DEVELOPMENT OF AN ENHANCED ADAPTIVE RESONANCE THEORY MAPPING SYSTEM FOR WATERSHED CLASSIFICATION

PU LI

Development of an Enhanced Adaptive Resonance Theory Mapping System for Watershed Classification

A Thesis

by Pu Li

Submitted to School of Graduate Studies

for the Degree of

Master

in Engineering

Memorial University of Newfoundland

April, 2009

St John's, Newfoundland

ABSTRACT

Watershed classification is a process that classifies watershed sub-basins into certain groups due to similarities and/or differences in their characteristics. Such a process is of necessity and importance to support the decision making and practice of watershed monitoring, modeling, and management and helps in reducing the set up and running cost and improving efficiency. A watershed system is usually characterized by a large variety of topographical, hydrological, and ecological features, which provides the basis for watershed classification and also makes it a challenging task. Furthermore, many of the features and their interrelationships are hardly measured or quantified accurately due to the complexity and uncertainty of the system. Numerous studies have been conducted on watershed classification but the comprehensive consideration of both systematic complexity and uncertainty in the classification process is lacking. There is a need of more efficient and reliable approaches of watershed classification to deal with complex and uncertain features.

This research aims to fill the gap by developing a novel classification system based on the enhanced adaptive resonance theory (ART) mapping approaches to classify complex watershed features under uncertainty for supporting watershed modeling and management. The developed system is composed of : (1) a two-stage adaptive resonance theory mapping (TSAM) approach by integrating multitier ART into the system to form an unsupervised learning module for cluster centroid calculation and a supervised learning module for normalized original input classification; and (2) an integrated rule-

based fuzzy adaptive resonance theory mapping (IRFAM) approach by incorporating fuzzy set theory and rule-based operation to the system to form an unsupervised learning module for cluster centroid calculation and two supervised learning modules for criteria combination and fuzzified input classification.

To test the feasibility and efficiency, the developed system was applied to a real-world case study in the Deer River watershed, Canada. The results indicated that the watershed sub-basins were properly classified into preset target groups by both approaches in the given conditions (e.g., vigilance = 0.7). The TSAM approach could efficiently solve the problem of difficulties in criteria generation by using ART unsupervised classification and centriod determination in the first stage and feed the criteria to the ARTMap supervised classification in the second stage. In comparison with the TSAM, the IRFAM approach could take advantages of fuzzy set theory to generate full criteria combinations to match the input patterns and use the rule-based operation to screen the matched patterns into the target groups. This can efficiently handle the classification for the input patterns with a high degree of uncertainty and wide ranges of variations. In the case that there are not sufficient information for generating fuzzy membership functions, the TSAM could be a better choice than the IRFAM from a feasibility perspective; otherwise, the IRFAM could provide more accurate classification results than the TSAM.

ACKNOWLEDGEMENTS

The first person I wish to express my heartfelt appreciation is my supervisor Dr. Bing Chen, whose extremely helpful guidance, support and encouragement have nourished this dissertation research from the beginning to the successful completion. I would also like to extend my special thanks to Dr. Tahir Husain for serving as my co-supervisor and providing his insightful suggestions that were important and helpful in guiding my research. I will always remember the precious time I spent under their excellent supervision during the course of my study at Memorial University of Newfoundland, and I will take it as a roadmap to my future research.

I gratefully acknowledge the Faculty of Graduate Studies and Research and the Faculty of Engineering at the Memorial University of Newfoundland for providing research scholarships, teaching assistantships, and research awards during the course of my study. My further appreciation goes to ArcticNet and Manitoba Hydro and Dr. Ken Snelgrove and Mr. Liang Jing at Memorial University of Newfoundland for data supply and technological advices, which helped me a lot in the case study.

The chain of my gratitude would be definitely incomplete if I would forget to thank my friends for their assistance in various aspects of my research and for their support with constant friendship. They include: Liang Jing, Jing Ping, Xiao Jiang, Abdullah Mofarrah, Youcheng Lin, Guangxu Liu, Ledong Guan, and Xuan Jiao. I appreciate the discussions and interactions I had with them during my master study.

Special thanks go to my parents in China, Guoxin Li and Yueming Chen, my sister and border, Cuiyi Li and Jianzheng Li, and my girl friend Jiyi Liu for their benevolent prayer, love and support to achieve my ambitions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

AI	Artificial Intelligence.`
ANN	Artificial Neural Networks.
ART	Adaptive Resonance Theory.
ARTMap	Adaptive Resonance Theory Mapping.
FAM	Fuzzy Adaptive Resonance Theory Mapping.
FNN	fuzzy-neural networks.
GIS	geographical information system.
IRFAM	Integrated Rule-based Fuzzy Adaptive Resonance Theory Mapping.
NDVI	normalized difference vegetation index.
NF	Neuro-Fuzzy.
NFS	neural-fuzzy systems.
NNs	Neural Networks.
TM	thematic mapper.
TSAM	Two-Stage Adaptive Resonance Theory Mapping.
G	group.
L	low level.
M	Medium level.
H	high level.
I_N	output normalized matrix.
m	number of features.
n	number of input patterns or number of data points in the cluster.
x	data set in the input matrix.
ρ	vigilance parameter.
I	input matrix to ART.
j	category index.
w	adaptive weights.
w_j	weight vector.

α	choice parameter.
β	learning rate.
J	chosen category.
T_j	choice function.
C	centroid matrix of each cluster.
y	value of the feature in each data point.
x_i	centroid value for each feature.
F^{ab}	map field.
F_2^a	feature representation field.
F_2^b	category representation fiel.
x^{ab}	output vector.
y^b	input pattern in ART_b.
ρ_{ab}	map field vigilance parameter.
$\overline{\rho}_a$	baseline vigilance.
I_a	input patterns to the ART_{2a}.
I_b	input patterns to the ART_{2b}.
μ	membership function.
X	a set of data points.
d	lower bound.
c	the point where $\mu_i(x) = 1$.
e	upper bound.
I_{a_0}	original input.
Y	fuzzy set.
I_a	fuzzified input.
I_{b0}	criteria combination matrix.
y_{ij}	inputs which membership function equal to 1.
I_b	fuzzified criteria combination matrix.

CHAPTER 1: INTRODUCTION

Watershed classification is a process that classifies watershed sub-basins into certain groups due to similarities and/or differences in their characteristics. Such a process is of necessity and importance to support the decision making and practice of watershed monitoring, modeling, and management and helps in reducing the set up and running cost and improving efficiency. A watershed system is usually characterized by a large variety of topographical, hydrological, and ecological features, which provides the basis for watershed classification and also makes it a challenging task. Furthermore, many of the features and their interrelationships are hardly measured or quantified accurately due to the complexity and uncertainty of the system.

Various classification methods have been developed in the past decades. For example, decision functions, distance functions and clustering, statistical approach, feature selection, fuzzy classification, and neural networks (Friedman and Kandel, 1999; Richard et al., 2001).

Decision functions are one kind of the traditional methods used for classification. When the number of classes is known and when the training patterns are such that there is geometrical separation between the classes a set of decision functions can often been used to classify an unknown pattern (Starseva, 1995). However, the relation between the complexity of the class of decision functions, the sample size, and the complexity of the distributions usually lead to statistical robustness problem and limit its application. Clustering presents another good example of the traditional classification methods. It is based on exploratory data analysis and aims to group a set of items into clusters such that items within a given cluster have a high degree of similarity (Bock, 1993; Jain et al., 1999). The commonly used clustering methods include hierarchical and partitioning (Spaeth, 1980; Gordon, 1999; Everitt, 2001), and dynamic clustering (Diday and Simon, 1976; Diday and Govaert, 1977). These methods can handle a high dimension of input factors, but need accurate description for the input patterns. If the input data become ambiguous, it will not be efficient by only using clustering methods.

In order to mitigate the impacts of uncertainties, the fuzzy set theory has been integrated with the traditional methods. Since the theory (Zadeh, 1965) is a generalization of the classical set theory, it has greater flexibility to capture various aspects of incompleteness or imperfection extensively existing in real life situations (Pedrycz, 1990; Pal et al., 2000). However, many studies indicated that the fuzzy set theory was week to handle high level of a system's complexities when the dimension of input factor is getting higher. Consequently, artificial neural networks (ANN) have been introduced to help deal with complexities and increase the speed of classification process (McCulloch and Pitts, 1943; Rosenblatt et al., 1962; Minsky and Papert, 1969; Hopfield, 1982; Hopfield and Tank, 1985; Rumelhart et al., 1986; Gorman and Sejnowski, 1988; Goodacre et al., 1992). These developments have brought neural network research to a new stage, resulting in a large number of ANN models and applications (Gopal et al., 1999; Lee et al., 1999; Han et al., 2002; Aires et al., 2004).

Since each classification method has both strengths and weakness, nowadays, combination of the traditional methods has become a common approach in the fields (Celeux and Mkhadri, 1992; Breiman, 1995; Bishop, 1995; Leblanc and Tibshirani, 1996; Raftery, 1996; Ferreira et al., 1999, 2000). Many attempts have been made in the last decade to design hybrid classification methods by combining the merits of individual techniques. Integration of neural networks (NNs) and fuzzy systems is one such hybrid technique and is known as neuro-fuzzy computing (NF) (Pal and Pedrycz, 1990; Ghosh, 1996; Pal and Mitra, 1999; Kuncheva, 2000; Abe, 2001). In the NF paradigm, much research effort has been made (Keller and Hunt, 1985; Kwon, Ishibuchi, and Tanaka, 1994; Ghosh and Pal, 1993; Pal and Ghosh, 1996; Pal and Mitra, 1999; Abe, 2001; Baraldi et al., 2001; Boskovitz and Guterman, 2002; Gamba and Dellacqua, 2003; Han et al., 2002; Qiu and Jensen, 2004). Among them, the adaptive resonance theory, a neural network that self-organize stable recognition codes in real time in response to arbitrary sequences of input patterns, has been regarded as a promising technique when it is incorporated with fuzzy set theory (Carpenter and Grossberg, 1987; Carpenter et al., 1991, 1992, 1997).

Traditional methods for watershed or land cover classification are based on remote sensing, which can be grouped into supervised and unsupervised classifications (Kartike, 1995). Some approaches are designed for the comparative changes of remote sensing data and image data (Myers et al., 1999; Kurihara et al., 2000). Geographical information system (GIS) and expert knowledge combined with the traditional method have gained recognition in watershed classification (Soheila et al., 2007). In order to obtain more accurate results under uncertain conditions, the fuzzy theory is usually integrated with the traditional methods (McMahan and Weber, 2003; Lucas et al., 2008). In order to handle the uncertainties and complexities in the system, fuzzy theory and ANN are introduced to help increase the accuracy and speed of the classification process (Daniel, 1993; Giles, 1995; Gopal et al., 1999; Tsoand Mather, 2001; Han et al., 2002; Varshney and Arora, 2004; Richards and Jia, 2006). However, there is still a lack of in-depth and comprehensive classification studies to deal with watershed classification when the selected features are complex and uncertain.

To help fill the gaps, this research aims at developing an integrated classification system to more efficiently and accurately classify watershed with high level of complexities and uncertainties by incorporating fuzzy set theory and ART mapping techniques. This objective entails the research of the following components: (1) development of a two-stage adaptive resonance theory mapping (TSAM) approach, which will consist of an unsupervised learning ART and a supervised learning ARTMap modules along with the normalization module and centroid determination modules; (2) development of an integrated rule-base fuzzy adaptive resonance theory mapping (IRFAM) approach, which will contain an unsupervised learning ART and two supervised learning ARTMap modules, accompanied with fuzzification, centroid determination, and rule-based operation modules; and (3) the application of the developed TSAM and IRFAM approaches to the Deer River watershed in Manitoba in order to test their efficiency and feasibility.

The thesis is composed of six chapters. Chapter 2 presents a comprehensive review on classification methods including traditional, fuzzy, neural network based, and hybrid approaches. The applications in watershed classification are discussed. In Chapter 3, a modified ART mapping approach, TSAM, fist developed by integrating three ART modules into the system to form an unsupervised learning module for cluster centroid calculation and a supervised learning module for normalized original input classification. Furthermore, the application of the TSAM approach to Deer River watershed is discussed.

In Chapter 4, an integrated ART approach, IRFAM, is develop by integrating a fuzzy interface and rule-based operation with multiple ART modules to form an unsupervised learning module for cluster centroid calculation, a supervised learning module for criteria combination classification, and a supervised learning module for fuzzified original input classification. In order to compare the classification efficiency, the IRFAM is applied to the same case as the one used for the TSAM. Chapter 5 presents the comparison between the TSAM and IRFAM approaches and discusses the differences from both statistical and realistic perspectives. Finally, conclusions of this dissertation research are drawn along with the recommendations for future work towards the end in Chapter 6.

CHAPTER 2: LITERATURE REVIEW

2.1 Classification Methods

The task of a classification system is to use the feature vector provided by the feature extractor to assign the object to a category. Because it is often impossible to conduct a perfect classification performance, a more general task is to determine the probability for each of the possible categories. The abstraction provided by the feature-vector representation of the input data enables the development of a largely domain-independent theory of classification (Friedman and Kandel, 1999; Richard et al., 2001).

The degree of difficulty of the classification problem depends on the variability in the feature values for objects in different categories. The variability of feature values for objects in the same category may be due to complexity which expresses a condition of numerous elements in a system and numerous forms of relationships among the elements as well as their dynamic changes, and/or noise which is often referred to uncertainty (Richard et al., 2001; Lloyd, 2006). The noise or uncertainty can be defined as follows: any property of the sensed pattern which is not due to the true underlying model but instead to randomness in the world or the sensors, furthermore, in practice is that it may not always be possible to determine the values of all of the features for a particular input (Schurmann, 1996; Friedman and Kandel, 1999; Richard et al., 2001).

In the pass decades, many classification methods have been developed under different conditions and some of them have been integrated to deal with complex situations. Some of these approaches have also been applied in to watershed classification.

2.1.1 Traditional Approaches

When the number of classes is known and when the training patterns are such that there is geometrical separation between the classes a set of decision functions can be often used to classify an unknown pattern (Starseva, 1995; Friedman and Kandel, 1999). The main obstacles to the achievement of high-quality classification are small sample sizes and complex distributions. On the one hand, a too strict limitation on the class of decision functions poses the question of whether this class is adequately consistent with the true distribution; the greater the inconsistency, the poorer is the classification. On the other hand, the more complex the class of functions used for a small sample size, the greater is the classification error. Consequently, the complexity of the chosen class of functions must match the existing sample size. The relation between the complexity of the class of decision functions, the sample size, and the complexity of the distributions comprises the sum and substance of the statistical robustness problem for classification decision functions (Richard et al., 2001).

Clustering (Bock, 1993; Jain, et al., 1999) is one of the most commonly used traditional classification approaches. It is an exploratory data analysis method that aims to group a set of items into clusters such that items within a given cluster have a high degree of similarity, while items belonging to different clusters have a high degree of dissimilarity. A number of cluster analysis techniques have been developed such as hierarchical, partitioning, and dynamic methods (Spaeth, 1980; Gordon, 1999; Everitt, 2001).

Hierarchical methods yield complete hierarchy, i.e., a nested sequence of partitions of the input data. Hierarchical methods can be either agglomerative or divisive.

Agglomerative methods start with trivial clustering, where each item is in a unique cluster, and end with the trivial clustering, where all items are in the same cluster. A divisive method starts with all items in the same cluster and performs divisions until a stopping criterion is met (Kraskov, 2003).

Partitioning methods try to obtain a single division of the input data into a fixed number of clusters. Often, these methods look for a partition that optimizes (usually locally) a criterion function. To improve the cluster quality, the algorithm is run multiple times with different starting points, and the best configuration obtained from all the runs is used as the output clustering. The partitioning methods mainly include: *K*-means clustering (Chris and Xiaofeng, 2004) and Fuzzy *c*-means clustering (Erminio and Guerrisi, 2002).

Dynamic cluster algorithms (Diday and Simon, 1976; Abrantes and Marques, 1998) are iterative two-step relocation algorithms involving, at each iteration, the construction of the clusters and the identification of the suitable representative of exemplar (means, exes, probability laws, groups of elements, etc.) of each cluster by locally optimizing an adequacy criterion between the clusters and their corresponding representatives. The k-means algorithm, with class representatives updated after all objects have been considered for relocation, is a particular case of dynamical clustering with the adequacy criterion being a variance criterion such that the class exemplar equals the center of gravity for the cluster.

The adaptive dynamic clusters algorithms (Diday and Govaert, 1977; Wang et al., 2006) also optimize a criterion based on a measure of fit between the clusters and their representation, but at each iteration there is a different distance for the comparison of each cluster with its representative. The idea is to associate each cluster with a distance which is defined according to the intra-class structure of the cluster. These distances are not determined once and for all, and they are different from one class to another. The advantage of these adaptive distances is that the clustering algorithm is able to recognize clusters of different shapes and sizes.

If the training patterns seem to form clusters the classifiers which use distance functions are often employed for classification. If each class is represented by a single prototype called the cluster center, a minimum-distance classifier can be used to classify a new pattern. A similar modified classifier is used if every class consists of several clusters. The nearest-neighbor classifier classifies a new pattern by measuring its distances from the training patterns and choosing the class to which the nearest neighbor belongs.

Each training pattern is in one of these classes but its specific classification is not known. In this case, some algorithms are used to determine the cluster (class) centers by minimizing some performance index. These centers are found iteratively and then a new pattern is classified using a minimum-distance classifier. One of such algorithms is c-Means where the exact number of classes is known. If there is a desired number k of clusters and the final number of classes which is determined by the algorithm cannot be much higher or much lower than k.

Besides clustering approaches, statistical approaches are also most used methods in classification. Many times the training patterns of various classes overlap for example when they are originated by some statistical distributions. In this case a statistical approach is appropriate, particularly when the various distribution functions of the classes are known. A statistical classifier must also evaluate the risk associated with every classification which measures the probability of misclassification. For example, the Bayes classifier based on Bayes formula from probability theory minimizes the total expected risk. This method is a fundamental statistical approach to the problem of pattern classification, which is based on quantifying the tradeoffs between various classification decisions using probability and the costs that accompany such decisions. It makes the assumption that the decision problem is posed in probabilistic terms, and that all of the revant probability values are known. To use Bayes classifier one must know the pattern distribution function for each class. If these distributions are not known they must be approximated using the training patterns. Sometimes the functional form of these distributions is known and one must only estimate its parameters. However, in some applications even the distribution's from is unknown and must be found (Friedman and Kandel, 1999).

The syntactic pattern classification which are also traditional classification approaches, utilizes the structure of the patterns. Typical patterns which are subject to syntactic pattern classification are characters, fingerprints, chromosomes, etc. In general, given a specific class, a grammar whose language consists of patterns in this class is designed. For an unknown new pattern a syntax classifier analyzes the pattern (a string) in a process called parsing and determines whether or not that string belongs to the language (class) (Friedman and Kandel, 1999).

2.1.2 Fuzzy Approaches

In order to obtain more accurate results under uncertain conditions, the fuzzy set theory is usually integrated with the traditional approaches. Quite often classification is performed with some degree of uncertainty. Modern control theory owes much in its development to mathematical models but when applied to real problems difficulties are often encountered in approximating real controlled objects by models because of the vagueness or fuzziness of the controlled objects. In addition, since most control theory is based on linear systems, it is difficult to develop control systems with good performance when real controlled objects have strong nonlinearity (Friedman and Kandel, 1999).

Since the fuzzy set theory is a generalization of the classical set theory, it has greater flexibility to capture various aspects of incompleteness or imperfection about real life situations (Zadeh, 1965). The significance of fuzzy set theory in the realm of pattern classification is effectively justified in various areas such as representing input patterns as an array of membership values denoting the degree of possession of certain properties, representing linguistically defined input features, representing multiclass membership of ambiguous patterns, generating rules and inferences in linguistic form, extracting ill-defined image regions, and describing relations among them (Pedrycz, 1990; Pal et al., 2000).

To apply fuzzy set theory to a system, experts' knowledge on the system needs to be expressed explicitly in if-then fuzzy rules. When the input to the fuzzy rules is given, the output is determined by inference using the fuzzy rules. This process of determining the output from input is one method of function approximation which is one of the major uses of multilayered networks. Function approximation is readily extended to pattern classification (Abe, 1997). Either the classification outcome itself may be in doubt, or the classified pattern may belong in some degree to more than one class. It is thus introduced fuzzy classification where a pattern is a member of every class with some grade of membership between 0 and 1(Friedman and Kandel, 1999)`.

2.1.3 Neural Network Approaches

Artificial neural network (ANN) has been introduced to handle the complexities in the system and help increase the speed of classification process. ANN is an attractive alternative to the statistical classifiers that can efficiently handle the complexities of the system (Schurmann, 1996; Dunne, 2007).

This kind of approach has attracted enormous attention, especially during the last decade, and is centered around the neuron as the basic building block of natural brains the unrivaled computational power and versatility of the biological brain are due to a complicated network of vast numbers of neurons. Biologists estimate that the human brain consists of 10^{11} neurons and that each of them is on the average connected to about 10^4 others (Patterson, 1996).

The neural network approach assumes as other approaches before that a set of training patterns and their correct classifications are given. The architecture of the net including input layer, output layer and hidden layers may be very complex. It is characterized by a set of weights and activation function which determine how any information (input signals) is being transmitted to the output layer. The neural network is trained by training patterns and adjusts the weights until the correct classifications are obtained. It is then used to classify arbitrary unknown patterns (Abe, 1997; Friedman and Kandel, 1999).

Neural networks have much in common with the structures needed for pattern classification. Pattern classification and neural networks go back to the same roots in the historic evolution of artificial intelligence (AI) techniques. The idea of neural networks is taken from biological systems performing pattern classification functions. It is no wonder that neural networks are considered to be predestined pattern classifiers. In this role they agree with the concepts developed in conventional pattern classification. (Mandic and Chambers, 2001; Dunne, 2007).

The neural network research, from the viewpoint of information processing, started from the neuron model proposed by McCulloch (1943) and Pitts in 1943. The output of the model takes the values of 1 and 0 as discussed afterwards, and when the input exceeds some predetermined threshold, the output changes stepwise from 0 and 1. From the end of the 1950's to the 1960's, Rosenblatt et al. (1962) developed preceptrons which connect the above neurons in layers and used them to study pattern classification. The perceptron is the origin of the now widely used multilayered network. Minsky and Papert (1969) showed the limitation of perceptrons, i.e., that they are only applicable when data belonging to different classes are linearly separable, interest in neural network rapidly shrank.

Neural networks have been shown (Cybenko, 1989; Funahashi, 1989; Hornik et al., 1989) to be able to approximate any continuous function arbitrarily well when sufficiently many hidden nodes are used. In the Bayesian context, the posterior is consistent (Lee, 2000). These properties make neural networks a good method for nonparametric regression, in that one does not have to choose a particular parametric form for the model.

In supervised classification tasks, a classification model is usually constructed according to a given training set. Once the model has been built, it can map a test data to a certain class in the given class set. Many classification techniques including decision tree (Qinlan, 1986; Freund, 1995), neural network (NN) (Lu et al., 1996), support vector machine (SVM) (Boser et al, 1992; Vapnik, 1995), rule based classifiers systems etc. have been proposed. Among these techniques, decision tree is simple and easy to be comprehended by human beings. SVM is a new machine learning method developed on the Statistical Learning Theory. SVM is gaining popularity due to many attractive features, and promising empirical performance. SVM is based on the hypothesis that the training samples obey a certain distribution which restricts its application scope. Neural network classification, which is supervised, has been proved to be a practical approach with lots of success stories in several classification tasks. However, its training efficiency is usually a problem, training on only the new silhouette could result in the network learning that pattern quite well, but forgetting previously learned patterns. Although retraining may not take as long as the initial training, it still could require a significant investment. Adaptive resonance theory (ART) was developed to solve this problem by using the short-term memory (STM) to storage the contrast-enhanced pattern, and the long-term memory (LTM) to implement an arousal mechanism, whereas the STM is used to cause gradual changes in the LTM (Grossberg, 1976).

2.1.4 Adaptive Resonance Theory Approaches

Adaptive resonance architectures are neural networks that self-organize stable recognition codes in real time in response to arbitrary sequences of input patterns. The basic principles of adaptive resonance theory (ART) were introduced by Carpenter and Grossberg (1987). ART networks self-organize stable recognition categories in response to arbitrary sequences of analogue (gray-scale, continuous-valued) input patterns, as well as binary input patterns.

There are two major ART paradigms distinguished by their forms of input data and processing. ART-1 is designed to accept only binary input vectors (Carpenter and Grossberg, 1987) whereas ART-2 (Carpenter and Grossberg, 1987) and Fuzzy ART (Carpenter et al., 1991, 1992) can also classify analog inputs. Both models can stably learn to categorize input patterns presented in an arbitrary order. There are many variations of ART models developed in different application domains, such as geomatic analysis, land cover classification, and Image analysis (Carpenter et al., 1991, 1997; Gopal et al., 1999; Chen, 1999).

ART models have been proposed under supervised learning conditions (Carpenter et al., 1991). ARTMap, a hierarchical network architecture, is able to rapidly self- organize stable categorical mapping between a given set of binary input vectors and binary output vectors while minimizing predictive error in an online setting. The Fuzzy ARTMap (FAM) model is an extension of ARTMAP that can learn stable recognition categories given both analog and binary input patterns. The ART modules of ARTMAP are replaced by Fuzzy ART modules in FAM. A brief description of ART that forms the basic modules in FAM architecture is given below. In Fuzzy ART the fuzzy logic AND connective, nun, is used to extend the method to real values in ART1 (Carpenter et al., 1991).

FAM processes uncertain (fuzzy) information and transforms it in terms of hyper-rectangles. Learning in FAM encompasses the recruitment of new hyper-rectangular protoypes and expansion of the boundary of existing prototypes in the feature space. Like other incremental ANNs, the growth criterion of FAM is subject to a similarity measure between the input pattern and the prototypes stored in the network. Given an input pattern, the prototype that has the highest degree of similarity with the input pattern is selected. A user-defined threshold is then used to decide whether or not the similarity level between the input pattern and the selected prototype is satisfactory to the user-defined level. If none of the prototypes can be found to meet the criterion, a new prototype is created and added to the network. However, a profound distinction between FAM and other incremental networks is that the Fuzzy ART (Carpenter and Grossberg, 1991) modules of FAM undergo a two-stage hypothesis selection and test process. On presentation of an input pattern, a feed forward pass is carried out to identify the most similar prototype according to a competitive selection process. The winning prototype is then tested against a vigilance threshold in the feedback pass. As long as the vigilance criterion is not satisfied, a new cycle of search (selection and test) for a new winning prototype will be initiated. This search process is continued until the criterion is satisfied by an existing prototype, or the creation of a new node that includes the input pattern. Obviously, this feedback mechanism has attributed to the formation of stable yet plastic knowledge structure in FAM (Carpenter et al., 1992).

The growth process of FAM allows boundary expansion of existing prototypes as well as inclusion of new nodes to the network without retraining. One of the undesirable effects of prototypical growth in FAM is the overlapping effect extended from node expansion (Simpson, 1992; Carpenter, 1997). It may not be a problem when overlapping occurs among prototypes of the same class.

However, if overlapping occurs among prototypes of different classes, it could bring an undesirable effect that causes the onset of ambiguity following hardly/undistinguishable regularities in the feature space. If the undesired overlapping prototypes of different classes remain unsettled, conflicting information would be retained as templates in the feature space.

2.1.5 Integrated Approaches

Different pattern classifiers trained for the same application can be viewed as approximations from different directions to the same goal, just as different starting point are possible to reach the same peak in a mountainous territory. Therefore, different pattern classifiers, derived from different concepts, using different sets of measurements, or designed with different constellations of their basic design parameters tend to behave differently in the individual case, even if they may exhibit the same long-term error rates. Under these circumstances combining different pattern classifiers developed for the same task bears the promise of improving the overall performance, just as in everyday life more than one expert is consulted if a difficult case is to be settled. Since different pattern classifiers have different strengths and weaknesses, classifier combination must be led by the goal of making the respective strengths effective and repelling the deficiencies (Schurmann, 1996).

Nowadays combining models has become a common approach in several fields, such as regression, neural networks, discriminant analysis, and so forth (Celeux and Mkhadri, 1992; Breiman, 1995; Bishop, 1995; Leblanc and Tibshirani, 1996; Raftery, 1996; Ferreira et al., 1999, 2000).

Many attempts have been made in the last decades to design hybrid systems for pattern classification by combining the merits of individual techniques. An integration of neural networks (NNs) and fuzzy set theory is one such hybrid technique and known as neuro-fuzzy (NF) computing (Pal and Ghosh, 1996; Pal and Mitra, 1999; Abe, 2001).

Both NNs and fuzzy approaches are adaptive in the estimation of the input-output function without any precise mathematical model. NNs handle numeric and quantitative information while fuzzy approaches can handle symbolic and qualitative data. Apart from this, in a fuzzy classifier patterns are assigned with a degree of belonging to different classes. Thus the partitions in fuzzy classifiers are soft and gradual rather than hard and crisp. Therefore, an integration of neural network and fuzzy approaches should have the merits of both and enable one to build more intelligent decision making systems. Fuzzy set theory is found to be more suitable and appropriate to handle these situations reasonably (Pedrycz, 1990; Kuncheva, 2000).

In the NF paradigm, much research effort has been made (Keller and Hunt, 1985; Ghosh and Pal, 1993; Kwon et al., 1994; Pal and Ghosh, 1996; Pal and Mitra, 1999; Abe, 2001; Baraldi et al., 2001; Boskovitz and Guterman, 2002; Han et al., 2002; Gamba and Dellacqua, 2003; Qiu and Jensen, 2004). NF hybridization is done broadly in two ways: NNs that are capable of handling fuzzy information (named as fuzzy-neural networks, FNN), and fuzzy systems augmented by NNs to enhance some of their characteristics such as flexibility, speed and adaptability (named as neural-fuzzy systems, NFS) (Pal and Ghosh, 1996; Pal and Mitra, 1999).

The NN and fuzzy approaches discussed so far can be applied to pattern classification and function approximation. Buckley et al (1992) reported that fuzzy systems and multilayered networks were mathematically equivalent in that they are convertible. But since the two approaches differ, they have their own advantages and disadvantages.

With multilayered networks, knowledge acquisition is done by network training. Namely, by gathering input-output data for pattern classification of function approximation and training the network using these data by the back propagation algorithm, the desired function is realized. On the other hand, fuzzy rules need to be acquired by interviewing experts. But for complicated system expert knowledge that is obtained intuition and experience is difficult to express in a rule format. Thus rule acquisition requires much time. As methods to extract fuzzy rules from numerical data, Wang and Mendel's method (1992) extracts fuzzy rules directly from data and Lin and Lee's method (1991) uses neural networks. Lin and Lee trained the neural network in which fuzzy rules were imbedded, extracted fuzzy rules from the trained network, and tuned the membership functions of extracted fuzzy rules using the same neural network.

The major shortcoming of neural networks is represented by their low degree of human comprehensibility. Many attempts have been made to solve this shortcoming of neural networks, by compiling the knowledge captured in the topology and weight matrix of a neural net work, into a symbolic form; most of them into sets of ordinary if-then rules (Towell and Shavlik, 1993; Yoo, 1993; Craven and Shavlik, 1993; Thrun, 1994), or into sets of fuzzy rules (Lin and Lee, 1991, Palade et al., 1996). The fuzzy neural networks are often used as an auto-tuning method for the determination and the adjustment of fuzzy rules.

2.2 Watershed Classification

The traditional classification of a watershed is based on basic hydrological characteristics which can be grouped into supervised and unsupervised classifications. The commonly used supervised classification methods are: K-Nearest Neighbor, Decision Tree Classifier, and Bayesian Classifier. Unsupervised classification includes Bayesian Learning, Maximum Likelihood Classification, and Clustering Classification.

To further improve geomatic analysis methodology by combining computer technologies and practical experience of researchers several other methods have also been developed (Kartike, 1995). For example, echelon analysis (Myers et al., 1997) is a useful technique to study the topological structure of a surface in a systematic and objective manner. The echelons are derived from the changes in topological connectivity. However, the traditional sub-basins classification methods are delineation based on land surface topography and the channel orders which always leads to large number of sub-basins. This leads to the difficulty of watershed modeling and management.

Geographical information system (GIS) and expert knowledge combined with the traditional supervised and unsupervised methods have gained recognition in watershed classfication (Soheila et al., 2007). Running et al. (1995) developed a simple logic for classifying global vegetation based on observable and unambiguous characteristics of vegetation structure that were important to ecosystem biogeochemistry and could be monitored on-site for model validation purposes. As land cover data is normally derived from remote sensing data by the application of a conventional statistical classification, these classification techniques are not always appropriate to address system complexities because they sometimes may make untenable assumptions for the data and produce misleading classification results.

Furthermore, the character of the sub-basins is based on many features such as area, elevation, channel length, channel slope, vegetation, soil type, etc. Therefore, a classification system that can reasonably classify watersheds is critical and desired to support more efficient watershed modeling and management practices. However, watershed systems are complex and usually featured by a variety of topographic, hydrologic, and ecologic conditions. Furthermore, many of the features are hardly measured accurately or measured by numeric character and hence linguistic or fuzzy of features can play an important role in handling system uncertainties.

In order to obtain more accurate results under uncertain conditions, the fuzzy set theory has been integrated with the traditional methods. McMahan and Weber (2003) used fuzzy classification to characterize the system complexity and watershed heterogeneity in more accurate predictions compared to supervised classification. Lucas et al. (2008) developed a fuzzy classifier which is the extension of the approach in which uncertainty is represented by the additional extra dimension in land cover classification. In order to handle the complexities in the system, ANN is also introduced to help increase the speed of classification process, and handle the complexities of the system as an attractive alternative to the statistical classifiers. Daniel (1993) applied an ANN approach to the problem of deriving land cover information from landset satellite thematic mapper (TM) digital imagery. It provided more accurate and valuable data for use with GIS than the traditional statistical methods. Gopal et al. (1999) used the ARTMap networks to conduct the classification of global land cover based on normalized difference vegetation index (NDVI). The overall results of classification suggested that ARTMap provided a viable technique for global land cover classification, and the authors also suggested that there is a great deal of uncertainty in global land cover types .

Fuzzy set theory and neural network approach can be integrated to handle the system where complexities and uncertainties coexist. Uncertainties can arise at any stage of a pattern classification system, resulting from incomplete or imprecise input information, ambiguity or vagueness in input data, ill-defined and/or overlapping boundaries among classes or regions, and indefiniteness in defining/extracting features and relations among them. It is therefore necessary for a classification system to have sufficient provision for representing uncertainties involved at every stage so that the final output (results) of the system is associated with the least possible uncertainty. Complexities grow up when the number of input features in the system is getting up and the interactions in these features getting more and more complex, as well as the activities from the outside system. The complexities could reduce the efficiency and increase the required time for classification process. Furthermore, complexities also may lead to low accuracy of the classification results. The uncertainty and complexity handling issue becomes more prominent in case of land cover classification of remote sensing imagery (Richards and Jia, 2006; Tsoand Mather, 2001; Varshney and Arora, 2004).

Some fuzzy neural network systems have been introduced in land cover classification. Giles (1995) used ANN as an alternative to the statistical classifiers and integrated a fuzzy classification output from a remote sensing data set that was preprocessed with ancillary data available in a GIS to increase the accuracy with which land cover was mapped. Lee et al. (1999) developed a neural-fuzzy classifier derived from the generic model of a 3-layer fuzzy perceptron for land cover classification and compared it with the maximum-likelihood classifiers. The result showed that neural-fuzzy classifier was considerably more accurate in general but less accurate in some particulate areas. They concluded that the neural-fuzzy model could be used to classify the mixed composition area. Han et al. (2002) conducted a comparative evaluation of Neural-Fuzzy, Neural Network, and Maximum Likelihood Classifiers for land cover classification. They concluded that that the neural-fuzzy classifier was the most accurate method for land cover classification and suitable under the condition of uncertainty and complexity.

2.3 Summary

In this chapter, classification reviews on traditional approaches, fuzzy approaches, neural network approaches, and adaptive resonance theory approaches have been made. And the integrated approaches based on these methods had also been discussed. Furthermore, the applications of the classification approaches and their integrated approaches in watershed classification have been discussed.

Different classification approaches have their own advantages in handling some situation and weakness in some of the others. Most of the time, the real world case various conditions are coexisting, thus it is not enough to use only one approach to process the classification. Therefore, different approaches are always integrated to solve the real world problems.

The traditionally watershed is delineation based on land surface and the channel orders which usually generates large number of sub-basins and leads to the difficulty of watershed modeling and management. Usually, there are full of uncertainties and complexities in watershed system. ART and fuzzy set theory can be integrated to handle the watershed classification where complexities and uncertainties coexist. For ART can efficiently handle the system complexity and obtain fast learning speed but it is weak in handling linguistic input data; and fuzzy set theory has high ability in handle uncertainties, but it will getting inefficiently if system getting complex.

CHAPTER 3: TWO-STAGE ADAPTIVE RESONANCE MAPPING (TSAM) APPROACH

3.1 Background

Adaptive Resonance Theory (ART) and ART mapping (ARTMap), theory of human cognitive information processing (Grossberg, 1976, 1980, Carpenter, 1992), has been introduced as a series of real-time neural network models for unsupervised and supervised classifications. It is capable of learning stable recognition categories in response to arbitrary input with either fast or slow learning.

ART unsupervised classification will generate relative large number and unpredictable results, while ARTMap can generate predictable results but it needs criteria for supervised learning. Usually, the criteria for classification are not easy to be obtained if the reference information is not enough. TSAM approach is develop to achieve the automatically classification by generating the criteria from ART getting the supervised classification results from ART mapping.

As shown in Figure 3.1, the developed TSAM approach includes two stages: The first stage is the centroid determination subsystem which can locate the centroids for the expected target groups by unsupervised ART module, and use the determined centroid as the criteria in the second stage; and the second stage is the classification subsystem which can classify the normalized original input. There are three ART modules integrated in the TSAM which are as follows: ART_1 is used for processing unsupervised classification for the normalized original input and generating the unsupervised classified groups; ART_{2a}

and ART_{2b} are used in an ART Mapping module for comparing the combinations determined in the first stage and the normalized original inputs, and classifying them.

3.2 Methodology

3.2.1 Normalization

Because the ART system can only handle data values between 0 and 1, the input data should be normalized before entry the system through the following equation:

$$I_N := \left(\frac{x_{ij} - x_{j\min}}{x_{j\max} - x_{j\min}} \right), i = 1,2,\cdots n;\ j = 1,2,\cdots m \tag{3.1}$$

where I_N is the output normalized matrix; x is the data set in the input matrix; n is the number of input patterns; and m is the number of features.

3.2.2 ART Systems

ART can adaptively create a new category corresponding to an input pattern if there is not any existing similar enough to the pattern. This process is called the vigilance test and incorporated into the adaptive backward network (Grossberg, 1976, 1980). Therefore, ART architecture allows the user to control the degree of similarity of patterns which are placed in the same category.

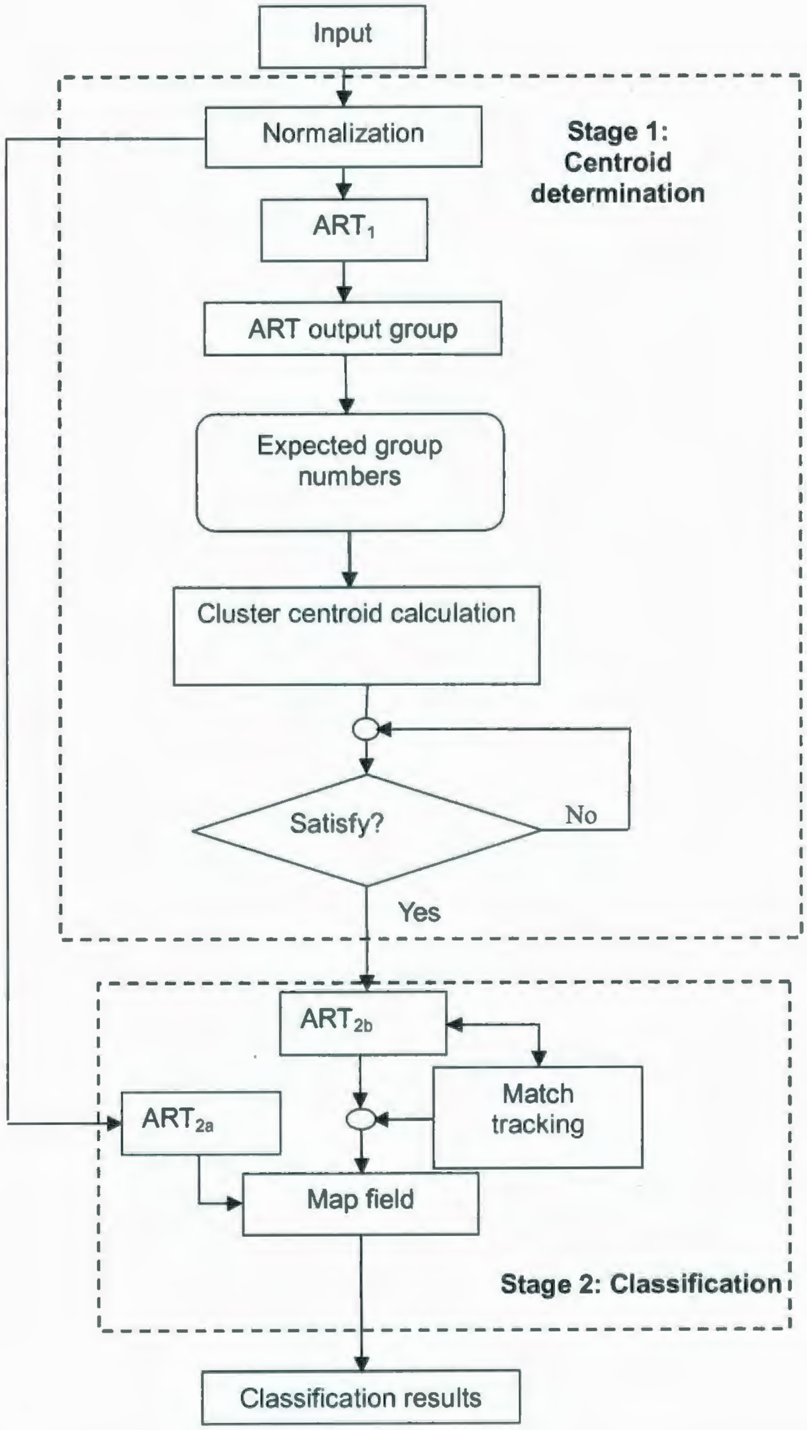


Figure 3.1 Flow chart of the Two-Stage ART-ARTMap Approach (TSAM)

Figure 3.2 shows the architecture of ART. The system consists of two layers, F1 and F2, which are connected to each other via the long term memory (LTM). The input pattern is received at F1, whereas classification takes place in F2. As mentioned before, the input is not directly classified. First a characterization takes place by means of extracting features, giving rise to activation in the feature representation field. The expectations, residing in the LTM connections, translate the input pattern to a categorization in the category representation field. The classification is compared to the expectation of the network, which resides in the LTM weights from F2 to F1. If there is a match, the expectations are strengthened, otherwise the classification is rejected. Each neuron in F1 is connected to all neurons in F2 via the continuous-valued forward LTM w^f, and vice versa via the binary-valued backward LTM w^b. The other modules are gain 1 and 2 (G1 and G2), and a reset module. Each neuron in the comparison layer receives three inputs: a component of the input pattern, a component of the feedback pattern, and a gain G1. A neuron outputs a 1 if and only if at least three of these inputs are high: the 'two-thirds rule.' The neurons in the recognition layer each compute the inner product of their incoming (continuous-valued) weights and the pattern sent over these connections. The winning neuron then inhibits all the other neurons via lateral inhibition. Gain 2 is the logical 'or' of all the elements in the input pattern x. Gain 1 equals gain 2, except when the feedback pattern from F2 contains any 1; then it is forced to zero. Finally, the reset signal is sent to the active neuron in F2 if the input vector x and the output of F1 by more than some vigilance level.

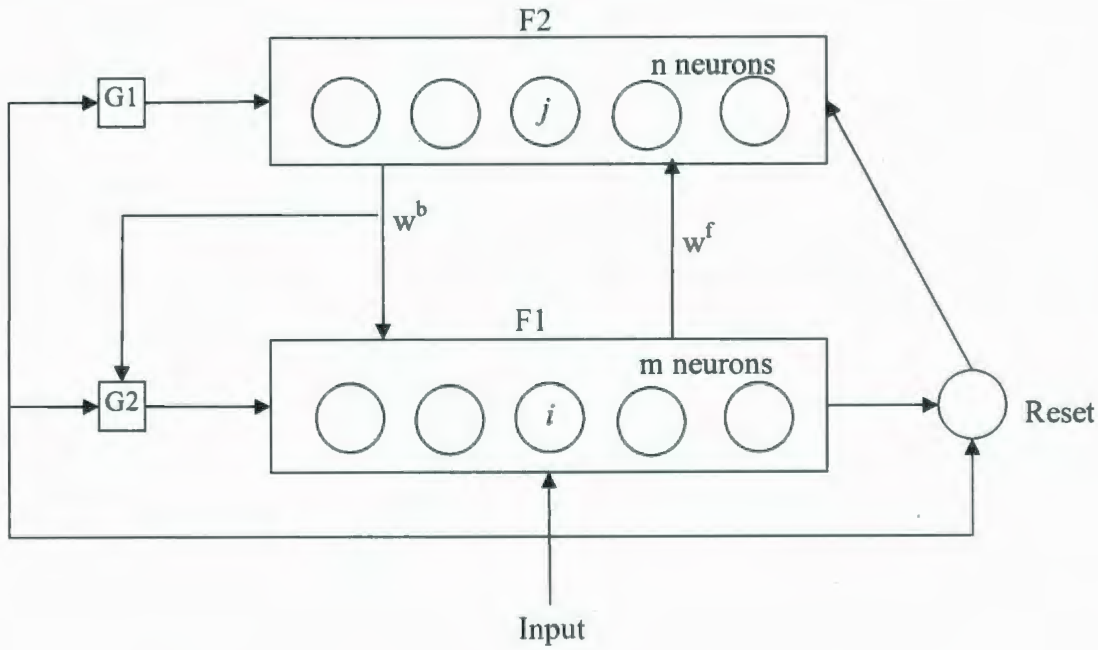


Figure 3.2 Architecture of ART neural network

The meaning of sufficiently similar pattern depends on the vigilance parameter ρ, where $0 < \rho < 1$. If ρ is small, the result tends to be a coarse categorization. The maximal vigilance parameter allows ART to classify input patterns into the highest classification speed. The evaluation of how the recognition rate of ART for a given pattern set is applied to find the maximal vigilance parameter. For the evaluation, a completely experimental design is used with various transformation effects and vigilance values. The maximal vigilance parameter can be found by using the differential function and bisection method (Grossberg, 1976, 1980). Furthermore, $\alpha>0$ is the choice parameter to break the tie when more than one prototype vector is a fuzzy subset of the input pattern, based on the winner-take-all rule (Xu et al., 2009).

Fuzzy operation is incorporated into the ART neural network which only learns to categorize binary input patterns to help ART learn and categorize analog patterns. The generalization of learning both analog and binary input patterns is achieved by replacing the appearance of the logical AND intersection operator ($\cap$) in ART by the min operator ($\wedge$) of fuzzy set theory. Assume each input I is an m-dimensional vector ($I_1, I_2, I_3, \ldots, I_m$). Let each category ($j$) correspond to a vector $w_j = w_{j1}, w_{j2}, w_{j3}, \ldots, w_{jm}$ of adaptive weights. The number of potential categories n ($j = 1, 2, \ldots, n$) is arbitrary. The ART weight vector w_j assumes both the bottom-up weight vectors (w^f) and the top-down weight vectors (w^b) of ART. The learning parameter β defines the degree to which the weight vector w_j learns characteristics of an input vector that is claimed by node J. Three parameters including choice parameter (α), learning parameter (β) and vigilance parameter (ρ) are to be adjusted to form the appropriate number of. The influence of these parameters to ART is noted as follows: 1) when the value of either the learning or vigilance parameter increases, the number of the categories increases; and 2) when the value of the choice parameter decreases, the number of the clusters increases. The Fuzzy ART algorithm is as follows (Carpenter and Grossberg, 2003):

Step 1: Initialize

Initially, each category is said to be uncommitted and the predetermined weight vector w_j is set as

$$w_{j1}(0) = w_{j2}(0) = \cdots = w_{jm}(0) = 1 \tag{3.2}$$

Then, a choice parameter α, a learning rate β, and a vigilance parameter ρ are set as:

$$1 > \alpha > 0,\ \beta \in [0,1],\ \rho \in [0,1] \tag{3.3}$$

Step 2: Complement coding

To improve the reliability of a category choice, input I_N is expanded with complement coding as follows:

$$I = (I_N, I_N^{\,c}),\ and\ I_N^{\,c} = 1 - I_N \tag{3.4}$$

Step 3: Category choice

For each I and category j the choice function T_j is defined by

$$T_j(I) = \frac{|I \wedge w_j|}{\alpha + |w_j|} \tag{3.5}$$

where $\wedge$ is the AND operator and defined by $(x \wedge y)_i = \min(x_i, y_i)$ and the $|\ |$ is defined by

$$|x| = \sum_{i=1}^{m} |x_i| \tag{3.6}$$

The system makes a category choice when at most one category can become active at a given time. The index J denotes the chosen category, where

$$T_J = \max\{T_j : j = 1, \cdots n\} \tag{3.7}$$

If more than one T_j is maximal, the system chooses the category with the smallest j index. In particular, nodes become committed in order $j = 1, 2, 3, \cdots n$.

Step 4: Resonance or reset

Resonance occurs if the match function of the chosen category meets the vigilance criterion; that is, if

$$\frac{|I \wedge w_J|}{|I|} \geq \rho \ , \tag{3.8}$$

learning then ensues according to the following equation (Carpenter et al., 1991);

$$w_J^{(new)} = \beta\left(I \wedge w_J^{(old)}\right) + (1 - \beta) w_J^{(old)} \ . \tag{3.9}$$

Fast learning corresponds to β = 1, which follows the learning rule presented by:

$$w_J^{(new)} = I \wedge w_J^{(old)} \tag{3.10}$$

A mismatch reset occurs if

$$\frac{|I \wedge w_J|}{|I|} < \rho \tag{3.11}$$

Then the value of the choice function T_J is reset to -1 for the duration of the input presentations to prevent its persistent selection during the search. A new index J is chosen by Eq 3.7. The search process continues until the chosen J satisfies Eq 3.8.

3.2.3 Centroids Determination

After the input patterns are classified by ART, the centroids are going to be located based on the expected target groups by the operation of the centroids locating module. For m expected target groups, the first m clusters which have the most data points in the clusters are selected. The centroid matrix of each cluster is given by:

$$C := (x_i)_{i=1,\cdots,m} \tag{3.12}$$

where m is the number of features of input data, and the centroid value for each feature c_i is given by:

$$c_i = \begin{cases} x_i & if\ i = 1 \\ \sum_2^n \left(\frac{i-1}{i}\sum_1^{i-1} x_{i-1} + \frac{1}{i}x_i\right) & if\ i > 1 \end{cases} \tag{3.13}$$

where n is the number of data points in the cluster and y is the value of the feature in each data point. The outputs of centroids are going to be used as the criteria to classify the normalized input pattern.

3.2.4 Map Field Activation

As shown in Figure 3.3, the map field F^{2ab} is activated whenever one of the ART_{2a} or ART_{2b} categories is active. If node J of feature representation field F_2^{2a} is chosen, then its weights w_J^{2ab} activate F^{ab}. If node K in category representation field F_2^{2b} is active, then the node K in F^{2ab} is activated by 1-to-1 pathways between F_2^{2b} and F^{2ab}. If both ART_{2a} and ART_{2b} are active, then F^{ab} becomes active only if ART_{2a} predicts the same category as ART_{2b} via the weights w_J^{2ab}. The F^{2ab} output vector x^{2ab} are calculated by:

$$x^{ab} = \begin{cases} y^b \wedge w_J^{2ab} & \text{if the Jth } F_2^{2a} \text{ is active and } F_2^{2b} \text{ is active} \\ w_J^{2ab} & \text{if the Jth } F_2^{2a} \text{ is active and } F_2^{2b} \text{ is inactive} \\ y^b & \text{if } F_2^{2a} \text{ is inactive and } F_2^{2b} \text{ is active} \\ 0 & \text{if } F_2^{2a} \text{ is inactive and } F_2^{2b} \text{ is inactive} \end{cases} \tag{3.14}$$

where y^b is the input pattern in ART_{2b} and also the input criteria for the whole approach. By Eq 3.14, $x^{ab} = 0$ if the prediction w_J^{ab} is disconfirmed by y^b.

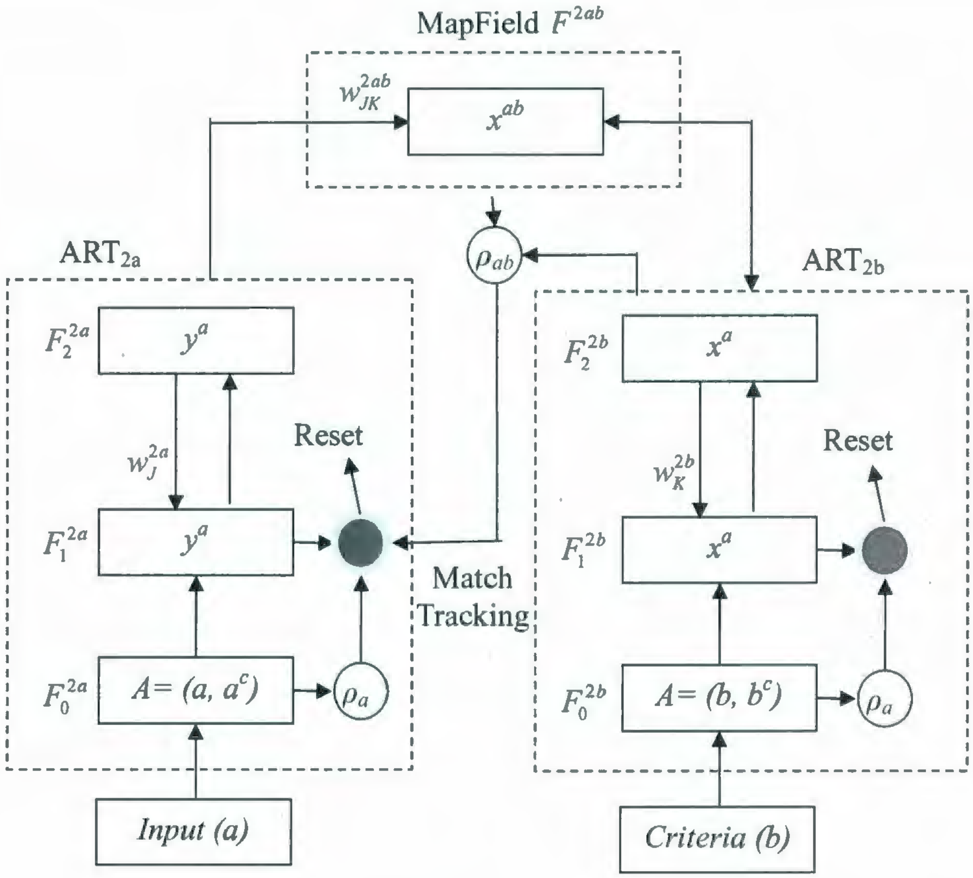


Figure 3.3 Architecture of ART mapping system

3.2.5 Match Tracking

At the start of each input presentation, the vigilance parameter ρ_a of ART_{2a} equals the baseline vigilance which was preset in Eq 3.3 in the initial step. The map field vigilance parameter is ρ_{ab}. If

$$\left|x^{ab}\right| < \rho_{ab}\left|y^{b}\right| \tag{3.15}$$

then ρ_a is increased until it is slightly larger than $\left|A \wedge w_J^{2a}\right| \cdot \left|A\right|^{-1}$, where A is the input to F_1^{2a} in a complement coding form. Consequently we have:

$$\left|x^{a}\right| = \left|A \wedge w_J^{2a}\right| < \rho_a\left|A\right| \tag{3.16}$$

where J is the index of the active F_2^{2a} node as in Eq 3.15. When this occurs, the ART_{2a} search leads to activation of another node J of F_2^{2a} with

$$\left|x^{a}\right| = \left|A \wedge w_J^{a}\right| \geq \rho_a\left|A\right| \tag{3.17}$$

and

$$\left|x^{ab}\right| = \left|y^{b} \wedge w_J^{ab}\right| \geq \rho_{ab}\left|y^{b}\right|, \tag{3.18}$$

or if no such node exists, F_2^{2a} will be ended for the remainder of the input presentation.

3.2.6 Map Field Learning

Learning rules determine how the map field weights w_{Jk}^{2ab} change through time. Weights w_{Jk}^{ab} in F_2^{2a} *to* F^{2ab} paths initially satisfy:

$$w_{jk}^{2ab}(0) = 1 \tag{3.19}$$

During the resonance step with the active ART_{2a} category J, w_{JK}^{2ab} approaches the map field vector x^{ab}. With fast learning, once J learns to predict the ART_{2b} category K, the association will be stored in the LTM not be changed.

The normalized input patterns are fed to the ART_{2a} as I_a and the centroid values are fed to the ART_{2b} as I_b. The I_a are compared with I_b in the classification subsystem screened into the preset target groups.

3.3 Application to Watershed Classification

3.3.1 Overview of the Study Area

In order to test the developed TSAM approach, a case study was conducted in the Deer River watershed in Manitoba were targeted. The Deer River is one of the major tributaries of the Churchill River and its drainage ranges from (-95.5°W, 57°N) to (-94°W, 58.5°N). The Deer River Watershed was delineated into 92 sub-basins based on a 3-arc-second digital elevation model (DEM) obtained from U. S. Geological Survey (USGS) (Figure 3.4). In order to efficiently support watershed modeling and management, these sub-basins need to be classified into certain groups and each group is supposed to present a type of combination of watershed features such as area, elevation, land cover, soul properties, and river channel shapes. In this study, four parameters that reflect the characteristics of the watershed were selected as the input patterns for the classification of the sub-basins. These parameters are area, elevation, along channel length, and along channel slope. Where along channel length is curvilinear distance measurement along the center of the channel and along channel slope is change in elevation divided by the length of channel along a channel distance

The original data for these parameters of 92 sub-basins are shown in Table 3.1.

3.3.2 Classification Process by TSAM

The original input data was first normalized and the results were shown in Table 3.2.

The vigilance parameter ρ for the ART modules was set as 0.7, which was commonly used value for ART learning (Yang and Yang, 2008). The learning rate β was set as 1.0 to enable the fast learning of the system. The choice parameter was set as 0.0001 to ensure that one category was active at one time. The normalized input data was fed to the ART1 for an unsupervised classification and the results are shown in Table 3.3.

Based on the unsupervised classification results and the preset target group number, the centroid value can be determined by the centroid determination module. The number of target groups in this case study was preset as 5, therefore the Class #3, #5, #8, #9, #10 were selected and numbered as Group 1 to 5 respectively. By using Eq 3.13, the centroid values were calculated for the selected groups (Table 3.4).

The centroid values were fed to ART_{2b} as the criteria for a supervised learning, and the normalized data was fed to ART_{2a} as the input for second stage classification. The final classification results were shown in Table 3.5.

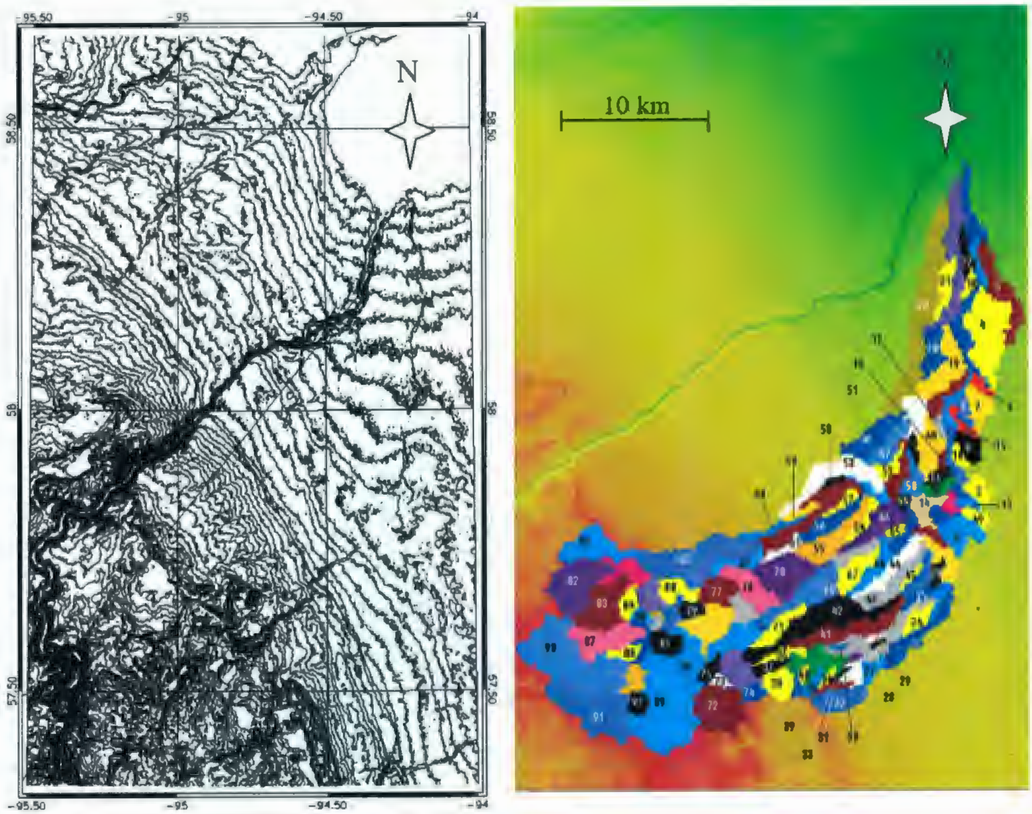


Figure 3.4 Sub-basins in the Deer River watershed

Table 3.1 Original input data for the 92 sub-basins in the Dear River watershed

Sub-basin #	Area (km^2)	Elevation (m)	Along channel length* (km)	Along channel slope** (%)
1	33.76	93	23.22	0.11
2	13.68	82	7.00	0.00
3	30.13	83	8.77	0.00
4	76.03	56	24.70	0.05
5	39.34	46	0.45	0.05
6	6.55	69	25.28	0.11
7	21.56	69	5.91	0.00
8	9.52	71	12.36	0.12
9	16.25	75	7.04	0.00
10	6.68	78	5.37	0.12
11	8.72	80	6.90	0.10
12	5.05	73	2.07	0.10
13	5.80	82	5.27	0.05
14	8.94	82	7.18	0.10
15	6.93	72	6.85	0.00
16	3.67	80	0.43	0.08
17	9.18	82	5.85	0.00
18	6.40	53	5.74	0.06
19	41.3	63	0.80	0.07
20	21.79	64	1.01	0.07
21	17.5	50	13.79	0.09
22	11.28	45	1.09	0.20
23	102.00	39	1.03	0.10
24	12.98	100	2.09	0.10
25	9.38	90	1.23	0.14
26	6.34	89	0.57	0.07
27	10.51	128	1.35	0.28
28	2.11	124	0.54	0.29
29	2.39	125	1.18	0.22
30	2.88	138	4.45	0.15
31	7.45	144	2.20	0.41
32	28.45	140	0.40	0.60
33	2.50	141	1.79	0.69

34	3.82	151	1.67	0.26
35	7.77	151	1.58	0.44
36	11.27	159	0.94	0.22
37	10.24	163	1.70	1.01
38	21.89	166	0.84	0.64
39	4.56	170	0.62	0.27
40	3.48	159	1.40	0.31
41	44.87	104	1.7	0.25
42	74.34	104	1.29	0.25
43	32.20	87	1.99	0.15
44	24.78	87	0.72	0.18
45	3.94	91	1.08	0.19
46	32.5	74	1.49	0.16
47	23.25	74	1.30	0.15
48	27.94	73	1.20	0.11
49	8.42	83	0.72	0.00
50	15.93	80	0.96	0.06
51	5.67	77	1.78	0.11
52	10.12	79	2.36	0.15
53	47.75	87	0.17	0.21
54	11.26	87	0.99	0.21
55	39.73	88	1.76	0.31
56	24.52	89	2.29	0.23
57	7.59	90	1.92	0.20
58	4.20	92	0.59	0.20
59	7.74	105	0.56	0.35
60	11.19	110	1.92	0.24
61	9.3	129	1.31	0.32
62	14.19	151	0.75	0.53
63	90.55	145	1.13	0.51
64	4.85	87	0.73	0.03
65	4.72	87	0.89	0.11
66	32.08	87	1.46	0.13
67	25.03	97	1.39	0.30
68	17.49	95	2.47	0.17
69	8.21	111	0.90	0.38
70	53.26	110	1.18	0.38
71	19.91	143	0.77	0.50
72	56.79	201	1.10	0.53
73	5.03	201	0.74	0.64
74	10.41	196	0.81	0.42

75	10.11	195	0.73	0.07
76	33.12	145	0.97	0.47
77	23.58	166	0.77	0.58
78	53.51	177	0.89	0.43
79	9.98	176	0.90	0.59
80	19.06	182	1.00	0.41
81	43.02	193	0.90	0.38
82	63.25	193	1.68	0.45
83	49.33	193	1.20	0.40
84	11.28	191	1.47	0.49
85	18.00	178	0.92	0.62
86	5.43	176	0.64	0.32
87	44.4	177	0.79	0.71
88	9.39	181	0.69	0.88
89	104.24	185	1.06	0.49
90	120.27	186	0.77	0.45
91	79.31	191	0.82	0.54
92	10.41	192	0.40	0.55

Note:

* Along channel length is curvilinear distance measurement along the center of the channel

** Along channel slope is change in elevation divided by the length of channel along a channel distance

Table 3.2 Normalized input data for the 92 sub-basins in the TSAM approach

Sub-basin #	Area	Elevation	Along channel length	Along channel slope
1	0.2679	0.3333	0.9180	0.1089
2	0.0979	0.2654	0.2720	0.0000
3	0.2371	0.2716	0.3425	0.0000
4	0.6256	0.1049	0.9769	0.0495
5	0.3151	0.0432	0.0112	0.0495
6	0.0376	0.1852	1.0000	0.1089
7	0.1646	0.1852	0.2286	0.0000
8	0.0627	0.1975	0.4855	0.1188
9	0.1197	0.2222	0.2736	0.0000
10	0.0387	0.2407	0.2071	0.1188
11	0.0559	0.2531	0.2680	0.0990
12	0.0249	0.2099	0.0757	0.0990
13	0.0312	0.2654	0.2031	0.0495
14	0.0578	0.2654	0.2792	0.0990
15	0.0408	0.2037	0.2660	0.0000
16	0.0132	0.2531	0.0104	0.0792
17	0.0598	0.2654	0.2262	0.0000
18	0.0363	0.0864	0.2218	0.0594
19	0.3317	0.1481	0.0251	0.0693
20	0.1666	0.1543	0.0335	0.0693
21	0.1302	0.0679	0.5424	0.0891
22	0.0776	0.0370	0.0366	0.1980
23	0.8454	0.0000	0.0342	0.0990
24	0.0920	0.3765	0.0765	0.0990
25	0.0615	0.3148	0.0422	0.1386
26	0.0358	0.3086	0.0159	0.0693
27	0.0711	0.5494	0.0470	0.2772
28	0.0000	0.5247	0.0147	0.2871
29	0.0024	0.5309	0.0402	0.2178
30	0.0065	0.6111	0.1705	0.1485
31	0.0452	0.6481	0.0808	0.4059
32	0.2229	0.6235	0.0092	0.5941
33	0.0033	0.6296	0.0645	0.6832
34	0.0145	0.6914	0.0597	0.2574
35	0.0479	0.6914	0.0562	0.4356
36	0.0775	0.7407	0.0307	0.2178
37	0.0688	0.7654	0.0609	1.0000

38	0.1674	0.7840	0.0267	0.6337
39	0.0207	0.8086	0.0179	0.2673
40	0.0116	0.7407	0.0490	0.3069
41	0.3619	0.4012	0.0609	0.2475
42	0.6113	0.4012	0.0446	0.2475
43	0.2547	0.2963	0.0725	0.1485
44	0.1919	0.2963	0.0219	0.1782
45	0.0155	0.3210	0.0362	0.1881
46	0.2572	0.2160	0.0526	0.1584
47	0.1789	0.2160	0.0450	0.1485
48	0.2186	0.2099	0.0410	0.1089
49	0.0534	0.2716	0.0219	0.0000
50	0.1170	0.2531	0.0315	0.0594
51	0.0301	0.2346	0.0641	0.1089
52	0.0678	0.2469	0.0872	0.1485
53	0.3863	0.2963	0.0000	0.2079
54	0.0774	0.2963	0.0327	0.2079
55	0.3184	0.3025	0.0633	0.3069
56	0.1897	0.3086	0.0844	0.2277
57	0.0464	0.3148	0.0697	0.1980
58	0.0177	0.3272	0.0167	0.1980
59	0.0476	0.4074	0.0155	0.3465
60	0.0768	0.4383	0.0697	0.2376
61	0.0608	0.5556	0.0454	0.3168
62	0.1022	0.6914	0.0231	0.5248
63	0.7485	0.6543	0.0382	0.5050
64	0.0232	0.2963	0.0223	0.0297
65	0.0221	0.2963	0.0287	0.1089
66	0.2536	0.2963	0.0514	0.1287
67	0.1940	0.3580	0.0486	0.2970
68	0.1302	0.3457	0.0916	0.1683
69	0.0516	0.4444	0.0291	0.3762
70	0.4329	0.4383	0.0402	0.3762
71	0.1506	0.6420	0.0239	0.4950
72	0.4628	1.0000	0.0370	0.5248
73	0.0247	1.0000	0.0227	0.6337
74	0.0702	0.9691	0.0255	0.4158
75	0.0677	0.9630	0.0223	0.0693
76	0.2624	0.6543	0.0319	0.4653
77	0.1817	0.7840	0.0239	0.5743
78	0.4350	0.8519	0.0287	0.4257

79	0.0666	0.8457	0.0291	0.5842
80	0.1434	0.8827	0.0331	0.4059
81	0.3462	0.9506	0.0291	0.3762
82	0.5174	0.9506	0.0601	0.4455
83	0.3996	0.9506	0.0410	0.3960
84	0.0776	0.9383	0.0518	0.4851
85	0.1345	0.8580	0.0299	0.6139
86	0.0281	0.8457	0.0187	0.3168
87	0.3579	0.8519	0.0247	0.7030
88	0.0616	0.8765	0.0207	0.8713
89	0.8643	0.9012	0.0354	0.4851
90	1.0000	0.9074	0.0239	0.4455
91	0.6534	0.9383	0.0259	0.5347
92	0.0702	0.9444	0.0092	0.5446

Table 3.3 Unsupervised classification results by the TSAM

Class #	Sub-basin #
1	1, 2, 3, 7, 9, 15, 17
2	4, 6
3	5, 8, 10, 11, 12, 13, 14, 16, 18, 19, 20, 21, 48, 65
4	22, 23,
5	24, 25, 27, 28 29, 30, 31, 33, 45, 57, 58, 59, 60, 61, 69
6	32, 34, 36, 37, 40
7	35, 38, 39, 62, 71, 76, 77
8	26, 41, 42, 43,44, 46, 47, 49, 50, 51, 52, 53, 54, 55, 56, 64, 66, 67, 68
9	63, 70, 72, 78, 81, 82, 83, 91
10	73, 74, 75, 79, 80, 84, 85, 86, 87, 92
11	88
12	89, 90

Table 3.4 Centroid values for the selected groups

Group	Area	Elevation	Along channel length	Along channel slope
1	0.49947	0.84182	0.03753	0.44802
2	0.19645	0.29695	0.04802	0.16258
3	0.1075	0.19224	0.17375	0.08699
4	0.1041	0.90988	0.02668	0.47723
5	0.03990	0.46626	0.05459	0.27459

Table 3.5 Final classification results by the TSAM

Group #	Sub-basin #
1	63, 72, 78, 81, 82, 83, 87, 89, 90, 91
2	1, 5, 19, 22, 23, 24, 25, 26, 41, 42, 43, 44, 45, 46, 47, 48, 49, 52, 53, 54, 55, 56, 57, 58, 64, 65, 66, 67, 68
3	2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 50, 51
4	32, 35, 37, 38, 39, 62, 71, 73, 74, 75, 76, 77, 79, 80, 84, 85, 86, 88, 92
5	27, 28, 29, 30, 31, 33, 34, 36, 40, 59, 60, 61, 69, 70
No group	4

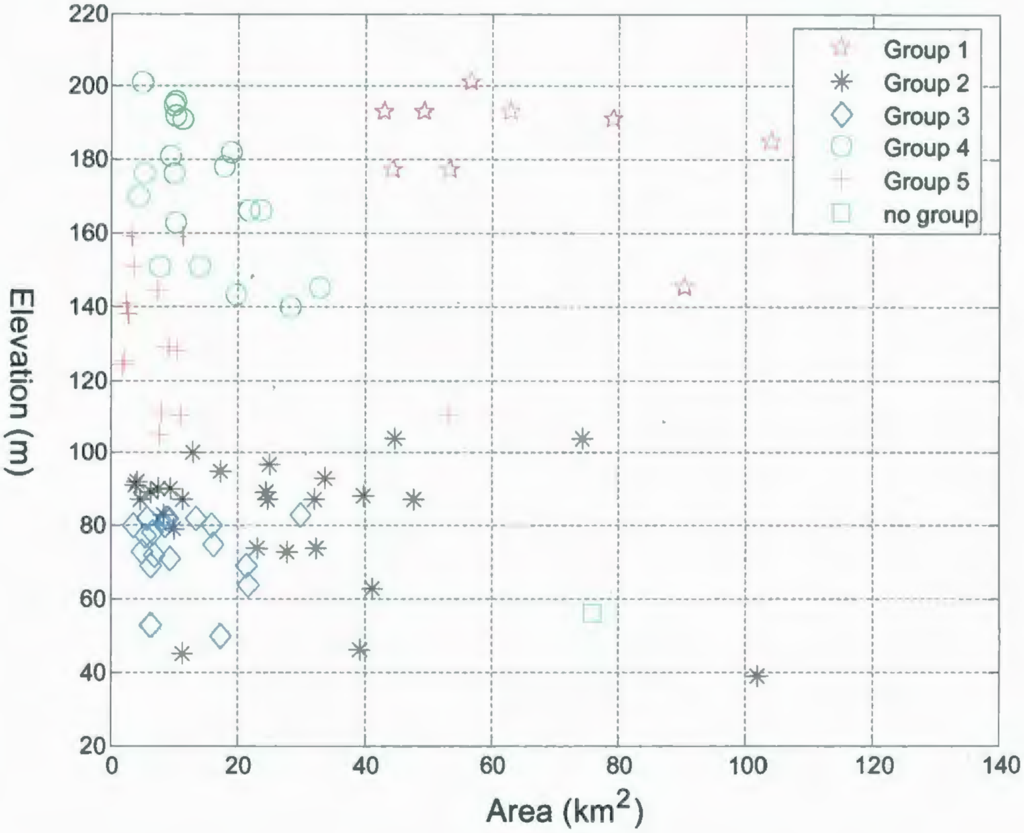


Figure 3.5 Distribution of Area vs. Elevation in the final classification results by TSAM

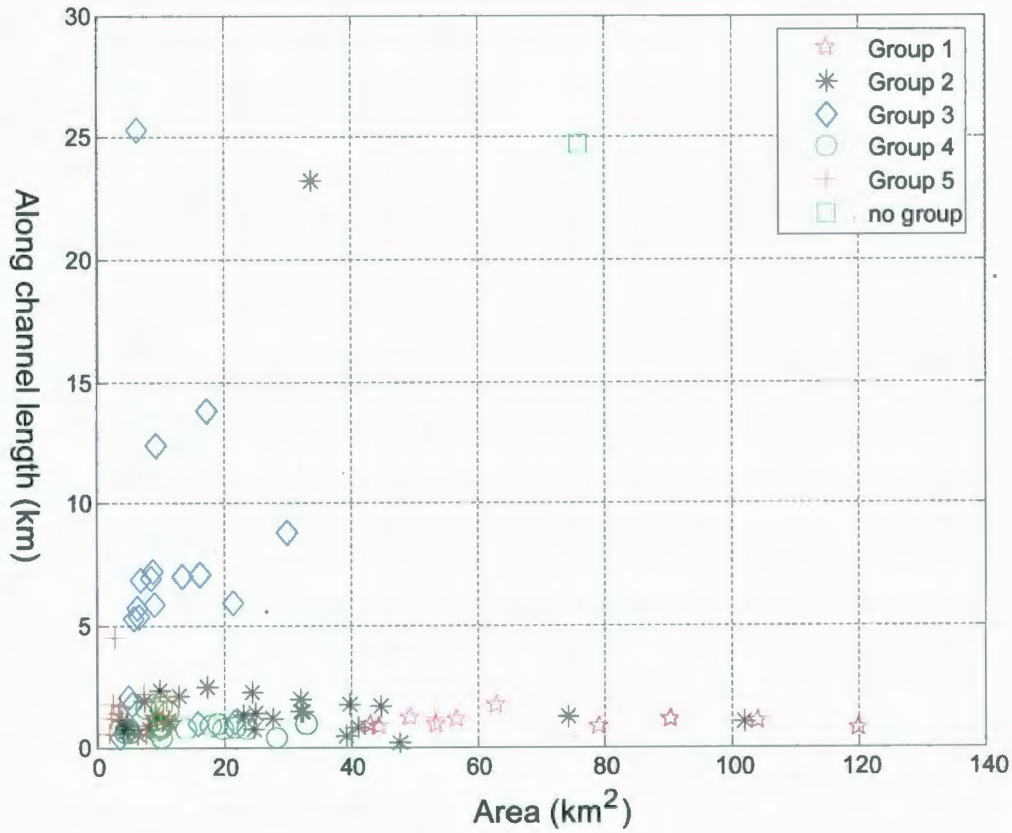


Figure 3.6 Distribution of Area vs. Along channel length in the final classification results by TSAM

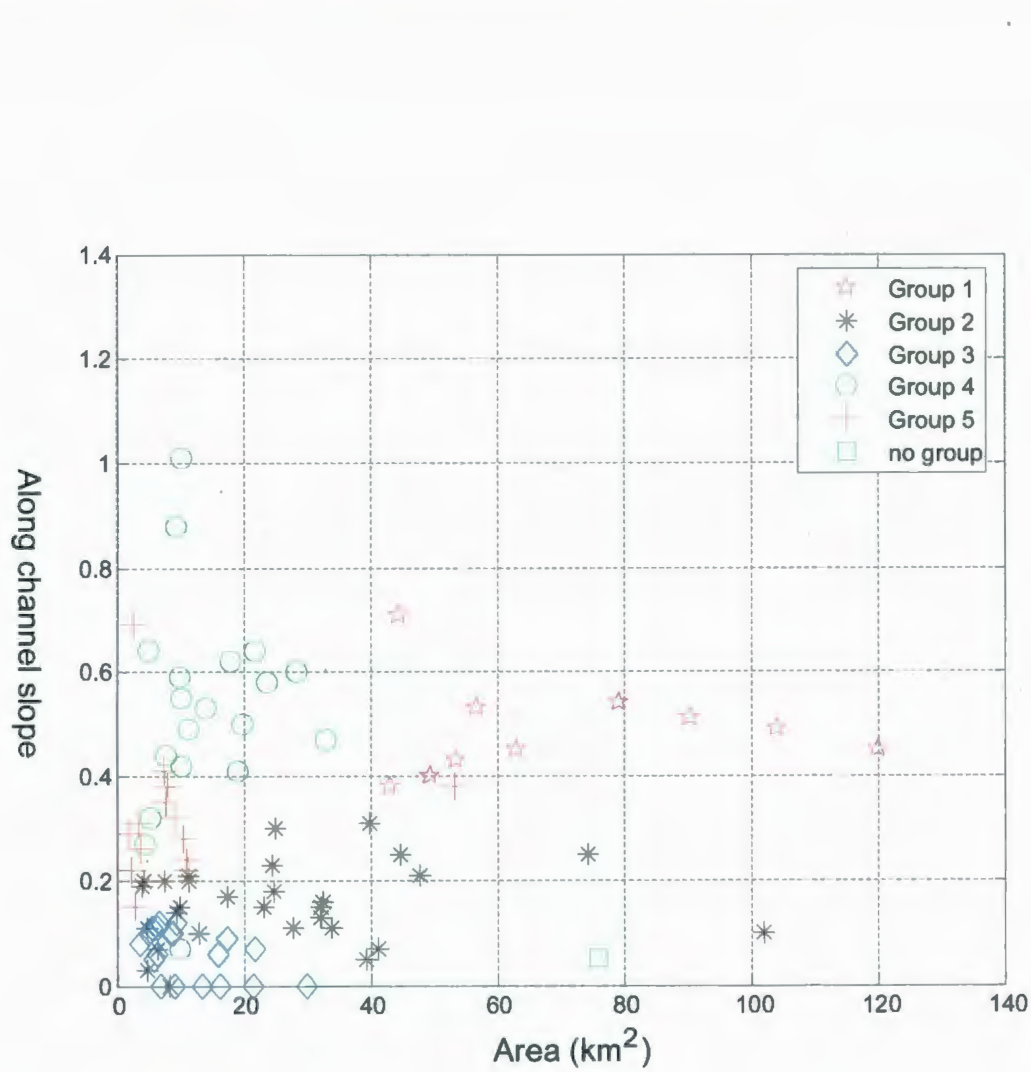


Figure 3.7 Distribution of Area vs. Along channel slope in the final classification results by TSAM

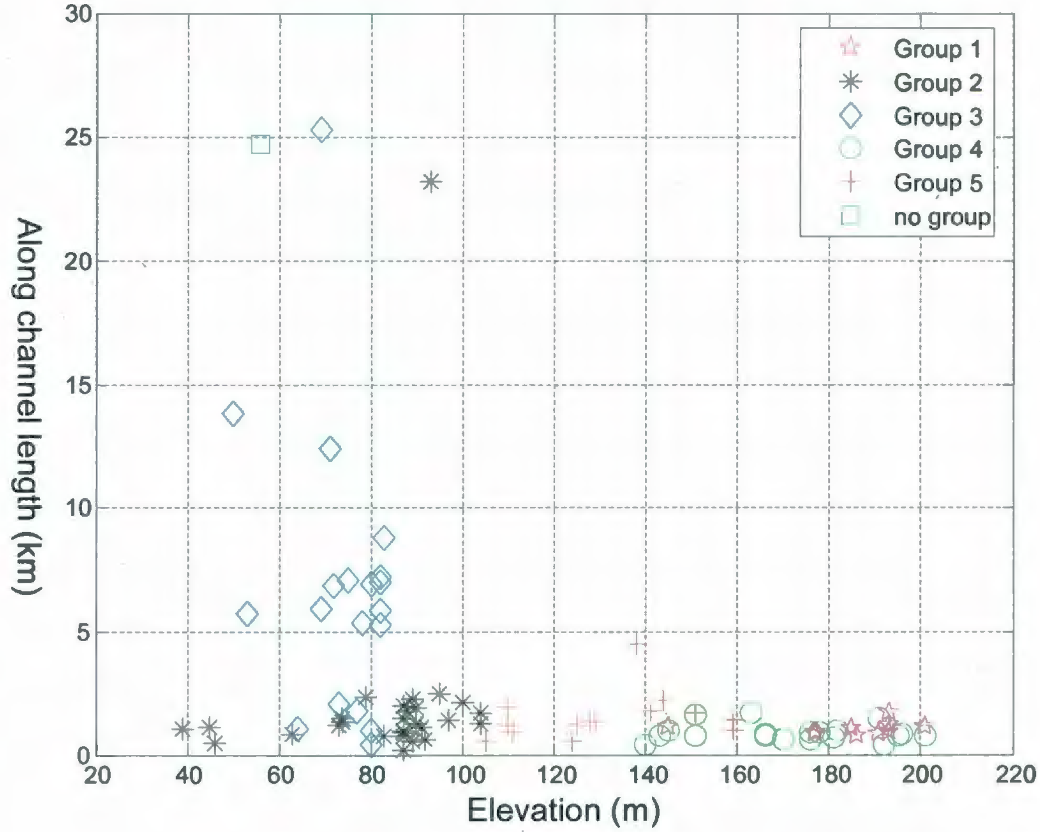


Figure 3.8 Distribution of Elevation vs. Along channel length the final classification results by TSAM

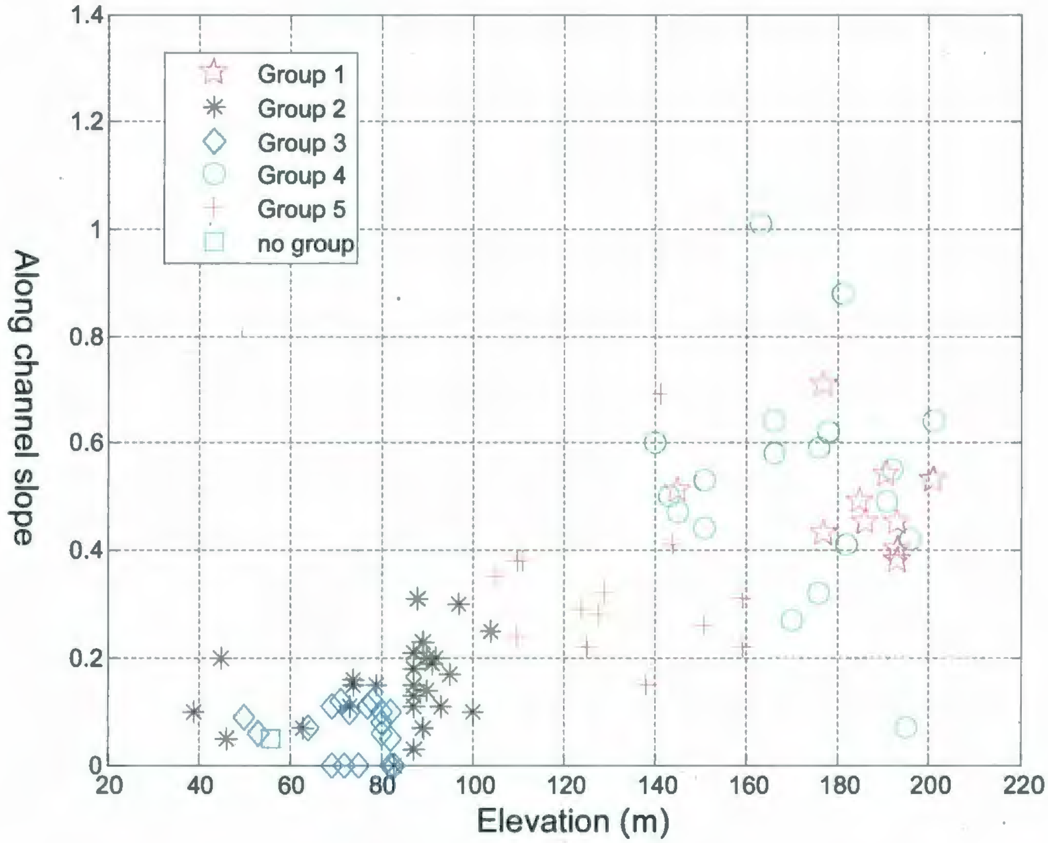


Figure 3.9 Distribution of Elevation vs. Along channel slope in the final classification results by TSAM

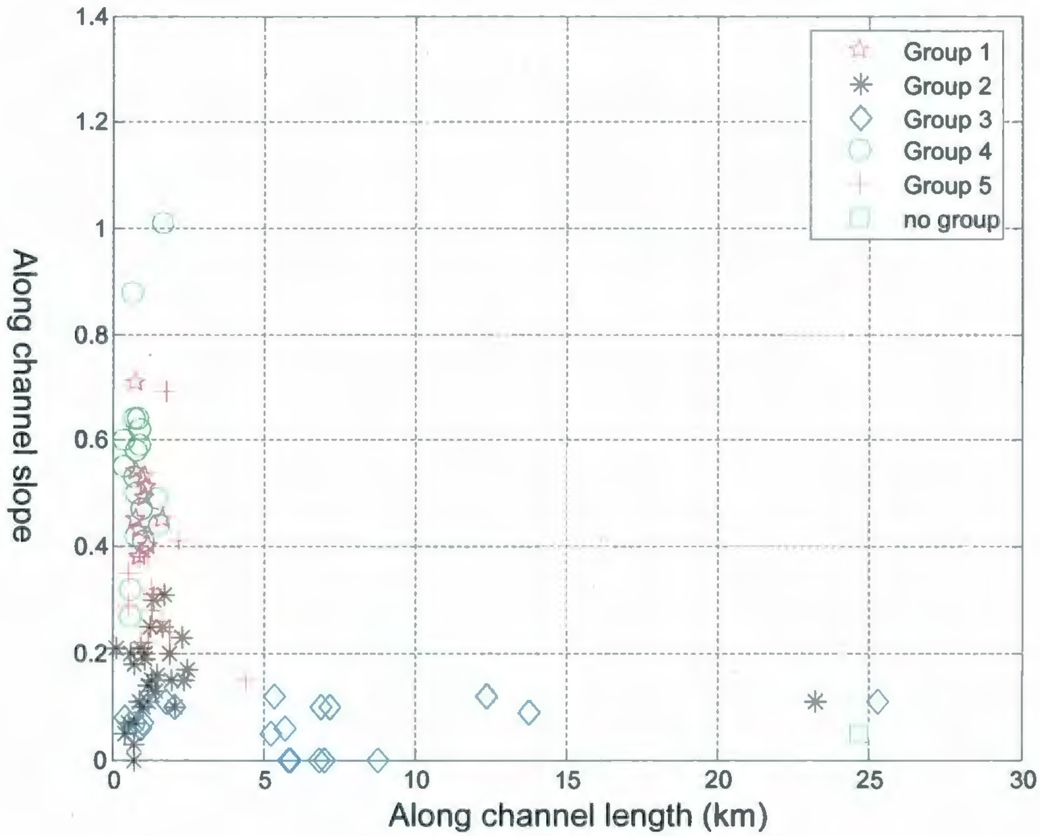


Figure 3.10 Distribution of Along channel length vs. Along channel slope in the final classification results by TSAM

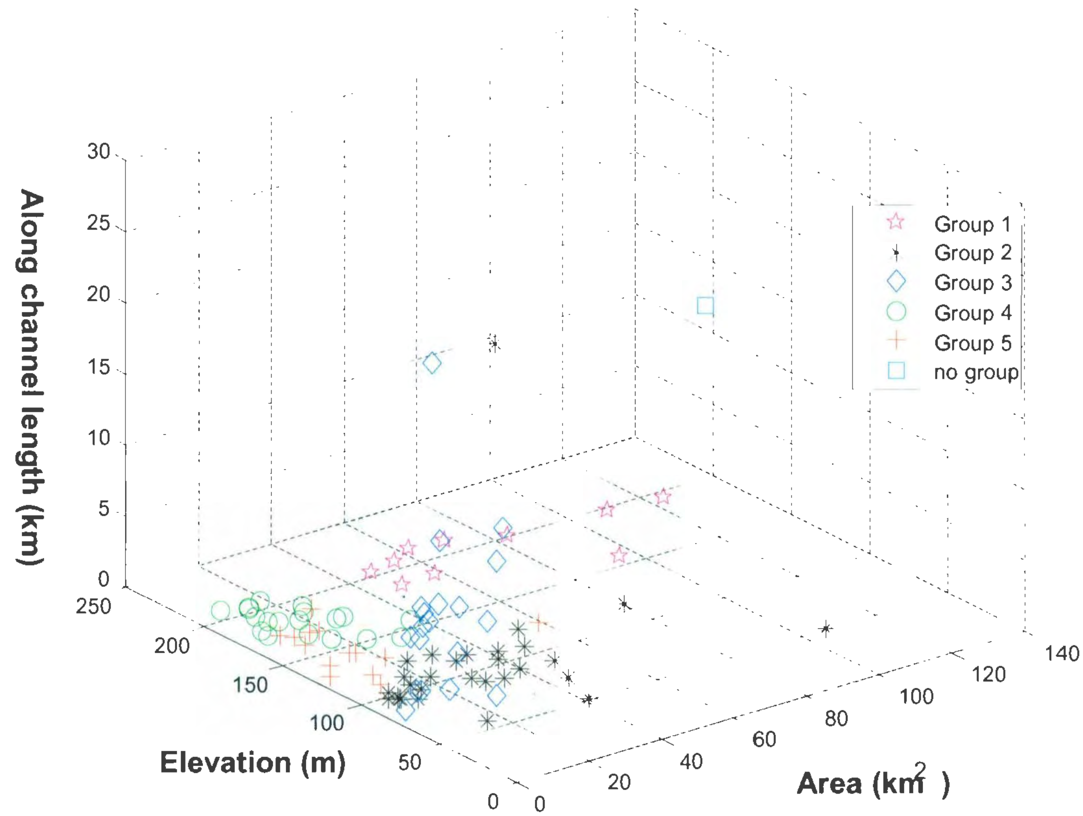


Figure 3.11 Distribution of Area vs. Elevation vs. Along channel length in the final classification results by TSAM

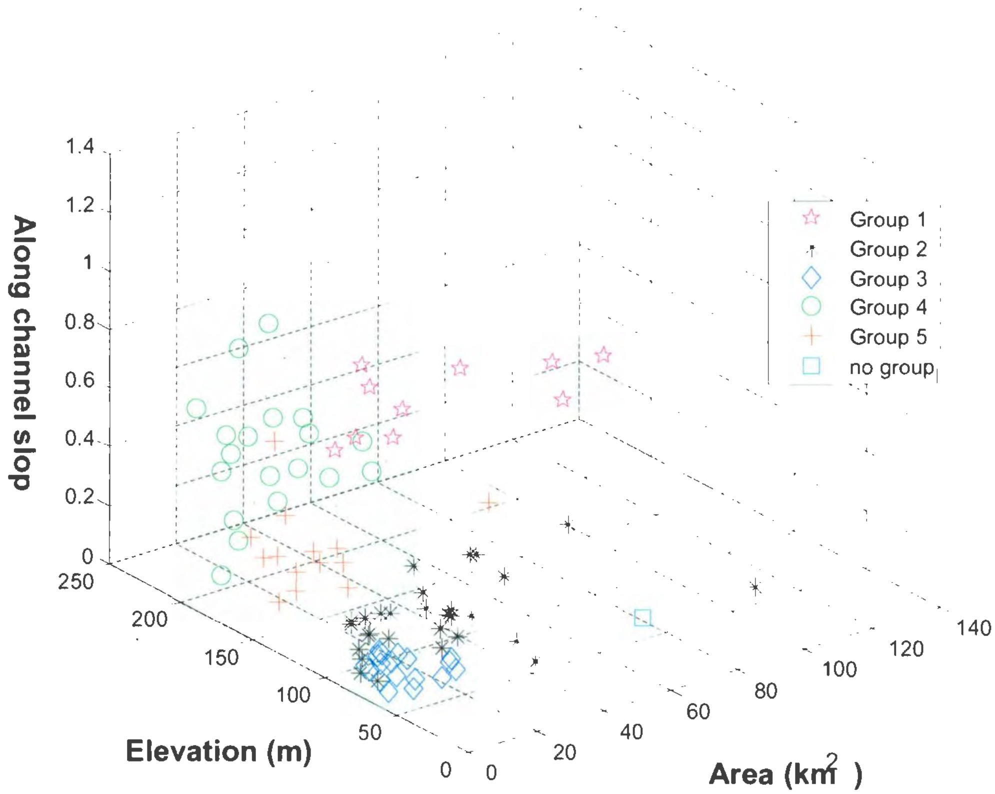


Figure 3.12 Distribution of Area vs. Elevation vs. Along channel slope in the final classification results by TSAM

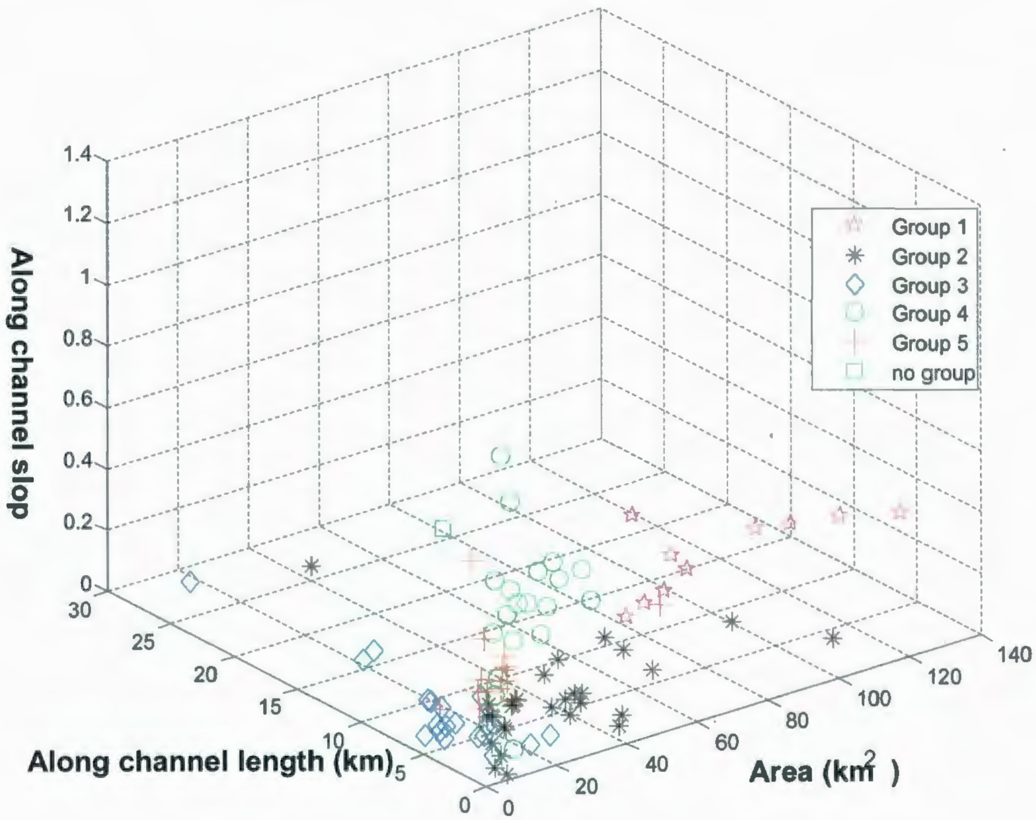


Figure 3.13 Distribution of Area vs. Along channel length vs. Along channel slope in the final classification results by TSAM

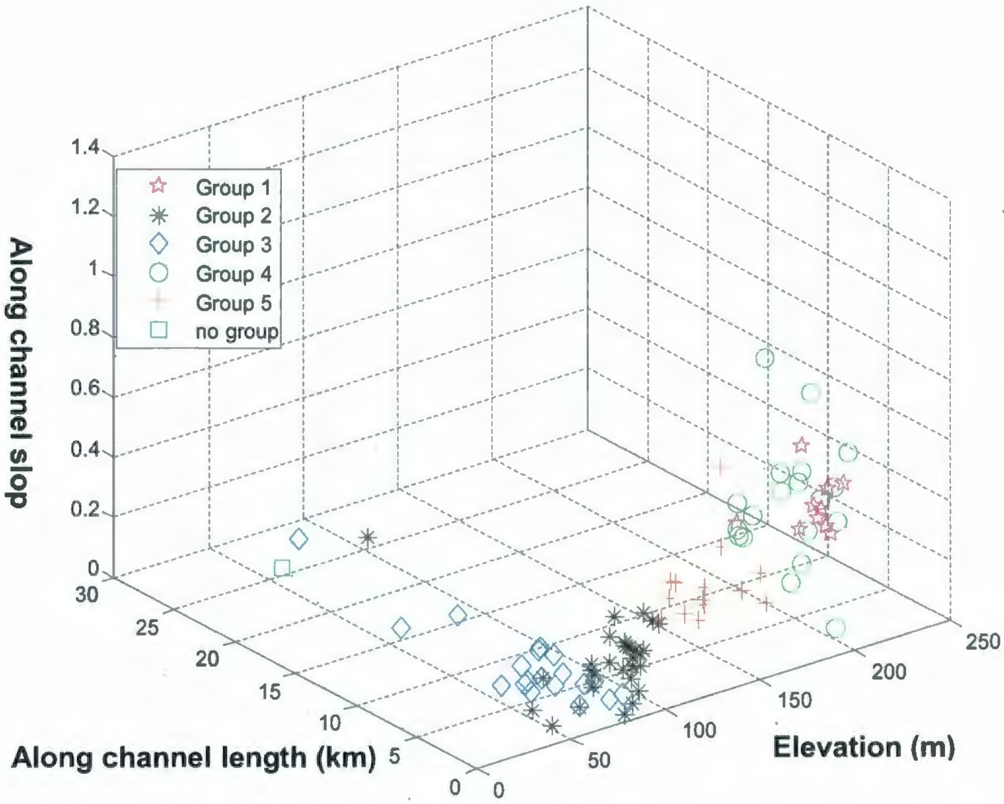


Figure 3.14 Distribution of Elevation vs. Along channel length vs. Along channel slope in the final classification results by TSAM

3.3.3 Result and Discussion

The classification result showed that 91 sub-basins can be properly classified into 5 preset target groups in the case of vigilance ρ equals to 0.7. However, sub-basin 4 cannot be classified into any group by the TSAM approach.

By using Eq 3.13, the new centroid values were obtained from the final classified groups. The old and new centroid values are compared in Table 3.6. It is observed that the new centroid values are very close ones, which means the representative characters of each group were unchanged through the final classification.

The distributions of different features in the classification results were shown in Figures 3.5 to 3.14. Figure 3.5 indicated that the double features of area and elevation had significant contribution to the classification results. Figure 3.6 indicated that the double features of area and along channel length only had significant contribution to classify Group 3. Figure 3.7 indicated that the double features of area and along channel slope had contribution to the classification results but less than the couple features of area and elevation. Figure 3.8 indicated that the double features of elevation and along channel length only had little contribution to the classification. Figure 3.9 indicated that the couple features of elevation and along channel slope had significant contribution to classify Group 2, 3 and 5, and no contribution to the others. Figure 3.10 indicated that the couple features of along channel length and along channel slope almost had no contribution to the classification. Figures 3.11 to 3.14 indicated that almost all the triple features had signification contributions to the classification.

The final classification results revealed that the sub-basins in Group 1 have the common features of large area, high elevation, short along channel length, and small along channel slope. These sub-basins mainly locate in the upstream of the river. The ratios of sub-basin length to width in this group are close to 1 which indicates a round shape of the sub-basins.

The sub-basins in Group 2 have the common features of low to medium area, low to medium elevation, very short along channel length, and low to medium along channel slope. These sub-basins mainly locate in the middle- and down-stream of the river. The ratios of sub-basin length to width in this group are much large than 1, which indicates the most sub-basins have an oblate shape and relatively short runoff concentration time.

The sub-basins in Group 3 have the common features of very small area, medium to low elevation, very short along channel length, and very low along channel slope. These sub-basins mainly locate in the downstream of the river. The ratios of sub-basin length to width in this group are close 1 which means the shape of the sub-basins in this group nearly likes a circle.

The sub-basins in Group 4 have the common features of small area, very high elevation, very short along channel length, and high along channel slope. These sub-basins mainly locate in the upstream of the river. The length and width ratios of most the sub-basins are close to 1.

The sub-basins in Group 5 have the common features of very small area, medium to high elevation, very short along channel length, and low to medium along channel slope. The sub-basins in this group locate in the midstream of the river. These length and width ratios of most of the sub-basins are very large which indicates an oblate shape of most of the sub-basins.

The no-group sub-basin 4 has an area of 76.03 km^2, elevation of 56 m, along channel length of 24.7 km, and along channel slope of 0.05%.the corresponding normalized data are 0.6256, 0.1049, 0.9769, and 0.0495. It respectively means that this sub-basin has features of medium to high area, low elevation, very long along channel length, and very small along channel slope. Because the characteristics of sub-basin 4 are quite different from the 5 groups, it could not be classified into any group. Viewed from Figure 3.2 it can be seen that the shape and area of sub-basin #4 is similar as #89 and 90, 91, but sub-basins #4 could not be classified and the others were classified into Group 1. This is because the elevation, along channel length, and along channel slope are quite different from sub-basin #4 and the other three ones. Furthermore, sub-basin 4 is located in the downstream of the river, where most of the sub-basins are classified into Group 3, with the features of very small area, medium to low elevation, very small along channel length, and very low along channel slope.

Table 3.6 New centroid values for final groups as well as old centroid values

Group	Area		Elevation		Along channel length		Along channel slope	
	Old	New	Old	New	Old	New	Old	New
1	0.4994	0.5785	0.8418	0.8956	0.03753	0.0344	0.44802	0.48416
2	0.1964	0.2033	0.2969	0.2716	0.04802	0.0742	0.16258	0.15466
3	0.1075	0.0801	0.1922	0.2147	0.17375	0.2647	0.08698	0.06096
4	0.1041	0.1036	0.9098	0.8193	0.02668	0.0282	0.47723	0.52267
5	0.0399	0.0644	0.4662	0.5679	0.05458	0.0540	0.27459	0.31825

3.4 Summary

This chapter presents a modified ART mapping approach (TSAM) by integrating three ART modules into the system classification process in two stages to form an unsupervised learning module for cluster centroid calculation and a supervised learning module for normalized input classification.

The ART unsupervised approach can provide an accurate classification result, but the number of final output groups cannot be controlled. The group number can be controlled by ARTMap supervised approach; however, it requires criteria for supervised learning. In traditional methods the criteria are usually obtained from literature or trough questionnaire survey, which could be inefficient and lead to errors in results. The TSAM approach can help solve these problems by using ART unsupervised classification and centriod determination modules in the first stage to generate criteria for the ARTMap supervised classification in the second stage. By this way, the approach can efficiently and accurately handle a complex classification problem, like the one in the case study.

The real world case study demonstrated that the TSAM approach has the ability in handing such problems and supports watershed modeling and management.

CHAPTER 4: INTEGRATED RULE-BASED FUZZY ADAPTIVE RESONANCE MAPPING (IRFAM) APPORACH

4.1 Background

Uncertainty and complexity are two major issues in watershed classification. Many studies have been conducted watershed classification, however many of which are lack of consideration in the conditions of uncertainty and complexity coexisting. IRFAM is developed aiming to achieve an efficient and reliable approach of watershed classification to deal with complex and uncertain features. In order to handle this problem, Fuzzy set theory which have high ability to deal with the uncertainty and the ART neural network which can efficiently handle complexity are incorporated into the IRFAM.

As shown in Figure 4.1, the IRFAM approach includes three subsystems: 1) centroid determination for locating the centroids of the expected target groups by unsupervised ART; 2) criteria combination subsystem for generating the fuzzy criteria combinations; and 3) classification subsystem for classifying the original inputs which have been converted into fuzzy set form. There are five ART modules integrated in the IRFAM approach as follows: ART_1 is used for processing unsupervised classification for the fuzzified inputs; ART_{2a} and ART_{2b} are used in an ART Mapping module for screening the criteria combinations into the preset target groups; ART_{3a} and ART_{3b} are used in an ART Mapping module for classification based on comparison of the criteria combined with the fuzzified inputs.

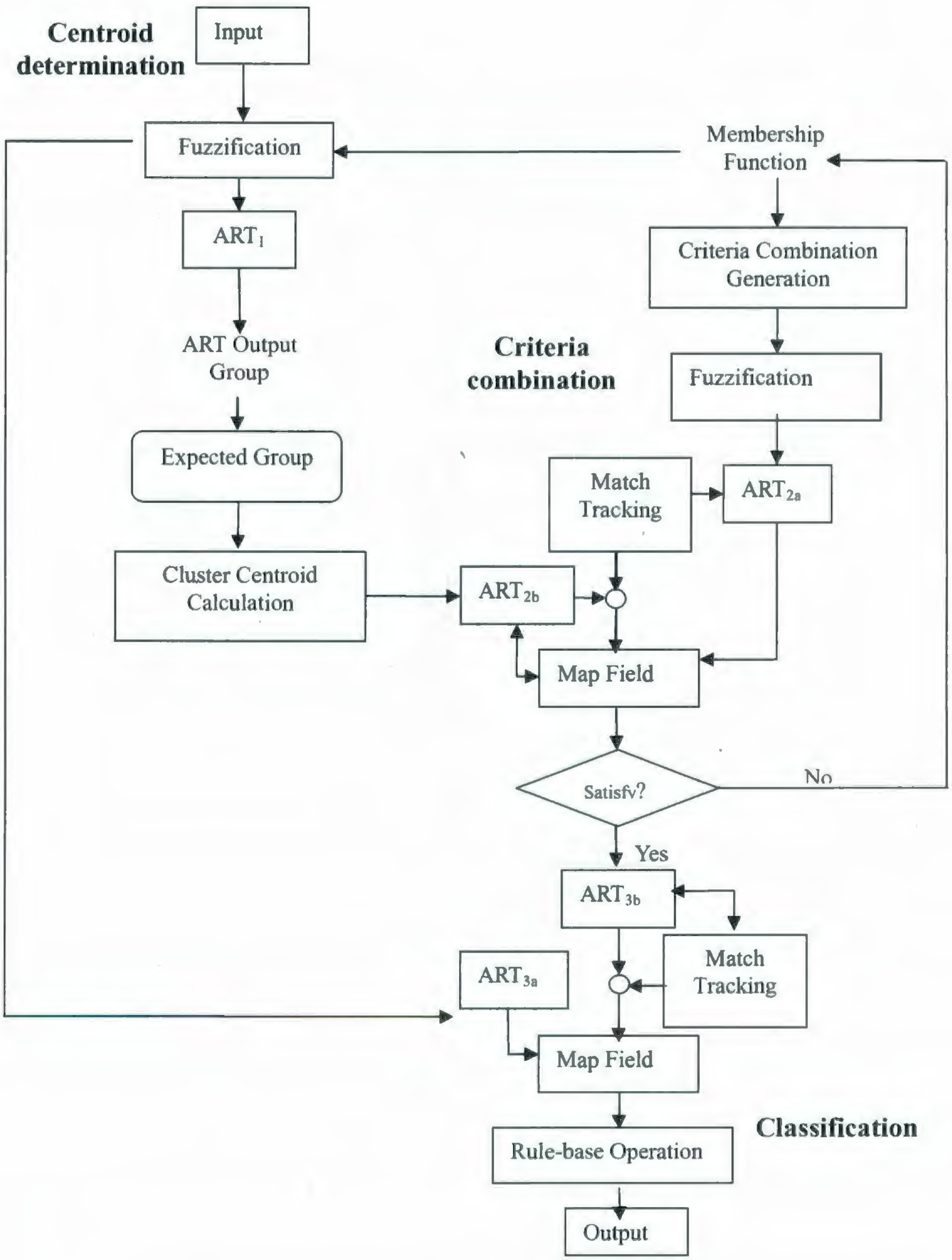


Figure 4.1 Flowchart of the Integrated Rule-base Fuzzy ART Mapping (IRFAM) System

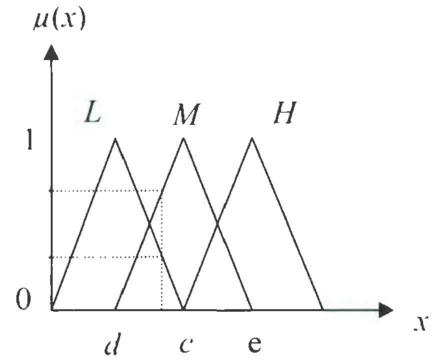


Figure 4.2 Triangular membership function

Note:
L for low level
M for medium level
H for high level

4.2 Methodology

A fuzzy set is a class of objects with continuous grades of membership which represents the degree of truth as an extension of valuation (Zadeh, 1965). Fuzzy sets generalize classical sets, the indicator functions of classical sets are special cases of the membership functions of fuzzy sets for the latter only take values 0 or 1.

A fuzzy set operation is an operation on fuzzy sets, which are generalization of crisp set operations. The most widely used operations are called standard fuzzy set operations, which include unions, complements, and intersections (Dubois and Prade 1988).

The membership function of the Union of two fuzzy sets A and B with membership functions μ_A and μ_B respectively is defined as the maximum of the two individual membership functions as shown in the follows. This is called the maximum criterion. The Union operation in Fuzzy set theory is the equivalent of the OR operation in Boolean algebra.

$$\mu_{A\cup B} = \max(\mu_A, \mu_B)$$

The membership function of the Intersection of two fuzzy sets A and B with membership functions μ_A and μ_B respectively is defined as the minimum of the two individual membership functions as shown in the follows. This is called the minimum criterion. The Intersection operation in Fuzzy set theory is the equivalent of the AND operation in Boolean algebra.

$$\mu_{A\cap B} = \min(\mu_A, \mu_B)$$

The membership function of the Complement of a Fuzzy set A with membership function μ_A is defined as the negation of the specified membership function as shown in the follows. This is called the negation criterion. The Complement operation in Fuzzy set theory is the equivalent of the NOT operation in Boolean algebra.

$$\mu_{\bar{A}} = 1 - \mu_A$$

4.2.1 Fuzzy Membership Function

Let X be a set of data points, with series of data points of x. therefore, $X=\{x\}$. A fuzzy set Y in X is characterized by a membership function $\mu(x)$. It can be used to describe the mean in measuring the degree of compatibility of a data value to a fuzzy set, or to describe the probability that this data value belongs to a fuzzy set Y in the interval [0, 1]. The $\mu(x)$ value at x indicates the grade of membership of x in Y. Therefore, the closer the value of $\mu(x)$ to 1, the higher the grade of membership of x in Y is appeared (Zadeh et al., 1968). The normal used membership functions are the triangular function, the trapezoidal function, and the bell shape function.

4.2.2 Fuzzification

A triangle membership function for x is given by:

$$\mu_i(x) = \begin{cases} \dfrac{x-d}{c-d} & if\ d \le x \le c \\ \dfrac{e-x}{e-c} & if\ c \le x \le e \\ 0 & otherwhiles \end{cases} \tag{4.1}$$

Where i is the level of membership functions which is the grade of difference for the feature; d is the lower bound of the *ith* level membership function; e is the upper bound of the *ith* level membership function; and c is the point where $\mu_i(x) = 1$ (Fig. 4.2).

The original input, I_{a0} is presented as follows:

$$I_{a_0} := (x_{kj})_{k=1,2,\cdots,n;\ j=1,2,\cdots,m} \tag{4.2}$$

where k is the number of samples, and j is the number of features in each sample.

Based on the membership function, the fuzzy set Y is given by the following:

$$Y := (\mu_i(x))_{i=1,2,\cdots,p} \tag{4.3}$$

where $\mu_i(x)$ is the *ith* level membership function

Then the original input I_{a0} is fuzzified as follows:

$$I_a := (\mu_i(x_{kj}))_{i=1,2,\cdots,m;k=1,2,\cdots,o;j=1,2,\cdots,n} \tag{4.4}$$

4.2.3 ART Systems

The fuzzified input I_a is feed to the ART_1 in the centroid determination subsystem. Based on the ART unsupervised learning (Eqs 3.2 to 3.11) the input patterns in I_a are automatically classified into certain groups.

4.2.4 Centroids Determination

After the input patterns are classified by ART_1, the centroids are going to be located based on the expected target groups by the operation of the centroids locating module (Eqs 3.12 and 3.13). The outputs of centroids are going to be used as the criteria to classify the criteria combinations from the fuzzy criteria combination subsystem.

4.2.5 Fuzzy Criteria Combination

The criteria combination is the combination of y_{ij} which has the membership function $\mu\ (y_{ij}) = 1$, where i is the level of membership function and j is the number of the feature. If there are m features with p levels of membership function, the criteria combinations will be in the number of p^m.

$$I_{b_0} := \left(y_j (y_i)_{i=1,2,\ldots,p}\right)_{j=1,2,\ldots,m} \tag{4.5}$$

After being operated by the fuzzification module, the criteria combination I_{b0} is converted to:

$$I_b := \left(\mu(y_j)(\mu(y_i))_{i=1,\ldots,p}\right)_{j=1,\ldots,m} \tag{4.6}$$

For example, for a series of input patterns with 2 parameters in each pattern e.g., area and elevation, and 3 levels for each parameter e.g., low, medium, and high, the criteria combination I_b will be:

$$I_b = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

I_b in Eq 4.6 and C in Eq 3.12 are used as inputs for the ART_{2a} and the ART_{2b} modules. Each input criteria combination in I_{b0} is compared with each centroid pattern by the operation of the ART Mapping model. Finally, the criteria combinations are classified into certain groups, the number of which normally is the same as the one of target groups. ART_{2a} and ART_{2b} are linked together by introducing an inter-ART module F^{ab} which is called the map field. Refer to section 3.2.4, this map field is applied to predict associations in categories and to achieve the match-tracking rule, whereby the vigilance parameter of ART_{2a} increases in response to a predictive mismatch at ART_{2b}. A loop achievement of the match-tracking rule that is used for local real-time processing is provided in (Carpenter et al., 1992):

Inputs to ART_{2a} are in the complement code from:

$$I_{2a} = A = (I_a, I_a^c) \tag{4.7}$$

Inputs to ART_{2b} are in the complement code from:

$$I_{2b} = B = (I_b, I_b^c) \tag{4.7}$$

4.2.6 Rule-Based Theory

A set of fuzzy if-then rules are used in the form of: *if a set of conditions can be satisfied, then a relative set of consequences can be determined.* The if-then rule is applied after the matching of input patterns and criteria combinations:

$$Rule\ R_r : if\ I_{as} \in y_t,\ and\ y_t \in G_r,\ then\ I_{as} \in G_r \tag{4.8}$$

where I_{as} is the sth input pattern; y_t is the tth criteria combination; and G_r is the rth group.

4.2.7 Classification

The classification subsystem consists of two modules: the mapping module which consists of ART_{3a}, ART_{3b} and the map field, and the rule-base operation. The mapping module in this subsystem is almost the same as the one in the centroid determination subsystem. The only difference between them is the vigilance. The vigilance for classification is higher than the one for centroid determination.

I_a and I_b are used as inputs for ART_{3a} and ART_{3b} modules. Each input pattern in I_a is compared with existing criteria combinations in I_b and is associated with the criteria combination that has the best match with it. By using the rule-based operation, the input patterns are classified into the group set which is preset by the if-then rule.

4.3 Application to watershed classification

In order to test the develop IRFAM classification approach, the Deer River watershed in Manitoba was targeted for a real-world case study. The watershed was first delineated into 92 sub-basins based on DEM and hydrological characteristics by Rivertools® (Figure 3.4).

In order to efficiently support watershed modeling and management, these sub-basins need to be classified into certain groups and each group is supposed to present a type of combination of watershed features such as area, elevation, land cover, soul properties, and river channel shapes. In this study, four parameters that reflect the characteristics of the watershed were selected as the input patterns for the classification of the sub-basins. These parameters are area, elevation, along channel length, and along channel slope. The original data for these parameters of 92 sub-basins are shown in Table 3.1.

4.3.1 IRFAM Application

Based on the historical information and the distributions of the features, the membership functions for the features are given in figures 4.3 to 4.6. Triangle membership Three levels of membership were set for each feature: Low (L), Medium (M), and high (H).

Figure 4.3 shows the membership function for the feature of "area", based on Eq 4.1:

$$\mu_L(area) = \begin{cases} 1 & if\ area \leq 5 \\ \dfrac{20 - area}{20 - 5} & if\ 5 \leq area \leq 20 \\ 0 & if\ 20 \leq area \end{cases} ;$$

$$\mu_{\mathrm{M}}(area)=\begin{cases}\dfrac{area-5}{20-5} & if\ 5\le area\le 20\\ \dfrac{40-area}{40-20} & if\ 20\le area\le 40\\ 0 & otherwise\end{cases};$$

$$\mu_{\mathrm{H}}(area)=\begin{cases}0 & if\ 20\le area\\ \dfrac{area-20}{40-20} & if\ 20\le area\le 40\\ 1 & if\ 40\le area\end{cases}.$$

Figure 4.4 shows the membership function for the feature of "elevation", based on Eq 4.1:

$$\mu_{\mathrm{L}}(elevation)=\begin{cases}1 & if\ elevation\le 70\\ \dfrac{120-elevation}{120-70} & if\ 70\le elevation\le 120\\ 0 & if\ 120\le elevation\end{cases};$$

$$\mu_{\mathrm{M}}(elevation)=\begin{cases}\dfrac{elevation-70}{120-70} & if\ 70\le elevation\le 120\\ \dfrac{170-elevation}{170-120} & if\ 120\le elevation\le 170\\ 0 & otherwise\end{cases}$$

$$\mu_{\mathrm{H}}(elevation)=\begin{cases}0 & if\ 120\le elevation\\ \dfrac{elevation-120}{170-120} & if\ 120\le elevation\le 170\\ 1 & if\ 170\le elevation\end{cases}$$

Figure 4.5 shows the membership function for the feature of "along channel length (ACL)", based on Eq 4.1:

$$\mu_{\mathrm{L}}(ACL)=\begin{cases}1 & if\ ACL\le 1\\ \dfrac{3-ACL}{3-1} & if\ 1\le ACL\le 3\\ 0 & if\ 3\le ACL\end{cases}$$

$$\mu_M(ACL) = \begin{cases} \dfrac{ACL-1}{3-1} & if\ 1 \le ACL \le 3 \\ \dfrac{7-ACL}{7-3} & if\ 3 \le ACL \le 7 \\ 0 & otherwise \end{cases}$$

$$\mu_H(ACL) = \begin{cases} 0 & if\ 1 \le ACL \\ \dfrac{ACL-3}{7-3} & if\ 3 \le ACL \le 7 \\ 1 & if\ 7 \le ACL \end{cases}$$

Figure 4.6 shows the membership function for the feature of "along channel slope (ACS)", based on Eq 4.1:

$$\mu_L(ACS) = \begin{cases} 1 & if\ ACS \le 0.1 \\ \dfrac{0.3-ACS}{0.3-0.1} & if\ 0.1 \le ACS \le 0.3 \\ 0 & if\ 0.3 \le ACS \end{cases}$$

$$\mu_M(ACS) = \begin{cases} \dfrac{ACS-0.1}{0.3-0.1} & if\ 0.1 \le ACS \le 0.3 \\ \dfrac{0.5-ACS}{0.5-0.3} & if\ 0.3 \le ACS \le 0.5 \\ 0 & otherwise \end{cases}$$

$$\mu_H(ACS) = \begin{cases} 0 & if\ 0.1 \le ACS \\ \dfrac{ACS-0.3}{0.5-0.3} & if\ 0.3 \le ACS \le 0.5 \\ 1 & if\ 0.5 \le ACS \end{cases}$$

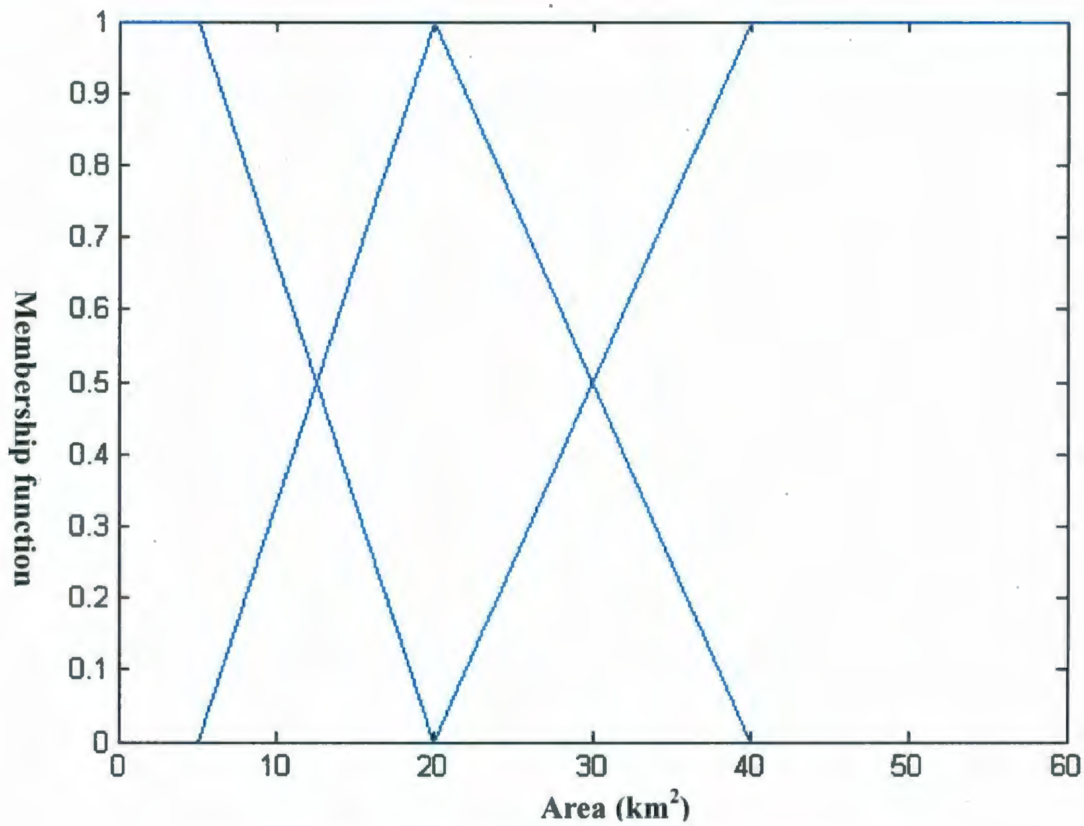


Figure 4.3 Membership function for the feature of "Area"

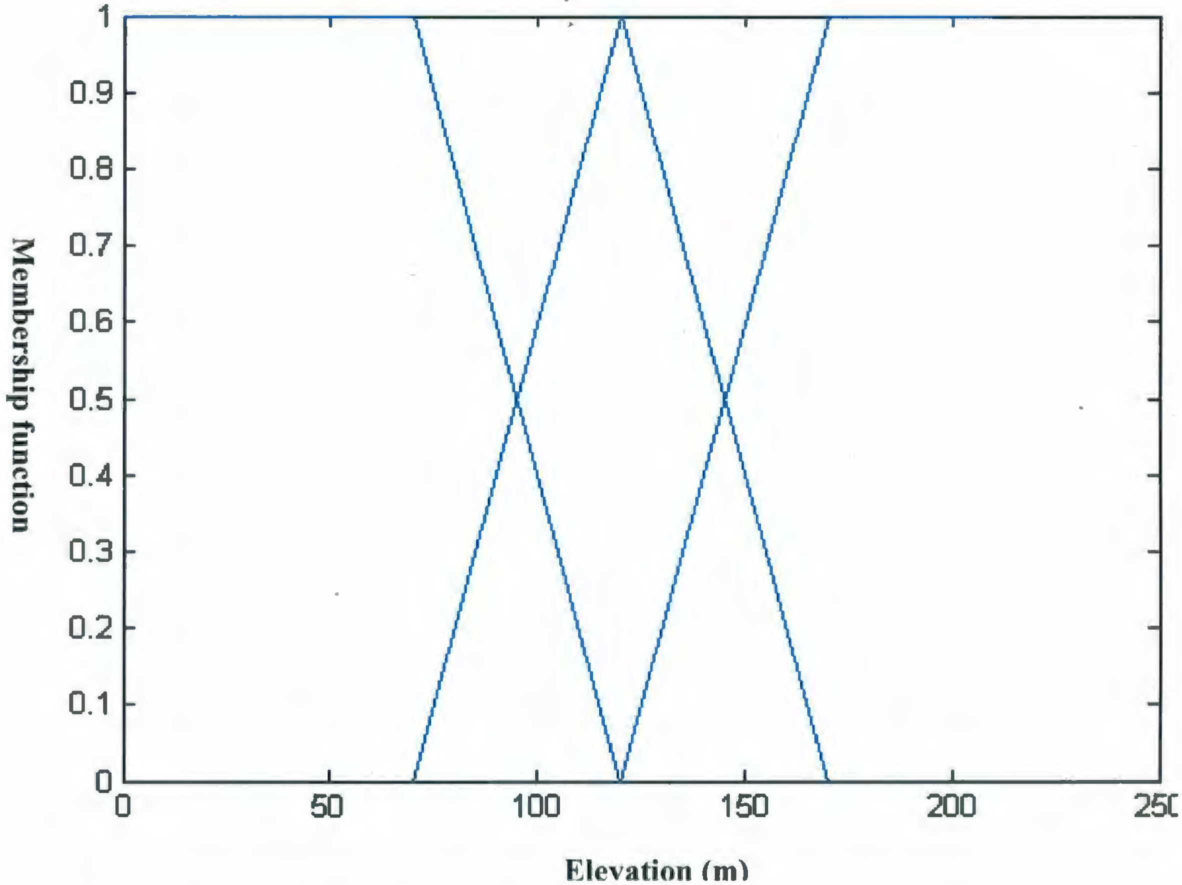


Figure 4.4 Membership function for the feature of "Elevation"

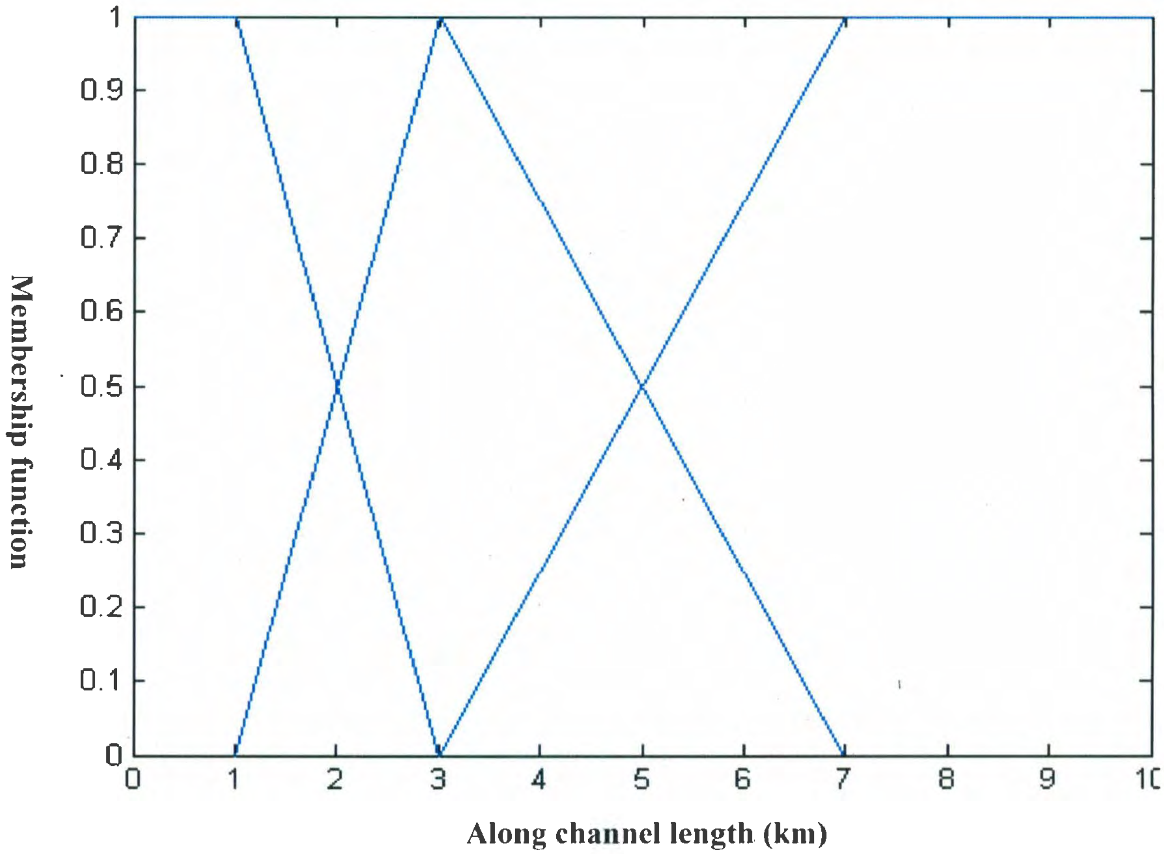


Figure 4.5 Membership function for the feature of "Along channel length"

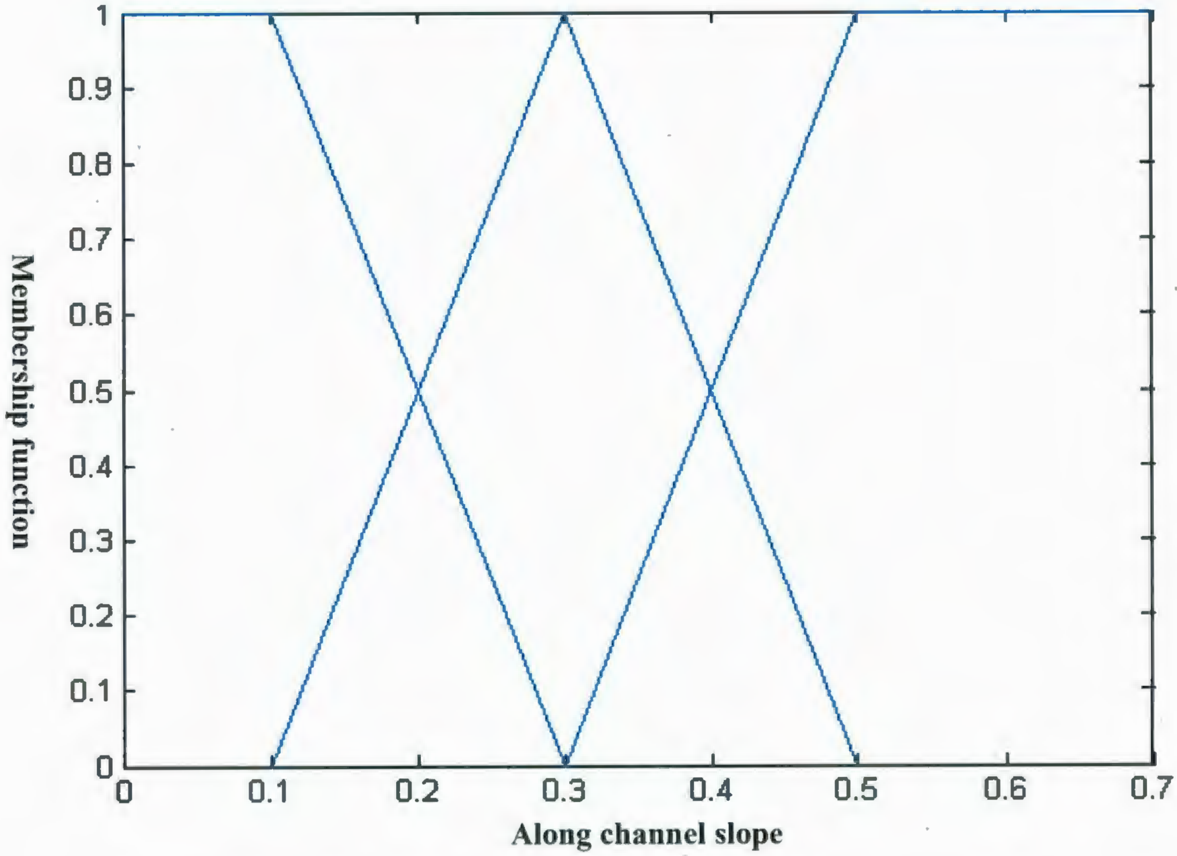


Figure 4.6 Membership function for the feature of "along channel slope"

The original data, I_{a0}, for these four features of 92 sub-basins are shown in Table 3.1. By Eqs 4.1 to 4.4, the original input I_{a0} is converted into the fuzzy sets.

In this case, the number of input data is 92, the number of parameters is 4, and the number of membership function levels is 3. Therefore, the original inputs were converted into the following format:

$$\begin{bmatrix} \mu_1(x_{11}) & \mu_2(x_{11}) & \mu_3(x_{11}) & \mu_1(x_{12}) & \mu_2(x_{12}) & \mu_3(x_{12}) & \mu_1(x_{13}) & \mu_2(x_{13}) & \mu_3(x_{13}) & \mu_1(x_{14}) & \mu_2(x_{14}) & \mu_3(x_{14}) \\ \mu_1(x_{21}) & \mu_2(x_{21}) & \mu_3(x_{21}) & \mu_1(x_{22}) & \mu_2(x_{22}) & \mu_3(x_{22}) & \mu_1(x_{23}) & \mu_2(x_{23}) & \mu_3(x_{23}) & \mu_1(x_{24}) & \mu_2(x_{24}) & \mu_3(x_{24}) \\ & & & & & \vdots & & & & & & \\ \mu_1(x_{921}) & \mu_2(x_{921}) & \mu_3(x_{921}) & \mu_1(x_{922}) & \mu_2(x_{922}) & \mu_3(x_{922}) & \mu_1(x_{923}) & \mu_2(x_{923}) & \mu_3(x_{923}) & \mu_1(x_{924}) & \mu_2(x_{924}) & \mu_3(x_{924}) \end{bmatrix}$$

The fuzzified values are shown in Table 4.1.

The vigilance for the ART_1 and ART_3 modules were $\rho = 0.7$, which is the maximum vigilance that can be used to fully classify all of the sub-basins via various attempts. The vigilance for ART_2 was $\rho = 0.5$, which indicates a similarity between combination and centroid pattern that if the match percentage of them is greater than 50%. The learning rates for all three modules were $\beta = 1$, which means a fast learning used for the approach. The choice parameter was set as 0.0001 to ensure that one category was active at one time.

The input data was normalized and fed to ART_1 for unsupervised learning. The unsupervised classification results are shown in Table 4.2.

Table 4.1 Fuzzified input data for the 92 sub-basins in the IRFAM approach

Sub-basin #	Area			Elevation			Along channel length			Along channel slope		
	L	M	H	L	M	H	L	M	H	L	M	H
1	**0.0000**	**0.3120**	**0.6880**	**0.54**	**0.46**	**0.00**	**0.000**	**0.0000**	**1.0000**	**0.95**	**0.05**	**0.00**
2	**0.4213**	0.5787	0.0000	0.76	0.24	0.00	0.000	0.0000	1.0000	1.00	0.00	0.00
3	**0.0000**	0.4935	0.5065	0.74	0.26	0.00	0.000	0.0000	1.0000	1.00	0.00	0.00
4	**0.0000**	0.0000	1.0000	1.00	0.00	0.00	0.000	0.0000	1.0000	1.00	0.00	0.00
5	**0.0000**	0.0330	0.9670	1.00	0.00	0.00	1.000	0.0000	0.0000	1.00	0.00	0.00
6	**0.8967**	0.1033	0.0000	1.00	0.00	0.00	0.000	0.0000	1.0000	0.95	0.05	0.00
7	**0.0000**	0.9220	0.0780	1.00	0.00	0.00	0.000	0.2725	0.7275	1.00	0.00	0.00
8	**0.6987**	0.3013	0.0000	0.98	0.02	0.00	0.000	0.0000	1.0000	0.90	0.10	0.00
9	**0.2500**	0.7500	0.0000	0.90	0.10	0.00	0.000	0.0000	1.0000	1.00	0.00	0.00
10	**0.8880**	0.1120	0.0000	0.84	0.16	0.00	0.000	0.4075	0.5925	0.90	0.10	0.00
11	**0.7520**	0.2480	0.0000	0.80	0.20	0.00	0.000	0.0250	0.9750	1.00	0.00	0.00
12	**0.9967**	0.0033	0.0000	0.94	0.06	0.00	0.465	0.5350	0.0000	1.00	0.00	0.00
13	**0.9467**	0.0533	0.0000	0.76	0.24	0.00	0.000	0.4325	0.5675	1.00	0.00	0.00
14	**0.7373**	0.2627	0.0000	0.76	0.24	0.00	0.000	0.0000	1.0000	1.00	0.00	0.00
15	**0.8713**	0.1287	0.0000	0.96	0.04	0.00	0.000	0.0375	0.9625	1.00	0.00	0.00
16	**1.0000**	0.0000	0.0000	0.80	0.20	0.00	1.000	0.0000	0.0000	1.00	0.00	0.00
17	**0.7213**	0.2787	0.0000	0.76	0.24	0.00	0.000	0.2875	0.7125	1.00	0.00	0.00
18	**0.9067**	0.0933	0.0000	1.00	0.00	0.00	0.000	0.3150	0.6850	1.00	0.00	0.00
19	**0.0000**	0.0000	1.0000	1.00	0.00	0.00	1.000	0.0000	0.0000	1.00	0.00	0.00
20	**0.0000**	0.9105	0.0895	1.00	0.00	0.00	0.995	0.0050	0.0000	1.00	0.00	0.00
21	**0.1667**	0.8333	0.0000	1.00	0.00	0.00	0.000	0.0000	1.0000	1.00	0.00	0.00
22	**0.5813**	0.4187	0.0000	1.00	0.00	0.00	0.955	0.0450	0.0000	0.50	0.50	0.00
23	**0.0000**	0.0000	1.0000	1.00	0.00	0.00	0.985	0.0150	0.0000	1.00	0.00	0.00

24	**0.4680**	0.5320	0.0000	0.40	0.60	0.00	0.455	0.5450	0.0000	1.00	0.00	0.00
25	**0.7080**	0.2920	0.0000	0.60	0.40	0.00	0.885	0.1150	0.0000	0.80	0.20	0.00
26	**0.9107**	0.0893	0.0000	0.62	0.38	0.00	1.000	0.0000	0.0000	1.00	0.00	0.00
27	**0.6327**	0.3673	0.0000	0.00	0.84	0.16	0.825	0.1750	0.0000	0.10	0.90	0.00
28	**1.0000**	0.0000	0.0000	0.00	0.92	0.08	1.000	0.0000	0.0000	0.05	0.95	0.00
29	**1.0000**	0.0000	0.0000	0.00	0.90	0.10	0.910	0.0900	0.0000	0.40	0.60	0.00
30	**1.0000**	0.0000	0.0000	0.00	0.64	0.36	0.000	0.6375	0.3625	0.75	0.25	0.00
31	**0.8367**	0.1633	0.0000	0.00	0.52	0.48	0.400	0.6000	0.0000	0.00	0.45	0.55
32	**0.0000**	0.5775	0.4225	0.00	0.60	0.40	1.000	0.0000	0.0000	0.00	0.00	1.00
33	**1.0000**	0.0000	0.0000	0.00	0.58	0.42	0.605	0.3950	0.0000	0.00	0.00	1.00
34	**1.0000**	0.0000	0.0000	0.00	0.38	0.62	0.665	0.3350	0.0000	0.20	0.80	0.00
35	**0.8153**	0.1847	0.0000	0.00	0.38	0.62	0.710	0.2900	0.0000	0.00	0.30	0.70
36	**0.5820**	0.4180	0.0000	0.00	0.22	0.78	1.000	0.0000	0.0000	0.40	0.60	0.00
37	**0.6507**	0.3493	0.0000	0.00	0.14	0.86	0.650	0.3500	0.0000	0.00	0.00	1.00
38	**0.0000**	0.9055	0.0945	0.00	0.08	0.92	1.000	0.0000	0.0000	0.00	0.00	1.00
39	**1.0000**	0.0000	0.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.15	0.85	0.00
40	**1.0000**	0.0000	0.0000	0.00	0.22	0.78	0.800	0.2000	0.0000	0.00	0.95	0.05
41	**0.0000**	0.0000	1.0000	0.32	0.68	0.00	0.650	0.3500	0.0000	0.25	0.75	0.00
42	**0.0000**	0.0000	1.0000	0.32	0.68	0.00	0.855	0.1450	0.0000	0.25	0.75	0.00
43	**0.0000**	0.3900	0.6100	0.66	0.34	0.00	0.505	0.4950	0.0000	0.75	0.25	0.00
44	**0.0000**	0.7610	0.2390	0.66	0.34	0.00	1.000	0.0000	0.0000	0.60	0.40	0.00
45	**1.0000**	0.0000	0.0000	0.58	0.42	0.00	0.960	0.0400	0.0000	0.55	0.45	0.00
46	**0.0000**	0.3750	0.6250	0.92	0.08	0.00	0.755	0.2450	0.0000	0.70	0.30	0.00
47	**0.0000**	0.8375	0.1625	0.92	0.08	0.00	0.850	0.1500	0.0000	0.75	0.25	0.00
48	**0.0000**	0.6030	0.3970	0.94	0.06	0.00	0.900	0.1000	0.0000	0.95	0.05	0.00
49	**0.7720**	0.2280	0.0000	0.74	0.26	0.00	1.000	0.0000	0.0000	1.00	0.00	0.00
50	**0.2713**	0.7287	0.0000	0.80	0.20	0.00	1.000	0.0000	0.0000	1.00	0.00	0.00

51	**0.9553**	0.0447	0.0000	0.86	0.14	0.00	0.610	0.3900	0.0000	0.95	0.05	0.00
52	**0.6587**	0.3413	0.0000	0.82	0.18	0.00	0.320	0.6800	0.0000	0.75	0.25	0.00
53	**0.0000**	0.0000	1.0000	0.66	0.34	0.00	1.000	0.0000	0.0000	0.45	0.55	0.00
54	**0.5827**	0.4173	0.0000	0.66	0.34	0.00	1.000	0.0000	0.0000	0.45	0.55	0.00
55	**0.0000**	0.0135	0.9865	0.64	0.36	0.00	0.620	0.3800	0.0000	0.00	0.95	0.05
56	**0.0000**	0.7740	0.2260	0.62	0.38	0.00	0.355	0.6450	0.0000	0.35	0.65	0.00
57	**0.8273**	0.1727	0.0000	0.60	0.40	0.00	0.540	0.4600	0.0000	0.50	0.50	0.00
58	**1.0000**	0.0000	0.0000	0.56	0.44	0.00	1.000	0.0000	0.0000	0.50	0.50	0.00
59	**0.8173**	0.1827	0.0000	0.30	0.70	0.00	1.000	0.0000	0.0000	0.00	0.75	0.25
60	**0.5873**	0.4127	0.0000	0.20	0.80	0.00	0.540	0.4600	0.0000	0.30	0.70	0.00
61	**0.7133**	0.2867	0.0000	0.00	0.82	0.18	0.845	0.1550	0.0000	0.00	0.90	0.10
62	**0.3873**	0.6127	0.0000	0.00	0.38	0.62	1.000	0.0000	0.0000	0.00	0.00	1.00
63	**0.0000**	0.0000	1.0000	0.00	0.50	0.50	0.935	0.0650	0.0000	0.00	0.00	1.00
64	**1.0000**	0.0000	0.0000	0.66	0.34	0.00	1.000	0.0000	0.0000	1.00	0.00	0.00
65	**1.0000**	0.0000	0.0000	0.66	0.34	0.00	1.000	0.0000	0.0000	0.95	0.05	0.00
66	**0.0000**	0.3960	0.6040	0.66	0.34	0.00	0.770	0.2300	0.0000	0.85	0.15	0.00
67	**0.0000**	0.7485	0.2515	0.46	0.54	0.00	0.805	0.1950	0.0000	0.00	1.00	0.00
68	**0.1673**	0.8327	0.0000	0.50	0.50	0.00	0.265	0.7350	0.0000	0.65	0.35	0.00
69	**0.7860**	0.2140	0.0000	0.18	0.82	0.00	1.000	0.0000	0.0000	0.00	0.60	0.40
70	**0.0000**	0.0000	1.0000	0.20	0.80	0.00	0.910	0.0900	0.0000	0.00	0.60	0.40
71	**0.0060**	0.9940	0.0000	0.00	0.54	0.46	1.000	0.0000	0.0000	0.00	0.00	1.00
72	**0.0000**	0.0000	1.0000	0.00	0.00	1.00	0.950	0.0500	0.0000	0.00	0.00	1.00
73	**0.9980**	0.0020	0.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.00	0.00	1.00
74	**0.6393**	0.3607	0.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.00	0.40	0.60
75	**0.6593**	0.3407	0.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	1.00	0.00	0.00
76	**0.0000**	0.3440	0.6560	0.00	0.50	0.50	1.000	0.0000	0.0000	0.00	0.15	0.85
77	**0.0000**	0.8210	0.1790	0.00	0.08	0.92	1.000	0.0000	0.0000	0.00	0.00	1.00

78	**0.0000**	0.0000	1.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.00	0.35	0.65
79	**0.6680**	0.3320	0.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.00	0.00	1.00
80	**0.0627**	0.9373	0.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.00	0.45	0.55
81	**0.0000**	0.0000	1.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.00	0.60	0.40
82	**0.0000**	0.0000	1.0000	0.00	0.00	1.00	0.660	0.3400	0.0000	0.00	0.25	0.75
83	**0.0000**	0.0000	1.0000	0.00	0.00	1.00	0.900	0.1000	0.0000	0.00	0.50	0.50
84	**0.5813**	0.4187	0.0000	0.00	0.00	1.00	0.765	0.2350	0.0000	0.00	0.05	0.95
85	**0.1333**	0.8667	0.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.00	0.00	1.00
86	**0.9713**	0.0287	0.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.00	0.90	0.10
87	**0.0000**	0.0000	1.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.00	0.00	1.00
88	**0.7073**	0.2927	0.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.00	0.00	1.00
89	**0.0000**	0.0000	1.0000	0.00	0.00	1.00	0.970	0.0300	0.0000	0.00	0.05	0.95
90	**0.0000**	0.0000	1.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.00	0.25	0.75
91	**0.0000**	0.0000	1.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.00	0.00	1.00
92	**0.6393**	0.3607	0.0000	0.00	0.00	1.00	1.000	0.0000	0.0000	0.00	0.00	1.00

Table 4.2 Unsupervised classification results by the IRFAM

Class #	Sub-basin #
1	1, 3, 4,
2	5, 12, 16, 19, 23
3	2, 6, 7, 8, 9, 10, 11, 13, 14, 15, 17, 18, 21
4	20, 22, 44,
5	24, 25, 26, 45, 57, 58
6	27, 28, 29, 34, 61
7	30, 31
8	32, 33, 62
9	35, 36, 40
10	37, 38, 71, 77, 84, 85
11	39, 59, 69, 86
12	41, 42, 43, 53
13	46, 47, 48, 49, 50,
14	51, 52,54
15	55,56, 67
16	60, 68, 87
17	63, 70, 72, 78, 81, 83, 89, 90, 91
18	64, 65, 66
19	73, 74, 75, 79, 88, 92
20	76, 80
21	82

Based on the unsupervised classification results and the preset target group number, the centroid value was determined by the centroid determination module. The number of target groups in this case was preset as 5, therefore Class 3, 5, 10, 17, and 20 were selected and the centroid values were obtained correspondingly (Table 4.3).

As number of input features was 4 and the 3 levels triangular membership function was used, the number of criteria combinations for the IRFAM approach was $3^4 = 81$. The details of criteria combination I_b are shown in Table 4.4 The criteria combinations were fed to ART_{2a} and the group centroids were used in ART_{2b}. Consequently the criteria combinations were classified into 5 groups. The classification results of combinations are shown in Table 4.5.

The fuzzified original input patterns were fed to the ART_{3a} and the criteria combinations were used in the ART_{3b}, and then each pattern was placed in a criteria combination which presented a best match. Then the fuzzy rule was applied for screening the data which matched with certain criteria combinations, and classifying them into the five preset groups. The final classification results are shown in Table 4.6.

Table 4.3 Criteria combinations in the IRFAM approach

Combination #	Area			Elevation			Along channel length			Along channel slope		
	L	M	H	L	M	H	L	M	H	L	M	H
1	1	0	0	1	0	0	1	0	0	1	0	0
2	1	0	0	1	0	0	1	0	0	0	1	0
3	1	0	0	1	0	0	1	0	0	0	0	1
4	1	0	0	1	0	0	0	1	0	1	0	0
5	1	0	0	1	0	0	0	1	0	0	1	0
6	1	0	0	1	0	0	0	1	0	0	0	1
7	1	0	0	1	0	0	0	0	1	1	0	0
8	1	0	0	1	0	0	0	0	1	0	1	0
9	1	0	0	1	0	0	0	0	1	0	0	1
10	1	0	0	0	1	0	1	0	0	1	0	0
11	1	0	0	0	1	0	1	0	0	0	1	0
12	1	0	0	0	1	0	1	0	0	0	0	1
13	1	0	0	0	1	0	0	1	0	1	0	0
14	1	0	0	0	1	0	0	1	0	0	1	0
15	1	0	0	0	1	0	0	1	0	0	0	1
16	1	0	0	0	1	0	0	0	1	1	0	0
17	1	0	0	0	1	0	0	0	1	0	1	0
18	1	0	0	0	1	0	0	0	1	0	0	1
19	1	0	0	0	0	1	1	0	0	1	0	0
20	1	0	0	0	0	1	1	0	0	0	1	0
21	1	0	0	0	0	1	1	0	0	0	0	1
22	1	0	0	0	0	1	0	1	0	1	0	0
23	1	0	0	0	0	1	0	1	0	0	1	0
24	1	0	0	0	0	1	0	1	0	0	0	1
25	1	0	0	0	0	1	0	0	1	1	0	0
26	1	0	0	0	0	1	0	0	1	0	1	0
27	1	0	0	0	0	1	0	0	1	0	0	1
28	0	1	0	1	0	0	1	0	0	1	0	0
29	0	1	0	1	0	0	1	0	0	0	1	0
30	0	1	0	1	0	0	1	0	0	0	0	1
31	0	1	0	1	0	0	0	1	0	1	0	0
32	0	1	0	1	0	0	0	1	0	0	1	0
33	0	1	0	1	0	0	0	1	0	0	0	1

34	0	1	0	1	0	0	0	0	1	1	0	0
35	0	1	0	1	0	0	0	0	1	0	1	0
36	0	1	0	1	0	0	0	0	1	0	0	1
37	0	1	0	0	1	0	1	0	0	1	0	0
38	0	1	0	0	1	0	1	0	0	0	1	0
39	0	1	0	0	1	0	1	0	0	0	0	1
40	0	1	0	0	1	0	0	1	0	1	0	0
41	0	1	0	0	1	0	0	1	0	0	1	0
42	0	1	0	0	1	0	0	1	0	0	0	1
43	0	1	0	0	1	0	0	0	1	1	0	0
44	0	1	0	0	1	0	0	0	1	0	1	0
45	0	1	0	0	1	0	0	0	1	0	0	1
46	0	1	0	0	0	1	1	0	0	1	0	0
47	0	1	0	0	0	1	1	0	0	0	1	0
48	0	1	0	0	0	1	1	0	0	0	0	1
49	0	1	0	0	0	1	0	1	0	1	0	0
50	0	1	0	0	0	1	0	1	0	0	1	0
51	0	1	0	0	0	1	0	1	0	0	0	1
52	0	1	0	0	0	1	0	0	1	1	0	0
53	0	1	0	0	0	1	0	0	1	0	1	0
54	0	1	0	0	0	1	0	0	1	0	0	1
55	0	0	1	1	0	0	1	0	0	1	0	0
56	0	0	1	1	0	0	1	0	0	0	1	0
57	0	0	1	1	0	0	1	0	0	0	0	1
58	0	0	1	1	0	0	0	1	0	1	0	0
59	0	0	1	1	0	0	0	1	0	0	1	0
60	0	0	1	1	0	0	0	1	0	0	0	1
61	0	0	1	1	0	0	0	0	1	1	0	0
62	0	0	1	1	0	0	0	0	1	0	1	0
63	0	0	1	1	0	0	0	0	1	0	0	1
64	0	0	1	0	1	0	1	0	0	1	0	0
65	0	0	1	0	1	0	1	0	0	0	1	0
66	0	0	1	0	1	0	1	0	0	0	0	1
67	0	0	1	0	1	0	0	1	0	1	0	0
68	0	0	1	0	1	0	0	1	0	0	1	0
69	0	0	1	0	1	0	0	1	0	0	0	1
70	0	0	1	0	1	0	0	0	1	1	0	0
71	0	0	1	0	1	0	0	0	1	0	1	0
72	0	0	1	0	1	0	0	0	1	0	0	1

73	0	0	1	0	0	1	1	0	0	1	0	0
74	0	0	1	0	0	1	1	0	0	0	1	0
75	0	0	1	0	0	1	1	0	0	0	0	1
76	0	0	1	0	0	1	0	1	0	1	0	0
77	0	0	1	0	0	1	0	1	0	0	1	0
78	0	0	1	0	0	1	0	1	0	0	0	1
79	0	0	1	0	0	1	0	0	1	1	0	0
80	0	0	1	0	0	1	0	0	1	0	1	0
81	0	0	1	0	0	1	0	0	1	0	0	1

Table 4.4 Centroid values for the 5 groups

Group	Area			Elevation			Along channel length			Along channel slope		
	L	M	H	L	M	H	L	M	H	L	M	H
1	0.00	0.00	1.00	0.02	0.13	0.85	0.97	0.03	0.00	0.00	0.23	0.77
2	0.72	0.28	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.17	0.06	0.77
3	0.64	0.36	0.01	0.89	0.11	0.00	0.00	0.14	0.86	0.98	0.02	0.00
4	0.23	0.73	0.05	0.00	0.14	0.86	0.90	0.10	0.00	0.00	0.01	0.99
5	0.82	0.18	0.00	0.56	0.44	0.00	0.81	0.19	0.00	0.73	0.27	0.00

Table 4.5 Classification results for the criteira combinations

Class #	Criteria Combination #
1	56, 57, 59, 60, 63, 64, 65, 66, 68, 69, 71, 72, 73, 74, 75, 76, 77, 78, 80, 81
2	3, 12, 15, 19, 20, 21, 22, 23, 24, 26, 27,
3	4, 6, 7, 8, 9, 16, 17, 18, 25, 31, 32, 34, 35, 36, 40, 43, 44, 52, 58, 61, 62, 70, 79
4	30, 33, 38, 39, 42, 45, 46, 47, 48, 49, 50, 51, 53, 54,
5	1, 2, 5, 10, 11, 13, 14, 28, 29, 37, 41, 55, 67,

Table 4.6 Final classification results by the IRFAM

Group	Sub-basin #
1	41, 42, 53, 55, 63, 70, 72, 76, 78, 81, 82, 83, 87, 89, 90, 91
2	31, 33, 34, 35, 36, 37, 39, 40, 73, 74, 75, 79, 84, 86, 88, 92
3	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 21, 24, 52, 56, 68
4	32, 38, 62, 67, 71, 77, 80, 85
5	5, 16, 19, 20, 22, 23, 25, 26, 27, 28, 29, 30, 43, 44, 45, 46, 47, 48, 49, 50, 51, 54, 57, 58, 59, 60, 61, 64, 65, 66, 69

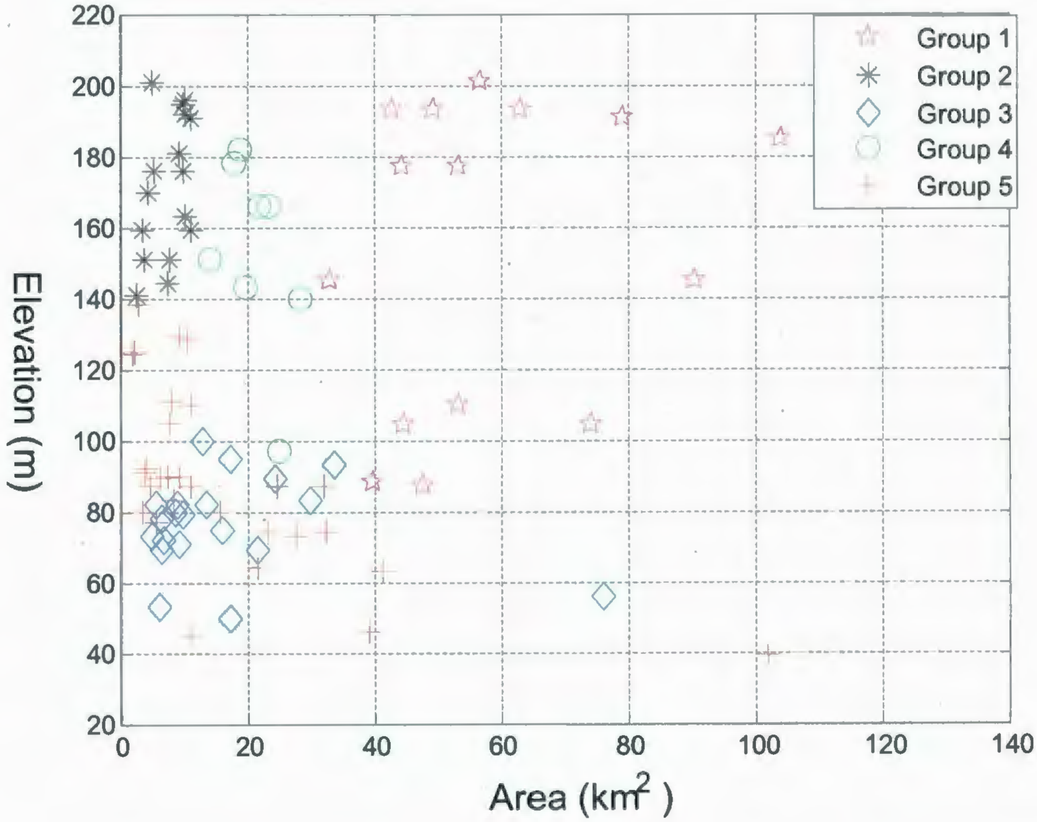


Figure 4.7 Distribution of Area vs. Elevation in the final classification results by IRFAM

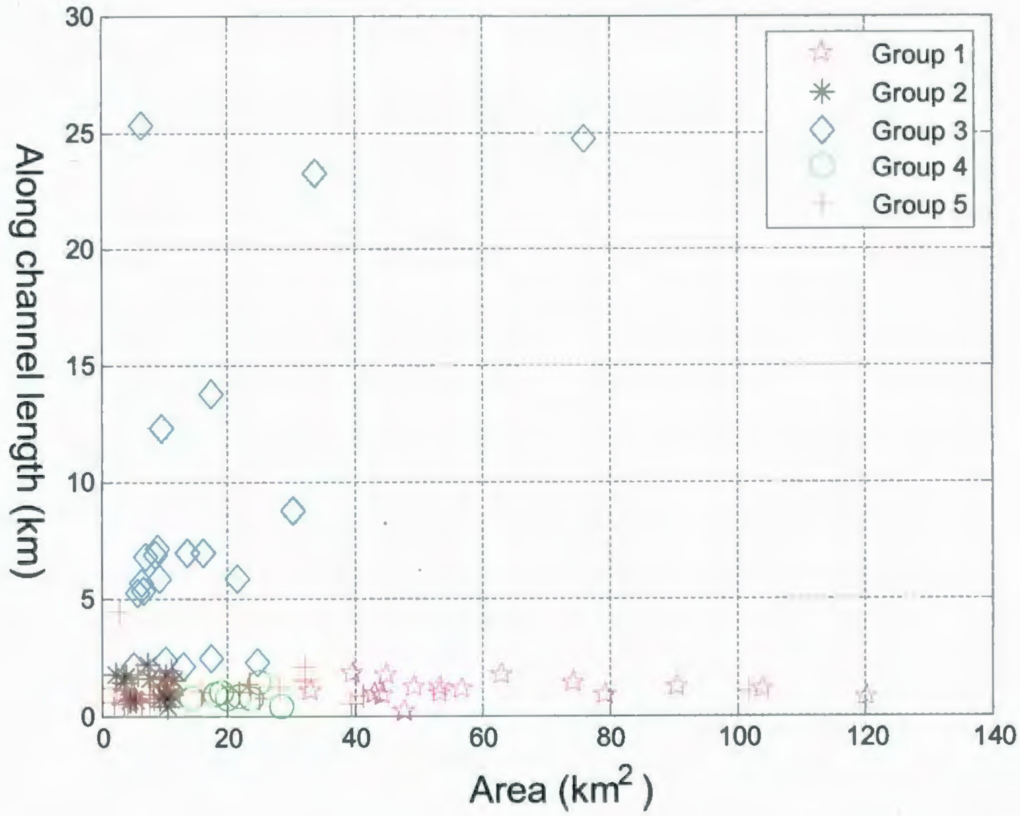


Figure 4.8 Distribution of Area vs. Along channel length in the final classification results by IRFAM

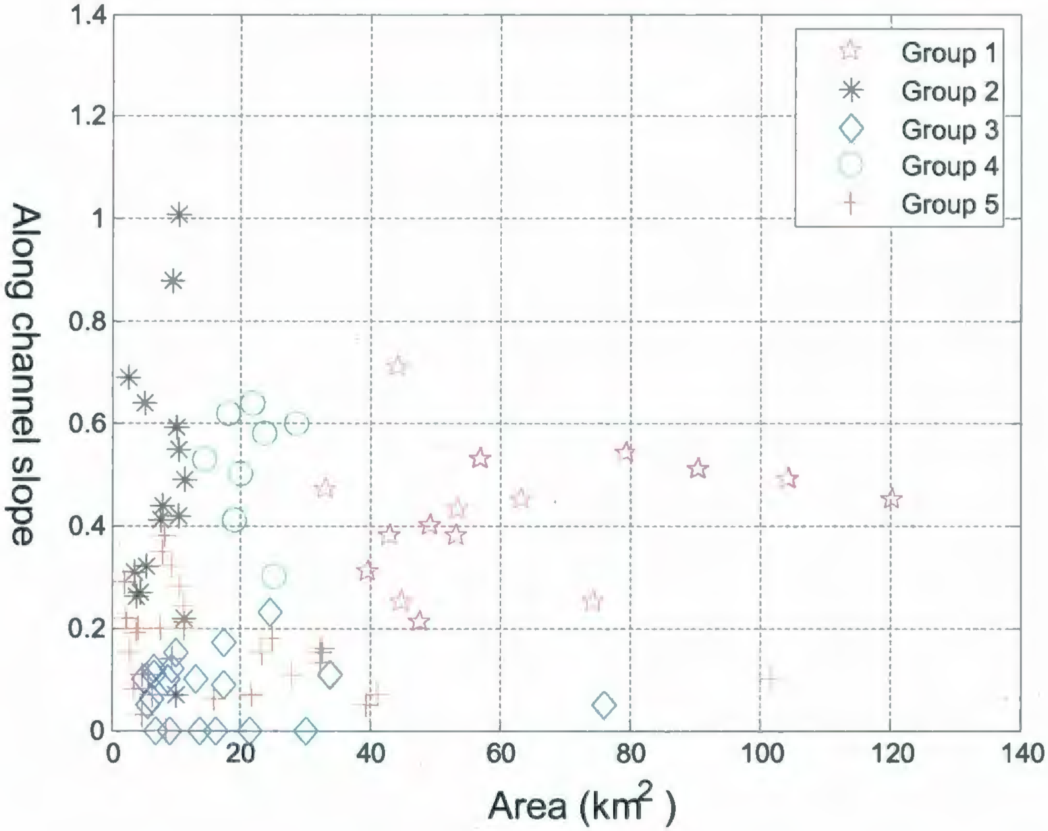


Figure 4.9 Distribution of Area vs. Along channel slope in the final classification results by IRFAM

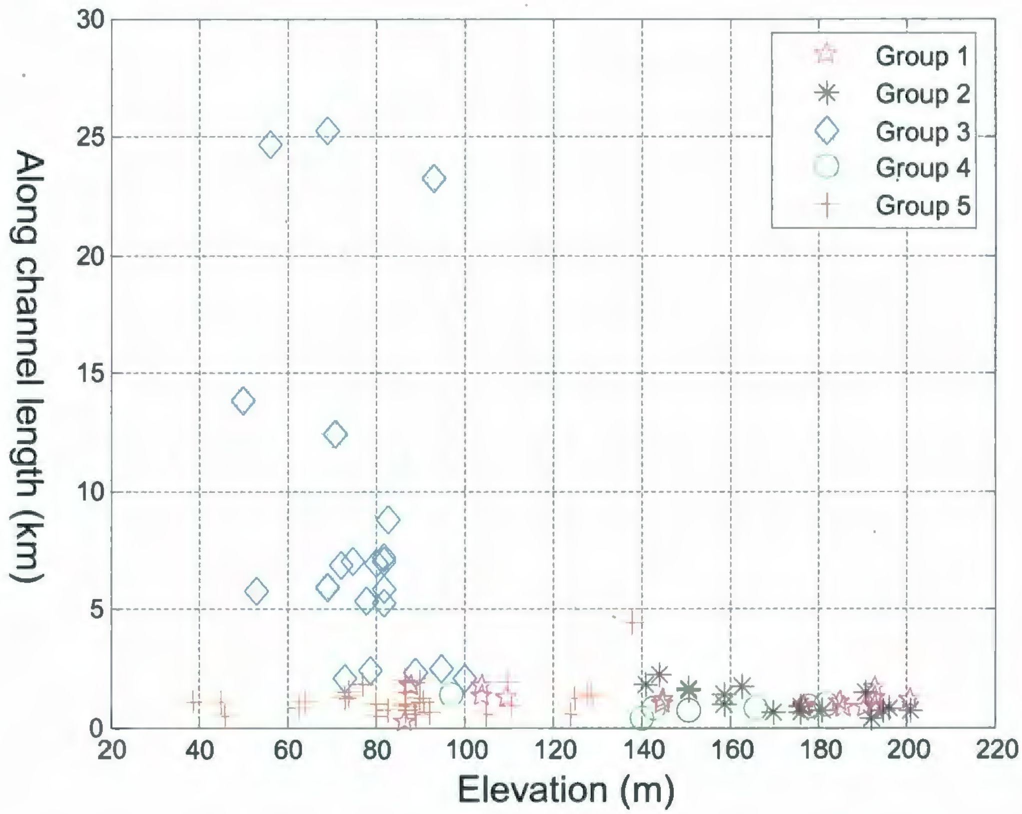


Figure 4.10 Distribution of Elevation vs. Along channel length the final classification results by IRFAM

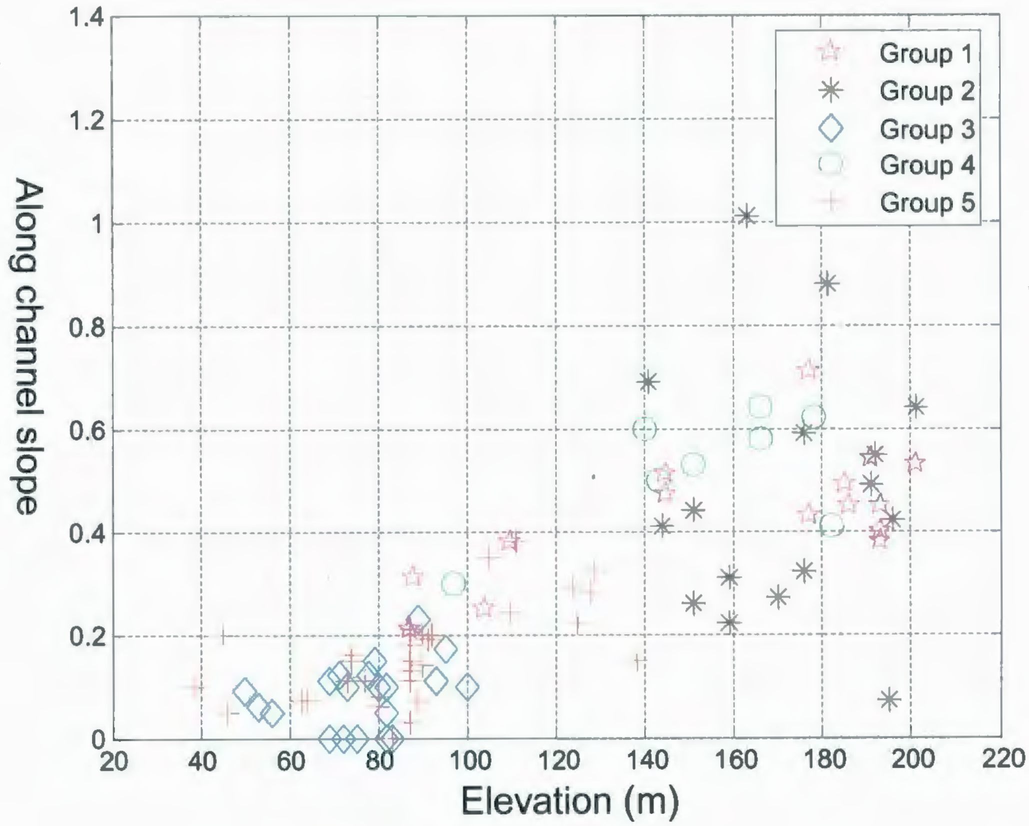


Figure 4.11 Distribution of Elevation vs. Along channel slope in the final classification results by IRFAM

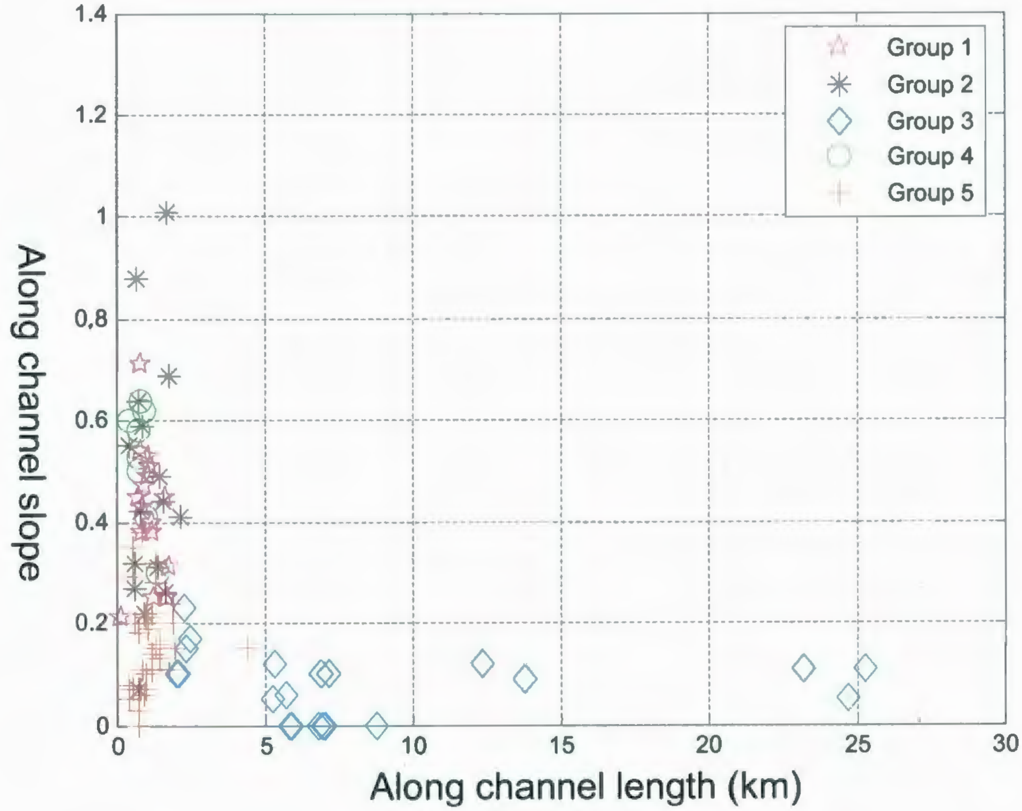


Figure 4.12 Distribution of Along channel length vs. Along channel slope in the final classification results by IRFAM

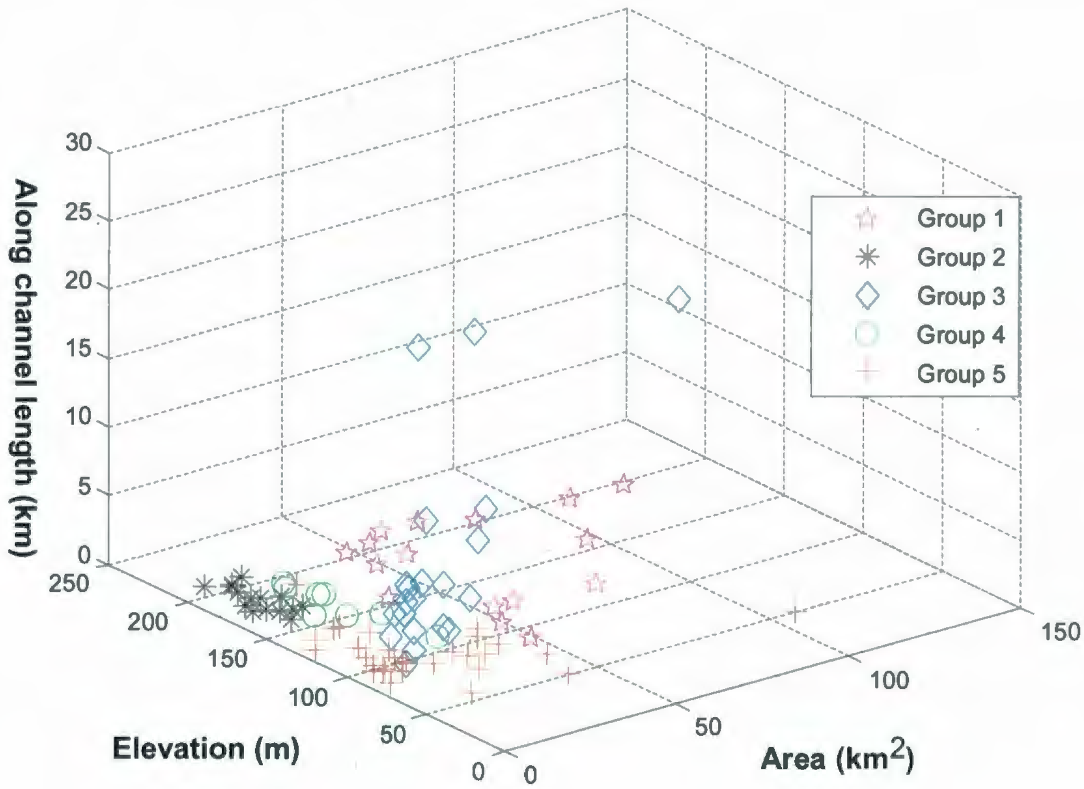


Figure 4.13 Distribution of Area vs. Elevation vs. Along channel length in the final classification results by IRFAM

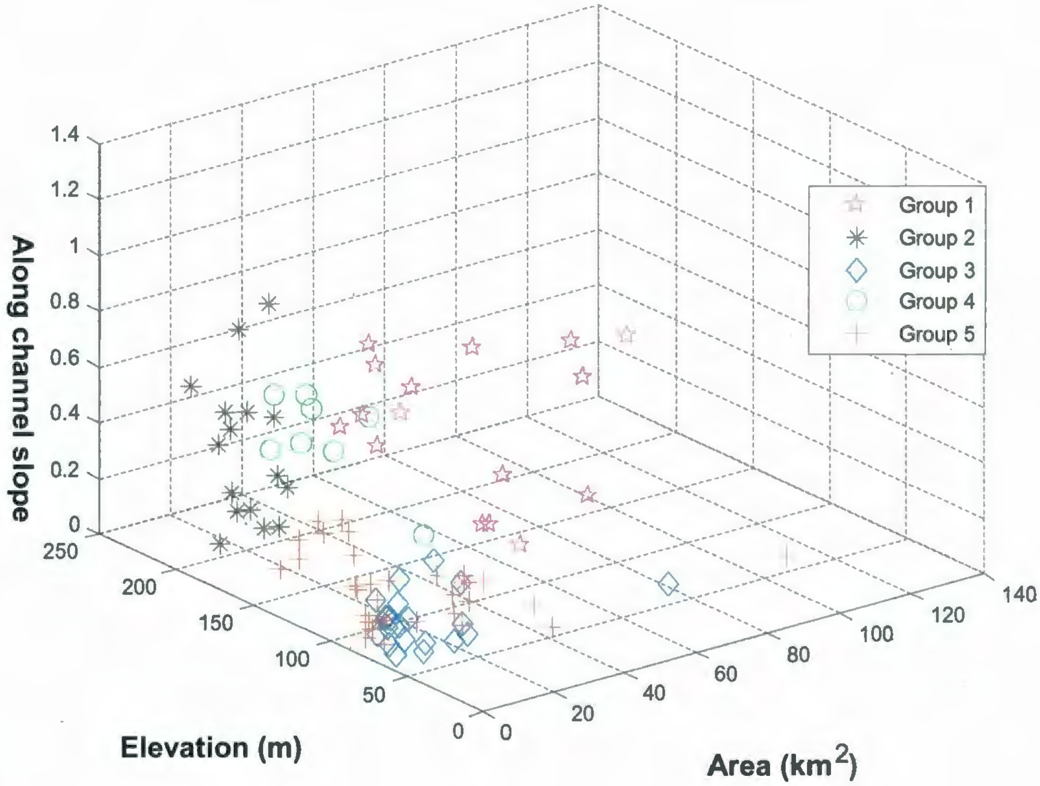


Figure 4.14 Distribution of Area vs. Elevation vs. Along channel slope in the final classification results by IRFAM

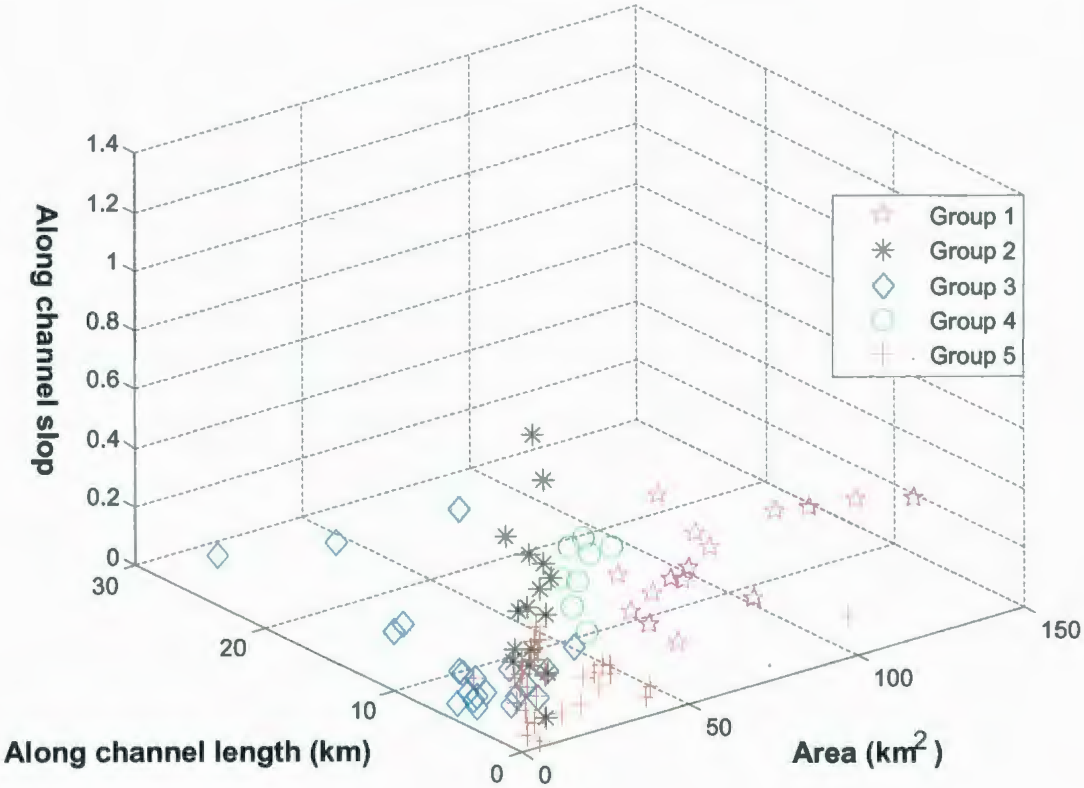


Figure 4.15 Distribution of Area vs. Along channel length vs. Along channel slope in the final classification results by IRFAM

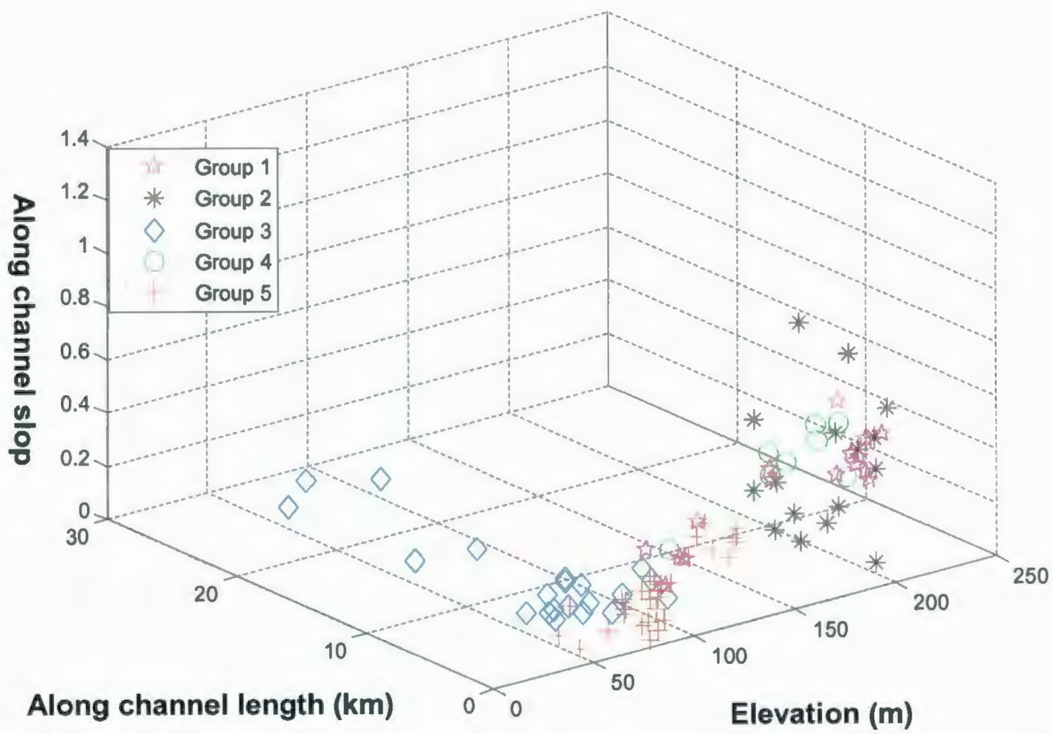


Figure 4.16 Distribution of Elevation vs. Along channel length vs. Along channel slope in the final classification results by IRFAM

4.3.2 Result Analysis and Discussion

The classification result showed that all the 92 sub-basins can be properly classified into 5 preset target groups in the case of vigilance ρ equals to 0.7.

By using Eq 3.13, the new centroid values were obtained from the final classified groups. The old and new centroid values are listed in Table 4.7. The new centroid values appear a little different from the old ones. This is because adaptation of uncertainty by introducing fuzzy set theory. From Figures 4.7 to 4.16 it can be seem in almost all the figures that each group clearly has its own boundary from the others, which approves a high ability of uncertainty handling for the IRFAM approach.

The distributions of different features in the classification results were shown in Figures 4.7 to 4.16. Figure 4.7 indicated that the double features of area and elevation had significant contribution to classify Group 1, 2 and 4, and little to the others. Figure 4.8 indicated that the double features of area and along channel length had significant contribution to classified Group 1 and 3, and less to the others. Figure 4.9 indicated that the double features of area and along channel slope had significant contribution to classify Group 1 and 4, and less to the others. Figure 4.10 indicated that the double features of elevation and along channel length only had contribution to classify Group 3, and almost no contribution to the others. Figure 4.11 indicated that the couple features of elevation and along channel slope only had little contribution to the classification. Figure 4.12 indicated that the couple features of along channel length and along channel slope had significant contribution to classify Group 4, and almost no contribution to the others. Figures 4.13 to 4.16 indicated that almost all the triple features had signification contributions to the classification.

Table 4.7 Comparison between old and new centroid value for the features of area, elevation, along channel length, and along channel slope

Group	Area					
	Old			New		
	L	M	H	L	M	H
1	0.00	0.00	1.00	0.00	0.02	0.98
2	0.72	0.28	0.00	0.80	0.20	0.00
3	0.64	0.36	0.01	0.50	0.38	0.12
4	0.23	0.73	0.05	0.07	0.81	0.12
5	0.82	0.18	0.00	0.55	0.26	0.18

Group	Elevation					
	Old			New		
	L	M	H	L	M	H
1	0.02	0.13	0.85	0.13	0.24	0.63
2	0.00	0.00	1.00	0.00	0.15	0.85
3	0.89	0.11	0.00	0.81	0.19	0.00
4	0.00	0.14	0.86	0.06	0.28	0.67
5	0.56	0.44	0.00	0.60	0.37	0.03

Group	Along channel length					
	Old			New		
	L	M	H	L	M	H
1	0.97	0.03	0.00	0.90	0.10	0.00
2	1.00	0.00	0.00	0.85	0.15	0.00
3	0.00	0.14	0.86	0.09	0.23	0.68
4	0.90	0.10	0.00	0.98	0.02	0.00
5	0.81	0.19	0.00	0.87	0.12	0.01

Group	Along channel slope					
	Old			New		
	L	M	H	L	M	H
1	0.00	0.23	0.77	0.06	0.36	0.58
2	0.17	0.06	0.77	0.11	0.33	0.56
3	0.98	0.02	0.00	0.93	0.07	0.00
4	0.00	0.01	0.99	0.00	0.18	0.82
5	0.73	0.27	0.00	0.66	0.32	0.02

From the final classification results revealed that the sub-basins in Group 1 have the common features of large area, medium to high elevation, short along channel length, and medium to high along channel slope. The sub-basins in this group mainly locate in the upstream of the river. They are upstream large area sub-basins.

The sub-basins in Group 2 have the common features of low to medium area, high elevation, short along channel length, and medium to high along channel slope. The sub-basins in this group mainly locate in the medium and downstream of the river. They are upstream small area sub-basins.

The sub-basins in Group 3 have the common features of low to medium area, low to medium elevation, medium to high channel length, and very low along channel slope. The sub-basins in this group mainly locate in the downstream of the river. They are downstream long channel sub-basins, which locate on plain.

The sub-basins in Group 4 have the common features of medium area, medium to high elevation, very short along channel length, and high along channel slope. The sub-basins in this group mainly locate in the upstream of the river.

The sub-basins in Group 5 have the common features of low to medium area, low to medium elevation, short along channel length, and low to medium along channel slope. The sub-basins in this group mainly locate in the midstream and downstream of the river. They are mid-downstream short channel sub-basins, which locate on plain.

4.4 Summary

This chapter presents a modified ART mapping approach by integrating fuzzy interface and rule-based operation with ART/ART mapping model to form the IRFAM approach. Five ART modules were employed in the approach to carry out an unsupervised learning for cluster centroid calculation, supervised learning for criteria combination classification, and a supervised learning module for fuzzified original input classification.

Since uncertainties and complexities are two major challenges faced by the traditional classification methods, the introduction of fuzzification and rule-based operation can effectively overcome these challenges. The developed IRFAM approach can generate full criteria combinations to match the input patterns and use the rule-based operation to screen the matched patterns into the target groups. The approach can efficiently handle the classification for the input patterns with a high degree of complex features and wide ranges of distributions.

The IRFAM approach is more powerful than the traditional clustering classification methods in handling the juncture problem. Because the datasets in the juncture are quite similar to each other, it is usually difficult to accurately classify them by clustering classification systems. By using full fuzzy combination, the IRFAM is able to classify the datasets with high similarities. The more levels of fuzzy membership functions used for the approach, the more accurate the classification results obtained from the approach.

Each criteria combination in the IRFAM approach represents a type of realistic meaning of sub-basins inputs. In the case study, once the input sub-basins are classified into a certain criteria combinations which mean that these sub-basins have similar features presented by the criteria combination, e.g. area, shape, and elevation. So the classification results have more practical meanings which would be helpful for the related practice such as watershed modeling and management.

CHAPTER 5: COMPARISON OF TSAM AND IRFAM APPROACHES

5.1 Statistical Analysis

Two classification approaches, TSAM and IRFAM, have been developed for supporting watershed modeling and management. This chapter focuses on discussion about the difference between the two approaches.

Based on the classification results of both approaches the mean and standard deviation of each parameter for each group were calculated. The results are shown in Tables 5.1 and 5.2.

By comparison of mean values and standard deviation values from the TSAM and the IRFAM results, it is observed that Group 1 in the TSAM is similar to Group 1 in the IRFAM, Group 2 in the TSAM is similar to Group 4 in the IRFAM, Group 3 in the TSAM is similar to Group 3 in the IRFAM, Group 4 in the TSAM is similar to Group 2 in the IRFAM, and Group 5 in the TSAM is similar to Group 5 in the IRFAM.

The higher standard deviation values show that the dispersion degree of area in most the groups in TSAM is higher than that in the IRFAM, which means the effect of area is more significant on watershed classification by the IRFAM then by the TSAM. The dispersion degree of elevation in all of the groups in IRFAM is higher than which in the TSAM, which means the effect of elevation is more significant on watershed classification by the TSAM then by the IRFAM. The effects of along channel length and along channel slope are similar in both the TSAM and the IRFAM.

Normal distribution approximation was used to examine the distribution of the parameters in each group. Due to the central limit theorem, normal distribution is an important model of quantitative phenomena in the natural and behavioral sciences (Berman, 1971). The probability density function of normal distribution is:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{5.1}$$

In parameter estimation, μ can be approximated as the mean of the data sets, and σ can be approximated as the standard deviation of the data sets. Based on the mean values and standard deviations in Tables 5.1 and 5.2, the distribution of each parameter in each group can be approximated. The distributions are shown in Figures 5.1 to 5.8. These figures also indicate more significant effect of area on sub-basins classification by the IRFAM, more significant effect of elevation on sub-basins classification by the TSAM, and similar effects of along channel length and along channel slope in both approaches. Furthermore, the along channel length has considerably low contribution in helping classification by using either approach. As shown in Figures 5.5 and 5.6, some of the along channel length distributions are too close which means that these groups are quite similar in along channel length, while the other distributions have very large σ which means the distributions of along channel length in these group are too fragmented to characterize the group.

Table 5.1 Mean value and standard deviation for the TSAM classification results

Group	Area		Elevation		Along channel length		Along channel slope	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
1	70.47	26.89	184.10	15.65	1.03	0.27	0.49	0.09
2	26.14	22.23	83.00	16.36	2.03	4.12	0.16	0.08
3	11.58	7.17	73.79	9.58	6.82	5.72	0.06	0.05
4	14.36	8.11	171.74	19.13	0.88	0.35	0.53	0.21
5	9.72	13.00	131.00	18.28	1.53	0.97	0.32	0.13

Table 5.2 Mean value and standard deviation for the IRFAM classification results

Group	Area		Elevation		Along channel length		Along channel slope	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
1	62.36	24.87	154.94	42.52	1.09	0.40	0.42	0.13
2	7.70	3.08	171.63	19.78	1.14	0.54	0.47	0.25
3	16.56	15.87	76.81	12.96	8.69	7.25	0.08	0.06
4	21.26	4.46	152.88	27.29	0.86	0.28	0.52	0.12
5	17.06	19.60	88.45	24.42	1.18	0.75	0.16	0.09

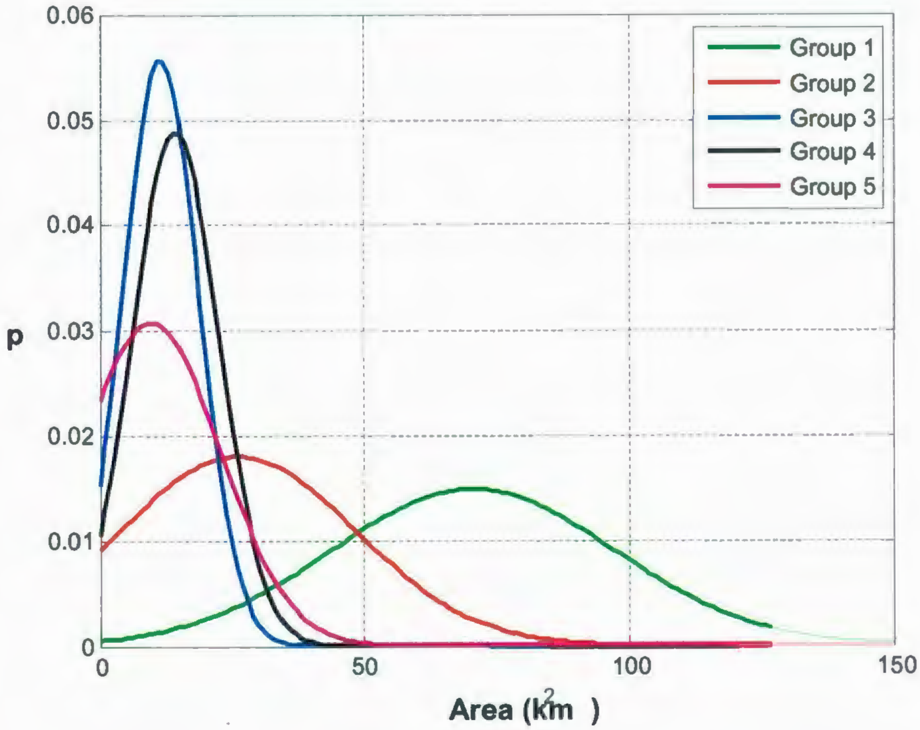


Figure 5.1 Distribution for the feature of "Area" in the TSAM classification results

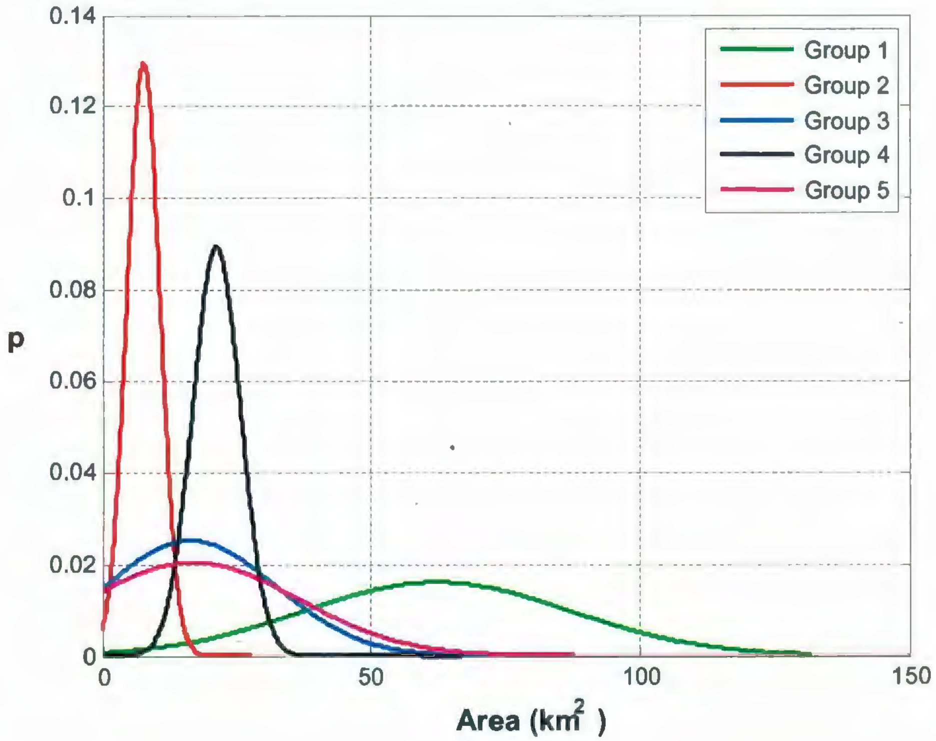


Figure 5.2 Distribution for the feature of "Area" in the IRFAM classification results

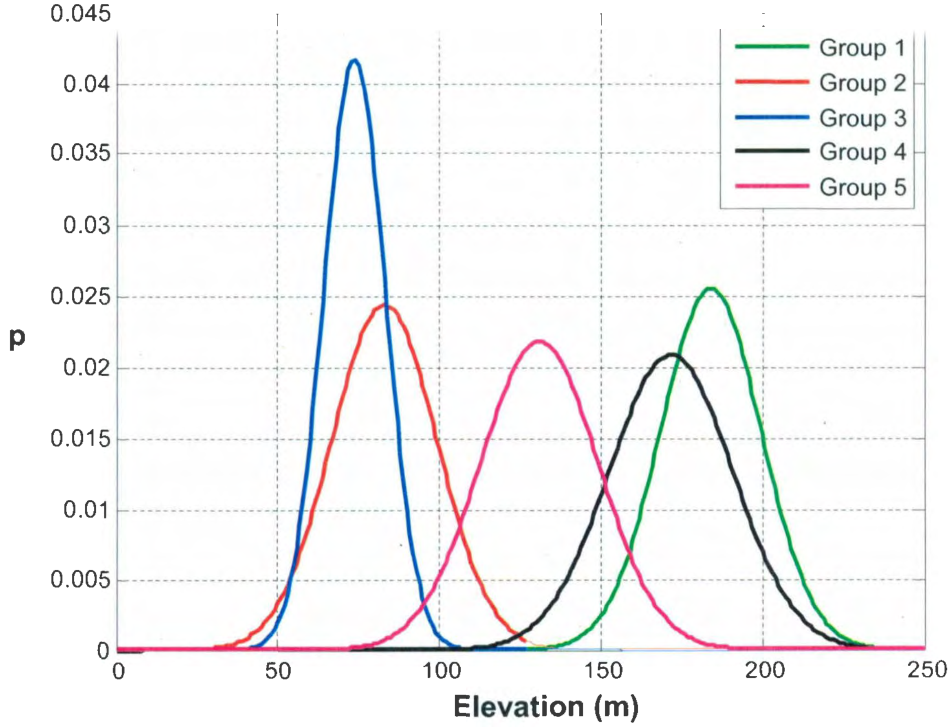


Figure 5.3 Distribution for the feature of "Elevation" in the TSAM classification results

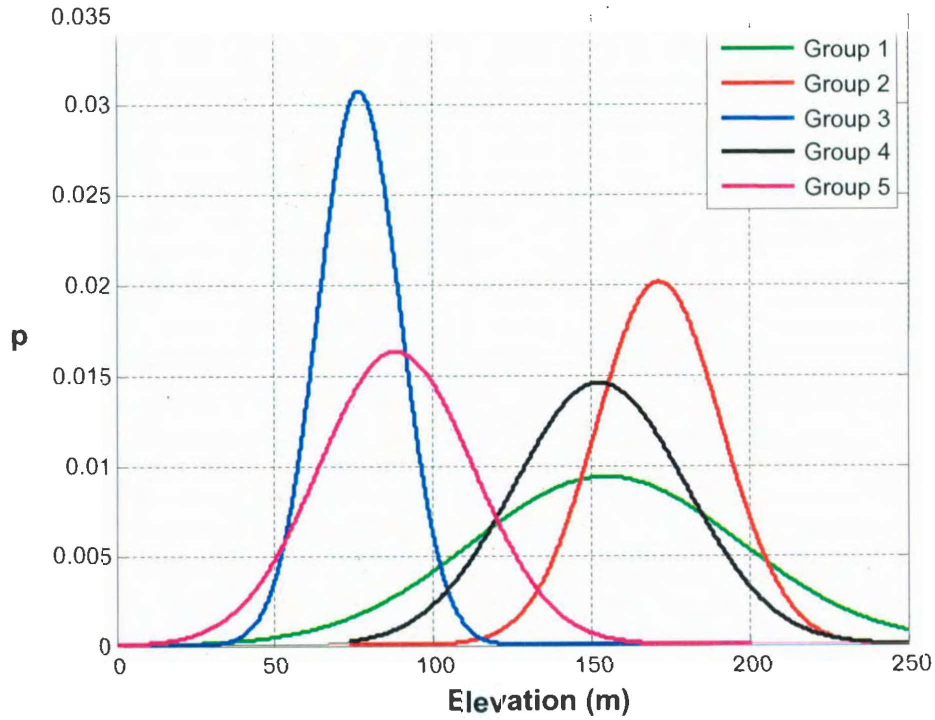


Figure 5.4 Distribution for the feature of "Elevation" in the IRFAM classification results

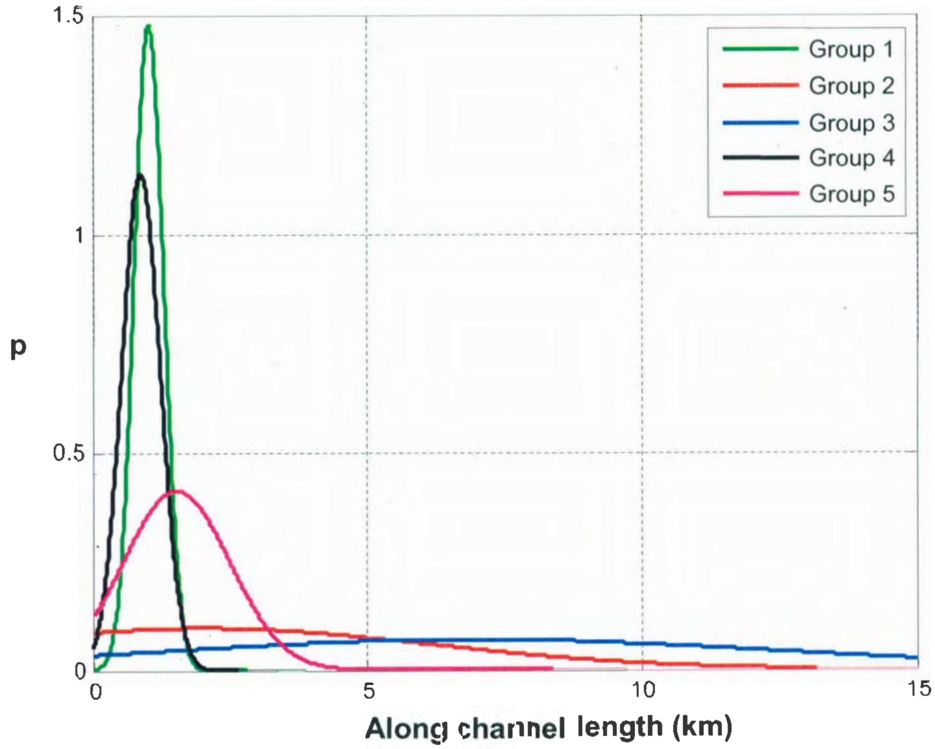


Figure 5.5 Distribution for the feature of "Along channel length" in the TSAM classification results

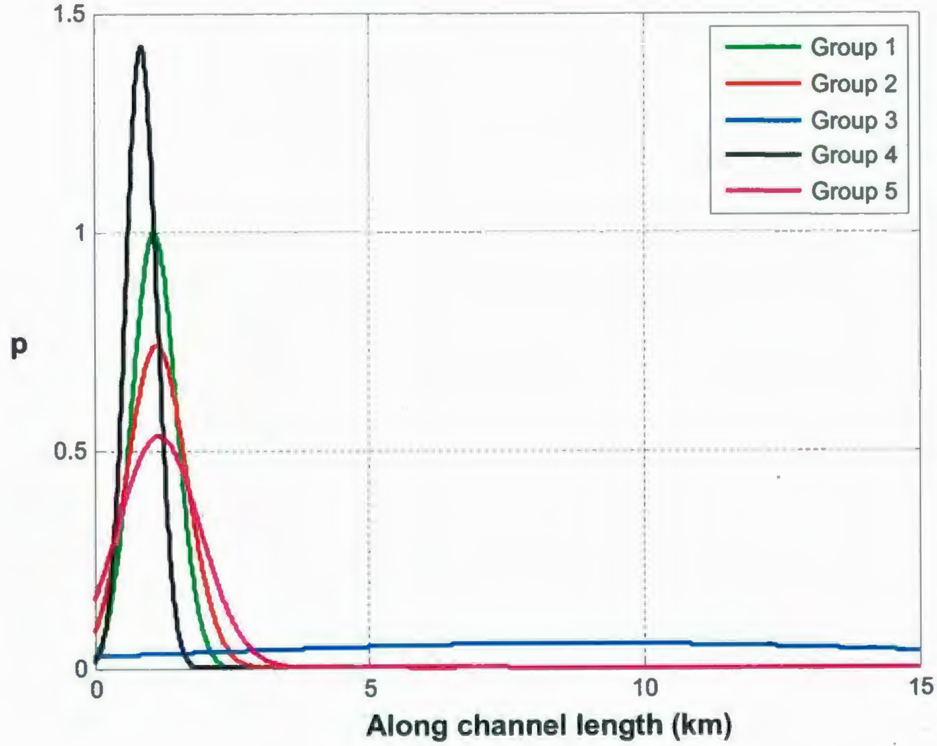


Figure 5.6 Distribution for the feature of "Along channel length" in the IRFAM classification results

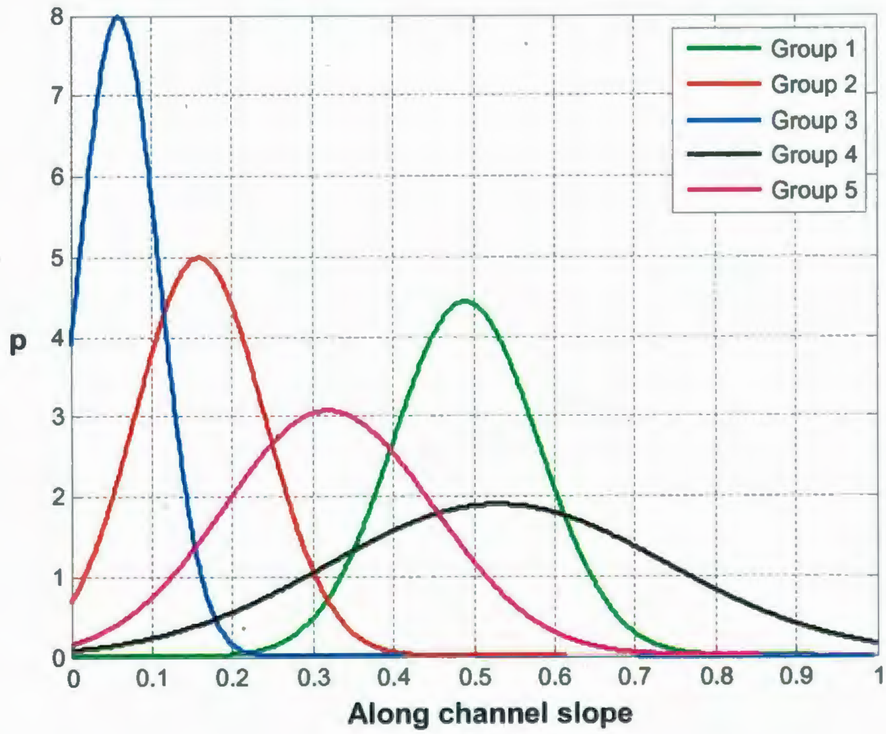


Figure 5.7 Distribution for the feature of "Along channel slope" in the TSAM classification results

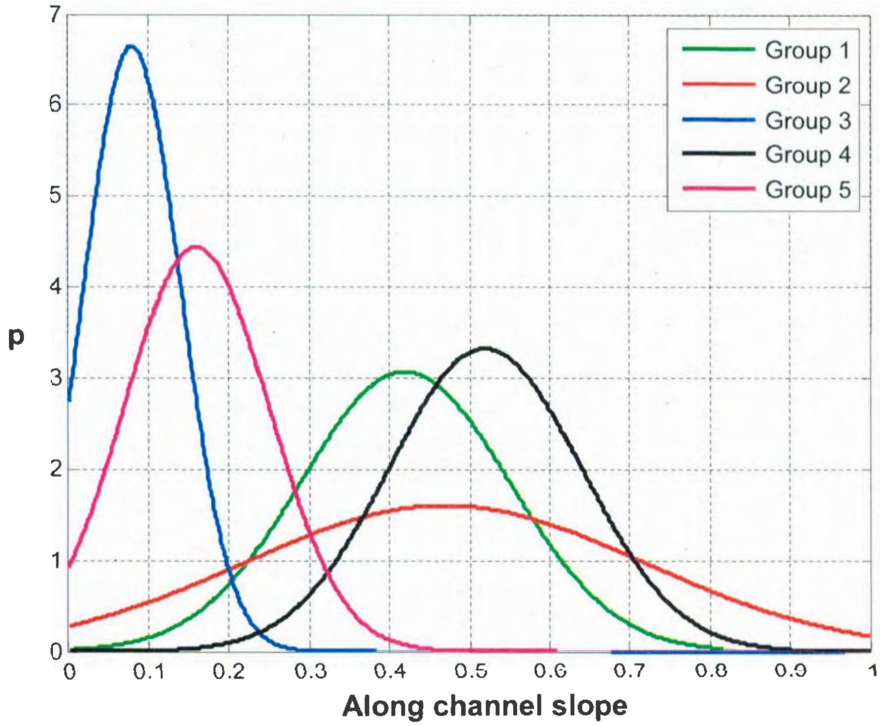


Figure 5.8 Distribution for the feature of "Along channel slope" in the IRFAM classification results

5.2 Realistic Analysis

The classification results show that the 91 sub-basins were classified and 1 sub-basin (#4) was left behind by the TSAM approach, while all the 92 sub-basins were classified by the IRFAM approach. The TSAM approach only use one combination for one criteria, which means that if the number of input features increases and the variation of the features is significant, the approach has to generate a large number of criteria combinations to make sure all of the input patterns can be properly classified. Consequently, a very large number of output groups will be produced. Furthermore, if the input patterns need to be classified into only small number of groups, some patterns will be left behind and not be classified into any groups. Comparatively, the IRFAM approach generates the full criteria combination to matching the input patterns and uses the rule-base operation to screen the matched patterns into the preset target groups, which can efficiently handle the input patterns with high degree complexities of features and wide range of their distributions.

Almost all of the relocated sub-basins are located in the junctures of the groups. Sub-basin number 4 cannot be classified by the TSAM because its features are quite different from all of the groups compared with the other sub-basins. However, they can still be properly classified by the IRFAM because the whole data area is divided by full fuzzy combinations which have been classified into certain groups beforehand. Each criteria combination represents certain common features of sub-basins, e.g., small area, high elevation, short along channel length, and large along channel slope. Once the input sub-basins are classified into a certain combination, it means that the input sub-basins have the common features.

Figure 5.9 shows the demonstration of rule-base screening of the IRFAM approach. Each grid in the figure refers to a fuzzy criteria combination that can be used to denote a type of sub-basin characteristics. From the figure it can be seen that point *a* is classified to Group *B* based on the clustering theory which is used by the TSAM. However, point *a* has the same characteristics as the grid that is classified to Group *A*, therefore it is more reasonable that point *a* is classified to Group *A* based on its physical meaning. Furthermore, point *b* cannot be classified to Group *A* or *B* based on the clustering theory which is used by the TSAM. However, point *b* has the same characteristics as the grid that is classified to Group *A*, therefore, point *b* can be classified to Group *A* based on IRFAM. The situation of sub-basin #4 is similar to point *b*, which is the reason why it can be classified by IRFAM.

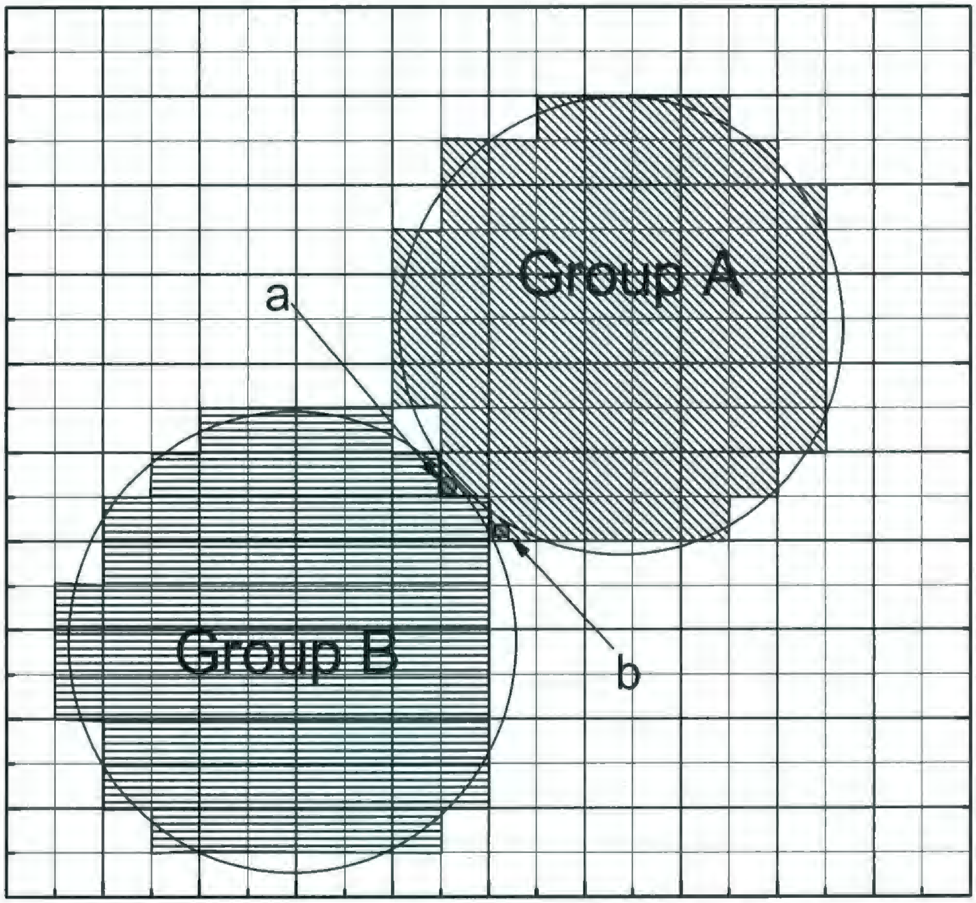
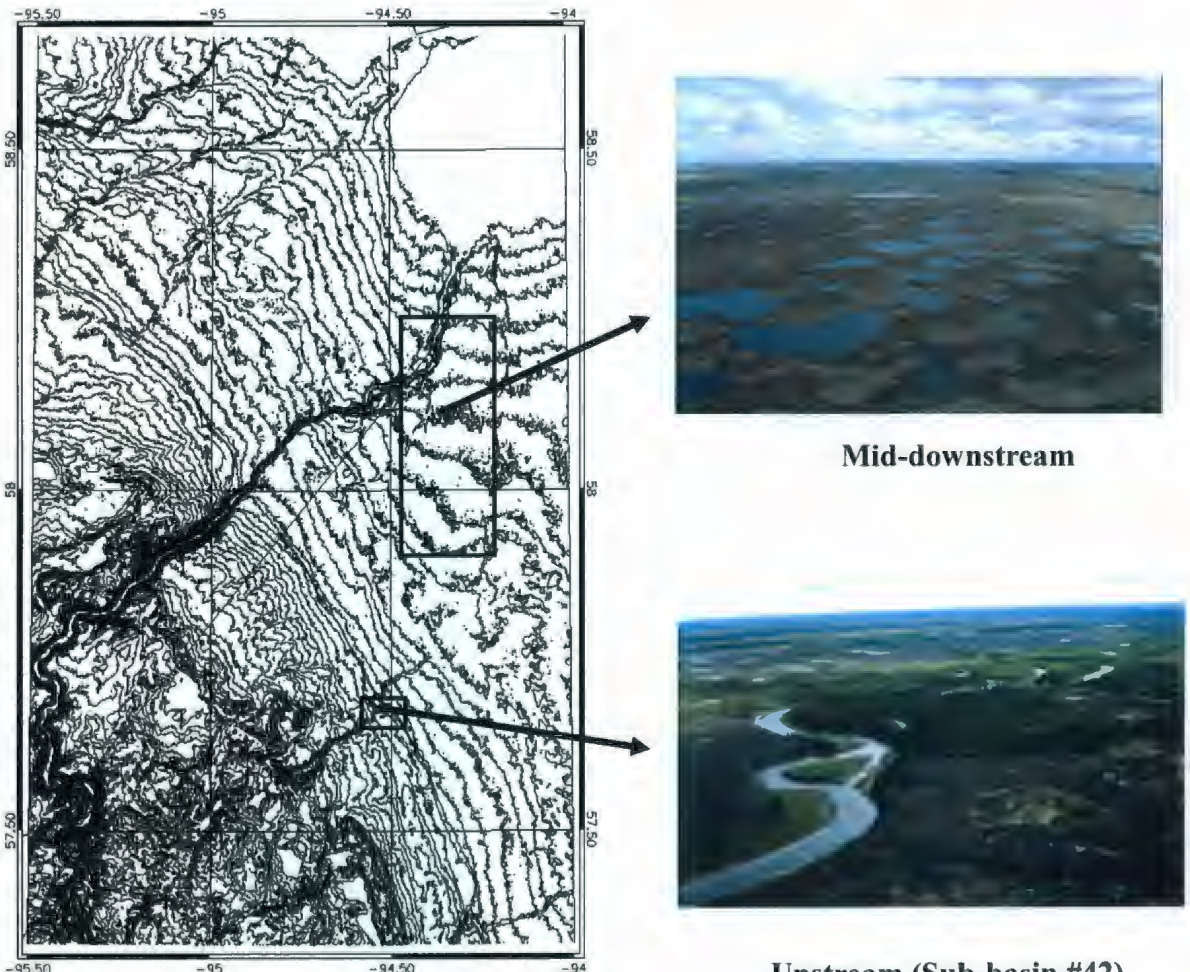


Figure 5.9 Demonstration of rule-base screening of the IRFAM approach



Mid-downstream

Upstream (Sub-basin #42)

Figure 5.10 On-site photos of mid-downstream and Sub-basin #42

In the case study, Sub-basin #42 was classified into Group 2 by the TSAM. The sub-basins in this group are mainly located in the middle to downstream of the river. In the IRFAM approach, it was classified into Group 1. The sub-basins in this group are mainly located in the upstream of the river. Figure 5.10 shows the on-site photo of mid-downstream and Sub-basin #42. View from the photos, it can be observed that the land cover in Sub-basin #42 is quite different form that in mid-downstream where tundra cover most of the land area. The vegetation is mainly forest which is common in the upstream area. Furthermore, there are extensive ponds stretching in the min-downstream area, but only a few in Sub-basin #42 which presents a common feature in the upstream area. Therefore, it is more reasonable that Sub-basin #42 should be classified to Group 1 in the IRFAM than to Group 2 in the TSAM. Similarly, it can be found that the other relocations of sub-basins are more reasonable from the TSAM to the IRFAM. This is because the TSAM only focus on the statistical information from data, but the IRFAM presents more realistic by introducing rule-base screening into the approach. This also partially contributes to the reason why the dispersion degree of classification result by the IRFAM is reasonably higher than that by the TSAM.

5.3 Summary

From the above discussion, the IRFAM approach has advantages in handle uncertainty and complexity than the TSAM approach. Furthermore, classification results from IRFAM appear more reasonability associating with the real-world condition. However, the IRFAM approach requires criteria for formulating membership function. The criteria are normally obtained from literature or questionnaire survey, which could introduce uncertainties into the approach and lead to errors in the result. The TSAM approach can efficiently solve these problems by using ART unsupervised classification and centriod determination subsystem in the first stage to automatically generate criteria for the ART mapping supervised classification in the second stage. In addition, the IRFAM approach needs longer time for classification than the TSAM approach. Therefore, in the case that there is not enough information for determining membership functions, at the same time there is a time limitation for the classification mission, the TSAM is a better choice than the IRFAM. Otherwise, the IRFAM could provide more accurate classification results than the TSAM.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

6.1 Summary

This dissertation research has focused on the development of an enhanced adaptive resonance theory (ART) mapping classification system to more efficiently and accurately classify watersheds with high level of complexities and uncertainties by incorporating fuzzy set theory, rule-based operation with multitier ART mapping techniques. A brief summary of the research is given as follows:

As one of the two new approaches in the developed system, the modified ART mapping (TSAM) approach has been developed by integrating three ART modules into the system classification process in two stages to form an unsupervised learning module for cluster centroid calculation and a supervised learning module for normalized input classification. The major functions of the three ART modules are: ART_1 is used for processing unsupervised classification for the normalized original input and generating unsupervised classified groups; ART_{2a} and ART_{2b} are used in ART mapping to comparing the combinations determined in the first stage and the normalized original inputs, and then classify them. The TSAM approach has been tested through a real-world case study conducted in the Deer River watershed, northern Manitoba, Canada. The results showed that the 91 watershed sub-basins could be properly classified into 5 preset target groups in the case of vigilance equals to 0.7. Only one sub-basin #4 could not be classified into any group by the TSAM approach..

The second new approach in the classification system is an integrated rule-based fuzzy adaptive resonance theory mapping (IRFAM) approach by integrating fuzzy interface and rule-based operation with ART/ART mapping techniques. Five ART modules are included to carry out the unsupervised learning for cluster centroid calculation and supervised learning for criteria combination and fuzzified original input classification. The IRFAM approach includes three subsystems: 1) centroid determination for locating the centroids of the expected target groups by unsupervised ART; 2) criteria combination for generating the fuzzy criteria combinations; and 3) classification for classifying the original inputs which have been converted into fuzzy set form. The same watershed has been used for testing this approach. The results showed that all the 92 watershed sub-basins could be properly classified into 5 preset target groups in the case of vigilance ρ equals to 0.7.

Comparison between the two classification approaches has been conducted based on the case study from both statistical and realistic perspectives. The statistical analysis indicated more significant effects of area on classification results by the IRFAM, more significant effects of elevation on classification results by the TSAM, and similar effects of along channel length and slope in both approaches. Furthermore, the along channel length had considerably low contribution to the classification by using either approaches. The realistic analysis indicated that the IRFAM approach presented remarkable advantages in handling both uncertainty and complexity existing in the watershed characteristics and produced reliable and accurate classification results. Generally, in the case that there are not sufficient information for generating fuzzy membership functions, the TSAM could be a better choice than the IRFAM from a feasibility perspective; otherwise, the IRFAM could provide more accurate classification results than the TSAM.

6.2 Research Achievements

A two-stage adaptive resonance theory mapping (TSAM) approach has been developed. It presents high automation and efficiency in generating the criteria for supervised learning and completing the classification process and low requirements in data inputs.

An integrated rule-base fuzzy adaptive resonance theory mapping (IRFAM) approach has been developed. It has advantages in effectively reflecting system uncertainties and complexities into the classification process leading to more reasonable results associating with the real-world conditions.

Through the application of the developed system to the Deer River watershed, the two new classification approaches have been tested and compared with each other for examining their efficiency and feasibility. The findings provide evidences that the developed system is able to meet the needs of more efficient and reliable approaches of watershed classification to deal with complex and uncertain features. The approaches should provide powerful tools for supporting the decision making and practice in the areas of watershed monitoring, modeling and management.

6.3 Recommendations for Future Research

The TSAM approach can provide high automation and efficiency in achieving classification tasks for a complex system, but its classification capacity would be impaired and decreasing when the level of uncertainties in the system increase. More efforts are needed in the future research to improve its ability in handling uncertainties.

The IRFAM approach is able to tackle both uncertainties and complexities through the integration of fuzzy sets and rule-based operation; however, limitations would occur in formulating membership functions in solving practical problems when sufficient information is not available. Research on how to more efficiently utilizing limited information and generating membership functions can be further conducted. Combination with other uncertainty analysis methods such as interval approach would be considered.

In the case study, only four features (i.e., area, elevation, along channel length, and along channel slope) which could mainly present the watershed hydrological characteristics were used. These features are assumed to be independent. However, interrelationships among these features may exist such as between elevation and along channel slope, and the associated impacts on classification results can be evaluated. Furthermore, more features can be considered in future research.

REFERENCES

Kartike B. and Majumder K. L. (1995), "An expert system for land cover classification", IEEE Traansaction on Geoscience and remote sensing, 33(1), 59-66.

Soheila S. H., Majed B., Mojtaba P., and Mostafa E. (2007), "GIS classification assessment for mapping soils by satellite images", Middle East Spatial Technology, 4th Conference and Exhibition, December 10 -12, 2007.

Running S. W., Loveland T. R., Pierce L. L., Nemani R. R., and Hunt E. R. (1995), "A remote sensing based vegetation classification logic for global land cover analysis", *J.* Remote sensing of environment, 51, 39-48.

Sauder J., McMahan J. B., and Weber K. T., (2003), "Fuzzy classification of heterogeneous vegetation in a complex arid ecosystem", African Journal of Range and Forage Science, 20(2), 126.

Lucas L. A., Centeno T. M., and Delgado M. R., (2008), "General type-2 fuzzy classifiers to land cover classification", Proceedings of the 2008 ACM symposium on Applied computing, Fortaleza, Ceara, Brazil, 2008.

Daniel L. C., (1993), "Artificial Neural Networks for Land cover Classification and Mapping", International Journal of Geographical Information Systems, 7(2), 173-186.

Gopal S., Woodcock C. and Strahler A. (1999), "Fuzzy ARTMAP classification of global land cover from the 1 degree AVHRR data set", Remote Sensing of Environment, 67, 23-43.

Giles M. F., (1995), "Land Cover Classification by an Artificial Neural Network with

Ancillary Information", International Journal of Geographical Information Systems, 9(5), 527-542.

Lee S. G., Jong G. H. Kwang H. C., Jae Y. S., Hee H. L., Michio M. and Kageo A. (1999), "A neuro-fuzzy classifier for land cover classification", Fuzzy Systems Conference Proceedings, August 11-25, 1999, Seoul, Korea.

Han J., Lee S., Chi K., and Ryu K., (2002), "Comparison of neuro-fuzzy, neural network, and maximum likelihood classifiers for land cover classification using IKONOS multispectral data", Geoscience and Remote Sensing Symposium, 6, 24-28.

Thandaveswara B. S. and Sajikumar N. (2000), "Classification of River Basins Using Artificial Neural Network", J. Hydrologic Engr., 5(3), 290-298.

Grossberg S. (1976), "Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors", Biological Cybernetics, 23, 121-134.

Grossberg S. (1976), "Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions", Biological Cybernetics, 23, 187-202.

Grossberg S. (1980), "Intracellular mechanisms of adaptation and self-regulation in self-organizing networks: The role of chemical transducers", Bulletin of Mathematical Biology, 42, 365-396.

Zadeh L. A., (1968), "Fuzzy algorithms", Information and Control, 12 (2), 94-102.

Carpenter G. A. and Grossberg S., (2003), "Adaptive resonance theory", ed. Arbib M. A., The Handbook of Brain Theory and Neural Networks, Second Edition, Cambridge, MA: MIT Press, 87-90.

Carpenter G. A. and Grossberg S. (1987), "Pattern Recognition by Self-organising Neural

Networks", Academic Press Inc., 399-410.

Carpenter G. A., and Grossberg S. (1987), "ART2: self-organization of stable category recognition codes for analogue input patterns", Appl. Opt. 26, 4919-4930.

Carpenter G., and Grossberg S. (1991), "Pattern Recognition by Self-Organizing Neural Networks", MIT Press, Cambridge, MA.

Carpenter G., Gjaja M., Gopal S., and Woodcock C. (1997), "ART networks in remote sensing", Journal of IEEE Trans. Geosci. Remote Sens. 35(2), 308-325.

Carpenter G. A., Grossberg S., and Rosen D.B., (1991), "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system", Neural Networks, 4, 759-771.

Carpenter G. A., Grossberg S., Markuzon N., Reynolds J. H., and Rosen D. B., (1992), "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps", IEEE Transactions on Neural Networks, 3, 698-713.

Carpenter G. A., and Grossberg S. (1987), "A massively parallel architecture for a self-organizing neural pattern recognition machine", Comput. Vision Graphics Image Process. 37, 54-115.

Carpenter G.A. (1997), "Distributed learning, recognition, and prediction by art and artmap neural networks", Neural Networks, 10, 1473-1494.

Simpson P.K. (1992), "Fuzzy min-max neural networks. I. Classification", IEEE Trans. Neural Networks, 3, 776-786.

Bock H. H. (1993), "Classification and clustering: problems for the Future", new Approaches in Classification and Data Analysis, eds. Diday E., et al., Berlin: Springer, 3-24.

Diday E., and Govaert G. (1977), "Classification Automatique avec distances Adaptatives", R.A.I.R.O. informatique Computer Science, 11, 329-349.

Diday E. and Simon J. J. (1976), "clustering Analysis", digital Pattern Recognition, ed. Fu K. S., Berlin: Springer, 47-94.

Everitt B. (2001), "Cluster Analysis", Halsted, New York.

Gordon A. D. (1999), "Classification Automatique et Distances Adaptatives", These de eme cycle, Mathematique appliqué, Universite Paris VI.

Jain A. K., Murty M. N., and Flynn P. J. (1999), "Data Clustering: A Review", ACM Computing Surveys, 31, 264-323.

Spaeth H. (1980), "Cluster Analysis Algorithms", Wiley, New York.

McCulloch W. S. and Pitts W. (1943), "A Logical Calculus of the Ideas Immanent in Nervous Activity", bulletin of Mathematical biophysics, 5, 115-133.

Rosenblatt F. (1962), "Principles of Neurodynamics", Spartan, New York.

Minsky M. L. and Papert S. A. (1969), "Perceptrons", MIT Press, Cambridge, Massachusetts.

Hopfield J. J. (1982), "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", in proceedings of the National Academy of sciences, 79, 2554-2558.

Hopfield J. J. and Tank D. W. (1985), "Neural Computation of Decisions in Optimization Problems", Biological Cybernetics, 52, 141-152.

Rumelhart D. E., McClelland J. L., and the PDP Research Group (1986), "Parallel Distributed Processing", MIT Press, Cambridge, Massachusetts.

Zadeh L. A. (1965), “Fuzzy sets”, Information and Control, 8, 338-352.

Buckley J. J., hayashi Y., and Czogala E., “On the Equivalence of Neural Networks and Fuzzy Expert Systems”, in Proceedings of International Joint Conference on Neural Networks, Baltimore, June 1992, 2, 691-695,.

Wang L. X. and Mendel J. M. (1992), “Generating Fuzzy Rules by Learning from Examples”, IEEE, Transactions on systems, Man, and Cybernetics, 26(12), 1182-1191.

Lin C. T. and Lee C. S. G. (1991), “Neural-Network-Based Fuzzy Logic Control and Decision System”, IEEE Transactions on Computers, 40(12), 1320-1336.

Katayama R., Kajitani Y., Fujiyama K., and Nishida Y. (1993), “Fuzzy Modeling Using the Gradient Method with Constraints on the Shape Parameters of membership Functions”, Japanese Journal of Fuzzy Theory and Systems, 5(6), 951-969.

Goodace R., Kell D. B., and Bianchi G. (1992), “neural networks and olive oil”, Nature, 359, 594.

Groman R. and Sejnowski T. (1988), “Analysis of hidden units in a layered network trained to classify sonar targets”, Neural Networks, 1, 75-89.

Aires F., Prigent C., and Rossow W. B. (2004), “neural network uncertainty assessment using Bayesian statistics: A remote sensing application”, Neural Computation, 16(11), 2415-2458.

Abe S. (2001), “Pattern classification: Neuro-fuzzy methods and their comparison”, Springer Verlag.

Acharyya M., De R. K., and Kundu M. K. (2003), “Segmentation of remotely sensed images using wavelet features and their evaluation in soft computing framework”, IEEE Transactions on Geoscience and Remote Sensing, 41(12), 2900-2905.

Anthony M., and Bartlett P. L. (1999), "Neural network learning: Theoretical foundations", Cambridge University Press.

Asuncion A., and Newman D. (2007), "UCI machine learning repository", University of California, Irvine, School of Information and Computer Sciences.

Baldi P. F., and Homik K. (1995), "Learning in linear neural networks: A survey", IEEE Transactions on Neural Networks, 6(4), 837-858.

Baraldi A., Binaghi E., Blonda P., Brivio, P. A., and Rampini A. (2001), "Comparison of the multilayer perceptron with neuro-fuzzy techniques in the estimation of cover class mixture in remotely sensed data", IEEE Transaction on Geoscience and Remote Sensing, 39(5), 994-1005.

Boskovitz, V., and Guterman,H.(2002), "An adaptive neuro-fuzzy system for automatic image segmentation and edge detection", IEEE Transaction on Fuzzy Systems, 10(2), 247-262.

Card D. H. (1982), "Using known map category marginal frequencies to improve estimates of thematic map accuracy", Photogrammetric Engineering and Remote Sensing, 48(3), 431-439.

Congalton R. G. (1991), "A review of assessing the accuracy of classifications of remotely sensed data", Remote Sensing of Environment, 37(1), 35-46.

Davies D. L., and Bouldin D. W. (1979), "A cluster separation measure", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1(4), 224-227.

Gamba P., and Dellacqua F. (2003), "Increased accuracy multiband urban classification using a neuro-fuzzy classifier", International Journal of Remote Sensing, 24(4), 827-834.

Ghosh A., Meher S. K., and Shankar B. U. (2008), "A novel fuzzy classifier based on product aggregation operator", Pattern Recognition, 41(3), 961-971.

Ghosh A., Pal N. R., and Pal S. K. (1993), "Self-organization for object extraction using a multilayer neural network and fuzziness measures," IEEE Transaction on Fuzzy Systems, 1(1), 54-69.

Haykin S. (1997), "Neural networks: A comprehensive foundation", (2nd ed.). Prentice Hall. IRS Data Users Hand Book (1989). Document no. IRS/NRSA/NDC/HB-02/89, NRSA, Hyderabad, India.

Keller J. K., and Hunt D. J. (1985), "Incorporating fuzzy membership functions in to the perceptron algorithm", IEEE Transaction on Pattern Analysis Machine Intelligence, 7(6), 693-699.

Kuncheva L. I. (2000), "Fuzzy classifier design", Springer-Verlag.

Ishibuchi H., Okada H. and Tanaka,H (1994), " Neural networks with interval weights for nonlinear mapping of interval vectors", IEICE Transactions on Information and Systems, 409-417.

Marinai S., Gori M., and Soda G. (2005), "Artificial neural networks for document analysis and recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(1), 23-35.

Mitra P., Shankar B. U., and Pal S. K. (2004), "Segmentation of multispectral remote sensing images using active support vector machines", Pattern Recognition Letters, 25(12), 1067-1074.

Pal S. K., and Ghosh A. (1996), "Neuro-fuzzy computing for image processing and pattern recognition", International Journal of Systems Science, 27, 1179-1193.

Pal S. K., Ghosh A., and Kundu M. K. (Eds.) (2000), "Soft computing for image processing", Heidelberg: Physica-Verlag.

Pal S. K., Ghosh A., and Shankar B. U. (2000), "Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation", International Journal of Remote Sensing, 21(11), 2269-2300.

Pal S. K., and Majumder D. D. (1977), "Fuzzy sets and decision making approaches in vowel and speaker recognition. IEEE Transaction on Systems, Man and Cybernetics, 7(1), 625-629.

Pal S. K., and Majumder D. K. D. (1986), "Fuzzy mathematical approach to pattern recognition", Wiley Eastern Ltd.

Pal S. K., and Mitra S. (1999), "Neuro-fuzzy pattern recognition: Methods in soft computing", New York, NY, USA: John Wiley & Sons, Inc.

Pedrycz W. (1990), "Fuzzy sets in pattern recognition: Methodology and methods", Pattern Recognition, 23(1), 121-146.

Qiu F., and Jensen J. R. (2004), "Opening the black box of neural networks for remote sensing image classification", International Journal of Remote Sensing, 25(9), 1749-1768.

Richards J. A., and Jia X. (2006), "Remote sensing digital image analysis: An introduction", Heidelberg: Springer-Verlag.

Ripley B. (1996), "Pattern recognition and neural networks", Cambridge University Press.

Rumelhart D. E., Hinton G. E., and Williams, R. J. (1986), "Learning representations by back-propagating errors", Nature, 323(6088), 533-536.

Tso B., and Mather P. M. (2001), "Classification methods for remotely sensed data", London: Taylor and Francis.

Varshney P. K., and Arora M. (2004), "Advanced image processing techniques for remote sensed hyperspectral data", Germany: Springer-Verlag.

Zhang G. P. (2000), "Neural networks for classification: A survey", IEEE Transactionson Systems, Man, and Cybernetics (Part C: Applications and Reviews), 30(4), 451-462.

Zurada J. M. (1992), "Introduction to artificial neural systems", PWS Publishing Company.

Berman S. M. (1971), "Mathematical Statistics: An Introduction Based on the Normal Distribution", Scranton, PA: Intext Educational Publishers.

Field S., Davey N., and Frank R. (1996), "Using neural networks to analyse software complexity", Australian Journal of Intelligent Information Processing System, 3(3), 14-31.

Friedman M. and Kandel A. (Eds) (1999), "Introduction to Pattern Recognition: Statistical, Structural, Neural, and Fuzzy Logic Approaches", Machine Perception and Artificial Intelligence, 32, 1-331.

Richard. O. D., Petter E. H., and David G. S. (Eds) (2001), "Pattern Classification", John Wiley & Sons, Inc., 2, 1-635.

Schurmann J. (Ed) (1996), "pattern Classification: a Unified View of Statistical and Neural Aproaches", John Wiley & Sons, Inc., 1-31, 330-357.

Lloyd, S. (2006), "Programming the Universe: A Quantum Computer Scientist Takes On the Cosmos", Alfred A. Knopf, Inc.

Kraskov A., Stögbauer H., Andrzejak G. R., and Grassberger P. (2003), "Hierarchical Clustering Based on Mutual Information", ArXiv q-bio.

Chris D. and Xiaofeng H. (2004), "K-means Clustering via Principal Component Analysis", Proceeding of international Conference Machine Learning (ICML 2004), 225-232. July 2004.

Erminio D. and Guerrisi F. (2002), "A Fuzzy Clustering Algorithm Based on the Analogy with Mechanical Physics: MeccFuz", Quality and Quantity, 36(3), 239-257.

Abrantes A. J. and Marques J. S. (1998), "A Method for Dynamic Clustering of Data", John N. Carter, Mark S. Nixon (Eds), Proceedings of the British Machine Vision Conference 1998, BMVC 1998, Southampton, UK.

Wang Y., Cai G. D., and Duan Q. (2006), "An Adaptive Dynamic Load Balancing Algorithm", Computer Engineering and Applications, 42(21).

Abe S. (Ed) (1997), "Neural Networks and Fuzzy Systems: Theory and Applications", Kluwer Academic Publishers, 1-216.

Dunne A. R. (Ed) (2007), "A Statistical Approach to Neural Networks for Pattern Recognition", John Wiley & Sons, Inc., Hoboken, New jersey, 1-30.

Patterson W. D. (Ed) (1996), "Artificial Neural Networks: Theory and Applications", Simon & Schuster Pte Ltd, 1-139.

Mandic D. and Chambers J. (2001), "Recurrent Neural Networks for Prediction: Architectures, Learning algorithms and Stability", Wiley.

Cybenko G. (1989), "Approximation by Superpositions of a Sigmoidal Function" , Mathematics of Control, Signals and Systems, 2, 303-314.

Funahashi K. (1989), "On the Approximate Realization of Continuous Mappings by Neural Networks", Neural Networks, 2(3), 183-192.

Hornik K., Stinchcombe M., and White H. (1989), "Multilayer Feedforward Networks are Universal Approximators", Neural Networks, 2(5) , 359-366.

Lee H. K. H. (2000), "Consistency of Posterior Distributions for Neural Networks", to appear in Neural Networks.

Qinlan J. R. (1986), "Introduction of decision trees", Machine Learning, 1, 86-106.

Freund Y. (1995), "Boosting a weak learning algorithm by majority", Information Computation, 121, 256-285.

Lu H. J., Setiono R., and Liu H. (1996), "Effect data mining using neural networks", IEEE Transaction on knowledge and data engineering, 8, 957-961.

Boser B. E., Guyon I. M., and Vapnik V. N. (1992), "A training algorithm for optimal margin classifiers", Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, ACM Press, 144-152.

Vapnik V. N. (1995), "The Nature of Statistical Learning Theory", Springer Verlag.

Grossberg S. (1976), "Adaptive pattern classification and universal recoding, 1: Parallel development and coding of neural feature detectors", Biological Cybernetics, 23, 187-202.

Chen J. L., Yeong J., and Tsay M. L. (1999), "Landscape Image Analysis Using Fuzzy Adaptive Resonance Theory", IEEE Transactions.

Tower G., and Shavlik J. W. (1993), "The Extraction of Refined Rules from knowledge-Based Neural Networks", Machine-Learning, 13.

Yoo J. H. (1993), "Symbolic Rule Extraction from Artificial Neural Networks", PhD thesis, Wayne State University.

Craven M. and Shavlik J. W. (1993), "Learning Symbolic Rules Using Artificial Neural Networks", Proceeding of the 10th International Conference on Machine Learning, Amherst.

Thrun S. B. (1994), "Extracting Symbolic Knowledge from Artificial Neural networks", Revised Version of Technical Research Report, Institut fur Informatik III, Universitat Bonn.

Lin C. T. and Lee C. S (1991), "Neural Network Based Fuzzy Logic Control and Decision System", IEEE Transactions on Computers, 40(12).

Palade V. V. A and Segal C. (1996), "Use of Genetic Algorithms for Tuning Fuzzy Models", The 1996 annuals of the University "Dunarea de Jos" of Galati, Fascicle nr. 3, Galati.

Yang M. S. and Yang J. H. (2008), "Machine-part cell formation in group technology using a modified ART1 method", European Journal of Operational Research, 188, 140-152.

Dubois D. and Prade H. (1988), "Fuzzy Sets and Systems", Academic Press, New York.

Xu R., Plessis du L., Damelin S. B., Sears M., and Wunsch D. (2009), "Analysis of Hyperspectral Data using Diffusion maps and Fuzzy Art", IJCNN.