Numerical Statistic Approach for Expert System in Rainfall Prediction Based On Data Series

Indrabayu[1,] Nadjamuddin Harun[2,] M. Saleh Pallu[3,] Andani Achmad[4]

[1] *Student of Doctoral Program Civil Engineering Hasanuddin University, Makassar Indonesia*
[1,2,4] *Department of Electrical Engineering, Hasanuddin University, Makassar, Indonesia*
[3] *Department of Civil Engineering, Hasanuddin University, Makassar, Indonesia*

Abstract

The potential of statistical approach in predicting rain fall is discussed in this paper. Two most implemented methods i.e. Auto-Regressive Integrated Moving Average (ARIMA) and Adaptive Splines Threshold Autoregressive (ASTAR) are compared in term of accuracy in prediction. Both methods are constructed to predict daily rainfall in the area of Makassar, Indonesia. Rain problem in Indonesia increasingly complex due to climate shifts that result in high intensity rainfall in the dry season so it is very influential on the development of many aspect of social-economy sector. A ten years daily data (2001-2010) obtained from BMKG (the Meteorology, Climatology and Geophysics). Several complementary data is also obtained from LAPAN (Government Space Agent). From various meteorological variables, four variables are selected for predicting rainfall- There are temperature, humidity, wind speed, and previous precipitation based on their high correlation to rain event.. These four variables are then input to the ARIMA and ASTAR. The accuracy of prediction is measured based on root mean square error (RMSE). ASTAR outperformed ARIMA with less RMSE which is 0.02 to 0.24.

Keywords: *ARIMA, ASTAR, Expert System, Rain Prediction,*

I. INTRODUCTION

Indonesia is a tropical country which has a high rainfall intensity. Rainfall is a stochastic process, which upcoming event depends on some other meteorology precursors. Many research have revealed these meteorology precursors especially those region affected by monsoonal. [1,2,3,4]. These precursors are sea surface temperature, land surface temperature, relative humidity, winds, geo potential height, and the surface pressure. A common methodology used in predicting daily rainfall intensity is harvesting abundant of previous daily rainfall data [5,6]. A statistical method is worth trying method for rain fall forecasting. This is due to the fact that statistical method can harvesting abundant of data and transform it into a simple line outlook i.e. auto regressive, moving average, and several other forms. A research has conducted in modelling ARIMA to forecast daily power consumption [7]. Though data forecast is highly deterministic but very complex since inflation rate is influenced by many parameters. Preliminary research on Daily Rainfall prediction has been conducted for Makassar which shows a promising results [8,9]. In this paper, ARIMA and ASTAR are compared in term of its special advantage in dealing stochastic data, in this case rainfall data, and future development for better predicting result. The paper is outline into 5 parts, i.e. (1) Introduction, (2)ARIMA modelling, (3) ASTAR Modelling, (4) Results and Discussion, and (5) Conclusions.

II. ARIMA MODELLING

ARIMA is used to predict a value in a response time series as a linear combination of its own past values, past errors, and current and past values of other time series. The ARIMA procedure provides a comprehensive set of tools for uni-variate time series model identification, parameter estimation, and forecasting, and it offers great flexibility in the kinds of ARIMA or ARIMA X models that can be analyzed.The ARIMA procedure supports seasonal, subset, and factored ARIMA models;; multiple regression analysis with ARMA errors; and rational transfer function models of any complexity. In general, the ARIMA procedure can be subtle as follows [8]:

Step 0) A class of models is formulated assuming certain hypotheses.
Step 1) A model is identified for the observed data.
Step 2) The model parameters are estimated.
Step 3) If the hypotheses of the model are validated, go to Step 4, otherwise go to Step 1 to refine the model.
Step 4) The model is ready for forecasting.

A. Step 0

In this step, a general ARIMA formulation is selected to model the rain fall data. This selection is carried out by careful inspection and selection of the main characteristics of the daily rain fall and other meteorological data. The corresponding data are: humidity, air pressure, surface land temperature and wind velocity (corresponding to daily respectively), among others.

B. Step 1

A trial model must be identified for the rain fall data. First, in order to make the underlying process stationary (a more homogeneous mean and variance), a transformation of the original rain fall data and the inclusion of factors of the form may be necessary. In this step, the checking process can be done using Autocorrelation function (ACF) or unit root test. A further check for lag residual and lag dependent tested from partial ACF.

C. Step 2

After the functions of the model have been specified, the parameters of these functions must be estimated. Good estimators of the parameters can be computed by assuming the data are observations of a stationary time series (Step 1). If a Moving Average (MA) pattern is identified then further optimization process needed by using maximum likelihood or least square estimation. A conditional likelihood function is selected in order to get a good starting point to obtain an exact likelihood function. Also, an option to detect and adjust possible unusual observations is selected. As these events are not initially known, a procedure that detects and minimizes the effect of the outliers is necessary. With this adjustment, a better understanding of the series, a better modeling and estimation, and, finally, a better forecasting performance is achieved.

D. Step 3

In this step, a diagnosis check is used to validate the model assumptions of Step 0. This diagnosis checks if the hypotheses made on the residuals (actual prices minus fitted prices, as estimated in Step 1) are true. Residuals must satisfy the requirements of a white noise process: zero mean, constant variance, uncorrelated process and normal distribution. These requirements can be checked by taking tests for randomness, such as the autocorrelation and partial autocorrelation plots. If the hypotheses on the residuals are validated by tests and plots, then, the model can be used to forecast prices. Otherwise, the residuals contain a certain structure that should be studied to refine the model in Step 1.

E. Step 4

In Step 4, the model from Step 2 can be used to predict future values of daily rainfall data. Due to this requirement, difficulties may arise because predictions can be less certain as the forecast lead time becomes larger. Based on the natural of data, time series forecasting is suit to short term forecasting (hourly or daily). For a long term period, a structural forecaster is more comply for the situation. The flowchart of corresponding steps above can be seen in Figure 1. Several historical daily data is collected from BMKG Makassar over 10 years periods (2001 -2010).

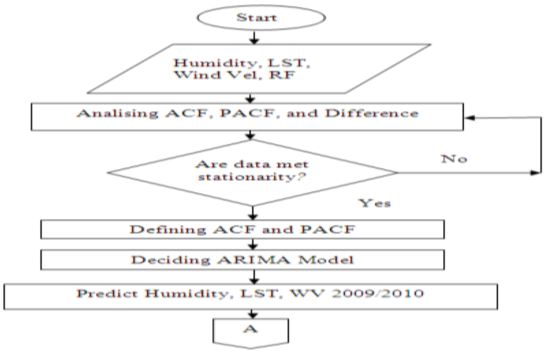
Start
Humidity, LST, Wind Vel, RF
Analising ACF, PACF, and Difference
Are data met stationarity?
No
Yes
Defining ACF and PACF
Deciding ARIMA Model
Predict Humidity, LST, WV 2009/2010
A

Figure 1. ARIMA process

2.1. DOUBLE REGRESSION

Double regression as part of multivariate analysis aiming on revealing the substantial relationship between two variable. A dependent variable Y is influenced by subsequent independent variable X. General view of process is shown in Figure 2.

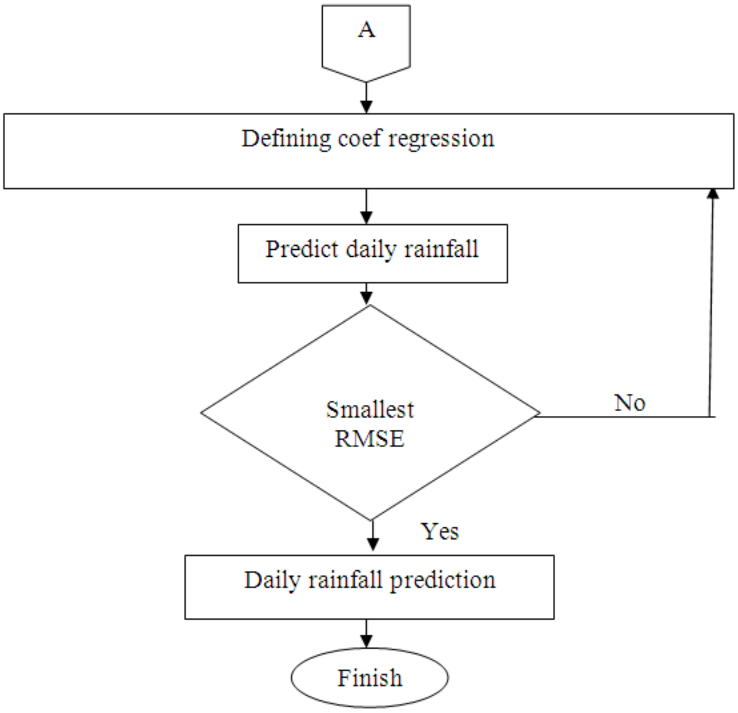

Figure 2. Double Regression Steps

The coefficient of independent variable Y (rain fall) and subsequent dependent variables Xn are formulated as follows:

$$Y = \beta0 + \beta1X1 + \beta2X2 + \beta3X3 \ldots + \beta nXni \qquad (1)$$

Y : Dependent/response Variable
X1 : Independence parame
ter 1
X2 : Independence parameter 2
Xn : Independence parameter n
β: Regression Coefficient

III. ASTAR Modelling

In modelling ASTAR several software are used and integrated to process the ASTAR result, I.e. Microsoft Excel, SPSS 16 and MARS 2.0 are the software for ASTAR planning system.Rain fall forecasting, as response variable (Y), Input variable, as predictor variable (X), is wind speed, humidity and temperature with X1, X2, and X3 respectively. All of predictor variables are applied to attain the best model of rain fall forecasting.The significant variable, influenced the next day condition with importance variable, is processed using MARS 2.0 Software.

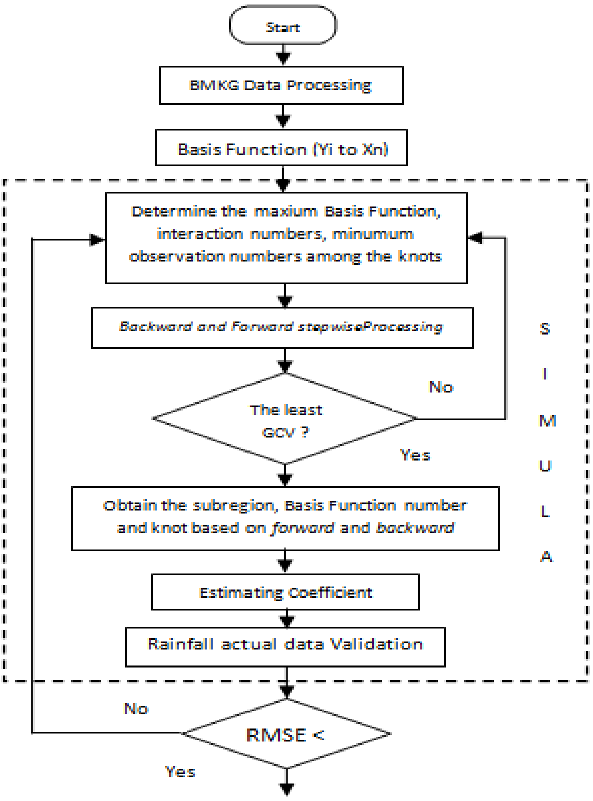
Start
BMKG Data Processing
Basis Function (Yi to Xn)
Determine the maxium Basis Function, interaction numbers, minumum observation numbers among the knots
Backward and Forward stepwiseProcessing
The least GCV ?
No
Yes
Obtain the subregion, Basis Function number and knot based on forward and backward
Estimating Coefficient
Rainfall actual data Validation
S I M U L A
RMSE <
No
Yes

Figure 3. Flowchart of ASTAR Methodology

Fuction Base

A Basis Function is distance between sequence knots. In ASTAR, Basis Function is a set of function to describe information that consist of one or two variables.

Max(0, x – t) or Min (0, t – x) is Basis Function value with t as a value to illustrate knot position and x as predictor variable. Every 1 knot will produce a couple of Basis Function.

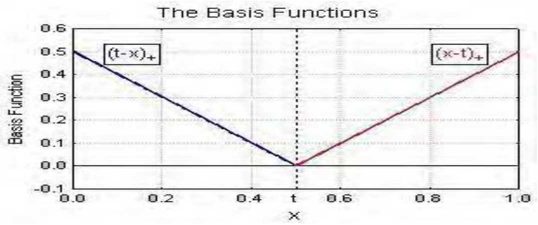

Figure 4. Basis Function

ASTAR Methods as data analysis technique to find the best model from a set of data. It is using past and present data to predict the short-term forecasting.

Modelling Stage of ASTAR

[1] Determine maximum Basis Function, maximum interaction numbers and minimum observation numbers between knots.

[2] Forward Stepwise Processing to obtain maximum number of Basis Function using MARS 2.0

[3] Backward Stepwise Processing to obtain Basis Function numbers from forward stepwise by minimizing the least GCV (Generalized Cross Validation) value.

[4] Knots selection using forward and backward algorithm.

[5] Estimating the coefficient of chosen Basis Function as a stage of response variable (Y) prediction (Y) to predictor variable (X).

Linearity on predictor variables is modelling main problem. One of MARS strategy to solve this problem, is reducing the modelling variable, thus it will diminish fake interaction due to its colinearity and will produce more stable forecasting.Variable declining could be completed by adding fine value in lack of-fit of knot selection. It is using forward algorithm. Model verification is applying RMSE (Root Mean Square Error) and MAE (Mean Absolut Error). This research is using RMSE to compare the accuracy of rain fall forecasting with observation data.

RMSE formula is shown bellow :

$$RMSE = \frac{\sqrt{\frac{1}{N}\sum_{t=h}^{N}(y_t - \hat{y}_t)^2}}{y\max - y\min} \qquad (2)$$

Where, yi = Observation value
ŷi = Forecasting value
n = Observation number
ymaks = maksimum observation data
ymin = minimum observation data

IV. RESULTS AND DISCUSSION

4.1. ARIMA

Stationary condition is checked for each variable. If variable does not meet the term, differencing process commencing. Figure 3 shows ACF of humidity from 2001-2008.

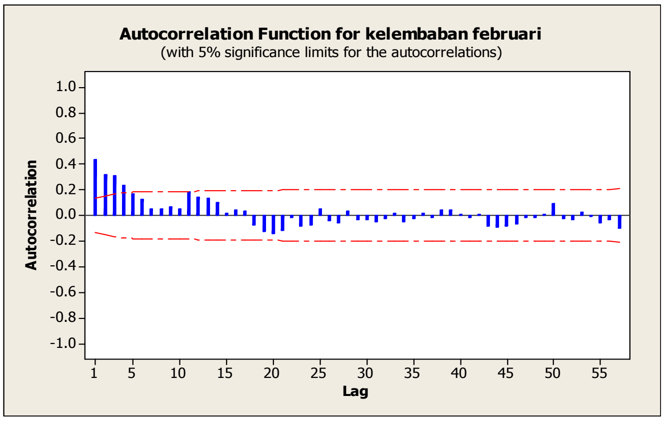

Figure 5. ACF of Humidity 2001-2008

The figures describe following terms:

[1] Time Lag 1-4 gain significance since out of range -0.130377<rk<0.130377.

[2] The ACF for humidity shows a not-stationary condition, hence differencing is needed.

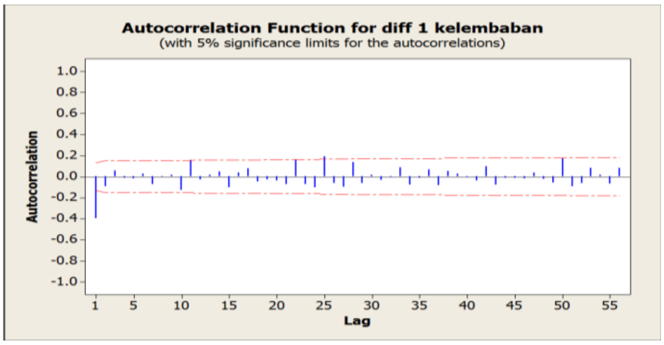

Figure 6. ACF Diff 1 of Humidity 2001-2008

Output of ACF diff 1 concluded points are:
[1] Time Lag 1,11, 25 gain significance since out of range -0.130377<rk<0.130377.
[2] As rule of thumbs, if the number of out of range time lags ≤ 3, then the ACF consider to be stationary.

Similar procedure imply for LST and Winds. When first differencing failed then next differencing need to be considered. The same procedure also conducted for PACF to find the MA value. The Mini Tab Software is used for calculating the coefficient of double regression. The outcome of processing is:

$$\text{Rain fall} = 1055 - 2.72\ H - 32.4\ LST + 3.53W \qquad (3)$$

From the formula above, daily rainfall can be predicted for 2009 and 2010.

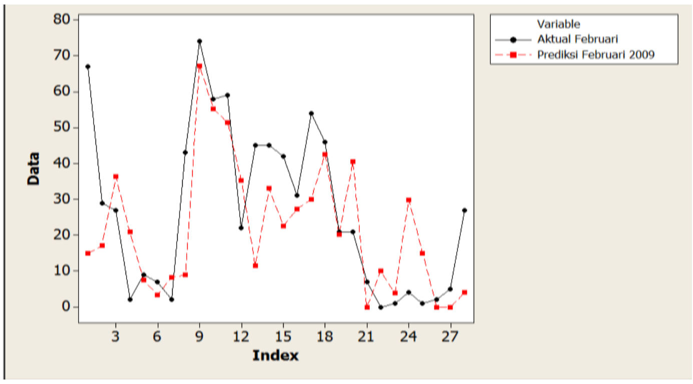

Figure 7. The actual vs prediction rainfall in february 2009

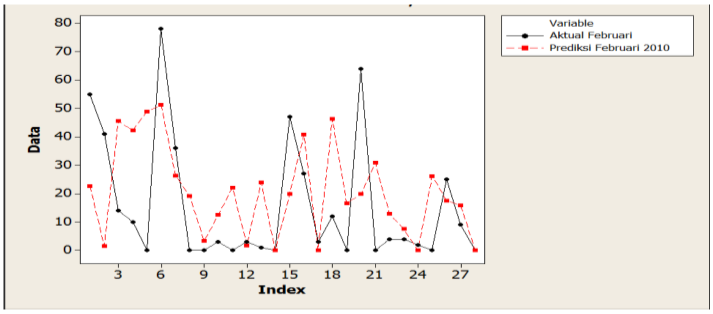

Figure 8. The actual vs prediction rainfall in february 2010

From both fig. 7 and fig 8, the prediction quite follow the actual data. The deviation range from 0-25 mm/hr which is acceptable if base on classification of rain by WMO. The root mean square (RMSE) calculated for year 2009 and 2010 are 0,242 and 0,301 respectively. A little drawback in 2010 is due to the fact that in 2010 the La Nina event occurred. There are also some disadvantages in the system since some data obtained from BMKG, particularly 2001-2003 periods has some blank events.

4.2. ASTAR

Before input parameter is determined, calculation of input variables correlation is important to know how those variables affect the rain fall. Input variable to rain fall condition is a variable with high correlation, thus the product could be used to forecast the rain fall. There are wind speed (knot), humidity (%) and temperature (C) as predictor variables. Rain fall forecasting for the year 2009 is using 2004 – 2008 BMKG Region IV, Makassar, data and forecasting of rain fall for the year 2010 is using 2004 – 2009 data. Relation of humidity, temperature and wind speed to rain fall is determined with MARS software. It will form a equation model to February 2009 prediction as follow:

$$Y = -9.296 + 36.493{*}BF2 + 5.351{*}BF4 + 1.302{*}BF5$$

While for Februari 2010 prediction has formulation model:

$$Y = -12.204 + 32.570{*}BF2 + 4.995{*}BF4 + 1.493{*}BF5$$ Where,

BF2 = max(0, 24.200 - Temperature);
BF4 = max(0, 12.000 – Wind, X1);
BF5 = max(0, Humidity - 69.000);

Rainfall modelling_FE = BF2 BF4 BF5

The next step is using Microsoft Excel to attain the nest modelling of rain fall forecasting for the year 2009 and 2010.

Y2009 = -9.296 + 36.493 * (24,2 - Tempt) + 5.351* (12- Wind) + 1.302 * (Humidity-69)
= 115,298 mm

Y2010 = (-12.204+(32.570*(24.3-Temprture))+(4.995*(12-WInd)) + (1.493* (Humidity-69)))
= 60.374 mm

Therefore, rain fall forecasting per 1 February 2009 and 2010 is 115,298 mm and 60,374 mm.

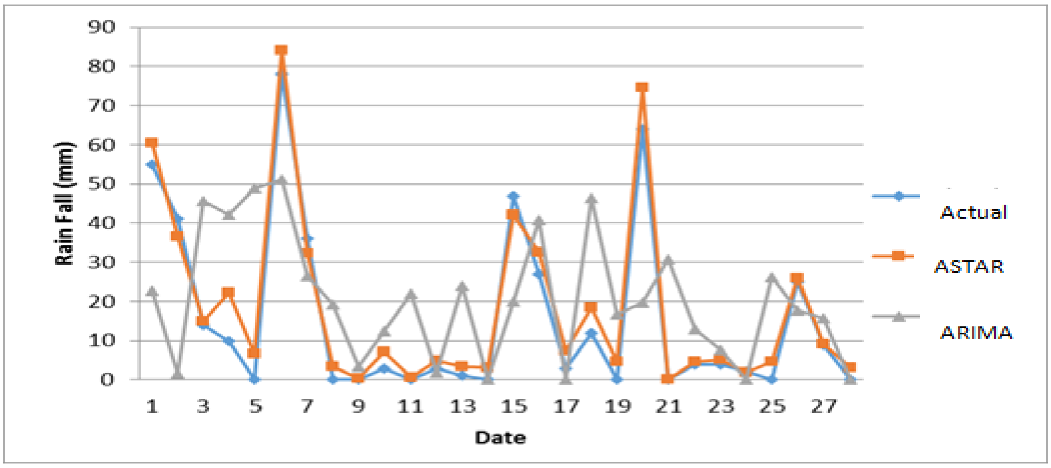

Figure 9. The actual vs prediction rainfall ARIMA and ASTAR in february 2010

From Fig. 9 it can be seen that ASTAR has better prediction compare to ARIMA. ASTAR also shows following trend to the actual data. It seem ARIMA cannot perform well in dealing stochastic data.

V. CONCLUSION

ASTAR method has a better prediction compare to ARIMA in general. It can been seen from the lower RMSE between 0,060757012 to 0,335681565 with average 0,1373 in year 2010 compare to ARIMA with RMSE from 0,19331303 to 0,440727825 with average 0,2942. ASTAR also shows a better following trend to the actual data since its feature in dealing with stochastic data.

REFERENCES

[1] E. Aldrian and Y.S Djamil, Application of Multivariate Anfis For Daily Rainfall Prediction: Influences Of Training Data Size, MAKARA, SAINS, VOLUME 12, NO. 1, APRIL 2008: 7-14.

[2] H. Wu, X. Lin, “Application of Fuzzy Neural Network to the Flood Season Precipitation Forecast”, International Joint Conference on Computational Sciences and Optimization IEEE, 2009.

[3] J.F. Nong, "Application of Nonparametric Methods in Short-range Precipitation Forecastng", International Joint Conference on Computational Sciences and Optimization IEEE, 2009.

[4] Fangqiong Luo, Chunmei Wu and Jiansheng Wu, "A Novel Neural Network Ensemble Model Based on Sample Reconstruction and Projection Pursuit for Rainfall Forecasting", ICNC, IEEE, 2010.

[5] I. Sonjaya, T. Kurniawan, “Uji Aplikasi HyBMG Untuk Prakiraan Curah Hujan Pola Monsunal”, Ekuatorial dan Lokal. BULETIN METEOROLOGI KLIMATOLOGI DAN Vol. 5 No. 3 SEPTEMBER 2009.

[6] Indrabayu, “Jaringan Sarat Tiruan dan Fuzzy Untuk Memprediksi Hujan”, Pertemuan Tahunan Nasional Teknik Elektro, FORTEI, 2011.

[7] J. Contreras, “ARIMA Models to Predict Next-Day Electricity Prices”, IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 18, NO. 3, AUGUST 2003.

[8] Indrabayu, N. Harun, M. S. Pallu, and A. Ahmad, Constructing Auto-Regressive Integrated Moving Average (ARIMA) as Expert System for Daily Precipitation Forecasting, The 2nd MICEEI International Conference, Makassar, Indonesia, 2011, pp.89.

[9] Indrabayu, N. Harun, M. S. Pallu, and A. Ahmad, Performance of ASTAR for Rainfall Forecasting, Proc. The 3rd MICEEI, Makassar, Indonesia, 2012, pp.327.