ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

MODELLING TIDE PREDICTION USING LINEAR MODEL AND ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (ANFIS)IN SEMARANG, INDONESIA

Alan Prahutama and Mustafid
Departement of Statistcs, Diponegoro University, Semarang, Indonesia
E-Mail: alan.prahutama@gmail.com

ABSTRACT

Semarang is an administrative city in Central Java province that is inevitably suffer from tidal flooding phenomenon. Tidal flooding is caused by the rising of sea level. Forecasting methods are techniques in Statistical tools for decision making. Therefore, a forecasting of sea level becomes important. One of the method to forecast time series data is ARIMA which require fulfilment of assumptions. One other way to put aside assumptions is by using ARIMAX. Meanwhile, non-linear approach that does not require assumptions fulfilment is ANFIS. The forecasting of sea level using ARIMAX is better than ARIMA and ANFIS. It shows that a certain complex forecasting methods does not guarantee to result the best model. The resulting model is ARIMAX (0, 1, [3]) $(1, 0, 0)^{12}$ with 7 outliers which produces 4.82 of RMSE.

Keywords: tide, linear model, ANFIS.

1. INTRODUCTION

Semarang is becoming the center of some transportation facilities including Ahmad Yani airport, Tawang and Poncol train stations, also Terboyo bus station. Moreover, in the northern Semarang, there are a lot of industries, for instance, Terboyo industrial area, Candi industrial area, also Mangkang industrial area. However, these areas are fairly potential for suffering disaster. One of the disasters threatening Semarang is flood caused by the rise of sea level or called as tidal flooding. Tidal flooding is a phenomenon that always happens in northern Semarang because it closes to Java Sea and decreasing level of mainland surface. Tidal flooding is a natural phenomenon in which sea level is high and reaches land area. Global warming is expected to accelerate the current rate of sea level rise, inundating many low-lying coastal and inter-tidal areas (Zhang and Gorelick, 2014). One of the impacts from that flood is the distraction of economic activities in Semarang, particularly in industrial areas. Therefore, forecasting or prediction of sea level is really needed.

Forecasting methods are techniques in Statistical tools for decision making. Forecasting approach for time series data can be done using two ways, the linear and non-linear approach. Forecasting methods with linear approach uses Autoregressive Integrated Moving Average (ARIMA). Meanwhile, forecasting methods with non-linear approach uses neural network, fuzzy and Adaptive neuro fuzzy inference system (ANFIS). Furthermore, ARIMA model that includes outlier is called ARIMAX. Liang *et al*. (2014) regarding the prediction of sea level use Neural Network (NN), Bang-Fuh *et al.* (2007) used Wavelet and artificial NN to forecast and supplement of tides arround Taiwan dan South China. Tsong-Lin (2004) developed Back Propagation NN for long term tidal prediction. Remya *et al.* (2012) used Genetic Algorithm for forecasting tidal. Guner and Yumuk (2014)developed fuzzy are application of a fuzzy inference system for the prediction of long shore sediment transport and Chang and Lain (2014) used ANFIS for the prediction of monthly shoreline changes in northeastern Taiwan. Chadsuthi *et al.* (2012) was modeling seasonal leptospirosis and its association with rainfall and temperature in Thailand using ARIMAX. Pektas and HK Cigizoglu (2013) comapare ARIMA method and ARIMAX method. One of this results is ARIMAX model have more superior performances than ARIMA model. Forecast sea level in Darwin Harbor Australia using Neuro fuzzy and neural network technique was developed by Karimi *at al.* (2013). The results demonstated that the Neuro fuzzy and neural network models had similar forecast accuracy, and their accuracy were better than the ARIMA model. Hence, in this research, it is intended to develop sea level prediction in Semarang using ARIMA, ARIMAX and ANFIS. ANFIS model does not need assumption of white noise residual and normally distributed residual, unlike ARIMA and ARIMAX.

2. MATERIALS AND METHODS

2.1 ARIMA model

ARIMA is one of forecasting method for time series data which supports linearity relation. ARIMA is the method first introduced by Box and Jenkins in 1976and until now become the most popular method for forecasting univariate time series data (Suhartono, 2011). ARIMA model includes seasonal and non-seasonal model. Seasonal ARIMA model (p,d,q)(P,D,Q) can be written as follow (Box *et.al*, 1994):

$$\varphi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D z_t = \theta_q(B)\Theta_Q(B^S)a_t$$

z_t is a time series data, S the lenght of seasonal lag, B backshift operator, $\phi_p(B)$ is non-seasonal component of AR, $\Phi_P(B^S)$ is seasonal component of AR with P order,

ARPN Journal of Engineering and Applied Sciences

$\theta_q(B)$ non-seasonal component of MA, $\Theta_Q(B^S)$ seasonal component of MA with Q order, and a_t is residual.

2.2 Outlier analysis in ARIMA model

Outlier is classified into Additive outlier (AO), Innovative outlier (IO), Level shift (LS), and Transitory Change (TC). AO only gives effect in T^{th} observation. Model with outlier is as follow (Bowerman and O'Connell, 1993):

$$z_t = \sum_{j=1}^{k} \omega_j v_j(B)\, I_j^{(T_j)} + \frac{\theta(B)}{\phi(B)} a_t$$

where $I_j^{(T)}$ shows the existence of outlier at the T_j period.

2.3 Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS modelling uses fuzzy set and fuzzy inference system. Fuzzy set does not have a precisely defined criterion of membership (Khashei *et al.*, 2009). Input space is mapped to be weight through a function (w), called membership function. Membership functions are as follows in Table-1.

Table-1. Membership function in ANFIS.

Name function	Function
Trianguler	$\max\left(\min\left(\frac{y-a}{b-a}, \frac{c-y}{c-b}\right), 0\right)$
Gausian	$\exp\left\{-\left(\frac{y-\mu}{\sigma}\right)^2\right\}$
Gbell	$\frac{1}{1+\left\lvert\frac{y-c}{a}\right\rvert^{2b}}$

a and c are the parameters on base of triangle, while b is on the peak; μ is the center point and σ is the width of the function; c is the center point, a is the width of the function, and b is the cross over point on slope.

ANFIS is a combination between ANN and fuzzy inference system (FIS) where neural network algorithm is used to determine FIS parameter (Chang and Lain, 2014). Sugeno's fuzzy model is a sistematic approach in generating fuzzy ruleusing IF x is A AND y is B THEN z=f(x,y) form of rule. ANFIS uses hybrid algorithm which means, merging the Least Square Estimator (LSE) on forward phase and error back propagation (EBP) on backward phase. For a clear explanation of the ANFIS model, a system with an input vector of 4 dimensions, one output value and a five layer network is taken by Figure-1.

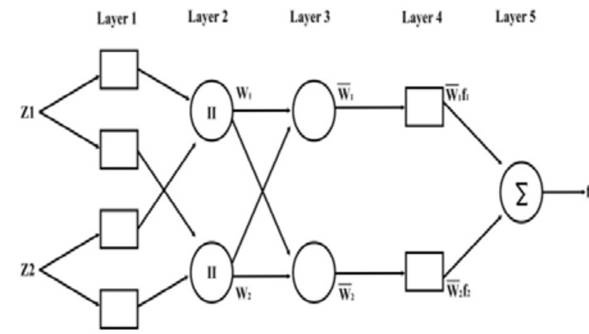

Figure-1. General architecture of the ANFIS.

Layer 1 (Input layer): map the input variables onto the fuzzy set. Layer 2 (Rule layer): conduct the fuzzy logic calculation (a total of sixteen rules in this example). Layer 3 (Normalization layer): normalize the results obtained from layer 2. Layer4 (Consequence layer): multiply the results obtained from layer 3 by the parameter of the Sugeno's fuzzy model. Layer 5 (Output layer): summarize the output results of layer 4.

3. RESULTS

3.1 Forecasting using ARIMA

The used data are average of sea levels in Semarang in period 2004 until 2013, or 120 in total. Data are divided into, 108 data for training and 12 data for testing.

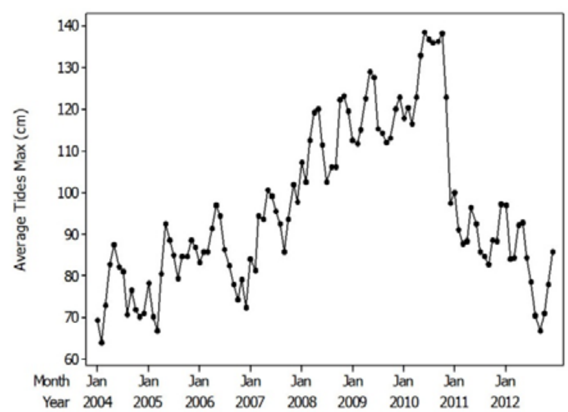

Figure-2. Scatter plot of sea level data in semarang 2004-2014.

Based on Figure-2, it shows that data have not reached stationary. So, they need to be differencedonce .Used ACF and PACF plots to identify the order of ARIMA model. ACF and PACF Plot as follow in Figure-3.

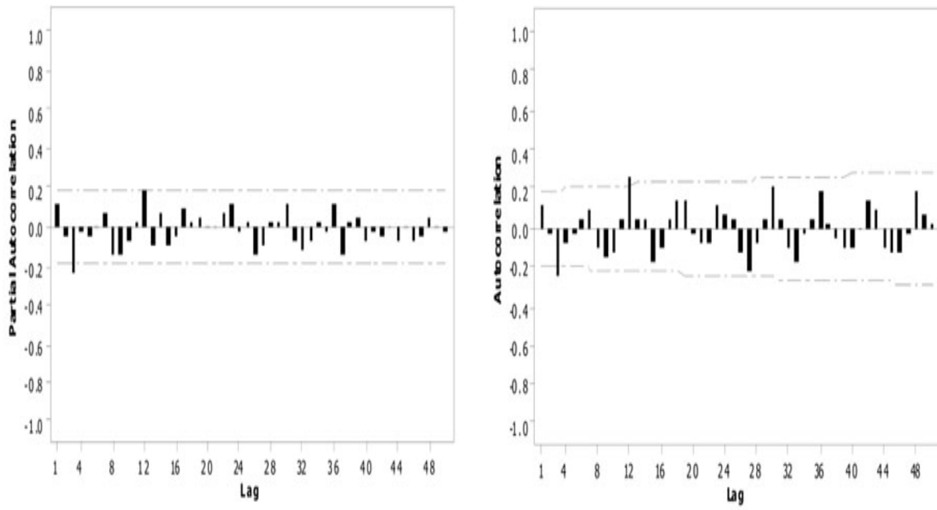

Figure-3. ACF and PACF plot.

Based on Figure-3, ACF and PACF plot show that data have reached stationary, lag that is off of limit is the 3rd lag, and indicates 12 seasonal pattern, so that the resulting model is ARIMA $(0,1,[3])(1,0,0)^{12}$. Then, estimation of ARIMA model is as follows in Table-2.

Table-2. Estimation of ARIMA model $(0, 1, [3])\ (1, 0, 0)^{12}$.

Parameter	Estimate	Std. Error	p-value	Decision
θ_3	0,20525	0,09617	0,0352	H_0 is rejected
Φ_1	0,24802	0,10089	0,0156	H_0 is rejected

Based on Table-2, the parameters of ARIMA $(0, 1, [3])\ (1, 0, 0)^{12}$ are significant, so that the model of ARIMA $(0, 1, [3])\ (1, 0, 0)^{12}$ is as follows:

$$z_t = \frac{(1-0.20525\, B^3)\, a_t}{(1-0.24802\, B^{12})(1-B)}$$

Based on Table-3, it shows that ARIMA $(0, 1, [3])\ (1, 0, 0)^{12}$ model has met white noise assumption and normal distribution.

Table-3. White noise assumption testing and normality testing for residual.

White noise testing				
Model	Lag	Chi-Square	p-value	Decision
SARIMA $(0,1,[3])(1,0,0)^{12}$	6	2,30	0,6800	H_0 is accepted
	12	6,69	0,7539	H_0 is accepted
	18	10,00	0,8664	H_0 is accepted
	24	13,54	0,9170	H_0 is accepted
Normality testing				
Model	**Kolmogorov Smirnov**		**p-value**	
SARIMA $(0,1,[3])(1,0,0)^{12}$	0, 047		> 0.15	

VOL. 11, NO. 11, JUNE 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

3.2. Forecasting using ARIMA with outlier (ARIMAX)

Modeling ARIMAX, the first step is detection outliers. Then determine the best model ARIMAX. Table-4 shows outlier detection of sea level data.

Table-4. Outlier detection.

Obs	Time ID	Type	Estimate	Chi-Square	p-value
84	**Dec2010**	**Shift**	**-28,55**	**22,94**	**<,0001**
83	Nov 2010	Shift	-17,62	9,64	0,0019
58	Oct 2008	Shift	11,98	4,58	0,0323
15	March 2005	Add	-8,23	4,66	0,0309
98	Feb2012	Shift	-11,53	4,31	0,0378
49	Jan 2008	Shift	10,98	4,31	0,0378
37	Jan2007	Shift	10,87	4,27	0,0387

Based on Table-4, there are 7 outliers, so that for modelling using ARIMA (0, 1, [3]) $(1, 0, 0)^{12}$ with outlier (ARIMAX), it can use 1, 2, until 7 outliers. To Determine the best model ARIMA with outlier use Akaieke Criterian Information (AIC). The best model is which has the lowest AIC. Combination of outliers can be shows in Table-5.

Table-5. AIC value of ARIMAX (0, 1, [3]) $(1, 0, 0)^{12}$ model.

Addition of outlier	AIC
84	681,0522
84, 83	669,5059
84, 83, 58	666,2085
84, 83, 58, 15	662,1137
84, 83, 58, 15, 98	659,2011
84, 83, 58, 15, 98, 49	653,9233
84, 83, 58, 15, 98, 49, 37	649,1083

Based on Table 5, ARIMAX (0, 1, [3]) $(1, 0, 0)^{12}$ model with 7 outliers gives the best model among others.

Based on Table 6 ARIMAX model with 7 outliers is as follows:

$$z_t = \frac{(1-0.319B^3)a_t}{(1-0.457B^{12})(1-B)} - 29.8\,I_T^{(84)} - 19.51\,I_T^{(83)} - 11.56\,I_T^{(58)} - $$
$$-8.09\,I_T^{(15)} - 11.23\,I_T^{(98)} + 15.85\,I_T^{(49)} + 11.7\,I_T^{(37)}$$

Table-6. Parameter estimation SARIMA model with 7 outliers.

Parameter	Type	Estimate	t-value	p-value
θ_3	-	0,31934	3,16	0,0021
Φ_1	-	0.45681	4,62	<,0001
I_{84}	Shift	-29,79528	-7,07	<,0001
I_{83}	Shift	-19,51136	-4,55	<,0001
I_{58}	Shift	-11,56469	2,74	0,0072
I_{15}	Add	-8,08988	-2,74	0,0074
I_{98}	Shift	-11,23359	-2,42	0,0172
I_{49}	Shift	15,85446	3,49	0,0007
I_{37}	Shift	11,6691	2,55	0,0124

3.3 Forecasting using Adaptive Neuro Fuzzy Inference System (ANFIS)

Membership Function using Gbell with 3 members, grouped into grid partition. In ANFIS modelling, the used data are input and output variables. Input variable used here is lag 12 (one input). Table-7 shows parameter premis with Gbell Function and 3 memberships. From Table-7, estimates parameters is as follows:

$$\hat{z}_t = \overline{w}_{1t}(0{,}5877z_{t-12} + 48{,}26) + \overline{w}_{2t}(1{,}886z_{t-12} - 80{,}37) +$$
$$+\overline{w}_{3t}(-0{,}577z_{t-12} + 158{,}8)$$

Hence, it is one input model.

4. DISCUSSIONS

Based on Figure-2 the highest tide period in 2010, which average is 126.6 cm. It show that tidal flood phenomena in Semarang City. Eventhough ARIMA models met assumptions in linier model, modeling with another methods can be used too. One of methods is ARIMA with outlier (ARIMAX). Based on Figure-2, scatterplot of data is fluctually, it can be sign that the data have outliers. ARIMAX model more complex than ARIMA model. Comparison of forecasting using ARIMA, ARIMAX, and ANFIS are as follow in Table-7. Based on Table-7, the Root Mean Square Error (RMSE) value can be counted to obtain the best model. RMSE values for every model are, ARIMA 10.98; ARIMAX 4.82; and ANFIS 5.76. ARIMAX and ANFIS have more similar accuracy than ARIMA model. The best model between ARIMAX and ANFIS is ARIMAX model. It shows that the complex model not be sured has good accuraty. It was supported by Makridakis and Hibon (2000) that the complex methods do not necessarily provide more accurate forecasts than simpler ones.

5. CONCLUSIONS

ARIMA model of sea level in Semarang results ARIMA (0, 1, [3]) (1, 0, 0)12 model. In outlier detection, there are 7 outliers, that results ARIMAX (0, 1, [3]) (1, 0, 0)12 model with outlier. ARIMAX (0, 1, [3]) (1, 0, 0)12 model with 7 outliers results the smallest AIC value. Forecasting using ANFIS, include Gbell membership function with 3 members, and obtain 5.76 of RMSE value, greater than what ARIMAX has. It shows that a complex

REFERENCES

AO Pektas and HK Cigizoglu. 2013. ANN Hybrid Model Versus ARIMA and ARIMAX Models of Runoff Coefficient. Journal of Hydrology. 500: 21-36.

Bang-Fuh C, Han-Der W, and Chin-Chu C. 2007. Wavelet and artificial neural network analyses of tide forecasting and Supplement of tides around Taiwan and South China. Ocean Engineering. 34: 2161-2175.

BL Bowerman, RT O'Connell. 1993. Forecasting and Time Series: An applied Approach 3rd edition. California: Duxburry Express.

FJ Chang, HC Lain. 2014. Adaptive neuro-fuzzy inference system for the prediction of monthly shoreline changes in northeastern Taiwan. Ocean Engineering. 84: 145-156.

GEP Box, GM Jenkins, and GC Reissel. 1994. Time Series Analysis Forecasting and Control, 3rd edition, Englewood Cliffs: Prentice Hall.

H Zhang, SM Gorelick. 2014. Coupled impact of sea level and tidal marsh restoration on endangered California clapper rail. Biological Conservation. 172: 89-100.

HAA Guner and HA Yumuk.2014. Application of a fuzzy inference system for the prediction of longshore sediment transport. Applied Ocean Research. 48: 162-175.

Makridakis S., danHibon M. 2000. The M3-Competition: Results, Conclusions, and Implications. International Journal of Forecasting, 16: Hal. 451-476.

M Khashei, M Bijari, GA Ardali. 2009. Improvement of Autoregressive Integrated Moving Average Models using Fuzzy Logic and Artificial Neural Networks. Neurocomputing. 72: 956-967.

PG Remya, Raj K, Sujit B. 2012.Forecasting tidal currents from tidal levels using genetic algorithm. Ocean Engineering. 40: 62-68.

S. Chadsuthi, C. Modchang, Y. Lenbury, S Iamsirithaworn, and W Triampo. 2012. Modeling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time-series and ARIMAX analyses. Asian pacific journal of Tropical Medicine. pp. 539-546.

S Karimi, O Kisi, J Shiri, O Makarynskyy.2013. Neuro-fuzzy and neural network techniques for forecasting sea level in Darwin Harbor Australia. Computers and Geosciences. 52: 50-59.

Suhartono. 2011. Time series forecasting by using seasonal autoregressive integrated moving average: subset, multiplicative or additive model. Journal of Mathematics and Statistics. 1: 20-27.

SX Liang, MC Li and ZC Sun. 2008. Prediction model for tidal level including strong meteorologic effect using a neural network.Ocean Engineering. 35: 666-675.

Tsong-Lin L. 2004. Back propagation neural network for long term tidal prediction. Ocean Engineering. 31: 225-238.