A Recommendation for Tropical Daily Rainfall Prediction Based on Meteorological Data Series in Indonesia

Indrabayu[1] and D.A. Suriamiharja [2]
[1] Informatics Department-Hasanuddin University, Makassar, Indonesia
[2] Physics Department-Hasanuddin University, Makassar, Indonesia

***Abstract*—The use of artificial intelligence (AI) and statistical methods for prediction based on data series have been widely used recently. In this paper both methods are implemented and compared in term of forecasting accuracy. The accuracy are measured in root mean square error (RMSE) and percentage of successful rainfall clustered to actual data. A hybrid AI i.e. Support Vector Machine and Fuzzy are combined for qualitative rainfall forecasting. A recommendation for sampling time and parameter used are also reviewed for better future forecast.**

***Index Terms*—Data Series, Rainfall, Precipitaion, Forcasting**

I. INTRODUCTION

Since time immemorial, experts have linked the rain events with other meteorological parameters. Factors such as temperature, humidity, air pressure, wind speed and solar radiation have been connected as precursor of rainfall rate. Many research has conducted to elucidate relationships among those factors. For developing country like Indonesia, rainfall forecasting using empirical data series are still used nowadays.

The development in methods in statistical analysis and massive used of artificial intelligence (AI) for expert system has significantly contributed in continuity of forecasting based on empirical data series. Many research worldwide have shown that AI and statistic do have extended capability in forecasting long and even short term of rainfall rate [1, 2, 3]. Several research also revealed the significance of other meteorological and climate factors in forecasting precipitation [4, 5, and 6]

In this paper, the authors focusing in improving existing methods used in Indonesia by Meteorology and Climatology and Geophysics Berau (BMKG) under expert system software so called HyBMG [7]. There are four algorithm used in HybMG i.e. ANFIS, Wavelet, TISEAN and Auto-Regressive Integrated Moving Average (ARIMA). The percentage accuracy for each methods is considerably still low. ARIMA shows best performance with percentage 79%.

The areas development are on the algorithms used in forecasting and also in viewing the impact of other meteorology factors recorded by weather instruments for decades. In this research, predictions based on qualitative approach are also introduced. Ten years of daily meteorological data from 2001-2010 are collected from BMKG Indonesia. Sub-Region A - Makassar City is chosen as research focus based on its strategic location i.e. center of Indonesia and close to equator.

This paper comprises of 5 parts i.e. Introduction; a new qualitative approach for precipitation forecasting; extensive comparison of powerful algorithms used for prediction based on data series; recommendation of sampling time and parameter used in forecasting; and conclusions..

II. SVM-FUZZY SCHEME

A. Proposed System

The proposed system combines the SVM and Fuzzy methods to achieve high accuracy of the rainfall prediction. Correlation for each parameter to rainfall are calculated. From five parameters recorded only three have significant correlation value. The input parameters chosen are humidity (*H*), wind velocity (*W*), temperature (*T*), and rainfall (*R*). The concept of the SVM-Fuzzy and its flowchart are shown in Fig.1. and Fig.2, respectively.

In SVM, eight years data (2001-2008) prepared as training data aiming on 2009 observation data. Re-training conducted over 2001-2009 periods for predicting the humidity, wind and temperature in 2010. The input data are used in kernel Radial Basis Function (RBF) to determine the support vectors. The position of support vectors is needed to figure out the weight vector (*w*) and bias (*b*) as prediction parameters.

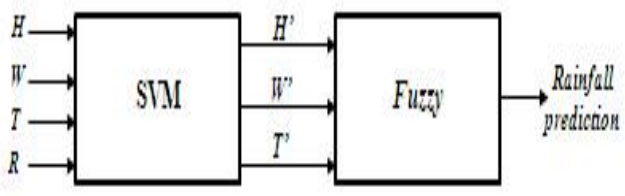

Fig.1 The concept of SVM-Fuzzy [8]

In fuzzy, there are three input variables, one output and 36 rules. Humidity and temperature have three membership functions, i.e. low, medium and high. While wind speed has four membership functions, i.e. calm,

moderate, strong, very strong. The rainfall prediction as an output of the fuzzy system is classified into sunny, very light rain, light rain, moderate, heavy rain, and very heavy rain. The BMKG rainfall classifications are shown in Table 1.

Table 1. Rainfall Classifications

Rain type	Rainfall (mm)	
	1 hour	24 hours
Very light (Clear Sky)	< 1	< 5
Light	1 s/d 5	5 s/d 20
Moderate	5 s/d 20	20 s/d 50
Heavy	10 s/d 20	50 s/d 100
Very heavy	> 20	> 100

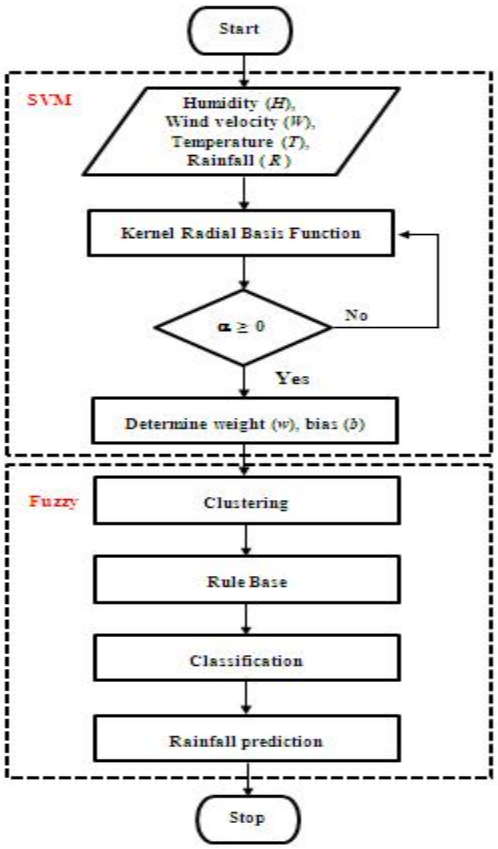

Fig.2 Flowchart of SVM-Fuzzy method [8]

B. Prediction of 3 meteorological input parameters

The prediction results of humidity, temperature, and wind speed of the SVM method in November 2009 are shown in Fig.3. to Fig.5, respectively. The selection of November as sample since it's the starting peak of rainy season.

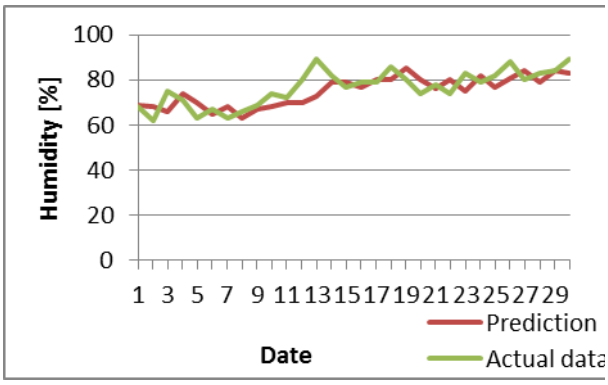

Fig.3 The prediction results of humidity [8]

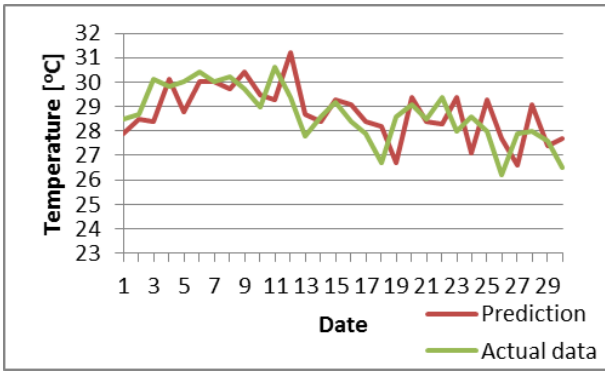

Fig.4 The prediction results of temperature [8]

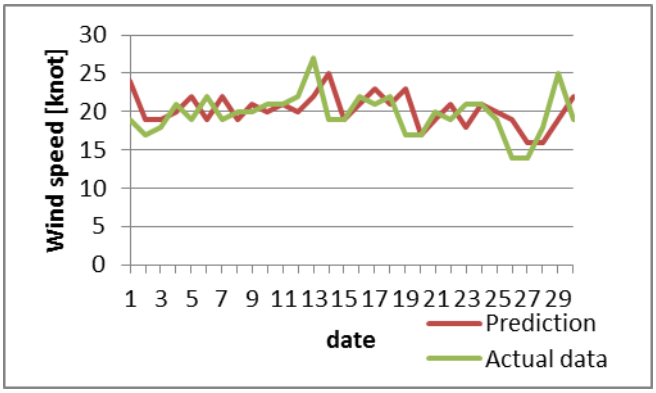

Fig.5 The prediction results of wind speed [8]

From Figure 3-5, a quantitative meteorological daily prediction are generated by SVM to be fed into Fuzzy. It can be seen that prediction is quite accurate and also following the trend of actual data. The forecast parameters are then inputted into Fuzzy for clustered. The result shown in Fig.2.

Table 2..*Clustered SVM-Fuzzy 2009*

Month	Data	Number of Events					Prediction		Acc (%)
		Clear Sky	Rain				T	F	
			Light	Moderate	Heavy	Very Heavy			
Jan	Act	9	3	12	7	0	17	14	
	Pred.	10	11	9	1	0			
Feb	Act	10	7	6	3	2	21	7	
	Pred.	9	11	6	2	0			
Mar	Act	25	4	2	0	0	24	7	
	Pred.	22	9	0	0	0			
April	Act	22	5	3	0	0	25	5	
	Pred.	23	7	0	0	0			
May	Act	23	5	3	0	0	24	7	
	Pred.	27	4	0	0	0			
June	Act	29	0	1	0	0	29	1	
	Pred.	30	0	0	0	0			
July	Act	29	2	0	0	0	25	6	
	Pred.	26	5	0	0	0			
August	Act	31	0	0	0	0	31	0	
	Pred.	31	0	0	0	0			
Sept	Act	27	1	2	0	0	27	3	
	Pred.	30	0	0	0	0			
Oct	Act	29	1	1	0	0	29	2	
	Pred.	29	2	0	0	0			
Nov	Act	26	2	1	1	0	24	6	
	Pred.	29	1	0	0	0			
Des	Act	18	7	2	4	0	15	16	
	Pred.	14	14	3	0	0			
Result							291	74	80

.

III. COMPARISON OF VARIOUS ALGORITHMS

There are two main approach in forecasting, quantitative and qualitative measure. In this paper, four quantitative approach are selected, i.e. hybrid Genetic Algorithm-Neural Network (GA-NN), Wavelet-Neural Network (W-NN), ARIMA and Adaptive Splines Threshold Autoregression (ASTAR). The performance of each algorithm are compared as linear validation curve and alsom from RMSE value.

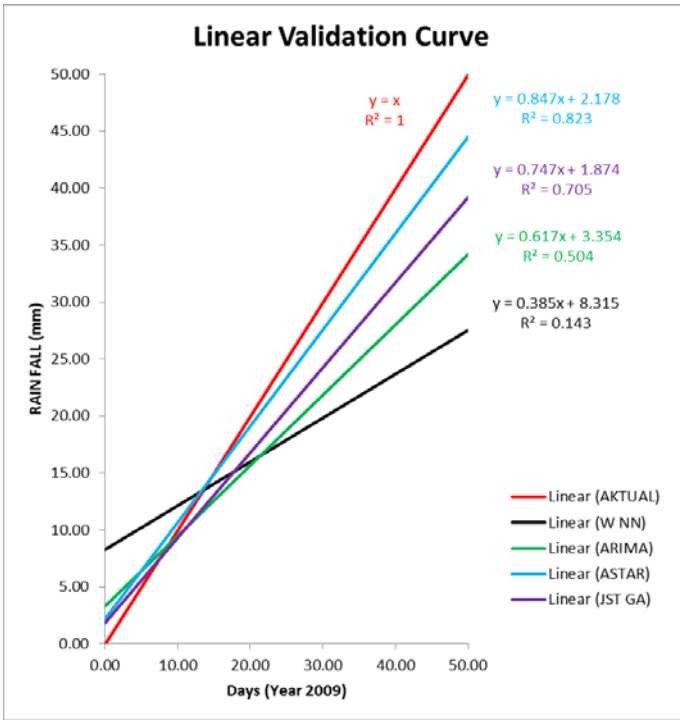

Fig.6 Linear Validation Curve

Result shows from Fig 6, that ASTAR outperform other algorithms with a tight linearity to actual data. This is due to the fact that ASTAR is best in dealing non-stationary data like the rain gauge. Moreover, for conformity, the RMSE value for each algorithm can be seen in Table 3.

Table 3.Comparison of RMSE Value

Methods	RSME
Genetic Algorithm – Neural Network	0,0921
Wavelet Neural Network	0,3219
ASTAR	0,0243
ARIMA	0,2392

It can be seen that ASTAR has lowest RMSE value compare to others. It seems that GA-NN has a close performance to ASTAR. A deeper review of Artificial Intelligence and Numerical Statistic Approach were discussed more in Indrabayu, et.al [9,10].

For Qualitative approach, three algorithm are used for comparison i.e. NN-Fuzzy, ANFIS and SVM-Fuzzy.

Unlike the previous approach, the performance of algorithms are measured on how well the system can clustered the forecast rainfall event to actual data.

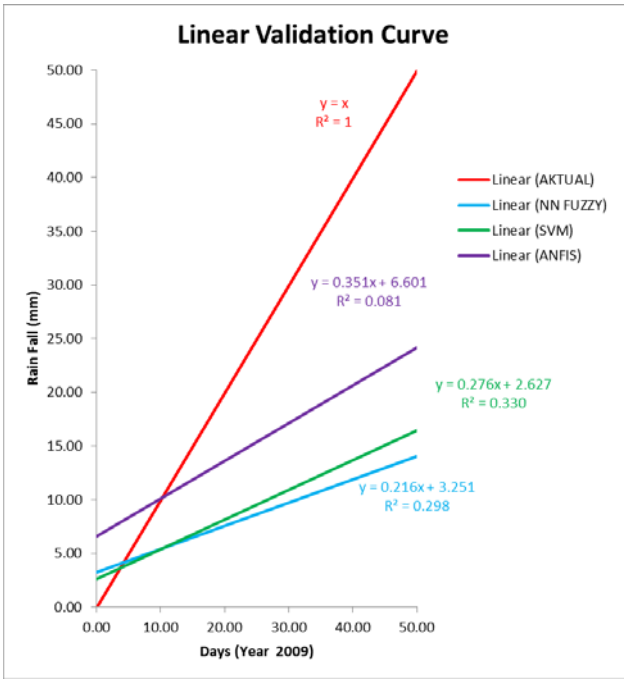

Fig. 7 Linear Validation Curve (Qualitative)

ANFIS shows the worst result compare to others.
In percentage accuracy point of view is shown in Table 4.

Methods	Accuracy %
ANFIS	65
Neural Network – *Fuzzy*	75,8
SVM – Fuzzy	80

.

IV. SAMPLING TIME OF DATA

BMKG has what so called "*taman alat*" which is a collection of weather monitoring instruments for each meteorological parameter. A Staff is assigned for each *taman alat* that spread out on several spot in the city. Daily data are recorded from averaging three times outlook of monitoring result.

From correlation check of parameter against rainfall event, only 3 parameters i.e. land surface temperature, humidity and wind speed showed better correlation property. Suspected variables like solar radiation and air pressure surprisingly have a very insignificant correlation to next day rainfall (less than 0.2). We believe that a smaller time sampling would give a better correlation and for sure would lead to better forecasting. A new radar was installed on Civil Engineering UNHAS by LAPAN (National Space and Flight Agency) in the late 2012. The radar is automated weather service monitoring which

SVM-Fuzzy has a better outfit to actual data in term of linearity.

required no conventional recorded from operator. The sampling resolution has been set to 15 minutes resolution. The result shows in fig 8 and 9.

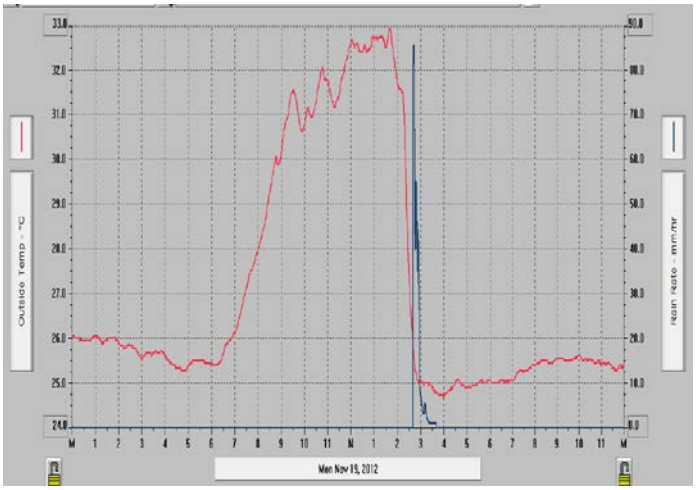

Fig 8. Overlay between Temperature and Rainfall

By overlaying both parameters, it can be found that temperature indicators shows a high fluctuation around 4-6 hours before rainfall event occurs. If smaller resolution time (2-3 hours) are used instead of averaging 1 day sampling time would give a better result in prediction. This is also true for wind speed and humidity. Fig 9 shows the relationship of parameters.

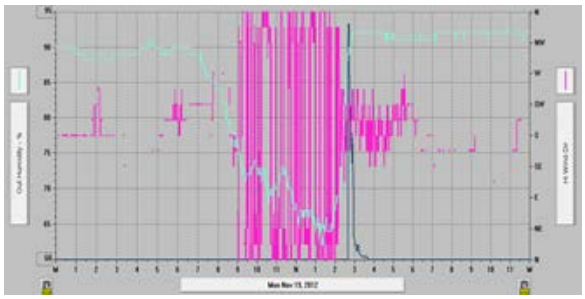

Fig. 9. Overlay Wind speed, Humidity, and Rainfall

Wind Gauge is represented by red lines and humidity is shown as light blue line. Wind Speed becomes a good indictor around 1-5 hours before rainfall event. A significant humidity declining are falling from 6 hours to the rainfall event. Again, the future research are so promising to gain a better accuracy with these new sampling time resolution.

V. CONCLUSION

Several powerful methods of data series predictor have been implemented in this research, both from AI and Statistical calculation. It seems ASTAR has shown better performance followed by hybrid AI (GA-NN) for quantitative based prediction. As qualitative based point of view, SVM-Fuzzy gives higher percentage of accuracy compare to NN-Fuzzy and ANFIS.

A smaller resolution sampling time should be implemented for all weather service to have a closer relationship among parameters. Future research will have not only smaller sampling time but also more suspected parameters that potentially linked to rainfall event.

REFERENCES

[1] Aldrian, E., Djamil, YS. Application of Multivariate Anfis for Daily Rainfall Prediction: Influences of Training Data Size. MAKARA, SAINS. Volume 12, No. 1, 2008.

[2] Hongxia, L., Chuanwei L. Construction and Application of Fuzzy Neural Network Model in Precipitation Forecast of Sanjiang Plain, China. International Conference on Wireless Communication, Networking and Mobile Computing (WiCOM), 2008.

[3] Gan, X. The research of rainfall prediction models based on Matlab neural network. Cloud Computing and Intelligence Systems (CCIS), 2011 IEEE.

[4] Kumar, A., Yang, F, Goddard L, Schubert, S., Differing Trends in the Tropical Surface Temperatures and Precipitation over Land and Oceans. Journal of Climate. Vol.17, 2004.

[5] Giannini, A., Robertson, W.A, Qian, J-H, A role for tropical tropospheric temperature adjustment to El Nin˜o–Southern Oscillation in the seasonality of monsoonal Indonesia precipitation predictability. Journal of Geophysical Research. VOL. 112, D16110, 2007.

[6] Edvin A. and Dwi S., "Identification Of Three Dominant Rainfall Regions Within Indonesia And Their Relationship To Sea Surface Temperature", International Journal Of ClimatologyInt. J. Climatol. 23: 1435–1452, 2003/

[7] I Sonjaya, T Kurniawan, "Uji Aplikasi Hybmg Versi 2.0 Untuk Prakiraan Curah Hujan Pola Monsunal Ekuatorial Dan Lokal", Buletin Meteorologi Klimatologi Dan Geofisika Vol. 5 No. 3 September 2009.

[8] Indrabayu, Nadjamuddin H, M. Saleh Pallu, and Andani A, A New Approach of Expert System For Rainfall Prediction Based On Data Series, International Journal of Engineering Research and Application (IJERA), Published Vol 3 Issue 2 april 2013 SSN 2248-9622

[9] Indrabayu, Nadjamuddin H, M. Saleh Pallu, and Andani A, Numerical Statistic Approach For Expert System In Rainfall Prediction Based On Data Series, International Journal of Computational Engineering Research (IJCER), Published Vol 3 Issue 4 April 2013, ISSN 2250-3005

[10] Indrabayu, Nadjamuddin Harun, M. Saleh Pallu, and Andani Achmad, Statistic Approach versus Artificial Intelligence for Rainfall Prediction Based on Data Series, International Journal of Engineering and Technology (IJETIY) 2013, Vol.5 No.2 May 2013. ISSN 0975-4024

AUTHORS

Indrabayu is with the Hasanuddin University, Informatics Department, Indonesia. Teaching Staff and also Head of Artificial Intelligence and Multimedia Laboratory. (e-mail: indrabayu@unhas.ac.id).

Dadang A. Sumiaharja, is with Department of Physic, Hasanuddin University. Teaching Staff (Professor) and Also Vice Rector of Academic affairs. (e-mail: vicerector1@unhas.ac.id).