School of Engineering and Information Technology

Interpretable Fuzzy Systems for Monthly Rainfall Spatial Interpolation and Time Series Prediction

Jesada Kajornrit

This thesis is presented for the degree of

Doctor of Philosophy of

Murdoch University

September, 2014

DECLARATION

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.

....
(Jesada Kajornrit)

ACKNOWLEDGEMENT

I wish to express my deepest gratitude and appreciation to my academic supervisor, Associate Professor Dr. Kevin Wong, for providing advice, assistance, and encouragement throughout this study, and especially for critically reviewing this thesis.

I would also like to thank the co-supervisor, Associate Professor Dr. Lance Fung, for his close association and advice as well as providing useful comments on my works. My extended thanks go to the staff at the School of Engineering and Information Technology and Murdoch University.

I am very grateful to the Dhurakij Pundit University for the financial support and the opportunity for my study. Without this support, my PhD study would not have been possible.

Finally, I am grateful to my family and my friends for their companionship, support and inspiration along the way of the study. I apologize if I have inadvertently forgotten to acknowledge those who have assisted me in any way.

ABSTRACT

This thesis proposes methodologies to analyze and establish interpretable fuzzy systems for monthly rainfall spatial interpolation and time series prediction. A fuzzy system has been selected due to its capability of handling the uncertainty in the data and due to its interpretability characteristic.

In the first part, this thesis proposes a methodology to analyze and establish interpretable fuzzy models for monthly rainfall spatial interpolation using global and local methods. In the global method, the proposed methodology begins with clustering analysis to determine the appropriate number of clusters, and fuzzy modeling and a genetic algorithm are then used to establish the fuzzy interpretation model. In the local method, the modular technique has been applied to improve the accuracy of the global models while the interpretability capability of the model is maintained.

In the second part, this thesis proposes a methodology to establish single and modular interpretable fuzzy models for monthly rainfall time series predictions. In the single model, the cooperative neuro-fuzzy technique and a genetic algorithm have been used. In the modular model, the modular technique has been applied to simplify the complexity of the single model. The whole system is decomposed into twelve sub-modules according to the calendar months. The proposed modular model consists of two functionally consecutive layers, the prediction layer and the aggregation layer. In the aggregation layer, Bayesian reasoning has been applied.

The case study used in this thesis is located in the northeast region of Thailand. The proposed models were compared with commonly-used conventional and intelligent methods in the hydrological discipline. The experimental results showed that, in the quantitative aspect, the proposed models can provide good prediction accuracy and, in the qualitative aspect, the proposed models can also meet the criteria used for model interpretability assessment.

LIST OF PUBLICATIONS RELATED TO THIS THESIS

A. Kajornrit, J., Wong, K.W., & Fung, C.C. (2014). *An interpretable fuzzy monthly rainfall spatial interpolation system for the construction of aerial rainfall maps*. Soft Computing (In Press).

B. Kajornrit, J., Wong, K.W., Fung, C.C., & Ong, Y.S. (2014). *An integrated intelligent technique for monthly rainfall time series prediction.* In Proceedings of the IEEE International Conference on Fuzzy Systems (China).

C. Kajornrit, J., Wong, K.W., & Fung, C.C. (2014). A modular spatial interpolation technique for monthly rainfall prediction in the northeast region of Thailand, *Advances in Intelligent Systems and Computing*, 265, 53–62.

D. Kajornrit, J., Wong, K.W., & Fung, C.C. (2013). An integrated intelligent technique for monthly rainfall spatial interpolation in the northeast region of Thailand, *Lecture Notes in Computer Science*, vol. 8227, pp. 384–391.

E. Kajornrit, J., & Wong, K.W. (2013). *Cluster validation method for localization of spatial rainfall data in the northeast region of Thailand*. In Proceedings of the IEEE International Conference on Machine Learning and Cybernetics (China).

F. Kajornrit, J., Wong, K.W., & Fung, C.C. (2013). *A modular technique for monthly rainfall time series prediction.* In Proceedings of the IEEE Symposium on Computational Intelligence in Dynamic and Uncertain Environments (Singapore).

G. Kajornrit, J. (2012). *Monthly rainfall time series prediction using modular fuzzy inference system with nonlinear optimization techniques*. In Proceedings of the Postgraduate Electrical Engineering and Computing Symposium (Australia).

H. Kajornrit, J., Wong, K.W., & Fung, C.C. (2012). Rainfall prediction in the northeast region of Thailand using cooperative neuro-fuzzy technique, *Asian International Journal of Science and Technology in Production and Manufacturing Engineering*, 5(3), 9–17.

I. Kajornrit, J., Wong, K.W., & Fung, C.C. (2012). *Rainfall prediction in the northeast region of Thailand using modular fuzzy inference system*. In Proceedings of the IEEE International Conference on Fuzzy Systems (Australia).

J. Kajornrit, J., Wong, K.W., & Fung, C.C. (2011). Estimation of missing rainfall data in northeast region of Thailand using spatial interpolation methods, *Australian Journal of Intelligent Information Processing Systems*, 13(1), 21–30.

K. Kajornrit, J., Wong, K.W., & Fung, C.C. (2011). *Estimation of missing rainfall data in northeast region of Thailand using kriging methods: A comparison study*. In Proceedings of the International Workshop on Bio-inspired Computing for Intelligent Environments and Logistic Systems (Australia).

SUMMARY OF PUBLICATIONS WITH RESPECT TO THE CHAPTERS

Chapters	Contributions	Publications
Chapter 1: Introduction		
Chapter 2: The problem of interpretability in hydrological models and interpretable fuzzy systems	• Successfully conduct a preliminary experiment to observe the performance of GIS-based spatial interpolation methods for monthly rainfall spatial interpolation.	[J], [K]
Chapter 3: An interpretable fuzzy system for monthly rainfall spatial interpolation	• Successfully develop a methodology to analyze and establish an interpretable fuzzy system for monthly rainfall spatial interpolation.	[A], [D], [E]
Chapter 4: A modular fuzzy system for monthly rainfall spatial interpolation	• Successfully improve the performance of the proposed model in the previous chapters by means of a modular technique.	[C]
Chapter 5: An interpretable fuzzy system for monthly rainfall time series prediction	• Successfully develop a methodology to establish an interpretable fuzzy system for monthly rainfall time series prediction.	[B], [H]
Chapter 6: A modular fuzzy system for monthly rainfall time series prediction	• Successfully improve the performance of the proposed model in the previous chapter by means of a modular technique and Bayesian reasoning.	[F], [G], [I]
Chapter 7: Conclusions		

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

ACF	Autocorrelation function
ADI	Alternative Dunn’s index
AM	Aggregation modules
ANFIS	Adaptive neuro-fuzzy inference system
ANN	Artificial neural network
AIC	Akaiki information criterion
ARMA	Autoregressive moving average
CK	Ordinary co-kriging
CNFIS	Cooperative neuro-fuzzy inference system
BPNN	Back-propagation neural network
DI	Dunn’s index
FCM	Fuzzy c-means
FIS	Fuzzy inference system
FL	Fuzzy logic
GA	Genetic algorithm
GAFIS	Genetic algorithm with fuzzy inference system
GIS	Geographic information systems
IDW	Inverse distance weighting
LP	Local polynomial
MF	Membership function
MFIS	Mamdani-type fuzzy inference system

MFIS–ORG	Mamdani-type fuzzy inference system without optimization
MFIS–OPT$_1$	Mamdani-type fuzzy inference system with first optimization
MFIS–OPT$_2$	Mamdani-type fuzzy inference system with second optimization
Mod FIS	Modular fuzzy inference system
Mod FIS–FSG	Modular fuzzy inference system with Gaussian function
Mod FIS–FST	Modular fuzzy inference system with triangle function
Mod FIS–BSA	Modular fuzzy inference system with Bayesian aggregation
Mod FIS–HSA	Modular fuzzy inference system with Hessian aggregation
OK	Ordinary kriging
PACF	Partial autocorrelation function
PM	Prediction modules
RBFN	Radial basis function network
S	Separation index
SC	Partition index
SD	Standard deviation
SFIS	Sugeno-type fuzzy inference system
SOM	Self organizing maps
SSE	Sum square error
TPS	Thin plate splines
TSA	Trend surface analysis
UK	Universal kriging
XB	Xie and Beni’s index

CHAPTER 1
INTRODUCTION

1.1. Rainfall Prediction

Prediction of hydrological variables is one of the important tasks in water management systems and planning (Araghinejad et al., 2011). In agricultural countries, such as Thailand, rainfall plays a vital role in the countries' economic development. The effectiveness of rainfall prediction is an important factor for allowing sustainable agricultural development, as well as flood and drought prevention (Wu et al., 2010). To accomplish this task, efficient rainfall prediction techniques as well as computational tools are essential.

In general, strategies used in the prediction of rainfall, including other climate variables, are based on spatial and temporal perspectives. In the former, the prediction is performed for any location by means of the spatial relationships (Isaaks & Srivastava, 1989), while in the latter the relationships along the time dimension are used for future prediction (Montgomery et al., 2008). In this thesis, the term "spatial interpolation" and "time series prediction" are adopted for both perspectives, respectively. As a result, rainfall prediction in this thesis is separated into two challenging issues.

Spatial interpolation is a method that estimates the values at unsampled points by using the values from neighbouring sampled points (Li & Heap, 2008). In geographic information systems (GIS), such a method is commonly used to create continuous surfaces from sampled points (Chang, 2006). This is important for water and agricultural management because such information is necessary in decision making, for example, irrigation planning, water flood way planning, assessment of dam and reservoir installation, and selection of agricultural products in a certain area (Sharma & Irmak, 2012).

Rainfall time series prediction is used to predict the value of rainfall in the near future. The prediction models are generated from the time series historical records (Wu et al., 2010). This is necessary in water management because it provides the lead-time extension to assess the amount of water coming into the basin. An accurate future rainfall prediction can have an impact on flood and drought prevention, reservoir operation, contract negotiation, and irrigation scheduling (Araghinejad et al., 2011).

Regardless of whether spatial interpolation or time series prediction is used, the common objective is to create an accurate rainfall prediction model. However, with respect to the rainfall data, creating an accurate rainfall prediction model is not an easy task because rainfall data collected can contain uncertainty and noisy information. They are also highly non-linear in nature. Compared to other climate variables such as humidity or temperature, prediction of rainfall is relatively more difficult due to various influential factors such as topology of the area (Kim & Pachepsky, 2010).

For decades, Box-Jenkins models have been commonly adopted in hydrological time series prediction. However, due to the complexity in rainfall data, the accuracy of the prediction models depends on the linearity and prior assumptions used. In the same manner, this problem has also been observed in the kriging methods when performing spatial interpolation on rainfall data. Consequently, the need of a better rainfall prediction model still exists.

Recently, with the advancement of modern computational modeling, many computational intelligent techniques have been proposed (Negnevitsky, 2011; Karray & Silva, 2004). These techniques have been applied to rainfall prediction. In most cases, computational intelligent techniques are able to provide considerable prediction accuracy when used to construct rainfall prediction models (Wu & Chau, 2013; Piazza et al., 2011; Wu et al., 2010; Hu & Zhang, 2008; Zhang & Wang, 2008; Somvanshi et al., 2006; Lee et al, 1998).

1.2. Understand the Rainfall Prediction Models

System identification involves the use of mathematical tools and algorithms to build dynamic models describing the behavior of the real-world systems from measured data (Zhou & Gan, 2008). One of the issues in system modeling is interpretability (or transparency) of the models. Model interpretability is defined as a property that enables users to understand and analyze the influence of each system parameter on the system output (Harris et al., 2002; Setnes et al., 1998; Brown & Harris, 1994). In general, there are three different strategies for system modeling addressing interpretability of the model: the white-box, black-box and grey-box models.

In the white-box model (e.g. Newton's laws), parameters are clearly presented and the model can be interpretable. However, when the problem becomes complex, the white-box model tends to be impractical due to the limitation of mathematic representation.

Contrastingly, the black-box model does not provide clear information on how the model performs the determination. The user normally has minimal options in understanding how the model works. However, this type of model can provide prediction without using prior knowledge. An example of the black-box model is an artificial neural network (ANN). The structure and parameters of the model may not reflect the behavior of the system to be modeled and sometimes provides questionable results (Abonyi et al., 2000).

The white-box model normally provides high interpretability with sometimes lower accuracy, whereas the black-box model may offer higher accuracy but poor interpretability. As a result, the grey-box model is deemed to be in between these two extremes when it comes to accuracy and interpretability, as prior knowledge of the system is considered, but it leaves the unknown parts of the system to be represented by the black-box modeling approach (Zhou & Gan, 2008). Interpretability is an important issue in a data-driven model because human analysts can gain insight into the complex real-world system to be modeled.

With respect to the hydrological area, no matter what rainfall or other hydrological variables are used, the established models usually try to achieve higher prediction accuracy, and most of the time the model transparency issue is overlooked (Wu & Chau, 2013; Wu et al., 2010). The trend is also observed with the large number of ANN techniques used to establish the hydrological prediction models (Singh & Imtiyaz, 2013; See et al., 2004; ASCE Task Committee on Application of Artificial Neural Networks in Hydrology [ASCE-TCAANNH], 2000a; ASCE-TCAANNH, 2000b).

ANN may provide considerable prediction accuracy and such a technique is easy to use and needs no prior knowledge of the system to establish the models. With these advantages, it is reasonable to understand why ANN has gained momentum and interest from many researchers in hydrological prediction in recent decades. On the other hand, fuzzy logic (FL), which is categorized as the interpretable grey-box model, formulates the system knowledge with rules in a transparent way to interpretation and analysis, and leaves the inference mechanism as the opaque part. For this reason, researchers have started to look at the use of FL to handle the issue of accuracy and interpretability of the rainfall prediction models (Asklany et al., 2011; Wong et al., 2003; Huang et al., 1998).

The aim of this thesis is to establish a rainfall prediction model that addresses the issues of accuracy and interpretability. This is important in the field of rainfall prediction because an understanding of rainfall behaviors is necessary for water management and planning. The interpretable rainfall prediction model can allow human analysts to practically enhance the model and, if possible, to gain insight into the rainfall data to be modeled.

1.3. Case Study Area and Monthly Rainfall Data

The case study area used in this thesis is located in the northeast region of Thailand (Figure 1.1), within the area of latitude between 14.11°N to 18.45°N and longitude between 100.83°E to 105.63°E. The total area size is 168,854 km^2 and it is a large plateau. The minimum and maximum altitudes are 17 m and 1799 m, respectively, above sea level.

The topology of the study area (the Khorat Plateau) tilts up toward the west region, the Phetchabun mountain range, down toward the east. The plateau consists of two main plains, the southern Khorat plain and the northern Sakon Nakhon plain, which are separated by the Phu Phan mountain range.

In general, there are three seasons a year, that is, hot, cool and rainy seasons. The rainy period gradually starts from March, reaches the highest level in between June and August, and then usually reduces by November. The average annual rainfall varies from 1270 to 2000 mm.

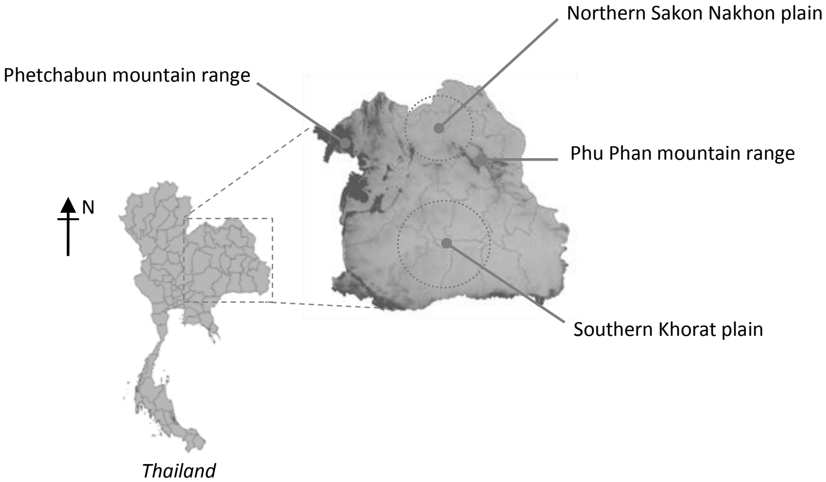


Figure 1.1. Case study area is located in the northeast region of Thailand.

According to the information reported by the Land Development Department of Thailand (n.d.), this area has experienced severe drought for many decades. The cultivation in this area is mainly subject to irrigation systems. Inefficient irrigation systems can result in poor agricultural produce. In comparison with other parts of Thailand, this area has to be further developed, and thus provides the motivation for this study.

The rainfall data used in this thesis are monthly rainfall data collected from rain gauge stations within the study area (Remote Sensing & GIS, n.d.). The number of rain gauge stations is approximately 295. The rainfall data ranges from 1981 to 2001. Some necessary information about the data will be presented in detail in the subsequent chapters.

1.4. Aim and Objectives of the Thesis

As aforementioned, the aim of this thesis is to propose a rainfall prediction model which is capable of handling the interpretability issue with considerable prediction accuracy to monthly rainfall spatial interpolation and time series prediction. A fuzzy system is therefore selected as a suitable technique to handle the complex rainfall data and to provide an interpretable mechanism at the same time. However, achieving high accuracy and high interpretability simultaneously is not an easy task; hybrid and modular techniques are therefore applied to achieve these goals.

1.4.1. Objectives for Spatial Interpolation

The main objective of the thesis is to develop a methodology to analyze and establish an interpretable fuzzy model for monthly rainfall spatial interpolation. In order to address this problem, two important issues should be taken into consideration. The first issue is how to localize a global area into local areas. The second issue is how to create a fuzzy model which is capable of providing good estimation accuracy and good model interpretability. However, a single fuzzy model may have limitations in achieving a high accuracy for a large area and maintaining the interpretability of the model. Therefore, a subsequent issue is how to increase the accuracy of the model while maintaining or enhancing the interpretability of the model. Solutions to all these issues are the key objectives of this study.

1.4.2. Objectives for Time Series Prediction

Another main objective of the thesis is to develop a methodology to create an interpretable fuzzy model for monthly rainfall time series prediction. The idiosyncrasy of this problem is different from spatial interpolation in that the input to the time series prediction model is not known in advance. As a result, the modeling process depends on the defined input, and this may complicate the modeling process. Furthermore, once an interpretable fuzzy model is established, such a model should provide an approach to enable human analysts to analyze into monthly time series data. In summary, all these issues will be addressed and they form the key objectives of this part of the study.

1.5. Overview of the Thesis

Chapter 1 provides an introduction to this thesis. In Chapter 2, related works about spatial interpolation and time series prediction in hydrological and related areas are presented. The problem of interpretability in hydrological models will be addressed. Furthermore, concepts of fuzzy systems and their interpretability features are discussed.

This thesis can also be considered in two parts. First, the thesis contributes towards the problem of monthly rainfall spatial interpolation as covered in Chapter 3 and Chapter 4. In Chapter 3, cluster analysis and global spatial interpolation methods are presented, whereas in Chapter 4 modular techniques used to improve the interpolation accuracy and model interpretability of the global method are described.

In the second part, the thesis contributes to the problem of monthly rainfall time series prediction as described in Chapters 5 and 6. In Chapter 5, a single model for monthly time series prediction is introduced, and in Chapter 6 modular techniques are used to simplify the complexity of the single model. These single and modular models are alternative solutions and each model has its own advantages. Finally, the conclusion is provided in Chapter 7.

CHAPTER 2

THE PROBLEM OF INTERPRETABILITY IN HYDROLOGICAL MODELS AND INTERPRETABLE FUZZY SYSTEMS

2.1. Introduction

Spatial interpolation and time series prediction of the rainfall play significant roles in water management and planning. Spatial interpolation provides the information of spatial distribution of the rainfall over a study area, while time series prediction provides the information for future projection used for flow forecasting. However, due to the complex nature of the rainfall, these tasks face many challenges.

Hydrological processes such as rainfall depend on many complex factors that are not clearly understood, and the conditions change from area to area. In this case, data-driven based models have been used to create the prediction models. Recently, many intelligent methods have been adopted to establish the hydrological models, due to the considerable ease of use and accuracy from these methods. Although most of these intelligent methods can provide satisfactory outcomes in terms of accuracy, such methods seem to disregard the interpretability capability.

Model interpretability is an important issue in data-driven modeling because interpretable models allow human analysts to understand the models. Furthermore, if the interpretable models are presented in an appropriate way, human analysts can gain insights of the data to be modeled. This thesis therefore focuses on how to create interpretable models for monthly rainfall spatial interpolation and time series prediction.

This chapter begins with literature reviews of spatial interpolation and time series prediction in hydrological and related areas. Next, the problem of interpretability will be examined and the importance of such systems will be highlighted. After that, backgrounds of fuzzy systems will be introduced, followed by the interpretability criteria of fuzzy modeling.

2.2. Spatial Interpolation in Hydrological and Related Areas

Spatial interpolation is a method that estimates the values at unsampled points by using the values from neighbouring sampled points (Li & Heap, 2008). In the discipline of geographic information systems (GIS), Burrough and McDonnell (1998) have classified spatial interpolation methods into global and local methods. The global methods use all sampled data points from the entire study area to establish spatial interpolation models, whereas the local methods use only a certain number of sampled data points to perform interpolation.

In turn, the local methods themselves can be classified into deterministic and stochastic (or geostatistic) methods (Li & Heap, 2008). The former does not provide the assessment of error with the estimated value. The latter, on the other hand, offers the assessment of errors with estimated variances. Spatial interpolation methods can also be grouped into exact and inexact methods. The interpolated values are the same as sampled values in the exact method whereas the interpolated values are different from sampled values in the inexact methods (Chang, 2006).

To date, many spatial interpolation methods have been proposed. However, none of the spatial interpolation methods are suitable to be applied to all spatial data in different locations and conditions. The success of a spatial interpolation method to provide reasonable interpolation of a spatial variable depends on several factors (Li & Heap, 2008) and sometimes a comparison study is necessary for most case studies (Li et al., 2011; Piazza et al., 2011). One recommended guideline of spatial interpolation methods for environmental variables is found in the work of Li and Heap (2008). Their guideline introduces a variety of commonly-used spatial interpolation methods and recommends methods for specific requirements.

According to the guideline, some of commonly-used spatial interpolation methods are adopted in this thesis for comparison purposes. These methods include trend surface analysis (TSA), inverse distance weighing (IDW), thin plate splines (TPS), ordinary kriging (OK), universal kriging (UK) and ordinary co-kriging (CK).

TSA is a global inexact spatial interpolation method that estimates values at unsampled points by using a polynomial equation. In practice, a third order polynomial equation is normally used because this order is appropriate for real-world data, which have both hill and valley surfaces (Chang, 2006). TSA has been applied to many environmental variables, for example, rainfall (Kajornrit et al., 2011), wind speed (Luo et al., 2008), and temperature (Collins & Bolstad, 1996). However, the accuracies of TSA reported in this literature are relatively poor in comparison with other methods.

IDW is an exact local spatial interpolation method that estimates the values at unsampled points by using a linear combination of values from nearby sampled points weighted by an inverse distance function (Li & Heap, 2008). IDW is the most commonly-used method and usually used as a control method (or standard method) for comparison (Li et al., 2011). IDW has been applied to many environmental variables, for example, precipitation (Luo et al., 2011; Goovaerts, 2000; Hartkamp et al., 1999; Nalder & Wein, 1998), wind speed (Cellura et al., 2008; Luo et al., 2008), solar irradiation (Sen & Sahin, 2001), seabed mud content (Li et al., 2011), depth of groundwater (Sun et al., 2009) and soil properties (Robinson & Metternicht, 2006).

Beside the standard IDW, some literature proposed additional techniques to IDW to enhance the interpolation accuracy. Some examples of these techniques are the clustering-assisted gradient plus inverse distance square (Tang et al., 2012), incorporation of fuzzy concept with a genetic algorithm (GA) (Chang et al., 2005), a classic weighting method with a cumulative semivariogram (Sen & Sahin, 2001) and the gradient plus inverse distance squared (Nalder & Wein, 1998).

TPS is an exact local spatial method that creates a surface passing through the sampled points with minimum curvature. Conceptually, TPS works as a flexible sheet of rubber passing through the sampled points (Chang, 2006). It was originally developed for climatic data analysis by Wahba and Wendelberger (1980). This technique has been applied to problems such as precipitation (Piazza et al., 2011; Hong et al., 2005; Hartkamp et al., 1999), temperature (Hancock & Hutchinson, 2006; Hong et al., 2005; Jeffrey et al. 2001), and wind speed (Luo et al., 2008).

Based on the literature, IDW and TPS have provided more accurate results than TSA in general. However, the accuracy between IDW and TPS cannot be clearly differentiated. This is because their accuracy is subject to many factors such as the topology of the study areas.

In addition to the deterministic methods, stochastic methods perform interpolation like IDW, except that such methods use spatially dependent variance of data instead of spatial distance. Kriging methods, developed by Matheron (1965) and based on the work of Krige (1951), are stochastic methods. Kriging does not only estimate data at unsampled points, but also assesses the quality of estimation.

The assumption of kriging methods is that the spatial variation of data is neither totally random nor deterministic. Instead, the spatial variation consists of three components, namely, a spatial correlation component that represents the variation of the regionalized variable, a drift or structure which represents a trend, and a random error term.

As far as kriging methods are concerned, one indispensable issue that has to be mentioned is the semivariogram. A semivariogram is the model representing the spatial correlation and is presented as:

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^{n} [z(x_i) - z(x_i + h)]^2 \tag{2.1}$$

where $\gamma(h)$ is the average semivariance between sampled points separated by lag h, n is the number of pairs of sample points, and z is the attribute value. In other words, a semivariogram is the relationship between lag distance and semivariance. A general model of a semivariogram is depicted in Figure 2.1.

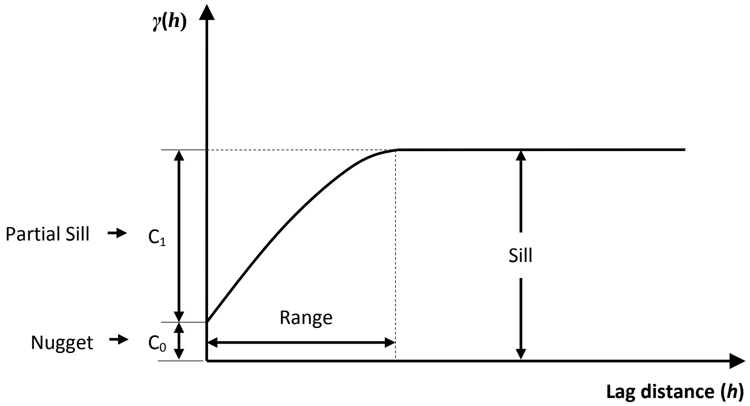


Figure 2.1. A general model of a semivariogram.

According to the figure, some important features are displayed: the nugget is the semi-variance at the distance of zero, representing the sampling error and/or spatial variance at a shorter distance than the minimum sample space; the range is the distance at which the semivariance starts to level off and beyond the range the semi-variance becomes a relatively constant value; and the sill is the semivariance at which the leveling takes place. The sill, in turn, consists of the partial sill (C_1) and the nugget *(C_0)*. In practice, an experimental semivariogram will be fitted by a mathematical model mainly for computational purposes. Usually, four types of mathematical models are preferred: spherical, exponential, Gaussian and linear models.

Presently, there are many kriging techniques proposed (Li & Heap, 2008). However, three commonly-used kriging methods are OK, UK and CK as mentioned before. OK interpolates data by using a fitted semivariogram and focuses only on the spatial correlation and absence drift, while UK interpolates data in the same way as OK except the drift of data is taken into account. CK performs similarly to OK except that a secondary variable is allowed. In practice, the most commonly-used second variable is the altitude for rainfall variable (Goovaerts, 2000).

Much work has contributed to the use of kriging methods to many hydrological and climatic variables such as: precipitation (Piazza et al., 2011; Luo et al., 2011; Bargaoui & Chebbi, 2009; Haberlandt, 2007; Yue et al., 2003; Jeffrey et al., 2001; Goovaerts, 2000; Hartkamp et al., 1999; Nalder & Wein, 1998), temperature (Hartkamp et al., 1999; Nalder & Wein, 1998), evaporation (Yue et al., 2003), wind speed (Cellura et al., 2008; Luo et al., 2008), and depth of groundwater (Sun et al., 2009). In addition to standard methods, some advanced kriging techniques such as kriging with external drift and indicator kriging with external drift may include secondary information from radars if they are available (Haberlandt, 2007).

One difficult task of using kriging methods is the modeling of the semivariogram. Fitting the experimental semivariogram with the mathematical model is rather subjective to users' experiences (Nalder & Wein, 1998; Huang et al., 1998). Besides, kriging methods also require the stationary condition of spatial data (Li et al., 2011). In other words, the success of kriging methods depends on the appropriate selection of a semivariogram and the stationary condition of spatial data. Furthermore, in comparison with deterministic methods, more computation is normally required, such as solving simultaneous equations being needed for every interpolated point (Huang et al., 1998).

With the advent of intelligent techniques such as ANN, much literature has adopted these techniques to perform spatial interpolation and has showed a common agreement that these techniques are promising approaches. In comparison with GIS-based methods (deterministic and stochastic), applications of these intelligent techniques are relatively new to those methods that have been discussed so far.

One of the commonly-used intelligent methods is the back-propagation neural network (BPNN). BPNN has been successfully applied to rainfall spatial interpolation (Piazza et al., 2011; Hu & Zang, 2008; Zhang & Wang, 2008) and wind speed spatial interpolation (Cellura et al., 2008). BPNN needs no prior knowledge or stationary condition to generalize the relationships of the spatial data. Also, it can easily incorporate the altitude variable to the model for rainfall spatial interpolation (Piazza et al., 2011; Kajornrit et al., 2011).

In addition to BPNN, radial basis function network (RBFN) is another ANN that is widely used for spatial interpolation. The architecture of RBFN is more interpretable and the training algorithm is faster than BPNN (Lin & Chen, 2004). The works of Liu et al. (2011), Luo et al. (2011), Lin and Chen (2004) and Lee et al. (1998) are successful examples of the applications of RBFN to spatial interpolation problems. Lee et al. (1998) mentioned in their study that RBFN showed superior results than BPNN. However, a comparison between these ANNs is difficult to justify due to different conditions of the study area and the available spatial data.

Another advantage of ANN is its flexibility as ANN can be efficiently integrated to other techniques. For example, in the case of RBFN, Liu et al. (2011) used RBFN with the bagging ensemble technique for soil content spatial interpolation. Luo et al. (2010) integrated RBFN to IDW for interpolating spatial precipitation data. Lin and Chen (2004) improved the original RBFN by combining it with semivariogram. In the case of BPNN, it was applied to the kriging method, in which BPNN was first used to capture the trend component of the spatial data and its residual was captured by the kriging method. This technique is commonly known as the neural kriging method and it has been applied to wind speed data (Cellura et al., 2008) and climatic data (Demyanov et al., 1998).

Fuzzy systems are the other approaches that have been applied to spatial interpolation. Huang et al. (1998) applied a dynamic fuzzy-reasoning-based function estimator to rainfall spatial interpolation, whereas Wong et al. (2001) used conservative fuzzy reasoning. Even though these fuzzy systems have showed promising results, the lack of a learning algorithm in the fuzzy systems makes it difficult to establish the prediction model.

Lately, hybrid techniques have also been adopted. Tutmez and Hatipoglu (2010) applied an adaptive neuro-fuzzy inference system (ANFIS) to spatial interpolation of nitrate in the groundwater and Wong et al. (2003) used a cooperative neuro-fuzzy inference system (CNFIS) for spatial interpolation of rainfall. Their results suggested that ANFIS and CNFIS are not only able to provide accurate results, but are also capable of providing the model interpretability for human analysts. These studies suggested that model interpretability is another issue that is equally important to the estimation accuracy.

Some other intelligent techniques have also been applied to spatial interpolation in the hydrological area. The support vector machine (SVM) (Li et al., 2011; Gilardi & Bengio, 2000), and the geographically weighted regression (GWR) (Piazza et al., 2011; Yu,

2009) are such examples. However, the number of applications from these techniques is still limited and more comparative studies are required to assess their performance.

As can be observed from the trend of research in this field, intelligent techniques have shown that they can be good alternative approaches from the conventional GIS-based methods. Such techniques also decrease the requirement of prior assumptions in the establishment process of stochastic methods and they provide relatively good prediction accuracy. In the next section, literature review in the area of time series prediction in hydrological and related areas is provided.

2.3. Time Series Prediction in Hydrological and Related Areas

In hydrological time series prediction, multiple linear regression (MLR) and conventional Box-Jenkins time series models (Box & Jenkins, 1970) have been widely adopted for decades. Applications of MLR, for example, can be found in the works of Wu et al. (2010) to predict daily and monthly rainfall time series; and in the works of Sudheer et al. (2002) to predict daily flow time series. The model was also applied to the monthly rainfall variable for drought forecasting in the work of Bacanli et al. (2009). Conventional Box-Jenkins time series models, that is, auto-regressive (AR) and autoregressive moving average (ARMA) have been applied to many hydrological variables for comparative purposes in the following literature.

The AR model was used to predict the daily streamflow time series in the works of Zounemat-Kermani and Teshnehlab (2008), and Firat and Güngör (2008), and it was also used to predict the monthly streamflow time series in the works of Firat and Turan (2009), Jain and Kumar (2007), and Raman and Sunilkumar (1995). The ARMA model was used in the work of Sudheer et al. (2002) and Nayak et al. (2004) to predict the daily streamflow time series, and was used in the work of Wu and Chau (2010) to predict the monthly streamflow time series. Wang et al. (2009) applied the ARMA model to the monthly discharge flow time series and Somvanshi et al. (2006) applied this model to predict the mean annual rainfall time series. Furthermore, Lohani et al. (2010) applied the Box-Jenkins linear transfer function for the daily rainfall-runoff model.

The methods reviewed so far have a common agreement between the researchers. From their experiments, the prediction accuracy of these models has been limited due to the linearity of the models. Such models suffer from the assumption of stationary (Montgomery et al., 2008), linearity (Jain & Kumar, 2007) and normal distribution conditions (Wang et al., 2009). These conditions prevent the models from representing the non-linear dynamic inherent in hydrological processes (Tokar & Johnson, 1999). Consequently, contemporary non-linear models such as ANN have been utilized in hydrological time series prediction.

ANN has been widely used in hydrological time series prediction, especially the one hidden layer BPNN. (Wu & Chau, 2013; Guo et al., 2011; Lohani et al., 2010; Wu et al., 2010; Wu & Chau, 2010; Wang et al., 2009; Firat & Turan, 2009; Bacanli et al., 2009; Firat & Güngör, 2008; Jain & Kumar, 2007; Somvanshi et al., 2006; Nayak et al., 2004; Sudheer et al., 2002; Raman & Sunilkumar, 1995). Based on the literature, ANN has proven to be an efficient technique and has provided more accurate results than those linear models in general. Furthermore, the ANN models are easier to establish and need no prior assumptions when compared to those linear models.

However, one comment about ANN is that such a model falls in the group of "Atheoretical model" (Sudheer et al., 2002). In other words, there is no consistent theory to define an appropriate input vector to the models for time series prediction. In the case of Box-Jenkins models, the establishment method employs the statistical theory to define the appropriate input to the model, that is, the autocorrelation function (ACF) and the partial auto-correlation function (PACF).

Sudheer et al. (2002) investigated the application of ACF and PACF to define an appropriate input vector to ANN. They suggested that ACF and PACF can be used as general criteria to define an appropriate input vector. This suggestion has been adopted later in some recent literature (Wu & Chau, 2013; Monira et al., 2011; Wu et al., 2010; Wu & Chau, 2010; Wang et al., 2009; Somvanshi et al., 2006).

Much literature has been dedicated to improve the prediction accuracy of ANN models. One approach is to apply pre-processing techniques to the time series data before feeding to ANN models. For example, Raman and Sunilkumar (1995) applied statistical normalization to the monthly flow time series whereas Jain and Kumar (2007) applied de-trended and de-seasonalized techniques. It seemed that these techniques can adjust the kurtosis and skewness of time series data to be more normally distributed (Wu & Chau, 2013; Jain & Kumar, 2007), and this resulted in improved accuracy. The smoothing techniques such as moving average (MA), principal component analysis (PCA), singular spectrum analysis (SSA) and wavelet analysis (WA) have recently been investigated in the work of Wu and Chau (2013), Wu et al. (2010) and Guo et al. (2011). These techniques have successfully improved the prediction accuracy of ANN by removing noise from time series data.

Besides pre-processing techniques, the adaptation in the architecture of the models is another approach to improve prediction accuracy of ANN. One common limitation of ANN is that the trained model may fall in the local minima and cannot efficiently generalize the training data. Modular and ensemble techniques have been applied to single ANNs to address this problem. The work of Wu and Chau (2013), Wu et al. (2010) and Raman and Sunilkumar (1995) are examples of the modular technique. An ensemble technique has been applied to ANN in the work of Monira et al. (2011) and applied to SVM in the work of Lu and Wang (2011) for rainfall time series prediction.

The fuzzy inference system (FIS) is another technique that has been adopted in the hydrological area (Asklany et al, 2011; Lohani et al., 2010; Toprak et al., 2009). In general, two commonly-used FISs are the Mamdani-type FIS (MFIS) and the Sugeno-type FIS (SFIS). In the case of time series, it seems that MFIS is not as popular as the SFIS model. MFIS was used as a method to predict water consumption time series by Firat et al. (2009) for comparison purposes. Based on their experiment, MFIS provided lower prediction accuracy when compared to the SFIS model.

An advantage of SFIS over MFIS is that such a model is capable of integrating with the back-propagation learning technique. ANFIS is an example of this capability. ANFIS has been applied to many hydrological time series variables, for example, streamflow (Firat & Turan, 2009; Wang et al., 2009; Firat & Gungor, 2008; Zounemat-Kermani & Teshnehlab, 2008; Keskin et al., 2006; Nayak et al., 2004), water consumption (Firat et al., 2009), drought index (Bacanli et al., 2009), and rainfall (Afshin et al., 2011). In these application examples, ANFIS has combined the learning ability of BPNN and FIS in achieving improved results.

From the literature reviewed, ANFIS provided more accurate results than BPNN and those linear models in their experiments. Furthermore, such a model is more interpretable than BPNN or Box-Jenkins models. ANFIS represents the model by fuzzy sets and fuzzy rules that are close to the nature of human linguistics. Fuzzy sets are close to human linguistic properties and fuzzy rules are close to logical inferences. Besides ANFIS, another method of combining BPNN and FIS is the fuzzy neural network, in which the fuzzy system is represented in the nodes of the BPNN to handle the uncertainty of data. The use of this technique in a hydrological study was reported in Alvisi and Franchini (2011). However, this technique was not intentionally applied to address the interpretability issue.

Recently, some other intelligent techniques have been used in hydrological time series prediction. SVM and support vector regression (SVR) are examples of these techniques. Some studies have shown that these techniques can be good alternative techniques (Wu & Chau, 2013; Guo et al., 2011; Lu & Wang, 2011). Singular spectrum analysis (SSA) was another technique that has been used for hydrological time series prediction (Marques et al., 2006). However, the SSA technique was applied to the time series data in order to decompose time series components for further analysis purposes, instead of enhancing prediction accuracy.

In summary, a number of techniques used in hydrological time series prediction have been reviewed. It is observed that the intelligent techniques such as ANN and its variants have gained much attention from researchers recently. Due to flexibility of model establishment as well as considerable prediction accuracy, the linear Box-Jenkins models have gradually been overtaken by the intelligent techniques as described in contemporary literature.

2.4. The Problem of Interpretability in Hydrological Models

Up to this point, some of literature concerning spatial interpolation and time series prediction in hydrological and related areas have been discussed. It is noted that intelligent techniques have gained attention from researchers and hydrologists, and these intelligent techniques have been widely adopted as alternative approaches over conventional methods. However, most of these intelligent techniques aim at achieving only the accuracy of the models and disregard the model interpretability issue.

As aforementioned, the interpretability of models is important because human analysts can gain insight into the data to be modeled when prior knowledge is unknown or unclear. For example, in cases when the data comes from natural phenomena, there is little knowledge available. Consequently, establishing an interpretable data-driven model for those natural phenomena is necessary.

As the need of interpretable models have been illustrated and emphasized, the objectives of this thesis can be formulated as follows. The main objective of this thesis is to develop a framework to establish interpretable data-driven models for spatial interpolation and time series prediction. From the literature review, it was suggested that fuzzy systems can be a promising solution to deal with the interpretability of the models as well as the complexity in the rainfall data. However, there are still many challenges that need to be addressed in order to create an accurate and interpretable fuzzy system, especially for hydrological application, and this thesis aims to address them.

2.4.1. Issues in Spatial Interpolation

One of the aims of many researchers in the field is to develop techniques that can improve global interpolation accuracy. Although the interpolator itself is in the form of a unified model, which is convenient for further analysis, such methods provide relatively lower accuracy than the local methods. Therefore, establishing an accurate interpretable fuzzy system for global spatial interpolation is one of the challenges that needs to be investigated.

Although intelligent techniques such as ANN can work well as global methods, the accuracy of these techniques can be improved by cooperating with the concept of divide and conquer as suggested in many research works (Wong et al., 2003; Huang et al., 1998; Lee et al., 1998). However, the procedure to localize (divide) the global area into local areas is still subjective (Huang et al., 1998; Lee et al., 1998). Although some clustering techniques such as the self-organizing maps (SOM) may cluster data automatically, it is sometimes difficult to interpret the results when applied to noisy spatial data. There is a need to investigate more systematic localization procedures.

For local deterministic and geostatistic interpolation methods, the accuracy can be improved from the global method, as reported in many studies (Luo et al., 2008; Collins & Bolstad, 1996). However, considerable computation is required (Huang et al., 1998). Furthermore, in the case of geostatistic methods such as the kriging method, fitting a semivariogram is rather subjectivity (Nalder & Wein 1998; Huang et al., 1998). Expert knowledge may be needed to examine and establish the appropriate semivariogram model. It would be helpful if intelligent techniques can enable human analysts to mitigate some of the subjectivity in the model establishment process.

2.4.2. Issues in Time Series Prediction

One problem of intelligent methods used for time series prediction is there is no consistent procedure to select appropriate inputs to the systems (Wang et al., 2009; Sudheer, et al., 2002). Although ACF and PACF can be used as a recommended criterion, it may affect the interpretability problem if the selected inputs are considerably large, especially for the FIS model. Feeding high dimensional inputs to the model results in the readability problem (Zhou & Gan, 2008) in the antecedent part of the FIS model. Thus, in this case, the first problem that needs to be considered is how to select appropriate and reasonable inputs for the monthly time series data.

Although the interpretability issue of the hydrological time series prediction model is not new, such an issue has been ignored by many researchers. That is because most of the recent literature aimed only to enhance the quantitative prediction results. Thus, the amount of literature related to this issue is rather limited. The interpretability issue of the hydrological prediction model can have a significant impact on time series data analysis. That is because the interpretable advantage of the model can provide a new approach to analyze the time series data, and this thesis aims to address this issue.

2.5. Solving the Issue of Interpretability with Fuzzy Systems

This thesis selects the fuzzy system to address the interpretability problem mentioned. The selection not only considers the interpretability of the fuzzy system, but also the capability of handling uncertainty in the data. The fuzzy system is an efficient approach to handle the uncertainty and complexity in rainfall data. To facilitate further discussion, this section provides the background on fuzzy systems.

The FIS processes a mapping of given inputs to outputs by using the fuzzy sets theory (Zadeh, 1965). FIS is an appropriate approach to be applied to the real-world problems because FIS allows for the variables "partial true" and/or "partial false", which reflect the uncertainty nature in physical processes (Negnevitsky, 2011).

In general, FIS consists of five basic components as shown in Figure 2.2 (Córdon, 2011; Nayak et al., 2004). These components include Rule Base, Database, Fuzzy Inference Engine, Fuzzification and Defuzzification Interfaces. The Rule Base and Database component are also termed as the Knowledge Base of the fuzzy inference systems.

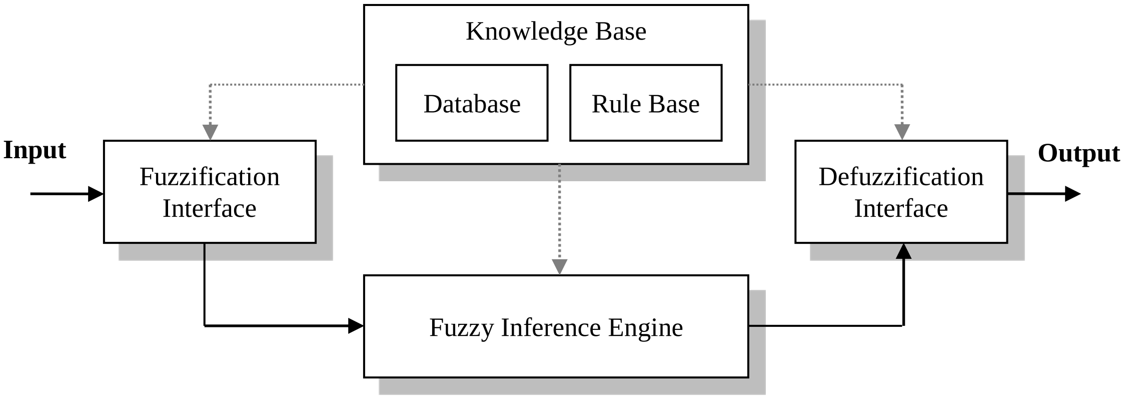


Figure 2.2. Five basic components of fuzzy inference systems.

Functions of these components are as follows: Rule Base involves IF-THEN rules for mapping the relationships between inputs and outputs of the system in the form of "IF *antecedent proposition* THEN *consequent proposition*". The Database is a collection of the fuzzy parameters or membership functions (MFs) for the input and output variables. The Fuzzification Interface component fuzzifies crisp inputs to fuzzy inputs and, on the other hand, the Defuzzification Interface component defuzzifies fuzzy outputs to crisp outputs. Finally, the Fuzzy Inference Engine derives a logical decision by using IF-THEN rules and handles the uncertainty by using MFs from the knowledge base.

In general, two typical approaches of FIS are used. They are the Mamdani-type FIS (MFIS) (Mamdani & Assilian, 1975) and the Sugeno-type FIS (SFIS) (Sugeno & Yasukawa, 1993). The difference between these FISs is the consequent part of fuzzy rules and how to defuzzify fuzzy sets outputs to crisp outputs. In MFIS, the fuzzy model is represented by linguistic rules with the following structure:

$$
\begin{aligned}
&Rule_i : IF\ x_1\ is\ A_{i,1}\ and\ \cdots\ x_n\ is\ A_{i,n} \\
&\qquad THEN\ y\ is\ B_i\ (i = 1, \ldots, L)
\end{aligned}
\tag{2.2}
$$

where $Rule_i$ denotes the i^{th} rule; L is the number of rules in the rule base; $x = (x_1, \ldots, x_n)^T$ and y are the inputs and output linguistic variables respectively; and, $A_{i,j}$ and B_i are the linguistic labels expressed as fuzzy sets that are specific to the system's behaviour. The MFIS defuzzifies output fuzzy sets by finding the centroid of a two-dimensional shape by integrating across a continuous variation function (see Appendix C).

In the SFIS, the consequent part of the fuzzy model is a linear equation and is represented in the following structure:

$$
\begin{aligned}
&Rule_i : IF\ x_1\ is\ A_{i,1}\ and\ \cdots\ x_n\ is\ A_{i,n} \\
&\qquad THEN\ y_i = a_{0i} + a_{1i}\, x_1 + \cdots + a_{ni}\, x_n\ (i = 1, \ldots, L)
\end{aligned}
\tag{2.3}
$$

where x and y are input and output variables respectively. Specifically, y is the local output set that determines local linear relationships by means of coefficients a_{ji}. The output of SFIS is in a form of singleton, a fuzzy set with unity membership grade at a singleton point and zero elsewhere on the universe of discourse. The output centroid is calculated by the weighted average method (see Appendix C).

In general, the MFIS is more intuitive and more well suited to human understanding than the SFIS, whereas SFIS works well with adaptive techniques and also guaranteed continuity of the output surface (Negnevitsky, 2011). The choice of the selected model is subject to the application's objectives. For example, in the case of system control or engineering applications, the SFIS seems to be more appropriate because such a model can provide smooth and continuous output. However, if the interpretability issue is taken into account, the MFIS seems to be more preferred than the SFIS model.

As mentioned, this thesis puts an emphasis on the interpretability of fuzzy systems (with acceptable accuracy). The MFIS is therefore selected as the base model in this thesis. Such a model will be applied to spatial interpolation and time series prediction as a solution to the interpretability problem.

One difficulty in establishing an interpretable fuzzy model is that accuracy and interpretability of the model can be contrasting objectives as presented in Figure 2.3 (Ishibuchi, 2007). To achieve higher accuracy, the fuzzy models can compromise the interpretability capability due to the increasing number of parameters in the models. On the other hand, in order to achieve higher interpretability, a fuzzy model may have reduced accuracy due to a decreased number of essential parameters in the models.

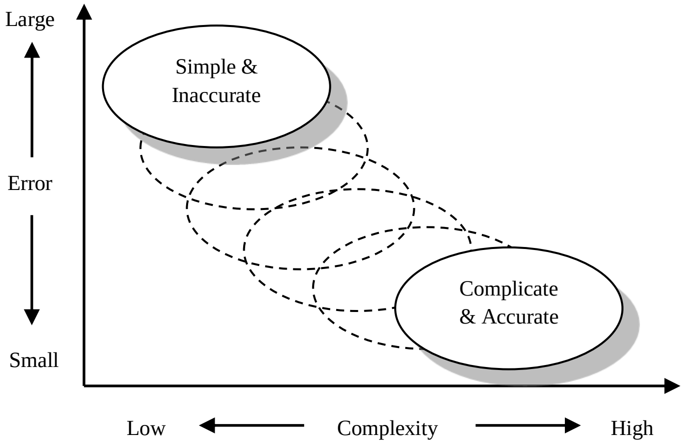


Figure 2.3. Contrasting problem in the interpretable fuzzy modeling (Ishibuchi, 2007)

Although establishing an interpretable fuzzy system is a difficult task, it can be seen as a challenging task as well. Figure 2.4 shows the conceptual relationship between the contrasting goals and the aim of this thesis. While this thesis aims to develop FIS models that are as close to the objective area as possible, criteria used for interpretability assessment have to be examined and established.

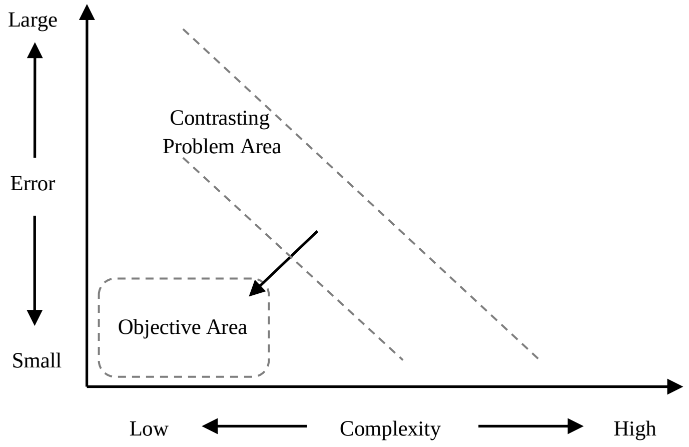


Figure 2.4. Contrasting problem area and the goal of this thesis.

2.6. Interpretability Criteria of Fuzzy Modeling

Interpretability of fuzzy systems has gained increased attention over the years (Córdon, 2011; Alcalá et al., 2006; Mikut et al., 2005; Casillas et al., 2003a, 2003b; Guillaume, 2001; Oliveira, 1999; Setnes et al., 1998a, 1998b). However, the semantic context of interpretability of fuzzy modeling has not been well defined. Consequently, qualitative justification of the interpretability of fuzzy modeling is rather arbitrary. Zhou and Gan (2008) proposed a unified framework to describe the interpretability of fuzzy modeling. This framework is re-presented here in Figure 2.5.

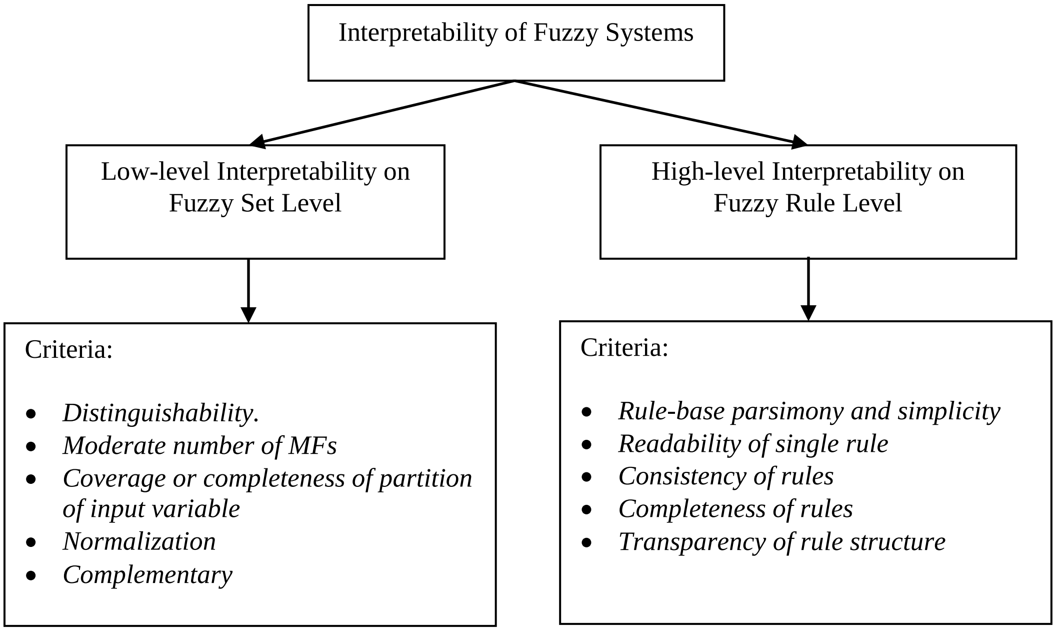

Interpretability of Fuzzy Systems
Low-level Interpretability on Fuzzy Set Level
High-level Interpretability on Fuzzy Rule Level
Criteria:
Distinguishability.
Moderate number of MFs
Coverage or completeness of partition of input variable
Normalization
Complementary
Criteria:
Rule-base parsimony and simplicity
Readability of single rule
Consistency of rules
Completeness of rules
Transparency of rule structure

Figure 2.5. A taxonomy of interpretability of fuzzy systems.

Their proposed framework categorizes fuzzy model interpretability into low-level and high-level criteria. The former criteria are defined on the fuzzy sets level, whereas the latter focus on the fuzzy rules level. As the framework distinguishes the interpretability of fuzzy systems into two levels, assessing the interpretability of the whole fuzzy systems can be possible (Alonso & Magdalena, 2011; Córdon, 2011; Alonso et al., 2009). This thesis therefore adopts this framework to assess the interpretable quality of fuzzy models established. The following are the brief contexts of these criteria.

2.6.1. Low-level Interpretability

- *Distinguishability*: In input space partitioning, fuzzy sets should be clearly defined in the distinctive ranges in the universe of discourse of variables. Each MF should be distinct enough from each other in representing a linguistic term with a clear semantic meaning.

- *Moderate number of MFs*: The number of MFs of a variable should not be selected arbitrarily, but they should be compatible with the number of conceptual entities a human can efficiently handle during the inferential activities. According to a popular suggestion in cognitive psychology, the number of different entities efficiently stored at the short-term memory should not exceed the limit of 7 ± 2 (Oliveira, 1999; Pedrycz et al., 1998).

- *Coverage or completeness of fuzzy partitioning*: The entire universe of discourse of a variable should be covered by the MFs generated, and every data point should belong to at least one of the fuzzy sets and have a linguistic representation. In other words, the membership value should not be zero for at least one of the linguistic labels.

- *Normalization*: Each MF of a variable is expected to represent a linguistic label with a clear semantic meaning. Thus, at least one data point in the universe of discourse should have a membership value equal to one, that is, MFs of a variable should be normal.

- *Complementary (optional)*: For each element of the universe of discourse, the sum of all its membership values should be equal to one. This guarantees uniform distribution of meaning among the elements. However, this requirement is only suitable for probability fuzzy systems. The possibility fuzzy systems do not consider this requirement.

2.6.2. High-level Interpretability

- *Rule base parsimony and simplicity*: The set of fuzzy rules must be as small as possible under the condition that the model performance is preserved at a satisfactory level. A large rule base would lead to a lack of global understanding of the system.

- *Readability of single rules*: The number of conditions in the premising part of the rule should not exceed the limit of 7 ± 2 distinct conditions, which is the number of conceptual entities a human can efficiently handle (Pena-Reyes & Sipper, 2003).

- *Consistency*: Rule base consistency means the absence of contradictory rules in the rule base in the sense that rules with similar promising parts should have similar consequent parts (Guillaume, 2001; Dubois et al., 1997).

- *Completeness*: For any possible input vector, at least one rule should be fired to prevent the fuzzy system from breaking inference (Guillaume, 2001).

- *Transparency of rule structure*: A fuzzy rule should characterize human knowledge or system behaviors in a clear way.

These low-level and high-level criteria suggest how the feature of interpretability in fuzzy systems can be assessed. However, to achieve all these criteria at the same time is not an easy task. Although some established fuzzy systems can satisfy all these criteria, it may not ensure that the accuracy of the established models is acceptable. This thesis aims to satisfy most of the criteria listed in Figure 2.5 and at the same time provides compatible results to other popular methods.

2.7. Conclusion

This chapter began with a review of the spatial interpolation and time series prediction techniques used in hydrological and related areas. The problems about the interpretability in hydrological models are highlighted in order to provide the background to the aims of this thesis. Next, the concepts of fuzzy systems have been introduced as a solution to the problems. This chapter also presented the interpretability criteria of fuzzy modeling to be used as a guideline to assess the interpretability quality of the established hydrological models in this thesis. Following this chapter, the next two chapters focus on the development of the interpretable fuzzy systems used for monthly rainfall spatial interpolation.

CHAPTER 3

AN INTERPRETABLE FUZZY SYSTEM FOR MONTHLY RAINFALL SPATIAL INTERPOLATION

3.1. Introduction

Interpretable data-driven models can enable human analysts to understand the nature of the data to be modeled. In the case of monthly rainfall spatial interpolation, interpretable models can provide human analysts with the understanding of the spatial distribution of rainfall data in a particular month. If interpretable data-driven models have been established in an appropriate way, the established models should be able to represent the spatial distribution of monthly rainfall data effectively via the interpretability advantage of the models.

Fuzzy systems have demonstrated their ability as an effective system identification tool (Ross, 2004). For an established fuzzy model, human analysts can interpret the nature of the data to be modeled through fuzzy sets and fuzzy rules, and this allows human analysts to enhance the model. The objective of this chapter is to propose a methodology to analyze and establish an interpretable fuzzy model for monthly rainfall spatial interpolation.

This chapter is organized as follows: Section 2 describes the case study area and datasets used in the chapter. Section 3 presents the proposed methodology to analyze and establish an interpretable fuzzy model for spatial interpolation of the monthly rainfall variable. Evaluation of the proposed methodology will be presented in Section 4. Finally, Section 5 is the conclusion for this chapter.

3.2. Case Study Area and Datasets

The case study area is located in the northeast region of Thailand. The distribution of rain gauge stations is shown in Figure 3.1. Eight months of spatial rainfall data are selected for eight case studies, that is, August 1998, September 1998, May 1999, September 1999, May 2000, August 2000, June 2001 and August 2001. These selected months are the months with relatively high rainfall for the year, and they also have a small amount of missing data. General information of the datasets (rainfall feature) is shown in Table 3.1.

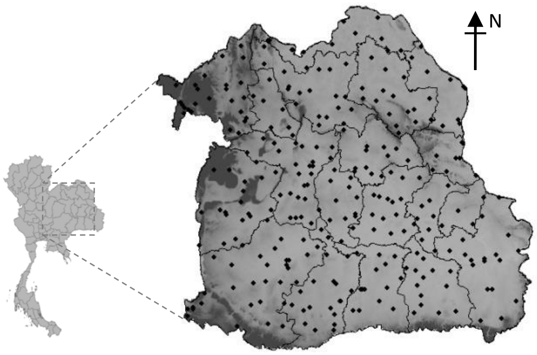


Figure 3.1. Case study area and the distribution of rain gauge stations.

In each case, 80 rain gauge stations (or approximately 30%) are randomly removed and are used to evaluate the established models. The dataset features comprise of information on the longitude (*x*), latitude (*y*), altitude (*v*) and the amount of monthly rainfall (*z*). The datasets are normalized by linear transformation for computational purposes. (Notice that the altitude feature is purposely used only in artificial neural network models in order to investigate the orographic effects of the study area.)

Table 3.1. General information (rainfall feature) of the eight case studies.

Statistics	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
Mean (mm.)	2325	2016	2448	2487	2669	2766	2397	3647
Standard Deviation	889	992	1213	1044	1175	1110	1369	1826
Kurtosis	1.529	2.548	0.983	2.835	0.162	1.261	1.902	0.332
Skewness	0.851	1.202	0.895	1.244	0.624	0.733	1.017	0.733
Minimum	380	184	60	455	48	368	90	341
Maximum	6118	7003	6912	7215	5956	7450	8207	10784
Training Data	198	195	200	197	200	197	188	178
Testing Data	80	80	80	80	80	80	80	80
Correlation	-0.047	-0.118	-0.001	0.007	-0.059	-0.241	-0.237	-0.046
R_x	0.121	0.235	0.229	0.134	0.396	0.323	0.638	0.408
R_y	0.033	0.032	0.563	0.309	0.127	-0.290	0.162	0.496

Note: R_x (and R_y) are correlation coefficients between the amount of rainfall and longitude (and latitude).

In Table 3.1, the row Correlation indicates the correlation coefficient between the amount of rainfall and the altitude of rain gauge stations. Since the correlation values are close to zero, it can be hypothesized that orographic effects are not strong in the study area (Goovaerts, 2000). Figure 3.2 shows an example of the scatter plot between the altitude and the amount of rainfall for Case 6, which has the highest magnitude of the correlation value. One can see that no linear relationship appears evident in the data. Thus, the altitude feature will not be used in the proposed models.

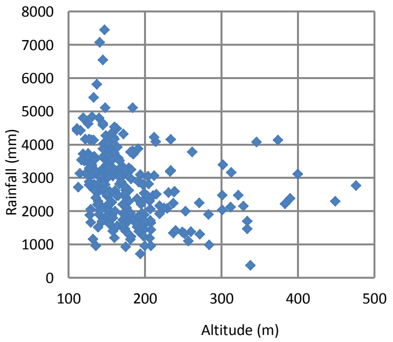


Figure 3.2. An example of the scatter plot between the altitude and amount of rainfall.

3.3. Establishing an Interpretable Fuzzy System

In general, several techniques have been proposed to establish interpretable fuzzy models. However, according to the recommendations by Zhou and Gan (2008), the "prototype-based fuzzy modeling" technique seems to be a good convenient approach to automatically construct both low-level and high-level interpretable fuzzy models in one model structure.

Such a technique makes use of a clustering method to partition data into important homogeneous regions (i.e. prototypes) that are characterized by multidimensional fuzzy sets. A rule is associated with each region (i.e. the premise part of each rule is a multi-dimensional fuzzy set). The MFs on individual variables can be obtained by projecting the multidimensional fuzzy set onto the corresponding antecedent individual variables.

This technique has been adopted in the proposed methodology due to its capability to effectively construct an interpretable fuzzy model in one process. Moreover, the structure of the fuzzy model constructed by this method is simple and flexible for further enhancement. However, a couple of prerequisite issues should be considered, that is, the type of clustering technique and how many clusters are required by the selected clustering technique.

3.3.1. Overview of the Proposed Methodology

The proposed methodology consists of four steps. Figure 3.3 illustrates an overview of the proposed methodology. The first step is to define the minimum number of clusters.

The second step is to determine the optimal number of clusters. The minimum number of clusters requires to localize the study area effectively. Once the optimal number of clusters is selected, a prototype-based MFIS model is created in step three by identifying the input-output mapping for each cluster. In the final step, the parameters of the created MFIS model are optimized. Overall, step one and step two can be considered as the clustering analysis, whereas step three and step four are the model establishment and optimization stages. The clustering analysis is necessary to determine the optimal number of cluster for the prototype-based fuzzy modeling. And the model's parameters (e.g. MFs) need to be optimized to provide better generalization of the model.

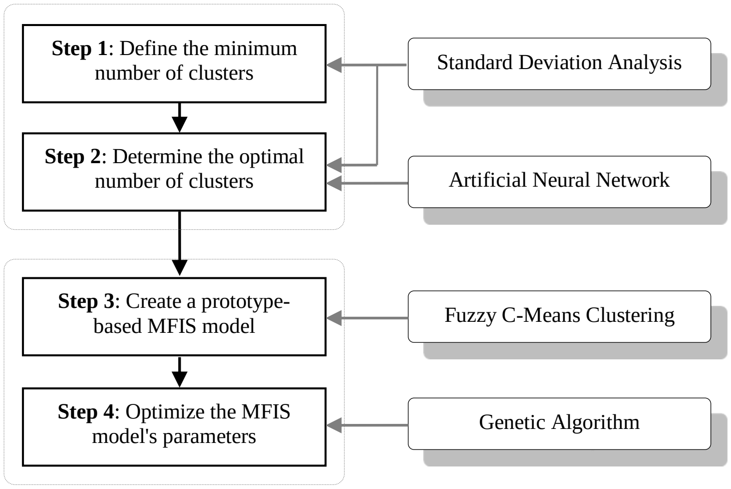

Step 1: Define the minimum number of clusters
Step 2: Determine the optimal number of clusters
Step 3: Create a prototype-based MFIS model
Step 4: Optimize the MFIS model's parameters
Standard Deviation Analysis
Artificial Neural Network
Fuzzy C-Means Clustering
Genetic Algorithm

Figure 3.3. An overview of the proposed methodology.

3.3.2. Fuzzy C-Means Clustering Analysis for Spatial Data

Clustering analysis for spatial data (or localization) is a step to partition a heterogonous global area into a group of homogenous local areas. In the case of spatial rainfall data, localization can be achieved by analyzing the rainfall pattern and the topography of the study area (Lee et al., 1998). However, if prior information about the study area is not known or difficult to analyze, clustering techniques are normally used (Wong et al., 2003).

Fuzzy c-means (FCM) clustering (Bezdek et al., 1984) are normally used to partition spatial data if there is uncertainty in determining the cluster boundaries (Hu et al., 2008). In many clustering algorithms, including FCM, first there is a need to know the number of clusters. However, there is no prior information about the number of clusters in general (Erilli et al., 2011).

To determine the optimal number of clusters, cluster validation indices are used. Some examples are: partition index (SC) (Bezdek, 1981), separation index (S) (Rezaee et al., 1998), Xie and Beni's index (XB) (Xie & Beni, 1991), Dunn's index (DI) (Dunn, 1973) and alternative Dunn's index (ADI) (Halkidi et al., 2001). However, these commonly-used FCM indices were developed for general purposes and were not specific to spatial data.

For spatial data, these indices do not take the characteristics of spatial data into account. Moreover, these indices sometimes show conflicting results, which makes it difficult to decide on what is the best number of clusters. As the proposed methodology adopts the FCM technique to localized global spatial data, a FCM validation method to determine the number of clusters for spatial data is firstly needed. Therefore, this thesis proposes two concurrent FCM cluster validation methods for spatial rainfall data. Assuming no prior knowledge of the number of clusters is available, using the two cooperative validation methods can make the decision more consistent.

The first method is based on statistical analysis whereas the second method is based on simulation. By using the cooperative method, the statistics-based method is proposed as the major criterion for determining the range of possible numbers of clusters. The simulation-based method is proposed as the decision support criterion to determine the best number of clusters.

3.3.2.1. The Statistics-Based Method

To analyze the spatial data, standard deviation (SD) is necessary to be used to estimate the variation of spatial data in the study area. Tutmez et al. (2007) suggested that the optimal number of clusters for FCM can be determined by:

$$\text{Minimize } n_c \text{ under, } \mathrm{Std}[z(x)] \approx \mathrm{Std}[z(c)] \tag{3.1}$$

where n_c is the optimal number of cluster, Std is the standard deviation, $z(x)$ are the observed values of the dataset and $z(c)$ are the observed values at the cluster centers (the computed central z values from all cluster centers). In this criterion the numbers of clusters are plotted against Std[$z(c)$]. The number of clusters satisfying constraint (3.1) is retained as the optimal number.

However, this method may not be appropriate if it is applied directly to the data using the global scale because the optimal cluster number identified can be too small. Therefore, a pre-conditional criterion to determine the minimum number of clusters before using Tutmez's method is added. This criterion analyzes the proportion between the mean of standard deviation of rainfall value (z) in all clusters and the number of clusters. The steps to determine the optimal number of clusters are listed as follows:

Step 1. Use FCM to partition data into *n* clusters. *n* starts from 2 to C_{max}, where C_{max} is the user-defined maximum number of clusters. C_{max} can range from 2 to any appropriate number as far as the model interpretability is concerned.

Step 2. For each number of cluster *n*:

Step 2a. For each cluster *i*, calculate SD_i, SD of rainfall values (*z*) in cluster *i*, and then calculate the proportion, P_n, between the mean of all SD_i and the number of clusters by

$$P_n = \mathrm{E}(SD_i) / n \tag{3.2}$$

Step 2b. Calculate the difference, D_n, between P_{n-1} and P_n by

$$D_n = P_{n-1} - P_n \tag{3.3}$$

Plot D_n against the number of clusters; the minimum number of clusters can be retained from the point that D_n becomes stable.

Step 3. For each *n*, calculate SD of $z(x)$ and $z(c)$ and plot it against *n*. Under the defined range, the appropriate number of clusters can be retained by constraint (3.1).

3.3.2.2. The Simulation-Based Method

The work of Erilli et al. (2011) proposed the use of ANN to determine the number of clusters for FCM by investigating the training performance (training error) when the network's inputs are the input-output pairs and network's output is the assigned cluster number. They showed that at the point when training performance leap up, this is an indication of the appropriate number of clusters. However, they did not clearly specify the architecture of the ANN used, and varying parameter values can affect the performance of the ANNs. Consequently, the leap up point may change if those parameters are changed.

However, the idea of using ANN to determine the number of clusters can be further investigated. In this thesis, another way of employing ANN to investigate the appropriate number of clusters will be proposed. In this method, one hidden layer BPNN is used. The steps of the proposed methods to determine the optimal number of clusters are listed as follows:

Step 1. Prepare the data matrix, input-output pairs. Use FCM to partition data into n clusters. n starts from 2 to C_{max}.

Step 2. For each number of cluster *n:*

Step 2a. Prepare the training data where the network's inputs are data matrix and the network's output is assigned a cluster number.

Step 2b. For $j = N_{min}$ to N_{max}, train BPNN with j hidden nodes, then evaluate the training performance of each BPNN.

Step 2c. Calculate the average value of the training performance of all trained BPNN, $Perf_n$.

Step 3. For each number of cluster n: calculate the performance proportion of the number of cluster n by

$$E_n = Perf_n / n \tag{3.4}$$

where E_n is the performance proportion of cluster n. The lower E_n indicates the more appropriate number of clusters.

The statistics-based method and the simulation-based method have been presented so far and cooperation of these two methods should alleviate the difficulty in the decision making stage. In practice, the lack of prior knowledge about the data on hand can make the modeling process difficult. Human analysts have to make the decision based on their own experiences. In the spatial interpolation method, fitting experimental semivariograms in kriging methods is an example of this subjectivity.

In clustering analysis, selection of the number of the clusters is a rather subjective task. Sometimes, using only the statistics-based method may not be enough to provide a confident selection. With the assistance of the intelligent technique of the simulation-based method, the selection can be improved and the optimal number of clusters is selected more confidently. This is the reason why two validation methods are used cooperatively in this thesis.

3.3.3. Model Establishment and Optimization

Once the number of clusters is determined, a prototype-based MFIS will be generated from the training data by using FCM. The clustering analysis in the previous two steps are important because if an appropriate clustering validation index for determining the number of prototypes is applied, FCM can generate a parsimonious rule base (Zhou & Gan, 2008). As two proposed validation indices are specifically developed for the spatial data, the MFIS generated should be related to a parsimonious rule base.

The MFs used are a Gaussian function because it provides smooth surfaces and has a low degree of freedom (Hameed, 2011). After the MFIS has been generated, it is then optimized by the GA (Holland, 1975) to improve interpolation accuracy. The chromosome of the GA consists of the sequence of input 1, input 2 and output, respectively. In turn, the inputs and output are the sequence of MFs which consist of two parameters (sigma and center).

The fitness function to be minimized is the sum square error (SSE) between the observed value (z) and the interpolated value (z') of the training data and it is given as

$$SSE = \sum_{i=1}^{S}(z_i' - z_i)^2 \tag{3.5}$$

An important point of optimization is how to control the diversity of individuals. In this process, the MFIS parameters are allowed to vary in certain controlled regions in order to preserve the structure of the prototypes of the generated MFIS.

Let α and β be user-defined control parameters, the center (c) parameters are allowed to vary in the range of [c - α, c + α] and the sigma (σ) parameters are allowed to vary within the range of [σ - β, σ + β] (Cordón et al., 2001; Ishibuchi et al., 1994). One benefit of these small controlled regions is that the structures of the prototypes for the optimized fuzzy systems will not be distorted too much from the original ones.

So far, the proposed methodology has been presented. In summary, such methodology consists of (i) the clustering analysis process and (ii) the model establishment process. In the next section, the proposed methodology will be evaluated in eight case studies.

3.4. Evaluation of the Proposed Methodology

In this section, the proposed methodology will be evaluated. In terms of the quantitative aspect (accuracy), as the established model works as the global method it will be compared to the commonly-used global spatial interpolation methods, that is, TSA, BPNN and RBFN. The orographic effects will also be tested with BPNN and RBFN. Besides this, the established model will be compared with ANFIS. The ANFIS model herein is an adaptive SFIS model generated by FCM with the same number of clusters from the established model. In terms of the qualitative aspect (interpretability), the interpretability criteria as mentioned in Chapter 2 will be used for the assessment.

3.4.1. Models Establishment

In the establishment process, the third order polynomial equation was adopted for TSA because it has both hill and valley surfaces which are suitable for most real world data (Chang, 2006). The one hidden layer network was adopted for BPNN (Piazza et al., 2011; Kajornrit et al., 2011) and the number of hidden nodes in RBFN was aligned with the number of training data (Lee et al., 1998).

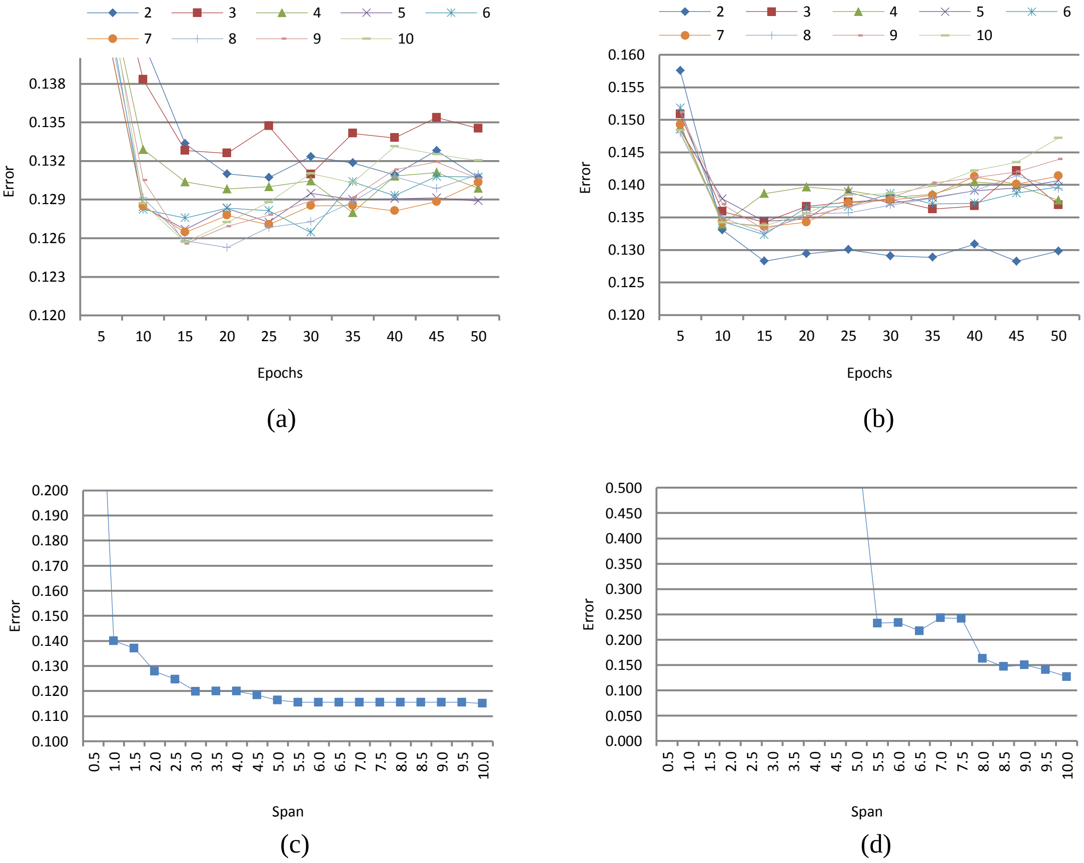
BPNN and RBFN were divided into two groups according to the inputs fed into the models. The first group (i.e. $BPNN_2$ and $RBFN_2$) used only spatial coordinates (x, y) as the inputs to the models. The second group (i.e. $BPNN_3$ and $RBFN_3$) used spatial coordinates (x, y) and altitude (v) as the model's inputs. This was to test the orographic effects of the study area.

The appropriate number of parameters for BPNN and RBFN were selected from the *K*-folds cross-validation method (Gilardi & Bengio, 2000). Briefly, the calibration data are divided into five partitions (or about 20%). For each partition *n*, $1 \leq n \leq 5$, a model is trained with other partitions and partition *n* is used for validation. The lowest average error from all *n* is used to indicate the appropriateness of the model's parameters.

Figure 3.4 shows an example of the *K*-folds cross validation process. The numbers in Figure 3.4(a) and Figure 3.4(b) are the numbers of hidden nodes of BPNNs. For BPNN, the parameter is the number of hidden nodes and training epochs. For RBFN the parameter is the span value (Beale et al., 2011). In Figure 3.4(a), for example, BPNN_2 showed the minimum error when the number of hidden node is 8 and the epoch is 20. And in Figure 3.4(c), the error of RBFN_2 becomes stable after span is 5.5. These values were selected as the models' parameters and there are shown in Table 3.2.

Figure 3.5 shows the results of the clustering analysis. The results of Case 1 to Case 8 are arranged from top to bottom, respectively. Graphs in the left-hand side are the results from the statistics-based method and graphs in the right-hand side are the results from the simulation-based method. Notice that, for graphs in Figure 3.5, P_n and D_n use the left-hand side axis for reference, whereas std(*z*(*c*)) and std(*z*(*x*)) use the right-hand side axis for reference. In addition, the results from the commonly-used FCM validation indices aforementioned are shown in the table in Appendix A.

In the experiments, C_{max} was set to 10. C_{max} was set to 10 in order to preserve the interpretability of the established fuzzy models. For the simulation-based method, N_{min} was set to 4 and N_{max} was set to 12. N_{max} was set to 12 because at the N_{max} greater than 12, the variations of training performance are too small to be observed. Some examples of clustering analysis are shown as follows.



(a)

(b)

(c)

(d)

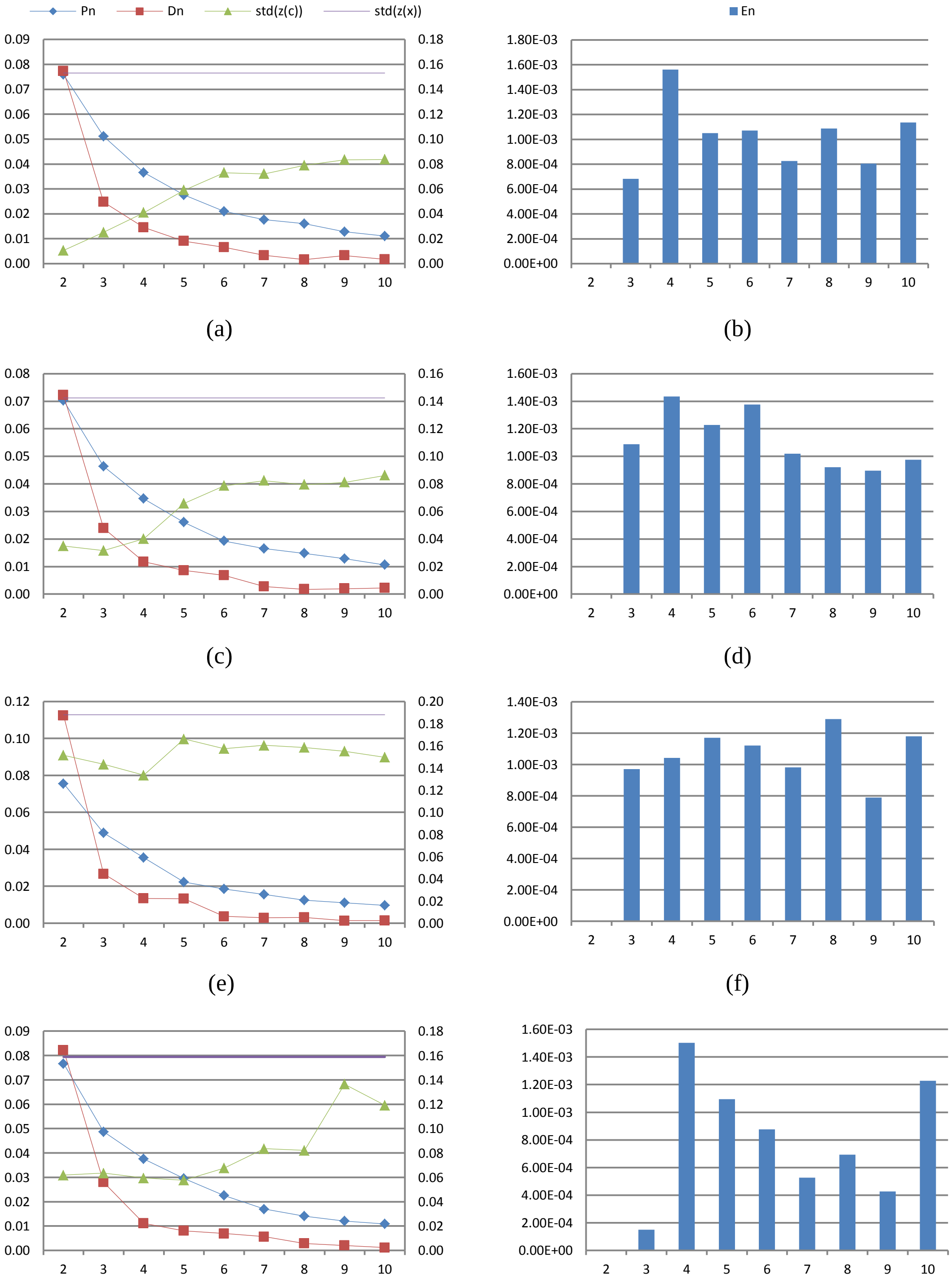
Note: (a) BPNN_2, (b) BPNN_3, (c) RBFN_2 and (d) RBFN_3.

Figure 3.4. An example of *K*-folds cross-validation method (Case 1).

Table 3.2. Summary of the number of parameters used in ANNs for each case study.

Case	BPNN_2	BPNN_3	RBFN_2	RBFN_3	Case	BPNN_2	BPNN_3	RBFN_2	RBFN_3
1	8 / 20	2 / 15	5.5	8.0	5	9 / 15	10 / 15	4.5	4.5
2	5 / 15	2 / 50	5.5	6.0	6	10 / 15	2 / 50	6.0	6.0
3	7 / 20	9 / 15	5.0	6.0	7	8 / 15	2 / 35	4.5	6.0
4	9 / 25	8 / 20	5.0	5.0	8	8 / 15	2 / 40	2.5	5.5

Note: BPNN = hidden nodes epoch, RBFN = span



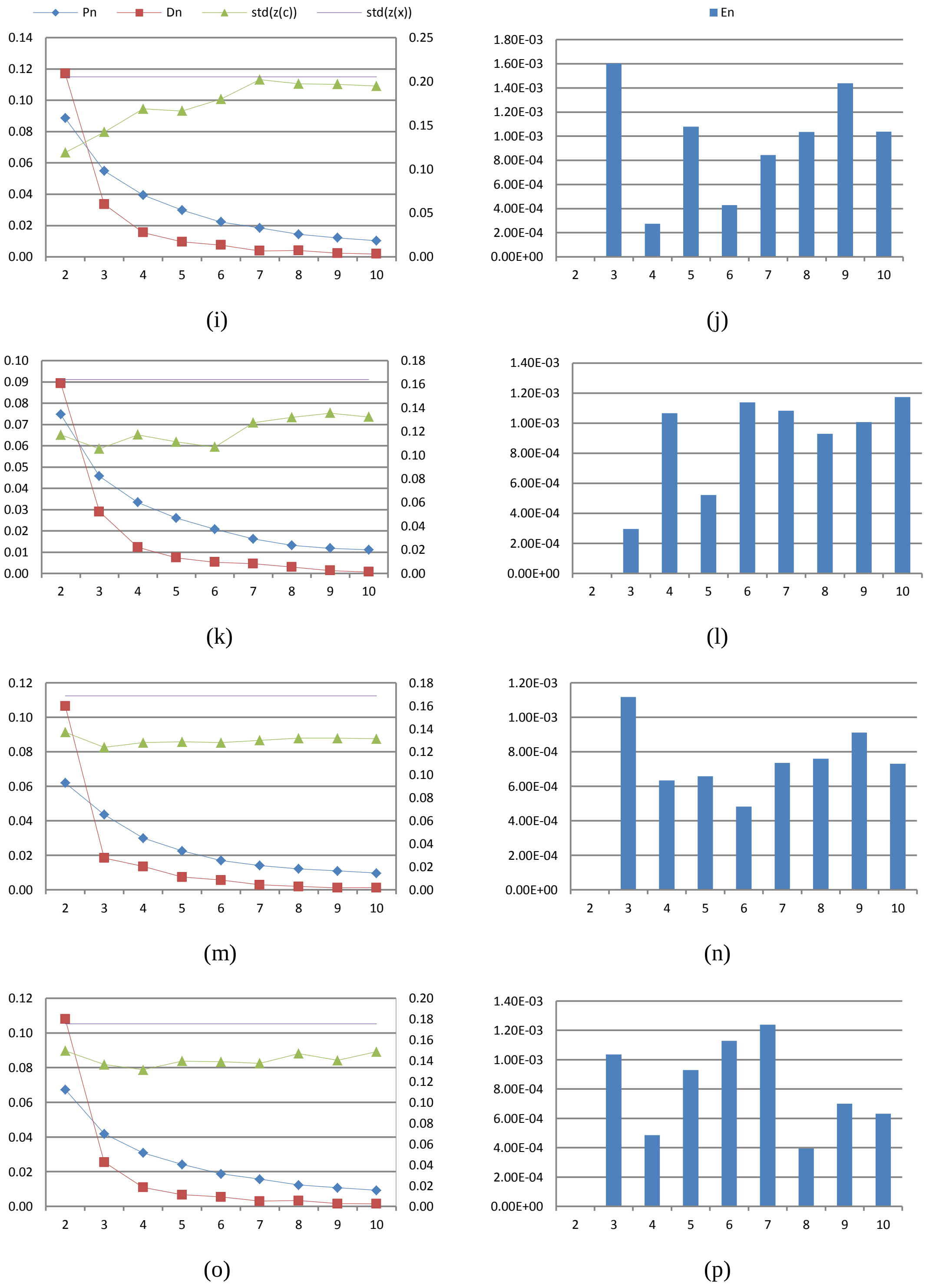
(a) (b)

(c) (d)

(e) (f)

(g) (h)

Figure 3.5. The results from the clustering analysis (Case 1 to Case 4).



(i)

(j)

(k)

(l)

(m)

(n)

(o)

(p)

Figure 3.5. (cont.) The results from the clustering analysis (Case 5 to Case 8).

In Case 1, Figure 3.5(a), D_n becomes stable at n = 7 and the difference between std(z(x)) and std(z(c)) continues to decrease until n = 6, and small variations are shown after that. Therefore, n = 7 is selected as the cluster number. In Fig 3.5(b), n = 7 and 9 shows relatively low E_n. It can be seen that n = 7 indicates the appropriate number of the clusters from both methods.

According to the table in Appendix A, the numbers of clusters determined from SC to ADI are 5, 5, 7, 8 and 8, respectively. At n = 5, SC becomes stable and S is the minimum. XB and ADI become stable at n = 7 and 8, respectively. DI is the maximum at n = 8. In this case, XB gives the same results to the proposed methods.

In Case 2, Figure 3.5(c), since D_n becomes stable at n = 7, the minimum number of clusters should be considered after this number. The difference between std(z(c)) and std(z(x)) shows small variation after n = 6 and has a minimum value at n = 7. Therefore, n = 7 is selected as the appropriate number of clusters. In Figure 3.5(d), after n = 7, E_n shows relatively small values. To keep the model from being too complex, n = 7 is selected for this case study.

According to the table in Appendix A, the numbers of clusters determined from SC to ADI are 7, 4, 8, 8 and 5, respectively. The SC value decreases and becomes stable at n = 7 and S value shows the minimum value at n = 4. At n = 8, XB shows the minimum value and DI shows the maximum value. ADI decreases and becomes stable at n = 5. The selected n in this case study is the same number as the SC index.

Table 3.3 summarizes the results from the clustering analysis. The selected numbers of clusters from Case 1 to Case 8 are 7, 7, 7, 9, 7, 8, 7 and 8, respectively. The last column of Table 3.3 shows the commonly-used FCM indices that provide the equivalent results to the selected number of clusters (see Appendix A).

Table 3.3. Summary of the results from the clustering analysis.

Case	**Defined Range**	**Statistics-based method**	**Simulation-based method**	**Selected *n***	**Commonly-used FCM indices**
1	n ≥ 7	n ≥ 7	n = 7	7	XB
2	n ≥ 7	n ≥ 7	n ≥ 7	7	SC
3	n ≥ 6	n ≥ 6 or n = 7	n = 7 or n = 9	7	DI, ADI
4	n ≥ 8	n = 9	n = 9	9	SC, S, DI
5	n ≥ 7	n = 7	n = 7	7	SC, S
6	n ≥ 8	n ≥ 8	n = 8	8	S, XB
7	n ≥ 7	n ≥ 7	n = 7	7	SC
8	n ≥ 7	n ≥ 7 or n = 8	n = 8	8	S, SC

Note: commonly-used FCM indices are SC, S, XB, DI and ADI, respectively.

In addition, the results from Case 7 and Case 8 point to the drawbacks of Tutmez's method if it is used directly. In Figure 3.5(m) and Figure 3.5(o), as the variation of the differences between std($z(c)$) and std($z(x)$) are small, the selected number of clusters can be too small for the global data (i.e. $n = 2$).

Once the number of clusters was determined, the prototype MFIS was generated from the FCM. The number of fuzzy rules and fuzzy sets were aligned with the number of clusters. In the optimization process, the parameters α and β were set to 0.1 and 0.05 respectively. These values are approximately 10 percent of the universe of discourse. As stated, the MFIS parameters were allowed to vary in certain controlled ranges. These controlled ranges are defined based on the reason that cluster centers should be varying inside its cluster in order to prevent indistinguishability of the optimized MFIS.

In this experiment, the number of the population of the GA was set to 200 (or about four times the number of genes in the chromosome) to ensure that there are at least four individuals for each parameter. The number of generations was set to 150, where the SSE became stable and did not show any considerable improvement. At this number of generations the best fitness value and average fitness values were met. From now on, the proposed model is called Genetic algorithm with fuzzy inference system (GAFIS).

In addition to GAFIS, the SFIS model is also generated with the same number of clusters using the FCM method. This SFIS will be used as the initial model for the ANFIS algorithm. The number of training epochs for ANFIS is approximately observed from the *K*-fold validation method in similar manner as the BPNN and RBFN. The approximate training epochs from Case 1 to Case 8 are 30, 15, 20, 10, 45, 25, 15 and 30, respectively, and are shown in Figure 3.6.

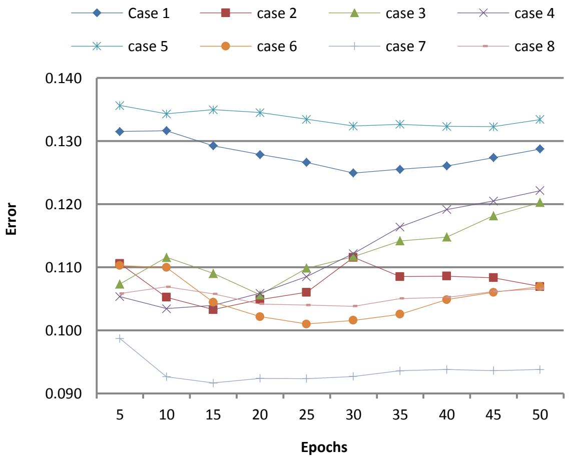


Figure 3.6. Results from the *K*-fold validation method for ANFIS models.

3.4.2. Quantitative Results

To evaluate the quantitative results, four quantitative measures have been adopted, that is, mean error (ME) or bias error, mean absolute error (MAE), root mean square error (RMSE) and correlation coefficient (R). These error measures are normalized by the mean values of the datasets for comparison purposes. These quantitative results are shown in Tables 3.4 to 3.7.

The Average rows in these tables are the average values from all cases. The Improvement rows in these tables are the improvement percentage based on TSA. These two values are also presented in Figure 3.7. Notice that the average values in the figures use the left axis for reference and the improvement values in the figures use the right axis for reference.

Table 3.4. Normalized mean error (bias error).

Case	TSA	BPNN$_2$	BPNN$_3$	RBFN$_2$	RBFN$_3$	ANFIS	GAFIS
1	0.026	0.000	0.028	0.023	0.045	-0.012	0.013
2	-0.018	-0.054	-0.057	-0.033	-0.053	-0.052	-0.049
3	0.077	0.049	0.014	0.060	0.069	0.056	0.049
4	0.074	0.082	0.070	0.059	0.110	0.075	0.054
5	0.013	0.024	0.022	-0.003	0.015	0.011	-0.007
6	-0.008	-0.011	-0.004	-0.001	-0.023	-0.022	-0.013
7	-0.002	0.004	0.009	0.005	0.020	0.008	0.007
8	-0.032	-0.028	-0.028	-0.019	0.000	-0.019	-0.023
Average	0.0313	0.0315	0.0290	0.0256	0.0419	0.0319	0.0268
Improvement	-	-0.89	7.20	18.12	-34.04	-1.92	14.10

Table 3.5. Normalized mean absolute error.

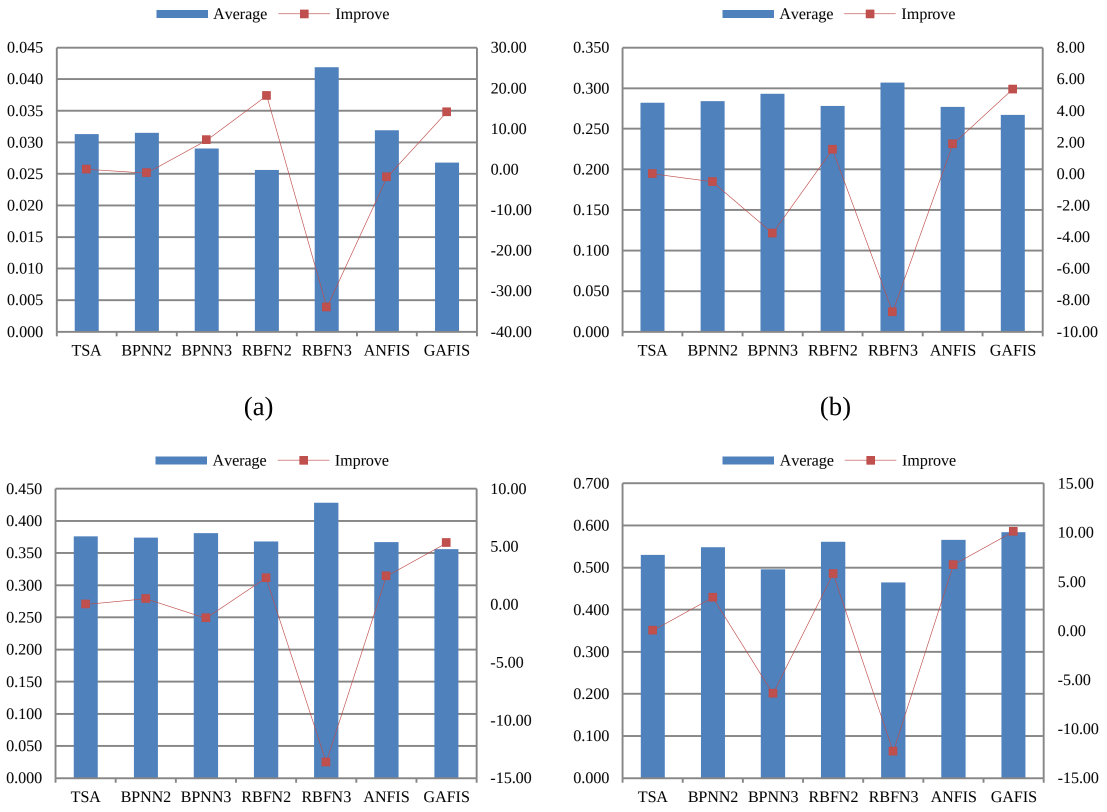
Case	TSA	BPNN$_2$	BPNN$_3$	RBFN$_2$	RBFN$_3$	ANFIS	GAFIS
1	0.279	0.300	0.307	0.286	0.313	0.286	0.284
2	0.330	0.336	0.354	0.336	0.368	0.333	0.322
3	0.253	0.236	0.258	0.227	0.280	0.223	0.224
4	0.301	0.315	0.303	0.314	0.325	0.307	0.284
5	0.286	0.277	0.283	0.267	0.287	0.269	0.266
6	0.282	0.276	0.287	0.266	0.288	0.290	0.267
7	0.270	0.275	0.300	0.266	0.286	0.263	0.256
8	0.256	0.254	0.250	0.260	0.306	0.242	0.233
Average	0.282	0.284	0.293	0.278	0.307	0.277	0.267
Improvement	0.00	-0.51	-3.77	1.54	-8.76	1.88	5.35

Table 3.6. Normalized root mean square error.

Case	TSA	BPNN$_2$	BPNN$_3$	RBFN$_2$	RBFN$_3$	ANFIS	GAFIS
1	0.368	0.368	0.395	0.361	0.388	0.352	0.366
2	0.419	0.421	0.440	0.420	0.481	0.414	0.403
3	0.327	0.318	0.329	0.309	0.373	0.296	0.290
4	0.408	0.416	0.398	0.421	0.532	0.419	0.396
5	0.370	0.365	0.359	0.347	0.382	0.359	0.348
6	0.367	0.365	0.371	0.353	0.379	0.372	0.353
7	0.398	0.393	0.407	0.388	0.454	0.393	0.369
8	0.353	0.349	0.345	0.345	0.432	0.333	0.326
Average	0.376	0.374	0.381	0.368	0.428	0.367	0.356
Improvement	0.00	0.49	-1.18	2.28	-13.64	2.43	5.31

Table 3.7. Correlation coefficient.

Case	TSA	BPNN$_2$	BPNN$_3$	RBFN$_2$	RBFN$_3$	ANFIS	GAFIS
1	0.357	0.383	0.067	0.404	0.382	0.472	0.370
2	0.628	0.612	0.542	0.616	0.409	0.611	0.657
3	0.674	0.691	0.692	0.712	0.650	0.734	0.749
4	0.262	0.298	0.343	0.245	0.195	0.245	0.315
5	0.427	0.474	0.467	0.534	0.433	0.516	0.537
6	0.461	0.478	0.428	0.515	0.449	0.470	0.513
7	0.720	0.730	0.703	0.736	0.648	0.729	0.769
8	0.712	0.720	0.728	0.725	0.554	0.749	0.760
Average	0.530	0.548	0.496	0.561	0.465	0.566	0.584
Improvement	0.00	3.38	-6.41	5.79	-12.31	6.70	10.08



(a)

(b)

(c)

(d)

Note: (a) normalized mean error, (b) normalized mean absolute error, (c) normalized root mean square error, and (d) correlation coefficient.

Figure 3.7. Plots of the average and improvement values.

First of all, the orographic effects have been confirmed through the interpolation accuracy of BPNN and RBFN. Overall, BPNN$_2$ and RBFN$_2$ provided better accuracy than BPNN$_3$ and RBFN$_3$. It is clear that the altitude feature can affect the accuracy of the models and it can also cause the models to provide inconsistent results in most cases. However, one issue was observed, that is, the robustness between BPNN and RBFN. BPNN is more tolerant to the irrelevant input than RBFN in these case studies.

In terms of ME, in Table 3.4 Cases 2, 3, 4, 6 and 8 show a similar bias direction. In Case 1, ANFIS tends to provide a negative bias whereas the others give a positive bias. In Case 5, RBFN$_2$ and GAFIS show a small negative bias whereas the others show a positive bias. In Case 7, only TSA shows a small negative sign value. However, these bias values do not show any uncommon conditions in general.

In the row Average, the average values of the absolute ME are calculated from all cases. The interpolation quality is better if the bias is close to zero. Therefore, interpolation quality can be ordered as RBFN$_2$ > GAFIS > TSA > BPNN$_2$ > ANFIS. However, these measures are not sufficient enough to justify the interpolation accuracy. It mainly tends to point out the tendencies of bias error.

In terms of MAE, in Table 3.5 GAFIS provides the best interpolation accuracy in five cases (2, 4, 5, 7 and 8). TSA gives the best accuracy in Case 1. ANFIS provides the best accuracy in Case 3 and RBFN$_2$ provides the best accuracy in Case 6. However, GAFIS also provides compatible accuracy at the best models in these two cases. Overall, the interpolation accuracy can be ordered as GAFIS > ANFIS > RBFN$_2$ > TSA > BPNN$_2$.

In terms of RMSE, in Table 3.4 GAFIS provides the best interpolation accuracy in five cases (i.e. 2, 3, 4, 7 and 8). ANFIS provides the best accuracy in Case 1, whereas RBFN$_2$ provides the best accuracy in Cases 5 and 6. However, GAFIS provides compatible accuracy to RBFN$_2$ in Case 6 as well. Overall, based on average results, the interpolation accuracy can be ordered as GAFIS > ANFIS > RBFN$_2$ > BPNN$_2$ > TSA.

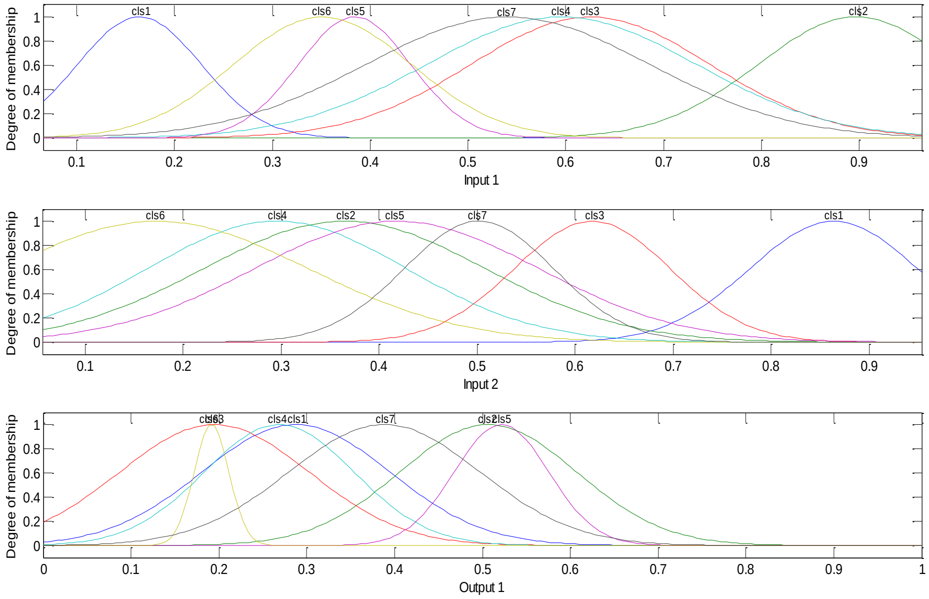
In terms of R, in Table 3.5 the best interpolator comes from GAFIS in six cases (i.e. 2, 3, 4, 5, 7 and 8). $RBFN_2$ provides the best accuracy in Case 6 and ANFIS provides the best accuracy in Case 1. Based on average values, the interpolation accuracy can be ordered as GAFIS >ANFIS > $RBFN_2$ > $BPNN_2$ > TSA. In summary, based on these quantitative measures, GAFIS can provide satisfactory interpolation accuracy when compared with commonly-used global methods.

3.4.3. Qualitative Results

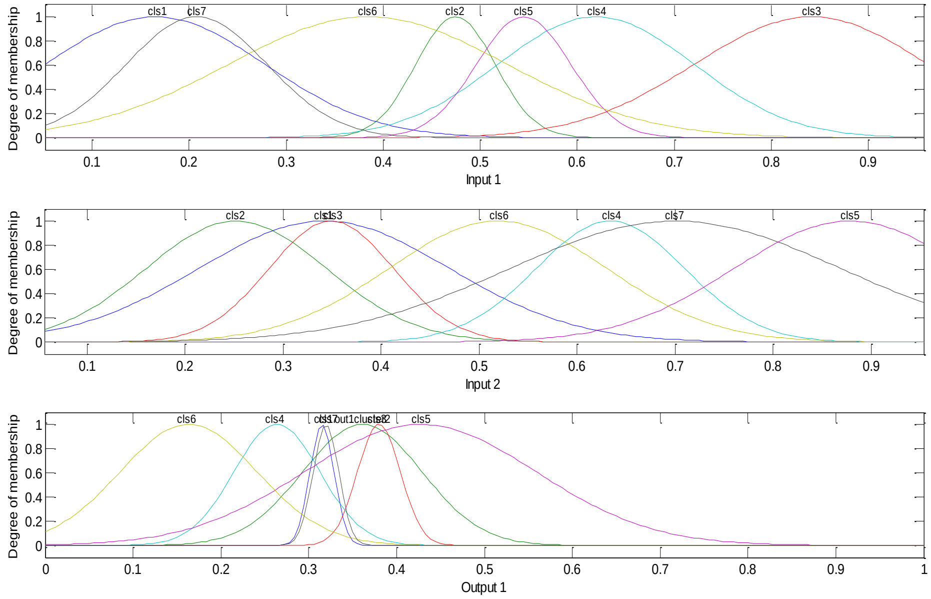
Figure 3.8 shows an example of the fuzzy rules of GAFIS models. One fuzzy rule is associated with one prototype. Figure 3.9 shows the optimized fuzzy parameters (MFs) of GAFIS models from Case 1 to Case 8. In each figure, MFs of input 1 (*longitude*), input 2 (*latitude*) and output (*rainfall*) have been presented.

IF longitude (x) = cls1 **AND** latitude (y) = cls1 **THEN** rainfall (z) = cls1
IF longitude (x) = cls2 **AND** latitude (y) = cls2 **THEN** rainfall (z) = cls2
IF longitude (x) = cls3 **AND** latitude (y) = cls3 **THEN** rainfall (z) = cls3
IF longitude (x) = cls4 **AND** latitude (y) = cls4 **THEN** rainfall (z) = cls4
IF longitude (x) = cls5 **AND** latitude (y) = cls5 **THEN** rainfall (z) = cls5
IF longitude (x) = cls6 **AND** latitude (y) = cls6 **THEN** rainfall (z) = cls6
IF longitude (x) = cls7 **AND** latitude (y) = cls7 **THEN** rainfall (z) = cls7

Figure 3.8. An example of fuzzy rules of the GAFIS model (Case 1).

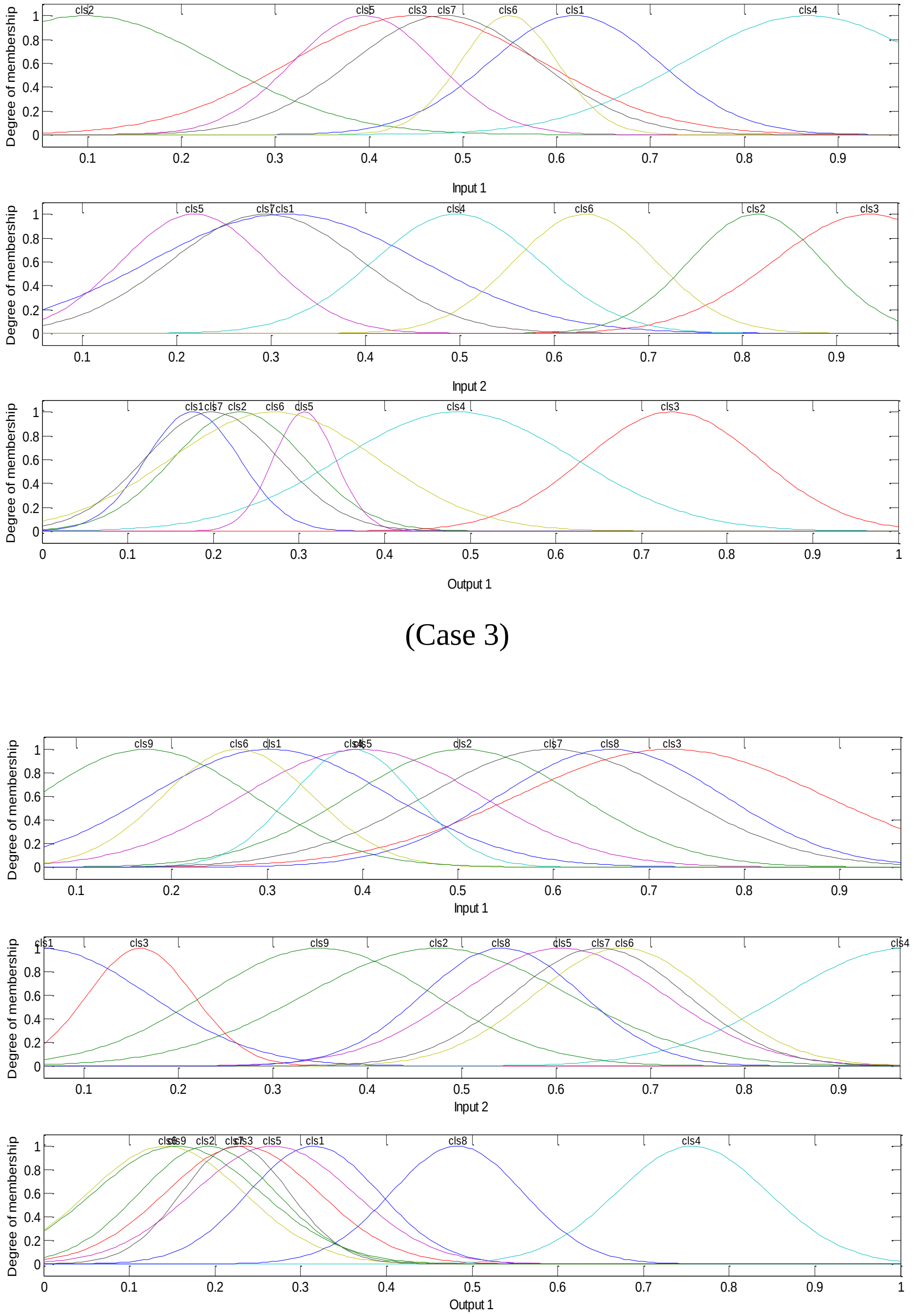


(Case 1)



(Case 2)

Figure 3.9. Optimized membership functions of the GAFIS models.

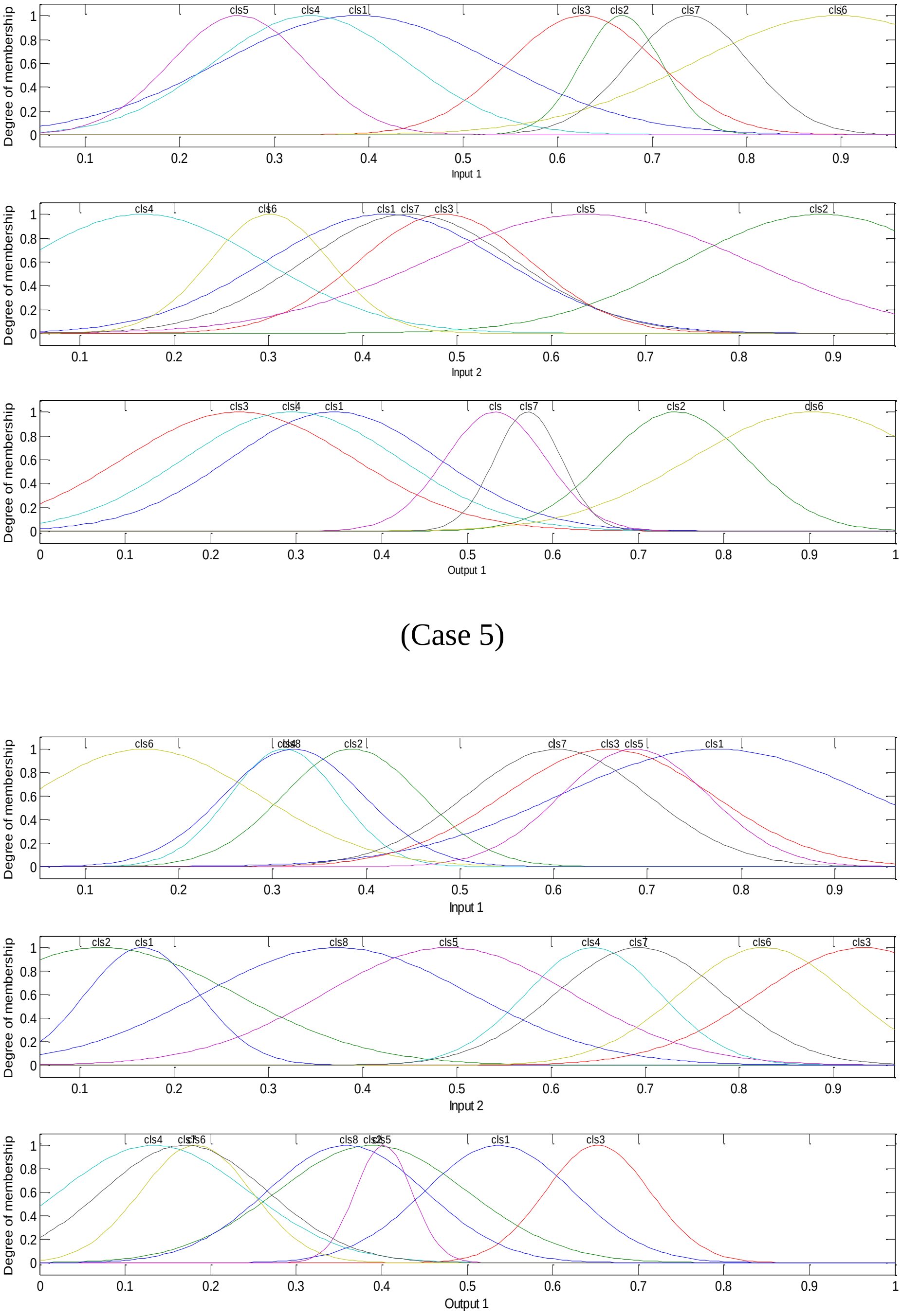


(Case 3)

(Case 4)

Figure 3.9. (cont.) Optimized membership functions of the GAFIS models.

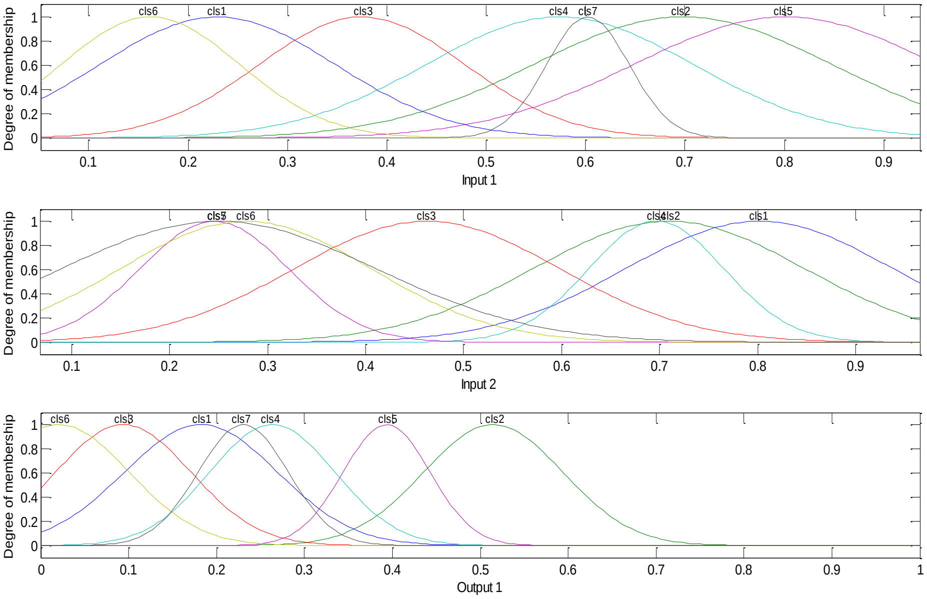
.



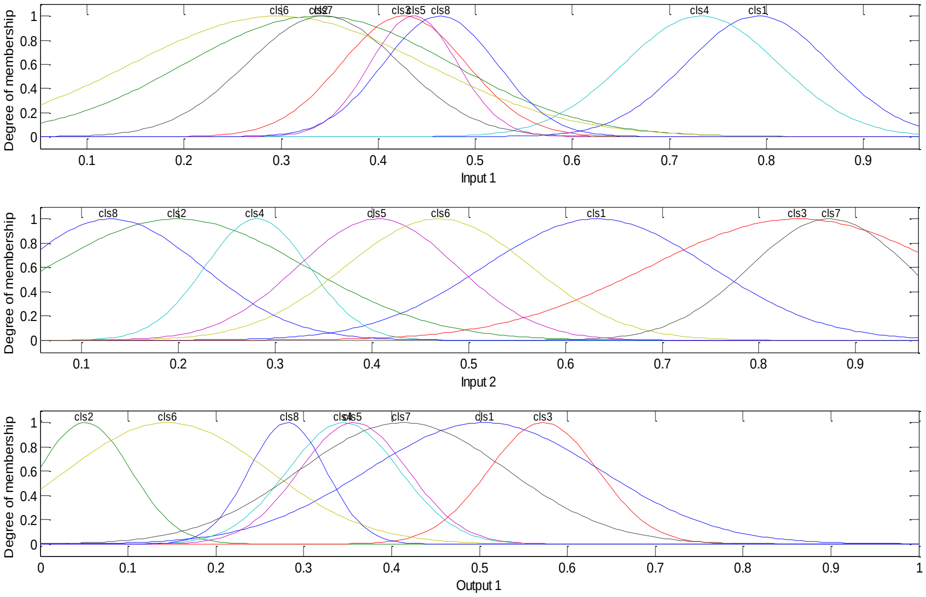
(Case 5)

(Case 6)

Figure 3.9. (cont.) Optimized membership functions of the GAFIS models.



(Case 7)



(Case 8)

Figure 3.9. (cont.) Optimized membership functions of the GAFIS models.

In terms of low-level interpretability (fuzzy sets level), MFs of GAFIS mostly satisfied *distinguishability* criterion. The ranges between two consecutive MFs are generally distinct enough to represent linguistic terms of the MFs. Actually the inputs of the models represent the spatial locations of the study area (i.e. longitude and latitude). Two inputs must be considered together to specify the locations. Although, the ranges of MFs in one input are not distinct enough, it does not mean that those MFs are not able to represent the spatial locations. For example, in Case 8 some MFs in the input 1 are not distinct but most MFs in the input 2 are distinct; thus, these two-dimensional MFs can represent the different spatial locations.

GAFIS models satisfied the moderate number of MFs criterion. The maximum numbers of MFs should not exceed 7 ± 2 (Oliveira, 1999; Pedrycz et al., 1998) in each input dimension. GAFIS models also satisfied normalization criterion, in which each MF has at least one point that has a membership value equal to one. GAFIS models showed satisfactory coverage of fuzzy partitioning criterion. The entire input space of the models has been covered by at least one MF. In the output space, four models satisfied this criterion; however, the other models appeared to be lacking MFs in the high rainfall. This situation can occur if few peak rainfall values appear in the rainfall data. These few values can be considered as outliers and can cause FCM to disregard these values in the establishment process.

In terms of high-level interpretability (fuzzy rule level), GAFIS models satisfy the rule base parsimonious and simplicity criterion. The global understanding of the models can be achieved in seven to nine fuzzy rules. The readability of single rules is considerably high since there are only two antecedent conditions in each rule. Since the GAFIS models are generated from the prototype-based method, so the completeness and consistency of fuzzy rule are qualified. No contradictory fuzzy rule appears in the system. As one fuzzy rule is aligned with one cluster, so at least one fuzzy rule will be fired. In summary, GAFIS models have provided good interpolation accuracy in comparison with commonly-used spatial interpolation methods and have also provided satisfactory model interpretability under the adopted criteria. Thus, the objective of this chapter has been achieved.

3.5. Conclusion

In this chapter, the methodology to analyze and establish an interpretable fuzzy model for global spatial interpolation has been proposed. Such methodology has been applied to monthly spatial rainfall data in the northeast region of Thailand. The proposed methodology begins with FCM clustering analysis to determine the optimal number of clusters. Next, a prototype-based MFIS is generated and is then optimized by using GA. The proposed methodology has been evaluated by eight case studies.

The established fuzzy models are evaluated from quantitative and qualitative aspects. Interpolation accuracy has been compared with commonly-used global spatial interpolation methods. Model interpretability has been assessed by using the fuzzy interpretability criteria. The experimental results demonstrated that the established models are capable of providing acceptable interpolation accuracy and interpretable models to human analysts.

CHAPTER 4

A MODULAR FUZZY SYSTEM FOR MONTHLY RAINFALL SPATIAL INTERPOLATION

4.1. Introduction

In the previous chapter, a methodology to analyze and establish interpretable fuzzy systems for global monthly rainfall spatial interpolation has been proposed. The experimental results showed that the established fuzzy models provided satisfactory interpolation accuracy in comparison with other global spatial interpolation methods. Furthermore, such models can also satisfy the model interpretability criteria presented by Zhou and Gan (2008).

However, the global spatial interpolation methods presented in the previous chapter normally provided lower accuracy when compared to local methods (Kajornrit et al., 2011; Luo et al., 2008; Collins & Bolstad, 1996). That is because the capability of a single model (under a certain constraint) may not be enough to capture the high complexity of spatial rainfall data of a global area. As a result, the focus of this chapter is to present local methods that can improve the overall accuracy.

On the issue of interpretability, if we suppose that the number of parameters of a fuzzy model is increased until the model is complicated enough to achieve better accuracy, the interpretability of the model may deteriorate (see Figure 2.3). The higher number of prototypes in the model can affect the indistinguishability issues. Furthermore, when the number of parameters increases, the training data may not be enough to efficiently optimize the model (Jang et al., 1997).

Due to this contradictory issue, the subsequent issue is how to increase a model's accuracy with minimum effects to the model's interpretability. This chapter will address this issue by using a modular technique. This chapter is organized as follows: Section 2 revisits the background information of the datasets; Section 3 presents the proposed methodology and in Section 4 the proposed methodology will be evaluated; finally, Section 5 provides a conclusion for this chapter.

4.2. Case Study Area and Datasets

The case study area and datasets used in this chapter are the same as in the previous chapter. General information about the eight case studies was shown in Table 3.1. Features of the data consist of longitude (x), latitude (y), altitude (v), and amount of monthly rainfall (z). The training data and testing data are also the same as in the previous chapter for comparison purposes. As mentioned, the low correlation between altitude and the amount of monthly rainfall indicates the weak orographic effects. However, it is worthwhile investigating the local methods again.

4.3. Establishing a Modular Fuzzy System

The conceptual architecture of the modular model adopted is similar to the multiple expert systems (Chris-Tseng & Almogahed, 2009), as shown in Figure 4.1. Such a model consists of a set of local modules and one gating module. Input data are fed into all the local modules and the gating module. The function of the local modules is to interpolate rainfall values, while the gating module combines the results from the selected local modules into the final result.

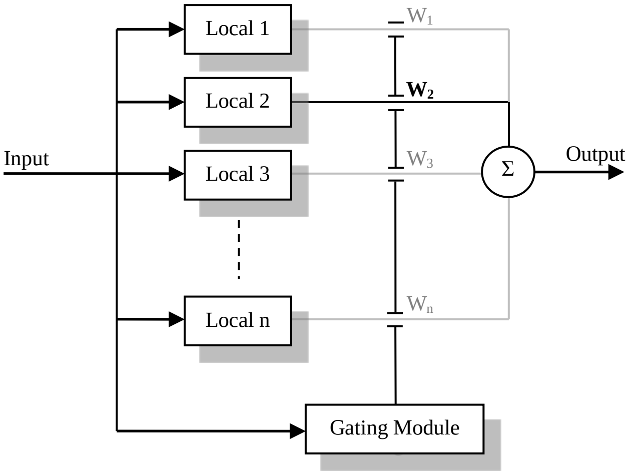


Figure 4.1. Conceptual architecture of the multiple expert systems.

However, to simplify the interpretability of this modular model, only one local module is selected to provide the final output. Therefore, this model can also be considered as a decision tree because the gating module works as a decision node and the local modules work as leaves of the decision tree. The proposed methodology consists of (i) localize the global area into local areas, (ii) establish local modules, and (iii) establish a gating module.

An overview of the establishment process is depicted in Figure 4.2. Global training data are first divided into local training data by FCM clustering. Each local training data is used to create one local module. Global training data and information gained from FCM clustering are then used to create the gating module.

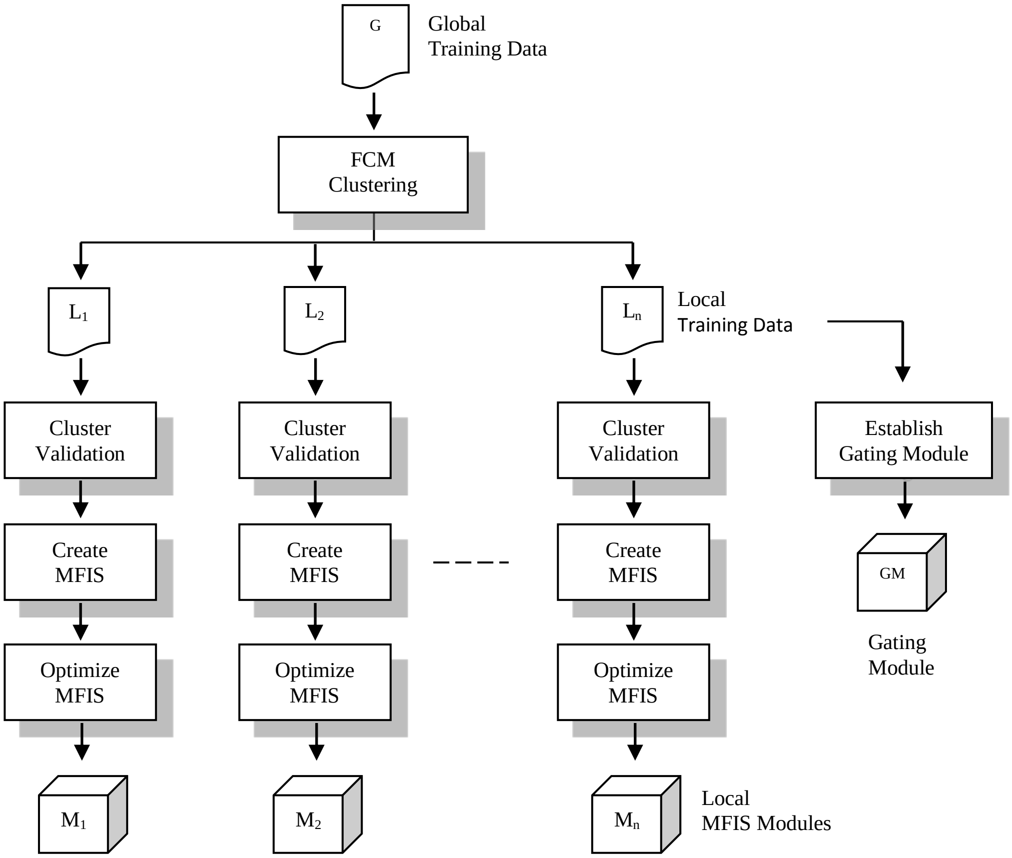


Figure 4.2. An overview of the establishment process.

4.3.1. Localize the Global Area

In the first step, the FCM clustering technique is applied to perform global localization. FCM is not only capable of dealing with uncertainty in the boundary, but it also provides the degree of membership values of the data that belong to other clusters. Global training data (*x*, *y*, *z*) are clustered into *n* clusters (i.e. *n* local areas). The number of *n* is determined by the validation methods proposed in Chapter 3, which is specifically developed for spatial data.

As mentioned, the selected number of clusters in the prototype-based fuzzy modeling is important to the interpretability of fuzzy models. If an appropriate clustering validation index for determining the number of clusters is applied, FCM can generate a parsimonious rule base (Zhou & Gan, 2008). In this chapter, the same method is used to determine the appropriate number of local modules for the modular approach.

4.3.2. Establish the Local Modules

The second step is to establish the local modules, using the training data in each cluster to determine the MFIS models. One concern is that using the modular model may introduce the problem of extrapolation between local areas when the boundaries are discrete, especially when rain gauge stations are sparse and/or have irregular distribution. Therefore, to prevent the problem of extrapolation between boundaries of local areas, small overlaps between local areas are needed.

Matrix U is an m x n matrix, the additional information from FCM, where m is the number of clusters and n is the number of data. Let μ_{ij}, the members of matrix U, be the degree of membership values of the data j to the cluster i and $\sum_{i=1}^{m} \mu_{ij} = 1$. Given x_j is the training data number j, x_j belongs to cluster i if (i) μ_{ij} in column j is maximum and (ii) μ_{ij} ≥ 0.5 of the maximum value of μ in column j. With this criterion, the overlap local data are created for the local modules.

To establish a local module, a prototype-based MFIS is created from FCM. The membership functions (MFs) used are the Gaussian function. The number of clusters is determined from Tutmez's criterion (Tutmez et al., 2007), which is determined by:

$$\text{Minimize } n_c \text{ under, Std}[z(x)] \approx \text{Std}[z(c)] \tag{4.1}$$

where n_c is the optimal number of clusters, Std is the standard deviation, $z(x)$ are the rainfall values of the data and $z(c)$ are the rainfall values at the cluster centers. The numbers of clusters are plotted against Std[$z(c)$]. The number of clusters that shows minimum distance between Std[$z(x)$] and Std[$z(c)$] is selected as the optimal number.

One constraint of this criterion is that Std[$z(c)$] $\leq$ Std[$z(x)$]. Rarely it is possible that Std[$z(c)$] > Std[$z(x)$] for all numbers of clusters. Consequently, all numbers of clusters do not satisfy this criterion. In this case, a default value must be defined. The proposed methodology sets the default values, n_{df}, as:

$$n_{df} = Floor(\ (n_{max} + n_{min}) \ / \ 2\ ) \tag{4.2}$$

The default values of n_{min} and n_{max} are 2 and 4, respectively. The proposed methodology selects the number of clusters as small as possible because the number of training data for each local module may not be enough to optimize the MFIS's parameters efficiently. Furthermore, as one local module represents one local area, the maximum number of fuzzy parameters should be at least less than the maximum number of fuzzy parameters recommended in the interpretability fuzzy criteria (Zhou & Gan, 2008).

Once the initial MFIS is created, the MFIS's parameters (sigma and center) are then optimized by GA. The chromosome of the algorithm consists of the sequence of input 1, input 2 and output respectively. In turn, the input and output are the sequence of MFs which consists of sigma and center parameters. The fitness function to be minimized is the sum square error between the observed value (z) and the interpolated value (z') of local training data and it is given as:

$$SSE = \sum_{i=1}^{S}(z_i' - z_i)^2 \tag{4.3}$$

In this process, the MFIS's parameters are allowed to be searched in a certain range in order to prevent indistinguishability of the MFs. Again, let α and β be user-defined control parameters, the center parameters (c) are allowed to be searched within [c - α, c + α] and the sigma parameters (σ) are allowed to be searched within [σ - β, σ + β].

4.3.3. Establish the Gating Module

In the proposed methodology, the function of the gating module is to activate the most appropriate local module to derive an interpolated value (z') from an input value (x, y). In other words, the final output of the system will be generated from one selected local module.

Normally, for a modular model, a few local modules may be activated at the same time. In this case, the final result comes from an aggregation of those activated local modules. However, at this point, to keep the established model simple and to maintain the interpretability of the model, only one local module will be activated. In this thesis, two methods of gating are proposed.

In the first method of gating, the generic gating module, the gating module selects one local module by determining the Euclidean distance between the input data and the cluster centers. The local module that gives the minimum distance is activated and is used to derive the final output. The formal expression of this method is:

$$z' = \sum_{i=1}^{n} w_i z_i' \tag{4.4}$$

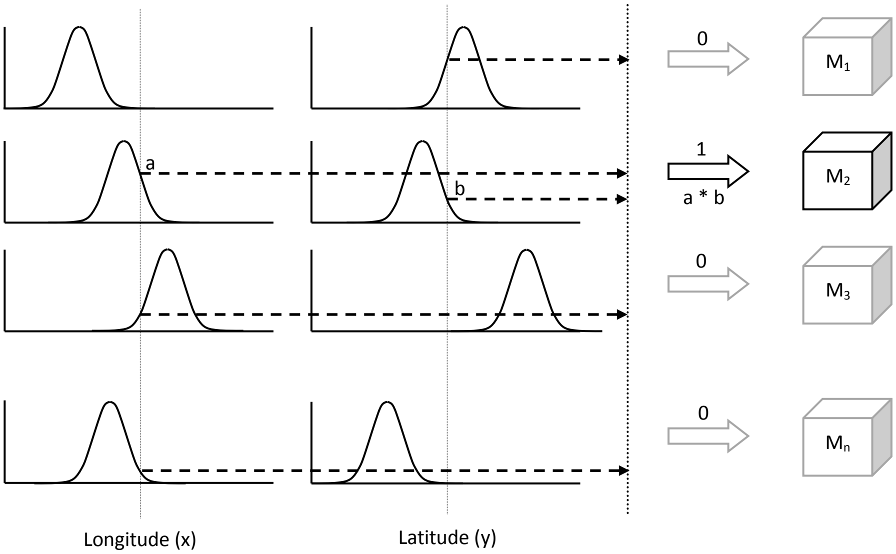
where z' is the final output, z_i' is the output of the local module i, and w_i is the weight associated with the local module i, which is evaluated by

$$w_i = \begin{cases} 1, & if\ d_i < d_{i'}\ where\ i \neq i' \\ 0, & elsewhere \end{cases} \tag{4.5}$$

In other words, the gating module uses a lookup table to calculate the geographic distances between the interpolated point (x_i, y_i) and the cluster centers of the local modules.

In the second method of gating, the fuzzy gating module, the decision is made by means of the antecedent part of the fuzzy system. The functional mechanism of the fuzzy gating module is the same as the general FIS, except the consequent part is not used (Figure 4.3). The number of fuzzy rules is associated with the number of local modules and, in turn, each fuzzy rule is associated with one MF in each input dimension.

For example, if there are seven local modules, the number of MFs of input 1, input 2 and the number of fuzzy rules are seven. Fuzzy rule$_1$ is associated with MF$_1$ in input 1 and input 2. When the input data is fed into the fuzzy gating module, the fuzzy inference process evaluates the firing strength of each rule *i*. The firing strength is the algebraic product of a_i and b_i, (or $a_i * b_i = r_i$), where a_i and b_i are the degree of membership values of input 1 and input 2 of rule *i*, respectively. The local module activated is associated with the rule that has the maximum firing strength.



IF *longitude (x)* = *cls$_1$* AND *latitude (y)* = *cls$_1$* THEN *local module$_1$ is activated*
IF *longitude (x)* = *cls$_2$* AND *latitude (y)* = *cls$_2$* THEN *local module$_2$ is activated*
IF *longitude (x)* = *cls$_3$* AND *latitude (y)* = *cls$_3$* THEN *local module$_3$ is activated*
....
IF *longitude (x)* = *cls$_n$* AND *latitude (y)* = *cls$_n$* THEN *local module$_n$ is activated*

Figure 4.3. Conceptual function of the fuzzy gating method.

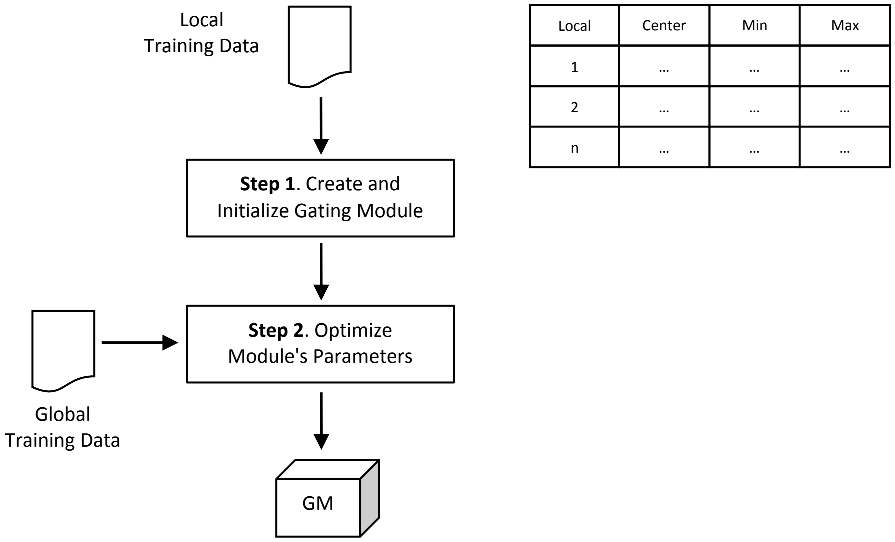
The formal expression of this method is:

$$z' = \sum_{i=1}^{n} w_i z_i' \tag{4.6}$$

where z' is the final output, z_i' is the output of local module *i*, and w_i is the weight associated with the local module *i*, which is evaluated by

$$w_i = \begin{cases} 1, & if\ r_i > r_{i'}\ where\ i \neq i' \\ 0, & elsewhere \end{cases} \tag{4.7}$$

Establishing a fuzzy gating module consists of two steps, that is, (i) create and initialize the module, and (ii) optimize parameters of the module as shown in Figure 4.4. The center (x_c, $\mathrm{y_c}$), minimum (x_{min}, y_{min}) and maximum (x_{max}, y_{max}) values of all local training data are used to create and initialize the MFs. The type of MF used is a Gaussian function. However, the triangle function is also used for comparison purposes.



Local	Center	Min	Max
1	...	...	...
2	...	...	...
n	...	...	...

Figure 4.4. Steps to create the fuzzy gating module.

To initialize Gaussian MFs, the centers of MFs are set based on the centers of the clusters. For example, suppose that the first cluster center (x_c, y_c) is (0.3, 0.7), the center of the first MF of input 1 is set to 0.3 and the center of the first MF of input 2 is set to 0.7. The sigma parameter of the first MF of input 1 and input 2 are set to 0.1424 × ($x_{max} - x_{min}$) and 0.1424 × ($y_{max} - y_{min}$), respectively. The constant value of 0.1424 was derived from a preliminary testing. The constant value of 0.1424 makes the width of the MF approximately close to the width of the data.

To initialize a triangle MF, such a MF consists of three parameters *a*, *b* and *c*, in which $a \leq b \leq c$. The *b* values of MFs are set based on the cluster center. For example, if the first cluster center (x_c, y_c) is (0.3, 0.7), the b value of the first MF of input 1 is set to 0.3 and the b value of the first MF of input 2 is set to 0.7. Similarly, the a value is set to the minimum value of the cluster and the c value is set to the maximum value of the cluster for all input dimensions. For both types of MFs, the number of fuzzy rules is associated with the number of clusters.

In the optimization step, parameters of the fuzzy gating module are optimized by GA. The chromosome consists of the MFs' parameters of input 1 and input 2. For the Gaussian MF, again, the α and β parameters are used to control the search space in the optimization process. Generally, the centroid of the cluster is not at the geometric center of the cluster. Thus, the parameter α allows a small space for the center of the MF to move closer to the geometric center of the cluster in order to let the MF cover the entire cluster.

For the triangle MF, the β_{in} and β_{out} parameters are used to control the search space of the a and c parameters. β_{in} is used to control the minimum search space toward the cluster centers and β_{out} is used to control the maximum search space toward the cluster centers (Figure 4.5). As the triangle MF is asymmetric, the b parameter is not required to be optimized. The fitness function used is to minimize sum square error between the observed values (z) and the interpolated values (z') of the global training data.

In summary, the proposed methodology begins with localizing the global area into local areas by FCM clustering. Next, one MFIS module is created for each local area to perform interpolation. After that, a gating module is created to perform the decision. Two gating methods are proposed. The first method is based on geographic distance and the second method used the concept of the fuzzy system. In the next section, the proposed methods will be evaluated.

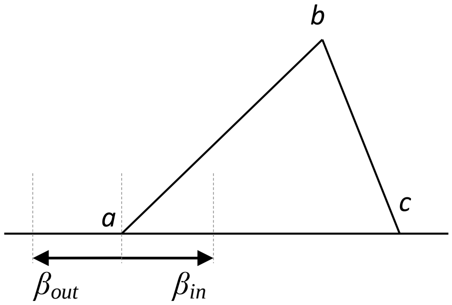


Figure 4.5. The search space of the β_{in} and β_{out} control parameters.

4.4. Evaluation of the Proposed Methodology

In order to evaluate the proposed methods, some commonly-used local spatial interpolation methods are adopted for comparison purposes. Those methods include inverse distance weighting (IDW), local polynomial (LP), thin plate splines (TPS), ordinary kriging (OK), universal kriging (UK) and ordinary co-kriging (CK). However, to simplify this comparison task, these methods will be firstly compared. After that, the best method of each case study will be selected to compare with the proposed methods.

IDW, LP and TPS are considered as deterministic methods, whereas OK, UK and CK are known as geostatistic methods. Among these methods, CK uses the altitude feature as the auxiliary variable to perform interpolation. The experiments of these methods are performed on *ArcGIS* application software (ESRI, n.d.), which is widely used in the GIS area.

4.4.1. Models Establishment

From now on, the local deterministic and geostatistic methods described above are called the GIS methods. The number of data points included to perform interpolation for the GIS methods are six, which was suggested by Zimmerman et al. (1999). The anisotropy feature is disabled in these experiments for the proposed methods used in the comparison.

For IDW, the k parameter was automatically optimized by *ArcGIS* (Luo et al., 2008) and it was used as the control method (Li et al., 2011), or standard benchmark. For LP, as the number of data points included was six, only the first and second orders are available. However, the second order provided large error; therefore, the first order is adopted instead.

TPS with tension is selected rather than regular TPS because it provided more accurate results. A spherical semivariogram is used for OK, UK and CK and the number of lags is twelve, which is automatically generated by the software. A first order polynomial is used for UK for representing the drift component. A higher order polynomial can remove some necessary spatial relationships (Chang, 2006).

To establish local modules, the α and β parameters are both set to 0.05. The α parameter is set to about 5 percent of the universe of discourse (UoD). This is set to a half of α parameter in GAFIS because the size of data in local modules decreases approximately more than a half in each dimension. The value of 0.05 should be large enough for the search space. For the β parameter, this setting allows the search space of MF's flank to be approximately 15 to 20 percent of UoD.

To establish the fuzzy gating module, for a Gaussian MF, the α and β parameters are set to 0.05 and 0.025 respectively. As mentioned, the α parameter allows a small space for the centroid of the cluster to move closer to the geometric center of the cluster in order to enable MFs to cover entire the cluster. The β parameter is set to a half of the setting in local modules because it allows the flank of MFs to vary in the smaller search space on the overlap area between clusters. For the triangle MF, the β_{in} and β_{out} parameters are both set to 0.1, which allows a search space approximately equal to the β parameter of the Gaussian MF.

In general, if α and β parameters are set too small, especially the β parameter, the optimal solution may not be met since the fuzzy models cannot handle the uncertainty in rainfall data efficiently under these constrained conditions. This is one reason that in these experiments the β parameter was set to a little bit larger search space than the α parameter. In the GA optimization process, the GA's population is set to 150 and the GA's generation is set to 30, where the best and average fitness values are met.

4.4.2. Quantitative Results

Similar to the previous chapter, four quantitative measures have been used to evaluate the interpolation accuracy, that is, mean error (ME) or bias error, mean absolute error (MAE), root mean square error (RMSE) and correlation coefficient (R). These measures are normalized by the mean values of the datasets for comparison purposes.

The experimental results of GIS methods are shown in Table 4.1 to 4.4. The Average rows in the tables are the average values from the eight case studies. The Improvement rows in the tables are the improvement percentage based on the IDW method. These two values are depicted in Figure 4.6. (Notice that for the graphs in Figure 4.6, the Average values refer to the left axis and the Improvement values refer to the right axis.)

According to the results, the best GIS method of each case is selected to compare with the proposed methods. From Case 1 to Case 8, the best methods are TPS, UK, IDW, OK, OK, TPS, UK and UK, respectively. From now on, Mod FIS, Mod FIS–FSG and Mod FIS–FST are the proposed models with the generic gating module, the Gaussian-MF fuzzy gating module and the triangle MF fuzzy gating module, respectively. Moreover, the results of the GAFIS model are also included in this comparison. Tables 4.5 to 4.8 show the experimental results of the proposed methods. The Average and Improvement values are depicted in Figure 4.7.

Table 4.1. Normalized mean error (GIS methods).

Case Study	IDW	LP	TPS	OK	UK	CK
1	0.005	0.027	0.004	0.015	0.001	0.015
2	-0.012	-0.018	-0.006	-0.030	-0.048	-0.030
3	0.035	0.039	0.040	0.041	0.042	0.041
4	0.079	0.058	0.082	0.059	0.044	0.059
5	-0.017	0.021	-0.011	0.000	-0.003	0.000
6	-0.018	-0.012	-0.016	-0.012	-0.012	-0.012
7	0.016	-0.012	0.013	0.014	0.005	0.014
8	-0.008	-0.038	-0.009	-0.025	-0.012	-0.024
Average	0.024	0.028	0.023	0.024	0.021	0.024
Improvement	0.00	-17.96	5.09	-3.14	11.69	-2.71

Table 4.2. Normalized mean absolute error (GIS methods).

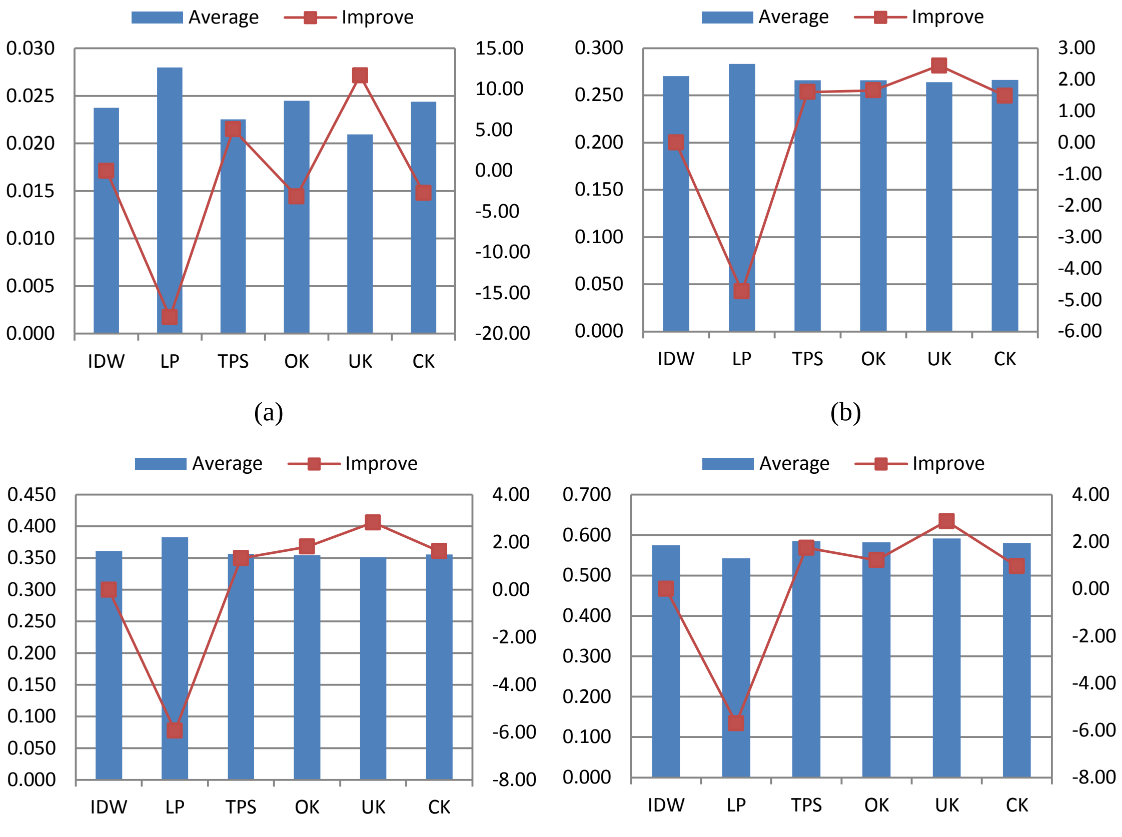
Case Study	IDW	LP	TPS	OK	UK	CK
1	0.287	0.293	0.283	0.290	0.290	0.290
2	0.337	0.364	0.331	0.331	0.327	0.333
3	0.201	0.231	0.200	0.205	0.217	0.205
4	0.299	0.319	0.291	0.281	0.289	0.281
5	0.269	0.286	0.271	0.259	0.257	0.260
6	0.270	0.291	0.271	0.275	0.272	0.275
7	0.271	0.270	0.272	0.264	0.258	0.266
8	0.229	0.213	0.211	0.222	0.203	0.222
Average	0.270	0.283	0.266	0.266	0.264	0.266
Improvement	0.00	-4.73	1.60	1.66	2.44	1.49

Table 4.3. Normalized root mean square error (GIS methods).

Case Study	IDW	LP	TPS	OK	UK	CK
1	0.357	0.361	0.355	0.372	0.360	0.372
2	0.431	0.478	0.423	0.422	0.409	0.425
3	0.275	0.330	0.277	0.279	0.289	0.279
4	0.402	0.439	0.397	0.384	0.393	0.384
5	0.360	0.386	0.360	0.336	0.341	0.337
6	0.358	0.372	0.357	0.358	0.358	0.358
7	0.387	0.396	0.387	0.378	0.375	0.380
8	0.320	0.300	0.297	0.309	0.284	0.308
Average	0.361	0.383	0.357	0.355	0.351	0.355
Improvement	0.00	-5.93	1.33	1.81	2.83	1.62

Table 4.4. Correlation coefficient (GIS methods).

Case Study	IDW	LP	TPS	OK	UK	CK
1	0.425	0.443	0.436	0.343	0.400	0.345
2	0.553	0.437	0.578	0.650	0.656	0.643
3	0.770	0.706	0.765	0.756	0.742	0.757
4	0.308	0.273	0.320	0.317	0.308	0.317
5	0.532	0.469	0.534	0.553	0.555	0.550
6	0.508	0.484	0.508	0.493	0.493	0.492
7	0.738	0.725	0.738	0.752	0.756	0.750
8	0.768	0.804	0.805	0.794	0.826	0.794
Average	0.575	0.543	0.585	0.582	0.592	0.581
Improvement	0.00	-5.70	1.74	1.22	2.87	0.96



(a) (b)

(c) (d)

Note: (a) normalized mean error, (b) normalized mean absolute error, (c) normalized root mean square error and (d) correlation coefficient.

Figure 4.6. Plot of the average and improvement values from GIS methods.

In terms of ME, Mod FIS–FSG provided a small negative bias in Case 1 while the others showed a positive bias. Mod FIS showed the same result in Case 7. In Case 8, Mod FIS–FST provided positive bias while the other showed negative bias. However, these bias errors are too small to be counted as suspicious condition. The Average values in Table 4.6 are calculated from the absolute value of ME. Based on this value, the quality of bias error can be ranked as GIS > Mod FIS–FSG = Mod FIS–FST > Mod FIS > GAFIS.

In terms of MAE, Mod FIS–FSG provided the best accuracy in Cases 1, 3 and 5, whereas Mod FIS–FST provided the best accuracy in Cases 4, 7 and 8. In Case 6, both of them showed compatible results at the best accuracy. Mod FIS provided the best accuracy in Case 2 and showed compatible results to Mod FIS–FSG in Case 3. Based on the average, the efficiency of all interpolators can be ranked as Mod FIS–FSG = Mod FIS–FST > Mod FIS > GIS > GAFIS.

In terms of RMSE, Mod FIS–FSG showed the best accuracy in Case 1 and Case 3, whereas Mod FIS–FST showed the best accuracy in Cases 5, 6 and 7. In Case 4, both of them provided compatible results for the best accuracy. Mod FIS showed the best accuracy in Case 2 and GIS showed the best accuracy in Case 8. In Case 3, however, Mod FIS–FST provided unsatisfactory results, which also occurred in MAE. Based on the average, the efficiency of all interpolators can be ranked as Mod FIS–FSG > Mod FIS–FST > Mod FIS > GIS > GAFIS.

Table 4.5. Normalized mean error (the proposed methods).

Case Study	GIS	GAFIS	Mod FIS	Mod FIS–FSG	Mod FIS–FST
1	0.004	0.013	0.024	- 0.009	0.012
2	- 0.048	- 0.049	- 0.034	- 0.024	- 0.033
3	0.035	0.049	0.002	0.020	0.016
4	0.059	0.054	0.070	0.083	0.060
5	0.000	- 0.007	- 0.068	- 0.021	- 0.040
6	- 0.016	- 0.013	- 0.012	- 0.004	- 0.010
7	0.005	0.007	- 0.005	0.000	0.016
8	- 0.012	- 0.023	- 0.034	- 0.032	0.007
Average	0.023	0.027	0.031	0.024	0.024
Improvement	0.00	-19.35	-38.43	-7.10	-7.77

Table 4.6. Normalized mean absolute error (the proposed methods).

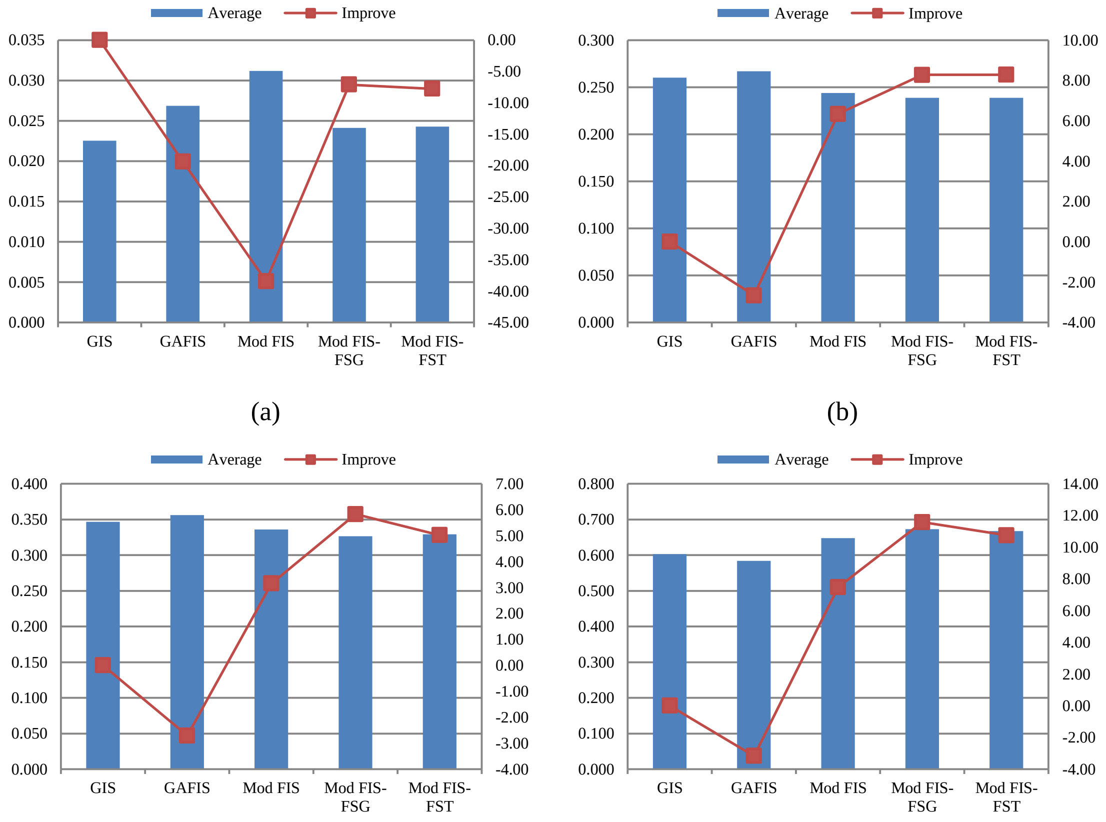
Case Study	GIS	GAFIS	Mod FIS	Mod FIS–FSG	Mod FIS–FST
1	0.283	0.284	0.258	0.248	0.251
2	0.327	0.322	0.283	0.290	0.294
3	0.201	0.224	0.198	0.198	0.203
4	0.281	0.285	0.255	0.254	0.251
5	0.259	0.266	0.241	0.227	0.232
6	0.271	0.267	0.260	0.249	0.249
7	0.258	0.256	0.249	0.238	0.237
8	0.203	0.233	0.205	0.204	0.192
Average	0.260	0.267	0.244	0.239	0.239
Improvement	0.00	-2.66	6.34	8.28	8.28

Table 4.7. Normalized root mean square error (the proposed methods).

Case Study	GIS	GAFIS	Mod FIS	Mod FIS–FSG	Mod FIS–FST
1	0.355	0.366	0.335	0.318	0.332
2	0.409	0.403	0.378	0.385	0.383
3	0.275	0.290	0.270	0.261	0.279
4	0.384	0.396	0.358	0.357	0.357
5	0.336	0.348	0.332	0.307	0.306
6	0.357	0.353	0.353	0.334	0.330
7	0.375	0.369	0.367	0.363	0.361
8	0.284	0.326	0.295	0.289	0.289
Average	0.347	0.356	0.336	0.327	0.329
Improvement	0.00	-2.71	3.16	5.83	5.02

Table 4.8. Correlation coefficient (the proposed methods).

Case Study	GIS	GAFIS	Mod FIS	Mod FIS–FSG	Mod FIS–FST
1	0.436	0.370	0.528	0.591	0.540
2	0.656	0.657	0.703	0.679	0.697
3	0.770	0.749	0.783	0.804	0.783
4	0.317	0.315	0.452	0.467	0.454
5	0.553	0.537	0.621	0.662	0.669
6	0.508	0.513	0.515	0.581	0.595
7	0.756	0.769	0.771	0.778	0.782
8	0.826	0.760	0.811	0.820	0.820
Average	0.603	0.584	0.648	0.673	0.668
Improvement	0.00	-3.16	7.48	11.58	10.75



(a)

(b)

(c)

(d)

Note: (a) normalized mean error, (b) normalized mean absolute error, (c) normalized root mean square error and (d) correlation coefficient.

Figure 4.7. Plot of the average and improvement values from the proposed methods.

In terms of R, Mod FIS–FSG showed the best in Cases 1, 3 and 4, whereas Mod FIS–FST showed the best in Cases 5, 6 and 7. Mod FIS provided the best accuracy in Case 2 and GIS showed the best accuracy in Case 8. The results from R and RMSE measure are rather compatible. Base on the average, interpolation accuracy are ranked as Mod FIS–FSG > Mod FIS–FST > Mod FIS > GIS > GAFIS.

Based on overall results, it can be concluded that the modular technique can improve the interpolation accuracy of GAFIS significantly. Furthermore, such a technique can provide satisfactory accuracy in comparison with the GIS methods. Among the modular models, Mod FIS–FSG and Mod-FIS–FST showed superior results to Mod FIS. This shows that the fuzzy gating module can improve accuracy compared with the generic gating module.

4.4.3. Qualitative Results

Table 4.9 shows the number of local modules and the number of prototypes created in the local modules of eight case studies. (Notice that the number of prototypes is associated with the number of fuzzy rules and the number of MFs of the inputs and the output of the fuzzy models.)

Table 4.9. Numbers of the local modules and the prototypes in the local modules.

Case Study	Number of Local Modules	Number of Prototypes in Local Modules
1	7	2, 3, 4, 4, 4, 4, 2
2	7	3, 4, 3, 3, 3, 2, 4
3	7	2, 4, 2, 4, 4, 4, 3
4	9	4, ,4, 3, 4, 4, 3, 2, 3, 3
5	7	3, 4, 4, 4, 4, 4, 3
6	8	2, 2, 2, 4, 4, 3, 4, 4
7	7	4, 3, 4, 4, 4, 3, 3
8	8	3, 4, 3, 4, 3, 4, 4, 3

Figures 4.8 and 4.9 show an example of the optimized Gaussian and the triangle MFs in the fuzzy gating module and the associated fuzzy rules. Figures 4.10 to 4.12 show an example of the optimized MFs and the associated fuzzy rules of the local modules with two, three, and four prototypes.

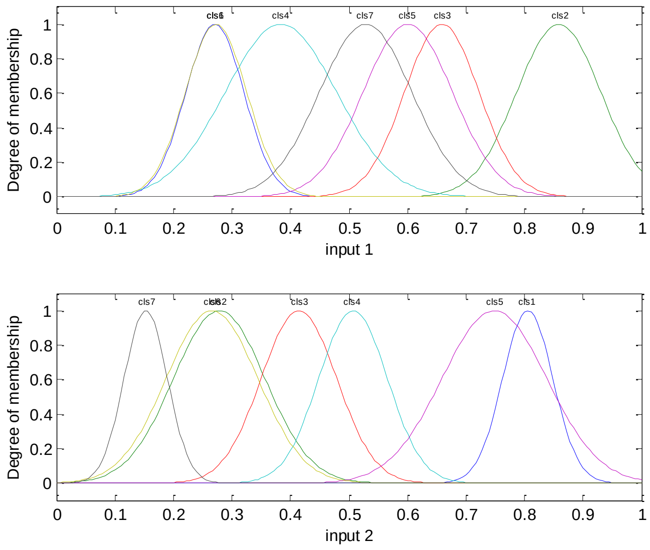
For local modules, in terms of low-level interpretability, the selected number of MFs is considered from standard deviation analysis. The maximum number of MFs varies from two to four in local modules. This satisfies the moderate number of MFs criterion. Furthermore, when the number of MFs is small, the distinguishability of MFs is also good. MFs of inputs directly refer to locations in the study area and MFs of the output directly explain the amount of monthly rainfall.

By using the prototyped-based fuzzy modeling, the normalization criterion of local modules is met, which at least one data point in the UoD having a membership value equal to one. Although one local module cannot satisfy the coverage of fuzzy partitioning criterion, whole local modules in the system can satisfy this criterion. All local modules are created from overlap data. Therefore, it should be enough to ensure that the entire UoD of variables is covered by the MFs generated.

In terms of high-level interpretability, the rule base parsimony and simplicity criterion is met since the number of rules in local modules does not exceed four. The readability of single rules criterion is also satisfied; however, that is because the input to the system is already readable. The consistency and completeness criteria can also be achieved. Under prototype-based fuzzy modeling, there is no conflicting rule and at least one rule has been fired. The transparency of rule structure is also clear to human analysts.

For the gating module, the interpretability of the generic gating module is simple and clear in itself. The mechanism to select the local module is not complicated. The selected local module is derived from the geometric distance between the interpolated point and the cluster centers. For the fuzzy gating module, on the other hand, such a module can be explained by the interpretability of GAFIS model. The mechanism of fuzzy gating is similar to the mechanism of GAFIS except for the derivation of the output. This difference does not affect the interpretability of the system in both the low-level and high-level, in general.

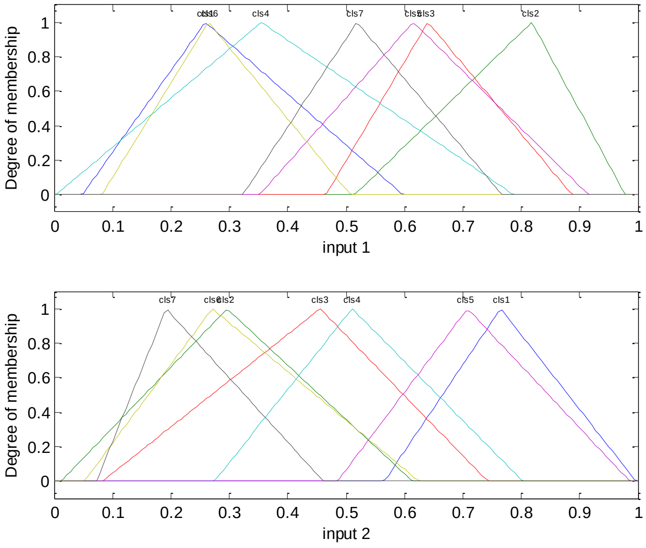
So far, the qualitative results have been presented. One can see that the interpretability of the local modules and the gating module is generally satisfactory. Although the number of fuzzy sets (MFs) and fuzzy rules of the whole system increases, the interpretability of the whole model is still satisfied, at least, in the modular approach. The entire system is clearly structured and well established. By using the modular concept, the accuracy of the system can be improved with less effect on the model interpretability.



IF x = cls1 AND y = cls1 THEN sub-module = 1
IF x = cls2 AND y = cls2 THEN sub-module = 2
IF x = cls3 AND y = cls3 THEN sub-module = 3
IF x = cls4 AND y = cls4 THEN sub-module = 4
IF x = cls5 AND y = cls5 THEN sub-module = 5
IF x = cls6 AND y = cls6 THEN sub-module = 6
IF x = cls7 AND y = cls7 THEN sub-module = 7

x = input 1 = longitude
y = input 2 = latitude

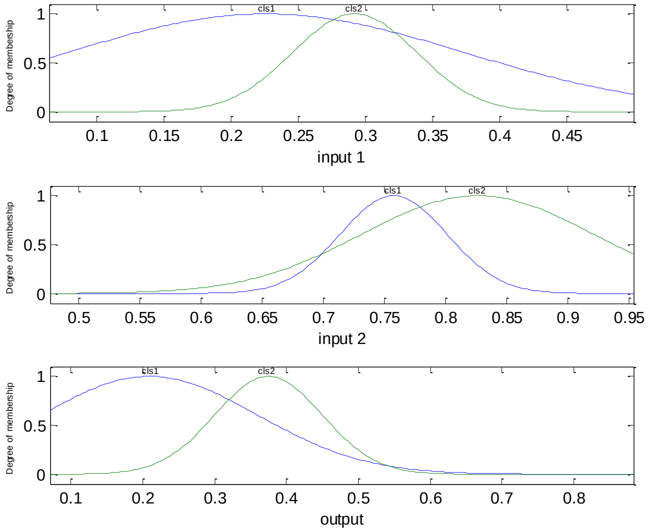
Figure 4.8. An example of Gaussian MFs in the fuzzy gating module (Case 1).



IF x = cls1 AND y = cls1 THEN sub-module = 1
IF x = cls2 AND y = cls2 THEN sub-module = 2
IF x = cls3 AND y = cls3 THEN sub-module = 3
IF x = cls4 AND y = cls4 THEN sub-module = 4
IF x = cls5 AND y = cls5 THEN sub-module = 5
IF x = cls6 AND y = cls6 THEN sub-module = 6
IF x = cls7 AND y = cls7 THEN sub-module = 7

x = input 1 = longitude
y = input 2 = latitude

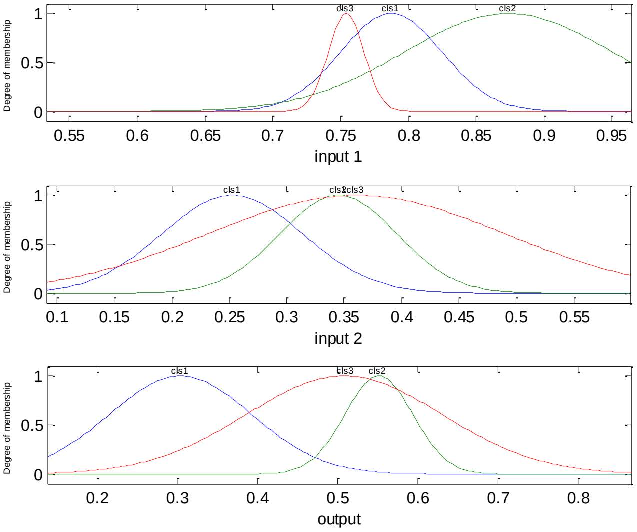
Figure 4.9. An example of the triangle MFs in the fuzzy gating module (Case 1).



IF x = cls1 AND y = cls1 THEN z = cls1
IF x = cls2 AND y = cls2 THEN z = cls2

x = input 1 = longitude
y = input 2 = latitude
z = output = rainfall

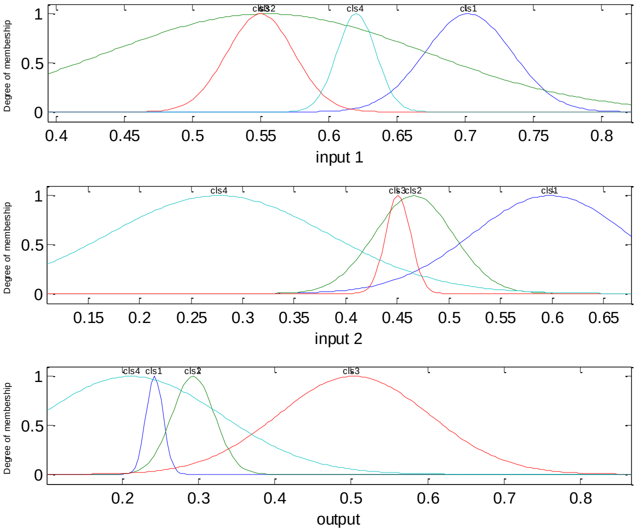
Figure 4.10. An example of MFs in the local module with two fuzzy rules (Case 1).



IF x = cls1 AND y = cls1 THEN z = cls1
IF x = cls2 AND y = cls2 THEN z = cls2
IF x = cls3 AND y = cls3 THEN z = cls3

x = input 1 = longitude
y = input 2 = latitude
z = output = rainfall

Figure 4.11. An example of MFs in the local module with three fuzzy rules (Case 1).



IF x = cls1 AND y = cls1 THEN z = cls1
IF x = cls2 AND y = cls2 THEN z = cls2
IF x = cls3 AND y = cls3 THEN z = cls3
IF x = cls4 AND y = cls4 THEN z = cls4

x = input 1 = longitude
y = input 2 = latitude
z = output = rainfall

Figure 4.12. An example of MFs in the local module with four fuzzy rules (Case 1).

4.5. Conclusion

This chapter proposes the use of a modular technique to improve the interpolation accuracy of GAFIS with small effect on the model's interpretability. The proposed methodology localizes the global area into several local areas by FCM clustering. The degree of membership values generated from FCM clustering is used to create overlap in local data. For each local area, a prototype-based fuzzy model is created by FCM and is then optimized by GA.

The fuzzy systems, again, are used to perform a gating function to activate the appropriate local modules. The established models are compared with commonly-used local spatial interpolation methods from GIS. The results have shown that the proposed model can provide accurate results when compared with those GIS methods. In addition, with the fuzzy gating module, the interpretable objective can be met at the global level and the local level in the modular way.

CHAPTER 5

AN INTERPRETABLE FUZZY SYSTEM FOR MONTHLY RAINFALL TIME SERIES PREDICTION

5.1. Introduction

As aforementioned, accurate time series prediction models of the rainfall variable are necessary for flow forecasting in river basins. In the same way, interpretable time series prediction models of the rainfall variable are also necessary for human analysts to understand the established models, so that the human analysts can gain an insight into the model, as well as adding any prior knowledge. Establishing an interpretable fuzzy system for monthly rainfall time series is therefore the aim of this part of the study.

Conceptually, at least two issues should be considered. Firstly, the inputs to the system should be readable and should clearly characterize human knowledge. In spatial interpolation, the inputs to the system have low dimensions and are directly related to the location. However, in time series prediction the inputs can be high dimensional vectors of historical information, and high complexity of the inputs most likely increases the complexity of the model.

The second issue is how to represent the system in a meaningful approach. Similar to what has been discussed before in the preceding chapters, the interpretability issue needs to be taken care of. The fuzzy models that satisfy interpretability fuzzy criteria are capable of providing a meaningful representation of the models. In this part of the thesis, these two issues will be taken into account as the key objectives.

This chapter is organized as follows: Section 2 is referenced to the case study and general information of the datasets; Section 3 presents the proposed methodology; evaluation of the proposed model will be presented in Section 4; and, finally, Section 5 is the conclusion.

5.2. Case Study Area and Datasets

The case study area is located in the northeast region of Thailand. This is the same area used in the case studies in Chapters 3 and 4; however, only the time series data will be used here. Eight monthly rainfall time series data collected throughout the study area are used in this chapter. Figure 5.1 shows the locations of the eight rain gauge stations.

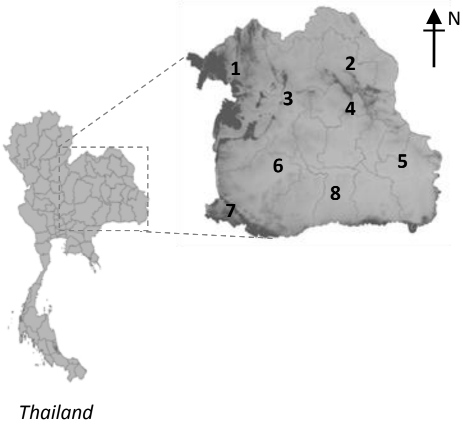


Figure 5.1. Locations of the eight rain gauge stations in the study area.

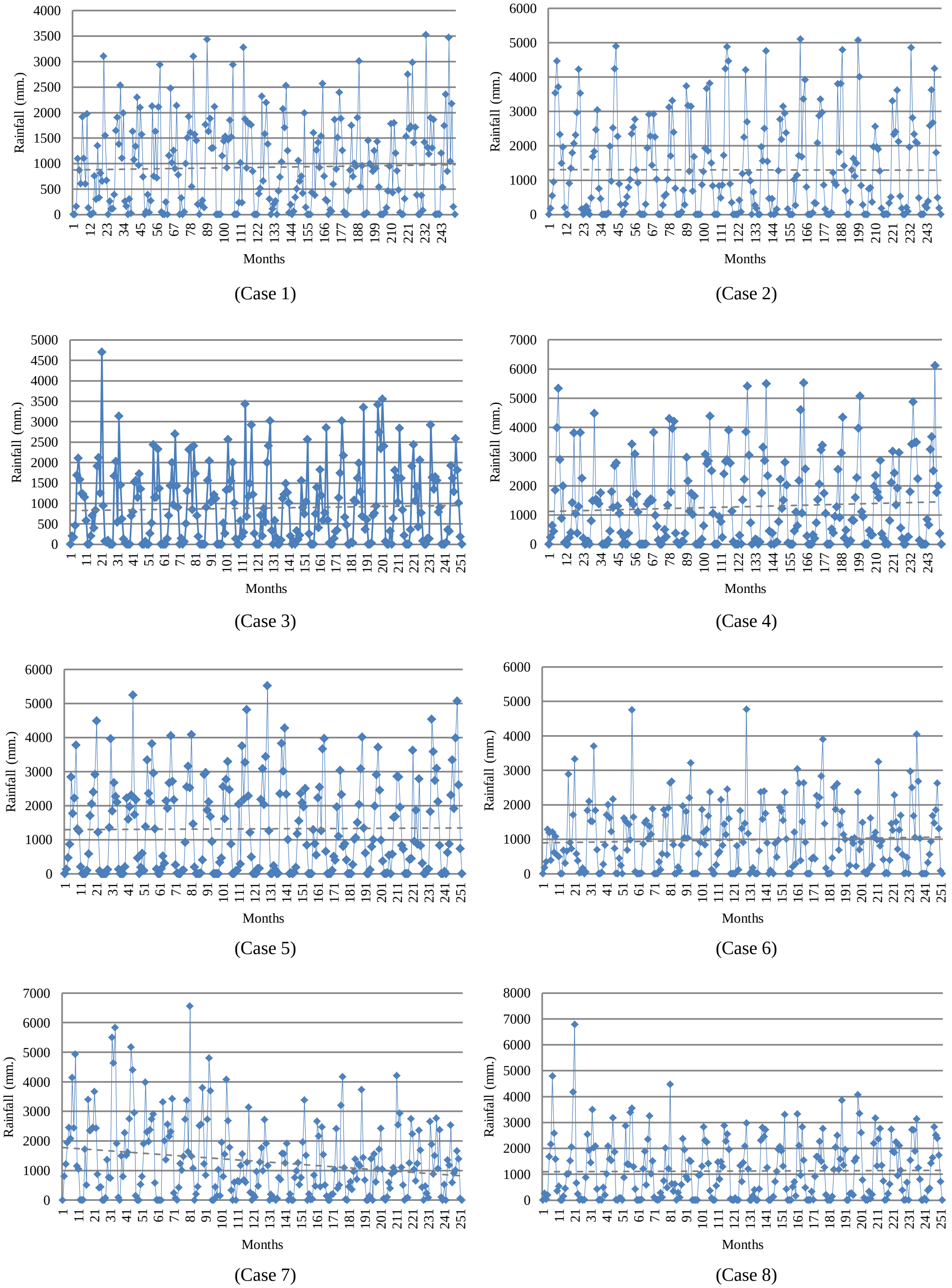
Figure 5.2 shows the rainfall time series graphs of the eight datasets. Although the monthly rainfall time series data are complex and noisy, there is no suspicious event such as dramatic shifts of trend appearing. Thus, all periods of the time series data will be used. The data from 1981 to 1998 are used to calibrate the models, and the data from 1999 to 2001 are used to validate the established models. The statistics and locations of the eight datasets are shown in Table 5.1. The data are normalized by linear transformation for computational purposes.

Table 5.1. Statistics and locations of the eight rainfall time series.

Statistics	**Case 1**	**Case 2**	**Case 3**	**Case 4**	**Case 5**	**Case 6**	**Case 7**	**Case 8**
Mean (mm.)	929	1303	889	1286	1319	981	1296	1124
SD	867	1382	922	1425	1346	976	1289	1153
Kurtosis	-0.045	-0.100	0.808	0.532	-0.224	1.229	1.590	1.725
Skewness	1.655	0.952	1.080	1.131	0.825	1.154	1.276	0.961
Minimum	0	0	0	0	0	0	0	0
Maximum	3527	5099	4704	6117	5519	4770	6558	6778
Latitude	17.25N	17.15N	16.66N	16.65N	15.50N	15.40N	14.63N	15.40N
Longitude	101.80E	104.13E	102.88E	104.05E	104.75E	102.35E	101.30E	103.40E
Altitude	283	176	164	155	129	152	476	152

5.3. Establishing an Interpretable Fuzzy System

An overview of the proposed methodology is depicted in Figure 5.3. The methodology consists of five steps. First, appropriate inputs to the system are selected. Second, a Mamdani-type FIS model and MFs are generated. Fuzzy rules are generated in the third step. The fourth and fifth steps are the optimization process. Actually, the optimization steps can be grouped into one step. In order to control the number of parameters in the optimization stage, and the fact that the objectives of these two steps are different, separating the optimization into two steps is more suitable.



(Case 1) (Case 2)

(Case 3) (Case 4)

(Case 5) (Case 6)

(Case 7) (Case 8)

Figure 5.2. The eight monthly rainfall time series graphs.

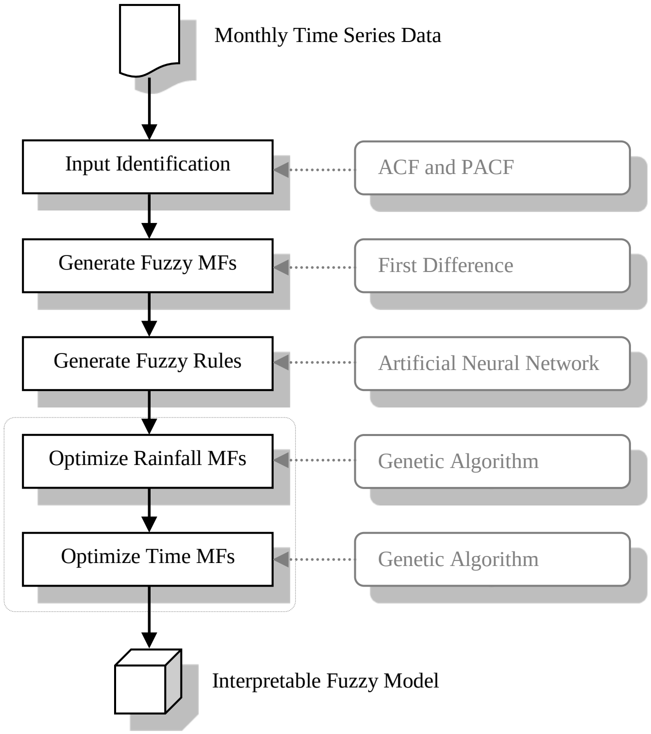


Figure 5.3. An overview of the proposed methodology.

5.3.1. Input Identification

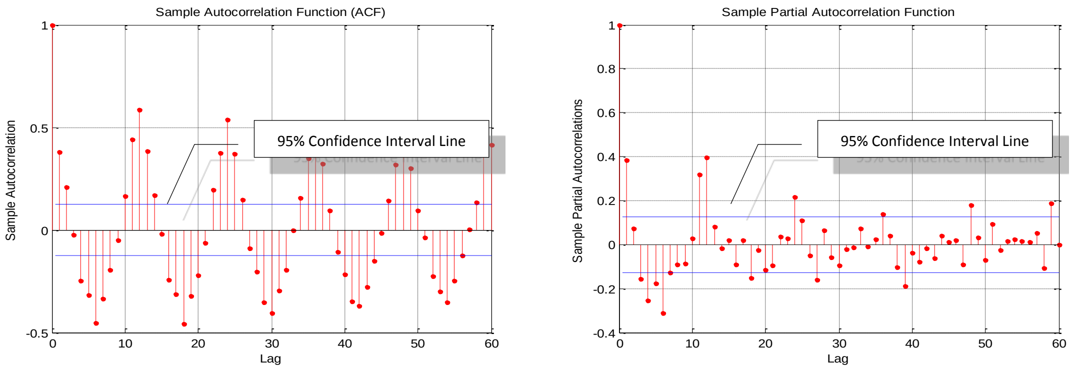
The objective of predicting rainfall using antecedent values is to generalize the relationships of the following form

$$y = f(x^m) \tag{5.1}$$

where x^m is an m-dimensional input vector representing rainfall values with different time lags and y is a one-dimensional output representing predicted rainfall value. Formerly, one difficulty of using non-linear models was that x^m was not known before and there was no consistence theory to define appropriate x^m (Wang et al., 2009).

Recently, two statistical methods, autocorrelation function (ACF) and partial autocorrelation function (PACF), have been employed to determine the dimension m of input vectors for non-linear models (Wu & Chau, 2013; Wu et al, 2010; Wang et al., 2009). In general, ACF and PACF are used to diagnose the order of the autoregressive process.

Figure 5.4 shows an example of ACF and PACF of the monthly rainfall time series data of Case 1 (notice that ACF and PACF of all cases are shown in Appendix B). ACF exhibits the peak value at lag 12 and PACF shows a significant correlation at 95% confidence level interval up to lag 12. Therefore, this suggests that twelve antecedent rainfall values contain sufficient information to predict future rainfall.



(a)

(b)

Figure 5.4. ACF (a) and PACF (b) of monthly rainfall time series data (Case 1).

However, for an FIS model, selecting 12 lags can result in an increase of complexity in fuzzy rules and will cause problems with readability, especially, in the antecedent part (Zhou & Gan, 2008). Furthermore, due to the issue of the curse of dimensionality, the number of fuzzy parameters can increase tremendously depending on the number of membership functions selected. Even using the phase space reconstruction to identify input may not be a good solution to this problem. However, as the monthly time series is periodic in nature, adding a time coefficient as a supplementary feature is a promising approach (Toprak et al., 2009; Keskin et al., 2006).

Time coefficient (C_t) is used to assist the model to scope prediction into a specific period. It may be C_t = 2 (wet and dry period) or C_t = 12 (calendar months). This study adopted C_t = 12 as a supplementary feature. Once the C_t is added into the system, using 12-lag antecedence as the model's inputs may be redundant. This study proposed the use of the first lag that crosses the confidence interval line as the minimum information for the model. Therefore, two first lags of rainfalls and C_t are considered as the model input. This selection conforms to the suggestion in the work of Keskin (2006) and Raman and Sunilkumar (1995) that 2-lag antecedence contains sufficient information for monthly hydrological time series prediction.

5.3.2. Generate Fuzzy Membership Functions

In order to create fuzzy MFs for the proposed methodology, two aspects need to be considered simultaneously. The created MFs should be distinguished enough and should reflect the characteristics of the time series data. Huarng (2001) suggested that the appropriate interval length between two consecutive MFs for time series data should be at least half of the average of fluctuations in the time series. The fluctuation, herein, is the absolute value of the first difference of any two consecutive data. This concept is adopted in this methodology. However, it has been adapted to fit to the characteristics of the monthly rainfall data.

In this methodology, the absolute values of the first difference of time series are calculated. The percentile at 25, 50 and 75 of these values are adopted to explain the fluctuation of the rainfall at low, medium and high periods. The low period of the rainfall is defined as zero to percentile 50 of the rainfall values; the medium period is defined as percentile 50 to 75; and above percentile 75 are defined as the high period. This procedure is applied to the 1st lag input, the 2nd lag input and the output of the fuzzy model.

The MFs of C_t are simple, that is, twelve MFs for twelve calendar months. A triangle MF is preferred to a Gaussian MF because the asymmetric characteristic of a triangle MF is more flexible. An example of the generated MFs is shown in Figure 5.5.

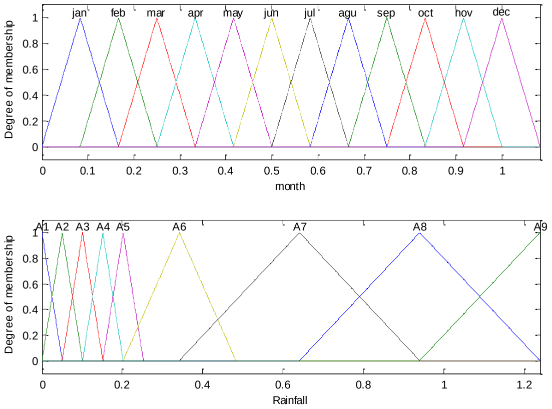


Figure 5.5. An example of the generated MFs of C_t and rainfall (Case 1).

5.3.3. Generate Fuzzy Rules

One drawback of fuzzy systems is the lack of self-learning ability to generalize the input-output relationships from the training data. In fact, many algorithms have been proposed for fuzzy systems to learn from training data. However, those algorithms are not suitable to be used for this method. Until now, the MFs have already been created and the next step is to construct the fuzzy rules.

The cooperative neuro-fuzzy inference system (CNFIS) (Wong et al., 2003) is a technique that combines the advantages of both ANN and FL. This technique uses the learning ability from ANN to learn from the training data and then it is used to extract the fuzzy rules. This approach is adopted to create the fuzzy rules. The procedure to create fuzzy rules is as follows:

Step 1. Use a one hidden layer BPNN to learn from the training data. The number of input nodes is three according to the system inputs, that is, C_t, *1*st-lag and *2*nd-lag inputs. The number of the output node is one corresponding to the system output, predicted rainfall. The number of hidden nodes is selected by trial and error.

Step 2. Prepare the set of input data. The set of input data is all the points in the input space where the degree of membership values is 1 in all dimensions. (*This input data is the antecedent part of the fuzzy rules*).

Step 3. Feed the input data into the BPNN, the output of BPNN are then mapped to the nearest MF in the output dimension of the fuzzy model. (*This output data is the consequence part of the fuzzy rules*).

The constructed readability fuzzy rules are generated in the form:

IF *month*=M_1 AND *1*st *lag*=A_1 AND *2*nd *lag*=B_1 THEN *rainfall*=C_1.
IF *month*=M_2 AND *1*st *lag*=A_2 AND *2*nd *lag*=B_2 THEN *rainfall*=C_2.
...

5.3.4. Optimize Fuzzy Membership Functions

In Figure 5.3, this process consists of the optimization of rainfall's MFs and time's MFs. The first step is to optimize MFs of the 1st lag input, the 2nd lag input and the output, whereas the second step is to optimize MFs of C_t. Actually, these processes can be done in a single process. However, to control the number of parameters in each optimization process, separating the optimization process into two processes is more appropriate. Moreover, the objectives of the optimizations are also different.

The objective of the first optimization is to fit the fuzzy rules and fuzzy MFs of the rainfall variable. As these two parameters come from two methods, they may not fit well. The objective of the second optimization is to capture the uncertainty in time dimension. The proposed methodology hypothesizes that the substantial uncertainty in time dimension will be well extracted when rainfall parameters are already fitted.

In the first optimization, the GA chromosome consists of the sequence of the 1st lag input, the 2nd lag input and the output respectively. In turn, the inputs and output are the sequence of MFs which consists of the three parameters of the triangle MF (a, b, c). The parameters are allowed to search in a small space (Cordon et al., 2001; Ishibuchi et al., 1994).

Let a, b and c be the initial values of the parameters of the triangle MF to be optimized, and let x be a parameter to be optimized (i.e. the parameters a, b or c), the search space of x is [x - α, x + α] and α is defined as

$$\alpha = \sigma * \frac{1}{2}(c - a) \tag{5.2}$$

where σ is the user's control parameter with the range of [0,1]. In other words, the search space α is dependent on the size of the initial MF.

In the second optimization process, the GA chromosome is the sequence of triangle MFs of the C_t. The search space is set in a different way. Figure 5.6 demonstrates a conceptual example of how to set the search space of parameters *a*, *b* and *c*.

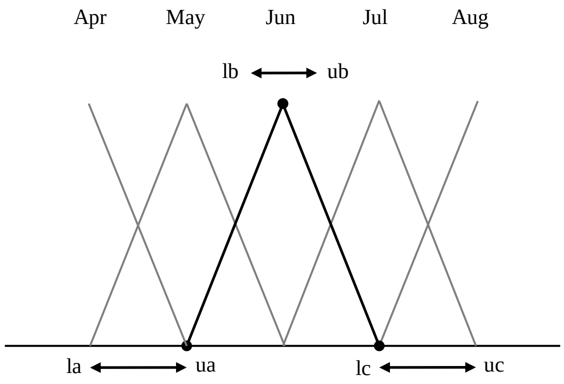


Figure 5.6. Search space of the triangle MFs of the input C_t.

Search space of parameters *a* and *c* are set in this manner in order to allow the FIS model to capture the uncertainty in time between months. The search space of parameter *b* is set in this manner in order to allow the FIS model to reduce some firing strength of that month. Due to these settings, the established model is capable of preventing the MFs of C_t from exhibiting the issue of indistinguishability.

In this approach, the search space for parameters *a* and *c* is equal to the intersect area between the two MFs and the search space for parameter *b* is equal to a half of that intersect area. However, this search space can be changed in accordance with the user's requirements. For both optimizations, the fitness function is to minimize sum square error between observed values (O) and predicted values (P) of the training data and it is given as

$$SSE = \sum_{i=1}^{S}(P_i - O_i)^2 \quad (5.3)$$

where S is the number of training data.

5.4. Evaluation of the Proposed Methodology

In order to evaluate prediction accuracy, the established models will be compared to commonly-used time series prediction models in hydrological study, that is, ARMA (Wu et al., 2010; Wang et al., 2009), BPNN (Wu & Chau, 2013; Wu et al., 2010; Jain & Kumar, 2007; Somvanshi et al., 2006) and ANFIS (Nayak et al., 2004; Zounemat-Kermani & Teshnehlab, 2008; Wang et al., 2009). Furthermore, the final models will also be compared to the BPNN models that are used to create the fuzzy rules as well as the established models before the first and the second optimization. This study uses the models to predict rainfall one step ahead (or one month).

5.4.1. Models Establishment

In order to select the optimal parameters for ARMA models, the Akaike information criterion (AIC) is adopted (Wu et al., 2010; Wang et al., 2009). This approach generated the ARMA models from the training data by replacing parameters p and q of the ARMA models from 0 to 12. The parameters that provide the lowest AIC value are used for the ARMA models. Table 5.2 shows the optimal parameters and the lowest AIC values for the ARMA models of the eight datasets.

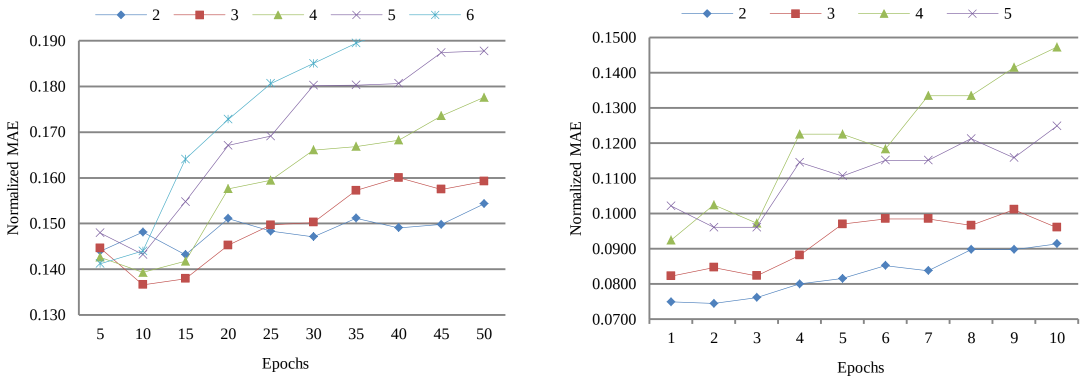
Table 5.2. The selected parameters and the lowest AIC values.

Case	(p, q)	AIC	Case	(p, q)	AIC
1	(4,4)	13.417	5	(5,3)	13.751
2	(10,9)	13.982	6	(12,1)	13.536
3	(6,3)	13.379	7	(12,0)	14.334
4	(8,11)	14.182	8	(11,2)	13.850

For BPNN and ANFIS, unlike Box-Jenkins models, there is no consistent theory to select the appropriate inputs. However, the work of Wu et al. (2010), Wu and Chau (2013), and Wang et al. (2009) recommended that ACF and PACF can be applied to select the appropriate inputs for these non-linear models. Considering ACF and PACF in Figure 5.4 (and also in Appendix B), they suggest that in general monthly rainfall time series in this study area show autoregressive process up to lag twelve. Therefore, 12-lag antecedence inputs should provide sufficient information for the models.

The architecture of BPNN and ANFIS are twelve inputs and one output. The optimal numbers of parameters were selected by a trial and error procedure. To investigate the optimal numbers of parameters, the training data are separated into two parts. The first part is used to train the models and the second part is used to test the models.

In the case of BPNN, the experiments varied the numbers of hidden nodes from two to six. An example of the results for BPNN is shown above (a) in Figure 5.7. From the experiment, the number of two or three hidden nodes can provide minimum error. Table 5.3 summarizes the number of hidden nodes (*hn*) of BPNN of the eight datasets. Furthermore, when the number of training epochs is larger than 15, error from the testing data started to increase. Therefore, the number of epochs is limited to 15.



(a) (b)

Figure 5.7. An example of trial and error processes to determine the optimal number of parameters of BPNN (a) and ANFIS (b).

In the case of ANFIS, the prototype-based fuzzy modeling was used. The Sugeno-type FIS was generated from FCM and was then optimized by the ANFIS procedure. An example of the results for ANFIS is shown above (b) in Figure 5.7. The experiments pointed out that a small number of clusters provided better prediction results. The effects of the number of epochs to the prediction error were more sensitive than for BPNN. Only two or three epochs were enough to generalize data. The number of selected cluster (*cls*) of ANFIS is presented in Table 5.3.

Table 5.3. The selected number of parameters of BPNN and ANFIS.

Case	hn / cls	Case	hn / cls
1	3 / 2	5	2 / 2
2	2 / 2	6	3 / 3
3	3 / 3	7	2 / 2
4	3 / 2	8	3 / 2

In the case of the proposed model, BPNN used to create fuzzy rules were selected in the same manner. The value of σ in the first optimization was set to 0.25 so as to preserve the shape of MFs after the first optimization. The number of population was set to 100 for both optimizations and the number of generations was set to 30 and 15 for the first and second optimization respectively, where the best and average fitness values were met. The reproduction scheme elite count was set to 2 and the crossover fraction was set to 0.8.

5.4.2. Quantitative Results

From now on, BPNN_{12} refers to BPNN with twelve antecedence lags input, BPNN_3 refers to BPNN with C_t and two antecedence lags input, MFIS–ORG is the proposed model before optimization, MFIS–OPT_1 and MFIS–OPT_2 are the proposed models after the first and the second optimization, respectively. Tables 5.4, 5.5 and 5.6 show the experimental results. The MAE and RMSE of each case are normalized by its mean of the dataset for comparison purposes.

The row Average in the tables refers to the average values from all cases and the row Improvement in the tables refers to the percentage improvement of the average values in comparison with the ARMA models. Figure 5.8 shows the average values (use the left axis for reference) and the improvement percentage (use the right axis for reference) of all models.

Table 5.4. Normalized mean absolute errors.

Case Study	**ARMA**	**BPNN$_{12}$**	**ANFIS**	**BPNN$_3$**	**MFIS–ORG**	**MFIS–OPT$_1$**	**MFIS–OPT$_2$**
1	0.740	0.566	0.555	0.512	0.530	0.486	0.483
2	0.480	0.484	0.387	0.393	0.390	0.386	0.381
3	0.595	0.620	0.580	0.498	0.531	0.499	0.420
4	0.549	0.616	0.522	0.490	0.529	0.477	0.472
5	0.623	0.611	0.518	0.547	0.515	0.502	0.464
6	0.570	0.660	0.595	0.527	0.540	0.525	0.515
7	0.518	0.567	0.510	0.493	0.448	0.443	0.440
8	0.419	0.526	0.420	0.432	0.486	0.436	0.410
Average	0.562	0.581	0.511	0.487	0.496	0.469	0.448
Improvement	0	-3.48	9.07	13.37	11.67	16.46	20.22

Table 5.5. Normalized root mean square errors.

Case Study	**ARMA**	**BPNN$_{12}$**	**ANFIS**	**BPNN$_3$**	**MFIS–ORG**	**MFIS–OPT$_1$**	**MFIS–OPT$_2$**
1	0.956	0.773	0.769	0.764	0.780	0.730	0.729
2	0.677	0.710	0.593	0.634	0.651	0.625	0.616
3	0.867	0.974	0.845	0.788	0.803	0.776	0.672
4	0.799	0.966	0.764	0.761	0.792	0.740	0.709
5	0.880	0.874	0.771	0.822	0.773	0.752	0.708
6	0.851	1.013	0.845	0.766	0.795	0.764	0.754
7	0.651	0.743	0.667	0.631	0.586	0.584	0.575
8	0.574	0.757	0.603	0.623	0.703	0.632	0.598
Average	0.782	0.851	0.732	0.724	0.735	0.700	0.670
Improvement	0	-8.87	6.37	7.42	5.95	10.43	14.32

Table 5.6. Correlation coefficient.

Case Study	**ARMA**	**BPNN$_{12}$**	**ANFIS**	**BPNN$_3$**	**MFIS–ORG**	**MFIS–OPT$_1$**	**MFIS–OPT$_2$**
1	0.539	0.731	0.764	0.749	0.748	0.794	0.800
2	0.787	0.761	0.844	0.822	0.824	0.838	0.844
3	0.543	0.572	0.583	0.632	0.623	0.645	0.732
4	0.797	0.740	0.835	0.846	0.809	0.837	0.855
5	0.666	0.656	0.738	0.703	0.741	0.755	0.781
6	0.587	0.465	0.613	0.695	0.663	0.704	0.718
7	0.466	0.371	0.468	0.570	0.612	0.630	0.634
8	0.776	0.664	0.772	0.757	0.691	0.750	0.767
Average	0.645	0.620	0.702	0.722	0.714	0.744	0.766
Improvement	0	-3.90	8.83	11.87	10.64	15.35	18.76

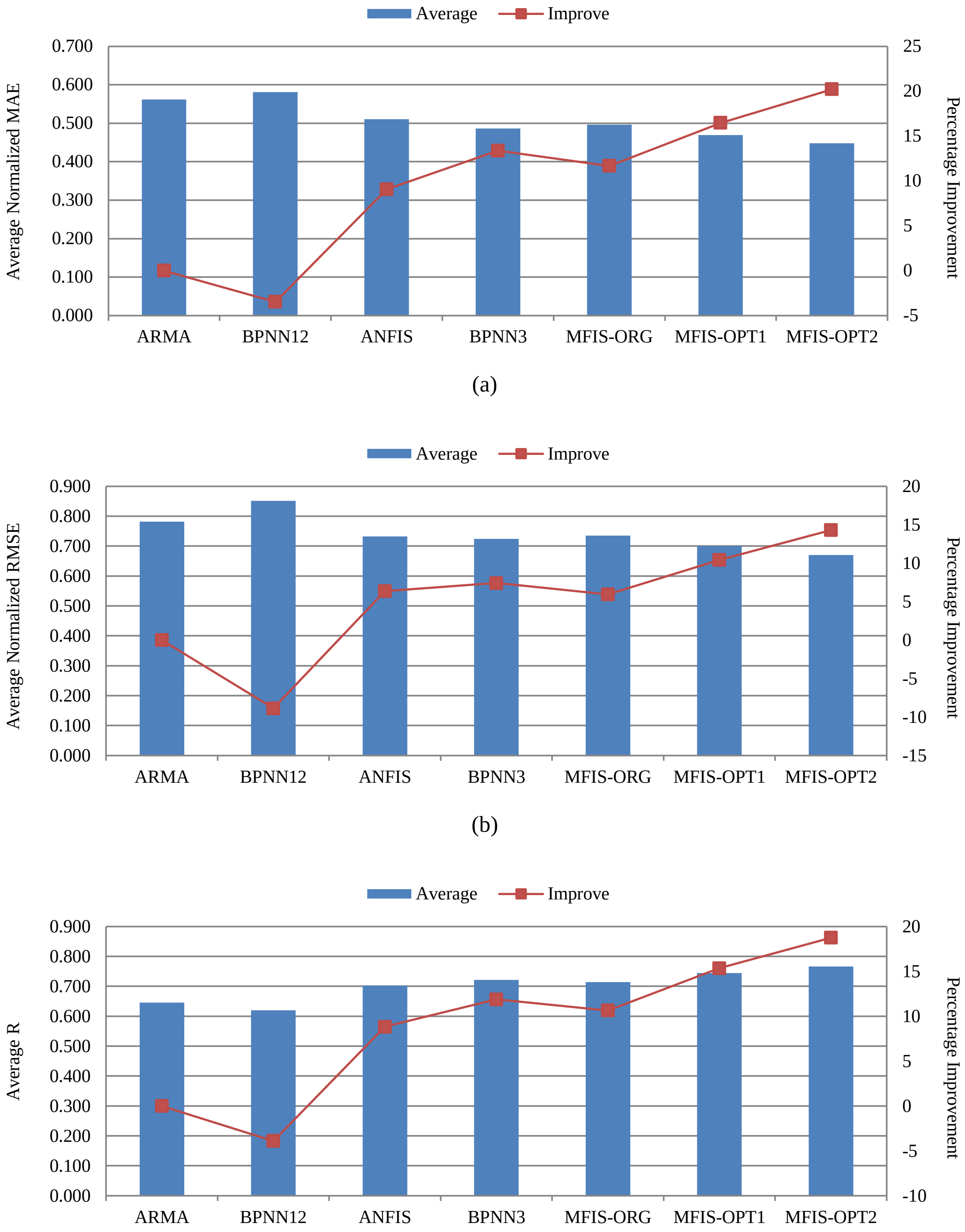


Figure 5.8. Average values and improvement of models.

In terms of MAE, the best accuracy comes from MFIS–OPT$_2$ in all cases. In contrast, the lowest accuracy appears in BPNN$_{12}$ in six cases and in ARMA in two cases. ANFIS shows relatively good accuracy in comparison with ARMA and BPNN$_{12}$. BPNN$_3$ provides better accuracy than ANFIS and MFIS–ORG in five cases. The accuracy of MFIS–ORG$_2$ is slightly better than MFIS–ORG$_1$ in general.

In terms of RMSE, the best accuracy comes from MFIS–OPT$_2$ in six cases, ANFIS and ARMA show the best accuracy in Case two and Case eight respectively. Again, BPNN$_{12}$ shows unsatisfactory prediction accuracy. Overall, ANFIS and BPNN$_3$ provide rather compatible results and prove slightly superior than MFIS–ORG. The accuracy of MFIS–OTP$_2$ has been improved from MFIS–OPT$_1$ in general.

In terms of R, the best accuracy comes from MFIS–OPT$_2$ in six cases. ANFIS shows the best on Case 8 and ARMA shows the best on Case 2. Overall, the results gained from the R measure corresponded to the RMSE measure. Overall the prediction accuracy can be ordered as MFIS–OPT$_2$ > MFIS–OPT$_1$ > BPNN$_3$ > ANFIS ≈ MFIS–ORG > ARMA > BPNN$_{12}$.

5.4.3. Qualitative results

Figure 5.9 shows an example of optimized fuzzy MFs and Figure 5.10 shows some parts of the generated fuzzy rules from MFIS–OPT$_2$. At the level of fuzzy set, the criteria of distinguishability, normality and completeness of partition in input space of MFs are preserved after optimization. Optimized MFs are well structured and clearly represented.

The number of fuzzy sets in each input dimension ranges from 9 to 13 depending on the fluctuation in the time series data. Although these numbers are higher than those recommended by Zhou and Gan (2008), an appropriate number of MFs in each input should not exceed 7 ± 2. These slightly higher numbers are necessary because they are a good explanation to the fluctuation in the time series data.

At the level of fuzzy rules, the proposed model provides good readability of single rules with only three conditions in antecedent part while ANFIS has twelve conditions. Since the fuzzy rules are extracted from generalized $BPNN_3$ by using a mapping procedure, the consistence and completeness of fuzzy rules are met. In the case of the transparency of the rules structure, as the proposed model presents the month feature as an input to the systems the fuzzy rule "*IF month = M AND 1st lag = A AND 2nd lag = B THEN rainfall = C*" can characterize or explain the monthly rainfall time series data in a clear way.

However, although many interpretable fuzzy criteria have been met, the number of generated fuzzy rules is still the problem because the proposed model has a large number of fuzzy rules. For example, if the number of MFs in the model is 9, the number of fuzzy rules generated is 972. This problem needs to be addressed.

For the monthly rainfall time series data, the number of redundant fuzzy rules (i.e. high rainfall in the dry period and vice versa) can be removed later by human analysts. This can be done by using expert knowledge or by observations from historical records. Due to the good readability structure of the fuzzy rules, the task of removal will be easier. Table 5.7 summarizes the qualitative results of the proposed models.

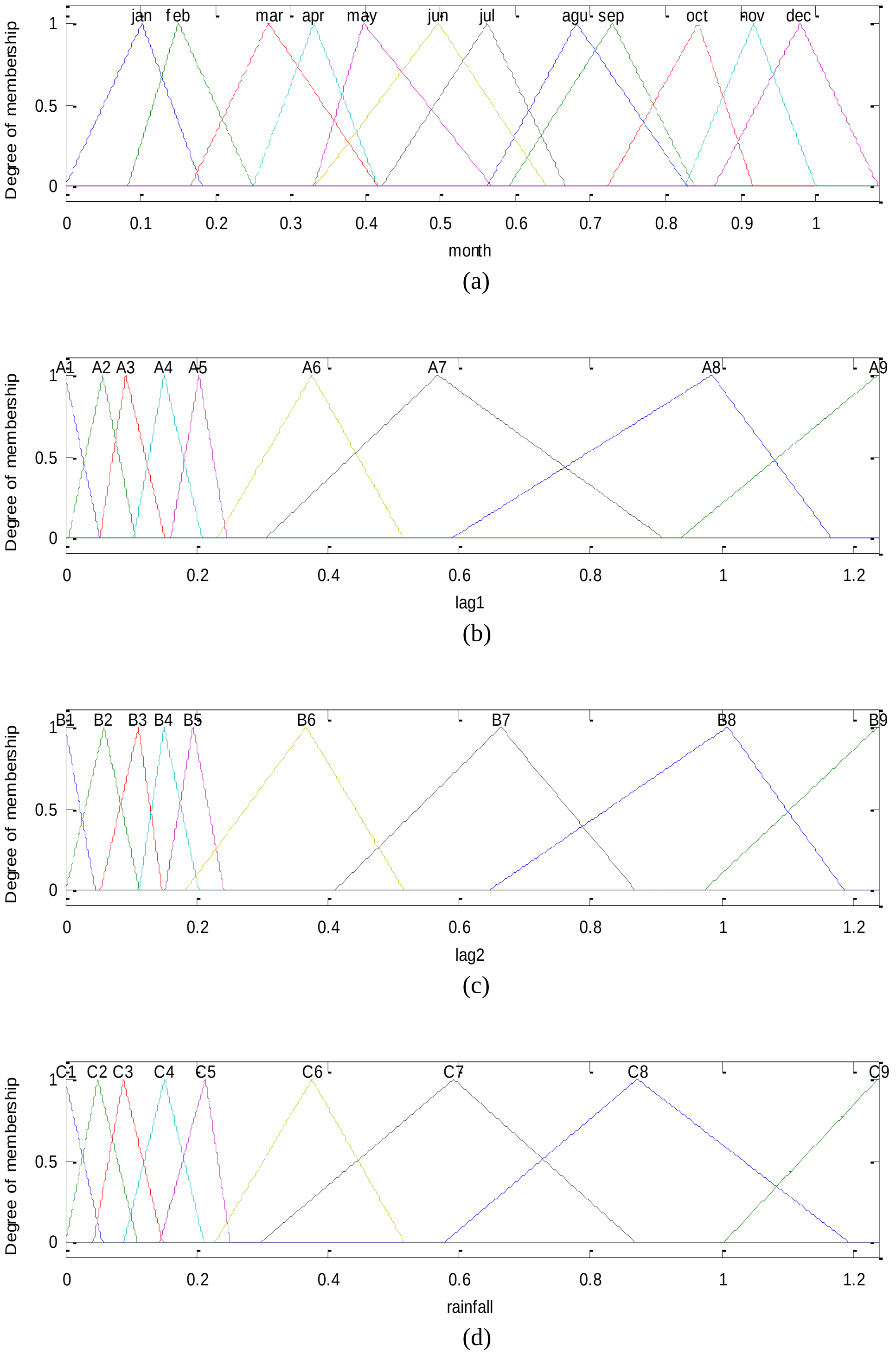
Table 5.7. Summary of qualitative results.

Level	Criterion	Weak	Fair	Good	Strong
Low-Level	Distinguishability				x
	Moderate number of MFs		x		
	Coverage or completeness of partition of input variable			x	
	Normalization				x
High-Level	Rule-base parsimony and simplicity	x			
	Readability of single rule			x	
	Consistency of rules				x
	Completeness of rules				x
	Transparency of rule structure			x	

Figure 5.11 represents the uncertainty in time dimension of the monthly rainfall time series data via fuzzy parameters. This interpretability characteristic is an advantage of the proposed model. These fuzzy MFs allow human analysts to investigate the uncertainty of rainfall data between months. As a consequence, further analysis into the monthly time series data can be enhanced.

Up to this point, the experimental results have been presented in terms of the quantitative and qualitative aspects. The results showed that the proposed model provided satisfactory prediction accuracy and acceptable model interpretability. These experimental results have suggested the following:

- Although BPNN$_{12}$ did not provide superior results than ARMA in this experiment, it does not mean that BPNN$_{12}$ is not an appropriate method. The dataset used is relatively small. The number of training data may not be enough for large inputs of BPNN$_{12.}$



(a)

(b)

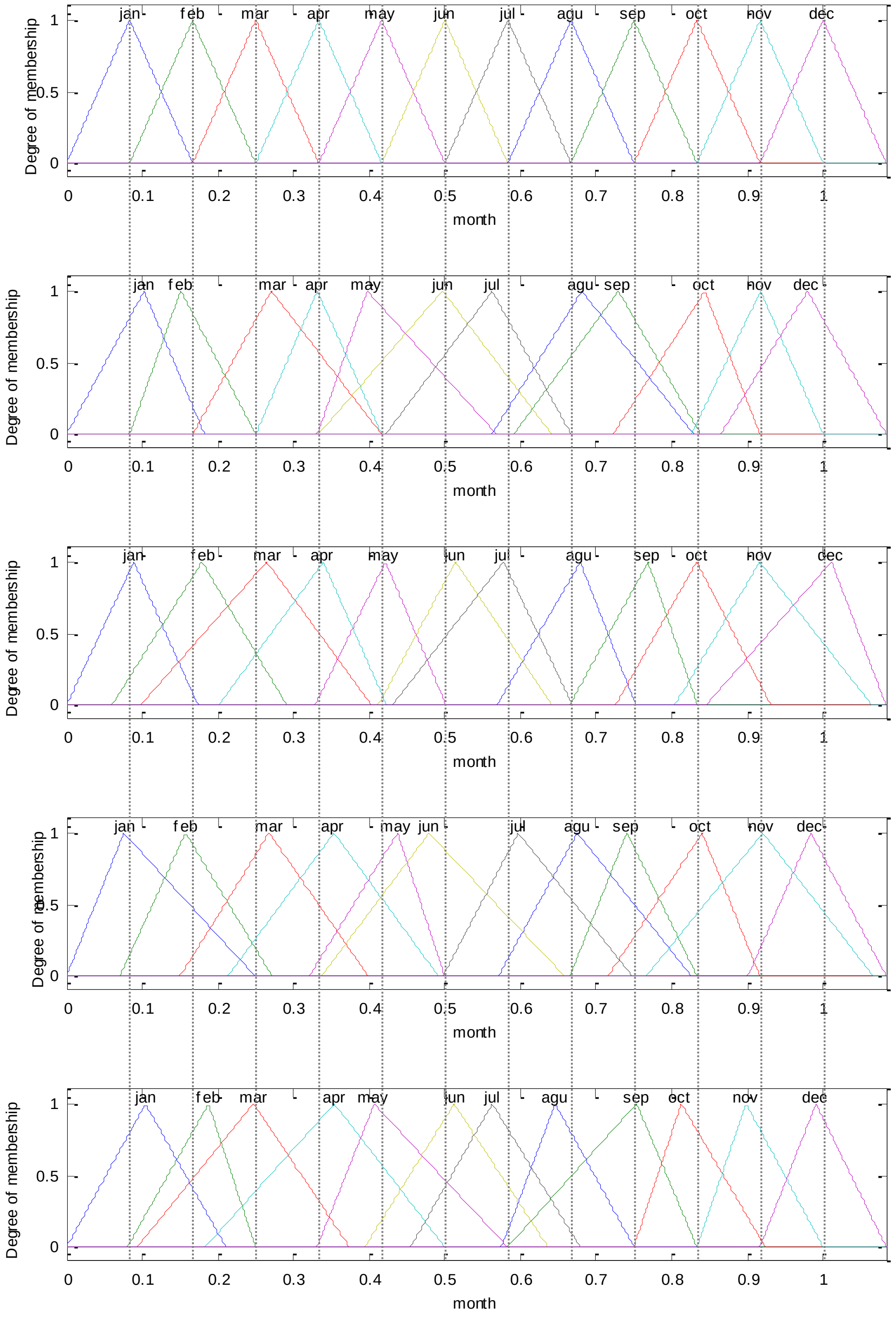
(c)

(d)

Figure 5.9. An example of optimized fuzzy MFs (Case 1).

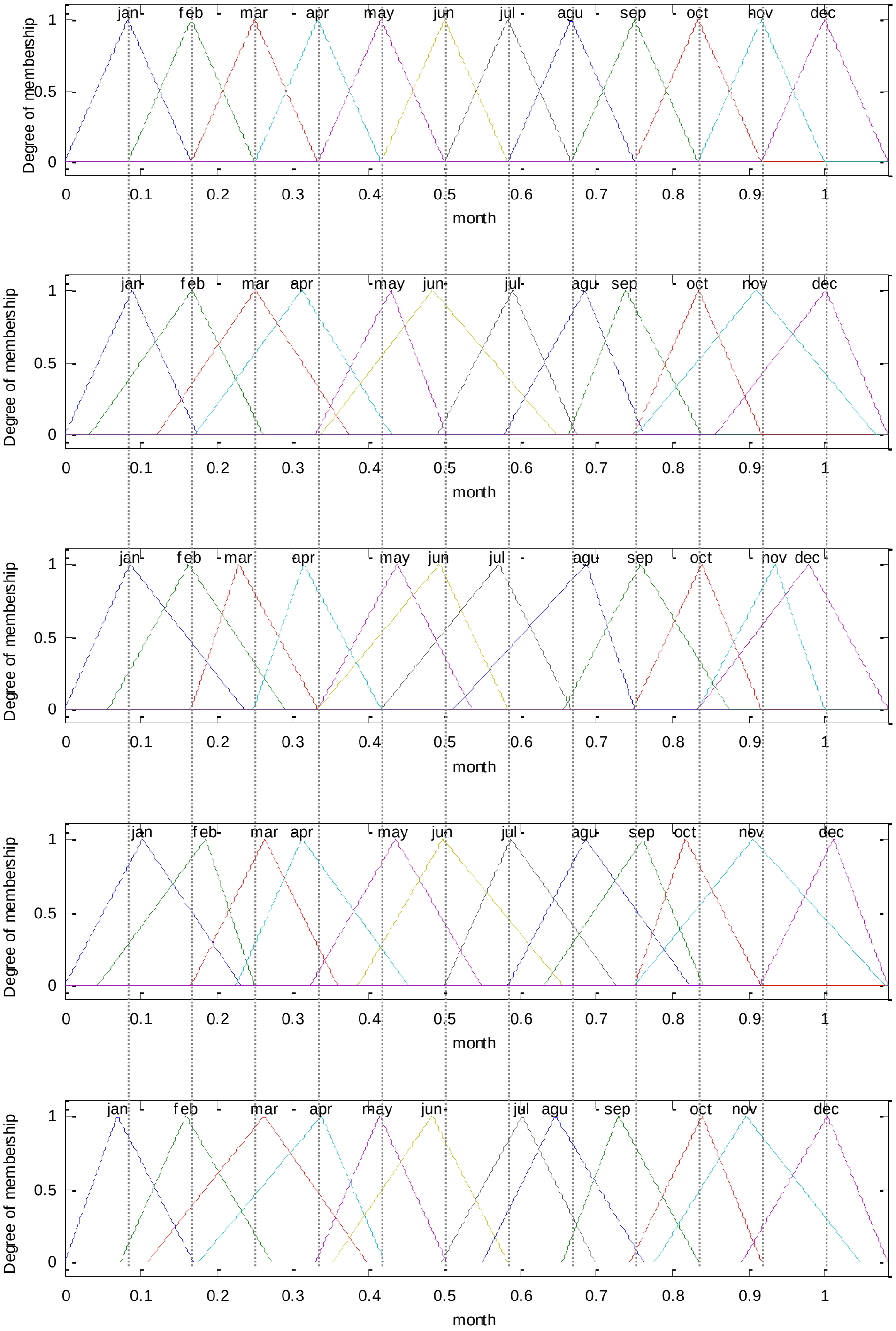
167. If (month is mar) and (lag1 is A1) and (lag2 is B5) then (rainfall is C3)
168. If (month is mar) and (lag1 is A1) and (lag2 is B6) then (rainfall is C3)
169. If (month is mar) and (lag1 is A1) and (lag2 is B7) then (rainfall is C2)
170. If (month is mar) and (lag1 is A1) and (lag2 is B8) then (rainfall is C1)
171. If (month is mar) and (lag1 is A1) and (lag2 is B9) then (rainfall is C2)
172. If (month is mar) and (lag1 is A2) and (lag2 is B1) then (rainfall is C4)
173. If (month is mar) and (lag1 is A2) and (lag2 is B2) then (rainfall is C4)
174. If (month is mar) and (lag1 is A2) and (lag2 is B3) then (rainfall is C4)
175. If (month is mar) and (lag1 is A2) and (lag2 is B4) then (rainfall is C4)
176. If (month is mar) and (lag1 is A2) and (lag2 is B5) then (rainfall is C3)
177. If (month is mar) and (lag1 is A2) and (lag2 is B6) then (rainfall is C3)
178. If (month is mar) and (lag1 is A2) and (lag2 is B7) then (rainfall is C2)
179. If (month is mar) and (lag1 is A2) and (lag2 is B8) then (rainfall is C2)
180. If (month is mar) and (lag1 is A2) and (lag2 is B9) then (rainfall is C2)
181. If (month is mar) and (lag1 is A3) and (lag2 is B1) then (rainfall is C4)
182. If (month is mar) and (lag1 is A3) and (lag2 is B2) then (rainfall is C4)
183. If (month is mar) and (lag1 is A3) and (lag2 is B3) then (rainfall is C4)
184. If (month is mar) and (lag1 is A3) and (lag2 is B4) then (rainfall is C3)
185. If (month is mar) and (lag1 is A3) and (lag2 is B5) then (rainfall is C3)
186. If (month is mar) and (lag1 is A3) and (lag2 is B6) then (rainfall is C3)
187. If (month is mar) and (lag1 is A3) and (lag2 is B7) then (rainfall is C2)
188. If (month is mar) and (lag1 is A3) and (lag2 is B8) then (rainfall is C2)
189. If (month is mar) and (lag1 is A3) and (lag2 is B9) then (rainfall is C2)
190. If (month is mar) and (lag1 is A4) and (lag2 is B1) then (rainfall is C4)
191. If (month is mar) and (lag1 is A4) and (lag2 is B2) then (rainfall is C4)
192. If (month is mar) and (lag1 is A4) and (lag2 is B3) then (rainfall is C4)
193. If (month is mar) and (lag1 is A4) and (lag2 is B4) then (rainfall is C3)
194. If (month is mar) and (lag1 is A4) and (lag2 is B5) then (rainfall is C3)
195. If (month is mar) and (lag1 is A4) and (lag2 is B6) then (rainfall is C3)
196. If (month is mar) and (lag1 is A4) and (lag2 is B7) then (rainfall is C2)
197. If (month is mar) and (lag1 is A4) and (lag2 is B8) then (rainfall is C2)
198. If (month is mar) and (lag1 is A4) and (lag2 is B9) then (rainfall is C2)
199. If (month is mar) and (lag1 is A5) and (lag2 is B1) then (rainfall is C4)
200. If (month is mar) and (lag1 is A5) and (lag2 is B2) then (rainfall is C4)
201. If (month is mar) and (lag1 is A5) and (lag2 is B3) then (rainfall is C4)
202. If (month is mar) and (lag1 is A5) and (lag2 is B4) then (rainfall is C3)
203. If (month is mar) and (lag1 is A5) and (lag2 is B5) then (rainfall is C3)
204. If (month is mar) and (lag1 is A5) and (lag2 is B6) then (rainfall is C3)
205. If (month is mar) and (lag1 is A5) and (lag2 is B7) then (rainfall is C2)
206. If (month is mar) and (lag1 is A5) and (lag2 is B8) then (rainfall is C2)
207. If (month is mar) and (lag1 is A5) and (lag2 is B9) then (rainfall is C1)
208. If (month is mar) and (lag1 is A6) and (lag2 is B1) then (rainfall is C4)
209. If (month is mar) and (lag1 is A6) and (lag2 is B2) then (rainfall is C4)
210. If (month is mar) and (lag1 is A6) and (lag2 is B3) then (rainfall is C4)
211. If (month is mar) and (lag1 is A6) and (lag2 is B4) then (rainfall is C3)
212. If (month is mar) and (lag1 is A6) and (lag2 is B5) then (rainfall is C3)
213. If (month is mar) and (lag1 is A6) and (lag2 is B6) then (rainfall is C2)
214. If (month is mar) and (lag1 is A6) and (lag2 is B7) then (rainfall is C2)
215. If (month is mar) and (lag1 is A6) and (lag2 is B8) then (rainfall is C3)
216. If (month is mar) and (lag1 is A6) and (lag2 is B9) then (rainfall is C1)
217. If (month is mar) and (lag1 is A7) and (lag2 is B1) then (rainfall is C4)
218. If (month is mar) and (lag1 is A7) and (lag2 is B2) then (rainfall is C4)
219. If (month is mar) and (lag1 is A7) and (lag2 is B3) then (rainfall is C3)
220. If (month is mar) and (lag1 is A7) and (lag2 is B4) then (rainfall is C3)
221. If (month is mar) and (lag1 is A7) and (lag2 is B5) then (rainfall is C3)
222. If (month is mar) and (lag1 is A7) and (lag2 is B6) then (rainfall is C2)
223. If (month is mar) and (lag1 is A7) and (lag2 is B7) then (rainfall is C2)
224. If (month is mar) and (lag1 is A7) and (lag2 is B8) then (rainfall is C2)
225. If (month is mar) and (lag1 is A7) and (lag2 is B9) then (rainfall is C1)
226. If (month is mar) and (lag1 is A8) and (lag2 is B1) then (rainfall is C4)
227. If (month is mar) and (lag1 is A8) and (lag2 is B2) then (rainfall is C4)

Figure 5.10. Samples of the generated fuzzy rules (Case 1).



Note: Case 1 to Case 4 are presented from the second graph to the bottom.

Figure 5.11. A presentation of uncertainty in time dimension through fuzzy MFs



Note: Case 5 to Case 8 are presented from the second graph to the bottom.

Figure 5.11. (cont.) A presentation of uncertainty in time dimension through fuzzy MFs.

- ANFIS is capable of capturing the uncertainty in the data because it provided better results than BPNN$_{12}$ and ARMA. However, the use of ANFIS should be handled with care because such a model showed higher sensitivity than BPNN$_{12}$. As can be seen in Figure. 5.7, ANFIS (b) tends to lose generalization in only a few epochs. This is one reason that BPNN was used instead of ANFIS to generate fuzzy rules in the proposed method.

- Using the time coefficient, C_t, as the supplementary feature for the periodic time series data is an effective way to improve the prediction accuracy. As shown in the results, BPNN$_3$ provided considerable improvement from BPNN$_{12}$ and ANFIS. However, the use of the time coefficient feature is limited to only periodic time series data.

- The conversion from BPNN$_3$ to MFIS–ORG inevitably decreases some prediction accuracy. However, this issue can be addressed by the optimization process. One can see that the prediction accuracy of MFIS–ORG was improved when fuzzy rules and fuzzy MFs were optimized (MFIS–OPT$_1$).

- The uncertainty in the time dimension has significant impact on the prediction accuracy of the proposed models. Once the MFs in the time dimension were optimized (MFIS–OPT$_2$), the prediction accuracy of the proposed models improved.

However, all these observation are based on the average results. In the details of converting from BPNN$_3$ to MFIS–OPT$_2$, the results showed that not all cases provided significant improvement. Cases 3, 5 and 7 provided large improvement, up to 10 percent.

Cases 1, 4 and 8 provided moderate improvement, about 3.5 to 5.5 percent. Cases 2 and 6 showed small improvement, approximately 2.5 percent. This difference is subject to the following:

- If the uncertainty in time series data is not strong, the prediction accuracy between those two models may not be different because BPNN is capable of handling weak uncertainty.
- In order to preserve the interpretability of the proposed model, search space is limited to a small region. GA may not be able to find a better optimal solution in the constrained search space.

5.5. Conclusion

This chapter proposed an integration of intelligent techniques, namely, fuzzy logic, an artificial neural network and a genetic algorithm to create interpretable fuzzy models for the monthly rainfall time series prediction. The proposed models were evaluated by eight monthly rainfall time series data in the northeast region of Thailand.

The experimental results illustrated that, in terms of the quantitative aspect, the proposed models provided satisfactory prediction accuracy in comparison with commonly-used time series prediction models in hydrology. In terms of the qualitative aspect, the proposed models can mostly satisfy the interpretability fuzzy criteria. Furthermore, the uncertainty in the time dimension of the data can be represented through fuzzy MFs.

However, one disadvantage of the proposed models is the large number of fuzzy rules generated. The high number of generated fuzzy rules will cause the interpretability of the proposed models to deteriorate. Although, in practice a number of redundant fuzzy rules can be removed subsequently by human analysts, it would be better if this problem is addressed. In the next chapter, one solution to this problem is presented.

CHAPTER 6

A MODULAR FUZZY SYSTEM FOR MONTHLY RAINFALL TIME SERIES PREDICTION

6.1. Introduction

In the previous chapter, a methodology to establish an interpretable fuzzy model for time series prediction has been proposed. The proposed methodology presented in Chapter 5 provided satisfactory prediction accuracy and provided adequate model interpretability. However, one fuzzy interpretability criterion is still not satisfied in the previous chapter, that is, rule base parsimony and simplicity, as the proposed models in the previous chapter can generate a large number of fuzzy rules.

In Chapter 4, the modular technique had been used to improve interpolation accuracy of the GAFIS model. Conceptually, the modular technique divides the whole data into several parts in order to reduce the complexity of the data. As a result, the complexity in the modeling process is decreased. This technique will be used herein again to simplify the complicacy of fuzzy parameters of the MFIS–OPT$_2$ used for time series prediction.

The chapter is organized as follows: Section 2 presents the case study and the datasets used in this chapter; Section 3 presents the modular fuzzy inference systems for monthly rainfall time series prediction; in Section 4 the proposed models will be evaluated and recommendations will be presented; and, finally, Section 5 is the conclusion.

6.2. Case Study Area and Datasets

The case study area and the datasets used in this chapter are the same as in the previous chapter. Some general information about the eight case studies has been shown in Table 5.1. The training data and testing data are also the same for comparison purposes.

6.3. Establishing a Modular Fuzzy System

In this chapter, again, the modular technique named multiple expert systems has been adopted as shown in Figure 4.1. However, this approach has been adapted according to the characteristics of monthly time series data. The details of the proposed model are as follows.

6.3.1. The Model's Architecture

The architecture of the proposed model is shown in Figure 6.1. The model consists of an input layer, a prediction layer, an aggregation layer and an output layer. The input layer is used to feed the input data into the associated prediction modules. The prediction layer consists of twelve prediction modules (predictors) associated with the calendar months. The function of these modules is to generalize the input-output relationships of the rainfall pattern of the month. The aggregation layer consists of twelve aggregation modules (aggregators) associated with the calendar months, which are the same as the prediction modules. The function of these modules is to aggregate the outputs from the associated prediction modules by using the combination weights. The output layer is used to derive the final prediction of the system.

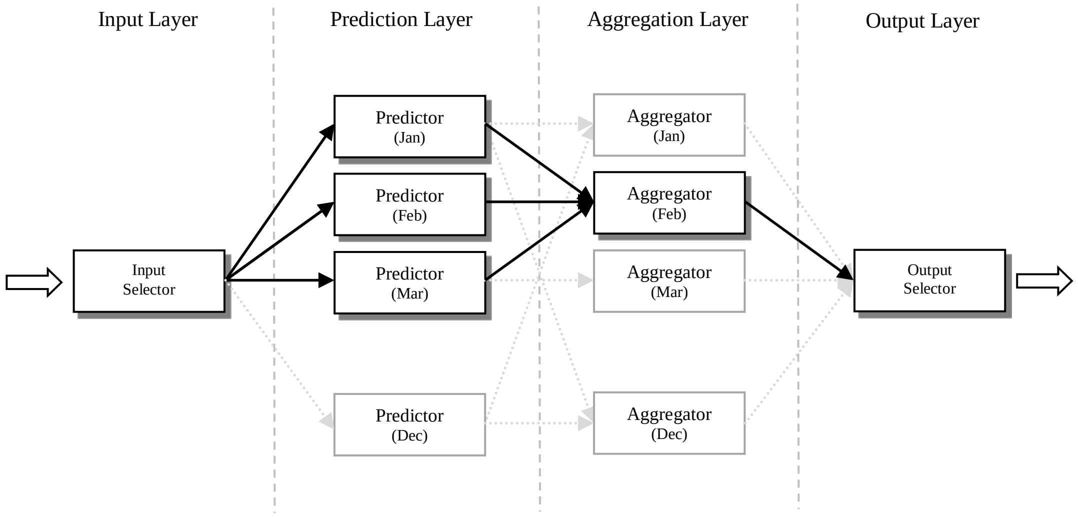


Figure 6.1. An overview of the proposed model's architecture.

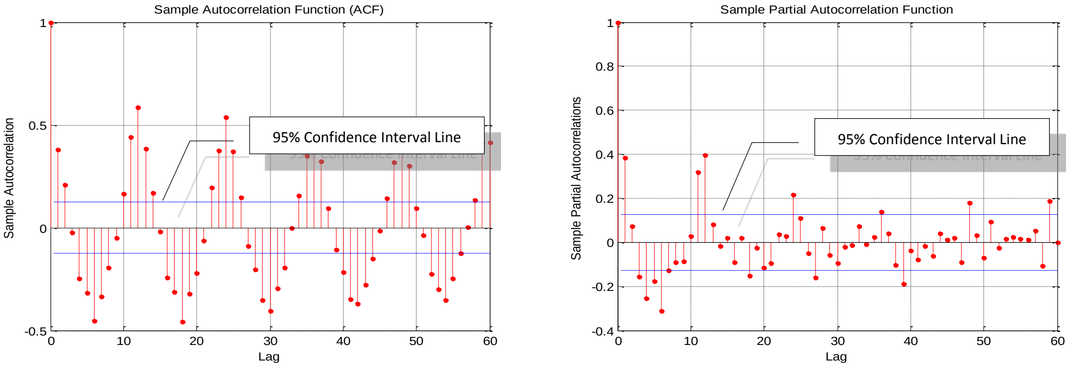
An example of the model's operation is as follows: suppose that the model is used to predict the rainfall value in February (see Figure 6.1). Firstly, the input selector feeds input data into the associated predictor (i.e. Feb), the previous predictor (i.e. Jan) and the next predictor (i.e. Mar). Secondly, the outputs from the predictors are aggregated by the associated aggregator (i.e. Feb). Finally, the output selector receives the aggregated output from the associated aggregator and provides the final output. Therefore, to perform the prediction, three consecutive predictors and one aggregator will be used.

6.3.2. Input Identification

As discussed before, the objective in predicting rainfall using antecedent values is to generalize the relationships of the form $y = f(x^m)$, where x^m is an m-dimensional input vector representing rainfall values with different time lags, and y is an one-dimensional output value representing the predicted rainfall. Since x^m is not known beforehand and there is no consistent procedure to define x^m for non-linear models, ACF and PACF are then used to identify an appropriate input again.

Figure 6.2 represents an example of ACF and PACF of the monthly rainfall time series data (ACF and PACF of all cases are shown in Appendix B). In general, ACF exhibits peak values at lag 12 and PACF exhibits a significant correlation at 95% confidence level interval up to lag 12. Thus, these functions suggest that twelve antecedent rainfall values contain sufficient information to predict future rainfall.

However, for the proposed model, as the whole system is decomposed into twelve sub-modules, 12-lag information may be redundant to the sub-module. Instead, the model suggests the first lag that crosses the confidence line in PACF as minimum information for each sub-module. Therefore, two antecedent rainfalls are considered as the input for each sub-module.



(a) (b)

Figure 6.2. ACF (a) and PACF (b) of monthly rainfall time series data (Case 1).

It can be observed that the inputs to the systems of the proposed models in the previous chapter and this chapter are identical. In both proposed models, the similarity is that the 1st lag and 2nd lag antecedence of the rainfall data and the time variable (month) are used as the inputs to the systems. The difference is that the proposed model in the previous chapter employs the time variable to provide the prediction, whereas the proposed model in this chapter employs the time variable to select the sub-modules.

6.3.3. Create Prediction Modules

In order to create interpretable fuzzy models for the prediction modules (PM), the prototype-based fuzzy modeling is used. Similar to those fuzzy models in Chapters 3 and 4, a Mamdani-type FIS model is created from the FCM technique. However, the method used to determine the number of clusters is different. As the number of training data in each sub-module is rather small (i.e. approximately 17 to 18 records), cluster analysis seems to be redundant, and only a simple clustering method is required.

Therefore, the number of clusters in the FCM method is determined by using the subtractive clustering (Jang et al., 1997). This technique is another commonly-used method in a hydrological study (Nayak & Sudheer, 2008). One parameter that has to be defined in the subtractive method is the vector that specifies the cluster center's range of influence in each of the data dimensions, assuming the data falls within a unit hyper box (radii). To ensure that the range of the subtractive method examines at least half of the range of data in a unit hyper box, this parameter is set to 0.5.

The subtractive method can make the modeling process more convenient by omitting the clustering analysis process. However, there is no report of this automatic technique to ensure the parsimonious number of created prototypes on the large dataset. In this case, for each sub-module, the number of training data is small and thus the complexity of training data is reduced. This technique is appropriate to reduce the complexity in the modeling process.

6.3.4. Create Aggregation Modules

In order to derive the final output of the system, the aggregation modules (AM) are used to combine the outputs y_i from $\{PM_i\}_{i=1}^{K}$ by using the combination weights. The combination formula is

$$y = \sum_{i=1}^{K} w_i y_i \tag{6.1}$$

where $w_i \geq 0$ and $\sum_{i=1}^{K} w_i = 1$. These weights can be viewed as the measure of "closeness" of the rainfall pattern, in which the rainfall pattern is close to the rainfall pattern of that PM. A larger combination weight indicates that the rainfall pattern is closer to that PM than the others. For comparison purposes, however, the sequential method and the non-sequential method are used to evaluate the combination weights. Bayesian learning is used for the sequential method and non-linear programming is used for the non-sequential method.

6.3.4.1. The Sequential Method

Wang et al. (2010) proposed the Bayesian learning method that aggregates information from modular neural networks in a sequential way and this method is adopted for the current model. Since the nature of time series data is sequential in time, the aggregation in a sequential fashion should be more appropriate than in a non-sequential fashion. The steps to create combination weights for the AM from associated PMs are as follows:

Step 1. Prepare training data for associated PMs.

Step 2. Ordering *S* training data records from oldest to newest.

Step 3. For *i* = 1 to *S*:

Step 3a. Calculate likelihood function (LF) values, $\omega_j^{\,i}$ *(j = 1,2, ... ,K)* as

$$\omega_\mathrm{j} = \frac{1/sse_j}{\sum_{k=1}^{K} 1/sse_k} \tag{6.2}$$

where sse_j is the training error of the j^{th} prediction module, *K* is the number of prediction modules aggregated (i.e. *K = 3*).

Step 3b. Update the combination weights by using Bayesian reasoning as

$$W_j^i = \begin{cases} w_j^i = w_j^{i-1} & if \;\; \sum_{j=1}^{K} w_j^{i-1}\omega_j^i = 0 \\ \frac{w_j^{i-1}\omega_j^i}{\sum_{j=1}^{K} w_j^{i-1}\omega_j^i} & otherwise \end{cases} \tag{6.3}$$

From Step 3b, it can be seen that the combination weights are constructed in a sequential way so that each training data processes a certain property of inheritance. The advantage of a Bayesian decision analysis is that it can model uncertainty information via the

Bayesian reasoning process (Wang et al., 2010), which can help human analysts to gain more insights into the system to be modeled.

6.3.4.2. The Non-Sequential Method

For the non-sequential method, constrained non-linear optimization (or constrained non-linear programming) is used to find the optimal combination weights. The algorithm attempts to find a constrained minimum of a scalar function of several variables starting from an initial estimate. The algorithm uses a Hessian, the second derivatives of the Lagrangian (Byrd et al., 2000). The problem can be specified by

$$\min_x f(x)\, such\ that \left[\begin{array}{c} A\,.\,x\ \leq b \\ Aeq\,.\,x\ \leq beq \end{array} \right. \tag{6.4}$$

where $A\,.\,x\ \leq b$ is set for constraint $w_i \geq 0$ and $Aeq\,.\,x\ \leq beq$ is set for constraint $\sum_{i=1}^{K} w_i = 1$. For this case, *A* = [*-1 0 0; 0 -1 0; 0 0 -1*]; *b* = [*0; 0; 0*]; *Aeq* = [*1 1 1*]; *beq* = [*1*]. The initial estimate vector is set to [*0 1 0*]$^{\mathrm{T}}$. In other words, the algorithm finds the optimal values of w_i are better than no aggregation method. The cost function *f(x)*, which has to be minimized, is as follows:

$$sse = \sum_{i=1}^{S}[(w_1 z'_{1i} +\ w_2 z'_{2i} + w_3 z'_{3i}) -\ z_i]^2 \tag{6.5}$$

Where SSE is error of training data, *S* is the number of training data, z'_i is the predicted value from PM$_i$, and z_i is the observed value.

6.4. Evaluation of the Proposed Methodology

This chapter continues from the previous one and all models are compared herein. These models include ARMA, $BPNN_{12}$, ANFIS, and MFIS–OPT_2. Also, the modular model without aggregation modules is included. Such a model derives the final output directly from the triggered PM module. Henceforth, Mod FIS refers to the modular models without an aggregation layer. Mod FIS–BSA and Mod FIS–HSA refer to the modular models with sequential and non-sequential aggregation layers, respectively.

6.4.1. Quantitative Results

To evaluate the prediction accuracy of the current models, three error measures have been used, MAE, RMSE and R. The experimental results are shown in Tables 6.1 to 6.3. In the tables, MAE and RMSE measure are normalized by the mean values of the datasets for comparison purposes. In the tables, the row Average refers to the average values from all case studies and the row Improvement refers to the improvement percentage of the average values based on the MFIS–OPT_2 model. Figure 6.3 shows these average values (use the left axis for reference) and improvement values (use the right axis for reference).

Table 6.1. Normalized mean absolute error.

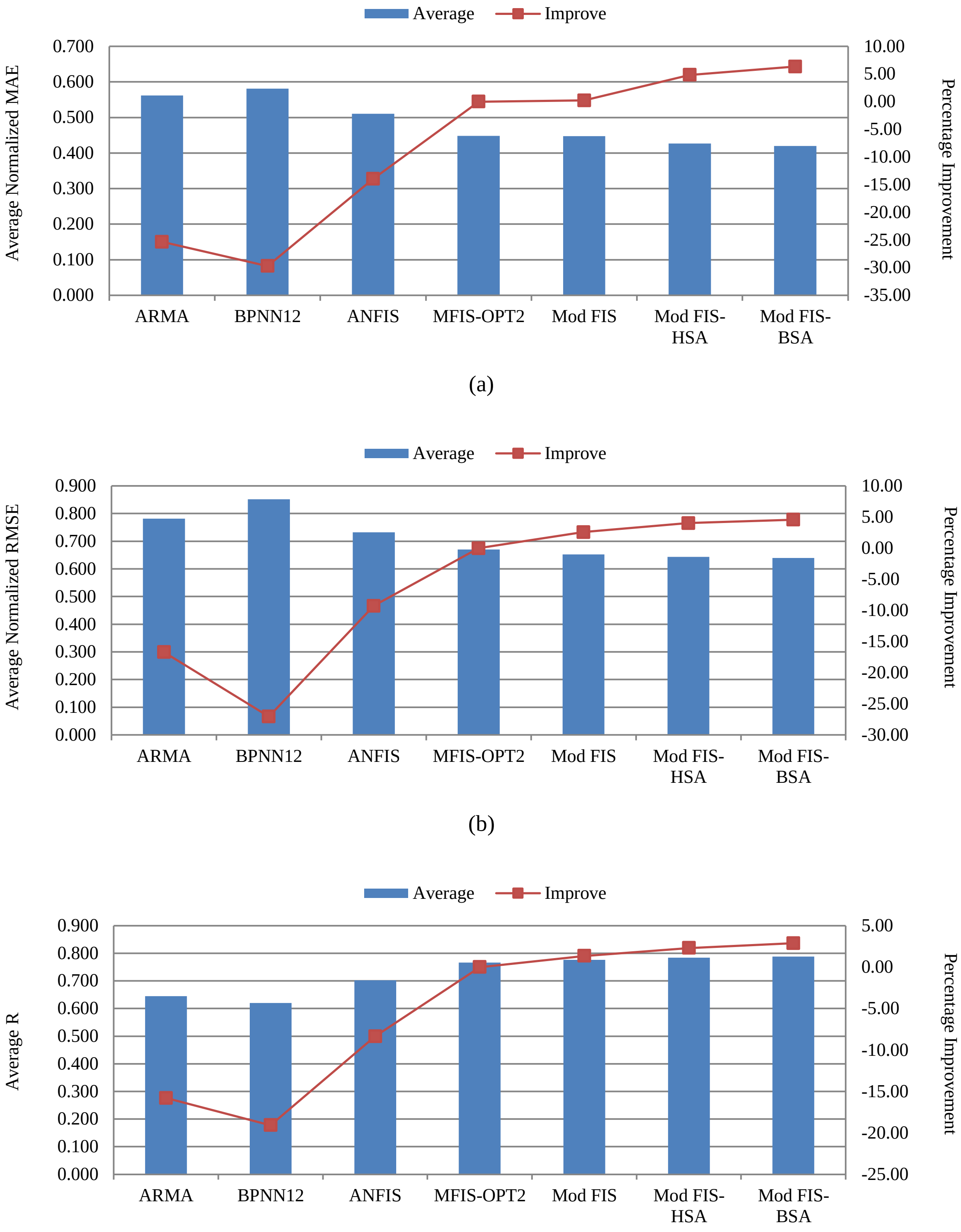
Case Study	ARMA	BPNN$_{12}$	ANFIS	MFIS–OPT$_2$	Mod FIS	Mod FIS–HSA	Mod FIS–BSA
1	0.740	0.566	0.555	0.483	0.463	0.492	0.448
2	0.480	0.484	0.387	0.381	0.384	0.350	0.331
3	0.595	0.620	0.580	0.420	0.456	0.409	0.398
4	0.549	0.616	0.522	0.472	0.474	0.468	0.461
5	0.623	0.611	0.518	0.464	0.455	0.456	0.440
6	0.570	0.660	0.595	0.515	0.495	0.464	0.465
7	0.518	0.567	0.510	0.440	0.491	0.424	0.460
8	0.419	0.526	0.420	0.410	0.360	0.349	0.354
Average	0.562	0.581	0.511	0.448	0.447	0.427	0.420
Improvement	-25.34	-29.71	-13.97	0.00	0.21	4.82	6.34

Table 6.2. Normalized root mean square error.

Case Study	ARMA	BPNN$_{12}$	ANFIS	MFIS–OPT$_2$	Mod FIS	Mod FIS–HSA	Mod FIS–BSA
1	0.956	0.773	0.769	0.729	0.662	0.688	0.658
2	0.677	0.710	0.593	0.616	0.537	0.565	0.513
3	0.867	0.974	0.845	0.672	0.727	0.688	0.699
4	0.799	0.966	0.764	0.709	0.640	0.648	0.641
5	0.880	0.874	0.771	0.708	0.694	0.695	0.673
6	0.851	1.013	0.845	0.754	0.767	0.758	0.763
7	0.651	0.743	0.667	0.575	0.665	0.588	0.648
8	0.574	0.757	0.603	0.598	0.528	0.515	0.520
Average	0.782	0.851	0.732	0.670	0.653	0.643	0.639
Improvement	-16.71	-27.06	-9.27	0.00	2.58	4.02	4.57

Table 6.3. Correlation coefficient.

Case Study	ARMA	BPNN$_{12}$	ANFIS	MFIS–OPT$_2$	Mod FIS	Mod FIS–HSA	Mod FIS–BSA
1	0.539	0.731	0.764	0.800	0.813	0.816	0.825
2	0.787	0.761	0.844	0.844	0.872	0.859	0.886
3	0.543	0.572	0.583	0.732	0.696	0.721	0.725
4	0.797	0.740	0.835	0.855	0.873	0.877	0.877
5	0.666	0.656	0.738	0.781	0.791	0.791	0.809
6	0.587	0.465	0.613	0.718	0.681	0.693	0.692
7	0.466	0.371	0.468	0.634	0.663	0.683	0.657
8	0.776	0.664	0.772	0.767	0.824	0.830	0.835
Average	0.645	0.620	0.702	0.766	0.777	0.784	0.788
Improvement	-15.80	-19.08	-8.36	0.00	1.35	2.31	2.87



(a)

(b)

(c)

Figure 6.3. Average values and improvement of models.

In terms of MAE, the best prediction accuracy comes from Mod FIS–BSA in five cases and comes from Mod FIS–HSA in three cases. Among the proposed models, the lowest accuracy appears on Mod FIS in four cases, MFIS–OPT$_2$ in three cases and Mod FIS–HSA in only one case. Based on the average values, the prediction accuracy can be ordered as Mod FIS–BSA > Mod FIS–HSA > Mod FIS ≈ MFIS–OPT$_2$ > ANFIS > ARMA > BPNN$_{12}$.

In terms of RMSE, the best prediction accuracy is in Mod FIS–BSA and MFIS–OPT$_{2,}$ three cases in each, and in Mod FIS and Mod FIS–HSA in one case each. Considering all the proposed models, the lowest accuracy appears in MFIS–OPT$_2$ in five cases and three cases in Mod FIS. Based on the average values, the prediction accuracy can be ranked as Mod FIS–BSA ≈ Mod FIS–HSA > Mod FIS > MFIS–OPT$_2$ > ANFIS > ARMA > BPNN$_{12}$. However, Mod FIS–BSA shows slightly better accuracy than Mod FIS–HSA.

In terms of R, the best prediction accuracy appears on Mod FIS–BSA in five cases, on MFIS–OPT$_2$ in two cases and on Mod FIS–HSA in one case. Among all proposed the models the lowest accuracy comes from MFIS–OPT$_2$ in six cases and from Mod FIS in two cases. Based on average values, the prediction accuracy can be ordered as Mod FIS–BSA ≈ Mod FIS–HSA > Mod FIS > MFIS–OPT$_2$ > ANFIS > ARMA > BPNN$_{12}$. For this measure, the accuracy gradually increases from MFIS–OPT$_2$ to Mod FIS–BSA.

As the results from MAE, RMSE and R measures are consolidated, these experimental results are rather consistent. Overall, the prediction accuracy can be ordered as Mod

FIS–BSA > Mod FIS–HSA > Mod FIS > MFIS–OPT$_2$ > ANFIS > ARMA > BPNN$_{12}$. It can be seen that major improvement comes from the use of a time coefficient (MFIS–OPT$_2$) and the modular technique (Mod FIS). For the modular model, the minor improvement comes from the use of an aggregation layer. In turn, the sequential method provided slightly more improvement than the non-sequential method.

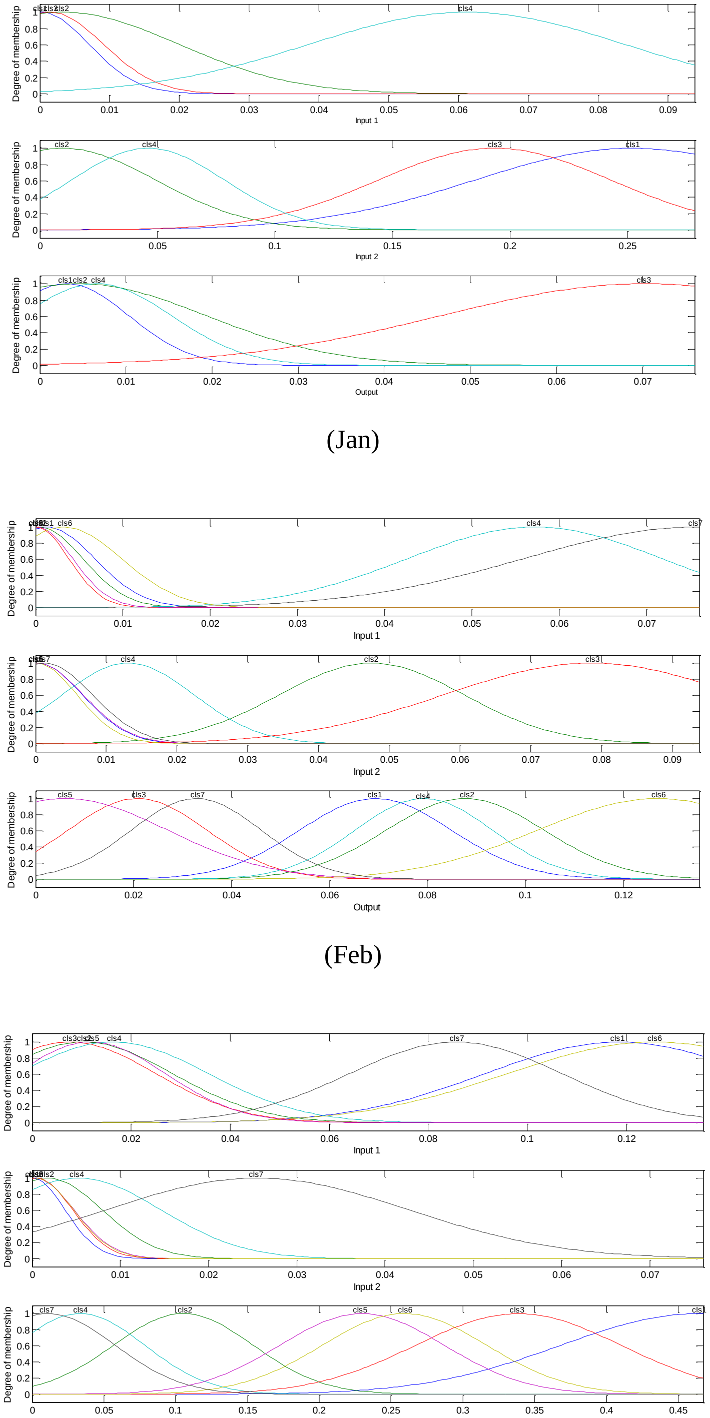
6.4.2. Qualitative Results

Table 6.4 shows the number of prototypes generated in twelve monthly sub-modules of the eight case studies. The number of prototypes is aligned with the number of fuzzy rules and the number of MFs of each input and output of the fuzzy models. The average number of prototypes of the fuzzy models from all cases is approximately 76.

Figure 6.4 shows an example of the fuzzy parameters of monthly PMs. The figure shows MFs of Case 1, which has the most number of prototypes (i.e. 94). The fuzzy rules of all PMs are in the form of "*IF 1stLag = cls$_i$ AND 2ndLag = cls$_i$ THEN Rainfall = cls$_i$*" (i =1, 2, 3, ..., n), where *n* is the number of generated prototypes.

Table 6.4. The numbers of prototypes generated in twelve sub-modules.

Case Study	Numbers of Local Modules	Numbers of Prototypes in Local Modules	Numbers of all Fuzzy Rules (MFs)
1	12	4, 7, 7, 8, 10, 10, 9, 7, 8, 9, 8, 7	94
2	12	1, 2, 6, 8, 8, 7, 8, 9, 9, 8, 5, 2	73
3	12	3, 4, 6, 8, 8, 6, 6, 6, 6, 7, 7, 5	72
4	12	1, 2, 5, 7, 8, 6, 8, 9, 9, 8, 8, 5	76
5	12	2, 1, 1, 2, 7, 7, 7, 8, 7, 7, 6, 5	60
6	12	2, 2, 5, 8, 8, 7, 8, 7, 8, 6, 9, 6	76
7	12	6, 5, 8, 7, 6, 9, 9, 8, 8, 7, 4, 6	83
8	12	2, 6, 8, 8, 7, 8, 7, 6, 5, 7, 6, 5	75

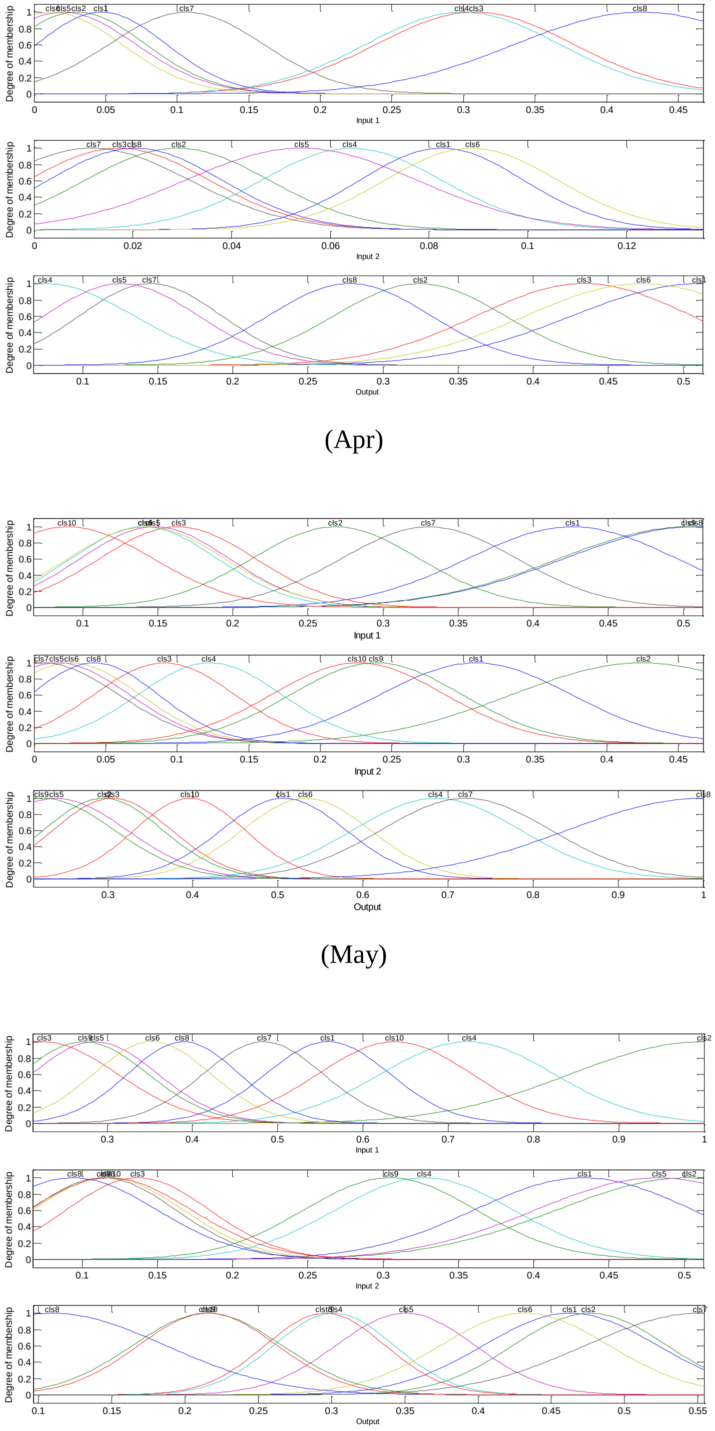


(Jan)

(Feb)

(Mar)

Figure 6.4. An example of fuzzy parameters of monthly sub-modules (Case 1).

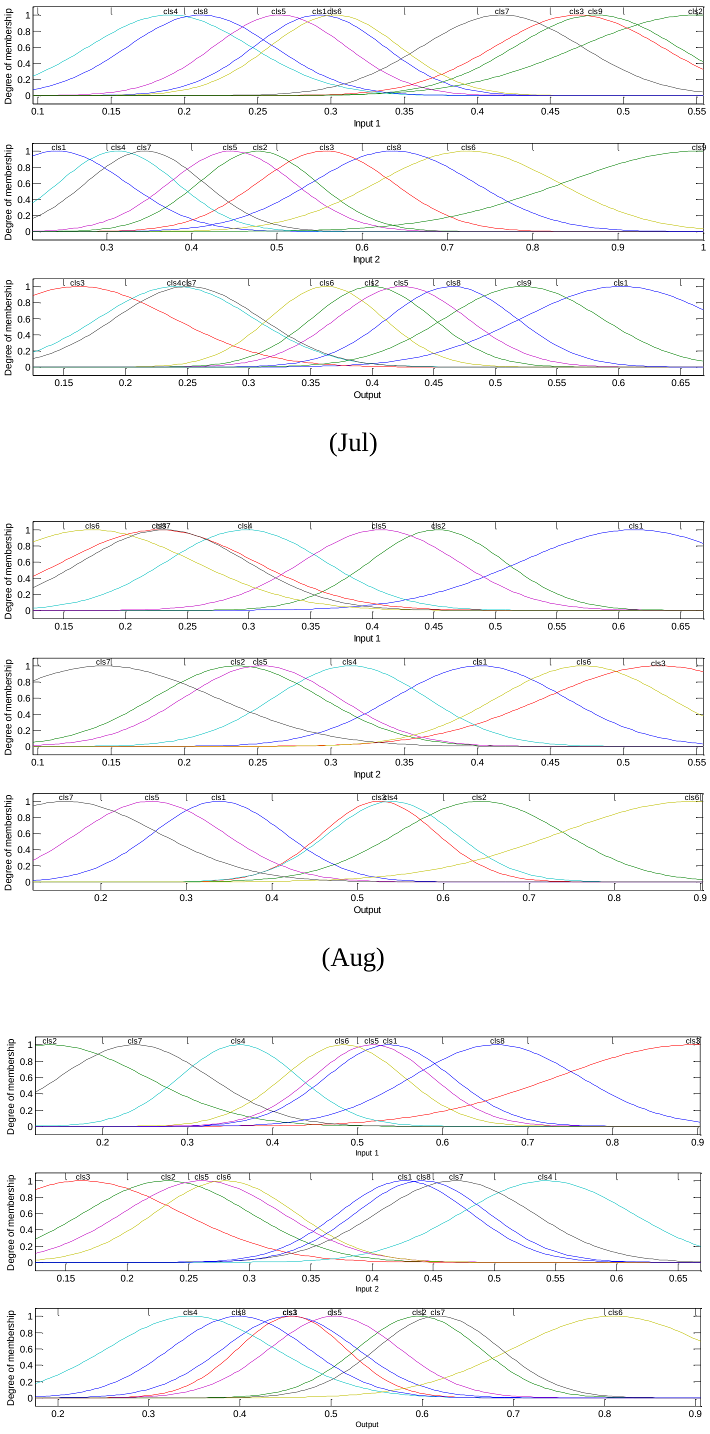


(Apr)

(May)

(Jun)

Figure 6.4. (cont.) An example of fuzzy parameters of monthly sub-modules (Case 1).

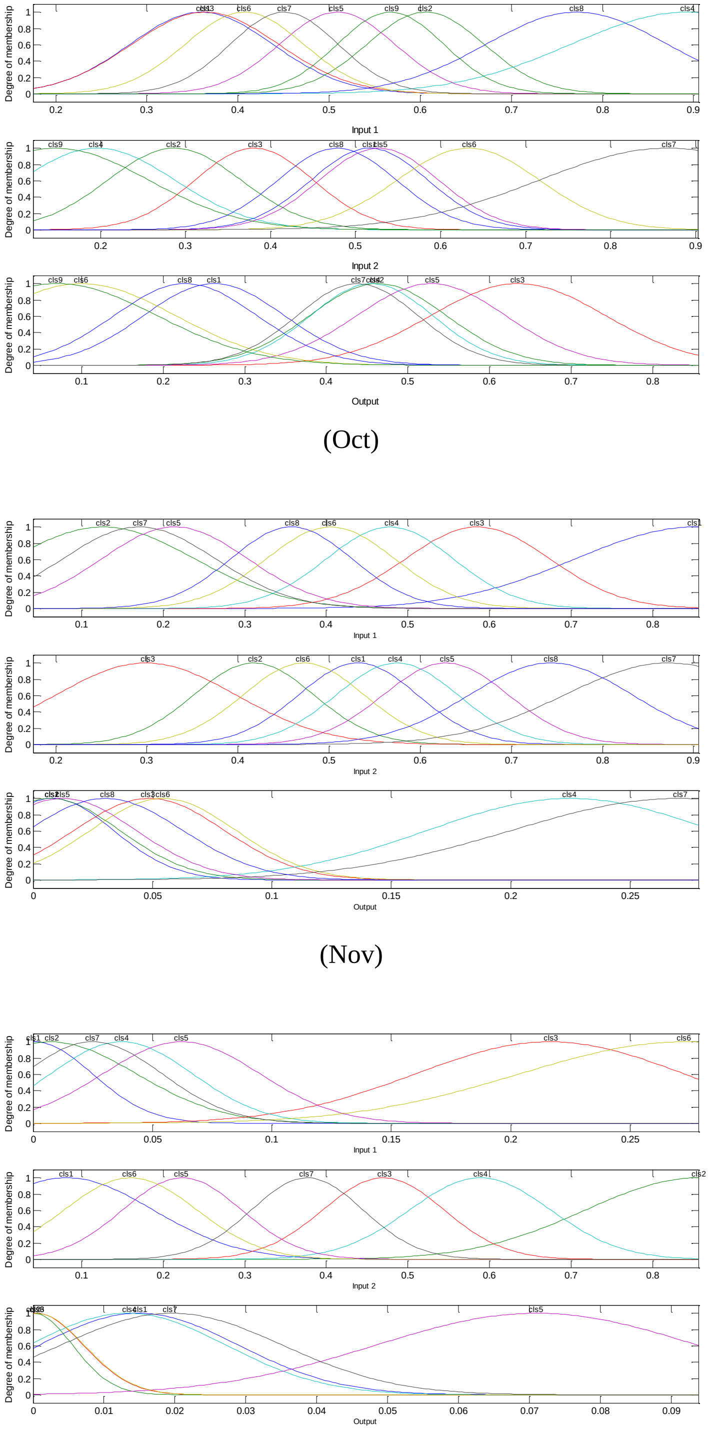


(Jul)

(Aug)

(Sep)

Figure 6.4. (cont.) An example of fuzzy parameters of monthly sub-modules (Case 1).



(Oct)

(Nov)

(Dec)

Figure 6.4. (cont.) An example of fuzzy parameters of monthly sub-modules (Case 1).

At the level of the fuzzy sets, according to Figure 6.5, the generated input's MFs satisfy the distinguishability criterion. In general, the generated MFs represent enough distinct space. However, some of the generated MFs are very close, for example, the first input of May's PM, and the second input of February's PM and June's PM. Overall, the input's MFs are distinct enough to represent linguistic terms.

As shown in Table 6.5, the numbers of MFs (prototype) in the sub-modules are close to the moderate number of MFs (7±2) criterion. Only two sub-modules in Case 1 have the number of MFs up to 10. Therefore, this criterion can be justified as accomplished. The normalization criterion is also met since at least one data point in the universe of discourse should have a membership value equal to one.

The case of the coverage or completeness of fuzzy partitioning criterion should be considered from both the entire system and the sub-modules. As the sub-modules are generated from prototype-based fuzzy modeling, this criterion should be satisfied for sub-modules. For the entire system, as the input is fed into three consecutive PMs, the possibility that the input vector does not belong to anyone of fuzzy sets is rather small. Thus, this criterion should be justified as acceptable for the entire system.

At the level of the fuzzy rules, the proposed model satisfies the rule base parsimonious and simplicity criterion. The number of fuzzy rules has considerably decreased in the proposed model from the proposed model in the previous chapter. As shown in Table 6.5, the average number of fuzzy rules for the models is about 64. The readability of single rules is satisfied since the number of conditions in the antecedent part is only two.

For the consistency and completeness criteria, as the models are generated by prototype-based fuzzy modeling, the contradictory fuzzy rules are absent and they guarantee at least one rule will be fired for any input vector. Also, the fuzzy rules of the proposed model satisfied the transparency of rule structure criterion. As the entire model is decomposed into twelve sub-modules, the two-dimensional input vector (i.e. first lag and second lag of rainfall data) is understandable by human analysts and can be represented in the three-dimensional space. Table 6.5 summarizes the qualitative results.

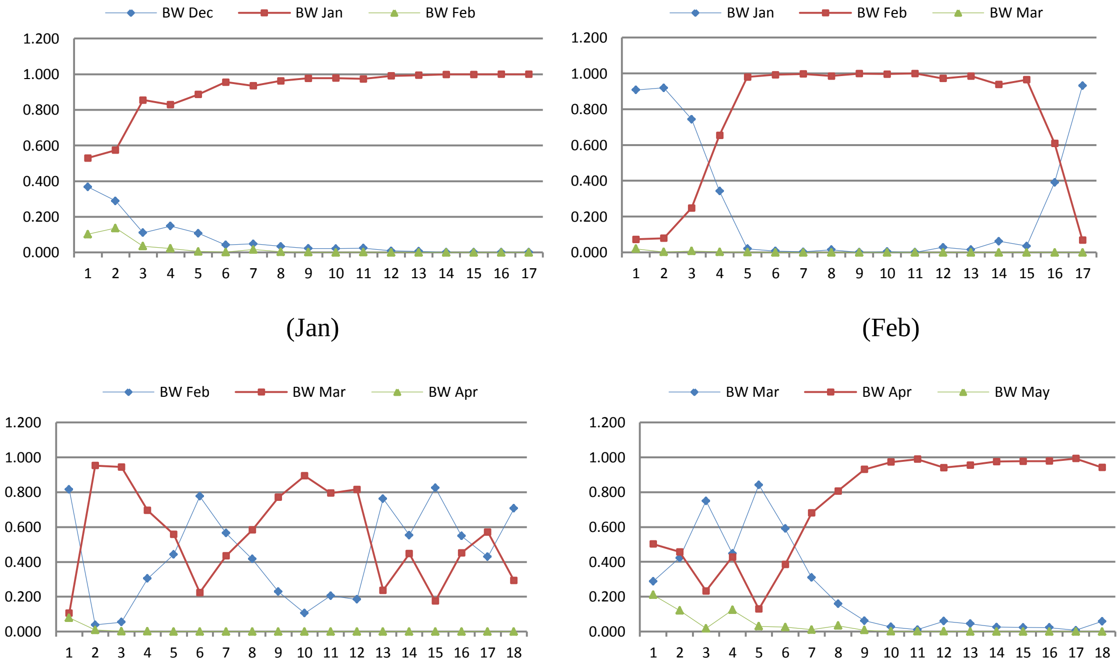
So far, the proposed modular models have been evaluated in quantitative and qualitative terms. The experimental results have suggested that the proposed modular models can improve prediction accuracy and simplify model interpretability of the proposed single model (i.e. MFIS–OPT$_2$). The substantial advantage of the Bayesian learning method is not only to improve prediction accuracy, but also to represent the uncertainty in time dimension inherited across the historical time.

This method proposes a new approach to analyze the uncertainty of time. Alternating to static representation of the uncertainty of time in MFIS–OPT$_2$, the Mod FIS–BSA proposes the dynamic representation of the uncertainty in time. Figure 6.5 shows an example of a combination of weights of the aggregation modules that are inherited across the time dimension.

Up to this point, the single and modular interpretable fuzzy models have been proposed. Each model has its own advantages and disadvantages. However, the question is which model is appropriate to the problem on hand. Table 6.6 summarizes some general issues that can be used as a guideline for the selection.

Table 6.5. Summary of qualitative results.

Level	Criterion	Weak	Fair	Good	Strong
Low-Level	Distinguishability		x		
	Moderate number of MFs			x	
	Coverage or completeness of partition of input variable			x	
	Normalization				x
High-Level	Rule-base parsimony and simplicity			x	
	Readability of single rule				x
	Consistency of rules				x
	Completeness of rules			x	
	Transparency of rule structure				x



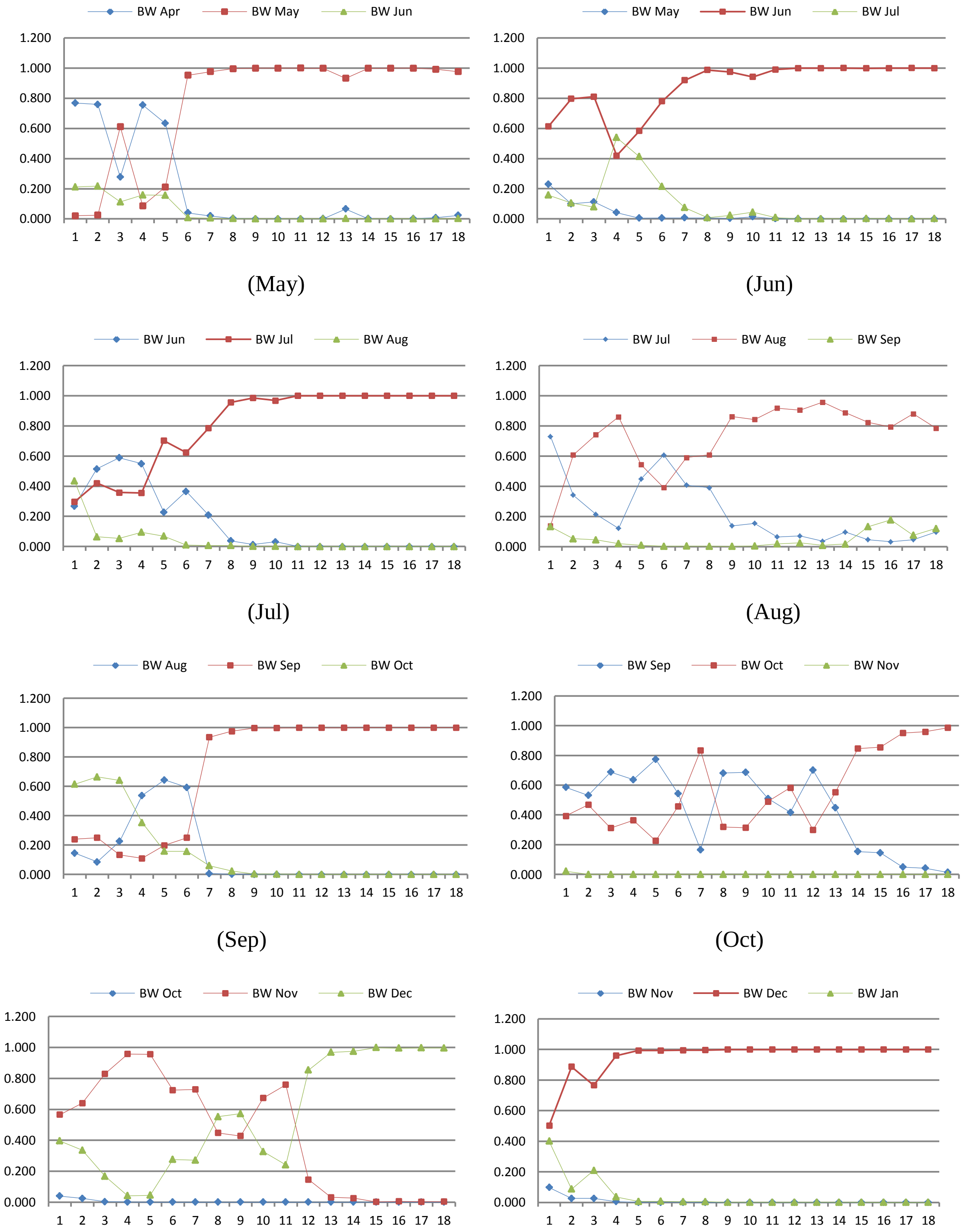
(Jan)

(Feb)

(Mar)

(Apr)

Figure 6.5. An example of the combination weights from the Bayesian method inherited across the time dimension (Case 1).



(May)

(Jun)

(Jul)

(Aug)

(Sep)

(Oct)

(Nov)

(Dec)

Figure 6.5. (cont.) An example of the combination weights from the Bayesian method inheriting across the time dimension (Case 1).

Table 6.6. Summary of observed characteristics of the proposed models for time series prediction.

No	characteristics	Single	Modular
1	The model that provides the direct process in the model establishment but the prediction accuracy may not be relatively high.	Preferred	
2	The model that provides relatively higher prediction accuracy but the model establishment is not complete in one structuring process.		Preferred
3	The model that provides the solidity in which the entire model consists of one clear structure and enables other techniques in each establishment step.	Selected	
4	The model that provides the flexibility in which an entire model consists of several structures and each of them can be modeled by different techniques.		Selected
5	The model that provides the static view of the uncertainty in time dimension for further analysis. (i.e. represented in the form of fuzzy parameters)	Selected	
6	The model that provide the dynamic view of the uncertainty in time dimension for further analysis. (i.e. represented in the form of Bayesian reasoning)		Selected
7	The model that provides the global understanding of monthly rainfall time series data. Human analysts may not need to make a detailed analysis of the data.	Selected	
8	The model that provides the local understanding of monthly rainfall time series data. Human analysts may be required to make a detailed analysis of the data.		Selected

6.5. Conclusion

This chapter proposes the use of a modular technique to create an interpretable fuzzy model for monthly rainfall time series prediction. This chapter continues from the previous chapter by proposing an alternative model which simplifies the complexity of the previous model. The proposed modular model consists of two main layers, that is, the prediction layer and the aggregation layer. A Mamdani-type FIS is used to predict monthly rainfall in the prediction layer, whereas the Bayesian reasoning method (or nonlinear programming method) is used to aggregate predicted results in the aggregation layer.

The proposed models are evaluated by eight monthly rainfall time series data in the northeast region of Thailand and are also compared with the previously proposed models. The experimental results pointed out that the use of the modular technique can improve the prediction from the single model's quantitative and qualitative terms. The prediction accuracy had been increased and the model interpretability had been simplified.

Consequently, the use of modular models allows the model to provide independent details of each part of the model to human analysts. Moreover, aggregation modules that employ the Bayesian reasoning method can be used to analyze the uncertainty in time dimension in a sequential fashion. Unlike the static representation of the single model, the modular models show dynamic representation instead. This sort of representation opens an alternative approach to analyze monthly time series data.

Finally, it would be more suitable if the model proposed in this chapter has been seen as an alternative to the model in the previous chapter. Both models provided comparably good prediction accuracy. But the uncertainty in time dimension is represented in different views. Each proposed model has its own prominence. This thesis has finalized the time series prediction technique by presenting some general idiosyncrasies of both single and modular models. They can be used as a guideline for selecting the appropriate models to be matched to user requirements.

CHAPTER 7
CONCLUSIONS

7.1. Introduction

This thesis proposed the methodologies to analyze and establish interpretable fuzzy systems for monthly rainfall spatial interpolation and time series prediction. The case study area in the experiments is located in the northeast region of Thailand. In each part of the thesis, eight case studies were used to evaluate the established models. The experimental results were presented by quantitative and qualitative points of view. A summary of each part follows.

7.2. Research Summary of Spatial Interpolation

- In the global method, the thesis proposes two FCM clustering validation indices to determine the number of clusters specific to the spatial data. The first method employs the statistical characteristics of spatial data by analyzing the standard deviation of clustered data. The second method employs the artificial neural network by analyzing the training performance of clustered data.

- The thesis then proposes the use of the prototype-based fuzzy modeling in cooperation with GA to create an interpretable fuzzy spatial interpolation model. The established model named GAFIS is the outcome of the methodology. The experimental results suggested that GAFIS provided good interpolation accuracy and model interpretability for the global method.

- In the local method, the thesis proposes a methodology to improve the accuracy of GAFIS while the model interpretability is maintained. At this point, a decision tree-like modular technique is applied. In turn, the thesis also proposes a fuzzy gating method to improve the accuracy of the generic gating method and also to maintain the interpretability in the gating module. The established models named Mod FIS–FSG/FST are the outcomes of the proposed methodology.

- The experimental results suggested that the established modular models can improve the interpolation accuracy from GAFIS and can be a good alternative to those of local spatial interpolation methods used in GIS. In terms of interpretability, due to the decision tree-like architecture, the established model can be interpreted effectively in a modular way.

- For future works, in the author's opinion, as the proposed methodology was developed for spatial data in general, one challenging task is to apply the proposed methodology to other spatial data which is not limited to the hydrological and environmental discipline. Another interesting point is to apply other intelligent techniques to the proposed methodology, for example, the memetic algorithm.

7.3. Research Summary of Time Series Prediction

- In the single model, this thesis utilizes a cooperation of FL, ANN and GA to create an interpretable fuzzy system for monthly rainfall time series prediction. The experimental results showed that the proposed single model provided satisfactory prediction accuracy in comparison with commonly-used models in the hydrological discipline. Thus, the proposed model can satisfy the quantitative requirement. In terms of the qualitative aspect, the proposed model provided good model interpretability. However, a large number of generated fuzzy rules are still unsatisfied.

- An advantage of the proposed single model is to provide an approach to analyze the uncertainty in time dimension of monthly time series data. Due to the advantage of the fuzzy system, this model allows human analysts to analyze the uncertainty in time dimension (between months) through the fuzzy MFs. By this approach, human analysts can gain insight into the data to be modeled.

- In the modular model, this thesis proposes the use of a modular technique to simplify the complexity in the single models. The experimental results showed that the modular model improved the prediction accuracy and also increased the interpretability of the single model. The number of fuzzy rules and fuzzy MFs considerably decreased. Furthermore, the modular model is more flexible than the single model in that users can use other techniques with the sub-modules independently.

- An advantage of the proposed model is to provide another approach to analyze the uncertainty in time dimension of monthly time series data. Due to the advantage of the Bayesian reasoning, this model allows human analysts to analyze the uncertainty in time dimension in a sequential fashion. By this approach, human analysts can understand how the uncertainty in time dimension varied along the calibration period.

- For future works, in the author’s opinion, the proposed models are developed for any monthly time series data. These models are applicable to other monthly time series data, actually any periodic time series data. One interesting work is to develop an algorithm to simplify the complexity of the established modular model by using the pruning method.

REFERENCES

Abonyi, J., Babuska, R., Verbruggen, H.B., & Szeifert, F. (2000). Incorporating prior knowledge in fuzzy model identification. *International Journal of Systems Science*, 31(5), 657–667.

Afshin, S., Fahmi, H., Alizadeh, A., Sedghi, H., & Kaveh, F. (2011). Long term rainfall forecasting by integrated artificial neural network-fuzzy logic-wavelet model in Karoon basin. *Scientific Research and Essays*, 6(6), 1200–1208.

Alcalá, R., Alcalá-Fdez, J., Casillas, J., Cordón, O., & Herrera, F. (2006). Hybrid learning models to get the interpretability–accuracy trade-off in fuzzy modeling. *Soft Computing*, 10(9), 717–734.

Alonso, J.M., & Magdalena, L. (2011). Special issue on interpretable fuzzy systems. *Information Sciences*, 181(20), 4331–4339.

Alonso, J.M., Magdalena, L., & González-Rodríguez, G. (2009). Looking for a good fuzzy system interpretability index: An experimental approach. *International Journal of Approximate Reasoning*, 51(1), 115–134.

Alvisi, S., & Franchini, M. (2011). Fuzzy neural networks for water level and discharge forecasting with uncertainty. *Environmental Modelling & Software*, 26(4), 523–537.

Araghinejad, S., Azmi, M., & Kholghi, M. (2011). Application of artificial neural network ensembles in probabilistic hydrological forecasting. *Journal of Hydrology*, 407(1–4), 94–104.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a). Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering*, 5(2), 115–123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b). Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering*, 5(2), 124–137.

Asklany, S.A., Elhelow, K., Youssef, I.K., & El-wahab, M.A. (2011). Rainfall events prediction using rule-based fuzzy inference system. *Atmospheric Research*, 101, 228–236.

Bacanli, U.G., Firat, M., & Dikbas, F. (2009). Adaptive neuro-fuzzy inference system for drought forecasting. *Stochastic Environmental Research and Risk Assessment*, 23(8), 1143–1154.

Bargaoui, Z.K., & Chebbi, A. (2009). Comparison of two kriging interpolation methods applied to spatiotemporal rainfall. *Journal of Hydrology*, 365(1–2), 56–73.

Beale, M.H., Hagan, M.T., & Demuth, H.B. (2011). Neural Network Toolbox™ User's Guide MATLAB. Retrieved from www.mathworks.com/help/pdf_doc/nnet/nnet _ug.pdf.

Bezdek, J.C. (1981) *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press.

Bezdek, J.C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy *c*-means clustering algorithm. *Computers & geosciences*, 10(2–3), 16–20.

Box, G.E.P., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.

Brown, M., & Harris, C. (1994). *Neuro fuzzy adaptive modelling and control*. New York: Prentice-Hall.

Burrough, P., & McDonnell, R. (1998). *Principles of geographical information systems*. New York: Oxford University Press.

Byrd, R.H., Gilbert, J.C., & Nocedal, J. (2000). A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89(1), 149–185.

Casillas, J., Cordón, O., Herrera, F., & Magdalena, L. (Eds.) (2003a). *Accuracy improvements in linguistic fuzzy modeling*. Springer.

Casillas, J., Cordón, O., Herrera, F., & Magdalena, L. (Eds.) (2003b). *Interpretability issues in fuzzy modeling*. Springer.

Cellura, M., Cirrincione, G., Marvuglia, A., & Miraoui, A. (2008). Wind speed spatial estimation for energy planning in Sicily: A neural kriging application. *Renewable Energy*, 33(6), 1251–1266.

Chang, C.L., Lo, S.L., & Yu, S.L. (2005). Applying fuzzy theory and genetic algorithm to interpolate precipitation. *Journal of Hydrology*, 314(1–4), 92–104.

Chang, K. (2006). *Introduction to geographic information systems* (3rd ed.). Singapore: McGraw-Hill.

Chris-Tseng, H., & Almogahed, B. (2009). Modular neural networks with applications to pattern profiling problems. *Neurocomputing*, 72, 2093–2100.

Collins, F.C., & Bolstad, P.V. (1996, January). *A comparison of spatial interpolation techniques in temperature estimation.* In Proceedings of the Third International Conference/Workshop on Integrating GIS and Environmental Modeling CD-ROM.

Córdon, O. (2011). A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems. *International Journal of Approximate Reasoning*, 52(6), 894–913.

Cordón, O., Herrera, F., & Villar, P. (2001). Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. *IEEE Transactions on Fuzzy Systems*, 9(4), 667–674.

Demyanov, V., Kanevsky, M., Chernov, S., Savelieva, E., & Timonin, V. (1998). Neural network residual kriging application for climatic data. *Journal of Geographic Information and Decision Analysis*, 2(2), 215–232.

Dubois, D., Prade, H., & Ughetto, L. (1997). Checking the coherence and redundancy of fuzzy knowledge bases. *IEEE Transaction on Fuzzy Systems*, 5(3), 398–417.

Dunn, J.C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3), 32–57.

Erilli, N.A., Yolcu, U., Eğrioğlu, E., Aladağ, C.H., & Öner, Y. (2011). Determining the most proper number of cluster in fuzzy clustering by using artificial neural networks. *Expert Systems with Applications*, 38(3), 2248–2252.

ESRI (n.d.). *ArcGIS*. Retrieved from http://www.esri.com/software/arcgis

Firat, M., & Gungor, M. (2008). Hydrological time-series modeling using an adaptive neuro-fuzzy inference system. *Hydrological Processes*, 22(13), 2122–2132.

Firat, M., & Turan, M.E. (2009). Monthly river flow forecasting by an adaptive neuro-fuzzy inference system. *Water and Environment Journal*, 24(2), 116–125.

Firat, M., Turan, M.E., & Yurdusev, M.A. (2009). Comparative analysis of fuzzy inference systems for water consumption time series prediction. *Journal of Hydrology*, 374(3–4), 235–241.

Gilardi, N., & Bengio, S. (2000). Local machine learning models for spatial data analysis. *Journal of Geographic Information and Decision Analysis*, 4(1), 11–28.

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, 228(1–2), 113–129.

Guillaume, S. (2001). Designing fuzzy inference systems from data: An interpretability-oriented review. *IEEE Transactions on Fuzzy Systems*, 9(3), 426–443.

Guo, J., Zhou, J., Qin, H., Zou. Q., & Li, Q. (2011). Monthly streamflow forecasting based on improved support vector machine model. *Expert Systems with Applications*, 38(10), 13073–13081.

Haberlandt, U. (2007). Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *Journal of Hydrology*, 332(1–2), 144–157.

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, 17(2), 107–145.

Hameed, I.A. (2011). Using Gaussian membership functions for improving the reliability and robustness of students' evaluation systems. *Expert Systems with Applications*, 38(6), 7135–7142.

Hancock, P.A., & Hutchinson, M.F. (2006). Spatial interpolation of large climate data sets using bivariate thin plate smoothing splines. *Environmental Modeling & Software*, 21, 1684–1694.

Harris, C.J., Hong, X., & Gan, Q. (2002). *Adaptive modelling, estimation and fusion from data.* Berlin: Springer.

Hartkamp, A.D., Beurs, K.D., Stein, A., & White, J.W. (1999). Interpolation techniques for climate variables. *National Resource Group Geographic Information Systems Series 99-01*, Mexico.

Holland, J.H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.

Hong, Y., Nix, H.A., Hutchinson, M.F., & Booth, T.H. (2005). Spatial interpolation of monthly mean climate data for China. *International Journal of Climatology*, 25(10), 1369–1379.

Hu, C., Meng, L., & Shi, W. (2008). Fuzzy clustering validity for spatial data. *Geo-Spatial Information Science*, 11(3), 191–196.

Hu, G., & Zhang, Q. (2008). *Application of adaptive variable structure of ANN to distributed rainfall interpolation*. In Proceedings of IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application, pp. 598–601.

Huang, Y., Wong, P., & Gedeon, T. (1998). Spatial interpolation using fuzzy reasoning and genetic algorithms. *Journal of Geographic Information and Decision Analysis*, 2(2), 204–214.

Huarng, K. (2001). Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy Sets and Systems*, 123(3), 387–394.

Isaaks, E.H., & Srivastava, R.M. (1989). *An introduction to applied geostatistics*. New York: Oxford University Press.

Ishibuchi, H. (2007, July). *Multiobjective genetic fuzzy systems: Review and future research directions.* In Proceedings of the IEEE International Conference on Fuzzy Systems, pp. 1–6.

Ishibuchi, H., Nozaki, K., Yamamoto, N., & Tanaka, H. (1994). Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms. *Fuzzy Sets and Systems*, 65(2–3), 237–253.

Jain, A., & Kumar, A.M. (2007). Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing*, 7(2), 585–592.

Jang, J.S.R., Sun, C.T., & Mizutani, E. (1997). *Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence*. Upper Saddle River, NJ: Prentice Hall.

Jeffrey, S.J., Carter, J.O., Moodie, K.B., & Beswick, A.R. (2001). Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling & Software*, 16(4), 309–330.

Kajornrit, J., Wong, K.W., & Fung, C.C. (2011). Estimation of missing rainfall data in northeast region of Thailand using spatial interpolation methods. *Australian Journal of Intelligent Information Processing Systems*, 13(1), 21–30.

Karray, F.O., & Silva, C.W. (2004). *Soft computing and intelligent systems design*. United Kingdom: Addison Wesley.

Zounemat-Kermani, M., & Teshnehlab, M. (2008). Using adaptive neuro-fuzzy inference system for hydrological time series prediction. *Applied Soft Computing*, 8(2), 928–936.

Keskin, M.E., Taylan, D., & Terzi, O. (2006). Adaptive neural-based fuzzy inference system (ANFIS) approach for modeling hydrological time series. *Hydrological Sciences Journal*, 51(4), 588–598.

Kim, J., & Pachepsky, Y.A. (2010). Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. *Journal of Hydrology*, 394(3–4), 305–314.

Krige, D.G. (1951) A statistical approach to some mine valuations problems at the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52, 119–138.

Land Development Department of Thailand (n.d.) *Rainfall data*. Retrieved [2013] from http://irw101.ldd.go.th/lib/images/dr_57th.pdf

Lee, S., Cho, S., & Wong, P.M. (1998). Rainfall prediction using artificial neural networks. *Journal of Geographic Information and Decision Analysis*, 2(2), 233–242.

Li, J., & Heap, A.D. (2008). A review of spatial interpolation methods for environmental scientists. *Geoscience Australia Record*, vol. 23.

Li, J., Heap, A.D., Potter, A., & Daniell, J.J. (2011). Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software*, 26(12), 1647–1659.

Lin, G., & Chen, L. (2004). A spatial interpolation method based on radial basis function networks incorporating a semivariogram model. *Journal of Hydrology*, 288, 288–298.

Liu, S., Zhang, Y., Ma, P., Lu, B., & Su, H. (2011). A novel spatial interpolation method based on the integrated RBF neural network. *Procedia Environmental Sciences*, 10(Part A), 568–575.

Lohani, A.K., Goel, N.K., & Bhatia, K.K.S. (2010). Comparative study of neural network, fuzzy logic and linear transfer function techniques in daily rainfall-runoff modeling under different input domains. *Hydrological Processes*, 25(2), 175–193.

Lu, K., & Wang, L. (2011, April). *A novel nonlinear combination model based on support vector machine for rainfall prediction*. In Proceedings of Fourth International Joint Conference on Computational Sciences and Optimization, Yunnan, China, pp. 1343–1346.

Luo, W., Taylor, M.C., & Parker, S.R. (2008). A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. *International Journal of Climatology*, 28(7), 947–959.

Luo, X., Xu, Y., & Shi, Y. (2011, June). *Comparision of interpolation methods for spatial precipitation under diverse orographic effects*. In Proceedings of 19th International Conference on Geoinformatics, Shanghai, China, pp. 1–5.

Luo, X., Xu, Y., & Xu, J. (2010, June). *Application of radial basis function network for spatial precipitation interpolation.* In Proceedings of 18th International Conference on Geoinformatics, Beijing, China, pp. 1–5.

Mamdani, E.H., & Assilian, S. (1975). An experiment in linguistic synthesis with fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1), 1–13.

Marques, C.A.F., Ferreira, J.A., Rocha, A., Castanheira, J.M., Melo-Goncalves, P., Vaz, N., & Dias, J.M. (2006). Singular spectrum analysis and forecasting of hydrological time series. *Physics and Chemistry of the Earth*, 31(18), 1172–1179.

Matheron, G. (1965). *Les variables régionalisées et leur estimation*. Paris: Masson.

Mikut, R., Jäkel, J., & Gröll, L. (2005). Interpretability issues in data-based learning of fuzzy systems. *Fuzzy Sets and Systems*, 150, 179–197.

Monira, S.S., Faisal, Z.M., & Hirose, H. (2011, September). *An adaptive ensemble method for quantitative rainfall forecast*. In Proceedings of SICE Annual Conference, Tokyo, pp. 149–154.

Montgomery, D.C., Jennings, C.L., & Kulahci, M. (2008). *Introduction to time series analysis and forecasting.* New Jersey: John Wiley & Sons.

Nalder, I.A., & Wein R.W. (1998). Spatial interpolation of climatic normals: Test of a new method in the Canadian boreal forest. *Agricultural and Forest Meteorology*, 92(4), 211–225.

Nayak, P.C., & Sudheer K.P. (2008). Fuzzy model identification based on cluster estimation for reservoir inflow forecasting. *Hydrological Processes*, 22(6), 827–841.

Nayak, P.C., Sudheer, K.P., Rangan, D.M., & Ramasastri, K.S. (2004). A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, 291(1–2), 52–66.

Negnevitsky, M. (2011). *Artificial intelligence: A guide to intelligent systems* (3rd ed.). United Kingdom: Addison Wesley.

Oliveira, J.V. de (1999). Semantic constraints for membership function optimization. *IEEE Transactions on Systems, Man and Cybernetics–Part A: Systems and Humans*, 29(1), 128–138.

Pedrycz, W., Bezdek, J.C., Hathaway, R.J., & Rogers, G.W. (1998). Two nonparametric models for fusing heterogeneous fuzzy data. *IEEE Transactions on Fuzzy Systems*, 6(3), 411–425.

Pena-Reyes, C.A., & Sipper, M. (2003). Fuzzy CoCo: Balancing accuracy and interpretability of fuzzy models by means of coevolution. In J. Casillas, O. Cordón, F. Herrera, & L. Magdalena (Eds.), *Accuracy improvements in linguistic fuzzy modeling studies in fuzziness and soft computing,* (vol. 129). Berlin: Springer.

Piazza, A.D., Conti, F.L., Noto, L.V., Viola, F., & Loggia, G.L. (2011). Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. *International Journal of Applied Earth Observation and Geoinformation*, 13(3), 396–408.

Raman, H., & Sunilkumar, N. (1995). Multivariate modeling of water resources time series using artificial neural networks. *Hydrological Sciences Journal.* 40(2), 145–163.

Remote Sensing & GIS (n.d.). GIS data of Thailand. Retrieved [2009] from http://www.rsgis.ait.ac.th/~souris/thailand.htm

Rezaee, M.R., Lelieveldt, B.P.F., & Reiber, J.H.C. (1998). A new cluster validity index for the fuzzy *c*-means. *Pattern Recognition Letters*, 19(3–4), 237–246.

Robinson, T.P., & Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture*, 50(2), 97–108.

Ross, T.J. (2004). *Fuzzy logic with engineering applications* (2nd ed.). United Kingdom: John Wiley & Sons.

See, L.M., Abrahart, R., & Kneale, P.E. (2004). *Neural networks for hydrological modeling.* London, UK: Taylor & Francis Group.

Sen, Z., & Sahin, A.D. (2001). Spatial interpolation and estimation of solar irradiation by cumulative semivariogram. *Solar Energy*, 71(1), 11–21.

Setnes, M., Babuska, R., & Verbruggen, H.B. (1998). Rule-based modeling: Precision and transparency. *IEEE Transactions on Systems, Man, and Cybernetics –Part C: Applications and Reviews*, 28(1), 165–169.

Setnes, M., Babuska, R., Kaymak, U., & van Nauta Lemke, H.R. (1998). Similarity measures in fuzzy rule base simplification. *IEEE Transactions on Systems, Man, and Cybernetics. Part B: Cybernetics*, 28(3), 376–386.

Sharma, V., & Irmak, S. (2012). Mapping spatially interpolated precipitation, reference evapotranspiration, actual crop evapotranspiration, and net irrigation requirements in Nebraska: Part II. Actual crop evapotranspiration and net irrigation requirements. *American Society of Agricultural and Transactions Biological Engineers,* 55(3), 923–936.

Singh, A., & Imtiyaz, M. (2013). *Hydrological modelling using process based and data driven models: Process-based and neural network modelling in hydrology.* Scholars' Press.

Somvanshi, V.K., Pandey, O.P., Agrawal, P.K., Kalanker, N.V., Prakash, M.R., & Chand, R. (2006). Modelling and prediction of rainfall using artificial neural network and ARIMA techniques. *Journal of Indian Geophysical Union*, 10(2), 141–151.

Sudheer, K.P., Gosain, A.K., & Ramasastri, K.S. (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrological Processes*, 16(6), 1325–1330.

Sugeno, M., & Yasukawa, T. (1993). A fuzzy-logic based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems*, 1(1), 7–31.

Sun, Y., Kang, S., Li, F., & Zhang, L. (2009). Comparison of interpolation methods for depth to groundwater and its temporal and spatial variations in the Minqin oasis of northwest China. *Environmental Modelling & Software*, 24(10), 1163–1170.

Tang, L., Su, X., Shao, G., Zhang, H., & Zhao, J. (2012). A clustering-assisted regression (CAR) approach for developing spatial climate data sets in China. *Environmental Modelling & Software*, 38, 122–128.

Tokar, A., & Johnson, P. (1999). Rainfall-runoff modeling using artificial neural networks. *Journal of Hydrologic Engineering*, 4(3), 232–239.

Toprak, Z.F., Eris, E., Agiralioglu, N., Cigizoglu, H.K., Yilmaz, L., Aksoy, H., Coskun, H.G., Andic, G., & Alganci, U. (2009). Modeling monthly mean flow in a poorly gauged basin by fuzzy logic. *Clean–Soil, Air, Water*, 37(7), 555–567.

Tutmez, B., & Hatipoglu, Z. (2010). Comparing two data driven interpolation methods for modeling nitrate distribution in aquifer. *Ecological Informatics*, 5(4), 311–315.

Tutmez, B., Tercan, A.E., & Kaymak, U. (2007). Fuzzy modeling for reserve estimation based on spatial variability. *Mathematical Geology*, 39(1), 87–111.

Wahba, G., & Wendelberger, J. (1980). Some new mathematical methods for variational objective analysis using splines and cross validation. *Monthly Weather Review*, 108, 1122–1143.

Wang, P., Xu, L., Zhou, S.M., Fan, Z., Li, Y., & Feng, S. (2010). A novel Bayesian learning method for information aggregation in modular neural networks. *Expert Systems with Applications*, 37(2), 1071–1074.

Wang, W.C., Chau, K.W., Cheng, C.T., & Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*, 374(3–4), 294–306.

Wong, K.W., Gedeon, T.D., & Wong, P.M. (2001, July). *Spatial interpolation using conservative fuzzy reasoning*. In Proceedings of Joint Ninth IFSA World Congress and Twentieth NAFIPS International Conference, Vancouver, vol. 5, pp. 2825–2829.

Wong, K.W., Wong, P.M., Gedeon, T.D., & Fung, C.C. (2003). Rainfall prediction model using soft computing technique. *Soft Computing*, 7(6), 434–438.

Wu, C.L., & Chau, K.W. (2010). Data-driven models for monthly streamflow time series prediction. *Engineering Applications of Artificial Intelligence*, 23(8), 1350–1367.

Wu, C.L., & Chau, K.W. (2013). Prediction of rainfall time series using modular soft computing methods. *Engineering Applications of Artificial Intelligence*, 26, 997–1007.

Wu, C.L., Chau, K.W., & Fan, C. (2010). Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *Journal of Hydrology*, 389(1–2), 146–167.

Xie, X.L., & Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8), 841–847.

Yu, D. (2009, August). *Spatial interpolation via GWR, a plausible alternative?* In Proceedings of Seventeenth IEEE International Conference on Geoinformatics, Fairfax, Virginia, pp. 1–5.

Yue, W., Xu, J., Liao, H., & Xu, L. (2003). Applications of spatial interpolation for climate variables based on geostatistics: A case study in Gansu province, China. *Geographic Information Science*, 9(1–2), 71–77.

Zadeh, L.A. (1965). Fuzzy Sets. *Information and Control*, 8, 338–353.

Zhang, Q., & Wang, C. (2008, October). *Integrated application of artificial neural network and genetic algorithm to the spatial interpolation of rainfall*. In Proceedings of Fourth International Conference on Natural Computation, Jinan, China, pp. 516–520.

Zhou, S., & Gan, J.Q. (2008). Low level interpretability and high level interpretability: A unified view of data-driven interpretable fuzzy system modelling. *Fuzzy Sets and Systems*, 159, 3091–3131.

Zimmerman, D., Pavlik, C., Ruggles, A., & Armstrong, M.P. (1999). An experimental comparison of ordinary kriging and universal kriging and inverse distance weighting. *Mathematical Geology*, 31(4), 375–399.

APPENDIX A

Results from commonly-used FCM validation indices.

			Case 1						Case 2		
n	**SC**	**S**	**XB**	**DI**	**ADI**	**n**	**SC**	**S**	**XB**	**DI**	**ADI**
2	3.74119	0.01889	2.40020	0.04630	0.00488	2	3.81831	0.01958	4.46798	0.05741	0.01094
3	2.10647	0.01195	8.01935	0.04723	0.00509	3	2.19085	0.01210	2.50868	0.04763	0.02252
4	1.67523	0.01129	2.49567	0.04254	0.00441	4	1.55153	**0.00912**	2.59978	0.04078	0.01181
5	**1.54040**	**0.00941**	1.97145	0.02546	0.00158	5	1.56487	0.01383	2.15173	0.04143	**0.00025**
6	1.56945	0.01327	2.08688	0.04925	0.00062	6	1.39405	0.01170	1.94041	0.04143	0.00144
7	1.46909	0.01082	**1.37428**	0.02754	0.00125	7	**1.21864**	0.01004	2.07136	0.05339	0.00005
8	1.44444	0.01252	1.14645	**0.07248**	**0.00025**	8	1.17731	0.00944	**1.37976**	**0.05957**	0.00053
9	1.34413	0.01184	1.27987	0.07019	0.00035	9	1.20541	0.01017	1.65637	0.04158	0.00003
10	1.35426	0.01161	1.24319	0.05144	0.00029	10	1.11637	0.00955	1.68305	0.04715	0.00004

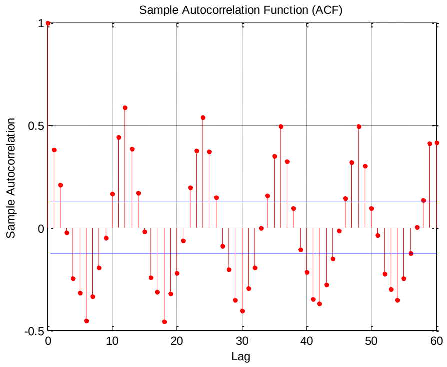
			Case 3						Case 4		
n	**SC**	**S**	**XB**	**DI**	**ADI**	**n**	**SC**	**S**	**XB**	**DI**	**ADI**
2	3.01261	0.01506	10.9572	0.00937	0.05451	2	3.91141	0.01985	4.33106	0.03907	0.21636
3	1.86610	0.01052	2.37837	0.06074	0.00276	3	2.21344	0.01223	3.49022	0.02466	0.00322
4	1.50754	0.00834	3.94808	0.03525	0.00222	4	1.62391	0.00984	2.37308	0.02327	**0.00042**
5	1.31016	**0.00820**	2.99020	0.06652	0.00149	5	1.64441	0.01490	2.13244	0.02376	0.00298
6	**1.20586**	0.01012	**1.71468**	0.04450	0.00363	6	1.41165	0.01191	1.84875	0.06552	0.00482
7	1.21073	0.01047	1.49790	**0.07494**	**0.00027**	7	1.39905	0.01107	1.92281	0.07084	0.00450
8	1.10700	0.00997	1.67046	0.07015	0.00179	8	1.32365	0.01134	**1.45646**	0.06849	0.00474
9	1.14038	0.00917	1.66426	0.07015	0.00036	9	**1.11106**	**0.00872**	1.56368	**0.07786**	0.00213
10	0.92099	0.00803	1.45111	0.07197	0.00018	10	1.06865	0.00924	1.48006	0.05864	0.00006

			Case 5						Case 6		
n	**SC**	**S**	**XB**	**DI**	**ADI**	**n**	**SC**	**S**	**XB**	**DI**	**ADI**
2	4.42081	0.02210	4.60425	0.04898	0.02841	2	2.97768	0.01512	4.06645	0.03108	0.00531
3	2.11103	0.01283	2.99875	0.03382	0.00097	3	1.99997	0.01240	1.91439	0.03418	0.00440
4	1.74476	0.01356	1.73027	0.04019	0.00116	4	1.47764	0.00937	2.46733	0.05147	0.00215
5	1.52363	0.01250	2.41909	0.03959	**0.00034**	5	1.53060	0.01366	1.86498	**0.07202**	**0.00001**
6	1.37723	0.01029	1.63524	0.04936	0.00074	6	1.49642	0.01152	2.09082	0.03542	0.00255
7	**1.22792**	**0.00898**	1.56688	0.04427	0.00004	7	**1.22999**	0.01051	1.90269	0.05126	0.00096
8	1.20654	0.00905	1.19568	0.05776	0.00017	8	1.15223	**0.00911**	**1.28961**	0.05068	0.00148
9	1.04025	0.00843	**1.11898**	**0.06938**	0.00014	9	1.08278	0.00922	1.20478	0.06540	0.00063
10	1.05519	0.00847	1.24648	0.06373	0.00001	10	1.08686	0.00930	1.29220	0.06540	0.00001

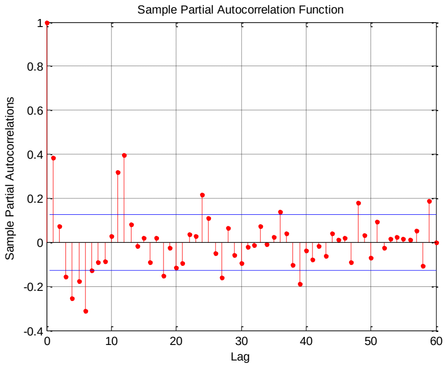
			Case 7						Case 8		
n	**SC**	**S**	**XB**	**DI**	**ADI**	**n**	**SC**	**S**	**XB**	**DI**	**ADI**
2	3.69848	0.01967	2.64381	0.04349	0.02559	2	3.13905	0.01764	6.79078	0.04602	0.01443
3	1.81911	0.00997	3.00923	0.02937	0.00852	3	1.76652	0.01057	2.72890	0.05869	0.00113
4	1.29284	**0.00780**	1.82292	0.03352	0.03788	4	1.45477	0.00924	2.47488	0.05511	0.00515
5	1.34172	0.01253	1.55378	0.04802	0.00557	5	1.37273	0.01202	2.24855	0.05538	0.00223
6	1.23881	0.01042	2.45065	0.03128	**0.00210**	6	1.25191	0.01028	1.66658	0.04574	0.00201
7	**1.12486**	0.00948	1.94834	0.04017	0.00146	7	1.20463	0.00951	**1.48238**	0.04574	**0.00003**
8	1.06936	0.00826	**1.38420**	**0.06700**	0.00078	8	**1.11148**	**0.00867**	1.88364	0.05566	0.00045
9	1.04575	0.00970	1.51423	0.06087	0.00146	9	1.10766	0.01091	2.21902	**0.08418**	0.00060
10	1.03915	0.00926	1.39951	0.04836	0.00027	10	1.10485	0.01054	1.57029	0.05024	0.00083

APPENDIX B

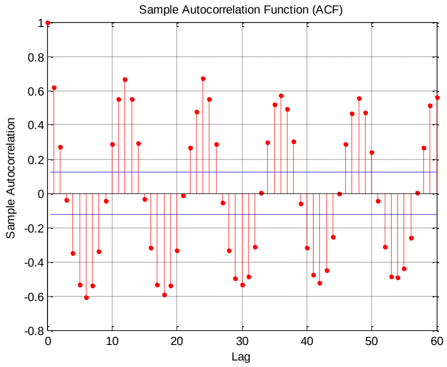
ACF and PACF of monthly rainfall time series data.



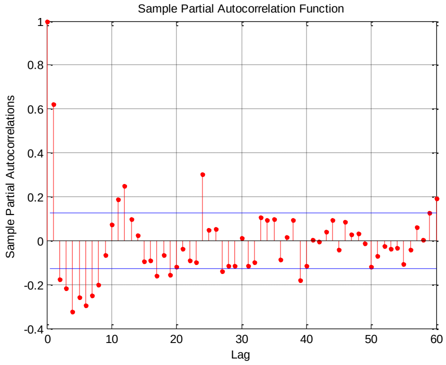
(Case 1’s ACF)



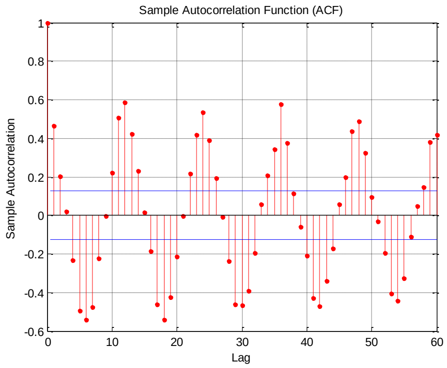
(Case 1's PACF)



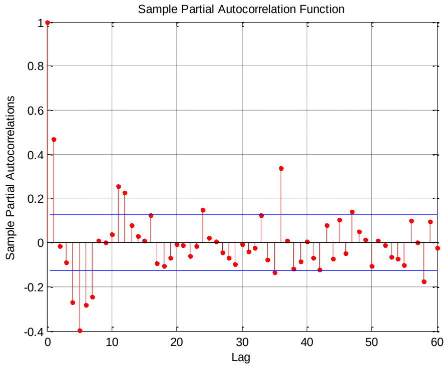
(Case 2’s ACF)



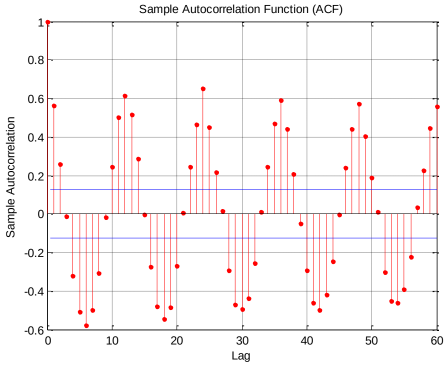
(Case 2's PACF)



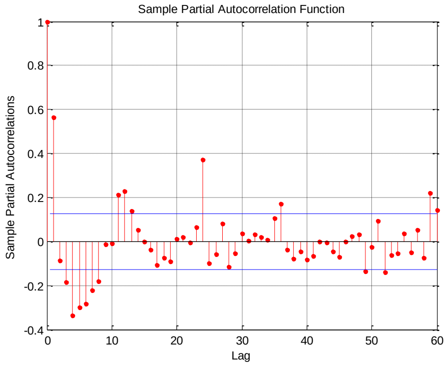
(Case 3’s ACF)



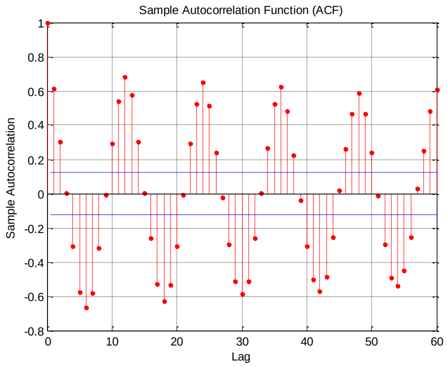
(Case 3's PACF)



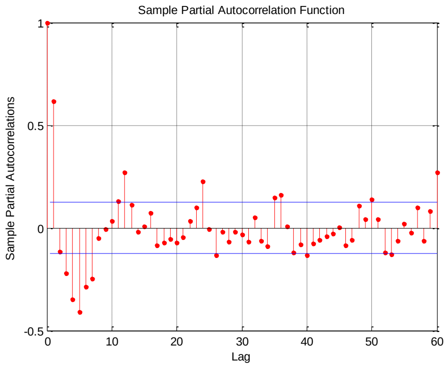
(Case 4’s ACF)



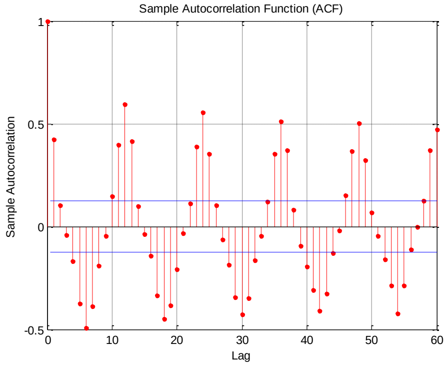
(Case 4's PACF)



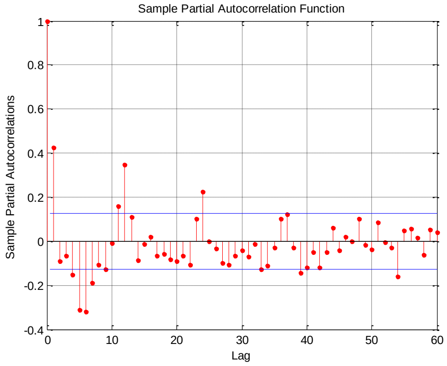
(Case 5’s ACF)



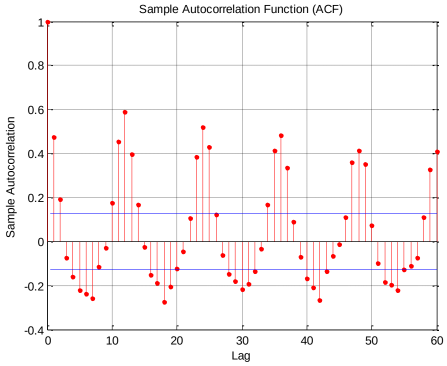
(Case 5's PACF)



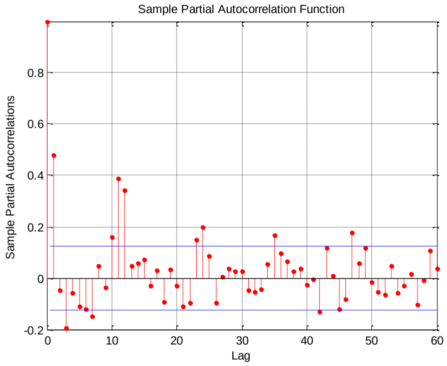
(Case 6’s ACF)



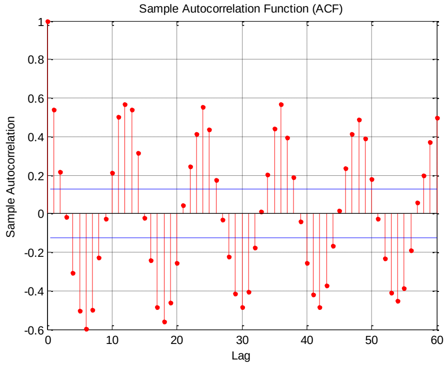
(Case 6's PACF)



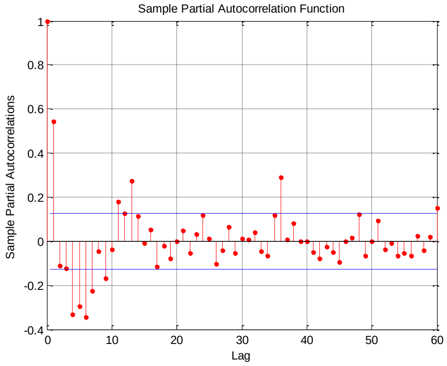
(Case 7's ACF)



(Case 7's PACF)



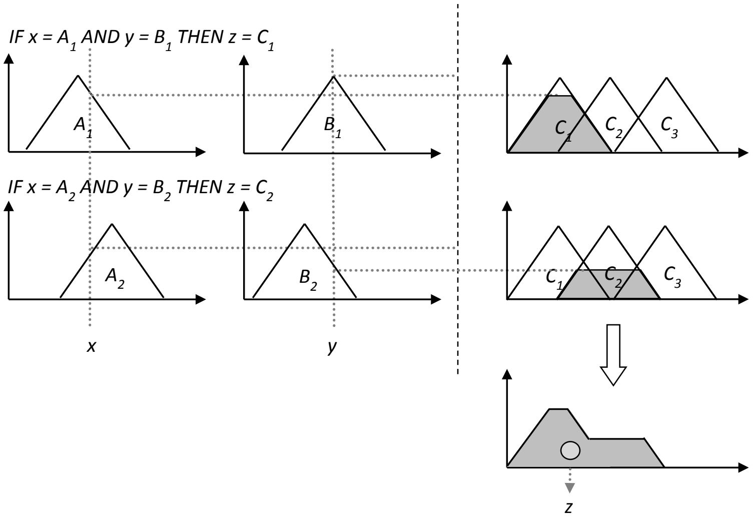
(Case 8's ACF)



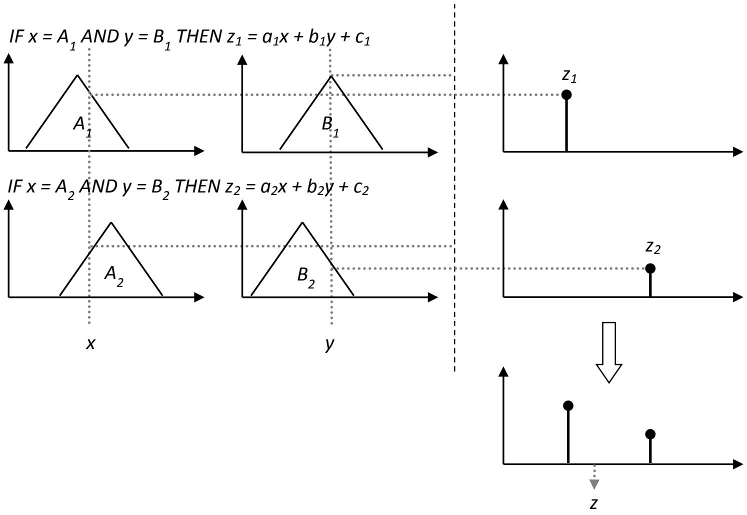
(Case 8's PACF)

APPENDIX C

Graphical Representations of MFIS and SFIS.


IF x = A1 AND y = B1 THEN z = C1
A1
B1
C1
C2
C3
IF x = A2 AND y = B2 THEN z = C2
A2
B2
C1
C2
C3
x
y
z

(a) Mamdani-Type Fuzzy Inference System (MFIS)


IF x = A1 AND y = B1 THEN z1 = a1x + b1y + c1
A1
B1
z1
IF x = A2 AND y = B2 THEN z2 = a2x + b2y + c2
A2
B2
z2
x
y
z

(b) Sugeno-Type Fuzzy Inference System (SFIS)