The following formula represents the weighted moving average:

$$WMA = \frac{\sum (\text{Total ships in previous } n \text{ period}) (\text{weight for period})}{\sum \text{Weight}} \quad (2)$$

The weight for each calculation represents ascending number of the number of ships in  $n$  period.

### 3.4. Exponential Smoothing

Exponential smoothing is also known as a weighted moving average as it involves very little record keeping of past data. The  $\alpha$  is a weight or smoothing constant that has a value greater than or equal to 0 and less than or equal to 1. This study will show the results of different value  $\alpha$  and obtain the best value  $\alpha$  in this forecast. The following formula show the exponential smoothing:

$$F_t = F_{t-1} + \alpha (A_{t-1} + F_{t-1}) \quad (3)$$

Where:

$F_t$  = new forecast

$F_{t-1}$  = previous period's forecast

$\alpha$  = smoothing constant ( $0 \leq \alpha \leq 1$ )

$A_{t-1}$  = previous period's actual number of ships

### 3.5. Multiple Regressions

Unlike time-series forecasting, associative forecasting models take into consideration several variables that are related to the quantity being predicted and make this method more reliable than others. Multiple regression analysis is a practical extension of the simple regression model, a straight-line mathematical model to describe the functional relationships between independent and dependent variables. It is represented by the following equation:

$$y = a + b_1 x_1 + b_2 x_2 \quad (4)$$

Where  $y$  is a dependent variable,  $a$  is a constant,  $x_1$ , and  $x_2$  are values of two independent variables, seasonal  $b_1$  and  $b_2$  are coefficients for the two independent variables.

### 3.6. Computational Approach of Multiple Regressions

The least-square method has been used to compute the independent variables to the dependent variables which are the density of traffic.

### 3.7. Mean Absolute Percent Error(MAPE)

The average of the absolute differences between the forecast and actual values is expressed as a percent of actual values. The formula for this error is as follows:

$$MAPE = \frac{\sum_i^n \frac{100 | \text{Actual}_i - \text{Forecast}_i |}{\text{Actual}_i}}{n} \quad (5)$$

Where  $n$  is a period for the forecast and actual values.

## 4. RESULTS AND DISCUSSION

### 4.1. Traffic Density Analysis

The statistical summary for all quarters of actual traffic within the study time frame of 499 day is tabulated in Table 1,

showing the trend line equation with maximum and minimum amount by quarter. The line graph of 6 quarter of years for July 2016 until December 2017, as shown in Figure 2-7, represents the distribution of daily traffic, which always fluctuates, and trend line of actual traffic density by each quarter.

### 4.2. Moving Average

For each group of four month data or seasonal group of data, the forecast for the consecutive days uses the first four days of the number of ships in each seasonal group. Based on the equation (1), a total of four days is divided by four and forecast value of the 5th day is obtained. This iteration is continuously being forecast until the 499th day. The MAPE was calculated for the comparison.

### 4.3. Weighted Moving Average

In this type of forecast, the more heavily weighted the latest days provide a more accurate projection. Weighted moving average for four days projection was calculated using the equation (2). The MAPE has been calculated for comparison with other methods.

### 4.4. Exponential Smoothing

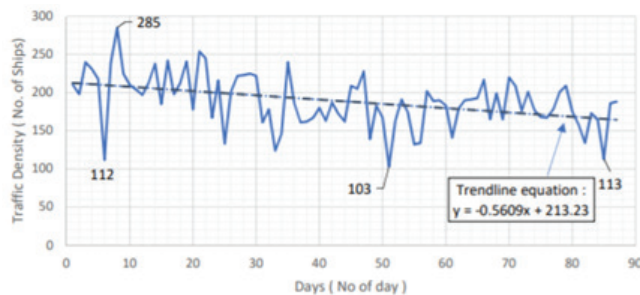
The equation (3), applying the forecast as the initial forecast for the first day, is 205. This consecutive iteration continues until the 499th day. In obtaining the best smoothing constant, try-and-error technique has been used to calculate the best alpha value and MAPE each difference smoothing constant,  $\alpha$  was calculated and tabulated as in Table 2.

From this table, the best value smoothing constant to match the forecasting or closest to the actual is  $\alpha = 0.8$ .

**Table 1.**

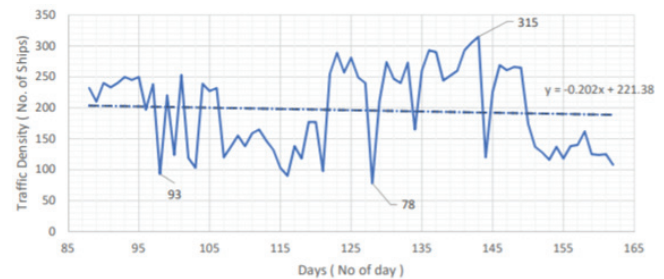
Traffic distribution with trend line equation.

Quarter	Period (months)	Trend line equation	Max. amount (number of ships)	Min. amount (number of ships)
3rd	Jul - Sep 2016	$y = -0.5609x + 213.23$	285	103
4th	Oct - Dec 2016	$y = -0.202x + 221.38$	315	78
1st	Jan - Mar 2017	$y = 0.7305x + 0.7559$	339	72
2nd	Apr - Jun 2017	$y = -0.3325x + 299.44$	349	36
3rd	Jul - Sep 2017	$y = 1.429x - 369.33$	367	65
4th	Oct - Dec 2017	$y = -0.0488x + 203.83$	282	77



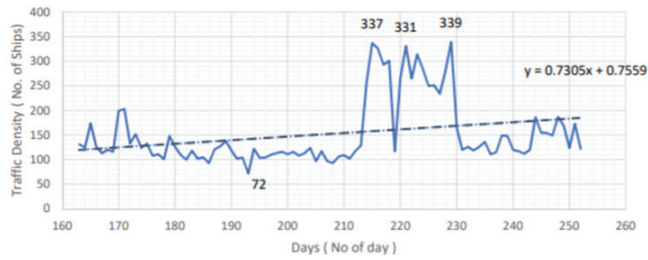
**Figure 2.**

The third quarter of 2016 traffic distribution from day 1 to day 87.

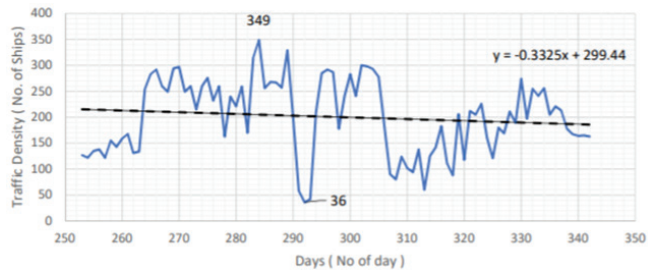


**Figure 3.**

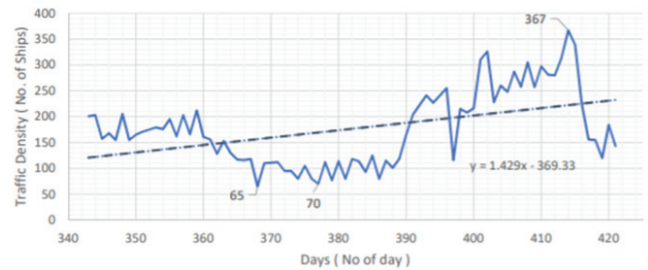
The fourth quarter of 2016 traffic distribution from day 88 to day 162.



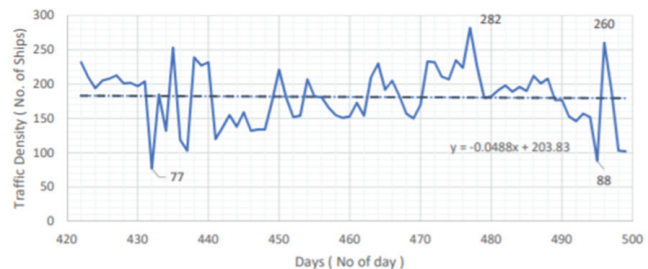
**Figure 4.**  
The first quarter of 2017 traffic distribution from day 163 to day 252.



**Figure 5.**  
The second quarter of 2016 traffic distribution from day 253 to day 342.



**Figure 6.**  
The third quarter of 2017 traffic distribution from day 343 to day 421.



**Figure 7.**  
The fourth quarter of 2016 of traffic distribution from day 422 to day 499.

**Table 2.**

MAPE for difference smoothing constant,  $\alpha$ .

Method	MAPE
Exponential smoothing, $\alpha = 0.4$	21.648
Exponential smoothing, $\alpha = 0.5$	21.147
Exponential smoothing, $\alpha = 0.6$	20.880
Exponential smoothing, $\alpha = 0.7$	20.738
Exponential smoothing, $\alpha = 0.8$	<b>20.701</b>
Exponential smoothing, $\alpha = 0.9$	20.827

#### 4.5. Associative Models

The multiple-regression is obtained as the equation considering the 4 variables, one variable representing the number of days, and the other 3 representing the set of seasonal variables. The following is the set of variables used. Table 3 shows the independent variable in which this density traffic is influenced by seasonal by quarter year.



**Figure 8.**  
Comparison of Time Series Forecasting of Day 5 to Day 20.

**Table 3.**

The variable represent the seasonal variables.

Seasonal	X1	X2	X3
1st quarter	0	0	0
2nd quarter	1	0	0
3rd quarter	0	1	0
4th quarter	0	0	1

From the general equation of Equation (4), applying the formula as the variables and the variables of  $x_1 = \text{days number}$ ,  $x_2 = X1$ ,  $x_3 = X2$ ,  $x_4 = X3$  yielding;



$$y = 159.327 - 0.0337x_1 + 51.211x_2 + 30.4209x_3 + 39.252x_4 \quad (6)$$

#### 4.6. Comparison by Forecasting Method

Mean Absolute Percentage Error (MAPE) values were compared to measure the effectiveness of the actual value for each of the methods. Applying Equation (5), data iteration for 499 days from dataset gives the following value respectively to the method used in this project as in Table 4. Trend line for each method is shown in Figure 8.

**Table 4.**

Comparison Using Different Method and its MAPE.

Method	MAPE (%)
Moving average	30.834
Weighted moving average	22.134
Exponential smoothing, $\alpha = 0.8$	20.701
Multiple regression	32.183

This comparison shows the exponential smoothing with alpha value 0.8 as the best method compared with other methods. The comparison by the plotting the graph for all method compared to actual data plotting is shown in Figure 9.

#### 4.7. Forecasting Future Traffic Density

From the analysis overall method, the reliable method in this study is the exponential smoothing alpha 0.8, as to perform the forecasting the extrapolating of recent data is required for continuity of the data projection. For the case, the date required is the 1<sup>st</sup> January 2018 and iteration using equation (3) to obtain the traffic density for each consecutive day. The plotting graph pattern of the previous week and the predicted day is shown in Figure 9.

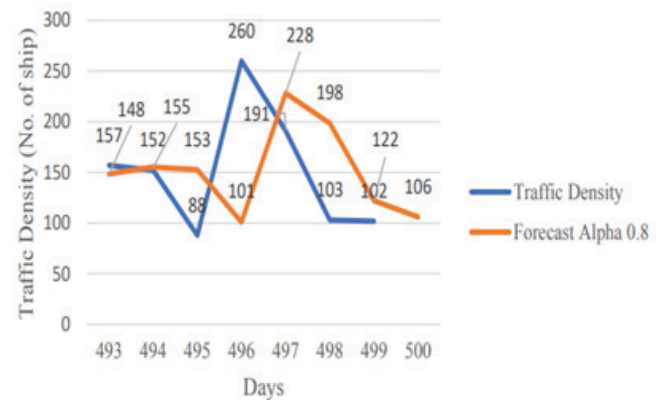
**Table 5.**

The iteration for the consecutive day in Jan 2018.

Date	Month	Num. of Days	Num. of Ship	Alpha=0.8	Absolute Error (%)
30	12	498	103	198.437	92.658
31	12	499	102	122.087	19.694
1	1	500		106.02	

## 5. CONCLUSION

In this study, the main objective of traffic forecasting analysis has been presented and discussed. The quantitative method has been used to forecast total density and by days as time horizon used to project the estimation density. The data are taken from July 2016 until December 2017, which only comprises 499 days as some of the data have not been completed due to some error in receiving the AIS signal. The quantitative method which is the time-series model; moving average, weighted moving average, exponential smoothing, and associative model; multiple-regression approach. These methods have been measured using the MAPE. The result showing the exponential smoothing alpha 0.8 gives the lowest MAPE as 20.701% and this will make this method reliable in terms of indicating or forecasting the future traffic density by the number of ships.



**Figure 9.**

Plotting of Forecasting using exponential smoothing alpha 0.8.

## REFERENCES

- Adhikari, R. & Agrawal, R.K., 2013. An Introductory Study on Time Series Modeling and Forecasting.
- Cheng, L. et al., 2018. Using big data to track marine oil transportation along the 21st-century Maritime Silk Road. *Science China Technological Sciences*, 62(4), pp.677–686. Available at: <http://dx.doi.org/10.1007/s11431-018-9335-1>.
- Fiorini, M., Capata, A. & Bloisi, D.D., 2016. AIS Data Visualization for Maritime Spatial Planning (MSP). *International Journal of e-Navigation and Maritime Economy*, 5, pp.45–60. Available at: <http://dx.doi.org/10.1016/j.enavi.2016.12.004>.
- Gao, X. & Makino, H., 2017. Analysis of anchoring ships around coastal industrial complex in a natural disaster. *Journal of Loss Prevention in the Process Industries*, 50, pp.355–363. Available at: <http://dx.doi.org/10.1016/j.jlp.2016.12.003>.

Heizer, J., 2016. Operations Management: Sustainability and Supply Chain Management. Pearson.

Kang, L., Meng, Q. & Liu, Q., 2018. Fundamental diagram of ship traffic in the Singapore Strait. *Ocean Engineering*, 147, pp.340–354. Available at: <http://dx.doi.org/10.1016/j.oceaneng.2017.10.051>.

Kundakçi, B. & Nas, S., 2018. Mapping Marine Traffic Density by Using AIS Data: An Application in the Northern Aegean Sea. *Polish Maritime Research*, 25(4), pp.49–58. Available at: <http://dx.doi.org/10.2478/pomr-2018-0131>.

Mustaffa, M. et al., 2019. Data mining analysis on ships collision risk and marine traffic characteristic of Port Klang Malaysia waterways from automatic identification system (AIS) data, *International Multi Conference of Engineers and Computer Scientists 2019 (IMECS 2019)*, 2019, vol. 2239, pp. 242–246.

Wu, L. et al., 2016. Mapping Global Shipping Density from AIS Data. *Journal of Navigation*, 70(1), pp.67–81. Available at: <http://dx.doi.org/10.1017/s0373463316000345>.

Xiao, F. et al., 2015. Comparison study on AIS data of ship traffic behavior. *Ocean Engineering*, 95, pp.84–93. Available at: <http://dx.doi.org/10.1016/j.oceaneng.2014.11.020>.

Zaman, M.B. et al., 2015. Risk of Navigation for Marine Traffic in the Malacca Strait Using AIS. *Procedia Earth and Planetary Science*, 14, pp.33–40. Available at: <http://dx.doi.org/10.1016/j.proeps.2015.07.082>.

Zhang, L., Meng, Q. & Fang Fwa, T., 2019. Big AIS data based spatial-temporal analyses of ship traffic in Singapore port waters. *Transportation Research Part E: Logistics and Transportation Review*, 129, pp.287–304. Available at: <http://dx.doi.org/10.1016/j.tre.2017.07.011>.