# Rainfall Prediction in the Northeast Region of Thailand using Cooperative Neuro-Fuzzy Technique

***Kajornrit J.***

*School of Information Technology, Murdoch University South Street, Murdoch, Western Australia, 6150*
*Email address: j_kajornrit@hotmail.com*

***Wong K. W.***

*School of Information Technology, Murdoch University South Street, Murdoch, Western Australia, 6150*
*Email address: k.wong@murdoch.edu.au*

***Fung C. C.***

*School of Information Technology, Murdoch University South Street, Murdoch, Western Australia, 6150*
*Email address: l.fung@murdoch.edu.au*

***Abstract***

*Accurate rainfall forecasting is a crucial task for reservoir operation and flood prevention because it can provide an extension of lead-time for flow forecasting. This study proposes two rainfall time series prediction models, the Single Fuzzy Inference System and the Modular Fuzzy Inference System, which use the concept of cooperative neuro-fuzzy technique. This case study is located in the northeast region of Thailand and the proposed models are evaluated by four monthly rainfall time series data. The experimental results showed that the proposed models could be a good alternative method to provide both accurate results and human-understandable prediction mechanism. Furthermore, this study found that when the number of training data was small, the proposed model provided better prediction accuracy than artificial neural networks.*

***Keywords**: Rainfall Prediction; Seasonal Time Series; Artificial Neural Networks; Fuzzy Inference System; Average-Based Interval.*

## 1 Introduction

Rainfall forecasting is indispensable for water management because it can provide an extension of lead-time for flow forecasting used in water strategic planning. This is especially important when it is used in reservoir operation and flood prevention. Usually, rainfall time series prediction has used conventional statistical models and Artificial Neural Networks (ANN) [8]. However, such models are difficult to be interpreted by human analysts, because the prediction mechanism is in parametric form. From a hydrologist's point of view, the accuracy of prediction and an understanding in the prediction mechanism are equally important.

Fuzzy Inference System (FIS) uses the process of mapping from a given set of inputs variables to outputs based on a set of human understandable fuzzy rules [19]. In the last decades, FIS has been successfully applied to various problems [3], [4]. An advantage of FIS is that its decision mechanism is interpretable. As fuzzy rules are closer to human reasoning, an analyst could understand how the model performs the prediction. If necessary, the analyst could also make use of his/her knowledge to modify the prediction model [5]. However, the disadvantage of FIS is its lack of learning ability from the given data. In contrast, an ANN is capable of adapting itself from training data. In many cases where human understanding in physical process is not clear, ANN has been used to learn the relationship between the observing data [6]. However, the disadvantage of ANN is its black-box nature, which is difficult to be interpreted. In order to combine the advantages of both models, this paper propose two rainfall time series prediction models, the Single Fuzzy Inference System (S-FIS) and the Modular Fuzzy Inference System (M-FIS), which use the concept of cooperative neuro-fuzzy technique.

This paper is organized as follows; Section 2 discusses the related works and Section 3 describes the case study area. Input identification and the proposed models are presented in Sections 4 and 5 respectively. Section 6 shows the experimental results. Finally, Section 7 provides the conclusion of this paper.

## 2 Soft Computing techniques in hydrological time series prediction

In the hydrological discipline, rainfall prediction is relatively difficult than other climate variables such as temperature. This is due to the highly stochastic nature in rainfall, which shows a lower degree of spatial and temporal variability. To address this challenge, ANN has been adopted in the past decades. For example, Coulibaly and Evora [7] compared six different ANNs to predict daily rainfall data. Among different types of ANN, they suggested that the Multilayer Perceptron, the Time-lagged Feedforward Network, and the Counter-propagation Fuzzy-Neural Network provided higher accuracy than the Generalized Radial Basis Function Network, the Recurrent Neural Network and the Time Delay Recurrent Neural Network. Another work was Wu et al. [8]. They proposed the use of data-driven models with data preprocessing techniques to predict precipitation data in daily and monthly scale. They proposed three preprocessing techniques, namely, Moving Average, Principle Component Analysis and Singular Spectrum Analysis to smoothen the time series data. Somvanshi et al. [1] confirmed in their work that ANN provided better accuracy than ARIMA model for daily rainfall time series prediction.

Time series prediction is not only used for rainfall data but also streamflow and rainfall-runoff modeling. Wang et al. [9] compared several computational models, namely, Auto-Regressive Moving Average (ARMA), ANN, Adaptive Neural-Fuzzy Inference System (ANFIS), Genetic Programming (GP) and Support Vector Machine (SVM) to predict monthly discharge time series. Their results indicated that ANFIS, GP and SVM have provided the best performance. Lohani [10] compared ANN, FIS and linear transfer model for daily rainfall-runoff model under different input domains. The results also showed that FIS outperformed linear model and ANN. Nayak et al. [11] and Kermani et al. [12] proposed the use of ANFIS model to river flow time series. In addition, Jain and Kumar [13] applied conventional preprocessing approaches (de-trended and de-seasonalized) to ANN for streamflow time series data.

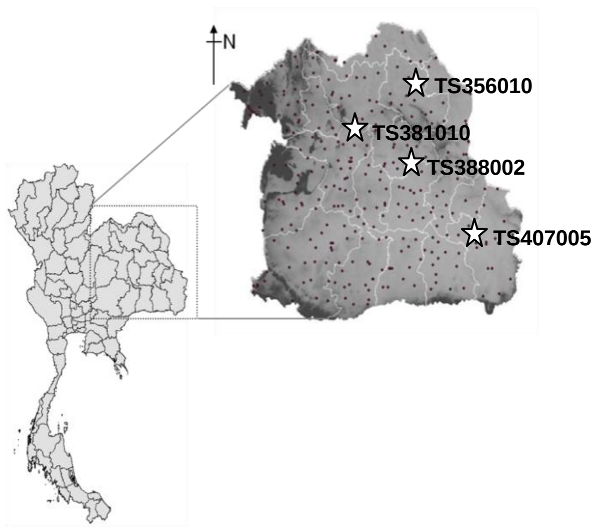



**Figure 1**: The case study area is located in the northeast region of Thailand. The positions of four rainfall stations are illustrated by star marks.

Up to this point, among all works mentioned, FIS itself has not been used as widely as ANN for time series prediction. Especially for rainfall time series prediction, reports on applications of FIS are limited. Thus, the primary aim of this study is to investigate an appropriate way to use FIS for rainfall time series prediction problem.

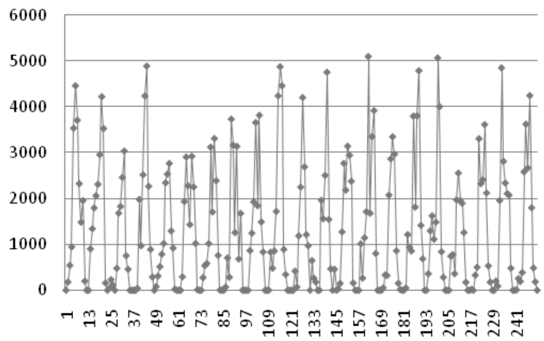
## 3 Case study area and data

The case study described in this study is located at the northeast region of Thailand (Figure 1). Four rainfall time series selected are depicted in Figure 2. Table 1 shows the statistics of the datasets used. The data from 1981 to 1998 were used to calibrate the models and data from 1999 to 2001 were used to validate the developed models. This study used the models to predict one step-ahead, that is, one month. To validate the models, Mean Absolute Error (MAE) is adopted as given in equation (1). The Coefficient of Fit (R) is also used to confirm the results. The performance of the proposed model is compared with conventional Box-Jenkins (BJ) models, Autoregressive (AR), Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) [1], [8], [10], [13] and [15].

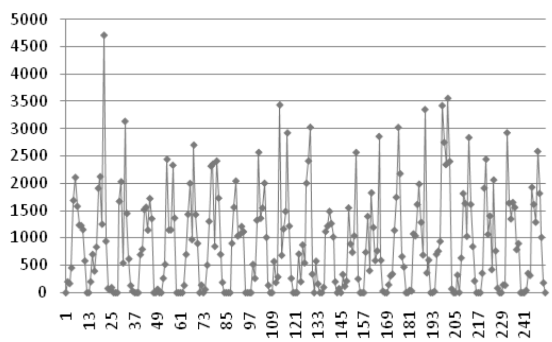
$$MAE = \sum_{i=1}^{m} |Oi - Pi| / m \qquad (1)$$

**Table 1**: DATASETS' STATISTICS

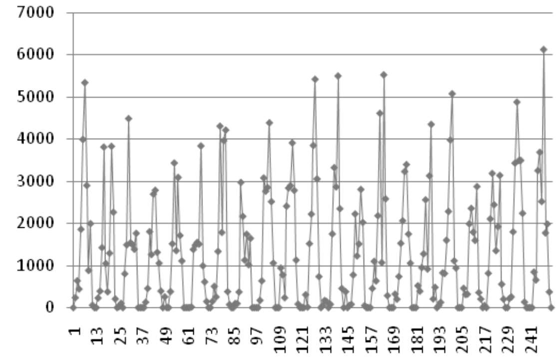
| Statistics | TS356010 | TS381010 | TS388002 | TS407005 |
|---|---|---|---|---|
| **Mean** | 1303.34 | 889.04 | 1286.28 | 1319.70 |
| **SD** | 1382.98 | 922.99 | 1425.88 | 1346.80 |
| **Kurtosis** | -0.10 | 0.808 | 0.532 | -0.224 |
| **Skewness** | 0.95 | 1.080 | 1.131 | 0.825 |
| **Minimum** | 0 | 0 | 0 | 0 |
| **Maximum** | 5099 | 4704 | 6117 | 5519 |
| **Latitude** | 104.13E | 102.88E | 104.05E | 104.75E |
| **Longitude** | 17.15N | 16.66N | 16.65N | 15.50N |
| **Altitude** | 176 | 164 | 155 | 129 |



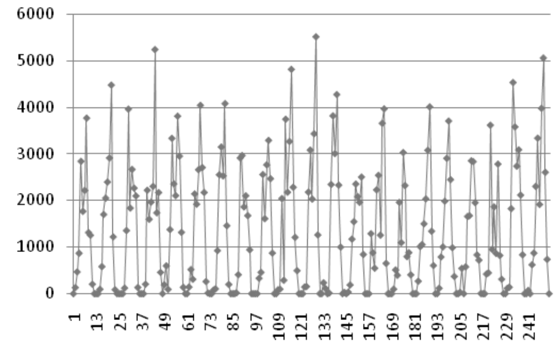

(TS356010)




(TS381010)




(TS388002)




(TS407005)

**Figure 2**: The four selected monthly rainfall time series used in this study.

## 4 Input Identification

In general, input of a time series model are normally based on previous data points (Lags). For BJ models, the analysis of autocorrelation function (ACF) and partial autocorrelation function (PACF) are used as a guide to identify the appropriate input. However, in the case of ANN or other related non-linear models, there was no theory to support the use of these functions [14]. Although some literatures addressed the applicability of ACF and PACF to non-linear models [15], other literatures preferred to conduct experiments to identify the appropriate input [11].

This study conducted an experiment to find an appropriate input based on data from five rainfall stations. Data from 1981 to 1995 were used for calibration and data from 1996 to 1998 were used for validation. By increasing the number of lags to ANNs, six different inputs models were prepared and tested. To predict $x_{(t)}$, first input model is $x_{(t-1)}$, second input model is $x_{(t-1)}$, $x_{(t-2)}$ and so on. Figure 3 shows the results from the experiment. In this figure, average normalized MAEs from five time series are illustrated in bold line. The results show that the MAE is the lowest at lag 5. The Five previous lags model is expected to be an appropriate input. Since increasing the number of input lags dose not significantly improve the prediction performance, additional methods may be needed.

In the case of seasonal data, there are other methods to identify an appropriate input to improve the prediction accuracy, for examples, using the Phase Space Reconstruction (PSR) [16] and adding time coefficient as a supplementary feature [2]. However, in the first method, large number of training data is needed. According to "*The Curse of Dimensionality*", when the number of input dimensions increases, the number of training data must be increased as well [17]. In this case study, the number of record is limited to 15 years, which could be considered as relatively small. Therefore it is more appropriate to add the time coefficient.

Time coefficient ($C_t$) was used to assist the model to scope prediction into specific period. It may be $C_t$ = 2 (wet and dry period), $C_t$ = 4 (winter spring summer and fall period), or $C_t$ = 12 (calendar months). This study adopted $C_t$ = 12 as supplementary features. In Figure 3, $C_t$ is added to original input data and test with ANNs (light line). The results show that using $C_t$ with 2 previous lags provided the lowest average MAE and it can improve the prediction performance up to 26% (dash line). So, the appropriate input used in this study should be rainfall from lag 1, lag 2 and $C_t$.

This experimental result is related to the work of Raman and Sunilkumar [18] who studied monthly inflow time series. In hydrological process, inflow is directly affected by rainfall, consequently, the characteristics of flow graph and rainfall graph are rather similar. They suggested using data from 2 previous lags to ANN models, however, instead of using a single ANN, they created twelve ANN models for each specific month and use "*month*" to select associated model to feed data in. If one considers this model as a black-box, one can see that their input is inflow from 2 previous lags and $C_t$ which relatively similar to this study

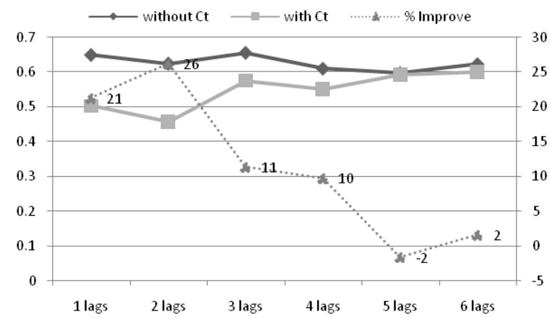



**Figure 3**: Average MAE measure of ANN models among different inputs.

## 5 The proposed models

This paper adopted the Mandani approach fuzzy inference system [20] since such model is more intuitive than the Sugeno approach [21]. To reduce the computational cost, triangular Membership Function (MF) is used. This study proposed two FIS models, namely, the Single Fuzzy Inference System (S-FIS) and the Modular Fuzzy Inference System (M-FIS), which use the concept of cooperative neuro-fuzzy technique. In S-FIS model, there is one single FIS model. Rainfall data from *lag 1*, *lag 2* and $C_t$ are feed directly in to the model. In M-FIS model, there are twelve FIS models associated to the calendar month. The $C_t$ is used to select associated model to feed in the rainfall data from *lag 1* and *lag 2*. The architectural overview of these two models is shown in the Figure 4.

Figure 5 shows the general steps to create these FIS models. The first step is to calculate the appropriate interval length between two consecutive MFs and then generate Mamdani FIS rule base model. At this step, *Average-Based Interval* is adopted. The second step is to create fuzzy rules. In this study, Back-Propagation Neural Networks (BPNN) is used to generalize from the training data and then used to extract fuzzy rules.

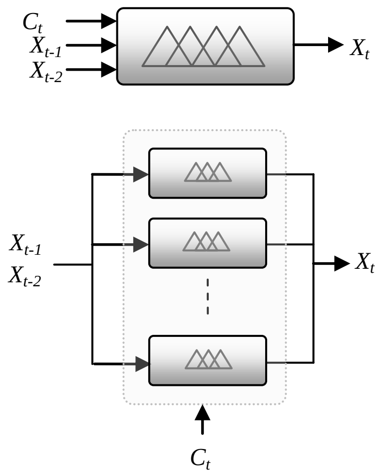



**Figure 4**: The architectural overview of the S-FIS (top) and M-FIS (bottom) models

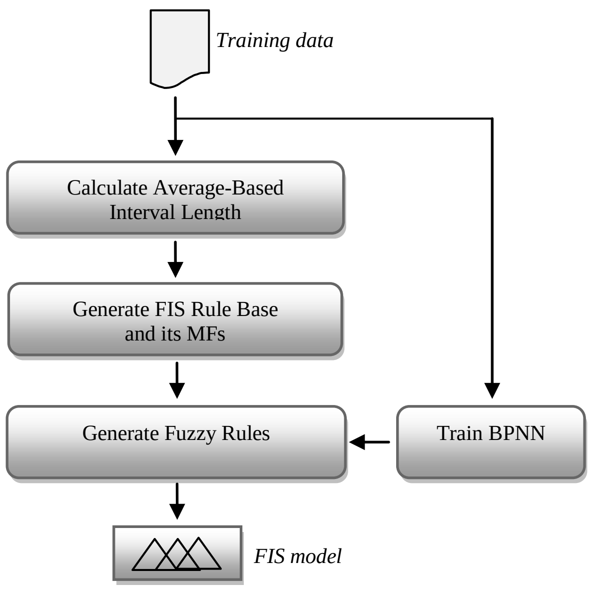



**Figure 5**: General steps to crate the S-FIS and M-FIS models

In the S-FIS model, the MFs of $C_t$ are simply depicted in Figure 6 (a). For rainfall input, interval length between two consecutive MFs is very important to be defined. When the length of the interval is too large, it may not be able to represent fluctuation in time series. On the other hand, when it is too small the objective of FIS will be diminished.

Huarng [22] proposed the A*verage-Based Interval* to define the appropriate interval length of MFs for fuzzy time series data based on the concept that "*at least half of the fluctuations in the time series are reflected by the effective length of interval*". The fluctuation in time series data is the absolute value of first difference of any two consecutive data. In this method, a half of the average value of all fluctuation in time series is defined as the interval length of consecutive two MFs. This method was successfully applied in the work reported in [23]. In this paper, this method is adapted a little bit more to fit to the nature of rainfall time series for this application.

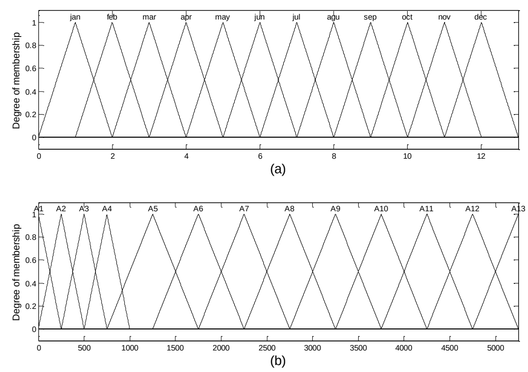



**Figure 6**: An example of membership functions in TS356010's S-FIS model, $C_t$ (a) and Rainfall (b)

Figure 6 (b) shows the rainfall's MFs of S-FIS from station TS356010. One can see that there are two interval lengths. The point that the interval length changes is around the 50 percentile of all the data. The data is separated into the lower area and the upper area by using 50 percentile as the boundary. Average-based intervals are calculated for both areas. Since the beginning and ending rainfall periods have smaller fluctuation than middle period, using smaller interval length is more appropriate [2]. In the M-FIS model, using two interval lengths is not necessary since each sub model is created according to the specific month.

As mentioned before, the drawback of FIS is the lack of learning ability from data. Such model needs experts or other supplementary procedure to help to create the fuzzy rules. In this study, the proposed methodology uses BPNN to learn the generalization features from the training data [5] and then is used to extract fuzzy rules. Once the BPNN was used to extract fuzzy rules, BPNN is not used anymore. The steps to create fuzzy rules are as follows:

*Step 1*: Training the BPNN with the training data. At this step, the BPNN is learned and generalized from the training data.

*Step 2*: Preparing the set of input data. The set of input data, in this case, are all the points in the input space where the degree of MF of FIS's input is 1 in all dimension. This input data are the premise part of the fuzzy rules.

*Step 3*: Feeding the input data into the BPNN, the output of BPNN are mapped to the nearest MF of FIS's output. This output data are consequence part of the fuzzy rule.

For example, considering the MFs in Figure 6, the input-output [3, 500, 750:1700] is replaced with fuzzy rule "*IF* $C_t$*=Mar and Lag1=A3 and Lag2=A4 THEN Predicted=A6"*. This step uses 1 hidden layer BPNN. The number of hidden nodes and input nodes are 3 for S-FIS and 2 for M-FIS.

## 6 Experimental results

The experimental results are shown in Table 2 and Table 3. In the tables, S-ANN and M-ANN are the neural networks used to create fuzzy rules for S-FIS and M-FIS respectively. In fact, the S-ANN and M-ANN themselves are also the prediction models. The performance between S-ANN and S-FIS is quite similar. It can be noted that the conversion from ANN-based to FIS-based does not reduce the prediction performance of the ANN. However, this conversion improves the S-ANN model from a qualitative point of view since M-FIS is interpretable with a set of human understandable fuzzy rules. The interesting point is the performance between M-ANN and M-FIS. This conversion can improve the performance of M-ANN.

Next, the proposed models have been compared with three conventional BJ models. The comparison results are depicted in Figure 7. Since the results from MAE and R measures are consolidated, these experimental results are rather consistent. Similar to the work by Raman and Sunilkumar [18], the AR model uses degree 2 because it uses the same input as the proposed models. The ARIMA and SARIMA models used in the study are automatically generated and optimized by statistical software. However, these generated models were also rechecked to ensure that they provided the best accuracy.

**Table 2**: MAE measure of validation period

| Datasets | S-ANN | S-FIS | M-ANN | M-FIS | AR | ARIMA | SARIMA |
|---|---|---|---|---|---|---|---|
| **TS356010** | 450.99 | 447.56 | 560.44 | 496.35 | 747.37 | 747.01 | 538.99 |
| **TS381010** | 332.71 | 343.88 | 439.91 | 442.32 | 534.32 | 402.42 | 503.99 |
| **TS388002** | 736.70 | 725.39 | 811.99 | 639.29 | 912.64 | 856.88 | 714.74 |
| **TS407005** | 636.37 | 634.65 | 776.63 | 661.30 | 901.76 | 672.35 | 799.34 |

**Table 3**: R measure of validation period

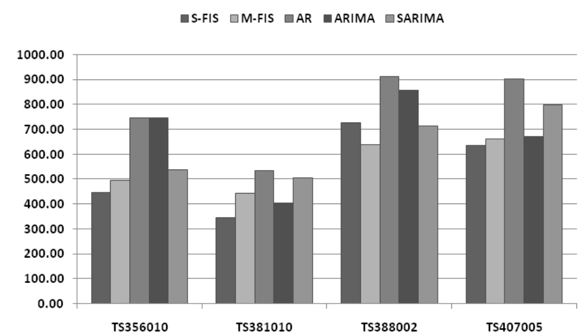
| Datasets | S-ANN | S-FIS | M-ANN | M-FIS | AR | ARIMA | SARIMA |
|---|---|---|---|---|---|---|---|
| **TS356010** | 0.884 | 0.887 | 0.755 | 0.850 | 0.650 | 0.759 | 0.837 |
| **TS381010** | 0.719 | 0.709 | 0.606 | 0.668 | 0.464 | 0.733 | 0.575 |
| **TS388002** | 0.760 | 0.773 | 0.712 | 0.871 | 0.606 | 0.685 | 0.769 |
| **TS407005** | 0.768 | 0.770 | 0.633 | 0.736 | 0.594 | 0.755 | 0.681 |

In term of MAE, among the three BJ models, the AR model provided the lowest accuracy in all datasets. ARIMA show higher accuracy than SARIMA in two of the datasets. In station TS356010 and TS407005 the proposed model shows higher performance than all BJ models, especially the S-FIS model. In station TS381010, the ARIMA model is better than M-FIS but the performance is lower than S-FIS. In station TS388002, SARIMA model showed better performance than S-FIS but lower than M-FIS. The average normalized MAE and average R measure from all datasets are shown in the Figure 8. It can be seen from the figure that, overall, the proposed models performed better than the results generated from AR, ARIMA and SARIMA model.

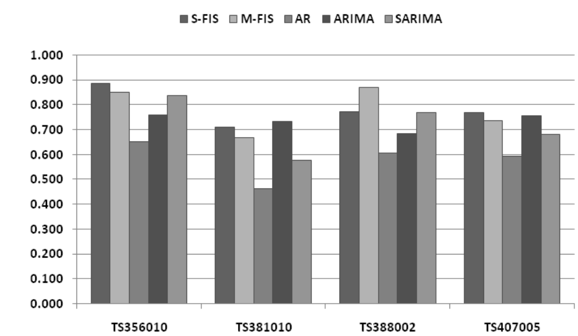
All aforementioned results are based on quantitative point of view in order to validate the experimental results. In qualitative point of view, the proposed model is easier to interpret than other models because the decision mechanism of such models is in the fuzzy rules form which is close to human reasoning [5]. Furthermore, when the models are in the form of rule base, it is easier for further enhancement and optimization by human expert. The advantage of S-FIS model is that time coefficient is expressed in term of MFs, so it is possible to apply optimization method to this feature. However, a large number of fuzzy rules are needed for single model. On the other hand, M-FIS model has smaller number of fuzzy rules when compared to S-FIS, but such model does not use any time feature.

## 7 Conclusions

Accurate rainfall forecasting is crucial for reservoir operation and flood prevention because it can provide an extension of lead-time of the flow forecasting and many time series prediction models have been applied. However, the prediction mechanism of those models may be difficult to be interpreted by human analysts. This study proposed the Single Fuzzy Inference System and the Modular Fuzzy Inference System, which use the concept of cooperative neuro-fuzzy technique to predict monthly rainfall time series in the northeast region of Thailand. The reported models used the *average-based interval* method to determine the fuzzy interval and use BPNN to extract fuzzy rules. The prediction performance of the proposed models is compared with conventional Box-Jenkins models. The experimental results showed that the proposed models could be a good alternative. Furthermore, the prediction mechanism can be interpreted through the human understandable fuzzy rules.

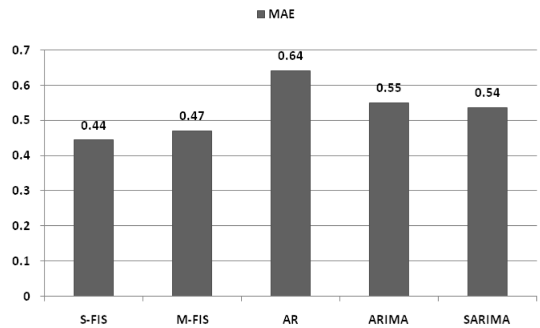



(a)

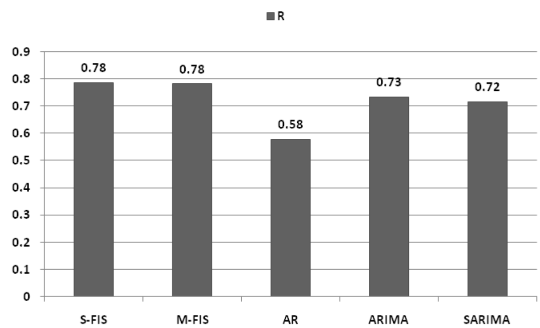



(b)

**Figure 7**: The comparison performance between the purposed models and conventional Box-Jenkins models: MAE (a) and R (b).




(a)




(b)

**Figure 8**: The average normalized MAE (a) and average R (b) of all datasets

## References


[1] V. K. Somvanshi, et al., “Modeling and prediction of rainfall using artificial neural network and ARIMA techniques.” J. Ind. Geophys. Union, vol. 10, no. 2, pp. 141-151, 2006.

[2] Z. F. Toprak, et al., “Modeling monthly mean flow in a poorly gauged basin by fuzzy logic,” *Clean*, vol. 37, no. 7, pp. 555-567, 2009.

[3] S. Kato and K. W. Wong, “Intelligent Automated Guided Vehicle with Reverse Strategy: A Comparison Study,” in Mario Köppen, Nikola K. Kasabov, George G. Coghill (Eds.) Advances in Neuro-Information Processing, Lecture Notes in Computer Science, Springer-Verlag, Berlin Heidelberg, pp. 638-646, 2009.

[4] K. W. Wong, and T. D. Gedeon, "Petrophysical Properties Prediction Using Self-generating Fuzzy Rules Inference System with Modified Alpha-cut Based Fuzzy Interpolation", Proceedings of The Seventh International Conference of Neural Information Processing ICONIP, pp. 1088-1092, November 2000, Korea

[5] K. W. Wong, P. M. Wong, T. D. Gedeon, C. C. Fung, “Rainfall Prediction Model Using Soft Computing Technique,” Soft Computing, vol 7, issue 6, pp. 434-438, 2003.

[6] C. C. Fung, K. W. Wong, H. Eren, R. Charlebois, and H. Crocker, “Modular Artificial Neural Network for Prediction of Petrophysical Properties from Well Log Data,” in IEEE Transactions on Instrumentation & Measurement, 46(6), December, pp.1259-1263, 1997.

[7] P. Coulibaly and N. D. Evora, “Comparison of neural network methods for infilling missing daily weather records.” Journal of Hydrology, vol. 341 pp. 27-41, 2007.

[8] C. L. Wu, K. W. Chau, and C. Fan, “Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques.” Journal of Hydrology, vol. 389, pp.146-167, 2010.

[9] W. Wang, K. Chau, C. Cheng and L. Qiu, “A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series.” Journal of Hydrology, vol. 374, pp. 294-306, 2009.

[10] A. K. Lohani, N. K. Goel and K. K. S. Bhatia, “Comparative study of neural network, fuzzy logic and linear transfer function techniques in daily rainfall-runoff modeling under different input domains.” Hydrological Process, vol. 25, pp. 175-193, 2011.

[11] P. C. Nayak, et al., A neuro-fuzzy computing technique for modeling hydrological time series, Journal of Hydrology, vol. 291. pp. 52-66, 2004.

[12] M. Z. Kermani and M. Teshnehlab, “Using adaptive neuro-fuzzy inference system for hydrological time series prediction.” Applied Soft Computing, vol. 8, pp. 928-936, 2008.

[13] A. Jain and A. M. Kumar, “Hybrid neural network models for hydrologic time series forecasting.” Applied Soft Computing, vol. 7, pp. 585-592, 2007.

[14] M. Khaashei, M. Bijari, G. A. r. Ardali, "Improvement of Auto-Regressive Integrated Moving Average models using Fuzzy logic and Artificial Neural Networks (ANNs)," *Neurocomputing*, vol. 72, pp. 956-967, (2009).

[15] K. P. Sudheer, "A data-driven algorithm for constructing artificial neural network rainfall-runoff models," Hydrological Precess, vol. 16, pp. 1325-1330, (2002).

[16] C. L. Wu and K. W. Chau, "Data-driven models for monthly streamflow time series prediction." *Engineering Applications of Artificial Intelligence*, vol. 23, pp. 1350-1367, 2010.

[17] S. Marsland, "Machine Learning An Algorithmic Perspective" CRC Press, 2009.

[18] H. Raman and N. Sunilkumar, "Multivariate modeling of water resources time series using artificial neural network*," Hydrological Sciences –Journal- des Sciences Hydroligiques*, vol. 40, pp.145-163, 1995.

[19] L. A. Zadeh, "Fuzzy Sets," Inform and Control, vol. 8, pp. 338 – 353. 1965.

[20] E. H. Mamdani, and S. Assilian, "An experiment in linguistic synthesis with fuzzy logic controller," International journal of man-machine studies, vol. 7 no. 1, pp.1-13, 1975.

[21] M. Sugeno, "Industrial application of fuzzy control," North-Holland, Amsterdam. 1985.

[22] K. Huarng, "Effective lengths of intervals to improve forecasting in fuzzy time series," *Fuzzy sets and system*, vol. 123. Pp. 387-394, 2001.

[23] H. Liu and M. Wei, "An improved fuzzy forecasting method for seasonal time series," *Expert System with Applications*, vol. 37, pp. 6310-6318, 2010.