Rainfall prediction using fuzzy inference system for preliminary micro-hydro power plant planning

B Suprapti¹, R Malani*¹, J Minardi²

¹Department of Information Technology, State Polytechnic of Samarinda, Samarinda, East Kalimantan, Indonesia
²Department of Information System, Faculty of Science and Technology, Nahdlatul Ulama of Islamic University, Jepara, Central Java, Indonesia

*Corresponding Author : anaogie@gmail.com

Abstract. East Kalimantan is a very rich area with water sources, in the form of river streams that branch to the remote areas. The conditions of natural potency like this become alternative solution for area that has not been reached by the availability of electric energy from State Electricity Company. The river water in selected location (catchment area) which is channelled to the canal, pipeline or penstock can be used to drive the waterwheel or turbine. The amount of power obtained depends on the volume/water discharge and headwater (the effective height between the reservoir and the turbine). The water discharge is strongly influenced by the amount of rainfall. Rainfall is the amount of water falling on the flat surface for a certain period measured, in units of mm³, above the horizontal surface in the absence of evaporation, run-off and infiltration. In this study, the prediction of rainfall is done in the area of East Kalimantan which has 13 watersheds which, in principle, have the potential for the construction of Micro Hydro Power Plant. Rainfall time series data is modelled by using AR (Auto Regressive) Model based on FIS (Fuzzy Inference System). The FIS structure of the training results is then used to predict the next two years rainfall.

1. Introduction

Hydropower is a system that utilizes water flow energy to be converted into electrical energy or mechanical energy. There are various ways to harness the energy of the water flow, one of which is the run of river system that does not require a large reservoir. This system is known as Microhydro. The power plant system using the microhydro is called as the Micro Hydro Power Plant (MHPP). MHPP is an alternative renewable energy system that meets energy conservation policies, especially for areas that have great potential for water energy sources.

East Kalimantan is a very rich area with water sources, in the form of river streams that branch to the remote areas. The conditions of natural potency like this become alternative solution for the area that has not been reached by the availability of electric energy from State Electricity Company. The river water in the selected location (catchment area) which is channelled to the canal, pipeline or penstock can be used to drive the waterwheel or turbine. The amount of power obtained by the water channelling method depends on the volume/water discharge and headwater (the effective height between the reservoir and the turbine). The water discharge (m³/s) is determined by the flow rate of the river water (m/s) and the cross-sectional area of the river (m).
Water discharge is strongly influenced by the amount of rainfall. Rainfall is the amount of water falling on the flat surface for a certain period measured, in units of mm³, above the horizontal surface in the absence of evaporation, run-off and infiltration. Measurements of instantaneous water discharge at the watershed location where the dam will be built are conducted to determine the smallest water discharge during the dry season through hydrological studies. Prediction of rainfall becomes very important for the initial planning of MHPP development.

In this study, the prediction of rainfall is done in the area of East Kalimantan which has 13 watersheds, which in principle has the potential for the construction of MHPP. Rainfall time series data is modeled using AR (Auto Regressive) Model based on FIS. Training is done to obtain the best FIS structure. The FIS structure of the training results is then used to predict the next two years rainfall.

2. Method

The number of rainy days in a month depends on the amount of rainfall (mm). The number of rainy days is what affects the condition of river water discharge. Rainfall or precipitation is an element of the hydrometer, which is a collection of liquid or solid particles that fall or float in the atmosphere resulting from the condensation process of water vapor (cloud) [1]. Rainfall data can be categorized as time series data with seasonal and non stationary patterns.

2.1 Fuzzy Logic

Fuzzy logic provides an inference structure that enables appropriate human reasoning capabilities. The theory of fuzzy logic provides an inference mechanism under cognitive uncertainty. The theory of fuzzy logic is based upon the notion of relative graded membership and so are the functions of mentation and cognitive processes. The utility of fuzzy sets lies in their ability to model uncertain or ambiguous data.

Fuzziness in a fuzzy set is characterized by its membership functions. It classifies the element in the set, whether it is discrete or continuous. The membership functions can also be formed by graphical representations. The graphical representations may include different shapes. There are certain restrictions regarding the shapes used. The rules formed to represent the fuzziness in an application are also fuzzy [2]. The “shape” of the membership function is an important criterion that has to be considered. There are many membership functions (MF) that can be used, one of which is Triangular MF as shown in Figure 1.

![Triangular MF](image1.png)
From Figure 1, fuzzy set is declared as

\[ \mu(x) = \text{trimf}(x, a, b, c) = \begin{cases} 
0 & x \leq a, x \geq c \\
\frac{x-a}{b-a} & a \leq x \leq b \\
\frac{c-x}{c-b} & b \leq x \leq c 
\end{cases} \]  

(1)

2.2 FIS (Fuzzy Inference System)

FIS is a system that processes the mapping formulation of a given input to produce an output using Fuzzy Logic. FIS is shown in Figure 2.

2.2.1 Fuzzify inputs. The first step is to take the inputs and determine the degree to which they belong to each of the appropriate fuzzy sets via membership functions. The input is always a crisp numerical value limited to the universe of discourse of the input variable and the output is a fuzzy degree of membership in the qualifying linguistic set. Fuzzification of the input is performed by considering either a table lookup or a function evaluation. In this way, by the rules, each input is fuzzified over all the qualifying membership functions required.

2.2.2 Applying Fuzzy Operator. After the inputs are fuzzified, the degree to which each part of the antecedent is satisfied for each rule. If the antecedent of a given rule has more than one part, the fuzzy operator is applied to obtain one number that represents the result of the antecedent for that rule. This number is then applied to the output function. The input to the fuzzy operator is two or more membership values from fuzzified input variables. The output is a single truth value. There are two commonly used fuzzy operators, that are OR (max) and AND (min). For example, there are two inputs (A and B) and one output (C) which have linguistic variables (Low, Medium, High) and have the same fuzzy set as shown in Figure 3. If \( A(x) = 20 \) and \( B(x) = 65 \) then for the rule:

\[ \text{IF } A \text{ is Low AND } B \text{ is Medium } \quad \text{THEN } C \text{ is Medium} \]  

(2)

it is illustrated in Figure 4.
2.2.3. Applying Implication Method. Before applying the implication method, we must determine the rule's weight. Every rule has a weight (a number between 0 and 1), which is applied to the number given by the antecedent. Generally, this weight is 1 (as it is for this example) and thus has no effect at all on the implication process. After proper weighting has been assigned to each rule, the implication method is implemented. A consequent is a fuzzy set represented by a membership function, which appropriately weights the linguistic characteristics that are attributed to it. The consequent is reshaped using a function associated with the antecedent (a single number). The input for the implication process is a single number given by the antecedent, and the output is a fuzzy set. Implication is implemented for each rule. For implication operation it is used AND method as shown in Figure 5.

2.2.4. Aggregate all outputs. Because decisions are based on the testing of all of the rules in a FIS, the rules must be combined in some manner in order to make a decision. Aggregation is the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set. Aggregation only occurs once for each output variable, just prior to the fifth and final step, defuzzification. The input of the aggregation process is the list of truncated output functions returned by the implication process for each rule. The output of the aggregation process is one fuzzy set for each output variable. For Fuzzy Mamdani, the aggregation stage is using OR (max) method. For example, it is shown in Figure 6.
2.2.5. Defuzzify. The input for the defuzzification process is a fuzzy set (the aggregate output fuzzy set) and the output is a single number. As much as fuzziness helps the rule evaluation during the intermediate steps, the final desired output for each variable is generally a single number. However, the aggregate of a fuzzy set encompasses a range of output values, and so must be defuzzified in order to resolve a single output value from the set. There are various methods of defuzzification, one of the most popular is the centroid method which returns the center of area under the curve. For a Fuzzy Set \( X \) with finite universe of discourse, the centroid method is expressed by:

\[
x_c = \frac{\sum_{i=1}^{n} x_i \mu(x_i)}{\sum_{i=1}^{n} \mu(x_i)}
\]

where \( x_c \) is the centroid value, \( x_i \) is \( x \) value of \( i \)th point correspond to \( \mu(x_i) \) from the aggregation result, \( \mu(x_i) \) is the degree of membership (fuzzy number) of \( x_i \), \( n \) is the number of \( (x_i, \mu(x_i)) \) pairs in the curve.

2.2.6. AR Model. AR model is declared as [4]:

\[
y(t) = -a_1 y(t-1) - \ldots - a_n y(t-n) + e(t)
\]

where \( a_i \) is a constant, \( n \) is order number of system, and \( e(t) \) is a white noise. Equation (4) can be decomposed into the following equation [4]:

\[
y(t) = (a_1 q^{-1} + \ldots + a_n q^{-n}) y(t) + e(t)
\]

\[
= - \left( \sum_{i=1}^{n} a_i q^{-i} \right) y(t) + e(t) = A(q^{-1}) y(t) + e(t)
\]

where \( q^{-1} \) is the delay operator, and \( A(q^{-1}) \) is polynomial to be estimated.

2.2.7. Data Labeling. For data modeling it is needed to use AR model based on fuzzy approach. Hence, it is required labeling data. The range of the universe of discourse is divided into several intervals where each interval represents a linguistic variable. For example, the range of the universe of discourse is \([0, 100]\) being divided into three intervals, i.e.: \([0, 33]\), \([33, 66]\), and \([66, 100]\). Each interval is labeled sequentially so that each interval corresponds to \([0, 33]\), \([33, 66]\), \([66, 100]\), or \([0,100]\). Data labeling is used to build Rule-Based. For example, a fuzzy set that uses Triangular MF has a range of universe of discourse \([0, 100]\) with 3 (three) linguistic variables (Low, Medium, High). Low is in the range \([0, 50]\), Medium: \([0, 50, 100]\), and High: \([50, 100, 100]\). The examples of data labeling are as follows:

\[
\begin{bmatrix}
65 & 25 & 60 \\
75 & 60 & 50
\end{bmatrix}
\Rightarrow
\begin{bmatrix}
2 & 1 & 2 \\
3 & 2 & 2
\end{bmatrix}
\Rightarrow
IF A is Medium (2) AND B is Low (1) THEN C is Medium (2)
\Rightarrow
\begin{array}{c}
\text{time series data} \\
\text{data labeling} \\
\text{Rule Based}
\end{array}
\]
The algorithm of prediction using AR model based on FIS is shown in Figure 7.

![Figure 7. The algorithm of prediction using AR model based on FIS](image)

2.2.8. Data Time Series Modeling. The monthly rainfall data are obtained from Statistics of Samarinda Municipality [1] in period 2006 – 2016 as shown in Table 1.

| Year | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 2006 | 228 | 207 | 215 | 207 | 307 | 185 | 24  | 98  | 108 | 70  | 191 | 110 |
| 2007 | 307 | 220 | 260 | 340 | 112 | 213 | 279 | 133 | 183 | 181 | 85  | 141 |
| 2008 | 143 | 194 | 211 | 259 | 51  | 205 | 333 | 149 | 153 | 208 | 501 | 350 |
| 2009 | 164 | 196 | 279 | 309 | 186 | 41  | 157 | 123 | 99  | 232 | 165 | 211 |
| 2010 | 148 | 162 | 157 | 164 | 223 | 320 | 259 | 144 | 202 | 235 | 207 | 224 |
| 2011 | 332 | 320 | 368 | 332 | 389 | 95  | 238 | 124 | 132 | 218 | 197 | 244 |
| 2012 | 330 | 206 | 257 | 371 | 128 | 172 | 147 | 140 | 110 | 116 | 228 | 200 |
| 2013 | 176 | 309 | 284 | 337 | 234 | 161 | 45  | 90  | 256 | 223 | 363 | 276 |
| 2014 | 273 | 197 | 319 | 126 | 190 | 211 | 50  | 81  | 82  | 111 | 300 | 449 |
| 2015 | 345 | 193 | 198 | 344 | 214 | 259 | 163 | 58  | 0   | 73  | 61  | 191 |
| 2016 | 159 | 99  | 318 | 369 | 225 | 202 | 163 | 99  | 226 | 175 | 292 | 357 |

By using 2nd order AR model then it can be stated:

\[ y(t) = -a_1 y(t-1) - a_2 y(t-2) + e(t) \]

By using FIS then it can be stated:

\[ y(t) = \text{Fuzzy}(y(t-1), y(t-2)) + e(t) \]

where \( y(t-2) \) is rainfall data from year 2006 – 2014, \( y(t-1) \) is rainfall data from year 2007 – 2015 as training input data, and \( y(t) \) is rainfall data from 2008 – 2016 as training target. By training Fuzzy \( (g) \) such that \( e(t) \rightarrow 0 \) then obtain Fuzzy \( (g) \rightarrow y(t) \).
In this study, the FIS uses 5 linguistic variables for initialization and MAPE target \( e(t) = 7\% \). The performance function of training result using MAPE (Mean Absolute Error) which is declared as:

\[
APE_i = \frac{\text{abs}(y(t_i) - y_{\text{fuzzy}}(t_i))}{y(t_i)} \times 100\%
\]

\[
\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} APE_i
\]

where \( N \) is the number of training data, \( y(t_i) \) is the \( i \)th training target, \( y_{\text{fuzzy}}(t_i) \) is the \( i \)th output of FIS. Implementation of FIS is done by using Matlab programming tool.

All the training data are organized into sequential form by year then by month so that there will be data. Because using 2nd order AR model then each variable will consist of 130 data. By using 5 linguistic variables for initialization then the interval width of each data label is expressed by:

\[
\text{interval} = \frac{\text{max}(\text{data})}{(\text{Lvar} - 1)}
\]

where \( \text{Lvar} \) is the number of linguistic variable for initialization. By using the rainfall data as shown in Table 1 so that: \( \text{interval} = \frac{\text{max}(\text{data})}{(5 - 1)} = 125.25 \). From labeling of data it is obtained the number of rule-based: \( 130 + 5 = 135 \). The training result is shown in Figure 8, and the prediction result for next 2 years is shown in Figure 9.
3. Result and Discussion
Summary of the training process is shown in Table 2. The training performance by MAPE is shown in Figure 10. From the training results obtained MAPE 6.28% achieved with 300 linguistic variables. The final interval width is 3.05 with the number of linguistic variables is 300. Table 2 and Figure 10 show that changes in the number of linguistic variables affect MAPE changes. Changes in the number of linguistic variables affect the change in interval width of each data label and number of rule-based.

To predict the next two years (2x12 = 24 predictive data) it is done by means of each prediction result used to add rule-based in such a way that it is done until it reaches 24 predictive data.

Table 2. Summary of training process

| LVar | 5   | 25  | 45  | 65  | 85  | 105 | 125 | 145 | 165 |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Rule-Based | 135 | 155 | 175 | 195 | 215 | 240 | 260 | 280 | 300 |
| Interval Width | 125 | 20.88 | 11.39 | 7.83 | 5.96 | 4.82 | 4.04 | 3.48 | 3.05 |
| MAPE | 162 | 47.07 | 26.77 | 21.13 | 11.98 | 8.58 | 7.45 | 7.36 | 6.28 |

Figure 9. Prediction Result of Rainfall years 2017 - 2018

Figure 10. Training performance
4. Conclusion
The time series data prediction by using AR model based on FIS depends on the order number of AR model that determine the number of FIS inputs and with one FIS output. The interval width of each data label depends on the number of linguistic variables. The number of rule-based relies on the amount of training data already labeled and the number of linguistic variables. The change in the number of linguistic variables depends on the specified MAPE target.

Acknowledgment
The authors would like to express heartfelt thanks to The Modern Computing Research Center, Department of Information Technology, State Polytechnic of Samarinda for providing all their support.

References
[1]. Statistics of Samarinda Municipality, 2007-2017 Samarinda Municipality in Figures 2007-2017, Regional Development of Planning of Samarinda City, BPS-Statistics of Samarinda Municipality, ISSN : 0215-2398
[2]. S N Sivanandam 2007 Introduction to Fuzzy Logic using Matlab, Springer-Verlag Berlin Heidelberg
[3]. Peter J Brockwell, 2002 Introduction to Time series and Forecasting – Second Edition. Springer-Verlag New York, Inc. ISBN 0-387-95351-5
[4]. Ming Tao Chou, 2011 Long-term Predictive Value Interval with The Fuzzy Time Series, Journal of Marine Science and Technology, 19(5) pp. 509-513
[5]. I Aydin, M Karakose, E Akin 2009 The Prediction Algorithm Based on Fuzzy Logic Using Time Series Data Mining Method, World Academy of Science, Engineering and Technology
[6]. Ashwini Tale, Arvind Singh Gusain, and Jyotirmoy Baguli 2017 Study of Load Forecasting Techniques using Fuzzy Logic International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering 6(2) ISSN (print): 2320-3765, (online): 2278-8875