# Comparative Study of Grammatical Evolution and Adaptive Neuro-Fuzzy Inference System on Rainfall Forecasting in Bandung

Fhira Nhita[1], Adiwijaya[2], Sheila Annisa[3], Sekar Kinasih[4]
School of Computing, Telkom University
Bandung 40257, Indonesia
fhiranhita@telkomuniversity.ac.id[1], adiwijaya@telkomuniversity.ac.id[2], lirasheilaannisa@gmail.com[3], sekarkinasihh@hotmail.com[4]

**Abstract – Rainfall is a very crucial weather parameter. The information on rainfall is also used for certain fields including farming, transportation, and flood early warning system. The significant fluctuation of rainfall in Bandung recently causes the difficulty in rainfall forecasting. The study analyzes and implements Soft Computing algorithm for rainfall forecasting in Bandung Regency. The algorithms belong to SC method are Fuzzy Logic, Neural Network, and Evolutionary Algorithms (EAs). The study compares the performance of the forecasting from two algorithms, Grammatical Evolution (GE) and Adaptive Neuro-Fuzzy Inference System (ANFIS). For GE algorithm, the comparison between two survivor selection methods is conducted, namely between generational replacement and steady state. The experiment used the rainfall data for Bandung regency obtained from Indonesian Agency for Meteorology Climatology and Geophysics (BMKG) for the current 10 years (2003-2012). The experiment shows the performance of forecasting result of 70.76% for GE that uses the generational replacement, 74.35% for GE that uses the steady state and 80% for ANFIS.**

***Keywords: rainfall, forecasting, Grammatical Evolution, Adaptive Neuro Fuzzy Inference System***

## I. INTRODUCTION

Rainfall is a very crucial weather parameter. Especially in Bandung, one of the cities that often floods although the city is in 900 meters above sea level. Also, the information on rainfall is needed in various fields like farming [5], transportation, etc. Therefore, an adaptive method that can forecast the rainfall data is essential. One of the methods that can be used is Soft Computing (SC).

SC is a group of methodologies that continues to grow, which is aimed at producing a system that has a tolerance towards the accuracy, uncertainty and partial truth. The algorithms belong to SC method are Fuzzy Logic, Neural Network [1], and Evolutionary Algorithms (EAs). SC method also combines two basic algorithms into a hybrid algorithm like Evolving Neural network and *Adaptive Neuro-Fuzzy Inference System* (ANFIS). Algorithm in SC has been widely used in solving various problems like the weather forecast that uses *Evolving Neural Network (ENN)* algorithm [2,3,5,7,9] dan *Evolving Fuzzy* [4] algorithms which produces performance under 70% respectively i.e. in rainfall forecasting for Kemayoran Jakarta area. Besides that, ANFIS algorithm in [8] has been used for weather forecasting in Dhaka compared with ARIMA models.

The study compares other SC algorithms to forecast the weather, especially rainfall in Bandung Regency, namely by using ANFIS and Grammatical Evolution (GE) that belongs to EAs algorithm. Both of these algorithms is often used in time-series forecasting among others in [8] and GE algorithms has been declared the best algorithm in forecasting [10].

## II. DATA AND METHODOLOGY

In this study, pre-processing is conducted first towards the data to be used before the step of learning GE and ANFIS is started. Fig.1 show the process diagram for the system to be built.

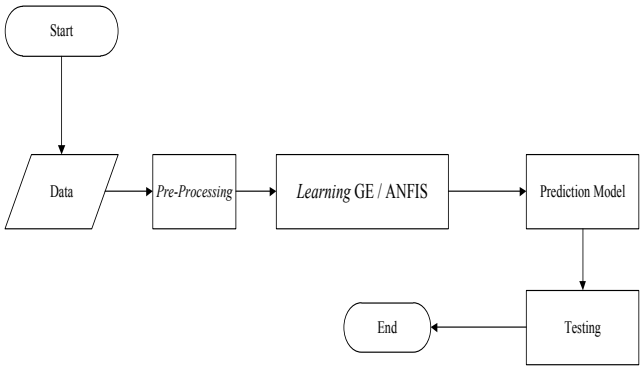


Fig. 1. System Block Diagram

### A. Data

Data used in the study is the rainfall data for Bandung regency obtained from Indonesian Agency for Meteorology Climatology and Geophysics (BMKG) for the current 10 years (2003-2012). The study uses input of data collected several months before the forecasting time (M, M-1, M-2, M-3, etc). The following is the pattern of rainfall for Bandung regency as depicted in the graphs in fig.2:

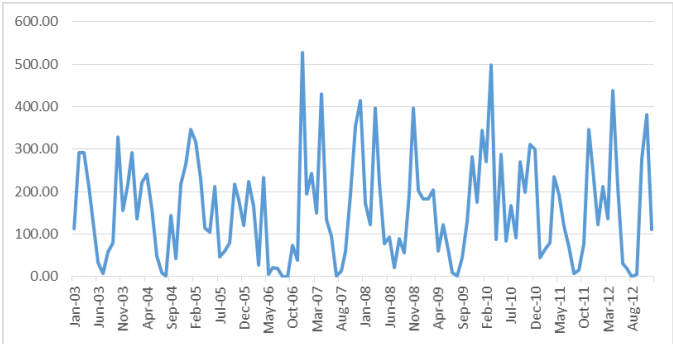


Fig. 2. Sample Data

### B. Pre-processing

#### 1) Moving Average

Moving Average is the method of forecasting the average score by taking the score of a group of research which is then proceeded as to find the average, which is then taken as the forecast for the following period. In this research, the use of moving average is aimed at finding the data pattern as well as eliminating the possibility of empty score for monthly rainfall data. The following is the formula for moving average [6]:

$$T_t = \frac{1}{k}\sum_{j=-m}^{m} Y_{t+j} \tag{1}$$

Where :
$T_t$ = Moving average result value with $Y$= year/record data;
$k$ = period of Moving Average which $k$ is an odd integer decided on observation ; and $m=(k-1)/2$

#### 2) Data Normalization

Normalization is altering the actual data to be *range* value with the interval of [0..1] which is aimed at minimizing the *error* score. After passing through the *moving average* step, the data are ready to be normalized. The following is the formula for normalization [4]:

$$X'_n = \left(\frac{X_n - Min_x}{Max_x - Min_x} \times 0.8\right) + 0.1 \tag{2}$$

Where :
$X'_n$ = result of normalization
$X_n$ = Actual data
$Min_x$ = minimum data for x variable
$Max_x$ = maximum data for x variable

#### 3) Data Partition

In this step, data will be divided into two parts, namely training and testing data. Proportion of data for training and testing is 70%: 30%. The grouping for the training and testing data is as follow:

- Training data : 84 data ( January 2003- December 2009)
- Testing data: 36 data ( January 2010 – December 2012)

### C. GE and ANFIS Training Process

#### 1) GE Training Process

After the preprocessing toward the data, the training process is conducted separately for GE and ANFIS algorithm as depicted in fig.3.

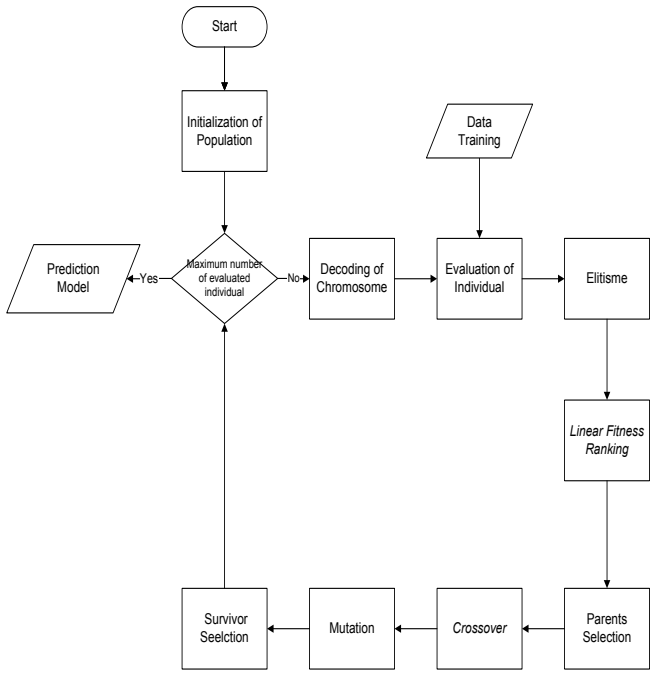


Fig. 3. GE Training Process

*2) ANFIS Training Process*

The training process towards ANFIS conducted through two phases namely forward and backward phases [10]. The forward phase will generate premise parameter score in the first iteration in the forward step and improve the consequent parameter in the fourth layer by using *Least Square Estimator* (LSE). While backward phase will repropagate forward score through *Backpropagation* by using *Gradient Descent* as to improve the premise parameter in the first layer. Fig.4 describes the training process on ANFIS.

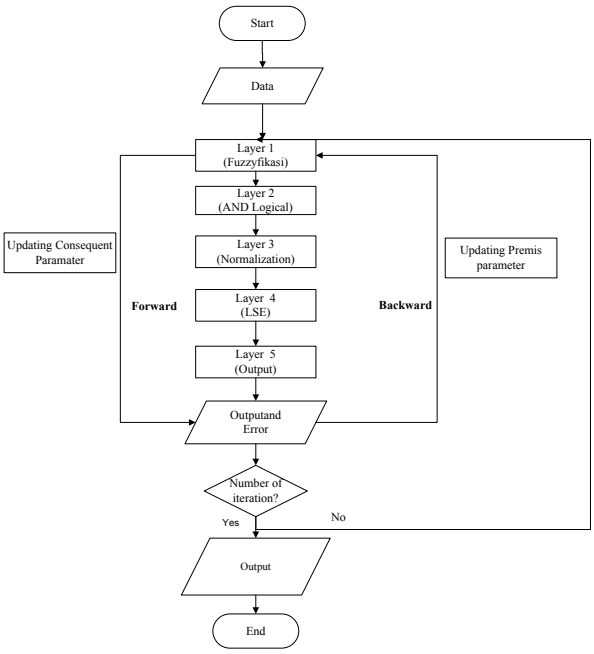


Fig. 4. ANFIS Training Process

## D. GE and ANFIS Testing Process

*1) GE Testing Process*

This phase is conducted to test the best forecasting function gained from the training process which is gained from GE training phase. The result of this process is optimum forecasted and the performance of the algorithm in forecasting rainfall for Month+1. Fig.5 illustrates the testing process in GE:

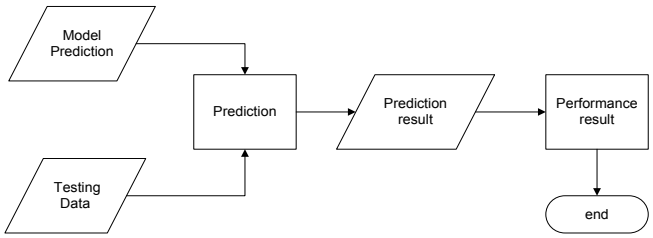


Fig. 5. GE Testing Process

*2) ANFIS Testing Process*

The testing process is conducted after premise parameter and consequent is obtained from training process. The training process only applies forward phase on ANFIS. The testing process is conducted in order to find out how well the established system build in the training process is as to give forecasted result which is close to the real score. The following illustrates the testing in ANFIS in Fig.6 :

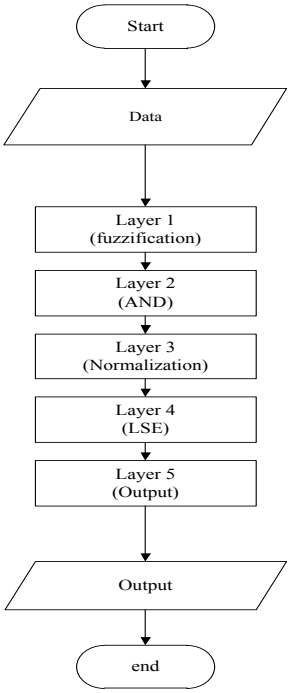


Fig. 6. ANFIS Testing Process

# III. EXPERIMENT SCENARIO AND RESULT ANALYSIS

## A. Grammatical Evolution

In this phase, the testing is conducted towards the use of BNF and the combination of evolution parameter. BNF is defined as the characteristics of data forecasting of *time series*, the pattern which is possibly drawn shows the addition of correlation of the rainfall data (stated in x variable) or the errors in the rainfall forecasting (stated in y variable). GE algorithm has been used in [2] that compared six ensemble forecast grammars or BNF. The sample of BNF Grammar used in the experiment describes in Fig.7.

```
N=          {expr, op, expr}
T=          {sin, cos, +, -, *, /, 0,1, 0,2 ,0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 2,
3, x1, x2, x3, x4, y1, y2, y3, y4}
S=          {expr}
1) <expr>:=          <expr><op><expr>   (A)
                     (<expr><op><expr>) (B)
                     <pre_op>(<expr>          (C)
                     <var>                    (D)
                     <const>                  (E)
2) <pre_op>:=        Sin                      (A)
                     Cos                      (B)
3) <op>::=+                           (A)
                     -                        (B)
                     *                        (C)
                     /                        (D)
4) <const>::=        0, 1     (A)     0, 6    (F)
                     0, 2     (B)     0, 7    (G)
                     0, 3     (C)     0, 8    (H)
                     0, 4     (D)     0, 9    (I)
                     0, 5     (E)     2       (J)
5) <var>::=          x1       (A)     y1      (E)
```

Fig. 7. BNF Grammar Samples

By using BNF above, several testing scenarios are conducted with two types of survivor selections, namely *Generational Replacement* and *Steady State*. the evolution parameters to be tested re crossover probability (Pc) and mutation probability (Pm) with the population size of 50 and 100 as can be seen in the table I below :

TABLE I. EVOLUTION PARAMETERS

| No | Population Size (PopSize) | Probability of Crossover (PC) | Probability of Mutation (PM) |
|---|---|---|---|
| 1 | 50 | 0,9 | 0,3 |
| 2 | | | 0,1 |
| 3 | | 0,7 | 0,3 |
| 4 | | | 0,1 |
| 5 | 100 | 0,9 | 0,3 |
| 6 | | | 0,1 |
| 7 | | 0,7 | 0,3 |
| 8 | | | 0,1 |

By using the two types of survivor selection, *generational replacement* and *steady state,* the following best scenario combination is obtained as can be seen in table II below :

TABLE II. THE BEST EVOLUTION PARAMETERS

| Survivor selection type | Pop Size | PC | PM | The Optimal Forecasting Model/Function |
|---|---|---|---|---|
| *Generational Replacement* | 50 | 0,7 | 0,3 | sin ( ( x4 + 0,7 - cos ( cos ( sin ( 2 * sin ( x2 ) - sin ( y4 ) ) - y4 ) ) ) ) / cos ( 0.1 ) |
| *Steady State* | 100 | 0,9 | 0,1 | x4 / ( 0,5 + x3 ) |

From the best scenario above, the performance of *generational replacement* is reached for 70.67% and for *steady state* is reached for 74.35%

### *B. Adaptive Neuro Fuzzy Inference System*

In this training process, several combinations are used to get the best architecture with the lowest *error*. The following table III shows the combination used:

TABLE III. EXPERIMENT SCENARIO

| Scenario | Input | *Fuzzy Set* | *Rule* | Iteration | *Learning Rate* |
|---|---|---|---|---|---|
| 1 - 36 | 2 | [2, 3, 4] | [min, max] | [50, 70, 100] | [0.1 , 0.01] |
| 37 - 72 | 3 | | | | |
| 73 - 108 | 4 | | | | |

After 108 scenarios are conducted based on the above table, the best result is obtained with the performance result of 80% on the testing or forecasting process. The following is the best architecture gained from various combination used in Fig.8 :

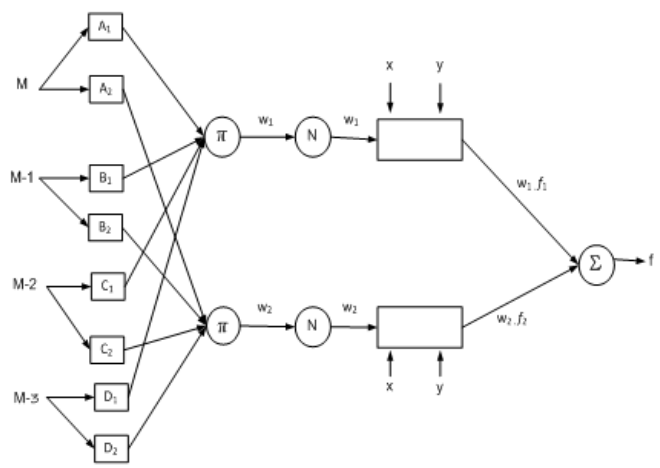


Fig. 8. Best Architecture of ANFIS

This architecture uses four *inputs* namely M, M-1, M-2, M-3, M-4 on input node which means the rainfall data monthly, then on the first layer, two nodes in every inputs which means it uses 2 linguistic scores, the second layer has mínimum *rule* (*half-connected*) which means that rule = the total linguistic score. The training process uses *learning rate* of 0.01 and iteration of 50 times.

From the experimental results obtained the best results on GE steady state is 74.35%, while the performance of ANFIS obtained result is 80%. This is partly due to the need for a deeper analysis to determine the BNF Grammar used on GE, because this will affect the prediction models.

## IV. CONCLUSION

Methods in Soft Computing areas has been widely used in solving various problems especially for data time series forecasting. This study compares two algorithms on Soft Computing, namely Grammatical Evolution and Adaptive Neuro-Fuzzy Inference System. The experiment shows the best performance of 70.76% for GE that uses survivor generational replacement selection, 74.35% for GE that uses survivor steady state selection and 80% for ANFIS in rainfall forecasting for next month based on the rainfall data for Bandung Regency. For future research is necessary to develop the methods to obtain better BNF Grammar in order to obtain better performance especially for GE algorithm.

## ACKNOWLEDGMENT

The authors would like to thank Telkom University for financial supporting this research. The authors would like to thank Indonesian Agency for Meteorology, Climatology and Geophysics (BMKG) for sharing information and supporting data.

## REFERENCES

[1] Adiwijaya, U.N. Wisesty, and F.Nhita, “Some Line Search Techniques on the Modified Backpropagation for Forecasting of Weather Data in Indonesia”, the Far East Journal of Mathematical Sciences 86:2 pp. 139-148, 2014.

[2] Dufek, Augusto, Dias, and Barbosa, “Evaluating the feasibility of grammar-based GP in combining meteorological forecast models”, IEEE Congress on Evolutionary Computation (CEC) 2013.

[3] F. Nhita, Adiwijaya, and Wisety U.N, “A Forecasting of Weather Data in Indonesia Using Evolving Neural Network based on Genetic Algorithm”, The Second International Conference on Technological Advances in Electrical, Electronic and Computer Engineering (TAEECE 2014), pp. 78-82, March 2014.

[4] F.Nhita and Adiwijaya, “A Rainfall Forecasting using Fuzzy System based on Genetic Algorithm”, International Conference on Information and Communication Technology (ICOICT 2013).

[5] F.Nhita, Adiwijaya, and U.N. Wisesty, “Planting Calendar Forecasting System using Evolving Neural Network”, the Far East Journal of Electronics and Communications 14:2 pp. 81-92, 2015.

[6] Makridakis, S., Wheelwright, S.C., & Hyndman, R.J.: Forecasting: Methods and Applications. John Wiley & Sons, Inc, United States of America (1998)

[7] Nurcahyo Septian, F.Nhita and Adiwijaya, “Rainfall Prediction in Kemayoran Jakarta Using hybrid Genetic Algorithm (GA) and Partially Connected Feedforward Neural Network (PCFNN)”, The Second International Conference on Information and Communication Technology (ICOICT 2014).

[8] Rahman, Islam, and Nadvi, “Comparative Study of ANFIS and ARIMA model for weather forecasting in Dhaka”, International Conference on Informatics, Electronics & Vision (ICIEV) 2013.

[9] Rismala Rita, Liong The Houw, Ardiyanti Arie (2013). “Prediction of Malaria Incidence in Banggai Regency Using Evolving Neural Network”, International Conference on Technology, Informatics, Management, Engineering & Environment (TIME-E 2013).

[10] Suyanto. 2008. Soft Computing: Membangun Mesin Ber-IQ Tinggi. Bandung: Informatika.