ABSTRACT

YUAN, HUI. Development and Evaluation of Advanced Classification Systems using Remotely Sensed Data for Accurate Land-Use/Land-Cover Mapping. (Under the direction of Dr. Siamak Khorram.)

Our general objective in this research was to design, develop, implement, and evaluate advanced classification approaches for more accurate land-use/land-cover mapping using remotely sensed data. The overall research consists of three interrelated studies.

In the first study, we developed Simulated Annealing (SA) based classification systems for land cover mapping. SA has been shown to be able to overcome the local minimum problem that is typical with many unsupervised classification approaches. *Our hypothesis in this study was that SA-based classification systems could help overcome the local minimum problem in one of such approaches, K-means, and thus improve the classification performance*. Two SA based classification systems have been developed. The Single SA-based (S-SA) system was developed based on the standard SA algorithm. The Integrated SA-based (I-SA) system was developed by combining the standard SA algorithm and *K*-means into a two-level classification system. We have used Landsat Thematic Mapper (TM) images to test the suitability of the SA-based systems. Experimental results have demonstrated that the SA-based systems significantly improved the classification accuracy over that of the *K*-means algorithm when appropriate parameters were chosen. The I-SA system was shown to produce a satisfactory classification more efficiently than the S-SA system.

In the second study, we developed an automated Artificial Neural Network (ANN) classification system including two classification modules: 1) a supervised Multilayer Perceptron (MLP) neural network module, and 2) an unsupervised Kohonen Self-Organizing Mapping (KSOM) neural network module. In the KSOM network module, we incorporated SA random search procedures into the standard KSOM learning algorithm to reduce the likelihood to be trapped by local minima. *Two hypotheses have been tested in this study: 1) the ANN system developed here was suitable for land-use/land-cover classification, and 2) the KSOM network refined by SA could help overcome the local minima problem and improve the classification performance of the standard KSOM network.* To test these hypotheses, we applied the ANN system to land-use/land-cove classifications of a Landsat TM image using a MLP network and two KSOM networks. The ANN classification system has been shown to be a robust and well-suited classification system for land cover mapping. Experimental results demonstrated that the incorporation of SA improved the classification performance of the standard KSOM network. Among the three classifications, the supervised MLP network had superior classification performance than the unsupervised KSOM networks.

ANNs have been shown to have great potential to fuse multiple source data sets. In the third study, we performed a two-stage multisource classification using the ANN system. The KSOM module in the ANN system was adapted as an Automated Data Selector (ADS) in the first stage to automatically select training data for the supervised MLP classification in the second stage. *Two hypotheses were tested in this study: 1) the ADS adapted from the KSOM network could improve the quality of training sets and the multisource classification performance, and 2) the fusion of Landsat TM and SPOT*

images could increase the classification accuracy of the derived land categories. Our experimental results demonstrated that the multisource classification using the ADS developed in this study was significantly better than that using manual data collection. The fusion of the two images significantly improved classification accuracy as compared to single source image classification when adequate and reliable training data sets were selected using the ADS.

In summary, our studies demonstrate that the automated SA and ANN classification systems are feasible and robust for land-use/land-cover mapping using remotely sensed data.

DEVELOPMENT AND EVALUATION OF ADVANCED CLASSIFICATION SYSTEMS USING REMOTELY SENSED DATA FOR ACCURATE LAND-USE/LAND-COVER MAPPING

by

Hui Yuan

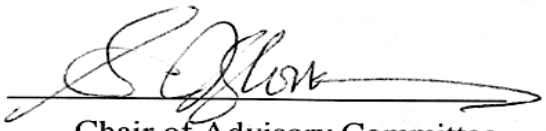
A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Forestry

Raleigh

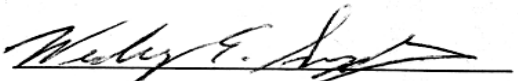
2002

Approved by

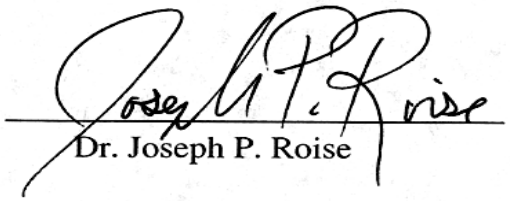


Chair of Advisory Committee
Dr. Siamak Khorram

Dr. Hugh A. Devine



Dr. Wesley E. Snyder



Dr. Joseph P. Roise

Dedication

To my parents and my husband

for their endless love, patience, and encouragement, which made it all possible.

Biography

Hui Yuan grew up in a small town of the mid-east of P. R. China. She first studied in No. 1 Middle school of ChuZhou in 1983 and then transferred to No.54 Middle School in Beijing, P. R. China. In September 1989 she attended NanJing Agricultural University and got her Bachelor of Arts in Information Management in July 1993. After graduation she took a job at the Information Center of Forestry Ministry (Currently Forestry Bureau) of P. R. China and worked as a computer engineer for three and half years. In August 1997, she started her Ph.D. study with specialties in Remote Sensing (RS), Geographic information systems (GIS), and spatial analysis at the North Carolina State University, Raleigh, NC under the direction of Professor Siamak Khorram. Upon completion of her degree, she plans to continue her career by utilizing her professional skills and expertise in environmental conservation and enhancement.

Acknowledgements

First of all I thank all my committee members, Dr. Siamak Khorram, Dr. Hugh Devine, Dr. Joseph Roise, and Dr. Wesley Snyder, for their supports and discussions throughout the development of this research. My first special gratitude goes to my advisor, Dr. Siamak Khorram, for his wisdom and enthusiastic encouragements to help me turn my dream into reality. I would also like to thank Dr. Snyder particularly for spending lots of time helping me clarify many technical issues in my dissertation. His input has greatly improved the completeness of my work. I owe a great debt to Dr. Xiaolong Dai, who taught me to walk the first step as a researcher and consistently provided me assistance throughout the five years of my degree program. I want to express my sincere gratefulness to Dr. Xinghe Yang from ERDAS IMAGINE, for his spiritual encouragements and technical supports. He had offered a lot of good suggestions to help me refine my dissertation.

In addition, I would like to take this chance to thank many graduate students and staff of the Center for Earth Observation for their friendly supports. Special credits should go to Joseph Knight for his great contribution of precious time hours in proofreading of many of my paper drafts. Thanks are also due to Cray Research, Inc. and North Carolina Supercomputing Center for their financial and technical support in my work. Without these people, I would not have made today possible.

Finally, with all my heart, I want to thank my parents and my husband. It is their unconditional love and understanding that support me to come so far.

Table of Contents

List of Tables viii

List of Figures ix

1. Introduction 1

1.1 Why advanced classification techniques are needed 2

1.2 Proposed classification systems and their applications 5

1.2.1 Study one: development of SA-based classification systems 6

1.2.2 Study two: development of an automated neural network classification system 8

1.2.3 Study three: development of a two-stage neural network-based multisource classification using multiple remotely sensed data 10

1.3 Dissertation Organization 12

References 13

2. Literature Review 16

2.1 Review of thematic land information extraction from remotely sensed data 16

2.1.1 Remotely sensed data and their applications for land cover classification 17

2.1.2 Classification algorithms and their applications for land cover classification using remotely sensed data 21

2.2 Review on Simulated Annealing (SA) and Applications for Land Cover Information Extraction using Remotely Sensed Data 30

2.2.1 Introduction 30

2.2.2 Fundamentals of Simulated Annealing 30

2.2.3 General discussion of SA and unsupervised classification 34

2.3 Review of Artificial Neural Networks Applications in Land cover Classifications 35

2.3.1 Introduction 35

2.3.2 Popular ANN models in remotely sensed classification applications 36

2.3.3 Multisource image classification using artificial neural networks 43

2.3.4 General discussion and summary 46

References 50

3. Paper #1 Development and Evaluation of Land-Use/Land-Cover Classification Systems using Simulated Annealing 67

3.1 Introduction 68

3.2 Simulated Annealing Adapted for Clustering Problems 72

3.2.1 Clustering by minimizing the distance function 72
3.2.2 Simulated Annealing: basic principle 75
3.2.3 Adaptation of Simulated Annealing for Clustering Issues 77
3.2.4 Definition of Parameters in SA 79
3.2.5 Integration of the SA and K-means for Clustering Refinement 81
3.3 Applications to Land-cover/Land-use Classification 83
3.3.1 Description of data sets and classification scheme 83
3.3.2 Study of the selection and roles of parameters in SA 86
3.3.3 Classification results and analysis 93
3.4 General Discussion and Conclusion 104
3.5 Recommendations for future research 107
References 108

4. #Paper 2 Development of an automated neural network classification system for remote sensing based land cover mapping 114
4.1 Introduction 115
4.2 Neural network Classification Approaches 118
4.2.1 Multilayer Perceptron (MLP) neural network 118
4.2.2 Kohonen's self-organizing mapping (KSOM) neural network 122
4.3 Development of an automated ANN classification system 127
4.4 Case study: neural network classification 134
4.4.1 Study area and classification scheme 134
4.4.2 Operational issues in neural network classification 136
4.4.3 Neural network classification and discussions 141
4.5 Conclusions and future work 149
References 152

5. #Paper 3 Development of a two-stage neural network-based land cover mapping system using multisource remotely sensing data 158
5.1 Introduction 159
5.1.1 Problem statement 161
5.1.2 A proposed two-stage neural network based multisource classification 163
5.2 Neural Network Methodologies and the Automated Data Selector 164
5.2.1 Kohonen's Self-Organizing Mapping (KSOM) neural network model 165
5.2.2 Multilayer Perceptron (MLP) neural network model 166
5.2.3 The adapted automated data selector 167
5.3 A Two-stage Neural Network based Multisource Classification 171
5.3.1 Study area and class definition 171
5.3.2 Image preprocessing 174
5.3.3 Two-stage multisource classification 177

5.3.4 Classification results and discussion 183
5.4 Summary and Conclusions 191
References 193

6. Summary and General Conclusions 198
6.1 Study One 198
6.2 Study Two 199
6.3 Study Three 200
6.4 General Summary 202
6.5 Recommended Future work 203

APPENDIX: Glossary 227

List of Tables

Table 3. 1 SA-related Parameters and Performance Indexes from TM-1 96
Table 3. 2 Error matrix on the classified result of TM-1 from the K-means 96
Table 3. 3 Error matrix on the classified result of TM-1 from the single SA-based algorithm 96
Table 3. 4 Error matrix on the classified result of TM-1 from the integrated SA-based Algorithm 97
Table 3. 5 Individual Kappa Analysis for the three Error Matrices from TM-1 97
Table 3. 6 Kappa Analysis Results for the Comparisons of the Error Matrices from TM-1 97
Table 3. 7 SA-related Parameters and best classification results from TM-2 99
Table 3. 8 Error matrix on the classified result of TM-2 from the K-means 99
Table 3. 9 Error matrix on the classified result of TM-2 from the Single SA-based algorithm 99
Table 3. 10 Error matrix on the classified result of TM-2 from the integrated SA-based algorithm 100
Table 3. 11 Individual Kappa Analysis for the Three Error Matrixes from TM-2 100
Table 3. 12 Kappa Analysis Results for the Comparisons of the Error Matrixes from TM-2 100
Table 4. 1 Classification scheme and category definition 135
Table 4. 2 Error matrix on the classified map from the MLP network 146
Table 4. 3 Error matrix on the classified map from the KSOM network 147
Table 4. 4 Error matrix on the classified map from the KSOM-SA 147
Table 4. 5 Individual Kappa Analysis for the three network error matrices 148
Table 4. 6 Kappa analysis results for the comparisons of the three error matrices 148
Table 5. 1 Classification scheme and category definition 174
Table 5. 2 Error matrix on the classified map of the two-stage multisource classification 184
Table 5. 3 Error matrix on the classified map of the one-stage MLP multisource classification 184
Table 5. 4 Error matrix on the classified map of the one-stage MLP-TM classification 185
Table 5. 5 Individual Kappa Analysis on the error matrices Table 5.2, Table5.3, and Table 5.4 185
Table 5. 6 Kappa analysis results for the comparisons of Table 5.2, Table5.3, and Table 5.4 185
Table 5. 7 Error matrix on the classified result of the fused image by dropping 40 reference points with time change from Table 5.2 189
Table 5. 8 Error matrix on the classified result of the fused image by dropping 9 reference points with significant misregistration error from Table 5.2 189
Table 5. 9 Error matrix on the classified result of the fused image by dropping 48 reference points either with class changes or with significant misregistration error from Table 5.2 190
Table 5. 10 Individual Kappa Analysis on Table 5.2, Table5.7, Table 5.8, and Table 5.9 190
Table 5. 11 Individual Kappa Analysis on the effects of the time change and the image registration on the classification of the fused image 190

List of Figures

Figure 1. 1 Flowchart and structure of the three-paper styled dissertation 6
Figure 1. 2 General system interface description of each ANN classifier module and interaction between sub-modules within system 9
Figure 2. 1 The structure of a three-layer neural network with 4 input nodes, 10 hidden nodes and 5 output nodes 38
Figure 2. 2 The structure of a Kohonen's Self-Organization Mapping Neural network with 4 input nodes and 9 output nodes 41
Figure 3. 1 The flowchart of K-means clustering algorithm 74
Figure 3. 2 The adapted single SA-based clustering algorithm 78
Figure 3. 3 Original Landsat TM-1 83
Figure 3. 4 Original Landsat TM-2 84
Figure 3. 5 Effect of T_0 on the magnitude of $J(V)$ 87
Figure 3. 6 Effect of μ on the magnitude of $J(V)$ 88
Figure 3. 7 Effect of the IET on magnitude of $J(V)$ 88
Figure 3. 8 Effect of the GP on magnitude of $J(V)$ 89
Figure 3. 9 Effect of T_0 on the Computational Time 89
Figure 3. 10 Effect of μ on the Computational Time 90
Figure 3. 11 Effect of the IET on the Computational Time 90
Figure 3. 12 Effect of the GP on the Computational Time 91
Figure 3. 13 The original image TM-1 and the classified maps 95
Figure 3. 14 The original image TM-2 and the classified maps 98
Figure 3. 15 Scatter plots of the 253 reference points from TM-1 and their class labels 102
Figure 3. 16 Scatter plots of the 299 reference points from TM-2 and their class labels 103
Figure 4. 1 The structure of three-layer neural network with 4 input nodes, 10 hidden nodes, and 5 output nodes 119
Figure 4. 2 The structure of Kohonen's Self-Organization Mapping Neural network 123
Figure 4. 3 The flowchart of the ANN-based Classification System 130
Figure 4. 4 Main Interface of the ANN System 131
Figure 4. 5 Pattern Conversion Sub-module Interfaces of the ANN System 132
Figure 4. 6 KSOM Training Sub-module Interfaces of the ANN System 132
Figure 4. 7 MLP Training Sub-module Interface of the ANN System 133
Figure 4. 8 Network Generalization Sub-module Interfaces of the ANN System 133
Figure 4. 9 The standard false color display of the acquired Landsat TM image 135
Figure 4. 10 Classified maps of the three network classifiers 144
Figure 4. 11 Learning curve of the mean squared error (MSE) vs. training epochs from the MLP network 145
Figure 4. 12 Topological maps from the KSOM and KSOM-SA networks 145
Figure 5. 1 The original KSOM Generalization Sub-module Interface of the ANN System 169
Figure 5. 2 The Modified Interface of the KSOM Generalization Sub-module 169
Figure 5. 3 The flowchart of the two-stage ANN classification 172

Figure 5. 4 Two multispectral images that are used in the fusion application..... 176
Figure 5. 5 Learning curves of the MLP training..... 181
Figure 5. 6 Classified maps of the study area 182

1. Introduction

Remote sensing is the science and art of deriving information about an object, area, or phenomenon by means of the analysis of data that are acquired without direct contact (Lillesand & Kiefer 2000; Jerson, 1996). Basically, the devices or sensors aboard satellites or aircrafts record electromagnetic energy emanating from the objects on multiple spectral bands, resulting in multispectral remotely sensed data (Swain & Davis 1978; Jerson, 1996; Lillesand & Kiefer 2000). Remotely sensed data have great potential to help solve many environmental problems (Swain & Davis 1978; Lillesand & Kiefer 2000): 1) management and conservation of natural resources, 2) urban and regional planning, 3) controlling and mitigating of environmental pollution, and 4) meteorological phenomena.

In analysis of remotely sensed data, there are numerous techniques developed to derive useful information from remotely sensed data to address these problems. Among them the computer-assisted digital imaging processing of remotely sensed data is the most popular technique used by most remote sensing researchers and scientists (Swain & Davis 1978). In recent decades, with the rapid emergence of many kinds of remotely sensed data, remote sensing has been widely used in Land-Use/Land-Cover (LU/LC) classification applications, such as urban planning, agricultural crop characterization, forest ecosystem classification, and flood monitoring (Jerson, 1996). The resulted land use and land cover mapping is one of the important sources for governments to manage and plan the natural resources at local and regional scales. It is therefore imperative to

investigate advanced computerized techniques to derive the most accurate and efficient information in a timely manner from currently available remotely sensed data.

1.1 Why advanced classification techniques are needed

It is crucial to maintain accurate and current information on land-cover and land-use conditions for timely planning, management and conservation of natural land resources. Multispectal classification of remotely sensed data offers a feasible tool to generate detailed land inventories especially for large, distant, and relatively ecologically complex areas. Various multispectal classification techniques have been used to produce high quality maps of the earth by assigning each pixel on the remotely sensed data to its appropriate thematic land category. An in-depth discussion of classification techniques can be found in numerous papers (Jain 1989; Jain & Dubes 1988; Schowengerdt 1983; Swain & Davis 1978; Tou & Gonzales 1977; Duda & Hart 1973). Among the most commonly used techniques in these remote sensing applications are *K*-means, minimum distance to means, maximum likelihood.

Image classification problems are often addressed as a minimization of the cost or energy functions by an iterative process, in which the optimal classification solution highly depends on the starting point and the convexity of the cost function (Klein & Dubes 1989; Laarhoven 1988). If the cost function is not convex, a search starting from a randomly selected point may be trapped in undesirable local minima and prevented from attaining the global or near-global minimum. Since many remotely sensed classification applications are very complicated and the related cost functions may not be convex, the

local minima problem could significantly affect the resulting classification performance. This issue has motivated many researchers to search for new techniques that can overcome the local minima problem and thus approximate the global or near-global minimal solution. As shown by Geman & Geman (1984) and Aarts & Laarhoven (1987), Simulated Annealing (SA) has the potential to find or approximate the global or near global minimum in a combinatorial optimization problem. The basic idea of SA is to incorporate some randomness into the assignments of cluster labels to pixels in the clustering procedure, thus reducing the likelihood of getting trapped in a local minimum. Based on this, we hypothesize that SA-based classification systems may be able to overcome the limitation of local minima so as to improve the classification accuracy for land cover classification. However, to date, SA has not been thoroughly investigated and has been rarely applied in remotely sensing applications because SA is computationally expensive and therefore not practical to deal with the large data sets typical of remotely sensed applications. However, the rapidly growing and available computing power in recent years enables faster processing of huge data sets and facilitates the use of more sophisticated and diverse methods for data classification. SA is one of the most interesting of these methods.

Many traditional classification methods are capable of image classification, but are found only to be robust under certain conditions. For example, some widely used traditional classification techniques such as the maximum likelihood classifier usually assume the Gaussian distribution of the input data. These techniques may become incapable when accurate estimation of the distribution model is difficult or impossible,

especially for multisource classification (Benediktsson *et al.*1990). It is thus necessary to develop advanced classification techniques that will have better classification performance under such circumstances. Classification approaches based on Artificial Neural Networks (ANN) have been extensively used in land cover and land use mapping during the past decade and are shown to be well-suited for remotely sensed image classifications (Zhang & Foody 2001; Serpico & Roli, 1995; Bischof *et al.* 1992; Heermann & Khazenie 1992; Benediktsson 1990). As compared to traditional statistical classification approaches, ANN-based approaches have many advantages: 1) independence of the statistical distribution of the input data, 2) the ability to estimate the non-linear relationship between the input data and desired outputs, 3) parallel computational potential, 4) fast generalization capability once a well-trained network is found, and 5) the ability to fuse different types of spatial data sets.

The rapid advancement of spatial and computer technologies in recent years has led to a dramatic boost in available remotely sensed data and Geographical Information System attribute data. The availability of multiple types of spatial data offers the potential to explore the real world more accurately because the fusion of complementary data can help reduce uncertainty and improve classification accuracy (Hagarat–Mascle *et al.* 1997). As a data fusion technique, ANN is a strong alternative to other multisource fusion techniques because there is no need to estimate the statistical distribution models for various data sources. Many previous studies on classification of multisource remote sensing data have demonstrated that ANN-based approaches can perform better and

simpler than traditional statistical classification methods such as maximum likelihood classifiers (Foody & Arora 1997; Bischof *et al.* 1992; Benediktsson *et al.*1990).

It has been shown that no single approach or procedure is optimal to completely solve a complex classification problem (Jain *et al.* 2000). The combination of classifiers has, therefore, become a heavily studied topic. An ideal classification system should be: 1) as fully automated and as fast as possible to process large data sets quickly, and 2) as robust and as flexible as possible to be able to handle various situations, for example, single source vs. multisource data classification, with minimal human interaction. Thus, the study of the complexity, feasibility, timeliness, and performance of different classification methods is our focus in this research.

1.2 Proposed classification systems and their applications

In this dissertation research, we propose to explore, design, develop, implement, and evaluate advanced LU/LC SA- and ANN-based classification systems that can handle single and multisource remotely sensed imageries. This research consists of three interrelated studies. The working flowchart for the three studies is shown in Figure 1.1. The specific objectives for each of the studies are as follows:

- To develop and evaluate SA-based classification systems in the first study.
- To develop an automated neural network classification system for remote sensing based land cover mapping applications in the second study.
- To develop a two-stage neural network based land cover mapping system using multisource remotely sensed data in the third study.

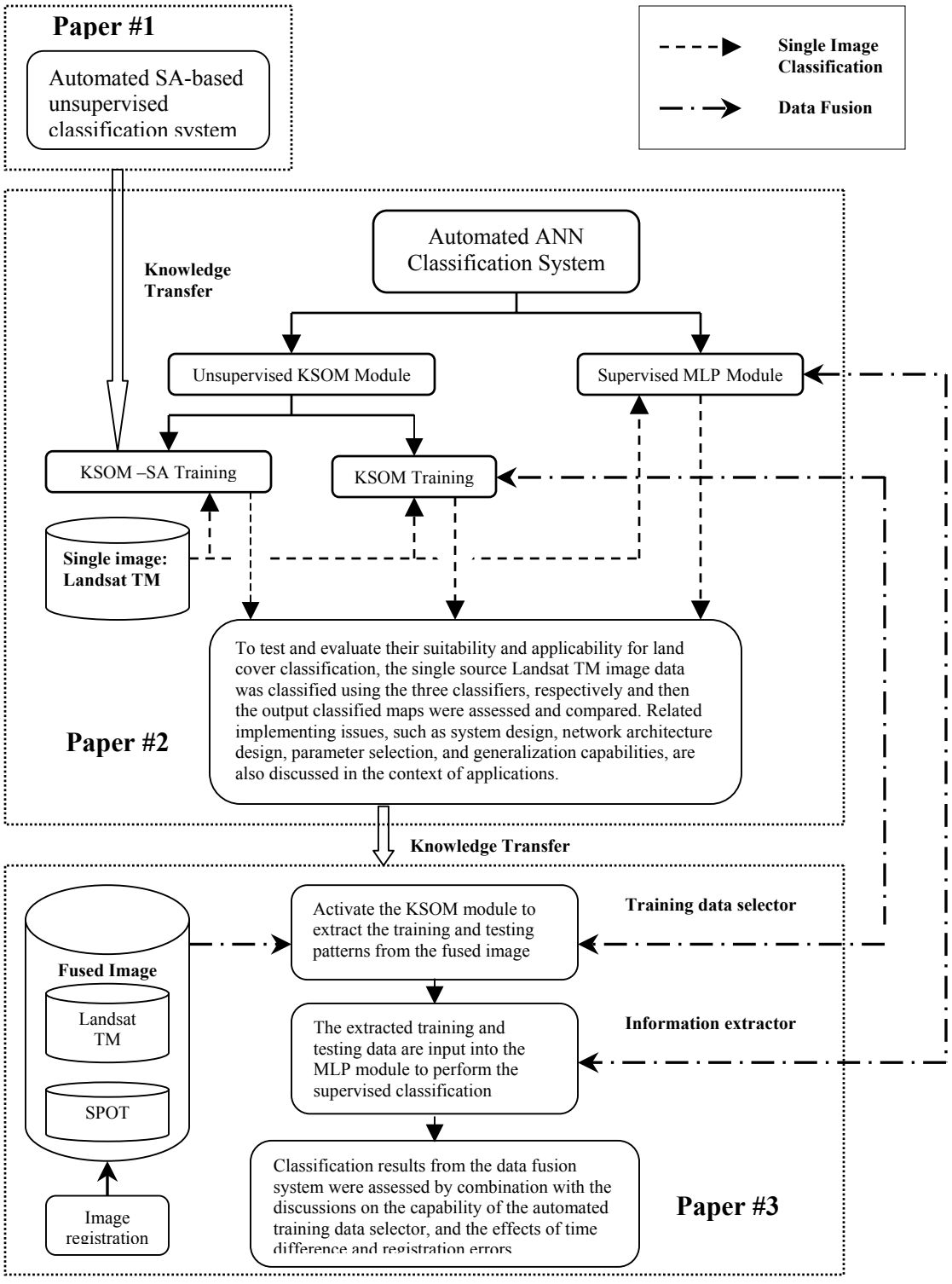


Figure 1. 1 Flowchart and structure of the three-paper styled dissertation

1.2.1 Study one: development of SA-based classification systems

Many previous studies have shown that one of the traditional iterative unsupervised approaches, *K*-means, suffers from the aforementioned local minima problem (Selim & Alsultan 1991; Klein & Dubes 1989). SA is proved to be able to overcome the local minimum problem (Geman & Geman 1984; Aarts & Laarhoven 1987). *Our hypothesis is that SA-based classification systems can solve the local minimum problem in K-means and thus improve classification performance.* In this study, two automated SA-based classification systems are proposed: 1) the Single SA-based (S-SA) system that is developed based on the standard SA algorithm, and 2) the Integrated SA-based (I-SA) system that is developed by combining the standard SA algorithm and *K*-means into a two-level classification system. The I-SA system is designed to produce a desired classification more efficiently than the S-SA system. To verify our hypothesis, we apply the two SA-based systems to classify two Landsat Thematic Mapper (TM) images with different landscape conditions. The classification performance of each algorithm is assessed by comparison with the collected reference data. Issues such as the effects of the complexity of the application and the controlling parameters in SA on the resulting classifications are one of the foci of this study. This study on SA-based classification systems is the first time that SA has been intensively investigated for land use/land cover classification purposes using satellite images. Working knowledge and experience obtained from this study are helpful for SA-based systems to be used in other classification applications using remotely sensed image data or ancillary data.

1.2.2 Study two: development of an automated neural network classification system

In the second study of this research, we propose to develop an automated ANN classification system that is suitable for LU/LC classification using remote sensed data. The ANN classification system consists of two ANN classification modules: 1) supervised Multilayer Perceptron (MLP) neural network module using the Backpropagation (BP) learning algorithm, and 2) an unsupervised Kohonen's Self-Organizing Mapping (KSOM) neural network module. In the KSOM network module, two competitive learning algorithms are provided: 1) the first learning algorithm is based on the standard KSOM competitive learning algorithm, and 2) the second algorithm is the refined KSOM-SA learning algorithm. Because the KSOM network suffers from the same local minimum problem as *K*-means, we refine the standard KSOM learning algorithm by incorporating SA global searching procedures to reduce the likelihood that it will be trapped in local minima. Within each ANN classification module, there are several sub-modules to perform the network pattern conversion, neural network training, and network generalization as shown in Figure1.2.

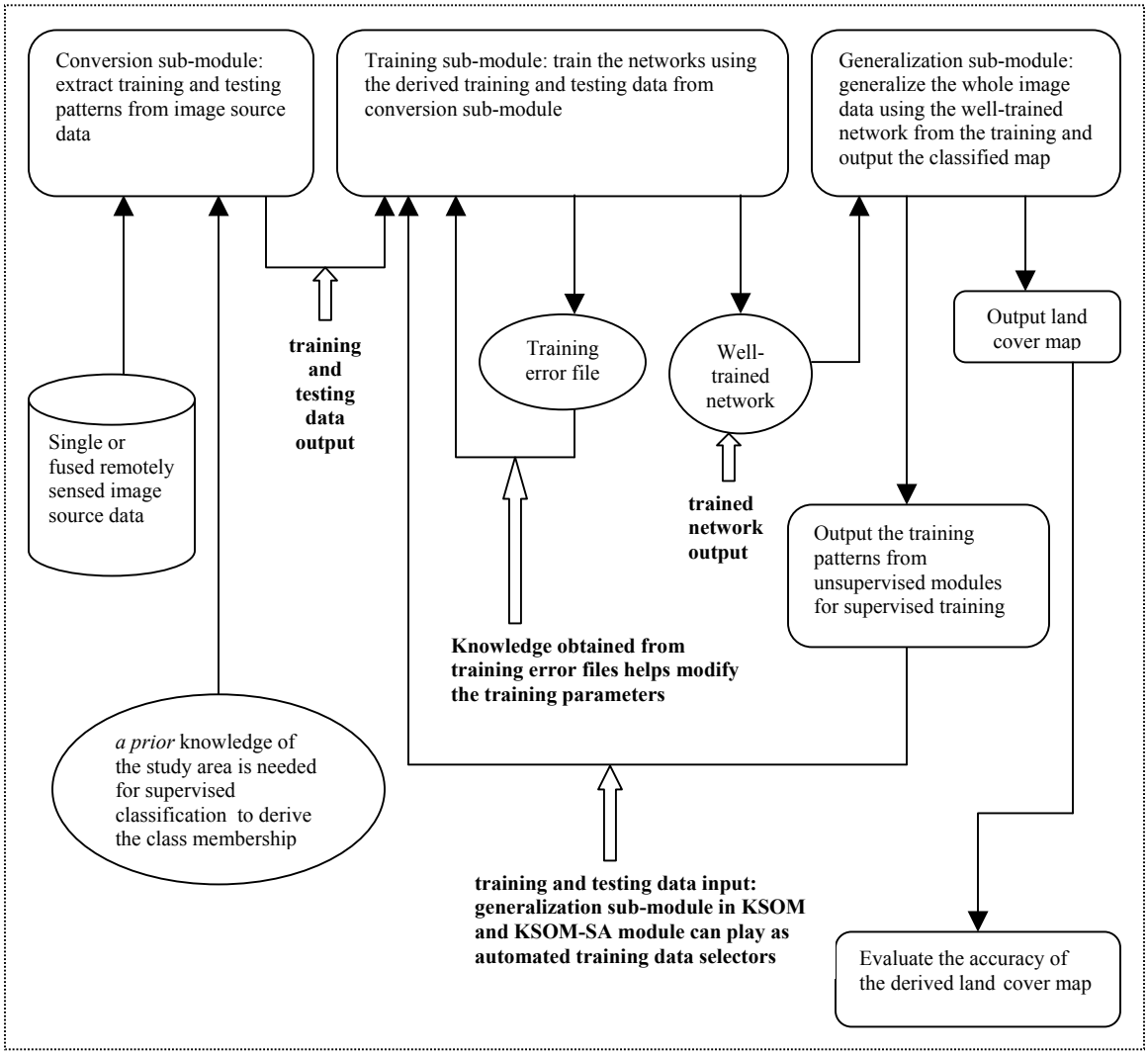


Figure 1. 2 General system interface description of each ANN classifier module and interaction between sub-modules within system

Two hypotheses are tested in this study: 1) the proposed ANN system is well-suited for LU/LC classification, and 2) the refined KSOM-SA network has the potential to overcome the local minima problem and thus improve the classification performance as compared with the standard KSOM network. To verify these two hypotheses, we apply the ANN system to perform LU/LC classifications of a Landsat TM image using the supervised MLP network and the two unsupervised KSOM networks. The training and testing data sets are manually extracted from the TM image to train the network classifiers. For the supervised MLP classification, the class labels are provided for each training and testing data. Once the training is complete, the three well-trained networks are used to classify the entire TM image into a land cover map with eight categories. The resulting classifications from the three neural networks are assessed by comparison with reference data derived from the high spatial resolution Digital Color Infrared (CIR) Digital Orthophoto Quarter Quad (DOQQ) data. To qualitatively evaluate the classification performance of the three ANN classification approaches in terms of the classification accuracy, Kappa analyses are performed on the resulting ANN classifications.

1.2.3 Study three: development of a two-stage neural network-based multisource classification using multiple remotely sensed data

The inclusion of complementary information from additional data sources, or "data fusion", can help improve classification accuracy (Hagarat–Mascle *et al.* 1997). ANN approaches have been shown to have great potential to fuse multiple source data sets (Foody & Arora 1997; Bischof *et al.* 1992; Benediktsson *et al.*1990). However, there

is a potential problem with ANNs. Their performance is very sensitive to the quality and quantity of collected training data sets. A well-trained ANN may fail to generalize well for those land categories that are poorly represented in the training data sets. This training data selection problem could become more serious in a data fusion application for two reasons. First, the required amount of the training data set for each category needs to be considerably increased along with the increase of input dimensions. Second, significant class changes during the time interval of acquired image data sets and misregistration error can compound the typical spectral signatures of certain land categories, thus making the visual interpretation of the scene problematic. For these reasons, it is necessary to take measures to assist in collecting adequate and reliable training data set for each desired category. In the third study, we propose to adapt the automated two-module ANN classification system to develop a two-stage multisource classification of Landsat TM and multispectral Satellite Pour l'Observation de la Terre (SPOT) images in the selected study area. In the first stage of this classification, the unsupervised KSOM neural network is adapted as an Automated Data Selector (ADS) to replace the manual process of training data collection to help improve the quality of the selected training data sets. In the second stage, the supervised MLP classification is trained using the automatically extracted training data sets and then used to perform the final classification. In addition to the two-stage multisource classification, two additional classifications of the same study area are also performed in this study: 1) a one-stage MLP multisource classification in which the MLP network is trained using the training data sets manually collected from the fused image and then used to classify the entire image, and 2) a one-stage MLP-TM

classification in which, instead of collecting training data sets from the fused image, we manually collect training data sets from the single Landsat TM image to train the network and then perform the image classification.

Two hypotheses are tested in this study: 1) the ADS adapted from the KSOM network provides adequate and reliable training data sets and thus improves the classification performance of the MLP classification, and 2) the fusion of the two Landsat TM and SPOT images using the ANN system increases the classification accuracy of the derived land categories. The comparison between the two-stage multisource classification using the ADS and the one-stage MLP multisource classification using the manually collected training data can show the effectiveness of the ADS and test the first hypothesis. By comparing the one-stage MLP multisource classification with the one-stage MLP-TM classification, the second hypothesis can be tested.

1.3 Dissertation Organization

The three studies are presented as a three publishable papers. Following the introduction chapter, three brief literature reviews are presented in Chapter 2. The three reviews provide a detailed description of previous and current work regarding the classification techniques in remote sensing. Then Paper #1, #2, #3, are arranged in the chapters 3, 4, and 5. A general summary and conclusions of the main results from the three papers are given as Chapter 6.

References

Aarts, E. H. L., and P. J. M. Van Laarhoven, "Simulated annealing: a pedestrian review of the theory and some applications", *Pattern recognition theory and applications,* pp.179-192, 1987.

Benediktsson, J.A., P.H. Swain, and O.K. Ersoy, "Neural network approaches versus statistical methods in classification of multi-source remote sensing data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 28, pp. 540-551, 1990.

Benediktsson, J.A., and J. R. Sveinsson, "Feature extraction for multisource data classification with artificial neural networks," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 727-740, 1997.

Bischof, H., W. Schneider, and A. J. Pinz, " Multispectral classification of landsat images using neural networks," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 30, no. 3, pp. 482 – 490, 1992.

Carpenter, G.A., M.N. Gjaja, S. Gopal, and C.E. Woodcock, "ART neural networks for remote sensing: vegetation classification from Landsat TM and terrain data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 35, no. 2, pp. 308-325, 1997.

Duda, R. O., and P. E. Hart. Pattern classification and scene analysis. *New York: Wiley-Interscience*. 1973.

Foody, G. M., and M. K. Arora, "An evaluation of some factors affecting the accuracy of classification by an artificial neural network," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 799-810,1997.

Geman, S., and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Proceeding. Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721 – 741, 1984.

Granville, V., M. Krivanek, and Jean-Paul Rasson, "Simulated annealing: a proof of convergence," *IEEE Transactions on Pattern Analysis and Machine Iintelligence*, vol. 16, no. 6, pp. 652-656, 1994.

Hegarat-Macel S. L., I. Bloch, and D. Vidal-Madjar, "Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 35, no. 4, pp.1018 – 1031,1997.

Heermann, P. D., and N. Khazenie, "Classification of multispectral remote sensing data using a back-propagation neural network," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 30, pp.81-88, 1992.

Highleyman, W. H., "Linear decision functions, with applications to pattern recognition," *Proc. IRE*, vol. 50, pp.1501 – 1514, 1962.

Jain, A. K., R. P. W. Dulin, and J. C. Mao, "Statistical pattern recognition: a review," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 22, no.1, pp. 4 – 37, 2000.

Jenson, J. R.. Introductory digital image processing. *Englewood Cliffs, NJ: Pretentice Hall.* 1996.

Klein, R. W., and R. C. Dubes, "Experiments in projection and clustering by annealing," *Pattern Recognition*, vol. 22, no. 2, pp. 213-220, 1989.

Kolonko, M., and M. T. Tran, "Convergence of simulated annealing with feedback temperature schedules," *Probability in the Engineering and Informational Sciences*, vol. 11, pp. 279-304, 1997.

Lillesand, T. M., and R. F. Keifer, Remote sensing and image interpretation. 4th. *NewYork: Wiley*, 2000.

Locatellli, M., "Convergence properties of simulated annealing for continuous global optimization," *Journal of Applied Probability*, vol. 33, pp. 1127-1140, 1996.

Serpico, S. B., and F. Roli, "Classification of multisensor remote-sensing images by structured neural networks," *IEEE Trans. Geosci. And Rem. Sens.*, vol.33, no.3, pp. 562-578, 1995.

Selim, S. Z., and K. Alsultan, "A simulated annealing algorithm for the clustering problem," *Pattern Recognition*, vol. 24, no. 10, pp.1003-1008, 1991.

Smith, F. W., " Pattern classifier design by linear programming," *IEEE Trans. Comput.*, vol. C-17, pp. 367 – 372, 1968.

Swain, P. H., and S. M. Davis, Remote sensing: the quantitative approach. *New York: McGraw-Hill*, pp. 136 – 188,1978.

Zhang, J., and G. M. Foody, "Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: statistical and artificial neural network approaches," *International Journal of Remote Sensing*, vol. 22, no. 4, pp. 615 – 628, 2001.

2. Literature Review

2.1 Review of thematic land information extraction from remotely sensed data

It is essential to maintain accurate and current information on the land-cover and land-use conditions for timely planning, management and conservation of the natural land resources. For this to be most effective, accurate thematic land information must be provided. Multispectral classification of remotely sensed data provides a feasible tool to generate such information. To produce a high quality thematic map using multispectral classification techniques, two questions need to be addressed: 1) which source data will be appropriate and easily acquired for the specific application, and 2) which classification techniques could be utilized to extract accurate land cover information in a timely manner. In this chapter, we first review the five popular remotely sensed data types including Advanced Very High Resolution Radiometer (AVHRR) data, Landsat Thematic Mapper (TM), Satellit Pour l'Observation de la Terre (SPOT), Radio Detecting and Ranging Satellite (RADARSAT), and I-KOH-NOHS (IKONOS) and their applicability in Land-Use/Land-Cover (LU/LC) applications based on their specific characteristics in section 2.1.1. Second, we review the five commonly used traditional classification techniques in remote sensing and their pros and cons in multispectral classification in the section 2.1.2.

2.1.1 Remotely sensed data and their applications for land cover classification

2.1.1.1 Advanced Very High Resolution Radiometer (AVHRR) Data

AVHRR is a five-channel scanner and is carried on the National Oceanic and Atmospheric Administration's (NOAA's) Polar Orbiting Environmental Satellites (POES), beginning in 1978. AVHRR data is capable of providing global and regional coverage for monitoring land use and land cover change. A number of studies have demonstrated the potential of AVHRR data for LU/LC applications at a regional scale (Malingreau *et al.* 1989; Townshen & Justice 1986). On the other hand, the coarse spatial resolution (1.1km at nadir) limits the uses of AVHRR in some applications because there exist many mixed pixels in an AVHRR image. Two approaches to use AVHRR data in land cover mapping are investigated: single-date and multi-temporal. However, the analysis of a series of single-date images for land-cover studies is often impractical, because it is very difficult to find cloud-free AVHRR images for a large area due to its high temporal resolution. Thus, increased efforts have been made to use composite AVHRR images from multi-temporal AVHRR data sets (Running *et al.* 1995; Townshend 1994; Brown *et al.* 1993; Evans *et al.* 1993; Loveland *et al.* 1991). To reduce the adverse effects of atmospheric conditions, some special techniques have been explored. Cihlar *et al.* (1996) used the Normalized difference Vegetation Index (NDVI) data derived AVHRR to reduce clouds and atmospheric effects and thus improve the vegetation classification accuracy. In addition, higher resolution satellite data are used to derive the detailed land categories within the general categories identified from the AVHRR data.

2.1.1.2 Landsat Thematic Mapper (TM) system

Since 1972, the National Aeronautics and Space Administration (NASA) have launched a series of Earth-monitoring satellites that are designed for the needs of resource managers and Earth scientists. The Landsat systems are one of the most widely used satellites for LU/LC applications. The improved spatial resolution (30 m) of TM data as compared to NOAA AVHRR data increases its capability to differentiate land cover types for LU/LC applications at a relatively local scale. The enhanced spectral resolution improves the spectral separability between LU/LC classes and makes it particularly useful to characterize areas with various vegetation covers (Solberg *et al.* 1994; Frank 1988; Strahler *et al.* 1978*)*. Furthermore, Landsat TM data has been demonstrated to well-suited for urban studies (Haack *et al.* 1987; Khorram *et al.* 1987). The enhanced characteristics of Landsat TM systems, and the reliable, low cost, data acquisitions have led the Landsat TM data to be one of the most popular satellite images in LU/LC applications.

2.1.1.3 Satellite Pour l'Observation de la Terre (SPOT) System

The SPOT satellite Earth Observation System was designed by the CNES (Center National d'Etudes Spatiales), in France. The development of the SPOT system with onboard High Resolution Visible (HRV) sensors was announced by Jaakkola *et al.* (1987) as a crucial step in the use of remote sensed data for LU/LC investigations. The SPOT system provides SPOT multispectral data with high resolutions in terms of both spatial resolution (20 meters) and spectral resolution (4 spectral bands). The improved spatial solution offers more accurate LU/LC information with sharper outlines and edges

that enhances the capabilities to detect the land surface details. This suggests that the SPOT system is particularly useful for land-cover mapping applications at a local scale. Merging of both SPOT multispectral and panchromatic data with a spatial resolution of 10m is suitable for urban/suburban mapping (Couloigner *et al.*, 1998).

2.1.1.4 Radio Detecting and Ranging Satellite (RADARSAT) System

The Canadian RADARSAT-1 was launched in November 1995 to acquire data for monitoring environmental changes and improving the management of natural resources. Its mapping capabilities enhance the classification performance of optical aerial and satellites. Optical aerial and satellites have been used for a long time to produce information about the current LU/LC characterization (Running *et al.* 1995; Jaakkola *et al.* 1987; Strahler *et al.* 1978). Unfortunately, its performance is restricted by adverse meteorological conditions, such as clouds, rain, or fog. Synthetic Aperture Radar (SAR) is a powerful sensor that can overcome these meteorological restrictions because it contains an electromagnetic microwave spectrum that is insensitive to rain, fog, hail, smoke, and most importantly, clouds (Elachi 1988). RADARSAT imagery has been utilized to study its applicability in the LU/LC applications in forest areas (Kux *et al.* 1998; Shimabukuro *et al.* 1998; Brisco 1985; Ulaby *et al.* 1983; Ahern *et al.* 1978). Fusion of RADARSAT data and optical satellite data has been investigated by many researchers (Paris & Kwong 1988; Brisco and Brown 1995). These studies show that the fusion of RADARSAT data and optical satellite data has the potential of taking advantage of the complementary characteristics in both data sets and improve the

resulting classification accuracy. However, RADARSAT data have to be rectified using some special image preprocessing procedures to remove speckle noise.

2.1.1.5 I-KOH-NOHS (IKONOS) System

The IKONOS satellite is the world's first commercial satellite to collect images with one-meter resolution. IKONOS is derived from the Greek word for "image". IKONOS-1 was successfully launched on 24th of September 1999. Space Imaging EOSAT (Colorado, USA) is operating the satellite in co-operation with a network of regional ground stations. IKONOS image data are distinct from other satellite data sets because of their very high spatial resolution (1-meter panchromatic and 4-meter multispectral). In the past, the applicability of remotely sensing data to many LU/LC applications, such as urban planning on a local scale and vegetation species discrimination, was limited because the high spatial resolution data were not available. The moderate spatial resolutions of Landsat TM and multispectral SPOT data have been shown to be inadequate for the accuracy and specificity required for many urban applications (Harrison & Richards 1988; Haack *et al.* 1987; Khorram *et al.* 1987). The very high-resolution IKONOS data offers a better choice in these classifications.

2.1.1.6 General discussions of satellite images

It is clear that different applications have different requirements for the spectral, temporal and spatial resolutions of the applied remotely sensed data, depending on the application and the level of land-cover complexity. Timely planning and management of natural resources on the earth require detailed and accurate thematic land information at

various mapping scales ranging from global to local. With the advantage of large coverage and high temporal resolution, the satellite imagery from NOAA AVHRR is more suitable for regional and global scale land mapping. For regional and local scale mapping, the most suitable satellite data are Landsat TM and SPOT image data sets. The IKONOS image data with very high spatial resolution makes urban planning at small local areas possible. In addition, the fusion of RADARSAT data and optical satellite images is useful to reduce the effects of adverse meteorological and atmospheric conditions. The use of the appropriate image data is very important for any specific application and ensures proper information to be generated. Unfortunately, in practice, there are many factors, such as the project budget and the timely requirement, that often compromise the accessibility of the most appropriate data. Under such circumstances, it is essential to use some powerful image analytical techniques to enhance the differentiation capability of the acquired image data.

2.1.2 Classification algorithms and their applications for land cover classification using remotely sensed data

In recent decades a variety of classification techniques have been developed to perform multispectral classification (Jensen 1996; Serpico & Roli 1995; Jain 1989; Schowengerdt 1983; Swain & Davis 1978; Duda & Hart 1973). Basically, they are fit into two general classification categories: supervised and unsupervised. In a supervised classification, *a priori* knowledge of the scene with respect to the identity and location of desired land categories is used to collect training data through field work and analysis of very high resolution remotely sensed data maps in house (Mausel *et al.* 1990).

Unsupervised classification is known as "clustering" which is "a generic labeling procedure designed to find natural groupings, or clusters, in multidimensional data based on measured or perceived similarities among patterns" (Jain 1989). It is the analysts' responsibility to relate the resulting spectral clusters from unsupervised classification to the desired land categories (Jensen 1996; Lark 1995; Schowengerdt 1983; Swain & Davis 1978).

In the following sections from 2.1.2.1 to 2.1.2.5, we briefly review the five most frequently used classifiers in remote sensing: three supervised (parallelpiped, maximum likelihood, and minimum distance), and two unsupervised (*K*-means and ISODATA). A general discussion on their pros and cons in remote sensing application is given in section 2.1.2.6.

2.1.2.1 Parallelpiped Classification

This method is based on a simple Boolean "and/or" logic decision rule (Jensen 1996). Training data in *n* spectral bands are used to estimate a *n*-dimensional mean vector, $\mu_k = (\mu_{k1}, \mu_{k2}, ..., \mu_{ki}, .\mu_{ki+1}, ..., \mu_{kn})$ with μ_{ki} denoting the mean value of the training data obtained from the k^{th} class in the i^{th} spectral band, and the standard deviation S_{ki} denoting the standard deviation of the training data of the k^{th} class in the i^{th} band. The parallelepiped algorithm assigns an unknown pixel Y_j denoted by an *n*-dimensional vector $(y_{j1}, y_{j2}, ..., y_{ji}, ... y_{jn})$ to the k^{th} class if and only if

$$\mu_{ki} - S_{ki} \le y_{ji} \le \mu_{ki} + S_{ki} \qquad \text{--- (2.1)}$$

where $k = 1,2,3,\ldots,m$ and m denotes the number of classes

$i = 1,2,3,\ldots,n$ and n denotes the number of spectral bands

Based on Equation 2.1, the decision boundary of each of the m classes is a rectangular area in a two-dimensional feature space. An unknown pixel is classified based on the decision boundary in which it lies, or as 'unknown' if it lies outside all decision boundaries (Lillesand & Kiefer 2000).

This parallelepiped method is computationally efficient (Lillesand & Kiefer 2000; Jensen 1996). However, the classifier may have high misclassification rates if there exist significant decision boundary overlaps between classes (Lillesand & Kiefer 2000). In addition, there may be many unclassified pixels outside any one of the m defined decision boundaries. Lillesand and Kiefer (2000) propose to solve the overlapping problem by dividing the single rectangles of the decision boundaries into a series of rectangles with stepped borders. Jensen (1996) suggests combining this classifier with other classifiers, such as the minimum distance and the maximum-likelihood, to classify those unclassified or ambiguous pixels to classes.

2.1.2.2 Maximum-Likelihood or Bayesian Classification

The maximum-likelihood (ML) is the most popular classifier in remote sensing (Foody *et al.* 1992; Maselli *et al.* 1992; Bolstad & Lillesand 1991; Richards 1986; Mather 1985; Shlien 1977).

Implementation of the maximum-likelihood classifier is based on the statistical assumption of a distribution model (e.g., Gaussian distribution) to estimate the density parameters such as the mean vector and covariance matrix using the acquired training

data. Based on the estimated probability density functions, an unknown pixel is assigned to the class with highest probability value. An extension of the maximum likelihood approach is the Bayesian classifier. The Bayesian classifier incorporates two weighting factors into classification: the *a priori* probability, i.e., the anticipated likelihood of occurrence for each class in the given scene, and a weight associated with the 'penalty' of misclassification.

Numerous successful applications using the maximum likelihood or Bayesian classifier have been reported (Ediriwickrema & Khorram 1997; Foody *et al.* 1992; Bolstad & Lillesand 1991; Mather 1985; Strahler 1980, Shlien 1977). The main drawback of the maximum-likelihood classification in remote sensing application is the large number of computation required to classify each pixel. Many authors (Bolstad & Lillesand 1991; Mather 1985; Strahler 1980; Shlien & Smith 1975; Eppler 1974) have used table look-up technique to address this problem based on a common knowledge that many of the pixels in the image have identical multispectral brightness values. Other improvement techniques include using feature extraction techniques such as principle component transformation to reduce dimensionality before performing the maximum likelihood classification (Lillesand and Kiefer 2000). Maselli *et al.* (1992) state that the maximum-likelihood classifier can have very good classification performance if the basic assumption of multi-normality in the training data is valid. However, the maximum-likelihood classifier may be invalid when the spectral distributions of desired classes deviate from normal, which is not an uncommon circumstance in complex classification problems.

2.1.2.3 Minimum distance to means Classification

This Minimum distance classifier is computationally simple (Jensen 1996). Like the parallelepiped classifier, the mean vectors and standard deviation for each class are estimated from the acquired training data. Using this method, an unclassified pixel is assigned to the nearest class in an *n*-band spectral space. The most commonly used distance measure to determine the nearest class is Euclidean distance.

The minimum distance classifier has been applied in many remote sensing applications (Lee & Landgrebe 1993; Franklin & Wilson 1992; Hodgson 1988). Jensen (1996) points out that the minimum-distance decision rule could produce good classification results that are comparable to those from the sophisticated classifiers such as maximum-likelihood when classes are spectrally separated.

2.1.2.4 *K*-means Classification

The *K*-means classifier is a very popular clustering algorithm and is the basis of many sophisticated clustering algorithms (Jain & Dubes 1988; Selim & Ismail 1984; Tou & Gonzales 1977; Duda & Hart 1973; Fukunga 1972). In a clustering problem, the generalized least-squared error function *J(V)* is defined as below to guide the search for an optimal *k* clustering:

$$J(V) = \sum_{j=1}^{N} d_{ij}^2 \qquad \text{--- (2.2)}$$

where d_{ij}^2 is the associated distance function ($d_{ij}^2 = \| y_j - v_i \|^2$) between a pattern y_j and a given cluster center v_i where $\| \ \|$ is a norm on R^n ,

$Y = \{y_j / j \in [1, N], y_j \in R^n\}$ is the data set,

N is the number of patterns within the data set,

$V = \{v_i / i \in [1, k], v_i \in R^n\}$ is the set of k cluster centers

The minimization of such a criterion function is usually sought by an iterative scheme, which starts with arbitrarily chosen initial cluster centers in the first iteration. Each pixel is then assigned to the closest cluster. A new set of cluster centers is calculated from the results of the previous assignments and the pixels reassigned to the clusters. The procedure continues until there is no significant change in pixel assignments from one iteration to the next or a certain number of iterations are attained.

The *K*-means classifier is computationally efficient. Although no general proof of convergence exists for this algorithm, it is able to produce good results if the clusters are compact, well separated in the feature space, and hyperspherical in shape if *Euclidean* distance is used, or hyperelliptical in shape if *Mahalanobis* distance is used (Jain *et al.* 2000). Other studies have shown that the *K*-means algorithm fails to produce satisfactory classifications due to the limitation of the local minimum (Kolonko & Tran 1997; Selim & Alsultan 1991; Fukunaga 1990; Klein & Dubes 1989). K-means is implemented via an iterative process to minimize the distance function defined in Equation 2.2. The resulting minimal solution is dependent on the randomly selected starting search point and the convexity of the error function. In a complex classification problem, the distance function may not be convex and thus there might be undesirable local minima to prevent the minimization from attaining the global or near-global minimum. Many techniques have been proposed to improve the performance of *K*-means by: 1) incorporating a fuzzy

criterion function in fuzzy K-means algorithm (Lillesand & Keifer 2000; Jerson 1996; Bezdek 1981), 2) using generic algorithms, such as simulated annealing (Yuan *et al.* 1999; Bandyopadhyay *et al.* 1998; Brown & Huntley 1992), and deterministic annealing to optimize the resulting clustering (Rose 1998), and 3) mapping it onto a neural network for parallel implementation (Mao *et al.* 1994).

2.1.2.5 ISODATA Classification Algorithm

The well-known Iterative Self-Organizing Data Analysis Technique (ISODATA) (Jain 1989; Selim & Ismail 1984; Tou & Gonzales 1977; Ball & Hall 1965) can be considered as a sophisticated version of *K*-means that incorporates a comprehensive set of heuristic procedures. The most important heuristic procedures in ISODATA are cluster lumping and cluster splitting. In the ISODATA classifier, a set of parameters is specified to create new clusters or merge existing clusters. A cluster is split if it contains too many pixels and has an unusually large variance within the cluster. Two clusters are merged if their cluster centers are sufficiently close. To perform the cluster split and merge operations in ISODATA algorithm, the parameters are usually specified as follows: (Jensen 1996; Tou & Gonzales 1977):

- The maximum number of clusters to be identified.
- The maximum number of pixels without cluster changes between iterations, and the maximum iterations. These two parameters control when to terminate the iteration.
- The maximum standard deviation and the maximum number of pixels within a cluster. These two parameters determine when to split clusters.

- The minimum distance and the minimum number of pixels within a cluster.

  These two parameters determine when to merge clusters.

The iterative clustering process of ISODATA is similar to that of *K*-means. However, the assignment of each pixel in ISODATA involves additional analyses using the described parameters to split or merge the clusters. There are many successful applications of ISODATA in remote sensing (Ouegan *et al.* 2000; Pierce *et al.* 1998; Jensen 1996). The cluster splitting and merging capability of the ISODATA classifier allows the algorithm to be, to some extent, less sensitive to poor initial clustering (Jain & Dubes 1988). Due to the heuristic nature of ISODATA, its effective implementation in a classification problem is highly dependent upon the analyst's understanding of the controlling parameters in clustering problems. Like *K*-means, there is no guarantee that this algorithm can converge to a globally optimal clustering.

2.1.2.6 Summary

Using these traditional classification methods, abundant successful supervised and unsupervised remote sensing application examples have been reported (Cihlar *et al.* 1998; Thomson *et al.* 1998; Bruzzone *et al.* 1997; Guneriussen *et al.* 1997; Remund *et al.* 1998; Foody *et al.* 1995; Benediktsson *et al.* 1990; Belward *et al.* 1990).

In a supervised classification, a set of labeled training samples must be collected. Using the collected training data, the spectral signatures of classes are estimated and then used to classify other unknown pixels throughout the image. For this reason, the classification performance of a supervised classifier highly depends on both the number and quality of training samples (Jain *et al.* 2000). However, the collection of reliable

training data may be very expensive and laborious in terms of time and economic cost. This drawback has limited the usability of supervised classification approaches and led us to explore the use of unsupervised methods. Unsupervised classification algorithms are to identify a set of clusters in the data with little supervision of the analyst. However, after clustering is completed, the analyst has to match the derived spectral classes to each of the desired land categories if the clustering purpose is to form a land cover map (Jensen 1996; Palylyk & Crown 1984). It is very likely that one general land category has large spectral variability and consists of many spectral classes (Lark 1995). The criteria that are used to match land categories with spectral classes will affect the accuracy of the resulting land cover map.

Our review shows that no single classifier is uniformly superior to all the others and so the user has to make a choice on a case-by-case basis. The combination of classifiers has, therefore, become a heavily studied topic (Jain *et al.* 2000). Supervised and unsupervised classifications can complement each other. In recent years hybrid classification that combines unsupervised and supervised classification has received wide attention (Laba *et al.* 1997; Jensen 1996; Sader *et al.* 1995; Lillesand & Keifer 1994). On the other hand, the rapid growing and available computing power, while enabling faster processing of huge data sets, has facilitated the use of elaborate and diverse methods for data classification. It is necessary to investigate more advanced classification techniques that have superior classification performance in terms of classification accuracy, speed, and cost.

2.2 Review on Simulated Annealing (SA) and Applications for Land Cover Information Extraction using Remotely Sensed Data

2.2.1 Introduction

Simulated Annealing (SA), which is developed based on an analogy between the physical annealing process of solids and the problems of solving large combinatorial optimization problems, has shown to have the potential to find or approximate the global or near-global optimal in a combinatorial optimization problem (Aarts & Laarhoven 1987;Geman & Geman 1984). However, SA often requires much greater computational time. SA was first described by Metropolis *et al.* in 1953 and then introduced by Kirkpatrick *et al.* (1983) and Cerny (1985) into combinatorial optimization problems. The special characteristics of SA are its general applicability and simplicity of implementation. SA has been proven applicable in many classification applications (Brown & Huntley 1992; Selim & Alsultan 1991; McErlean *et al.* 1990; Klein & Dubes 1989). In section 2.2.2, the fundamental theory behind the SA algorithm is introduced. In section 2.2.3, the relationship between SA and clustering problems is established and the incorporation of SA into clustering problems is proposed to reduce the likelihood of traditional clustering methods to be trapped by local minima in the same section.

2.2.2 Fundamentals of Simulated Annealing

Metropolis *et al.* (1953) first introduced the well-known *Metropolis algorithm* by simulating the physical annealing process of solids. The algorithm is described as follows (Laarhoven 1988):

1) For a solid at a given temperature *T,* its current state *i* has energy E_i.

2) Perturb the solid by randomly selecting a particle of the solid and relocating it. The perturbation of the particle creates a new state *j* of the solid with energy E_j.

3) Calculate the energy difference, ΔE, before and after the perturbation. Our decision of accepting the new state *j* is made based on the criterion: if ΔE is less than or equal to 0, the new state *j* is accepted as the current state, otherwise, the state *j* is accepted with a probability that is given by

$$\exp(-\frac{\Delta E}{k_B T}) \qquad \text{--- (2.3)}$$

where $\Delta E = E_j - E_i$, *T* denotes the temperature, and k_B is a constant called the *Boltzmann constant*.

4) Repeat 2) and 3) until a sufficient number of perturbations are applied and the thermal equilibrium at this temperature is reached.

The thermal equilibrium of a solid at a given temperature *T* is characterized by the *Boltzmann distribution* (Toda *et al.* 1983). The *Boltzmann distribution* gives the probability of the solid being in a state *i* with energy E_i at the temperature *T,* and is given by

$$p_T\{X = i\} = \frac{1}{Z(T)} \exp\left(\frac{-E_i}{k_B T}\right) \qquad \text{--- (2.4)}$$

where *X* is a stochastic variable that denotes the current state of the solid. *Z(T)* is a normalization function that is defined as

$$Z(T) = \sum_i \exp\left(\frac{-E_i}{k_B T}\right) \qquad \text{---(2.5)}$$

where the summation sums all possible states. Based on the *Boltzmann distribution*, the thermal equilibrium state of the solid denotes that the minimal energy state is reached at the given temperature *T*.

The *Metropolis algorithm* that was originally designed for the physical annealing of solids is adapted to solve combinatorial optimization problems based on two comparisons between them: 1) the generated sequences of states of the solid by perturbations corresponds to a sequence of combinatorial solutions, and 2) the resulting energy of each state of the solid corresponds to the cost or error function in the combinatorial optimization problem. The SA algorithm is composed of a number of iterations of *Metropolis algorithm* by starting the temperature *T* sufficiently high at the beginning and then cooling down gradually. At each given temperature *T*, a number of perturbations are performed and the thermal equilibrium is reached. Then the temperature is decreased to the next lower temperature in a given schedule. This process is called the cooling schedule.

Mathematically, the new state generation and the cooling process in SA can be modeled by Markov chains (Feller 1950). In SA, the generation of each new state *j* depends only on the previous state *i*. The sequence of states can be represented by Markov chains that are described as a set of conditional probabilities $P_{ij}(k)$ for each pair of new and previous states (*i, j*). Let *X(k)* denotes the generate state of the k^{th} perturbation, then we have

$$P_{ij}(k) = \Pr\{X(k) = j \mid X(k-1) = i\} \qquad \text{--- (2.6)}$$

where $P_{ij}(k)$ is the conditional probability that the generated state of the k^{th} perturbation is j, given that the previous state of $(k\text{-}1)^{th}$ perturbation is i. If the conditional probabilities are independent of k or the order of perturbations, the corresponding Markov chains are homogeneous. Otherwise, they are inhomogeneous. SA can be thought to consist of both homogeneous Markov chains generated at a fixed temperature and inhomogeneous Markov chains generated when the thermal equilibrium is obtained at a given temperature and the temperature is decreased to the next lower temperature. Based on the mathematical Markov chain model, SA is proven to be able to converge to a global optimum in combinatorial optimization problems (Geman & Geman 1984; Laarhoven 1988; Aarts & Laarhoven 1987). However, the global convergence of SA requires a large number of iterations or perturbations. It has been shown that the global convergence of SA can be approximated in polynomial time by specifying four parameters in a cooling schedule (Laarhoven 1988; Aarts & Laarhoven 1987): 1) the initial temperature, 2) the decrement function that decreases the temperature, 3) the final temperature to terminate the cooling schedule, and 4) a sufficient number of perturbations at each temperature. When the suitable parameters are chosen and the cooling schedule is properly performed, the global or near-global optimal solution to a combinatorial optimization problem can be attained by SA.

2.2.3 General discussion of SA and unsupervised classification

Unsupervised classification, or 'clustering' is the search for optimal groupings, or clusters based on the similarities between patterns in terms of a distance measure such as Euclidean distance (Jain & Dubes 1988; Tou & Gonzales 1977). Clustering is one type of combinatorial optimization problem. Generally, clustering approaches classify input data by minimizing the distance function defined in Equation 2.2 to estimate the cluster centroids. *K*-means is one of the most popular clustering algorithms applied in remotely sensed applications The *K*-means clustering algorithm has been shown to be computationally efficient and to be able to produce good results if the clusters are compact, well-separated in the feature space, and hyperspherical in shape (Jain *et al.* 2000). However, the classical *K*-means algorithm suffers from the local minimum problem that prevents it from producing satisfactory classification in many occasions.

Jain *et al.* (2000) suggest using generic methods to solve the local minimum problem. One such method is SA. SA can incorporate some randomness to the assignments of cluster labels to pixels in traditional clustering algorithms. The incorporation of SA is likely to reduce the likelihood of traditional clustering algorithm being trapped in a local minimum, and thus improve the resulting classification accuracy. However, to date, SA has not been thoroughly investigated and rarely applied in remote sensing applications due to the fact that SA is computationally expensive and therefore not practical to deal with the large data sets typical in remote sensed applications. The rapidly growing and available computing power in recent years, while enabling faster processing of huge data sets, has facilitated the use of more elaborate and diverse

methods for data classification. SA is one of the most interesting of these methods. It is necessary to conduct some research to develop SA-based classification systems and test their applicability in LU/LC classification using remotely sensed data.

2.3 Review of Artificial Neural Networks Applications in Land cover Classifications

2.3.1 Introduction

Artificial Neural networks (ANNs) originally emerged in 1940's to simulate the behavior of the human nervous system. In late 1980's and early 1990's, ANNs were first used for remotely sensed data classification (Benediktsson *et al.* 1990; Key *et al.* 1989). Subsequently, a number of successful classification applications in remote sensing have been reported (Bruzzone *et al.* 1997; Foody & Arora 1997; Kanellopoulos & Wilkinson 1997; Paola & Schowengerdt 1997; Foody 1995; Bischoff *et al.* 1992; Heermann & Khazenie 1992).

ANNs are fundamentally network systems in which a large number of small and interconnected processors or neurons work together to solve difficult classification and optimization problems. ANN approaches have a distinct advantage over statistical classification methods in that they are non-parametric and require little *a priori* knowledge of the input data distribution (Benediktsson *et al.* 1990). Other superior advantages of ANNs are the ability to estimate the non-linear relationship between the input data and desired outputs, the capability of parallel computation, and the fast generalization capability.

In the following sections, two ANN neural network models in remote sensing are reviewed in section 2.3.2. In section 2.3.3, we present a review on multisource classification using neural networks. A general discussion and summary concerning neural network classification applications in remote sensing are given in section 2.3.4.

2.3.2 Popular ANN models in remotely sensed classification applications

A variety of ANN approaches have been found to be well-suited for a wide variety of LU/LC classification applications in remote sensing (Carpenter *et al.* 1997; Bishop1995; Bezdek *et al.* 1992; Benediktsson *et al.* 1990). Among them, two most popular network models in remote sensing, supervised Multilayer Perceptron network model (Rumelhart & McClelland 1986) and unsupervised Kohonen's Self-Organizing Mapping neural network model (Kohonen 1982), are reviewed respectively in section 2.3.2.1 and 2.3.3.2.

2.3.2.1 Multilayer Perceptron (MLP) neural network model

The MLP neural network using single or multilayer perceptrons is the most popular network model for image classification in remote sensing. MLP networks are usually trained by the supervised Backpropagation (BP) algorithm. The BP learning algorithm was first introduced by Werbos (1974) and then extended by Rumelhart and McClelland (1986). The BP training or learning process requires a set of training patterns with inputs and corresponding desired outputs. Basically, the training process is a search for a network interconnected with a suitable set of weights that can minimize the total differences between the network outputs and desired outputs for the training patterns.

Through the search, the resulting network is believed to be able to approximate the inherent input-output relation. Based on the minimization goal, the mean square error (MSE) J is used as a classification performance criterion given by

$$J = \frac{1}{2N}\sum_{i=1}^{N}\varepsilon_i^2 \qquad \text{--- (2.7)}$$

where N is the number of training patterns. ε_i^2 is the Euclidean distance between the network output of the pattern and the desired output.

A typical MLP consists of one input layer, one or more hidden layers and one output layer. Figure 2.1 shows a typical three-layer MLP neural network architecture with four input nodes in the input layer, 10 hidden nodes in the hidden layer, and 5 output nodes in the output layer. This network architecture is usually denoted as 4 -10 - 5. All nodes in different layers are connected by the associated weights w_{ij}.

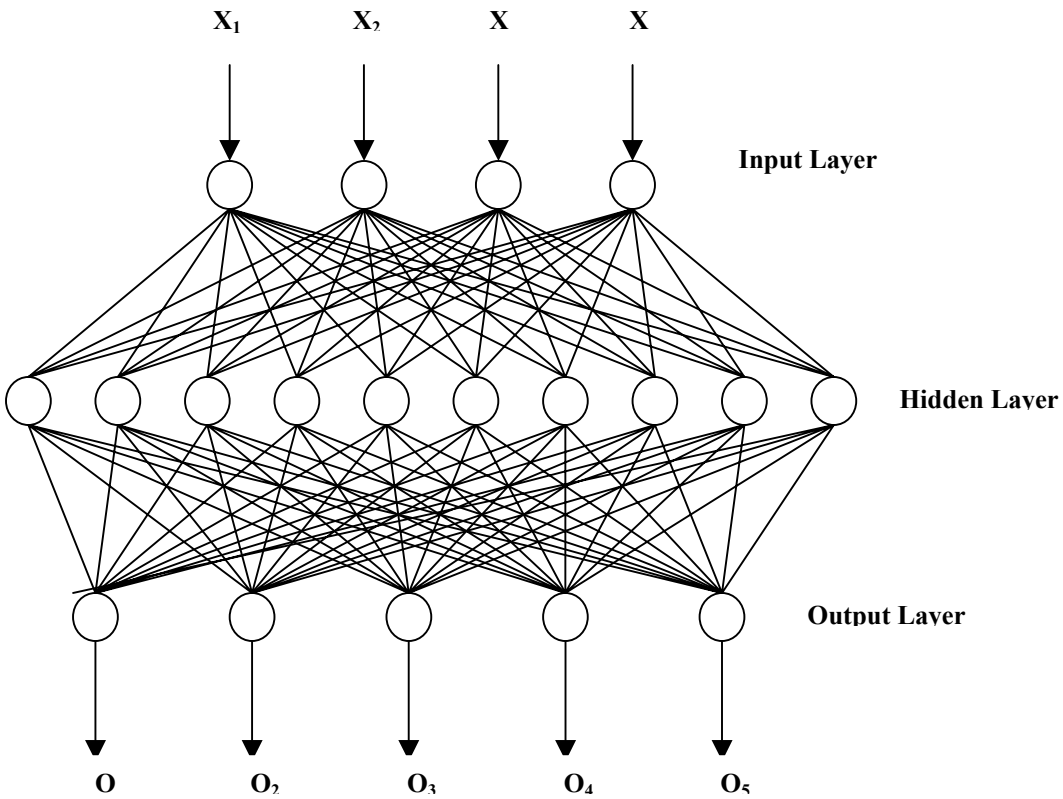


Figure 2. 1 The structure of a three-layer neural network with 4 input nodes, 10 hidden nodes and 5 output nodes

For each input pattern presented to the network, the current network output of the input pattern is computed as follows. First, the net input to any node j in the hidden or output layer is calculated from the nodes in the previous layer as follows:

$$net_j = \sum_{i=1}^{k} (w_{ij} x_i) + b_j \qquad \text{--- (2.8)}$$

where w_{ij} is the weight linking the i^{th} node in the previous layer and the j^{th} node in the next layer. b_j is the bias term associated with this node and x_i is the output or activation

value of the i^{th} node from the previous layer. Then, an activation sigmoid function, *f*, of the net value for each node *j* in hidden and output layers are calculated:

$$f(net_j) = \frac{1}{1 + e^{(-net_j)}} \qquad \text{--- (2.9)}$$

The outputs of the nodes in the hidden layer and the output layer depend on the inputs from previous layer, their activation function and the bias. In the next step, the outputs from the networks are compared with the desired outputs and the error or difference between them is then backpropagated to adjust the weights between layers based on the following generalized delta rule:

$$w_{ij}(n+1) = w_{ij}(n) + \eta \delta_j(n) x_i(n) \qquad \text{--- (2.10)}$$

where $w_{ij}(n+1)$ and $w_{ij}(n)$ are the weight at iteration *n+1* and iteration *n*;

n: iteration,

η is the learning rate,

$x_i(n)$ is the activation value at the i^{th} node in the previous layer,

$\delta_j(n)$ is the backpropagated error of the j^{th} node in the current layer that is defined as:

$$\delta_j(n) = O_j(1 - O_j)\sum_{i=1}^{k} \delta_i w_{ij} \qquad \text{--- (2.11)}$$

where O_j is the output of the j^{th} node,

δ_i is the backpropagated error of the i^{th} node in the previous layer,

k is the number of nodes in the previous layer.

The generalized delta rule used to update the weights is computationally slow. Principe *et al.* (1999) suggest including a memory term (the past increment to the weight) to speed up and stabilize the BP learning. In this *momentum learning*, the generalized delta rule in Equation 2.10 is modified by taking incremental changes as follows:

$$w_{ij}(n+1) = w_{ij}(n) + \eta \delta_j(n) x_i(n) + \alpha \Delta w_{ij}(n-1) \qquad \text{--- (2.12)}$$

where α is the momentum constant,

$\Delta w_{ij}(n-1)$ is the weight change from the previous iteration.

Detailed implementation of MLPs is described by Principe *et al.* (1999). A number of studies have been performed to apply MLP neural networks to classify remotely sensed data (Kanellopoulos & Wilkinson 1997; Paola & Schowengerdt, 1997; Foody *et al.* 1995).

2.3.2.2 Kohonen's Self-Organizing Mapping (KSOM) Neural Network Model

Kohonen's Self-Organization Mapping (KSOM) network, proposed by Kohonen (1982), is a relatively simple clustering tool to group dimensionally complex data into clusters through competitive learning. KSOM network is a competitive two-layer network composed of: 1) an input layer with a number of input nodes, 2) an output or competitive layer with a number of competitive nodes arranged by rows and columns, and 3) a set of connection weights linking input nodes with competitive nodes. Figure 2.2 shows the structure of a KSOM network with four input nodes and a competitive layer with 3 rows and 3 columns. The training or learning process of KSOM is to search for a set of optimal weight vectors that minimize their distances from all input patterns.

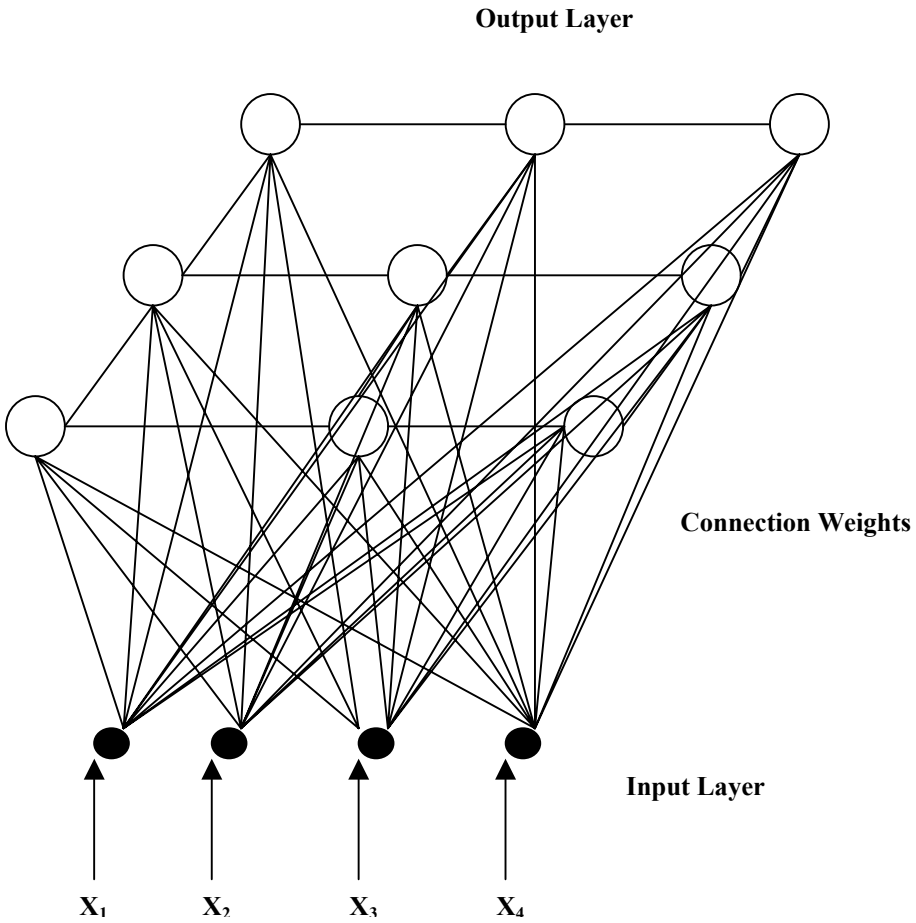


Figure 2. 2 The structure of a Kohonen's Self-Organization Mapping Neural network with 4 input nodes and 9 output nodes

During the competitive learning, when an input pattern is presented to the KSOM network, the node in the competitive layer with a weight vector closest to the input vector in terms of Euclidean distance is declared as the 'winning' node. Only this winning node and the nodes in its neighborhood are allowed to update their weight vectors. It is the spatial neighborhood property that makes KSOM networks different from other competitive networks (Principe *et al*, 1999). Via the spatial neighborhood property, KSOM networks have the capability to preserve and reveal the inherent topological

relationships in the input data. In remotely sensed classification applications, the topological relationships are demonstrated in a way that the pixels with similar spectral values are assigned to the neighboring output nodes in the competitive layer. The KSOM algorithm (Lippmann 1987; Chen *et al.* 1999) is described as below:

a) Initialize weights from *n* inputs to the *k* output nodes to the interval [0,1] randomly, and set the initial radius of neighborhood $N_g(t)$. Select the iteration limit for the algorithm, *iteration_max.*

b) Present new input pattern to network.

c) Compute the Euclidean distance $d_{ij}(t)$ between the input pattern X_i and each output node *j* using the equation as below

$$d_{ij} = \sum_{i=1}^{n} \left(x_i(t) - W_{ij}(t)\right)^2 \qquad \text{--- (2.13)}$$

d) Select the output node with the minimal distance as the winning node.

e) Update the weights of the winning node and all the nodes within its neighborhood

$$W_{ij}(t+1) = W_{ij}(t) + \mathrm{N_g}(t)\eta(t)\left(x_i(t) - W_{ij}(t)\right) \qquad \text{--- (2.14)}$$

where $\eta(t)$ denotes the learning rate at the iteration *t*, which is usually decreased as *t* increases, and $N_g(t)$ the neighborhood radius at the iteration *t*.

f) Send the next input pattern to the input layer, and then go to step c) until all learning patterns are sent once.

g) Update the learning rate and the neighborhood radius.

h) Increment *t* by 1, and then go to step b) until *t* = *iteration_max* or the average distance between the input pattern and the winning node drops below a specific threshold.

The KSOM algorithm can be viewed as a neural network version of *K*-means. However, unsupervised KSOM approach has some important characteristics over *K*-means (Goncalves *et al.* 1998; Nour & Madey 1996):

- Self-organizing and adaptive learning process
- Preservation and revelation of inherent topological relationship among input data.
- Parallel architecture

In many previous studies, KSOM networks have been shown to be effective in remotely sensed classification applications (Babu 1997; Baraldi & Parmiggiani 1995; Bezdek *et al.* 1992). However, Bezdek *et al.* (1992) have demonstrated that the performance of the KSOM network model heavily depends on an optimal combination and initialization of the specified parameters.

2.3.3 Multisource image classification using artificial neural networks

During recent decades advances in space and computer technologies have made accessible large amounts and various types of data about the Earth and its environment. Abundant data from various remote sensors offer the potential to fuse data from various sources to address different environmental problems. Inclusion of complementary information from additional data sources, or "data fusion", can help improve classification accuracy (Hegarat-Mascle *et al.* 1997; Lou & Kay 1989). Thus, it is

mandatory to develop effective data fusion techniques to take advantage of the complimentary information from a variety of spatial data sets in multisource classification.

Multisource classifications using statistical methods and ANN approaches are a heavily studied topic in remote sensing (Paola & Schowengerdt 1997; Benediktsson *et al.* 1990; Benediktsson & Swain 1989). Based on our review, general statements with respect to these two types of fusion techniques are drawn as follows:

- Traditional statistical methods such as maximum-likelihood and minimum-distance are generally incapable in multisource classification using remotely sensed data for two reasons: 1) the statistical distributions of multiple source data sets may significantly vary from both classes and data sources and thus are unlikely to be accurately estimated by a multivariate statistical model (Paola & Schowengerdt 1997), and 2) most traditional statistical methods do not have the capability to weight the data sources even though multiple data sources are not equally reliable (Benediktsson *et al.* 1990). Some advanced statistical methods have been developed and shown to be able to perform very well in multisource classification (Solberg *et al*. 1996). However, these statistical methods are very complicated requiring the derivation of a sufficiently accurate multivariate statistical model and the proper weighting and modeling of multiple data sources (Benediktsson *et al.* 1990).
- Comparatively, ANN approaches are surprisingly simple because they are nonparametric and distribution-free. By an iterative learning process, ANN

approaches can extract sufficient information from the training data to approximate the inherent input-output relationships without the need to estimate statistical distribution models of multiple source data sets. Thus, ANN approaches have become particularly attractive for data fusion. Many studies have shown that ANN approaches have similar or superior classification performance to statistical methods in multisource classifications (Bruzzone *et al.* 1999; Serpico & Roli 1995; Benediktsson *et al.* 1990). However, their performance is sensitive to the quality and quantity of collected training data. They may fail to generalize well for those samples poorly represented in training set.

- In recent years some hybrid methods have been investigated to combine statistical and neural network models (Bruzzone *et al.* 1999; Benediktsson & Kanellopoulos 1999; Benediktsson *et al.* 1997). These hybrid methods are generally very complicated and require analysts to have very good understanding of both the statistical and neural network models.
- In summary, both the statistical and ANN techniques have their pros and cons. In practice, the analyst should be clearly aware of their limitations and makes a wise choice that is appropriate for their applications in terms of classification accuracy, feasibility, and cost.

In many multisource classification applications using remotely sensed data, two factors may have significant effects on the resulting classification performance. The first factor is the temporal difference between the acquired multiple data sources. The time interval between the acquisitions of the multiple data sets is usually ignored by assuming

that all the images are acquired at the same time, or there is no class change occurring during the time interval (Serpico & Roli 1995). However, this ideal condition is not always satisfied in practice. Significant class changes occurring during the period may cause a great deal of interpretation confusion and result in high misclassification error. The second factor is the effect of the geometrical co-registration of multiple images with different spatial, spectral, and projection systems into one common reference system. Dai and Khorram (1998) point out that the misregistration error has significant effects on multisource analysis such as data fusion and change detection. Large misregistration error may render the fused data useless for classification, even resulting in poorer classification accuracy than that using single remote sensing data.

2.3.4 General discussion and summary

The introduction of ANN is one of the most significant advancements in the field of image classification. Numerous studies have demonstrated that ANN approaches perform well over a range of traditional statistical methods (Carpenter *et al.* 1997; Foody *et al.* 1995; Bischof *et al.* 1992; Benediktsson *et al.* 1990). On the other hand, ANN approaches suffer from many limitations: 1) no generally applicable criteria for defining a suitable network architecture, 2) high dependence on training conditions, i.e., quality and quantity of training data, initial weights, and training parameters, and 3) low training speed (Bischof & Leonardis 1998; Foody & Arora 1997; Serpico & Roli 1995). To date, there are no explicit rules to be derived to solve these problems. Based on our review on a large number of ANN peer studies, these issues are discussed.

- Network architecture

The design of network architecture has influence on the classification performance of a neural network (Foody 1999). In a MLP network, we commonly have the number of input nodes equal to the number of input variables, i.e., spectral bands, and let the number of output nodes equal to the number of desired classes. Once the input and output layers are determined, the analyst must decide how many hidden layers and hidden nodes to use (Foody & Arora 1997). Although there is no universal guideline to determine the number of hidden layer and hidden nodes, there are a number of 'rules of thumb' that may be used. Kanellopoulos and Wilkinson (1997) and Lippman (1989) state that single hidden layer networks are sufficient for most classification problems and the number of hidden nodes should be at least four times the number of input nodes or twice the number of the output nodes.

- Quality and quantity of training data

The importance of the selection of training data in neural network has been recognized and emphasized by many researchers. Benediktsson *et al.* (1990) suggest that it is essential that the acquired training data sets have the ability to provide a complete representation of each of the desired classes in ANN classifications. Otherwise, the neural networks having high classification accuracy for the training data may fail to generalize well for those samples poorly represented in training data sets. The size of the training set should increase considerably along with the increases of the number of the input dimension or spectral bands, the spectral variability of desired classes, and the desired classification accuracy (Principe *et al.* 1999; Foody and Arora 1997; Zhuang *et al.* 1991).

- Network training speed

One shortcoming of ANNs is that they demand long training times (Benediktsson & Sveinsson 1997). The training speed is generally associated with the learning rate and the selected training algorithm. There is a trade-off between fast training using a larger learning rate and the stabilization of network training. Learning rates are application-related and normally determined by experimentation. To stabilize the network training, the learning rate must be kept sufficiently small. However, a very small learning rate will make the training very slow, which may not be realistic for practical implementation. Thus, it is suggested that a slightly larger learning rate is used at the early stage of training to speed the training speed and gradually decreased to stabilize the training. An advanced training algorithm such as the conjugated gradient (CG) algorithm has been investigated to speed up network training and improve the training performance (Goncalves *et al.* 1998; Benediktsson and Sveinsson 1997; Fitch *et al.* 1991). In the CG training algorithm, the optimal learning rate is calculated at each iteration during the training stage.

- Generalization and overfitting

In classification applications, the success of an ANN approach is evaluated not only based on its performance on the training data, but also, more importantly, on its generalization capability on the entire population data beyond the acquired training data. In neural network classification, it is very common that a well-trained network with a very low training error fails to generalize well when applied to classify those data beyond the training data. This phenomenon is called overfitting (Principe *et al.* 1999). Overfitting

arises when the classes are not properly represented in training data sets, or the network is over-trained. The over-trained network learns specifics in the particular training set that are not the typical characteristics of the whole data set. To avoid the overfitting, a cross-validation approach is proposed to terminate the training before the overfitting occurs (Principe *et al.* 1999). This cross-validation approach is shown to be able to overcome the overfitting problem and improve the generalization capability of the trained network (Paola & Schowengerdt 1995). Furthermore, this approach can avoid wasting time to training an ineffective network.

- Local minima problem

Amar *et al.* (1995) suggest that the neural network training that starts with a random weight initialization may suffer a potential local minimum problem. Yuan *et al.* (1999) have shown that Simulated Annealing (SA) has the potential to overcome the local minima problem in the *K*-means clustering algorithm. We incorporate SA into neural networks to reduce the likelihood to get trapped by local minima in Chapter 4.

- Summary

Given the pros and cons of neural networks, many recent studies have been moved in the direction of developing hybrid approaches and fuzzy approaches to enhance the classification capability of neural networks. Hybrid classification approaches have been investigated in recent years to improve the quality of acquired training data (Goncalves *et al.* 1998; Nogami *et al.* 1997; Yoshida and Omatu 1994). The common practice in these studies is to use the clustering capability of unsupervised KSOM

networks to assist the manual training data collection. However, no automated training data selection is observed from any of them. Thus, it is necessary to develop an Automated Data Selector (ADS) to help select sufficient high quality training data providing complete representation of classes. Although many successful application examples have demonstrated that ANN approaches have great potential in remotely sensed classification applications using single and multiple remotely sensed data sets, their uses are still limited to those neural network professionals. To popularize the usability of neural network in remote sensing, a fully automated neural network system is needed.

References

Aarts, E. H. L., and P. J. M. Van Laarhoven, "Simulated annealing: a pedestrian review of the theory and some applications", *Pattern recognition theory and applications,* pp.179-192, 1987.

Anderson, J. R., *et al.*, "A land use and land cover classification system for use with remotely sensed data," *Washington, DC: U.S. Geological Survey Professional Paper 964*, pp. 28, 1976.

Ahern, F.J., *et al.*, "Simultaneous microwave and optical wavelength observations of agricultural targets", *Canadian Journal of Remote Sensing*, vol. 4, no. 2, pp. 127 – 142, 1978.

Amar, F., *et al.*, "Comparison of neural network algorithms for remote sensing applications," *IEEE Proceedings. IGARSS 1995*, vol. 1, pp. 694 – 696, 1995.

Babu, G. P., "Self-organizing neural networks for spatial data," *Pattern Recognition Letters*, vol. 18, pp. 133 –142, 1997.

Ball, G. H., and D. J. Hall, "ISODATA: a novel method data analysis and pattern classification," *Technical report, Stanford Research Institute, Menlo Park, California, USA*, 1965.

Bandyopadhyay, S., S. K. Pal, and C. A. Murthy, "Simulated annealing based pattern classification," *Journal of Information Sciences*, vol. 109, pp. 165 – 184, 1998.

Baraldi, A. and F. Parmiggiani, "A neural network for unsupervised categorization of multivalued input patterns: an application to satellite image clustering," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 33, pp. 305 - 316, 1995.

Belward, A. S., *et al.*, "An unsupervised approach to the classification of semi-natural vegetation from Landsat Thematic Mapper data," *International Journal of Remote Sensing*, vol.11, pp.429 – 445, 1990.

Benediktsson, J. A., and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 37, pp. 1367 - 1377, 1999.

Benediktsson, J. A., and J. R. Sveinsson, "Feature extraction for multisource data classification with artificial neural networks," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 727-740, 1997.

Benediktsson, J.A., *et al.*, "Parallel consensual neural networks," *IEEE Transaction on Neural Networks*, vol. 8, pp. 54 - 64, 1997.

Benediktsson, J.A., P. H. Swain, and O. K. Ersoy, "Neural network approaches

versus statistical methods in classification of multi-source remote sensing data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 28, pp. 540-551, 1990.

Benediktsson, J.A., P. H. Swain, "A method of statistical multisource classification with a mechanism to weight the influence of the data sources," *IEEE Proceedings. IGARSS 1989*, pp. 517 – 520, 1989.

Bezdek, J. C., *et al.* "Fuzzy Kohonen clustering networks," *IEEE International Conference on Fuzzy Systems*, pp. 1035 – 1043, 1992.

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

Bilbo, G. L., *et al.*, "Mean field annealing: a formalism for constructing GNC-like algorithms," *IEEE Transaction on Neural Networks*, vol. 3, no. 1, pp. 131 – 138, 1992.

Bischof, H., W. Schneider, and A. J. Pinz, " Multispectral classification of landsat images using neural networks," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 30, no. 3, pp. 482 – 490, 1992.

Bishop, C. M, Radial basis functions. In *Neural networks for pattern recognition*. *Oxford, New York: Clarendon Press*, 1995.

Bolstad, P.V, and T. M. Lillesand, "Rapid maximum likelihood classification," *Photogrammetic Engineering & Remote Sensing*, vol. 57, pp. 64 – 74, 1991.

Brisco, B., and R.J. Brown, "Multidate SAR/TM synergism for crop classification in western Canada", *Photogrammetic Engineering & Remote Sensing*, vol. 61,

no. 8, pp. 1009 – 1014, 1995.

Brisco, B., "Multi-dimensional approaches to the RADAR remote sensing of land cover", *Ph.D. thesis, Department of Geography, University of Kansas, Lawrence, KS*, 1985.

Brown, D. E., and C. L. Huntley, "A practical application of simulated annealing to clustering," *Pattern recognition*, vol. 25,no. 4, pp. 401-412, 1992.

Brown, J. F., *et al.*, "Using multisource data in global land-cover characterization: concepts, requirments, and methods", *Photogrammetic Engineering & Remote Sensing*, vol. 59, pp. 977-987, 1993.

Bruzzone, L., D. F. Prieto, and S. B. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1350 – 1359, 1999.

Bruzzone, L., C. Conese, F. Maselli, and F. Roli, "Multisource classification of complex rural areas by statistical and neural network approaches," *Photogrametric Engineering and Remote Sensing*, vol. 63, pp. 523 – 533, 1997.

Carpenter, G.A., M.N. Gjaja, S. Gopal, and C.E. Woodcock, "ART neural networks for remote sensing: vegetation classification from Landsat TM and terrain data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 35, no. 2, pp. 308-325, 1997.

Cerny, V., "Thermodynamical approach to the traveling salesman problem: an efficient

simulation algorithm," *Journal of Optimization Theory and Applications*, vol. 45, pp. 45 –51, 1985.

Chen, Z., *et al.*, "Texture segmentation based on Wavelet and Kohonen network for remotely sensed images," *IEEE SMC'99 Conference Proceedings*, vol. 6, pp. 816 – 821, 1999.

Cihlar, J., *et al.*, "Classification by progressive generalization: a new automated methodology for remote sensing multichannel data," *International Journal of Remote Sensing*, vol. 19, no. 14, pp. 2685 - 2704, 1998.

Couloigner, *et al.*, "Benefit of the future SPOT-5 and of data fusion to urban roads mapping", *International Journal of Remote Sensing*, vol. 19, no. 8, pp. 1519 – 1532, 1998.

Dai, X. L., and S. Khorram, "Data fusion using artificial neural networks: a case study on multitemporal change analysis," *Computers, Environments, and Urban Systems*, vol. 23, pp. 19 – 31, 1999.

Dai, X. L., and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 36, no. 5, pp. 1566 - 1577, 1998.

Dai, X. L., and S. Khorram, "Development of a new automated land cover change detection system from remotely sensed imagery based on artificial neural networks," *IEEE Proceedings. IGARSS 1997*, pp. 1029 – 1031, 1997.

Dubes, R., and A. K. Jain, "Cluster methodologies in exploratory data analysis," Advances in Computing, vol. 19, pp. 113 – 228, 1980.

Dubes, R., and A. K. Jain, "Cluster techniques: the user's dilemma," *Pattern Recognition*, vol. 8, pp. 247 – 260, 1976.

Duda, R. O., and P. E. Hart. Pattern classification and scene analysis. *New York: Wiley-Interscience*. 1973.

Ediriwickrema, J., and S. Khorram, "Hierarchical maximum-likelihood classification for improved accuracies," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 810- 816, 1997.

Elachi, C.T., "Spaceborne RADAR Remote sensing: applications and techniques", *IEEE Press*, New York, pp.255, 1988.

Eppler, W. G., "An improved version of the table look up algorithm for pattern recognition," in *Proceedings, of 9th International Symposium on Remote Sensing of Environment*, Ann Arbor, MI, pp. 793 – 812, 1974.

Ersoy, O. K., and D. Hong, "Parallel self-organizing hierarchical neural networks," *IEEE Transaction on Industrial Electronics*, vol. 40, no. 2, pp. 218 – 227, 1993.

Evans, D.L., *et al.*, "Mapping forest distributions with AVHRR data", *World Resource Rev.,* vol. 5, pp. 66-71, 1993.

Feller, W., An introduction to probability theory and its applications. *New York: Wiley*, 1950.

Fitch, J. P., *et al.*, "Ship wave detection procedure using conjugate gradient trained neural network," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 29, 1991.

Foody, G. M., "The significance of border training patterns in classification by a

feedforward neural network using back propagation learning," *International Journal of Remote Sensing*, vol. 20, no. 18, pp. 3549 – 3562, 1999.

Foody, G. M., and M. K. Arora, "An evaluation of some factors affecting the accuracy of classification by an artificial neural network," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 799-810, 1997.

Foody, G. M., "Approaches for the production and evaluation of fuzzy land cover classifications from remotely sensed data," *International Journal of Remote Sensing*, vol. 17, pp. 1317 – 1340, 1996.

Foody, G.M., M.B. McCulloch, and W.B. Yates, "The effect of training set size and composition on artificial neural network classification," *International Journal of Remote Sensing*, vol. 16, pp. 1707 – 1723, 1995.

Foody, G. M., et al., "Derivation and application of probabilities of class membership form the maximum-likelihood classification," *Photogrammetic Engineering & Remote Sensing*, vol. 58, pp. 1335 – 1341, 1992.

Frank, T. D., "Mapping dominant vegetation communities in the Colorado Rocky Mountain front range with Landsat Thematic Mapper and digital terrain data," *Photogrammetic Engineering & Remote Sensing*, vol.54, no.12, pp.1727 – 1734, 1988.

Franklin, S. E. "Terrain analysis from Digital patterns in geomorphometry and Landsat MSS spectral response", *Photogrammetic Engineering & Remote Sensing*, vol. 53, no. 1, pp. 59 – 65, 1987.

Fukunaga, K., Introduction to statistical pattern recognition. 2nd. *New York: Academic*,

1990.

Geman, S., and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Iintelligence*,, vol. 6, pp. 721 – 741, 1984.

Goncalves, M. L., *et al.*, "A neural architecture for the classification of remote sensing imagery with advanced learning algorithms," *Proceedings of IEEE Signal Processing Society Workshop*, pp. 577 – 586, 1998.

Gong P. *et al*., "An assessment of some factors influencing multispectral land-cover classification", *Photogrammetic Engineering & Remote Sensing*, vol.56, no. 5, pp. 597 – 603, 1990.

Guneriussen, T., *et al.*, "Snow monitoring using EMISAR and ERS-1 data within the European multi-sensor airborne campaign EMAC-95," *the 1997 IEEE International Geoscience and Remote Sensing Symposiums*, vol.2, pp. 631 –633,1997.

Haack, B.N., *et al.*, "Assessment of Landsat MSS and TM data for urban and near-urban landcover digital classification", *Remote Sensing of Environment*, vol. 21, no. 2, pp. 201 – 213, 1987.

Harrison, A.R., and T.R. Richards, "Multispectral classification of urban land use using SPOT HRV data", *Digest – International Geoscience and Remote Sensing Symposium, Edinburgh, U.K.*, pp. 205 – 206, 1988.

Heermann, P. D., and N. Khazenie, "Classification of multispectral remote sensing

data using a back-propagation neural network," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 30, pp.81-88, 1992.

Hegarat-Mascle, S. L., S. D. Vidal-Madjar, and P. Olivier, "Applications of simulated annealing to SAR image and classification problems," *International Journal of Remote Sensing*, vol. 17, no. 9, pp. 1761-1776, 1996.

Jaakkola, S., *et al.*, "Applicability of SPOT for forest inventory, mapping and change monitoring*", SPOT –1 First In-flight results (Toulouse Cepadues-Editions)*, pp. 259 – 264, 1987.

Jain, A. K., R. P. W. Dulin, and J. C. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22, no. 1, pp. 4 – 37, 2000.

Jain, A. K., Fundamentals of digital image processing. *Englewood Cliffs, New Jersy: Prentice-Hall,* pp. 414 – 421, 1989.

Jain, A.K., and R. Dubes, Algorithms for Clustering Data. *Englewood Cliffs, New Jersy: Prentice-Hall*, pp. 92 – 117,1988.

Jenson, J. R., Introductory digital image processing. *Englewood Cliffs, NJ: Pretentice Hall,* 1996.

Kanellopoulos, I., and G.G. Wilkinson, "Strategies and best practice for neural network image classification," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 711 – 725, 1997.

Key, J., *et al.*, "Classification of merged AVHRR and SMMR arctic data with neural networks," *Photogrammetic Engineering & Remote Sensing*, vol.55,

pp. 1331 – 1338, 1989.

Khorram, S., *et al.*, "Comparison of Landsat MSS and TM data for urban land-use classification", *IEEE Transactions on Geoscience and Remote Sensing*, vol.25, no. 2, pp.238 – 243, 1987.

Kirkpatrick, S., C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671-688, 1983.

Klein, R. W., and R. C. Dubes, "Experiments in projection and clustering by annealing," *Pattern Recognition*, vol. 22, no. 2, pp. 213-220, 1989.

Kohonen, T., "Self-orgazning formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, pp. 56 – 69, 1982.

Kolonko, M., and M. T. Tran, "Convergence of simulated annealing with feedback temperature schedules," *Probability in the Engineering and Informational Sciences*, vol. 11, pp. 279-304, 1997.

Kux, H. J. H., *et al.*, "Evaluation of RADARSAT for land use and land cover dynamics in the Southern Brazilian Amazon State of Acre", *Canadian Journal of Remote Sensing*, vol. 24, n. 4, pp. 350 –359, 1998.

Laarhoven, P.J.M., Theoretical and computational aspects of simulated annealing. *Amsterdam, Netherlands: Center for mathematics and computer science,* 1988.

Laba, M., S. D. Smith, and S. D. Degloria, "Landsat-based land cover mapping in the lower Yuna River watershed in the Dominican Republic," *International Journal of Remote Sensing*, vol. 18, no. 14, pp. 3011 – 3025, 1997.

Lark, R. M, "A reappraisal of unsupervised classification, I: correspondence between

spectral and conceptual classes," *International Journal of Remote Sensing*, vol. 16, no. 8, pp. 1425 – 1443, 1995.

Lee, J., *et al.*, "A neural network approach to cloud classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, pp.846 – 855, 1990.

Lee, T., J. A. Richards, and P. H. Swain, "Probablistic and evidential approaches for multisource data analysis," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 283 – 293, 1987.

Lillesand, T. M., and R. F. Keifer, Remote sensing and image interpretation. 4th. *NewYork: Wiley*, 2000.

Lippmann, R. P, "Pattern recognition using neural networks," *IEEE Communications Magazine*, pp. 47 – 64, 1989.

Lippmann, R. P., An introduction to computing with neural networks. *IEEE Acoust. Speech Signal Process.*, pp. 4 – 22, 1987.

Loveland, T. R., *et al.*, "Development of landcover characteristics database for the conterminous U.S.", *Photogrammetic Engineering & Remote Sensing*, v.(57), pp. 1453-1463, 1991.

Luo, R. C., and M. G. Kay, "Multisensor integration and fusion in intelligent systems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, pp. 901 – 931, 1989.

Malingreau, J. P., *et al.*, "AVHRR for monitoring tropical deforestation", *International Journal of Remote Sensing*, vol.10, pp. 855 – 867, 1989.

Mao, J., K. Mohiuddin, and A. K. Jain, "Parsimonious network design and feature

selection through node pruning," *Proc. 12th International Conference on Pattern Recognition*, pp. 622 –624, 1994.

Marth, P. M., "A computationally-efficient maximum-likelihood classifier employing prior probabilities for remotely sensed data," *International Journal of Remote Sensing*, vol. 6, no. 2, pp. 369 – 376, 1985.

Maselli, F., C. Conese, L. Petkov, and R. Resti, "Inclusion of priori probabilities derived from nonparametric process into the maximum-likelihood classifier," *Photogrammetic Engineering & Remote Sensing*, vol. 58, pp. 201 – 207, 1992.

Mausel, P.W., W.J., Kamber, and J. K. Lee, "Optimum band selection for supervised classification of multispectral data," *Photogrammetic Engineering & Remote Sensing*, vol.56, no.1, pp.55 – 60, 1990.

McErlean, F. J., D. A. Bell, and S.I. McClean, "The use of simulated annealing for clustering data in databases," *Information Systems*, vol. 15, no. 2, pp. 233-245, 1990.

Metropolis, N., *et al.*, "Equation of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087-1092, 1953.

Nour, M. A., and G. R. Madey, "Heuristic and optimization approaches to extending the Kohonen slef-organizing algorithm," *European Journal of Operational Research*, vol. 93, pp. 428 – 448, 1996.

Quegan, S., *et al.*, "Combining unsupervised and knowledge-based method in large-scale forest classification," *IEEE Proceedings. IGARSS 2000*, vol. 1, pp. 426 – 428, 2000.

Palylyk, C. L, and P. H., Crown, “Application of clustering to Landsat MSS digital data for Peatland inventory,” Canadian Journal of Remote Sensing, vol. 10, pp. 201 – 208, 1984.

Paola, J. D., and R. A. Schowengerdt, “The effect of neural-network structure on a multispectral land-use/land-cover classification,” *Photogrametric Engineering and Remote Sensing*, vol. 63, pp. 535 – 544, 1997.

Paola, J. D., and R. A. Schowengerdt, “A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, pp. 981 – 996, 1995.

Paris, J. F., and H. H. Kwong, “Characterization of vegetation with combined Thematic Mapper (TM) and shuttle imaging RADAR (SIR-B) image data”, *Photogrammetic Engineering & Remote Sensing*, vol. 54, no. 8, pp. 1187 – 1193, 1988.

Pierce, L., *et al.*, “An automated unsupervised/supervised classification methodology,” *IEEE Proceedings. IGARSS 1998*, vol. 4, pp. 1781 – 1783, 1998.

Principe, J. C., N. R. Euliano, and W. C. Lefebvre, Neural and adaptive systems: fundamentals through simulations. *John Wiley & Sons, Inc.*, pp. 656, 1999.

Remund, Q. P., D. G. Long, and M. R. Drinkwater, “Polar sea-ice classification using enhanced resolution NSCAT data,” ," *IEEE Proceedings. IGARSS 1998*, vol. 4, pp. 1976 - 1978, 1998.

Richards, J. A., Remote sensing digital image analysis. *Springer-Verlag*, 1986.

Rose, K., "Deterministic annealing for clustering, compression, classification, regression, and related optimization problem," *IEEE Proceeding*, vol. 86, pp. 2210 – 2239, 1998.

Rumelhart, D. E., and J. L. McClelland, Parallel distributed computing. *PDP research group, Cambridge, MA: MIT Press*, vol. 3, 1986.

Running, S. W., *et al.*, " A remote sensing based vegetation classification logic for global land cover analysis", *Remote sensing environment*, v.(50), pp.39-48, 1995.

Sader, S. A., D. Ahl, W. S. Liou, "Accuracy of Landsat-TM and GIS rule –based methods for forest wetland classification in Marine," *Remote Sensing of Environment*, vol.53, pp. 133 – 144, 1995.

Schalkoff, R. J., Pattern recognition: statistical, structural, and neural approaches. *John Wiley & Sons, Inc.*, 1992.

Schowengerdt, R. A., Techniques for image processing and classification in remote Sensing. *New York: Academic Press*, pp. 129 – 214,1983.

Selim, S. Z., and K. Alsultan, "A simulated annealing algorithm for the clustering problem," *Pattern Recognition*, vol. 24, no. 10, pp.1003-1008, 1991.

Selim, S. Z., and M.A. Ismail, "K-means-type algorithms: a generalized convergence theorem and characterization of local optimality," *IEEE Transactions onPattern Analysis and Machine Intelligence*, vol. 6, pp. 81 - 87, 1984.

Serpico, S. B., and F. Roli, "Classification of multisensor remote-sensing images by structured neural networks," *IEEE Transactions on Geoscience* and Remote Sensing, vol. 33, no. 3, pp. 562-578, 1995.

Shimabukuro, *et al.*, "Land cover classification from RADARSAT data of the Tapajos National forest, Brazil", *Canadian Journal of Remote Sensing*, vol. 24, no. 4, pp. 393 – 401, 1998.

Shlien, S., and A. Smith, "A rapid method to generate spectral theme classification of Landsat imagery," *Remote Sensing of Environment*, vol. 4, pp. 67 – 77, 1977.

Snyder, W., *et al.*, "Segmentation brain images using mean field annealing," *Inform. Process. Med. Imag., Proc. 12th Int. Conf. IPMI, Wye, UK,* pp. 218 – 226, 1991.

Solberg, A. H. S., T. Text, and A. K. Jain, "A Markov random field model for classification of multisource satellite imagery,"*IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, pp. 100 – 112, 1996.

Solberg, A. H. S., *et al.*, "Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images*", IEEE Transactions on Geoscience* and Remote Sensing, vol. 32, no. 4, pp.768 – 778, 1994.

Strahler, A.H., "The use of prior probabilities in maximum likelihood classification of remotely sensed data," *Remote Sensing of the Environment*, vol. 10, pp. 135 – 163, 1980.

Swain, P. H., and S. M. Davis, Remote sensing: the quantitative approach. *New York: McGraw-Hill*, pp. 136 – 188,1978.

Thomson, A.G., R.M. Fuller, and J.A. Eastwoods, "Supervised versus unsupervised methods for classification of coasts and river corridors from airborne remote sensing," *International Journal of Remote Sensing*, vol. 18, pp. 3423 – 3431, 1998.

Toda, M., R. Kubo, and N. Saito, Statistical physics. *Springer-Verlag, Berlin*, 1983.

Townshend, J. R., *et al.*, "The 1km resolution global data set: needs of the international geosphere bioshpere programme," *International Journal of Remote Sensing*, vol. 15, pp. 3417 - 3441, 1994.

Townshend, J. R., "Global data sets for land applications from the advanced very high resolution radiometer: introduction", *International Journal of Remote Sensing*, vol. 15, pp.3319-3332, 1994.

Townshend, J.R., and C.O. Justice, "Analysis of the dynamics of African vegetation using the normalized difference vegetation index", *International Journal of Remote Sensing*, vol.7, pp. 1435 – 1445, 1986.

Tou, J. T., and R. C. Gonzales, Pattern recognition principles, *Addison- Wesley Publishing Company,* 1977.

Ulaby, F.T., *et al.*, "Crop classification using airborne RADAR and Landsat data", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 20, pp.518 – 528, 1982.

Xiao, R. R., R. Wilson, and R. Carande, "Neural network classification with IFSAR and multispectral data fusion," *IEEE Proceedings. IGARSS 1998*, vol. 3, pp. 1327 – 1329, 1998.

Yoshida, T., and S. Omatu, "Nueral network approach to land cover mapping," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 32, pp.1103 - 1109, 1994.

Yuan, H., S. Khorram, and X. L. Dai, "Applications of simulated annealing minimization technique for unsupervised classification of remotely sensed data," *IEEE Proceedings. IGARSS 1999*, vol. 1, pp. 134 – 136, 1999.

Zhuang, X., B. A. Engel, M. F. Baumgardner, and P. H. Swain, "Improving classification of crop residues using digital land ownership data and Landsat TM imagery," *Photogrammetic Engineering & Remote Sensing*, vol. 57, no. 11, pp. 1487 – 1492, 1991.

3. Paper #1 Development and Evaluation of Land-Use/Land-Cover Classification Systems using Simulated Annealing

ABSTRACT –

Many previous studies have shown the classical unsupervised approaches *K*-means failed to produce satisfactory classification due to the limitation of local minima. To tackle this problem, this research was designed to develop algorithms and methodologies of Simulated Annealing (SA) systems based on *K*-means and to test their applicability for land cover classification using remotely sensed data. SA is developed based on the analogy between the physical annealing process of solids and combinatorial optimization problems and is proved to have the potential to find or approximate the global or near global optimal in a combinatorial optimization problem. The basic idea of SA is to incorporate some randomness to the assignments of cluster labels to pixels in the clustering procedure. *We hypothesize in this study that SA-based systems can reduce the likelihood of converging on a local minimum and thus improve the final classification accuracy.* In this research, two automated SA-based classification systems were developed and applied to remotely sensed Landsat Thematic Mapper (TM) data. The first one, called the single SA-based (S-SA) system, was developed based on the standard SA algorithm. To reduce the computational intensity, the second one, called the integrated SA-based (I-SA) system, was designed to combine the standard SA algorithm and *K*-means into a two-level classification system. The I-SA system was expected to produce the desired classification more efficiently than the S-SA system. The two SA-based

systems and the *K*-means were used to classify two multispectral Landsat TM images with different ground condition complexities. The classification performance of each algorithm was assessed first by a mathematical distance index and then by error matrices to evaluate a set of the randomly sampled pixels on the basis of reference data. Other related important issues, such as the effects of the complexity of the application and the controlling parameters in SA on the resulting classifications, were also discussed based on the experimental results and analysis. Our Kappa statistical analysis on the resulting error matrices demonstrated that the SA-based systems had significantly improved the classification accuracy over that of the *K*-means algorithm in this study when appropriate parameters were chosen. The knowledge and insights on SA learned from our study can facilitate the incorporation of SA random search procedures into other classification approaches that suffer from the same local minimum problem.

3.1 Introduction

Remotely sensed data of the Earth have been utilized to extract useful thematic information about land. Land information derived from remote sensing is useful for modeling and monitoring of numerous natural and environmental processes. In particular, satellite imagery have been proven to be suited for the automated generations of land-cover inventories, and many classification methods using such remotely sensed data have been applied and investigated (Foody 1995; Sergi *et al.* 1995; Serpico & Roli 1995; Bischof *et al.* 1992; Benediktsson *et al.* 1990; Belward *et al.* 1990; Schowengerdt 1983). Basically, these classification methods fit into two categories: supervised and

unsupervised. In a supervised classification, *a priori* knowledge about each land cover type, such as urban, forest, and agriculture, is used to support collecting training samples for each class. To generate an accurate classification, training samples must be reliable and of a sufficient number to estimate the characteristics of classes with adequate accuracy. However, in many remote sensing applications, training sample collection is problematic due to insufficient or unreliable information available and complicated or heterogeneous landscape condition. Comparatively, unsupervised classification requires only a minimal amount of information and therefore becomes more attractive especially in the case that reliable information and adequate knowledge about the study area are not available. Thomson *et al.* (1998) point out that unsupervised classification has operational advantages over supervised methods in terms of reduced interaction time by the producer; however, it offers less direct control over resulting classes.

As defined by Jain and Dubes (1988), unsupervised classification is known as "clustering" which is a generic labeling procedure designed to find natural groupings, or clusters, in multidimensional data based on measured or perceived similarities among patterns. The simplest and most intuitive clustering is based on the concept of similarity: patterns that are similar should be assigned to the same class. The similarity between pattern vectors, which we also consider as pixels in Euclidean space, could be established by determining their proximity to a selected cluster center vector in terms of a distance measure such as the Euclidean (Tou & Gonzales 1977). A desired classification or clustering could be obtained by iteratively minimizing the total Euclidean distance between pattern vectors and their associated cluster center vector. Once clustering is

completed, it is the analysts' responsibility to relate the resulting clusters to the land-cover classes on the ground (Jensen 1996).

Many clustering algorithms have been used for remotely sensed classification applications (Thomson *et al.* 1998; Guneriussen *et al.* 1997; Remund *et al.* 1997; Belward *et al.* 1990). Generally, clustering approaches classify input data by means of minimizing an error function via an iterative process to estimate the cluster characteristics. *K*-means is one of the most popular clustering algorithms applied in remotely sensed applications (Jain & Dubes 1988; Selim & Ismail 1984; Tou & Gonzales 1977; Duda & Hart 1973; Fukunga 1972). Unfortunately, Laarhoven (1988) points out that this kind of iterative clustering algorithms suffers from a serious problem: the solution is strongly dependent on the starting point and convexity of the distance function in clustering problems. If the distance function is not convex, a search starting from a randomly selected point may be trapped by undesirable local minima and prevented from attaining the global or near-global minimum. Because of the potential limitation of local minima, one does not know whether the resulting classification from *K*-means is close to the global minimum or not, or whether there are other better classification solutions to this particular classification problem (Hegarat-Mascle *et al.* 1996; Klein & Dubes 1989). Since many remote sensing classification applications are complicated and their distance functions may not be convex, the local minima problem often becomes more problematic. This motivates us to find an unsupervised approach that could overcome the local minima problem and closely approximate the global or near-global optimal solution.

Simulated Annealing (SA) was developed on the basis of an analogy between the physical annealing process of solids and the large combinatorial optimization problems (Kirkpatrick *et al.* 1983; Cerny 1985). Geman and Geman (1984) have proved that SA has the potential to find or approximate the global or near-global optimal in a combinatorial optimization problem. However, SA often requires much greater computational time. The basic idea of SA is to incorporate some randomness to the assignments of cluster labels to pixels in the clustering procedure, thus reducing the likelihood of getting trapped in a local minimum. A SA-based classification system may be able to reduce the limitation of local minima so as to improve the classification accuracy for land cover classification. However, to date, SA has not been thoroughly investigated and rarely applied in remote sensing applications due to the fact that SA is computationally prohibitive and therefore not practical to deal with the large data sets typical in remote sensed applications. The rapidly growing and available computing power in recent years, while enabling faster processing of huge data sets, has also facilitated the use of more elaborate and diverse methods for data classification. SA is one of the most interesting of these methods.

Our overall objective is to explore the development of algorithms and applications to use SA in unsupervised classifiers for improvement of land cover classification. The specific objectives of this paper are:

(1) To investigate applicability of SA for land cover classification using remotely sensed Landsat Thematic Mapper (TM) multispectral data;

(2) To study the effects of controlling parameters in SA on classification accuracy and efficiency and;

(3) To evaluate and compare the resultant classification results from the SA-based algorithms versus the classical unsupervised *K*-means algorithm.

This paper consists of five sections. In Section II, the clustering approaches SA and *K*-means are explained, and detailed procedures for adapting SA for an unsupervised classification are also described. Experimental results from two Land-Use/Land-Cover (LU/LC) classification applications are presented and analyzed in section III. In section IV, the general discussion and conclusions are presented. Finally, in section V, recommendations for future research are discussed.

3.2 Simulated Annealing Adapted for Clustering Problems

3.2.1 Clustering by minimizing the distance function

In most clustering problems, we begin with *N* patterns, but wish to partition these patterns into *K* clusters. Algorithms that minimize the sum of squared Euclidean distance between patterns and cluster centers have been shown to have many theoretical and practical advantages (Tou & Gonzales 1977). This minimization can be expressed as the generalized least-squared error function *J(V)*:

$$J(V) = \sum_{j=1}^{N} d_{ij}^2 \qquad \text{--- (3.1)}$$

Where

$Y = \{y_j / j \in [1, N], y_j \in R^n\}$ *is the data set,*

N is the number of patterns within the data set,

$V = \{v_i / i \in [1,k], v_i \in R^n\}$ *is the set of cluster centers,*

k is the number of clusters; $2 \le k < N$,

N_i *denotes the number of patterns assigned to the* i^{th} *cluster center* v_i, *which is defined as follows:*

$$v_i = \frac{\sum_{j=1}^{N_i} y_j}{N_i} \qquad \text{--- (3.2)}$$

d_{ij}^2 *is the associated distance function (*$d_{ij}^2 = \| y_j - v_i \|^2$*) between a pattern* y_j *and a given cluster center* v_i *where* $\| \ \ \|$ *is a Euclidean norm on* R^n;

The aforementioned minimization is a hard *k*-partition of *Y*: any pattern y_j among *Y* must be assigned to a cluster and can only belong to this cluster. The objective of the squared distance clustering method is to find a partition containing *K* clusters that minimizes $J(V)$ for a fixed *K*. In the real implementations, to realize the objective, we minimize d_{ij}^2 for each pattern y_j.

Based on the minimization objective described above, the *K*-means algorithm is shown in Figure 3.1 (Tou & Gonzales 1977).

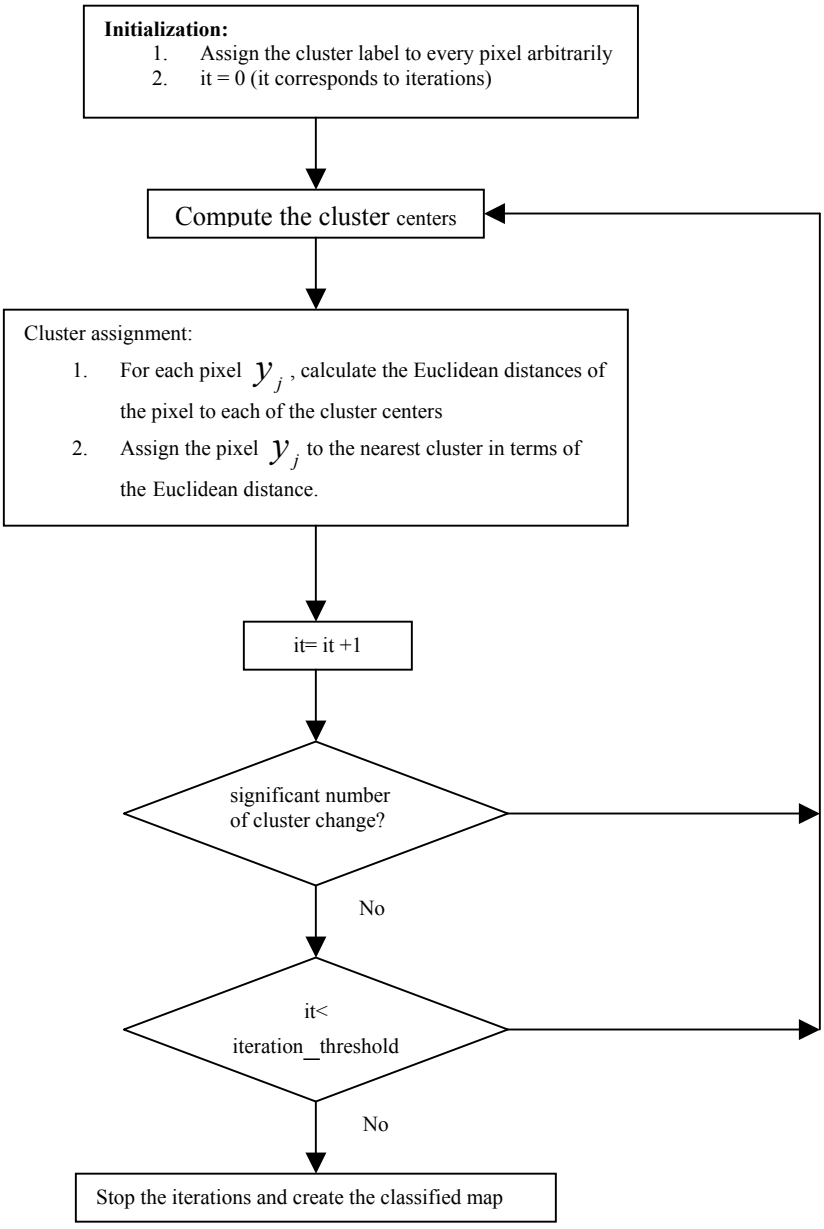


Figure 3. 1 The flowchart of K-means clustering algorithm

The *K*-means clustering algorithm described above is computationally efficient and can produce good results if the clusters are compact, well-separated in the feature space, and hyperspherical in shape when *Euclidean* distance is used (Jain 2000). However, these conditions may not be satisfied in many complicated remote sensing applications. A number of previous studies have shown that the classical *K*-means algorithm failed to produce satisfactory classification due to the limitation of the local minimum (Kolonko & Tran 1997; Hegarat-Mascle *et al.* 1996; Selim & Alsultan 1991; Klein & Dubes 1989).

3.2.2 Simulated Annealing: basic principle

SA was first described by Metropolis *et al.* in 1953 and then introduced by Kirkpatrick *et al.* (1983) and independently Cerny (1985) into combinatorial optimization problems. The SA approach originates from the analogy between the physical annealing process of solid and the large combinatorial optimization problems. In the physical annealing process, if the temperature decreases sufficiently slowly, the ground state, i.e., the particles of the solid are arranged in a highly structured lattice and the energy state of the solid is minimal, can be found. The physical annealing process is simulated in a combinatorial optimization problem to search for a globally optimal configuration so as to minimize a predefined cost function. Clustering, or unsupervised classification, is one type of combinatorial optimization problem. SA has been proven applicable in many such classification applications (Brown & Huntley 1992; Selim & Alsultan 1991; McErlean *et al.* 1990; Klein & Dubes 1989). However, there is much research still to be done in order to use SA for LU/LC mapping purpose using remotely sensed data. Many clustering

approaches have been used to perform LU/LC classifications using remotely sensed data, but were found to suffer from the local minimum minimization (Kolonko & Tran 1997; Hegarat-Mascle *et al.* 1996; Klein & Dubes 1989). Since SA has been shown to be able to overcome the local minimum problem (Geman and Geman 1984), it is worthwhile to develop and investigate SA-based classification systems for LU/LC classification applications using remotely sensed data.

Mathematically, SA can be modeled by means of a Markov chain: a sequence of trials, for which the outcome of each trial depends only on the outcome of the previous one, i.e., the current configuration (Feller 1950). The basic procedure involves a cooling schedule, in which a parameter called temperature T starts out sufficiently high and is gradually lowered in a given schedule to minimize an energy or cost function associated with a specific problem formulation. At each temperature T, a small, randomly generated, perturbation is repetitively applied many times until the system reaches thermal equilibrium at that temperature. Then the algorithm moves to the next temperature in the given schedule. The rule of accepting a perturbation is based upon the *Metropolis criterion*: 1) If ΔE, the energy difference before and after the perturbation, is less than 0, then the perturbation is accepted with probability 1 and the process is continued with the perturbed state; 2) otherwise, the perturbation is accepted with the probability $\exp(-\Delta E / T)$. As a result of a properly designed cooling schedule, the algorithm eventually evolves into a stable state with minimal energy based on the *Boltzmann distribution* (Laarhoven 1988; Geman & Geman 1984).

3.2.3 Adaptation of Simulated Annealing for Clustering Issues

To adapt SA to clustering problems, the energy function is defined by Equation 3.1. We start with a high controlling temperature denoted as T_0 and an arbitrary initial clustering assignment. For a given randomly selected pattern y_j, we reassign it from its previous cluster *m* to another randomly selected cluster *n* based on a uniform probability. The squared Euclidean distances d_{mj}^2 and d_{nj}^2 of the selected pattern y_j from the cluster centers v_m and v_n, and Δd_{mn}, the distance change between d_{mj}^2 and d_{nj}^2, are respectively computed. We accept the new assignment either with a distance decrease Δd_{mn}<0, or with a distance increase according to a positive probability $\exp(-\Delta d_{mn}/T)$ where *T* is the temperature at this state. Each trial of generating and accepting a reassignment of a randomly selected pattern is considered as a "transition". The adapted SA algorithm is stated in Figure 3.2. This described SA algorithm will be denoted as the single SA-based (S-SA) algorithm in the following sections.

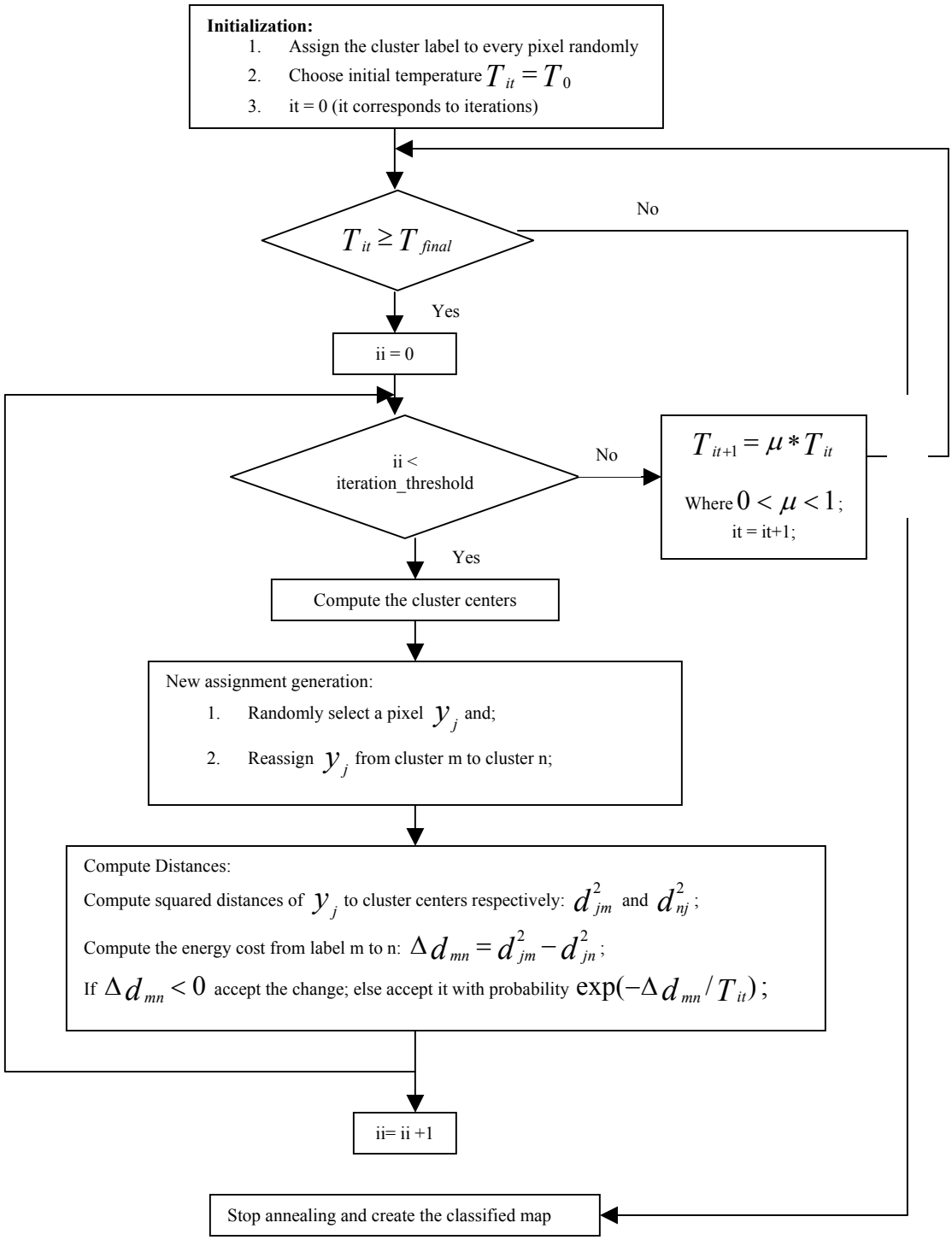


Figure 3. 2 The adapted single SA-based clustering algorithm

3.2.4 Definition of Parameters in SA

The SA algorithm is quite simple to implement. However, the quality of the final solution depends on the convergence of the algorithm, which is governed by a set of parameters, i.e., the cooling schedule. Geman & Geman (1984) and Aarts & Laarhoven (1987) had proved the convergence of the SA algorithm based on the theory of both heterogeneous and homogenous Markov chains and discussed the conditions under which asymptotic convergence is guaranteed. The cooling schedule works as a Boltzmann machine if the temperature decreases sufficiently slowly. As the temperature decreases, the Boltzmann machine will have a converging tendency toward the low-energy states and finally, when the temperature approaches zero, only the minimum-energy states have a non-zero probability of occurrence. A finite-time implementation of the SA algorithm could obtain the approximation of an optimal solution of the combinatorial optimization problem by specifying a set of appropriate parameters that control the convergence of the algorithm (Laarhoven 1988). These parameters are combined as a cooling schedule: 1) the initial value of the temperature, 2) the decrement function for decreasing the value of the controlling parameter, 3) the final value of the temperature, and 4) a way to guarantee a finite number of accepted transitions at each temperature. These parameters are chosen so as to 'imitate the asymptotic behavior of the homogeneous algorithm in polynomial time, thereby removing any guarantees with respect to the optimality of the configuration returned by the algorithm' (Laarhoven 1988). The choice of cooling schedule is a very difficult problem for which there is no "best" choice for all problems. The search for an adequate cooling schedule has been one of the main topics in many papers published over

the years (Collins *et al.* 1987; Aarts & Laarhoven 1987; Laarhoven 1988; Aarts & Korst 1989). According to these previous studies, these parameters were specified and chosen for their relative simplicity and fast computation:

- *The initial value of the temperature* T_0. The basic rule is that the selected initial value should be sufficiently large such that virtually all transitions can be accepted. Several researchers used the average difference in cost of subsequent configurations occurring in a Markov chain (Leong 1985; White 1984). Some smaller candidates of T_0 might be applicable in simple clustering applications.
- *The decrement factor* μ *for decreasing the controlling parameter* T_n *where n is the iteration number*. To approximate the global optimal solution, the controlling temperature T_n has to decrease sufficiently slowly. Convergence in probability to the global minimum has only been proven (Geman & Geman 1984) for the logarithmic annealing schedule: $T_n = \frac{1}{\ln n}$. This inverse-log schedule, though guarantees the global convergence for a wide class of combinatorial problems, is far too slow for almost any practical implementation. The simple geometric function of temperature: $T_{n+1} = \mu * T_n$ is used where μ, the decrement factor, is a small constant typically between 0.80 and 0.99. This decreasing function has been proposed in many SA research works and proven to be efficient and applicable (Hegarat-Mascle & Olivier 1996; Aarts & Korst 1989; Laarhoven 1988; Lundy & Mees 1986; Bonomi & Lutton 1984; Burkard & Rendl 1984).

- *The final value of the temperature* T_{final}. It is used to determine when to stop the annealing process. In our study, we fixed T_{final} as a small number close to zero. Via a proper cooling schedule, while T_n is approaching T_{final}, a configuration with minimal energy might be targeted.
- *A finite number of accepted transitions at each temperature* T_n. This factor actually determines whether the thermal equilibrium at each temperature, an important prerequisite for the globally optimum approximation, can be restored. To reduce the computational complexity in this study, this factor is specified by two parameters. The first parameter is the iterations, or the times of image scan at each temperature. The second parameter is the generation probability, which is used to randomly select a number of pixels per image. For each pixel at each image scan, a random probability is drawn from a uniform probability distribution. If the probability is larger than the predefined generation probability, this pixel will be operated and reassigned to another randomly selected cluster; otherwise the pixel will be passed. The acceptance of the new reassignment depends on the evaluation of the incurred energy change before and after this reassignment. The iterations at each temperature and generation probability altogether determine the length of the homogeneous Markov chain at each T_n.

3.2.5 Integration of the SA and *K*-means for Clustering Refinement

As we mentioned above, SA is simple but computationally demanding. This is one of the main reasons preventing the wide applicability of SA in remote sensing applications dealing with large data sets. It is therefore imperative to investigate a way to implement SA more efficiently. In clustering problems, intuitively we can believe that if the initial clustering is as close to the optimal clustering as possible, then searching for the global optimum will be faster and more reliable. We hypothesize that if the SA starts with an initialization close to the final optimal solution, a small initial temperature value and a small number of perturbed pixels might be sufficient conditions to find the global or near-global optima with a much quicker cooling schedule. Although *K*-means may not be able to find the global optimum for a specified problem, the resulting clustering solution from it might be more likely to be closer to the global optimum than a random initialization. Based on the assumptions above, we proposed an integrated algorithm of SA and *K*-means called the integrated SA-based (I-SA) system. In this I-SA system, the best result from *K*-means was used to initialize the cluster centers in SA and then the SA algorithm was implemented using a faster cooling schedule, by which an improved clustering solution might be found at a shorter time. To some degree, the I-SA system keeps the random search capability from the SA to reduce the chance to be trapped by undesirable local minima.

3.3 Applications to Land-cover/Land-use Classification

3.3.1 Description of data sets and classification scheme

In our study, three unsupervised classification systems were developed based on the *K*-means algorithm, the single SA algorithm, and the I-SA algorithm. These systems were applied to classify two remotely sensed Landsat TM images. Landsat TM data have been successfully utilized for LU/LC mapping applications (Dai & Khorram 1999; Schistad Solberg *et al.* 1994; Frank 1988; Khorram *et al.* 1987; Haack *et al.* 1987; Strahler *et al.* 1978). The TM image shown in Figure 3.3 was taken in summer 1994 and covers an area of 2,856.15 *ha.* The study area is located near the Raleigh/Durhum International Airport in Raleigh, North Carolina. The TM image shown in Figure 3.4 was collected in winter 1991. The study area is about 7,831.89 *ha* and is located in the middle of Georgia near Macon. For reference convenience, in the following sections, the first TM image will be denoted as TM-1 while the second one as TM-2.

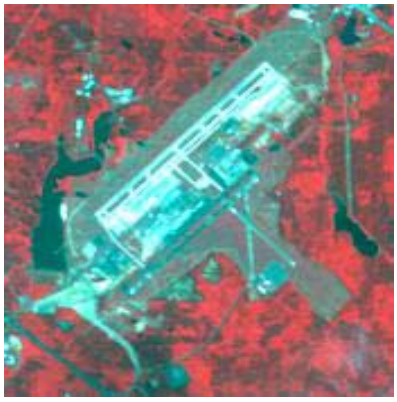


Figure 3. 3 Original Landsat TM-1

Figure 3. 4 Original Landsat TM-2

Both of the TM images were analyzed visually assisted by aerial photos and available *a priori* knowledge of the study areas. Using the reference information from each specific study area, the classification scheme for each application was initially developed on the basis of the standard land use/land cover classification scheme proposed by Anderson *et al.* (1976). As a further step, classification schemes were adapted depending on the image complexity and separation capability so that the spectrally derived clusters could be easily translated into real land cover classes. From the standard false color composite TM-1 image, we observed that the landscape condition was relatively simple. It was easy to determine that five land cover classes dominate this area, which are urban, mixed forest, evergreen forest, Grassland/Agriculture, water. By studying the TM-2 image and referring to ancillary data, we found that this study area was much more heterogeneous than TM-1 and more land cover classes in this study area.

Seven classes were eventually derived including urban/residential, evergreen forest, deciduous forest, mixed forest, bare soil or fallow crop field, transitional area or grassland, water. To reduce the computational complexity, only three bands were used for classification for each application. Jensen (1996) indicated that either the combination of band2 (Optical Green Band), band3 (Optical Red Band), and band4 (Reflected Near-infrared Band) or the combination of band3, band4, and band5 (Reflected Middle-infrared Band) are the optimal three-band combinations for land cover classification. In TM-1, band2, band3, and band4 were used to classify the image while band3, band4, and band5 were used in TM-2.

The classification applications using TM-1 and TM-2 were to test the wide applicability and potential of the SA-based systems and their reliability in applications with different classification complexities. In this study, all the experiments were run on the Origin 2000 produced by Silicon Graphics, Inc (SGI), which has 32 processors and 16 Gigabytes of memory. To evaluate the classification performances, the resulting classification results were first evaluated by its mathematical performance index $J(V)$. Based on the aforementioned clustering formulation, our minimization objective is to minimize the sum of squared Euclidean distance $J(V)$. The resulting classified map with lowest $J(V)$ is believed to be more accurate than others. To compare the overall performance of tested clustering algorithms in this paper, the classification result with the lowest $J(V)$ was selected from each of the tested algorithms and further evaluated for its

true accuracy by comparing with reference data manually derived from aerial photos, the original TM images, and other ancillary data.

3.3.2 Study of the selection and roles of parameters in SA

To study the role of the aforementioned parameters in remote sensing applications, a large number of experiments using various parameter candidates were conducted by applying the S-SA system to image TM-1. The following candidates for each parameter were used:

Initial values of temperature T_0: 10, 100, 1000, 10,000

Decrement factors μ: 0.80, 0.85, 0.90, and 0.99

Iterations at each temperature (IET): 10, 20, 30, and 40

Generation probability (GP): 0.85, 0.90, and 0.95

To visualize the effects of these parameters, several groups of experimental data for each tested parameter were selected to compare the resulting $J(V)$ and the required computational time. Each group of parameter combinations was ran with four random initializations. In the following figures, $J(V)$ and computational time for each group are the averages of these four instances. Within each selected data group, only the studied parameter is allowed to change among the selected range while all other parameters are fixed. For example, in Figure 3.5, each data group represents a randomly selected data series in which T_0 ranges from 10 to 10,000 while the other three parameters were fixed. The selected groups of data are graphed on the studied parameter vs. $J(V)$ in Figure 3.5

–3.8 while on the studied parameter vs. computational time in seconds in Figure 3.9 – 3.12.

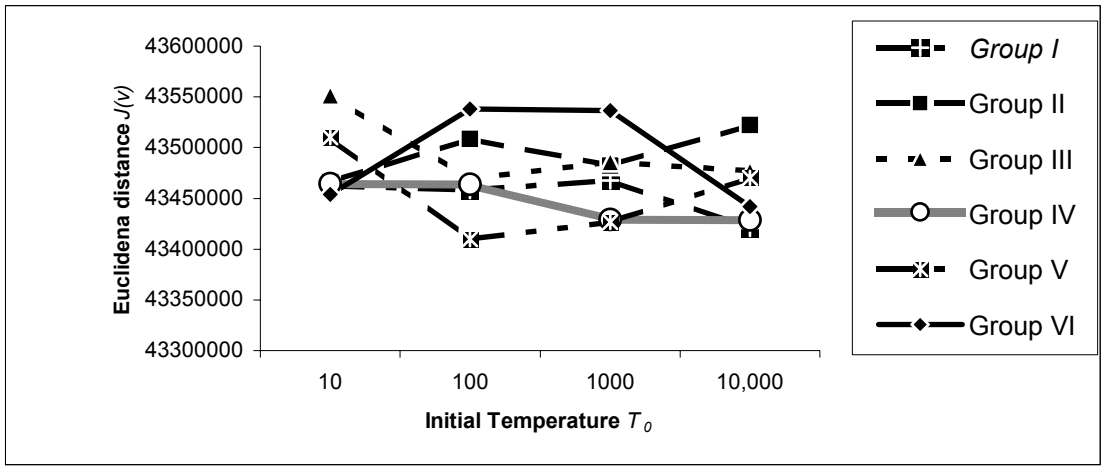


Figure 3. 5 Effect of T_0 on the magnitude of $J(V)$

Group I: μ = 0.99, IET = 20, GP = 0.90

Group II: μ = 0.90, IET = 20, GP = 0.90

Group III: μ = 0.85, IET = 20, GP = 0.90

Group IV: μ = 0.80, IET = 20, GP = 0.90

Group V: μ = 0.90, IET = 20, GP = 0.85

Group VI: μ = 0.90, IET = 40, GP = 0.90

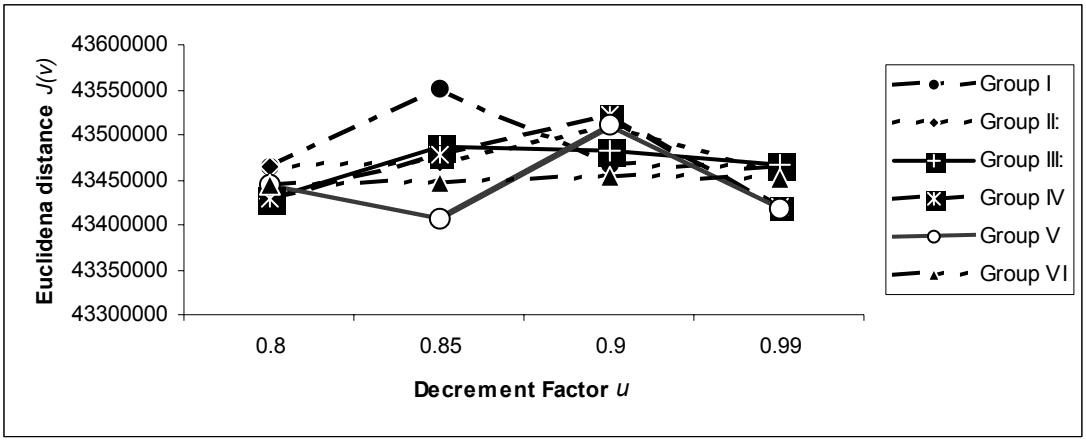


Figure 3. 6 Effect of μ on the magnitude of $J(V)$

Group I: T_0 = 10, IET = 20, GP = 0.90

Group II: T_0 = 100, IET = 20, GP = 0.90

Group III: T_0 = 1000, IET = 20, GP = 0.90

Group IV: T_0 = 10,000, IET = 20, GP = 0.90

Group V: T_0 = 10, IET = 20, GP = 0.85

Group VI: T_0 = 10, IET = 40, GP = 0.90

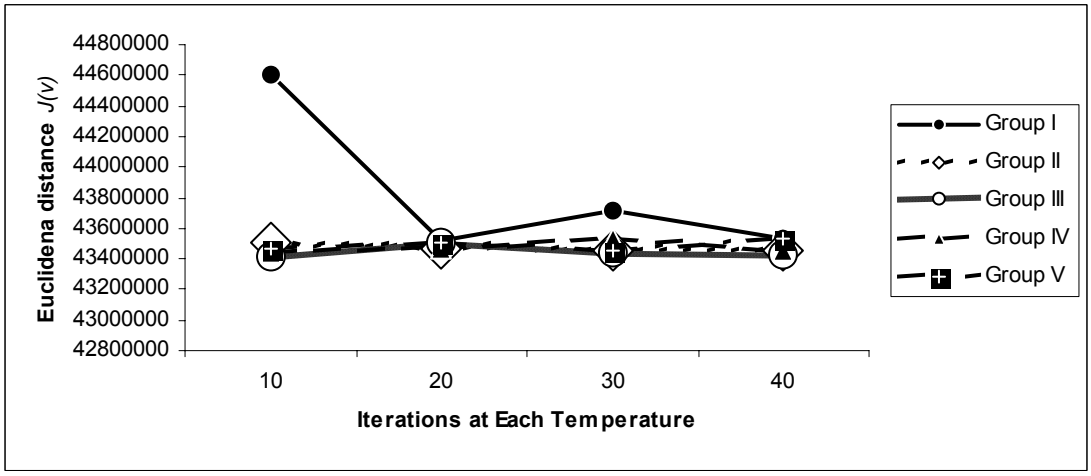


Figure 3. 7 Effect of the IET on magnitude of $J(V)$

Group I: T_0 = 10, μ = 0.90, GP = 0.95

Group II: T_0 = 10, μ = 0.90, GP = 0.90

Group III: T_0 = 10, μ = 0.90, GP = 0.85

Group IV: T_0 = 10, μ = 0.99, GP = 0.90

Group V: T_0 = 100, μ = 0.90, GP = 0.90

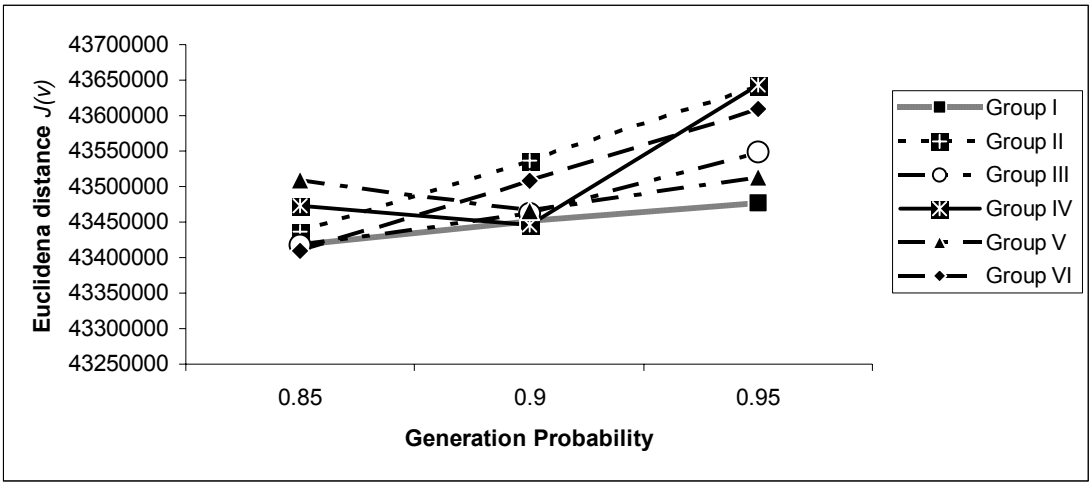


Figure 3. 8 Effect of the GP on magnitude of $J(V)$

Group I: T_0 = 10, μ = 0.99, IET = 40

Group II: T_0 = 10, μ = 0.99, IET = 30

Group III: T_0 = 10, μ = 0.99, IET = 20

Group IV: T_0 = 10, μ = 0.99, IET = 10

Group V: T_0 = 10, μ = 0.90, IET = 20

Group VI: T_0 = 100, μ = 0.90, IET = 20

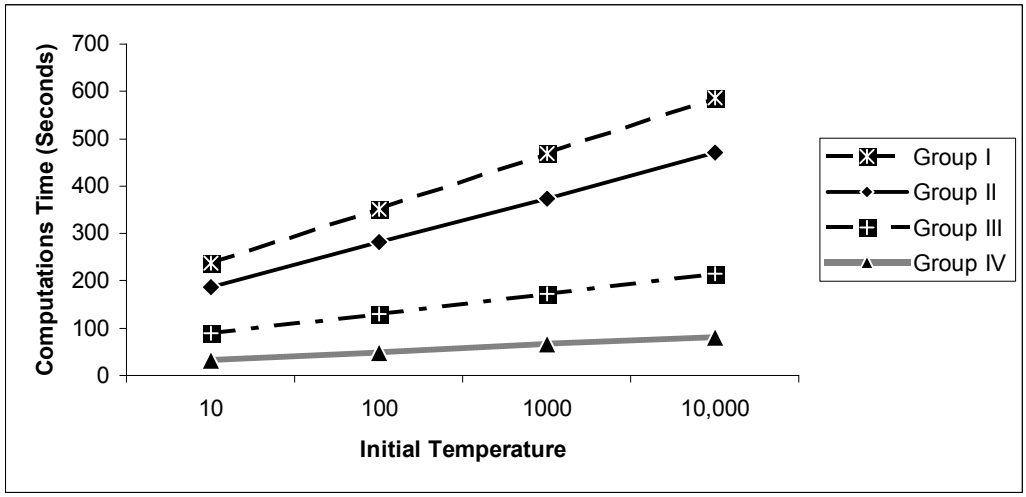


Figure 3. 9 Effect of T_0 on the Computational Time

Group I: μ = 0.90, IET = 40, GP = 0.95

Group II: μ = 0.90, IET = 30, GP = 0.90

Group III: μ = 0.85, IET = 20, GP = 0.85

Group IV: μ = 0.80, IET = 10, GP = 0.85

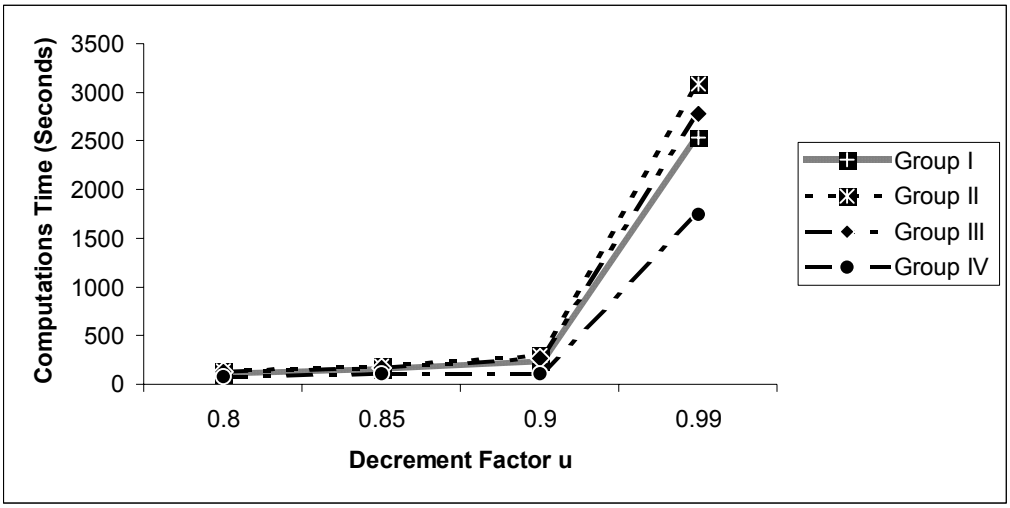


Figure 3. 10 Effect of μ on the Computational Time

Group I: T_0 = 10, IET = 40, GP = 0.95

Group II: T_0 = 100, IET = 30, GP = 0.85

Group III: T_0 = 1000, IET = 20, GP = 0.85

Group IV: T_0 = 10,000, IET = 10, GP = 0.90

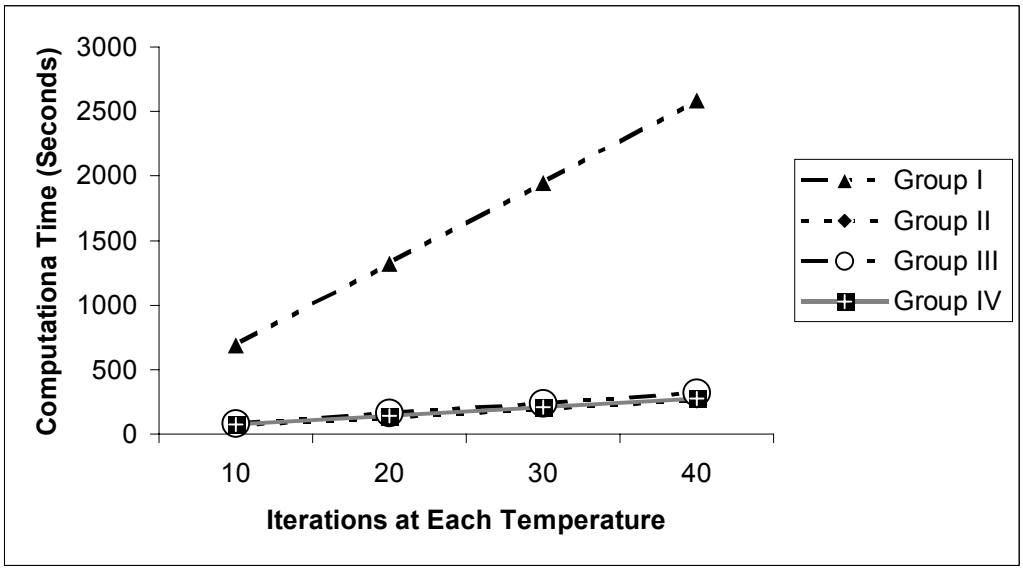


Figure 3. 11 Effect of the IET on the Computational Time

Group I: T_0 = 10, μ = 0.99, GP = 0.90

Group II: T_0 = 100, μ = 0.85, GP = 0.85

Group III: T_0 = 1000, μ = 0.85, GP = 0.90

Group IV: T_0 = 10,000, μ = 0.80, GP = 0.95

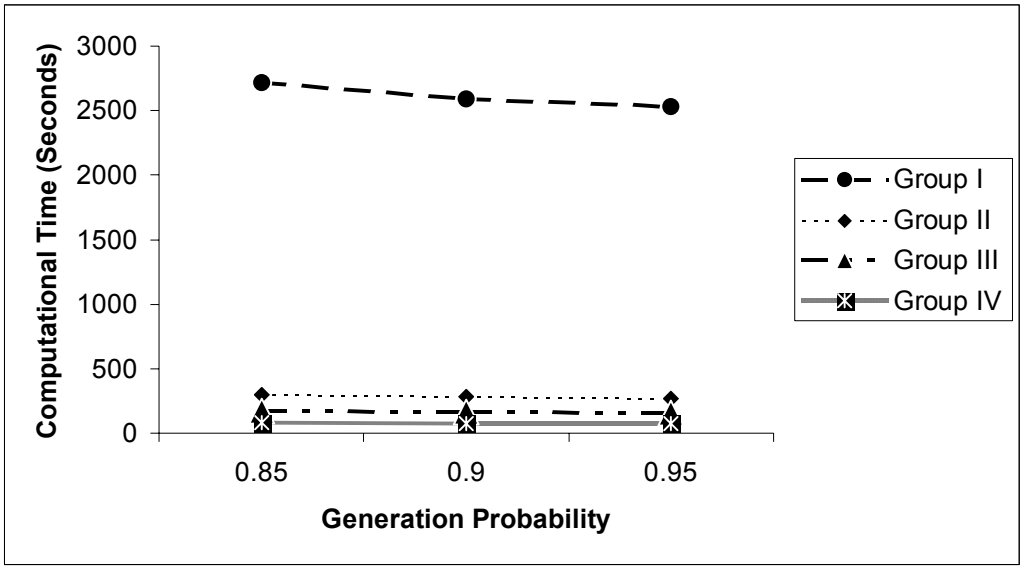


Figure 3. 12 Effect of the GP on the Computational Time
Group I: T_0 **= 10,** μ **= 0.99, IET = 40**
Group II: T_0 **= 100,** μ **= 0.90, IET = 30**
Group III: T_0 **= 1000,** μ **= 0.85, IET = 20**
Group IV: T_0 **= 10,000,** μ **= 0.80, IET = 10**

1) Effect of the initial value of temperature T_0 and the decrement factor μ

In this study, the largest T_0 candidate (10,000) was chosen to guarantee all transitions to be accepted at the beginning of clustering while the lowest T_0 candidate (10) can only allow a small number of transitions to be accepted. This decrement factor μ is used to decrease the temperature in the cooling schedule by multiplying it with the current temperature. The SA implementation requires that temperature decreases sufficiently slowly but there is no criterion to specify how slow the decrease needs to be. Although Laarhoven (1988) and Aarts & Korst (1989) suggest that T_0 should be large enough that virtually all transitions can be accepted, there is no clear evidence from

Figure 3.5 to show that the magnitudes of T_0 make a significant difference in this clustering application in terms of reducing $J(V)$. Lower initial values like 10, 100 may also be applicable in this application. This might due to all these selected T_0 candidates being too high for this application. From Figure 3.6, we observe that comparatively the slowest decrement factor μ (0.99) has the most stable performance, but there is only a small performance improvement observed among the decrement factor candidates.

2) Effect of the iterations at each temperature and generation probability

In this study, the IET and the GP is combined together to determine the length of the homogeneous Markov chain at each temperature. A finite number of accepted transitions (i.e., reassignment of class labels) at each temperature are essential to the convergence of SA. The GP is a uniform probability between 0.0 and 1.0. The larger the IET and the smaller the GP, the more transitions operated and, of course, the greater likelihood that more transitions are accepted at each temperature. From the behaviors of the five groups of data in Figure 3.7, it is found that there are no significant differences among the four IET candidates in all groups except the Group I. Figure 3.8 demonstrates the effect of the GP on minimizing $J(V)$. For each group of data, the lowest $J(V)$ s in four groups were obtained with GP (0.85) while in other two groups the lowest $J(V)$ s were obtained with GP (0.90).

3) Effect of parameters on Computational Time (CT)

The effects of the parameters on CT are studied as well using the same graph method. From Figure 3.9 to Figure 3.12, we see that the CT increases when T_0, μ and

IET increase, and when GP decreases. There is a compromise between computational time and classification performance. Among the parameters, T_0 and μ are the factors to escalate the CT most significantly.

In this specific application, the magnitude of the initial temperature did not have a significant effect on the clustering performance. The slowest decrement factor candidate (0.99) in our study performed better than other candidates, but not significantly. From the graphical analysis above, we saw that the appropriate selection of the GP was comparatively more important among the four studied parameters, which verified that a certain number of accepted transitions are one of the essential conditions for successful SA implementations. However, the search for appropriate parameters is specific for each individual application and is certainly related to the structure and cluster numbers of the classified data. Based on our parameter analytical results, we conclude that the choice of a single parameter is not vital for a successful implementation. The optimal and efficient implementation depends on an optimal parameter combination for these four parameters.

3.3.3 Classification results and analysis

In this section we present the classification results of the three clustering algorithms using two TM images: TM-1 and TM-2. To obtain the best relative performance for each of the algorithms, a number of experiments were performed using either different initializations in *K*-means or different parameter combinations in the two SA-based algorithms. Then the classification result with the lowest $J(V)$ from each algorithm was selected for further accuracy assessment.

Figure 3.13 shows the classified maps from TM-1. For each of the selected classified maps, Table 3.1 gives the SA-related parameters and their performance in terms of $J(V)$ and computational time. The I-SA algorithm obtained the best classification performance in TM-1. Both SA-based algorithms performed better than the *K*-means in this application. However, *K*-means was the most efficient algorithm, taking only a few seconds to run. To validate the three classified maps in terms of land-cover/land-use classification accuracy, 253 sample points were randomly selected and manually interpreted based on the reference data derived from false composite TM-1 imagery and aerial photos. The resulting error matrices are presented in Table 3.2-3.4. Among the three classified maps, the highest overall accuracy (91.3%) was obtained by the I-SA algorithm. The Kappa analysis was thus used to statistically determine if one of the derived error matrices is significantly different than another (Bishop *et al.* 1975). The KHAT values shown in Table 3.5 demonstrate that there are strong agreements between the classification results and the reference data for all these three error matrices. From the *Z*-tests shown in Table 3.6, it is indicated that at the 90% confidence level, the I-SA algorithm is significantly better than the *K*-means, while there is no significant difference between S-SA and the other two.

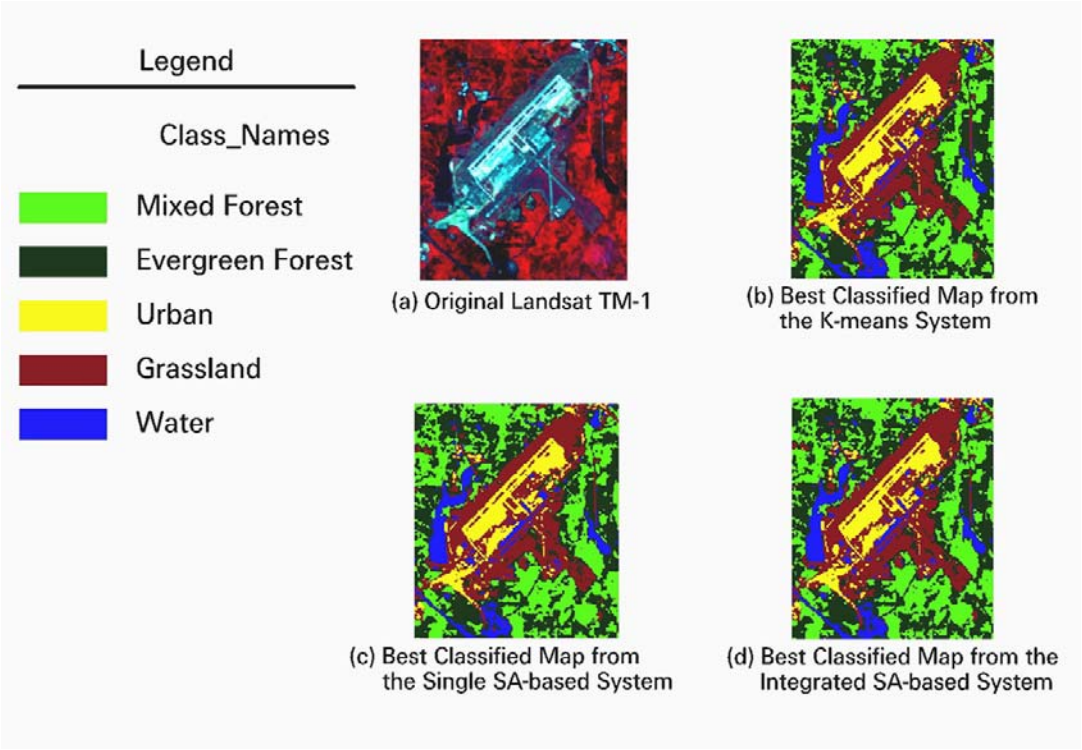


Figure 3. 13 The original image TM-1 and the classified maps

Table 3. 1 SA-related Parameters and Performance Indexes from TM-1

	T_0	μ	IET	GP	J(v)	Computational Time (Seconds)
K-means	-	-	-	-	43442871	10
Single SA-based	10	0.99	20	0.85	43383734	1391
Integrated SA-based	5	0.90	30	0.80	43383246	204

Table 3. 2 Error matrix on the classified result of TM-1 from the K-means

		Reference Data						
Classified Image		**Mixed Forest**	**Evergreen Forest**	**Urban**	**Grassland**	**Water**	**Classified Totals**	**Users' Accuracy**
	Mixed Forest	**60**	8				68	88.24%
	Evergreen Forest	11	**60**		5		76	78.95%
	Urban			**22**	1		23	95.65%
	Grassland		2	2	**63**	1	68	92.65%
	Water		1		4	**13**	18	72.22%
	Reference Totals	71	71	24	73	14	**253**	
	Producers' Accuracy	84.51%	84.51%	91.67%	86.30%	92.86%		
		Overall Accuracy: 218/253 = 86.17%						

Table 3. 3 Error matrix on the classified result of TM-1 from the single SA-based algorithm

		Reference Data						
Classified Image		**Mixed Forest**	**Evergreen Forest**	**Urban**	**Grassland**	**Water**	**Classified Totals**	**Users' Accuracy**
	Mixed Forest	**63**	8				71	88.73%
	Evergreen Forest	8	**59**		6		73	80.82%
	Urban			**23**	1		24	95.83%
	Grassland		3	1	**62**		66	93.94%
	Water		1		4	**14**	19	73.68%
	Reference Totals	71	71	24	73	14	**253**	
	Producers' Accuracy	88.73%	83.10%	95.83%	84.93%	100%		
		Overall Accuracy: 221/253 = 87.35%						

Table 3. 4 Error matrix on the classified result of TM-1 from the integrated SA-based Algorithm

		Reference Data						
Classified Image		Mixed Forest	Evergreen Forest	Urban	Grassland	Water	Classified Totals	User's Accuracy
	Mixed Forest	**65**	2		1		68	95.59%
	Evergreen Forest	5	**67**	1	4	1	78	85.90%
	Urban			**23**	1		24	95.83%
	Grassland	1	2		**63**		66	95.45%
	Water				4	**13**	17	76.47%
	Reference Totals	71	71	24	73	14	**253**	
	Producers' Accuracy	91.55%	94.37%	95.83%	86.30%	92.86%		
		Overall Accuracy: 231/253 = 91.30%						

Table 3. 5 Individual Kappa Analysis for the three Error Matrices from TM-1

	K-means	Single SA-Based	Integrated SA-Based
KHAT	0.82	0.83	0.88
Kappa Variance	0.00085	0.00078	0.00056
Z-Value	28.06	29.80	37.42

Table 3. 6 Kappa Analysis Results for the Comparisons of the Error Matrices from TM-1

	K-means	Single SA-Based	Integrated SA-Based
***K*-means**		0.40	1.87
Single SA-Based			1.43
Integrated SA-Based			

The classified maps created from TM-2 are shown in Figure 3.14. Their SA-related parameters and the performance indexes are presented in Table 3.7. Among the

three studied algorithms, the S-SA algorithm obtained the best classification result. To conduct the accuracy assessments, 299 sample points were randomly selected from TM-2 and manually interpreted. The resulting error matrices are presented in Table 3.8-3.10. The highest overall accuracy 75.59% was obtained by the S-SA algorithm. The Kappa analysis shown in Table 3.11-3.12 demonstrates that the agreement between the classification results and the reference data in TM-2 is moderate, and the S-SA algorithm had significantly improved the classification of TM-2 as compared to the *K*-means and the I-SA algorithms. However, unlike with TM-1, the performance of the I-SA algorithm in TM-2 is poorer than the other two.

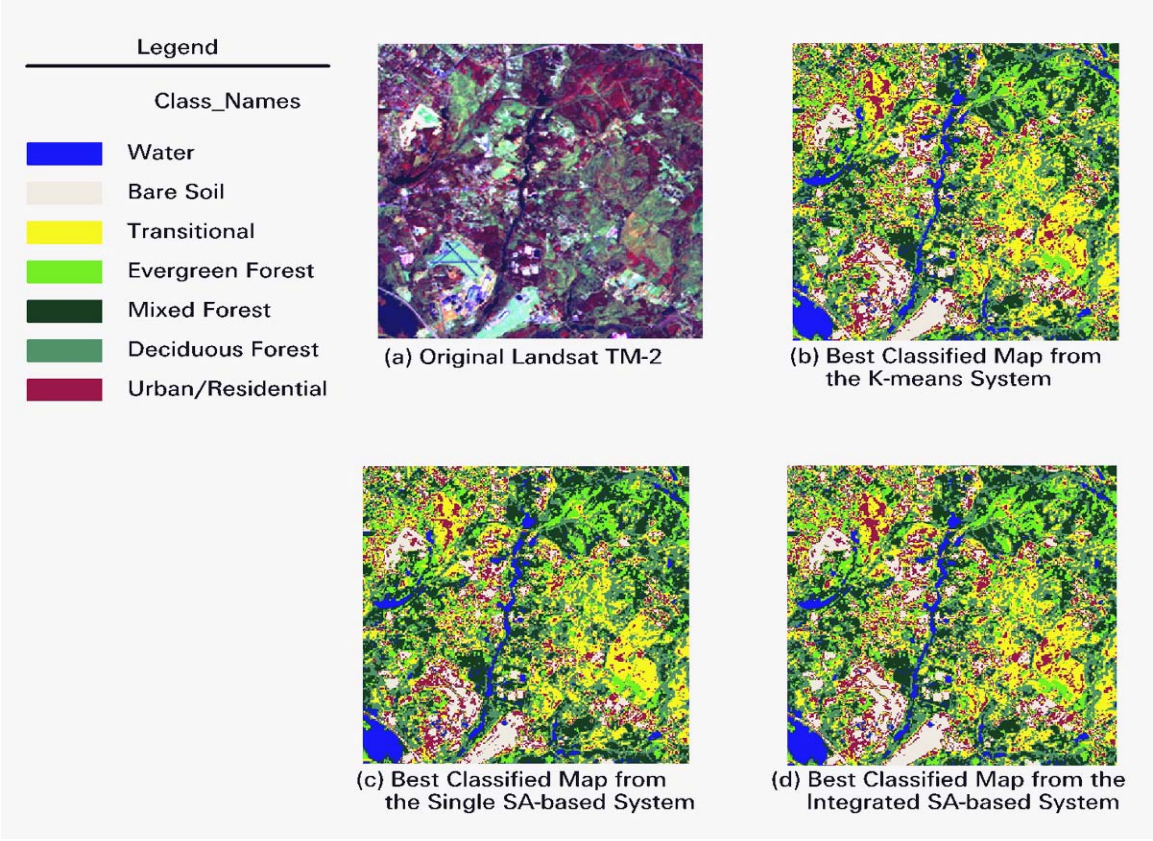


Figure 3. 14 The original image TM-2 and the classified maps

Table 3. 7 SA-related Parameters and best classification results from TM-2

	T_0	μ	IET	GP	*J(v)*	Computational Time (Seconds)
***K*-means**	-	-	-	-	107561539	20
Single SA-based	20	0.80	50	0.90	107218988	510
Integrated SA-based	5	0.75	40	0.65	107566200	978

Table 3. 8 Error matrix on the classified result of TM-2 from the K-means where C1=Urban/Residential, C2 = Evergreen Forest, C3 = Deciduous Forest, C4 = Mixed Forest, C5 = Bare Soil/Crop, C6 = Transitional/Other Grassland, C7 = Water

		Reference Data								
		C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**Classified Totals**	**Users' Accuracy**
Classified Image	**C1**	**18**			4	2	15		39	46.2
	C2		**45**	4	7				56	80.4
	C3	1	5	**34**	7		2		49	60.4
	C4	1	6	1	**24**				32	75.0%
	C5	5		1	1	**24**	1		32	75.0%
	C6	3		20	3	1	**35**		62	56.5
	C7	1		5	1			**22**	29	75.9
	Reference Totals	29	56	65	47	27	53	22	299	
	Producer's Accuracy	62.0%	80.4%	52.3%	51.1%	88.9%	66.0%	100%		
	Overall Classification Accuracy = 202/299 = 67.56%									

Table 3. 9 Error matrix on the classified result of TM-2 from the Single SA-based algorithm where C1 = Urban/Residential, C2 = Evergreen Forest, C3 = Deciduous Forest, C4 = Mixed Forest, C5 = Bare Soil/Crop, C6 = Transitional/Other Grassland, C7 = Water

		Reference Data								
		C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**Classified Totals**	**Users' Accuracy**
Classified Image	**C1**	**20**			2	1	8		31	64.52%
	C2		**50**	11	9	1			71	70.42%
	C3	2		**44**	5		3		54	81.48%
	C4		6		**24**				30	80.00%
	C5	3			2	**24**			29	82.76%
	C6	3		8	4	1	**42**		58	72.41%
	C7	1		2	1			**22**	26	84.62%
	Reference Totals	29	56	65	47	27	53	22	299	
	Producer's Accuracy	68.97%	89.29%	67.69%	51.06%	88.89%	79.25%	100%		
	Overall Classification Accuracy = 226/299 = 75.59%									

Table 3. 10 Error matrix on the classified result of TM-2 from the integrated SA-based algorithm where C1 = Urban/Residential, C2 = Evergreen Forest, C3 = Deciduous Forest, C4 = Mixed Forest, C5 = Bare Soil/Crop, C6 = Transitional/Other Grassland, C7 = Water

		Reference Data								
Classified Image		C1	C2	C3	C4	C5	C6	C7	Classified Totals	Users' Accuracy
	C1	**15**			4	2	15		36	41.67%
	C2		**46**	6	7				59	77.97%
	C3	1	3	**34**	7		2		47	72.34%
	C4	1	7	1	**23**				32	71.88%
	C5	8		1	2	**24**	1		36	66.67%
	C6	3		20	3	1	**35**		62	56.45%
	C7	1		3	1			**22**	27	81.48%
	Reference Totals	29	56	65	47	27	53	22	299	
	Producer's Accuracy	68.97%	89.29%	67.69%	51.06%	88.89%	79.25%	100%		
	Overall Classification Accuracy = 199/299 = 66.56%									

Table 3. 11 Individual Kappa Analysis for the Three Error Matrixes from TM-2

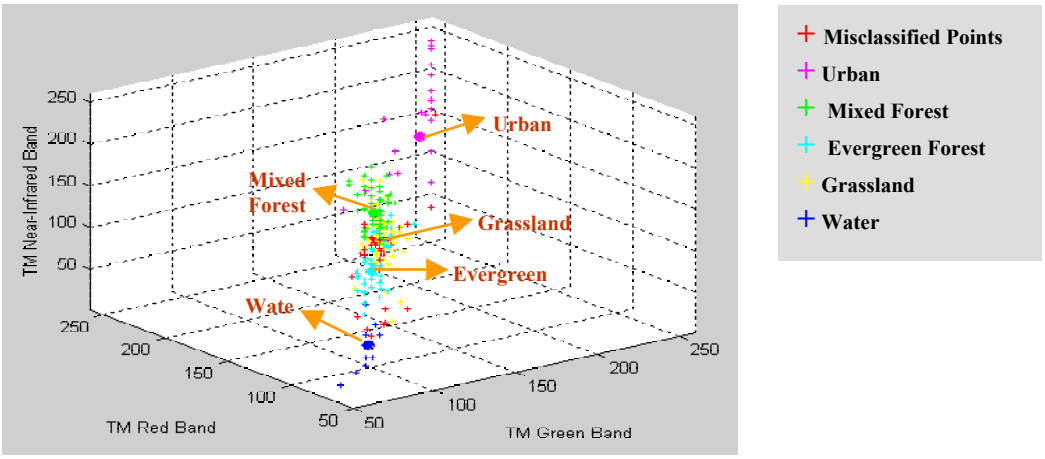
	K-means	Single SA-Based	Integrated SA-Based
KHAT	0.62	0.71	0.60
Kappa Variance	0.0010	0.00088	0.0010
Z-Value	19.21	24.03	18.74

Table 3. 12 Kappa Analysis Results for the Comparisons of the Error Matrixes from TM-2

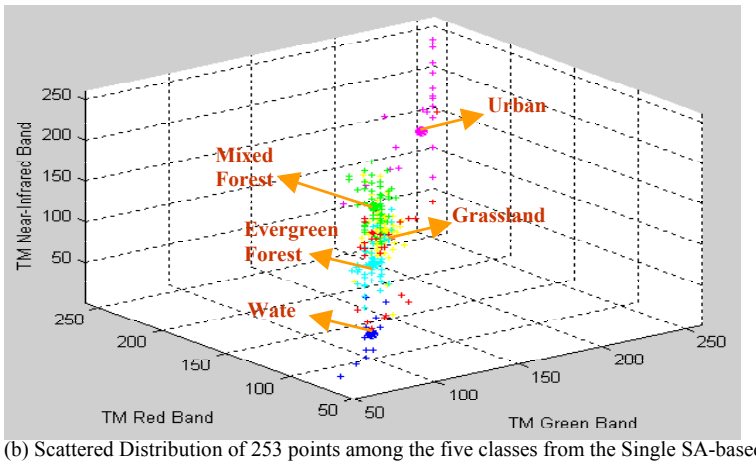
	K-means	Single SA-Based	Integrated SA-Based
***K*-means**		2.13	0.27
Single SA-Based			2.40
Integrated SA-Based			

By analyzing the classification results, we see that the classification performances of the three clustering algorithms using TM-1 and TM-2 are different in several aspects. First, the classification accuracy in TM-1 is relatively high and satisfactory while the classification accuracy in TM-2 is moderate. Second, the I-SA algorithm in TM-1 had the

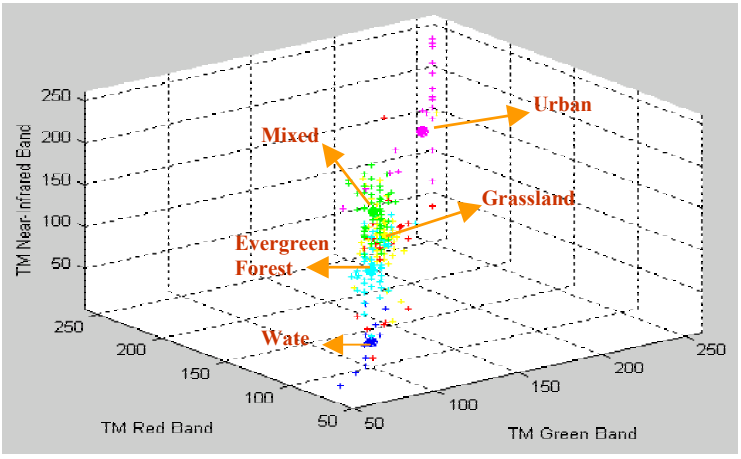
best classification performance while its performance in TM-2 is poor with the S-SA algorithm producing the best classification. Third, the resulting SA-related classification improvement in TM-2 is more significant than that in TM-1. We speculate that these differences may come from the differences of data structures and distributions between TM-1 and TM-2. Visually we have observed that the landscape condition in TM-1 is relatively simple. Five typical land cover classes can be easily differentiated by the naked eye. From the distribution plots shown in Figure 3.15, it is further demonstrated that the five classes in TM-1 are spectrally compact and well-separated. The comparatively homogeneous study area reduces the classification complexity and helps obtain a relatively high classification accuracy in TM-1 classification. The landscape conditions in TM-2 are more complicated than TM-1 because the study area is more heterogeneous and contains more LU/LC classes. The spectral complexity in TM-2 is demonstrated as a more scattered distribution among the seven classes shown in Figure 3.16. The increase of spectral complexity explains why the accuracy of resulting classification of TM-2 is lower than that of TM-1. In a more complicated classification case like TM-2, it often suffers more local minima than a simpler case. This is one of the reasons why the SA-based algorithm has improved the classification accuracy more significantly in TM-2 than TM-1. It is due to the same reason that the I-SA algorithm has failed in TM-2. In a complicated classification application, T_0 and transitions at each temperature must be specified sufficiently large to help the algorithm jump out of the trap of many local minima.



(a) Scattered Distribution of 253 points among the five classes from the *K*-means

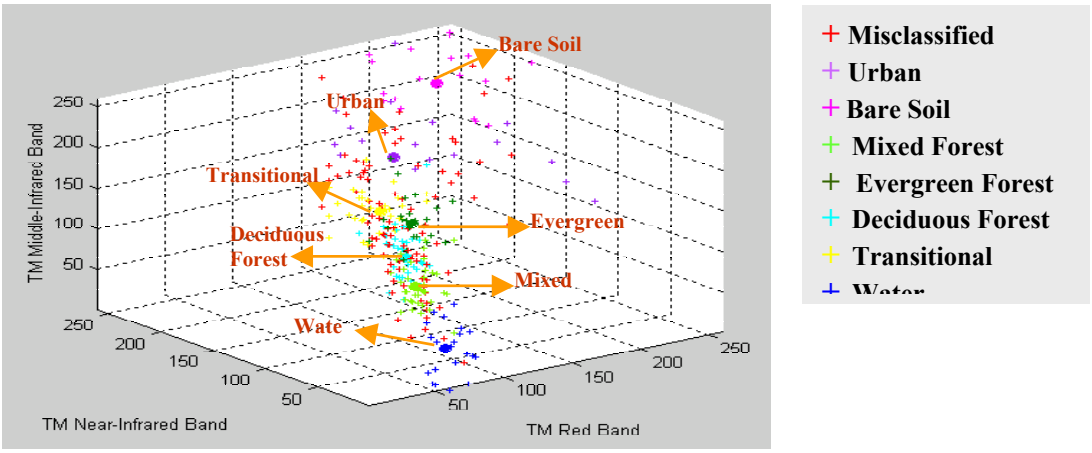


(b) Scattered Distribution of 253 points among the five classes from the Single SA-based Algorithm

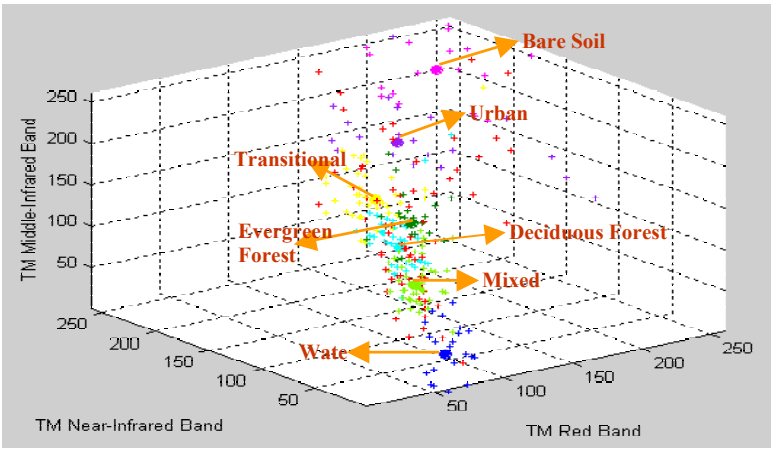


(c) Scattered Distribution of 253 points among the five classes from the Integrated SA-based Algorithm

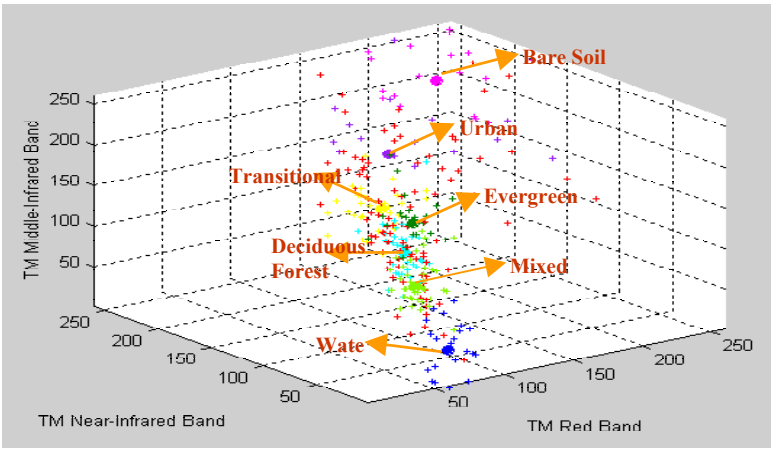
Figure 3. 15 Scatter plots of the 253 reference points from TM-1 and their class labels



(a) Scattered Distribution of 299 points among the seven classes from the *K*-means



(b) Scattered Distribution of 299 points among the seven classes from the Single SA-based Algorithm



(c) Scattered Distribution of 299 points among the seven classes from the Integrated SA-based Algorithm

Figure 3. 16 Scatter plots of the 299 reference points from TM-2 and their class labels

In both applied cases, TM-1 and TM-2, either the I-SA or the S-SA has improved the classification as compared to *K*-means but with significantly greater computational efforts. This verifies our assumption that the SA-based algorithms are able to overcome the local minimum limitations of *K*-means and improve the classification accuracy. The experimental results and analysis described above lead to the conclusion that SA-based classification systems are applicable for land cover classification using remotely sensed data and can significantly improve the classification accuracy when applied to relatively complicated classification applications.

3.4 General Discussion and Conclusion

In this study, algorithms and methodologies of SA originating from the analogy between the physical annealing process of a solid and combinatorial optimization problems have been investigated and adapted to solve the local minimum problem in unsupervised classification and thus improve the classification accuracy. Two SA-based unsupervised classification systems were developed. Their applicability and implementations for LU/LC classification were studied using two remotely sensed images, TM-1 and TM-2, and their classification performances were compared with the classical clustering algorithm *K*-means. Based on the experimental results and analysis, two factors are found essential for the success of the SA-based systems: one is the complexity of the application and another is the selection of appropriate controlling parameters. The nature of unsupervised classification is to establish decision boundaries based on the unlabeled data by optimizing certain error criteria so as to naturally group

the data into clusters. The decision boundaries should be difficult to find in many applications because the search is often very complicated and suffers from local minima. The incorporation of SA procedure proves to be helpful to reduce the adverse effects from these local minima. The local minimum problem may be worse for complex applications because they may suffer more local minima than simple applications. Our experimental analysis suggests that, for simpler applications, the I-SA system is a good choice because it is likely to produce better classification more efficiently. However, for more complicated applications, either when more land classes are desired or the landscape conditions are very heterogeneous, the I-SA system is not appropriate and the S-SA system should be used.

Theoretically SA has been proven to be able to provide the global optimal solution in clustering problems but only in an infinite run time. In practice, the globally optimizing nature of the SA can be taken advantage of to improve the classification accuracy and reduce the uncertainties for land cover classification with appropriate selections of initial temperature and cooling schedule. *K*-means can implement efficiently and produce very satisfactory results if classified patterns are distributed in compact, hyperspherical, and contain well-separated clusters. The performance of SA-based systems is very sensitive to the choices of controlling parameters. If appropriate parameters are chosen, the resultant classification can be improved significantly. However, if the selected cooling schedule is not appropriate, the resulting classification may be far from the global optima. Although effective implementation of the SA-based systems in classification problems is still a delicate issue for which there is no analytical

solution by far, a couple of general guidelines are given from our study. The main aspects are concluded as follows:

1) The selection of each single parameter in the cooling schedule may not be so critical. It may be more important to choose an optimal parameter combination for accurate and efficient classification using SA. The speed of the cooling schedule should be slowed while the application complexity increases.

2) The initial value of temperature T_0 should be sufficiently large, but some lower T_0, which may not allow all transitions to be accepted, may also be applicable in simple applications. T_0 needs to be increased while the application complexity increases.

3) The decrement factor μ plays an important role in SA. The decreasing speed must be slow enough that thermal equilibrium could be reached at each temperature. Candidates at least above 0.80 are highly suggested from our study.

4) The iterations at each temperature IET and the generation probability GP determine the length of the Markov chain at each temperature. The more the IET and the lower GP, the longer the Markov Chain is. There is a trade-off between large decrements μ and small IET and higher GP. We either could choose smaller μ with greater IET and lower GP or choose larger μ with smaller IET and larger GP.

3.5 Recommendations for future research

SA-based systems have been studied for their potential for improving classification accuracies in LU/LC classification applications using Landsat TM images. This study is the first time that SA has been intensively investigated for land cover mapping using remotely sensed Landsat TM images. To test the general applicability and robustness of SA-based classification systems for LU/LC mapping using remotely sensed data, it is necessary to apply the SA-based classification systems for various LU/LC classification applications using many other remotely sensed data other than Landsat TM data. Additional classification applications would be valuable to provide more evidences to demonstrate the potential of SA-based classification systems to improve classification accuracy as compared to many traditional classification approaches suffering from the local minimum problem.

The main advantages of SA are its potential to find near-optimal solutions and its ease of implementation. The disadvantage is the potentially large amount of time required for its application in remote sensing. It is necessary to investigate possibilities of speeding up the SA-based algorithm, such as design of faster sequential algorithm or parallel algorithm. However, rapidly growing and available computing power will lessen the computational problem and facilitate the use of SA in remotely sensing applications. The potential of approximating a global optimum, using the SA-based approaches is very attractive and it is worthwhile to explore other possible ways to combine the SA idea with other classification approaches that suffer from the local minimum problem. In

Chapter 4, we incorporate SA random searching process into the standard Kohonen's Self-Organizing (KSOM) neural network to improve its classification performance.

References

Aarts, E. H. L., and P. J. M. Van Laarhoven, "Simulated annealing: a pedestrian review of the theory and some applications", *Pattern recognition theory and applications,* pp.179-192, 1987.

Aarts, E., and J. Korst, Simulated annealing and Boltzmann machines: a stochastic approach to combinatorial optimization and neural computing. *John Wiley & Sons,* 1989.

Anderson, J. R., *et al.*, "A land use and land cover classification system for use with remotely sensed data," *Washington, DC: U.S. Geological Survey Professional Paper 964*, pp. 28, 1976.

Aragon, C. R., *et al.*, "Optimization by simulated annealing: an experimental evaluation," *working paper*, 1984.

Bandyopadhyay, S., S. K. Pal, and C. A. Murthy, "Simulated annealing based pattern classification," *Journal of Information Sciences*, vol. 109, pp. 165 – 184, 1998.

Belward, A. S., *et al.*, "An unsupervised approach to the classification of semi-natural vegetation from Landsat Thematic Mapper data," *International Journal of Remote Sensing*, vol.11, pp.429 – 445, 1990.

Benediktsson, J.A., P.H. Swain, and O.K. Ersoy, "Neural network approaches versus statistical methods in classification of multi-source remote sensing data," *IEEE*

Transaction on Geoscience and Remote Sensing, vol.28, pp.540-551, 1990.

Bischof, H., W. Schneider, and A. J. Pinz, " Multispectral classification of landsat images using neural networks," *IEEE Transaction on Geoscience and Remote Sensing*, vol.30, no.3, pp.482 – 490, 1992.

Bishop, Y., S. Fienberg, and P. Holland, Discrete multivariate analysis – theory and practice, *MIT Press: Cambridge, Massachusetts*, pp.575, 1975.

Bonomi, E., and J. L. Lutton, " The N-city traveling salesman problem: statistical mechanics and Metropolis algorithm," *SIAM Review*, v.26, pp. 551 – 568, 1984.

Brown, D. E., and C. L. Huntley, "A practical application of simulated annealing to clustering," *Pattern recognition*, vol. 25,no. 4, pp. 401-412, 1992.

Burkard, R.E., and F. Rendl, " A thermodynamically motivated simulation procedure for combinatorial optimization problems," *European Journal of Oper. Res.*, vol.17, pp. 169 – 174, 1984.

Cerny, V., "Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm," *Journal of Optimization Theory and Applications*, vol. 45, pp. 45 –51, 1985.

Dai, X., and S. Khorram, "Data fusion using artificial neural networks: a case study on multitemporal change analysis," *Comput., Environ., and Urban Systems*, vol. 23, pp.19 – 31, 1999.

Duda, R. O., and P. E. Hart. Pattern classification and scene analysis. *New York: Wiley-Interscience*. 1973.

Foody, G.M., M.B. McCulloch, and W.B. Yates, "Classification of Remotely Sensed data

by an Artificial Neural Network: issues related to training data characteristics," *Photogrametric Engineering and Remote Sensing*, vol.61, n.4, pp. 391-401, 1995.

Frank, T.D., "Mapping dominant vegetation communities in the Colorado Rocky Mountain front range with Landsat Thematic Mapper and digital terrain data," *Photogrammetic Engineering & Remote Sensing*, vol. 54, no. 12, pp. 1727 – 1734, 1988.

Fukunaga, K., Introduction to Statistical Pattern Recognition. *New York: Academic*, 1990.

Geman, S., and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Proceeding. Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721 – 741, 1984.

Granville, V., M. Krivanek, and Jean-Paul Rasson, "Simulated annealing: a proof of convergence," *IEEE Transactions on Pattern Analysis and Machine Iintelligence*, vol. 16, no. 6, pp. 652-656, 1994.

Guneriussen, T., *et al.*, "Snow monitoring using EMISAR and ERS-1 data within the European multi-sensor airborne campaign EMAC-95," *the 1997 IEEE International Geoscience and Remote Sensing Symposiums*, vol. 2, pp. 631 – 633, 1997.

Haack, B.N., *et al.*, "Assessment of Landsat MSS and TM data for urban and near-urban landcover digital classification", *Remote Sensing of Environment*, vol. 21, no. 2, pp. 201 – 213, 1987.

Hegarat-Mascle, L., S. D. Vidal-Madjar, and P. Olivier, "Applications of simulated annealing to SAR image and classification problems," *International*

Journal of Remote Sensing, vol. 17, no. 9, pp. 1761-1776, 1996.

Jain, A., and R. Dubes, Algorithms for Clustering Data. *Prentice-Hall, Englewood Cliffs, New Jersy*, 1988.

Jenson, J. R., Introductory digital image processing. *Englewood Cliffs, NJ: Pretentice Hall,* 1996.

Khorram, S., *et al.*, "Comparison of Landsat MSS and TM data for urban land-use classification", *IEEE Transactions on Geoscience and Remote Sensing*, vol.25, no. 2, pp.238 – 243, 1987.

Kirkpatrick, S., C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671-688, 1983.

Klein, R. W., and R. C. Dubes, "Experiments in projection and clustering by annealing," *Pattern Recognition*, vol. 22, no. 2, pp. 213-220, 1989.

Kolonko, M., and M. T. Tran, "Convergence of simulated annealing with feedback temperature schedules," *Probability in the Engineering and Informational Sciences*, vol. 11, pp. 279-304, 1997.

Laarhoven, P.J.M., Theoretical and computational aspects of simulated annealing. *Amsterdam, Netherlands: Center for mathematics and computer science,* 1988.

Leong, H.W., D.F. Wong, and C. L. Liu, "A simulated-annealing channel router," *Proc. ICCAD*, Santa Clara, pp. 226 – 229, 1985.

Locatellli, M., "Convergence properties of simulated annealing for continuous global optimization," *Journal of Applied Probability*, vol. 33, pp. 1127-1140, 1996.

Lundy, M., and A. Mees, “Convergence of an annealing algorithm,” *Math. Prog.*, vol. 34, pp. 111 – 124, 1986.

McErlean, F. J., D. A. Bell, and S.I. McClean, “The use of simulated annealing for clustering data in databases,” *Information Systems*, vol. 15, no. 2, pp. 233-245, 1990.

Metropolis, N., *et al.*, “Equation of state calculations by fast computing machines,” *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087-1092, 1953.

Remund, Q. P., D. G. Long, and M. R. Drinkwater, “Polar sea-ice classification using enhanced resolution NSCAT data,” ," *the 1998 IEEE International Geoscience and Remote Sensing Symposiums*, vol. 4, pp. 1976 - 1978, 1998.

Rose. K., “Deterministic annealing for clustering, compression, classification, regression, and related optimization problem,” *IEEE Proceeding*, vol. 86, pp. 2210 – 2239, 1998.

Schistad Solberg, A.H., *et al.*, “Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images*”, IEEE Transactions on Geoscience* and Remote Sensing, vol. 32, no. 4, pp. 768 – 778, 1994.

Schowengerdt, R.A., Techniques for Image processing and classification in remote sensing, *Academic Press*, pp. 129 – 214,1983.

Selim, S. Z., and K. Alsultan, “A simulated annealing algorithm for the clustering problem,” *Pattern Recognition*, vol. 24, no. 10, pp.1003-1008, 1991.

Sergi, R., *et al.*, “Landsar-TM images classification using principal components analysis and neural networks,” *Geoscience and Remote Sensing Symposium Proceedings*

of IGARSS'95. *Quantitative remote sensing for science and applications*, vol.3, pp.1927 – 1929, 1995.

Serpico, S. B., and F. Roli, "Classification of multisensor remote-sensing images by structured neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 3, pp. 562-578, 1995.

Shlien, S., and A. Smith, "A rapid method to generate spectral theme classification of Landsat imagery," *Remote Sensing of Environment*, vol. 4, pp. 67 – 77, 1975.

Snyder, W., *et al.*, "Segmentation brain images using mean field annealing," *Inform. Process. Med. Imag., Proc. 12th Int. Conf. IPMI, Wye, UK,* pp. 218 – 226, 1991.

Strahler, A.H., *et al.*, "Improving forest cover classification accuracy from Landsat by incorporating topographic information", *Proc., 12th International Symposium on Remote Sensing of the Environment*, pp. 927 – 942, 1978.

Thomson, A.G., R.M. Fuller, and J.A. Eastwoods, "Supervised versus unsupervised methods for classification of coasts and river corridors from airborne remote sensing," *International Journal of Remote Sensing*, vol. 18, pp. 3423 – 3431, 1998.

Tou, J. T., and R. C. Gonzales, Pattern recognition principles, *Addison- Wesley Publishing Company,* 1977.

White, S. R., "Concepts of scale in simulated annealing," *Proc. ICCD*, Port Chester, pp. 646 – 651, 198

4. #Paper 2 Development of an automated neural network classification system for remote sensing based land cover mapping

ABSTRACT --

Given the rapid advances and increasing popularity of Artificial Neural Networks (ANNs) for image classification in remote sensing, there is a need for an automated ANN classification system. To fill this gap, we have developed an automated ANN classification system consisting of two ANN modules: 1) an unsupervised Kohonen's Self-Organizing Mapping (KSOM) neural network module, and 2) a supervised Multilayer Perceptron (MLP) neural network module using the Backpropagation (BP) training algorithm. In the KSOM network module, two training algorithms were provided: 1) the standard KSOM competitive learning algorithm, and 2) a refined KSOM learning algorithm. The refined KSOM learning algorithm incorporated Simulated Annealing (SA) global searching procedures into the standard KSOM neural network to reduce the likelihood that it will be trapped in local minima. To test the applicability of the developed ANN system, we applied the automated ANN system to perform Land-Use/Land-Cover (LU/LC) classifications of a Landsat Thematic Mapper (TM) image using a supervised MLP network and two unsupervised KSOM networks. Our case study shows that the ANN classification system fulfills the tasks of network training pattern creation, network training, and network generalization. To evaluate the classification performance of the ANN classification approaches, the resulting classifications from the

three networks were assessed by the reference data derived from the high spatial resolution Digital Color Infrared (CIR) Digital Orthophoto Quarter Quad (DOQQ) data. Among the three performed networks, the supervised MLP network obtained the superior classification performance than the two unsupervised KSOM networks. Between the two unsupervised KSOM networks, the classification performance of the refined KSOM network was shown to be significantly better than that of the KSOM network. This demonstrates that the incorporation of SA improves the classification performance of the standard KSOM network. We conclude from this study that the automated ANN classification system is robust for land cover mapping applications. This system will be particularly useful when traditional statistical classification methods such as maximum likelihood classifiers cannot be used because the statistical distribution model of the input data is not normal.

4.1 Introduction

In recent decades multispectral classification of remotely sensed data has been widely used to generate thematic Land-Use/Land-Cover (LU/LC) inventories for various purposes, such as urban planning, agricultural crop characterization, forest ecosystem classification (Solberg *et al.* 1994; Khorram *et al.* 1987; Haack *et al.* 1987; Ulaby *et al.* 1982)**.** A variety of classification algorithms have been developed to perform such classification tasks (Ediriwickrema & Khorram 1997; Schowengerdt 1983; Swain & Davis 1978; Duda & Hart 1973). Among them, classification approaches based on Artificial Neural Networks (ANNs) have emerged in recent years and have been shown

to be well-suited for image classification (Dai & Khorram 1999; Heermann & Khazenie 1992; Benediktsson *et al.* 1990). Many previous studies on classification of multispectral images have demonstrated that ANNs perform better than many traditional classification methods, such as maximum likelihood classifiers (Foody & Arora 1997; Bischof *et al.* 1992; Benediktsson *et al.* 1990). ANNs are fundamentally network systems simulating the working process in human mind, in which a large number of small and interconnected processors or neurons work together to solve difficult classification and optimization problems. ANN approaches have a distinct advantage over statistical classification methods in that they are non-parametric and require little or no *a priori* knowledge of the distribution model of input data (Benediktsson *et al.* 1990). Some other superior advantages of ANNs include parallel computation, the ability to estimate the non-linear relationship between the input data and desired outputs, and fast generalization capability.

Despite the rapid advances and increasing popularity of ANNs for image classification in remote sensing, there is no automated ANN classification system provided by commercial remote sensing software packages. To fill this gap, we have developed an automated ANN classification system in this study and investigated its applicability in remotely sensed classification applications. This system offers more feasible and robust classification alternatives than traditional classification methods such as the maximum-likelihood and the Iterative Self-Organizing Data Analysis Technique (ISODATA) that are widely available with most remote sensing software packages. The automated ANN classification system consists of two ANN modules: 1) an unsupervised

Kohonen's Self-Organizing Mapping (KSOM) neural network module, and 2) a supervised Multi-Layer Perceptron (MLP) network module. The MLP network module is trained using the very popular Backpropagation (BP) training algorithm. In the KSOM network module, two training algorithms are provided: 1) the standard KSOM competitive learning algorithm, and 2) a refined KSOM learning algorithm by Simulated Annealing (SA). Because the KSOM network suffers from the same local minimum problem as *K*-means (Klein & Dubes 1989; Laarhoven 1988), the refined KSOM learning algorithm incorporates SA global searching procedures to reduce the likelihood that it will be trapped in local minima.

In this study, two hypotheses have been tested: 1) the ANN classification system developed in this study is well-suited for LU/LC classification purpose, and 2) the refined KSOM-SA network has the potential to overcome the local minima problem and improve the classification performance of the standard KSOM network. To test these two hypotheses, we have applied this ANN system to perform LU/LC classifications based on Landsat Thematic Mapper (TM) data using the supervised MLP network and two unsupervised KSOM networks.

We introduce the MLP and KSOM neural network models in section 4.2. In section 4.3, the automated ANN classification system developed in this study and the associated classification procedures are described. A case study using the ANN system is reported in section 4.4, where the classification performance of each of the classifiers is compared and evaluated based on the experimental results. Along with this case study, related implementing issues are also discussed. Conclusions are given in section 4.5.

4.2 Neural network Classification Approaches

A variety of ANN approaches have been applied to many LU/LC classification applications using remotely sensed data (Carpenter *et al.* 1997; Bishop1995; Bezdek *et al.* 1992; Benediktsson *et al.* 1990). Among them, the two most popular network models are the supervised MLP network model (Rumelhart & McClelland 1986) and the unsupervised KSOM neural network model (Kohonen 1982). In section 4.2.1, the MLP network model and the supervised BP training algorithm are introduced. The KSOM network model and two unsupervised learning algorithms are presented in section 4.2.2.

4.2.1 Multilayer Perceptron (MLP) neural network

The MLP neural network is the most commonly used network model for image classification in remote sensing (Kanellopoulos & Wilkinson 1997; Paola & Schowengerdt 1997; Foody *et al.* 1995). Basically, MLP is a supervised neural network classification model using single or multilayer perceptrons to approximate the inherent input-output relationships. MLP networks are usually trained with the supervised Backpropagation (BP) training algorithm (Rumelhart & McClelland 1986). The training or learning process requires a set of training patterns with inputs and corresponding desired outputs. Basically, the training process is a search for a network interconnected with a suitable set of weights that can minimize the total differences between the network outputs and desired outputs for the training patterns. Based on the minimization goal, the mean square error (MSE) J is used as a classification performance criterion given by

$$J = \frac{1}{2N}\sum_{i=1}^{N}\varepsilon_i^2 \qquad \text{--- (4.1)}$$

where N is the number of training patterns. ε_i^2 is the Euclidean distance between the network output of each training pattern and the desired output. Once the training process is completed, the well-trained MLP network can be used to classify new patterns.

A typical MLP consists of one input layer, one or more hidden layers, and one output layer. Figure 4.1 shows a typical three-layer MLP neural network architecture with 4 input nodes in the input layer, 10 hidden nodes in the hidden layer, and 5 output nodes in the output layer. This network architecture is often denoted as 4 -10 - 5.

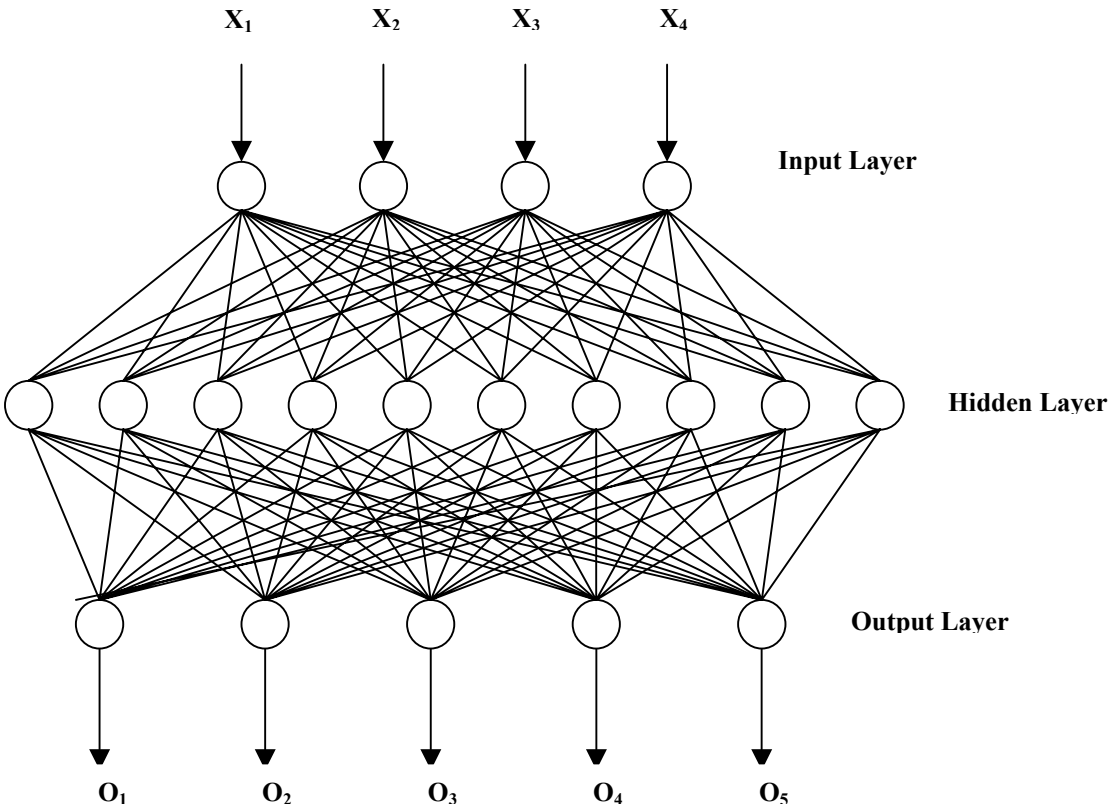


Figure 4. 1 The structure of three-layer neural network with 4 input nodes, 10 hidden nodes, and 5 output nodes

All nodes in different layers are connected by associated weights. For each input pattern presented to the network, the current network output is computed. First, the net input to any node *j* in the hidden or output layer is calculated from all the nodes in the previous layer as follows:

$$net_j = \sum_{i=1}^{k} (w_{ij} x_i) + b_j \qquad \text{--- (4.2)}$$

where w_{ij} is the weight linking the i^{th} node in the previous layer and the j^{th} node in the next layer, b_j is the bias term associated with this node, x_i is the output or activation value of the i^{th} node from the previous layer, and *k* is the number of the nodes in the previous layer. Then, an activation sigmoid function, *f*, of the net value for any node *j* in hidden and output layers are calculated:

$$f(net_j) = \frac{1}{1 + e^{(-net_j)}} \qquad \text{--- (4.3)}$$

The outputs of the nodes in the hidden layer and the output layer depend on the inputs from the previous layer, associated weights, activation function, and the bias. In the next step, the calculated outputs are compared with the desired outputs and the error or difference between them is then backpropagated to adjust the weights between layers so as to move the network output closer to the desired output using the following generalized delta rule:

$$w_{ij}(n+1) = w_{ij}(n) + \eta \delta_j(n) x_i(n) \qquad \text{--- (4.4)}$$

where $w_{ij}(n+1)$ and $w_{ij}(n)$ are the weight at iteration *n+1* and iteration *n*;

n: iteration,

η is the learning rate,

$x_i(n)$ is the activation value of the i^{th} node in the previous layer,

$\delta_j(n)$ is the backpropagated error of the *jth* node in the current layer that is defined as:

$$\delta_j(n) = O_j(1 - O_j)\sum_{i=1}^{k} \delta_i w_{ij} \qquad \text{--- (4.5)}$$

where O_j is the output of the j^{th} node in the current layer,

δ_i is the backpropagated error of the i^{th} node in the previous layer,

k is the number of nodes in the previous layer.

In the standard BP algorithm, the generalized delta rule used to update the weights is usually very slow and unstable. Principe *et al.* (1999) suggest incorporating a *momentum term* (the past increment to the weight) to speed up and stabilize the BP learning. In this *momentum learning*, the generalized delta rule in Equation 1.4 is modified by taking incremental changes as follows:

$$w_{ij}(n+1) = w_{ij}(n) + \eta \delta_j(n) x_i(n) + \alpha \Delta w_{ij}(n-1) \qquad \text{--- (4.6)}$$

Where α is the momentum constant,

$\Delta w_{ij}(n-1)$ is the weight change from previous iteration.

Detailed implementation procedures for MLP networks can be found with Principe *et al.* (1999). Although there are many successful MLP application examples (Dai & Khorram 1999; Kanellopoulos & Wilkinson 1997; Paola & Schowengerdt, 1997; Foody *et al.* 1995; Bischof *et al.* 1992; Lee *et al.* 1990), it is widely recognized that

MLPs are sensitive to many operational factors, such as the size and quality of training data set, network architecture, training parameters, and overfitting problems. These factors are application-dependent and have to be addressed by means of trial and error. We will discuss these operational issues together with our case study in section 4.4.

4.2.2 Kohonen's self-organizing mapping (KSOM) neural network

Kohonen's Self-Organization Mapping (KSOM) network, proposed by Kohonen (1982), is a relatively simple clustering tool to group data into clusters through competitive learning. A KSOM network is basically a competitive two-layer network that consists of: 1) an input layer with a number of input nodes, 2) an output or competitive layer with a number of competitive nodes arranged by rows and columns, and 3) a set of connection weights linking input nodes with each competitive node. The number of input nodes normally equals the number of input dimensions while the number of competitive nodes is contingent upon the number of spectrally separable clusters. Figure 4.2 shows the typical structure of a KSOM network with an input layer with 4 input nodes and a competitive layer with 9 competitive nodes that are arranged by 3 rows and 3 columns.

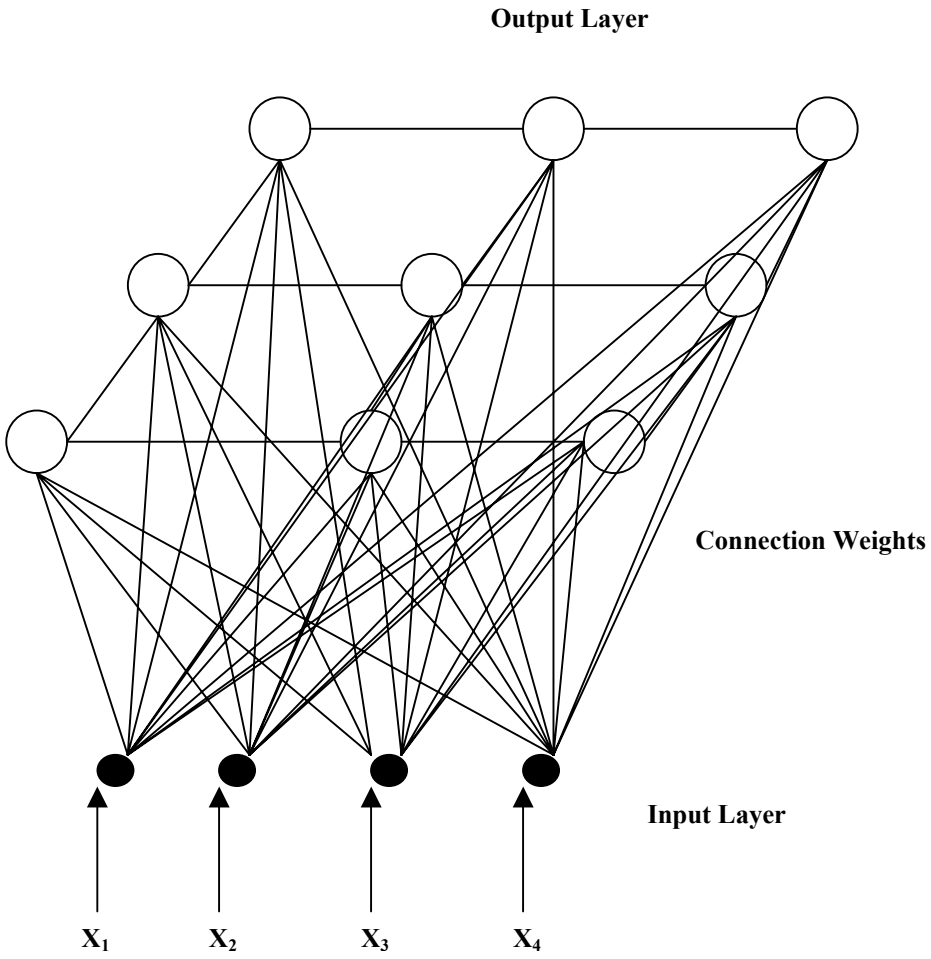


Figure 4. 2 The structure of Kohonen's Self-Organization Mapping Neural network

In a KSOM network, when an input pattern is presented to the KSOM network, the node in the competitive layer with the weight vector closest to this input vector in terms of Euclidean distance is called the "winning" node. Only this winning node and the nodes in its neighborhood update their weight vectors during the training procedure. This spatial neighborhood property makes the KSOM network different from other competitive networks (Principe *et al*, 1999). Due to this property, KSOM networks have the ability to preserve the topological relationships in the original input data, by which

the clusters with similar spectral signatures in image classification are assigned to the neighboring nodes in the competitive layer (Goncalves *et al.*1998). The standard KSOM algorithm is summarized by Lippmann (1987) and Chen *et al.* (1999) as follows:

1) Initialize weights from *n* inputs to the *k* output nodes to the interval [0,1] randomly, and set the initial radius of neighborhood $N_g(t)$. Select the iteration limit for the algorithm, *iteration_max.*
2) Present an input pattern X_i to network.
3) Compute the Euclidean distances $d_{ij}(t)$ between the input pattern X_i and each output node *j* using the equation as below

$$d_{ij} = \sum_{i=1}^{n} \left(x_i(t) - W_{ij}(t) \right)^2 \qquad \text{--- (4.7)}$$

4) Select the output node with the minimal distance as the winning node.
5) Update the weights of the winning node and all the nodes within its neighborhood

$$W_{ij}(t+1) = W_{ij}(t) + \mathrm{N}_{\mathrm{g}}(t)\eta(t)\left(x_i(t) - W_{ij}(t) \right) \qquad \text{--- (4.8)}$$

where $\eta(t)$ denotes the learning rate at the iteration *t*, which is usually decreased as *t* increases, and $N_g(t)$ the neighborhood radius at the iteration *t*.
6) Send the next input pattern to the input layer, and then go to step 3) until all learning patterns are sent once.
7) Update the learning rate and the neighborhood radius.
8) Increment *t* by 1, and then go to step 2) until *t* = *iteration_max* or the average distance between the input pattern and the winning node drops below a specific threshold.

KSOM network is one of the most commonly used unsupervised neural network approaches in remotely sensed classification applications (Babu 1997; Baraldi & Parmiggiani 1995; Bezdek *et al.* 1992). As one of the many unsupervised classification techniques, the KSOM neural network is ideal for classification problems where class labels for training patterns are impossible or very expensive to obtain, for example, when the landscape condition is very heterogeneous. KSOM has operational advantages over supervised methods in terms of reduced interaction time by the analyst; however, it offers less control over resulting classes (Thomson *et al.* 1998). For LU/LC mapping purposes, it is the analysts' responsibility to relate the resulting spectral clusters to the corresponding land-cover classes on the ground (Jensen 1996).

A KSOM network can obtain similar classification results as the *K*-means clustering algorithm because they have a similar clustering objective to minimize the distances between the input patterns and the assigned clusters centers via a gradient descent-based searching process. It is generally known that these *K*-means type algorithms can produce good results only if the clusters are compact, well separated in the feature space, and hyperspherical in shape when *Euclidean* distance is used. However, in a complex problem that cannot be solved by a simple convex cost function, the local minima problem is inevitable. Simulated Annealing (SA) was developed on the basis of an analogy between the physical annealing process of solids and the large combinatorial optimization problems (Kirkpatrick *et al.* 1983; Cerny 1985). Geman and Geman (1984) have proved that SA has great potential to find or approximate the global or near-global optimal in a combinatorial optimization problem. The basic idea of SA is

to incorporate some randomness to the assignments of cluster labels to pixels in the clustering procedure, thus reducing the likelihood of getting trapped by local minima. For such reason, a SA-based classification system has potential to reduce the limitation of local minima so as to improve the classification accuracy for land cover classification. In Chapter 3, we have empirically shown that the incorporation of SA into assignments of cluster labels to pixels in the standard *K*-means clustering procedure has reduced the likelihood that it is trapped by local minima and thus improved the classification accuracy.

In this paper, we have refined the standard KSOM learning algorithm by incorporating SA global searching procedures. The refined KSOM-SA uses a cooling schedule required by most SA-related applications. In the KSOM-SA training, we introduce a control parameter denoted as the temperature *T*. The KSOM-SA procedure is described as follows:

1) We start with a high temperature *T* that decreases gradually and an arbitrary initial assignment of each training pattern to an arbitrary output node.
2) An input pattern X_i is randomly selected and presented to the KSOM-SA network. The input pattern presented to this network is selected based on a parameter of generation probability.
3) For each randomly selected input pattern X_i, we reassign it to an output node *n* that is different from its previous assigned output node *m*.

4) The Euclidean distances d_{mi} and d_{ni} of the input pattern X_i from the associated weight vectors w_{mi} and w_{ni}, and Δd_{mn}, the distance change between d_{mi} and d_{ni}, are respectively calculated using Equation 4.7.

5) Instead of simply choosing the output node with the closest weight vector in the standard KSOM learning algorithm, we determine the winning node either with a distance decrease Δd_{mn} <0, or with a distance increase according to a positive probability $\exp(-\Delta d_{mn}/T)$ where T is the temperature at this state.

6) Repeat 2) to 5) until a certain number of iterations are reached at this temperature.

7) Decrease the temperature *T* in a given schedule.

8) Go to 2) until the temperature *T* approaches zero.

The KSOM-SA learning algorithm is less computationally efficient than the standard KSOM learning algorithm, but is very likely to escape the local minima limitations and improve classification performance. In the following sections, we use the KSOM network to refer to the standard KSOM network and the KSOM-SA network to the refined KSOM network with the incorporation of SA.

4.3 Development of an automated ANN classification system

As we have mentioned before, although ANN approaches have become increasingly popular for image classification in remote sensing in the past decade, their

applications in remote sensing are still limited to professional researchers. This is due to the lack of an automated and robust ANN system in existing commercial remote sensing software packages. Thus, we proposed to develop an automated ANN classification system to offer ANN classification alternatives in addition to those traditional classification methods commonly used. This ANN classification system was built within the working environment provided by the commercial remote sensing software ERDAS IMAGINE.

The automated ANN classification system consists of two ANN classification modules: 1) a KSOM module based on unsupervised KSOM neural networks, and 2) a MLP module based on supervised MLP neural networks. In the ANN system, each classification module is composed of several sub-modules: the pattern conversion sub-module, the network training sub-modules, and the network generalization sub-module. A working flowchart of the ANN-based classification system and functions of each sub-module are summarized in Figure 4.3. The pattern conversion sub-module performs the following functions: 1) sampling a certain number of training and testing patterns from a number of selected image subsets, 2) scaling the input pattern into the network operational interval, and 3) generating training or testing pattern files. In the pattern conversion sub-module of the supervised MLP network, the corresponding class label must be provided for each pattern. Network training sub-modules provide the graphical user interfaces to allow the user to interactively define the network architecture and training parameters needed for the network training and to perform the training once all the parameters are set. In the KSOM module, two training sub-modules are provided

including a standard KSOM training sub-module and a KSOM-SA training sub-module. In the MLP module, the BP training sub-module is used. During a network training trial using each of the training sub-modules, an error file is created to record the training MSEs to assist in monitoring the training behavior and selecting appropriate network and parameters. After training is completed, network generalization sub-modules are implemented to generalize the entire image using the trained network and produce the classified map.

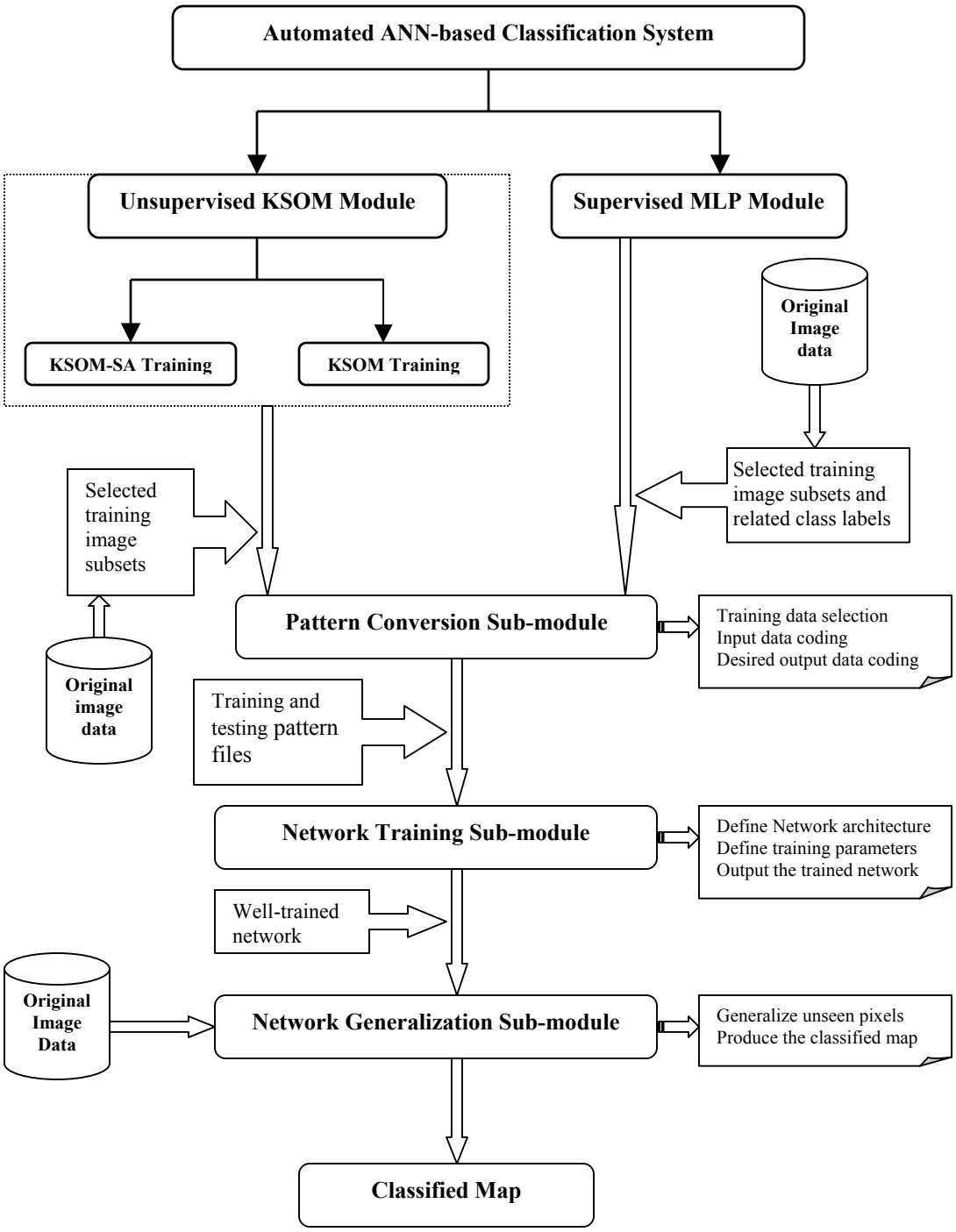


Figure 4. 3 The flowchart of the ANN-based Classification System

The ANN classification system is fully automated, menu-driven, and provides user-friendly graphical interfaces. The graphical user interfaces of the ANN system that are shown through Figure 4.4 to Figure 4.8 were built using ERDAS IMAGINE EML (ERDAS 1994). The ANN classification algorithms were implemented using C/C++ and the ERDAS IMAGINE Toolkit (ERDAS 1994).

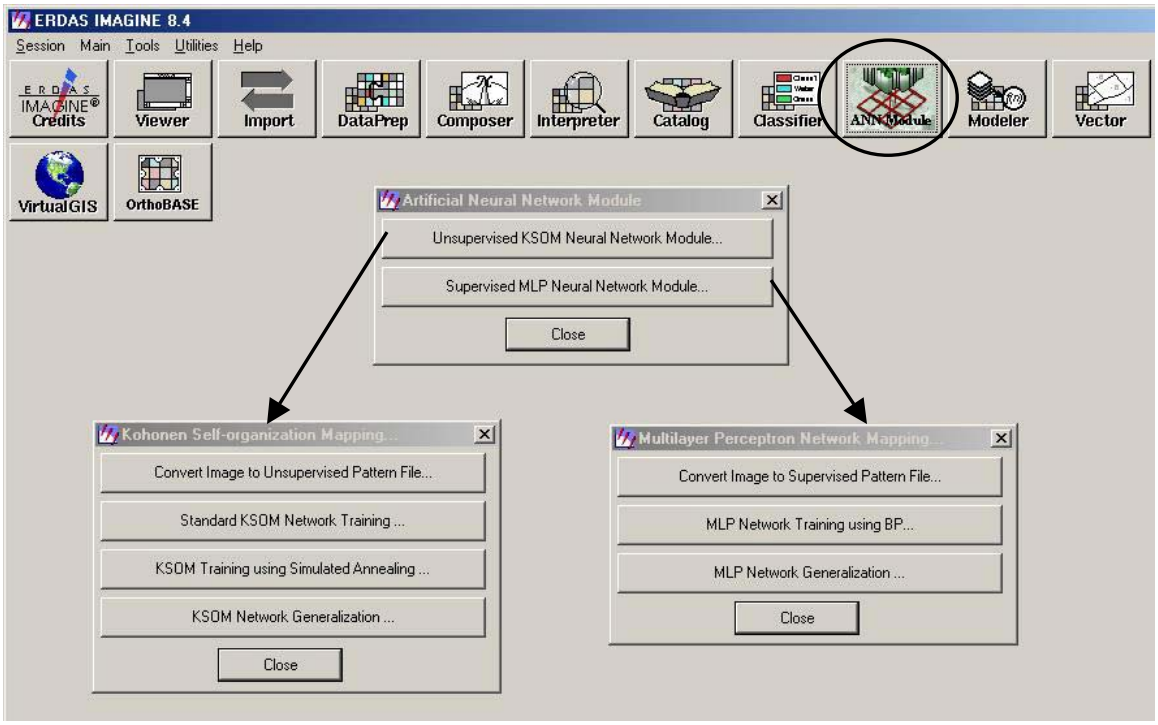


Figure 4. 4 Main Interface of the ANN System

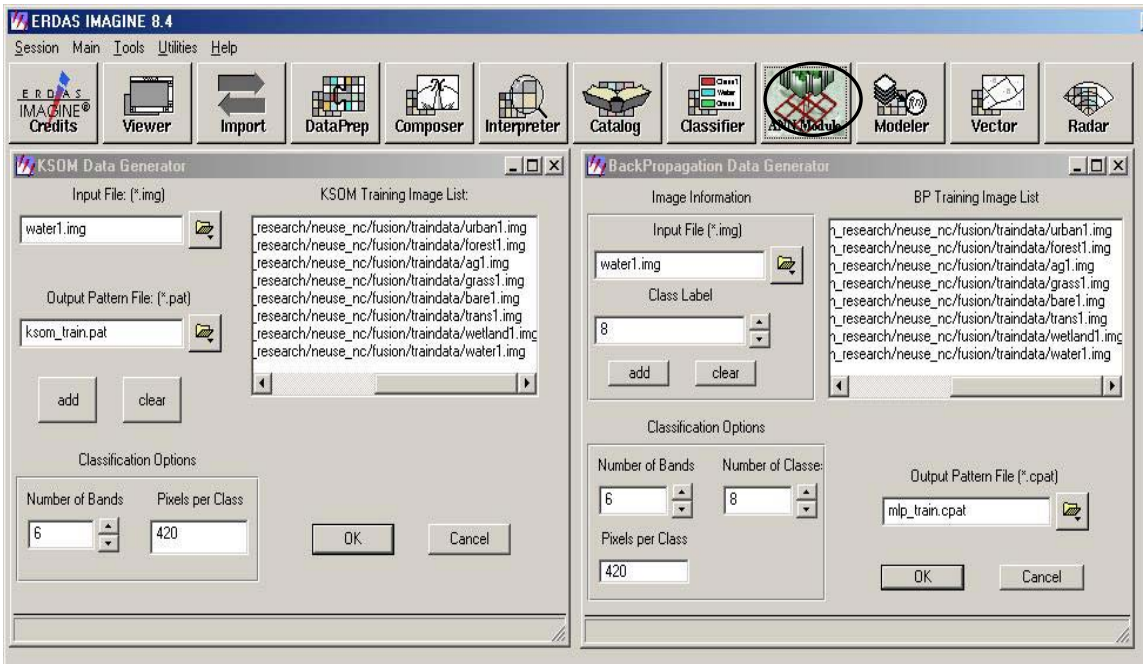


Figure 4. 5 Pattern Conversion Sub-module Interfaces of the ANN System

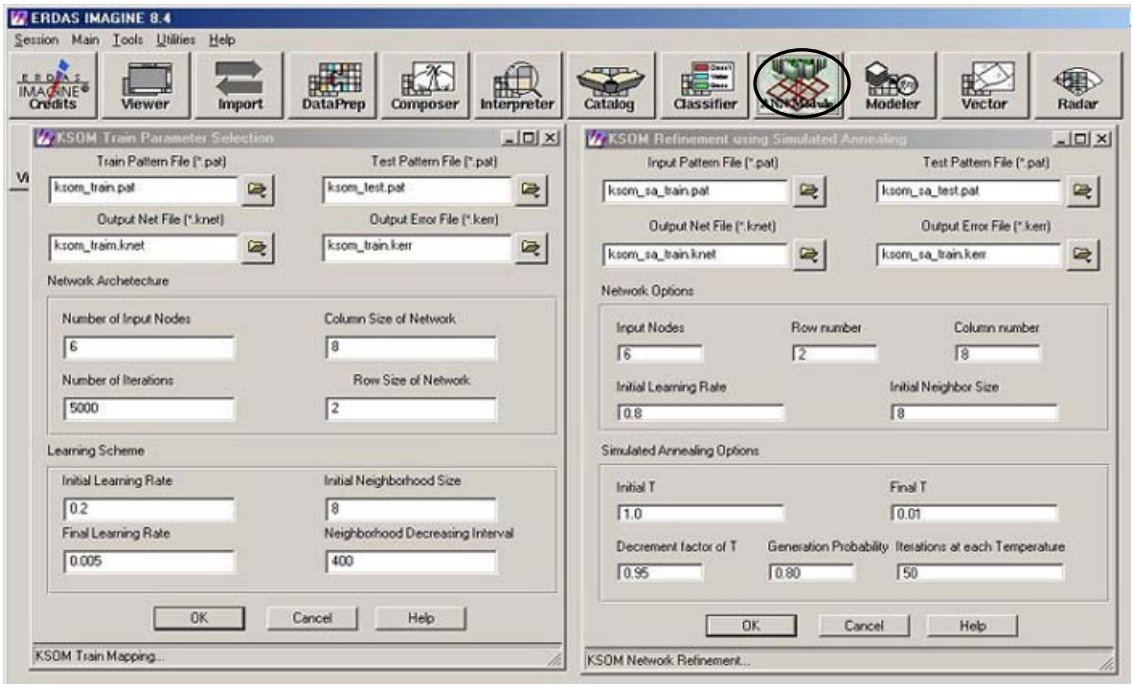


Figure 4. 6 KSOM Training Sub-module Interfaces of the ANN System

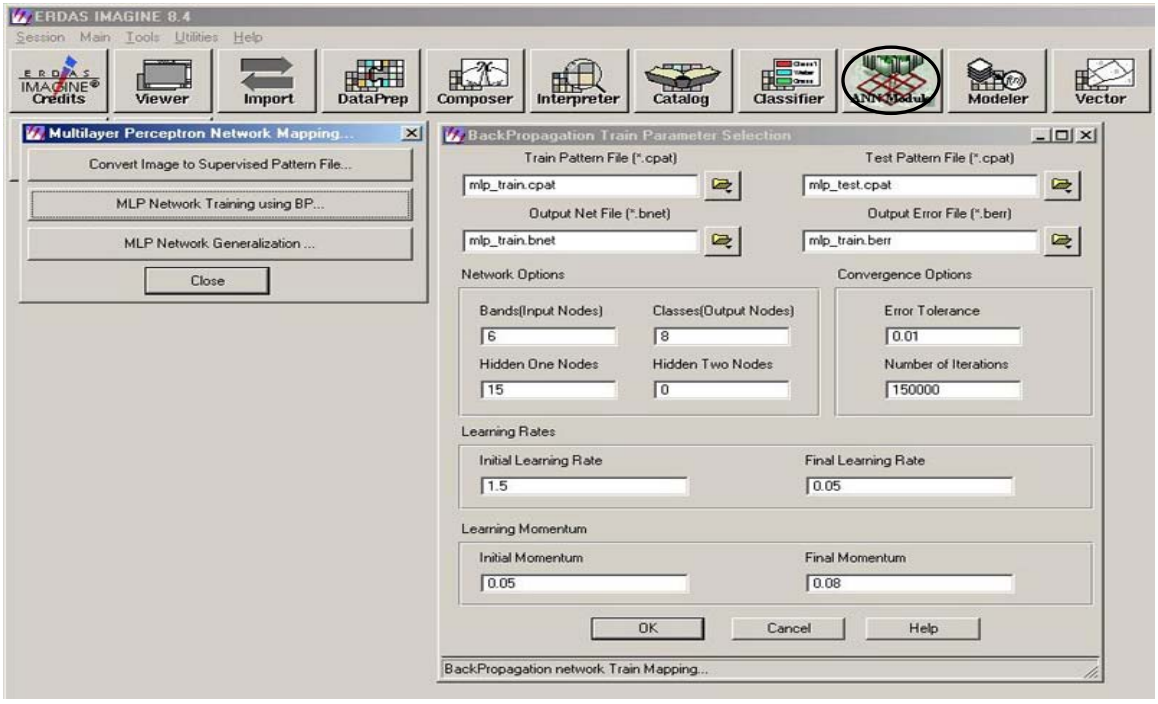


Figure 4. 7 MLP Training Sub-module Interface of the ANN System

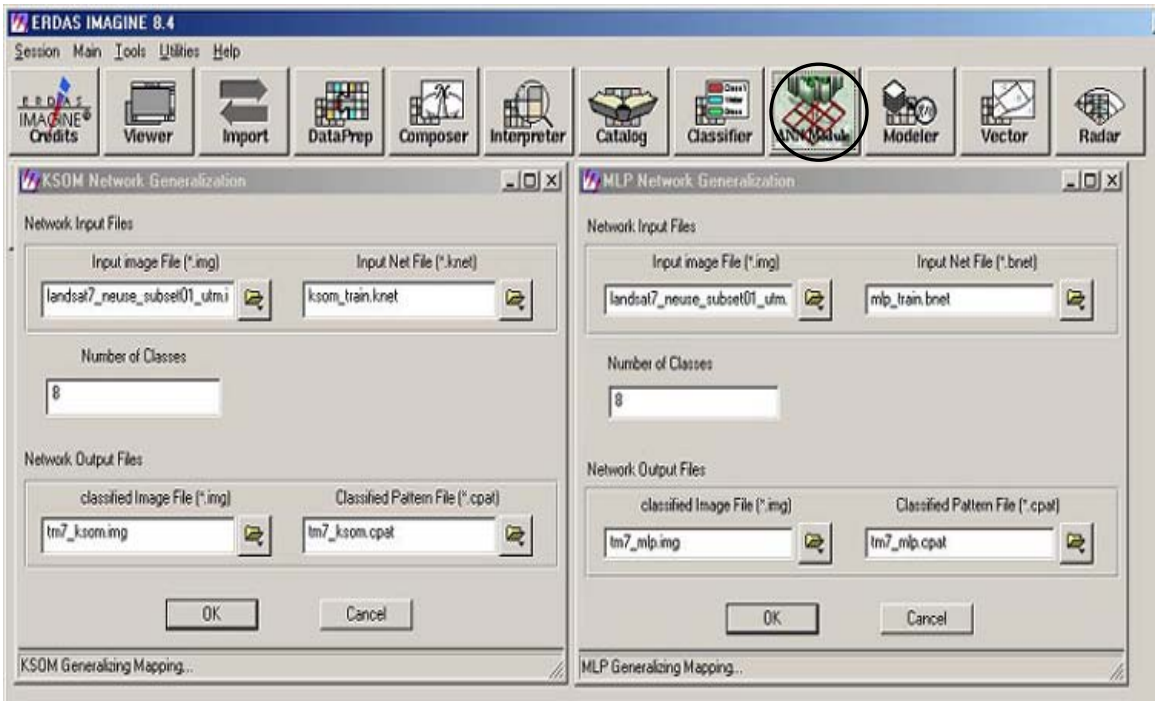


Figure 4. 8 Network Generalization Sub-module Interfaces of the ANN System

4.4 Case study: neural network classification

To test the suitability of the developed ANN system in land cover mapping, we applied the system to perform LU/LC classifications of a Landsat TM image in the selected study area using the supervised MLP network and two KSOM networks.

4.4.1 Study area and classification scheme

In this case study, the lower Neuse Basin area in North Carolina was used to perform the land-use/land cover classification. The study area covers a large region consisting of four counties including Craven, Jones, Pamlico, and Onslow. The study area is chosen because there are a variety of urban, agricultural, hydrologic thematic features in the coastal plain. In addition to the dominating classes of forest, agricultural lands, and open water, there exist large areas of woody wetlands and transitional lands. The Landsat TM image of the study area, acquired in September 1999 (Figure 4.9), was used to perform the classification. By visually interpreting the Landsat TM image and high spatial resolution Digital Color Infrared (CIR) Digital Orthophoto Quarter Quad (DOQQ) data of the study area acquired between January and March 1998, the classification scheme (Table 4.1) was determined by reference to the land use/land cover classification scheme proposed by Anderson *et al.* (1976). Eight major land categories were defined. In this application, six TM bands excluding the thermal band were used for the classification. We exclude the thermal band because its spatial resolution (60m) is different from other bands (30m), which requires additional preprocessing work.

Figure 4. 9 The standard false color display of the acquired Landsat TM image

Table 4. 1 Classification scheme and category definition

Class Number	Class Name	Class Definition
1	Urban	Commercial/Industrial/Residential/transportation
2	Forest	Natural Forested Upland including evergreen, deciduous, and mixed forests
3	Planted crop field	Planted crop fields for the production of crops
4	Grass/pasture	Vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes, or hay crops or pasture
5	Bare/fallow area	Bare construction sites, rock, sand, or fallow agricultural land
6	Transitional area	Areas dynamically changing from one land cover to another.
7	Woody wetland	Areas of forested or shrubland vegetation where soil or substrate is periodically saturated with or covered with water.
8	Water	All areas of open water

4.4.2 Operational issues in neural network classification

Although neural network approaches have been shown to be powerful in many studies, the performance of the ANN classifiers is found to be sensitive to many factors: 1) the size and complexity of network, 2) controlling parameters such as learning rate, and 3) the quality and quantity of training data sets (Bischof & Leonardis 1998; Foody & Arora 1997; Serpico & Roli 1995).

- Training data collection

Among all the factors affecting neural network classification performances, the quality and size the training data set have the most influence on the generalization capability of the resulted network classifier and the final overall classification accuracy (Foody *et al.* 1995; Zhuang *et al.* 1994). The training data set should provide as complete a representative description of each land category as possible. The size of the training set should considerably increase along with the increases of the spectral variability of desired classes, the number of associated weights, and the desired classification accuracy (Principe *et al.* 1999). Foody (1999) points out that incorporation of some boundary patterns from each class is particularly useful in achieving satisfactory classification accuracy. These boundary patterns refer to those patterns in each class that lie farther away from the class center and close to the decision boundary. A network trained with boundary patterns may have lower training accuracy, but can have better generalizing performance than the network trained with pure patterns.

- Network architecture

In a typical MLP neural network, the number of input and output nodes are usually determined by the specific application, i.e., the number of input nodes equals the input dimension and the number of output nodes equals the number of desired land categories. Single hidden layer networks are found to be sufficient for most classification problems (Kanellopoulos & Wilkinson 1997; Paola & Schowengerdt 1997; Lippman 1989). Thus the remaining problem is to determine the number of hidden nodes in the single hidden layer. For a particular problem, the number of hidden nodes is usually determined by trial and error. Networks with varying number of hidden nodes, e.g., the same, twice, or three time the number of input nodes or output nodes, are experimented. Then the network architecture with the best performance is chosen. In the KSOM network, the number of the input nodes equals the input dimensions. However, the number of the KSOM output nodes does not always match the number of the desired land categories and is normally determine by experimentation. The optimal number of the output nodes in the KSOM network indicates the number of separable spectral clusters and is closely related to the geometrical characteristics of the input data. Our system provides the ability to determine the network architecture by trial and error. The error file recording the MSEs of training and testing data sets is created from each network training, which is used to decide which architecture is optimal.

- Network input/output coding

Most neural network algorithms are designed to deal with continuous data ranging from 0 to 1. Since most spectral band value ranges from 0 to 255, each input data is

scaled by 255 before it is presented to the network. For a similar reason, network output values range from 0 to 1, which must be translated to represent the LU/LC classes in remote sensing classification application. In our system, the input coding is conducted by the pattern conversion sub-module, which automatically scales each input pattern into an input vector within the range while sampling training or testing patterns from a number of image subsets. In the pattern conversion sub-module of the MLP classification module, the desired class label of each input pattern is coded as a unit vector with the same size as the number of classes, in which only the element with the class label is equal to 1 and all other elements are assigned to 0. The scaled input vectors and the coded desired output unit vector (supervised case) are stored in pattern files, which are presented to network training. Once the training is completed, the network generalization sub-module performs the final classification of the entire image by assigning each input pixel to the class of the output node with the highest output value or closest weight vector.

- Training parameters

The performance of neural network-based classifiers is sensitive to the selection of training parameters. The number of training parameters varies with the type of network and training algorithm used. In our system, the required parameters of the MLP network include: 1) initial learning rate, 2) final learning rate, 3) the momentum rate, and 4) the number of training epochs. To make the network training stable, the learning rate must be kept small enough. However, the computational cost of using a very small learning rate is high. Thus, we usually start with a slightly larger learning rate to run the training

faster at the early stage of training and then gradually lower the learning to stabilize the training. The momentum rate should also be chosen via experimentation. The chosen number of iterations must be large enough to learn sufficient class membership knowledge from the training data set, but not too large to have the training data over-trained. We use epoch training in the MLP network. In the epoch training, the weight adjustment from each input pixel is computed and stored without changing the weights. After the whole training set passes through the network, the average weight adjustment of the whole training set is used to update the weight. Epoch training is more efficient and stable than pixel-by-pixel training (Principe *et al.* 1999).

Standard KSOM network training requires the definition of the following parameters: 1) the initial learning rate, 2) the final learning rate, 3) the initial neighborhood radius, 4) the neighborhood decrement interval, and 5) the number of training iterations. The selection of learning rates and the number of training iterations is similar to that in MLP network. The initial neighborhood radius is usually set equal to the larger size of row or column and decreases after a certain number of iterations. The determination of the initial neighborhood radius and the neighborhood decreasing factor is crucial for KSOM networks to achieve a topology-preserving map from the input space to the discrete output. In addition to these standard KSOM training parameters, KSOM-SA training requires more parameters. The following parameters are combined as a cooling schedule including: 1) the initial value of the control temperature T, 2) the decrement factor for decreasing T, 3) the final value of T, 4) the number of iteration at each temperature, and 5) a generalization probability to control which pixel in the

training set is selected to train the network. In Chapter 3, we have developed a couple of guidelines for selecting these SA-related parameters. An optimal selection of parameter combination is believed to be critical to obtain good classification accuracy.

- Overfitting problem

In neural network classification, when the neural network classifier is over-trained, it will start to focus on learning specifics of the particular training set that are not the typical characteristics of the whole data set. This problem is called overfitting. The problem can considerably decrease generalization accuracy when some land classes are not properly represented in training data sets. One of the most effective methods to avoid overfitting is to use a cross-validation approach to stop the training at an appropriate time (Principe *et al.* 1999). Basically, two data sets should be collected: one is for training the network and the other one is for testing the trained network. During network training, only the training data set is used to train the network and update the weights. However, the classification performances with both testing and training data are computed and checked. If training error keeps decreasing, while testing error continuously increases, the training will be terminated. Via the cross-validation approach, the training can be stopped before the overfitting occurs (Paola & Schowengerdt 1995). Furthermore, the cross-validation approach can provide clues as to whether the collected training data sets completely represent land cover classes. This cross-validation is the default training method in our system.

4.4.3 Neural network classification and discussions

For each class, multiple image subsets for each class were first selected from the TM image using ERDAS IMAGINE tools. Our goal is not only to collect homogeneous image subsets but also to collect some image subsets from heterogeneous areas, in which many boundary pixels are included. The number of the image subsets for each land category varied from the spectral variability within that category. The categories with high within-class spectral variability, such as urban and crop field, had more image subsets than others. Then the selected image subsets were processed using the pattern conversion sub-modules in the ANN classification system. By using the pattern conversion sub-module, a certain number of training and testing patterns per class were extracted from these image subsets, coded, and saved into training and testing pattern files. In MLP classification, the corresponding class label related to each sampled pattern was provided. In this application, a total of 3360 training pixels and 360 testing pixels were chosen with 420 training and 45 testing pixels for each of the eight classes. During the network training, only the 3360 training patterns were used to train the network. However, after each training iteration, the training and testing MSEs of training and testing sets were calculated and recorded into an error file, which were used to monitor the training behaviors and assist selecting optimal network architecture and training parameters.

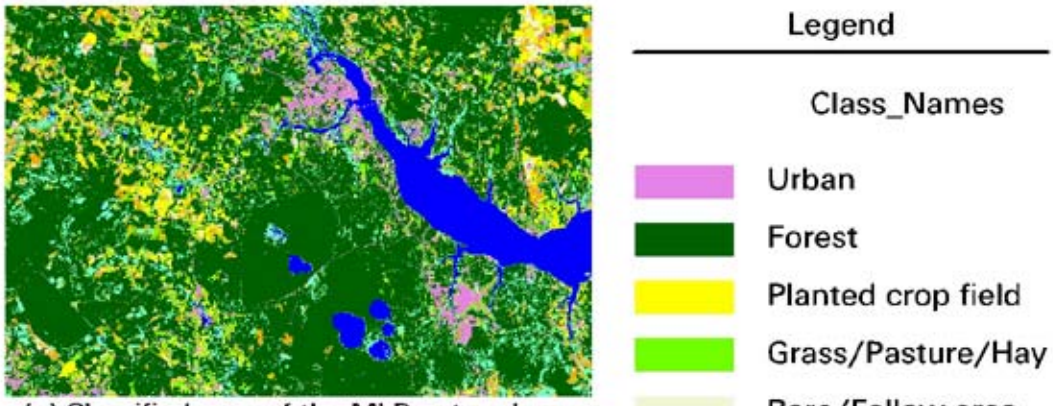
In the MLP classification, we used the three-layer perceptron network, consisting of 6 input nodes and 8 output nodes. We ran a number of MLP experiments using different hidden nodes and training parameters. From each trial, an error file was created

to record the training and testing errors of each iteration along the training. Based on the resulting error files, the optimal network architecture and training parameters was chosen. The optimal MLP network architecture in this application was found to be a single layer network with 15 hidden nodes. The best MLP training was obtained with training parameters are set as below: i) training epochs (150,000), ii) initial learning rate (1.5), iii) final learning rate (0.05), and iv) momentum (0.08). The well-trained MLP network was input into the network generalization sub-module and then used to classify the TM image into a classified map with 8 LU/LC classes shown in Figure 4.10(a). The learning curve of the mean squared errors vs. training epochs for both training and testing sets is plotted in Figure 4.11, indicating a good generalization of the trained MLP network. From Figure 4.11, we observed that although we had set 150,000 training epochs, the training method actually stopped the training at the 122,302th epoch because there was no further training improvement in the testing data set. The cross-validation training method proves effective in this application in terms of avoiding overfitting.

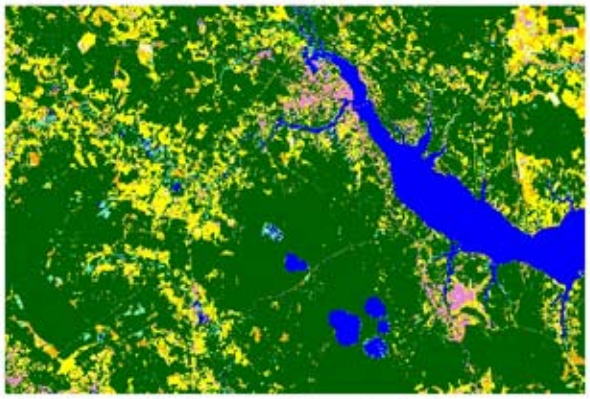
The optimal KSOM network training used 16 output nodes arranged as 2 rows and 8 columns. The parameters in the optimal KSOM training were set as below: i) training iterations (5000), ii) initial learning rate (0.2), iii) final learning rate (0.005), iv) initial neighborhood size (8), and v) the neighborhood decrement interval iterations (400). The optimal KSOM-SA training was obtained using the same architecture as the KSOM network. The parameters in the optimal KSOM-SA training were set as below: i) initial T (1.0), ii) final T (0.01), iii) decrement factor of T (0.95), iv) generation probability (0.80), v) iterations at each T (50), vi) initial learning rate (0.8), vii)

decrement factor of learning rate (0.95), viii) initial neighborhood size (8), and ix) neighborhood decrement factor (0.75). Using the two KSOM based networks, instead of the eight classes we expected, 16 spectral clusters were resulted from the two trained KSOM and KSOM-SA networks. We manually interpreted the 16 clusters and matched them with one of the eight classes. The two feature maps from the KSOM and KSOM-SA network plotted in Figure 4.12(a) and Figure 4.12(b) were created to reflect the information class membership of each spectral cluster in this classification application. The resulting classified maps with eight classes from the KSOM and KSOM-SA networks are shown in Figure 4.10(b) and Figure 4.10(c).

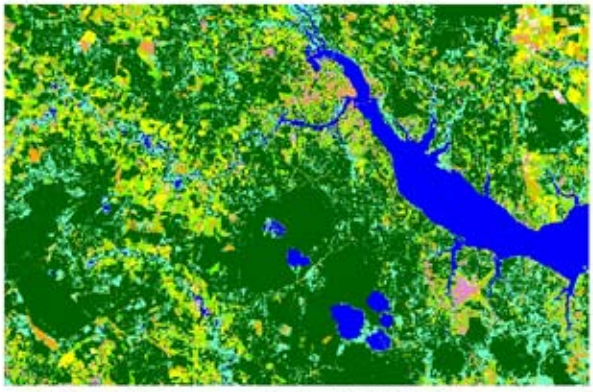
The KSOM and KSOM-SA networks have the topological preservation capability in a way that pixels with similar spectral values are assigned to the neighboring classes, and nodes in the competitive layer with similar spectral signatures are located as near neighbors (Goncalves *et al.*1998). Our observation of these two feature maps shown in Figure 4.12(a-b) verifies this phenomenon. Furthermore, by analyzing the relative location of each spectral cluster and its represented information class to others, we obtain some knowledge of the inherent spectral relationship between classes such as the within-class and between-class spectral variability. From Figure 4.12(a) and Figure 4.12(b), we see that the urban class consists of four spectral clusters, three of which are within the close neighborhood and one of which is farther away, indicating its widely scattered within-class variability.



(a) Classified map of the MLP network



(b) Classified map of the KSOM network



(c) Classified map of the KSOM-SA network

Figure 4. 10 Classified maps of the three network classifiers

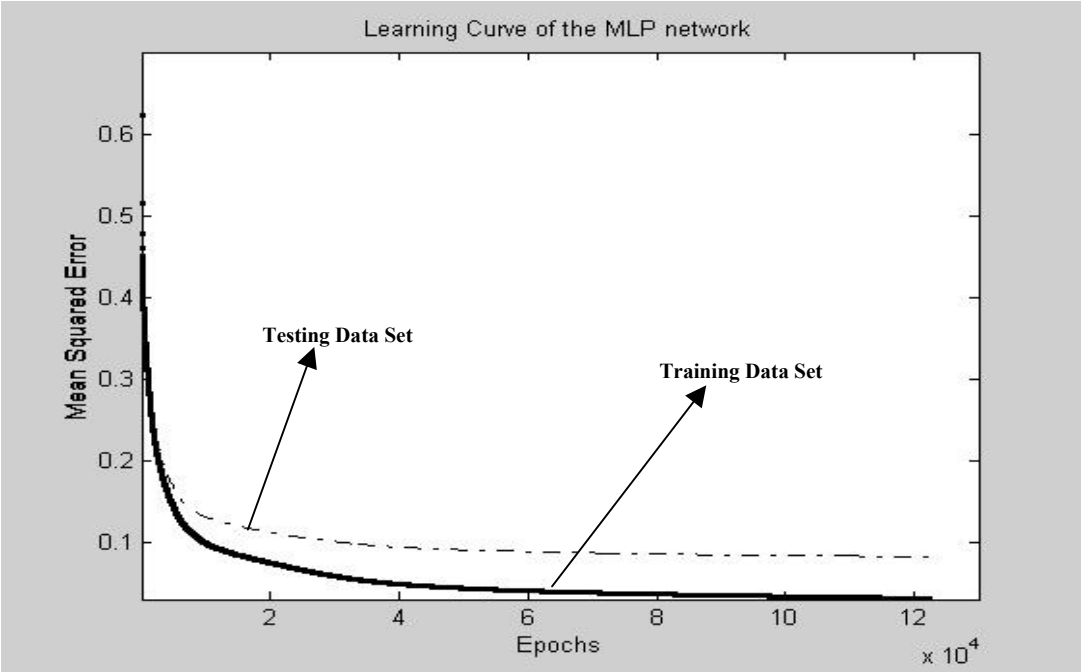
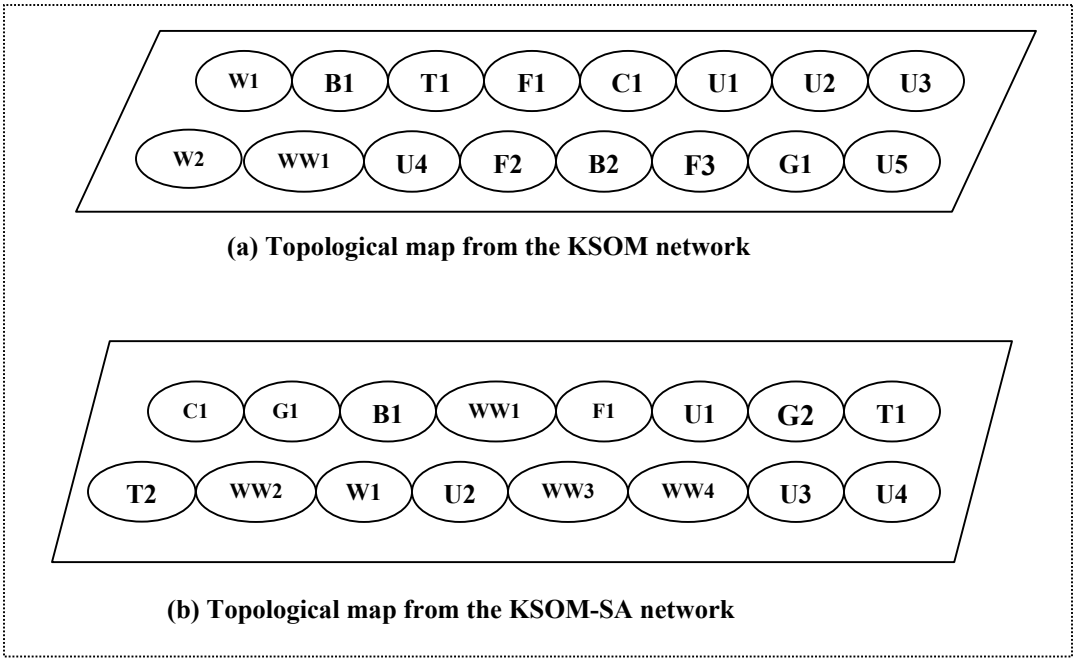


Figure 4. 11 Learning curve of the mean squared error (MSE) vs. training epochs from the MLP network



(a) Topological map from the KSOM network

(b) Topological map from the KSOM-SA network

Figure 4. 12 Topological maps from the KSOM and KSOM-SA networks where U: Urban, F: Forest, C: Crop field, G: Grassland, B: Bare soil, T: Transitional area, WW: Woody wetland, W: Water

To evaluate the three resultant classified maps, we assessed the classification accuracy based on a pixel-by-pixel comparison. By a stratified sampling method, 60 pixels per class were randomly selected from the MLP classified map and a total of 480 pixels were used for the accuracy assessments of the three network classifiers. By manually interpreting the CIR DOQQ and original TM image for each of the 480 pixels, three error matrices corresponding to each of the classified maps from Table 4.2 to Table 4.4 were generated. The Kappa statistic was calculated for each error matrix to test if any of the classifiers had significantly improved classification accuracy over the others. By studying these error matrices, we see that the MLP network has obtained the best classification accuracy 88.13% among the three classifiers, approximately 29.8% higher than the KSOM network and 25% higher than the KSOM-SA network. The Kappa statistics and analysis shown in Table 4.5 and Table 4.6 further confirm this.

Table 4. 2 Error matrix on the classified map from the MLP network

		Reference Data									
Classified Image		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Classified Totals**	**Users' Accuracy**
	1	**46**	1	1	1	11				60	76.7%
	2		**60**							60	100.0%
	3		2	**52**	6					60	86.7%
	4		2	4	**53**	1				60	88.3%
	5	2		1	1	**56**				60	93.3%
	6		4	2			**52**	2		60	86.7%
	7		12				3	**44**	1	60	73.3%
	8								**60**	60	100.0%
	Reference Totals	48	81	60	61	68	55	46	61	**480**	
	Producers' Accuracy	95.8%	74.1%	86.7%	86.9%	83.4%	94.6%	95.7%	98.3%		
		Overall Accuracy: 423/480 = 88.13%									

Table 4. 3 Error matrix on the classified map from the KSOM network

		Reference Data									
Classified Image		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Classified Totals**	**Users' Accuracy**
	1	**22**		1		10				33	66.7%
	2	8	**75**		1	1	13	30		128	58.6%
	3	12	5	**59**	60	27	9	1		173	34.1%%
	4	1				4				5	0.0%%
	5					**19**				19	100.0%%
	6	5	1			7	**33**	3		49	67.4%%
	7							**12**	1	13	92.3%%
	8								**60**	60	100.0%
	Reference Totals	48	81	60	61	68	55	46	61	**480**	
	Producers' Accuracy	45.8%	92.6%	98.3%	0.0%	27.9%	60.0%	26.1%	98.4%		
		Overall Accuracy: 280/480 = 58.33%									

Table 4. 4 Error matrix on the classified map from the KSOM-SA

		Reference Data									
Classified Image		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Classified Totals**	**Users' Accuracy**
	1	**13**				22				35	37.1%
	2	1	**56**			1				58	96.6%
	3	3		**32**	12	1				48	66.7%
	4	14	5	27	**48**	15	5	1		115	41.7%
	5					**12**				12	100.0%
	6	17	4	1	1	16	**42**	1		82	51.22%
	7		16			1	8	**40**	1	66	60.6%
	8							4	**60**	64	93.6%
	Reference Totals	48	81	60	61	68	55	46	61	**480**	
	Producers' Accuracy	27.1%	69.1%	53.3%	78.7%	17.7%	76.4%	87.0%	98.4%		
		Overall Accuracy: 303/480 = 63.13%									

Table 4. 5 Individual Kappa Analysis for the three network error matrices

	MLP	KSOM	KSOM-SA
KHAT	0.86	0.52	0.58
Kappa Variance	0.0003	0.0006	0.0006
Z-Value	51.32	20.70	23.48

Table 4. 6 Kappa analysis results for the comparisons of the three error matrices

	MLP	KSOM	KSOM-SA
MLP		11.44	9.55
KSOM			1.72
KSOM-SA			

Between the KSOM and KSOM-SA networks, the overall classification accuracy from the KSOM-SA network is 4.8% higher than that of the KSOM network. The Kappa analysis from Table 4.6 shows that at the 90% confidence level, the KSOM-SA network is moderately better than the KSOM network. We postulate that the improvement is due to the incorporation of SA into the standard KSOM network. We also observed a fact that both of the KSOM-based networks have low individual classification accuracies in those classes involving a lot of human activities, such as urban and grass land. We postulate that this is due to their unsupervised nature, having less human control on the assignment of pixels to classes. Therefore, in complex LU/LC mapping applications, we recommend the use of supervised MLP networks for image classification, with the assistance of unsupervised KSOM networks to analyze the inherent spectral relationships between classes.

4.5 Conclusions and future work

In this paper, we presented an automated two-module ANN classification system, consisting of an unsupervised KSOM network module and a supervised MLP neural network module. The MLP network module is trained using the BP algorithm. In the KSOM network module, two training sub-modules are provided including the standard KSOM training sub-module and the refined KSOM-SA training sub-module. To verify the operational suitability of the developed ANN system, we performed three network classifications of a selected Landsat TM image. Based on the experimental results and our analysis of these results, we summarize our classification results from this study as follows:

1) The supervised MLP network obtained superior classification results than the two unsupervised KSOM networks. We postulate that this is due to the unsupervised nature of the two KSOM networks and complex spectral conditions in some land use categories.
2) For the two unsupervised KSOM networks, the classification performance of the KSOM-SA network was shown to be moderately better than that of the KSOM network, indicating that the incorporation of SA could help improve the classification performance of the standard KSOM network in this classification application.
3) The KSOM and KSOM-SA networks were demonstrated to have topological preservation capability. In the resultant feature maps, clusters with similar spectral

signatures were located as neighboring nodes in the output layer. The feature maps from KSOM networks could provide useful information regarding the representation, the variability, and the similarity of spectral classes related to the desired land categories. This information may be used to assist in selecting training data for supervised classification.

Several important conclusions are drawn from this study:

- In this study, we have developed an automated ANN classification system and embedded it within the working environment of ERDAS IMAGINE. The system has been shown to be a fully automated and user-friendly classification system that can be easily operated by a remote sensing expert who is not a neural network expert.
- The ANN system has been shown to be well-suited for land cover mapping using remotely sensed data and could be especially robust and useful when the distribution of the input data is not normal. A normal distribution is required by many statistical classification approaches such as the maximum likelihood classification.
- This study provided an additional case study to verify the superior classification capability of ANN approaches using remotely sensed data. Based on the knowledge obtained from this case study, we recommend that in complex LU/LC mapping applications, supervised MLP networks are used to derive detailed and accurate image classification, and unsupervised KSOM networks are used to

assist in analyzing the inherent spectral characteristics between and within classes.

- The classification performance of neural network approaches is very sensitive to the selection of many operational parameters, such as the size and quality of training data set, network architecture, training parameters, and overfitting problems. This ANN system developed in this study was designed such as the selection of training parameters and network architecture by means of experimentation to be simple. Network input and output coding were done automatically in the pattern conversion sub-module. Furthermore, the overfitting problem has been effectively avoided using a cross-validation training method.

To further test the suitability of the developed classification systems in a variety of land cover mapping applications, it is necessary to apply them to other classification applications using a range of remotely sensed data sets at different mapping scales in the future study. Unlike traditional classification approaches, neural network approaches do not require the estimation of the distribution model in the input data. Because of this, we postulate that the ANN system is suited for multisource classification because there is no need to estimate statistical distribution model of various data sources. It is worthwhile to adapt the system for a data fusion application to fuse multisource remotely sensed images. In Chapter 5, we introduce a two-stage neural network based fusion system based on the ANN classification system using Landsat TM and multispectral Satellite Pour l'Observation de la Terre (SPOT) images.

References

Anderson, J. R., *et al.*, "A land use and land cover classification system for use with remotely sensed data," *Washington, DC: U.S. Geological Survey Professional Paper 964*, pp. 28, 1976.

Babu, G. P., "Self-organizing neural networks for spatial data," *Pattern Recognition Letters*, vol. 18, pp. 133 –142, 1997.

Baraldi, A. and F. Parmiggiani, "A neural network for unsupervised categorization of multivalued input patterns: an application to satellite image clustering," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 33, pp. 305 - 316, 1995.

Benediktsson, J. A., and J. R. Sveinsson, "Feature extraction for multisource data classification with artificial neural networks," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 727-740, 1997.

Benediktsson, J.A., P. H. Swain, and O. K. Ersoy, "Neural network approaches versus statistical methods in classification of multi-source remote sensing data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 28, pp. 540-551, 1990.

Bezdek, J. C., *et al.* "Fuzzy Kohonen clustering networks," *IEEE International Conference on Fuzzy Systems*, pp. 1035 – 1043, 1992.

Bischof, H., and A. Leonardis, "Finding optimal neural networks for land use classification," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 36, no.1, pp. 337 – 341, 1998.

Bischof, H., W. Schneider, and A. J. Pinz, " Multispectral classification of landsat images

using neural networks," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 30, no. 3, pp. 482 – 490, 1992.

Bishop, C. M, Radial basis functions. In *Neural networks for pattern recognition*. *Oxford, New York: Clarendon Press*, 1995.

Carpenter, G.A., M.N. Gjaja, S. Gopal, and C.E. Woodcock, "ART neural networks for remote sensing: vegetation classification from Landsat TM and terrain data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 35, no. 2, pp. 308-325, 1997.

Cerny, V., "Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm," *Journal of Optimization Theory and Applications*, vol. 45, pp. 45 –51, 1985.

Chen, Z., *et al.*, "Texture segmentation based on Wavelet and Kohonen network for remotely sensed images," *IEEE SMC'99 Conference Proceedings*, vol. 6, pp. 816 – 821, 1999.

Dai, X. L., and S. Khorram, "Data fusion using artificial neural networks: a case study on multitemporal change analysis," *Computers, Environments, and Urban Systems*, vol. 23, pp. 19 – 31, 1999.

Duda, R. O., and P. E. Hart. Pattern classification and scene analysis. *New York: Wiley-Interscience*. 1973.

Ediriwickrema, J., and S. Khorram, "Hierarchical maximum-likelihood classification for improved accuracies," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 810- 816, 1997.

ERDAS, ERDAS field guide. Atlanta, GA: ERDAS, Inc., 1994.

Foody, G. M., "The significance of border training patterns in classification by a feedforward neural network using back propagation learning," *International Journal of Remote Sensing*, vol. 20, no. 18, pp. 3549 – 3562, 1999.

Foody, G. M., and M. K. Arora, "An evaluation of some factors affecting the accuracy of classification by an artificial neural network," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 799-810, 1997.

Foody, G.M., M.B. McCulloch, and W.B. Yates, "The effect of training set size and composition on artificial neural network classification," *International Journal of Remote Sensing*, vol. 16, pp. 1707 – 1723, 1995.

Geman, S., and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Iintelligence*,, vol. 6, pp. 721 – 741, 1984.

Goncalves, M. L., *et al.*, "A neural architecture for the classification of remote sensing imagery with advanced learning algorithms," *Proceedings of IEEE Signal Processing Society Workshop*, pp. 577 – 586, 1998.

Haack, B.N., *et al.*, "Assessment of Landsat MSS and TM data for urban and near-urban landcover digital classification", *Remote Sensing of Environment*, vol. 21, no. 2, pp.201 – 213, 1987.

Heermann, P. D., and N. Khazenie, "Classification of multispectral remote sensing data using a back-propagation neural network," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 30, pp.81-88, 1992.

Jenson, J. R., Introductory digital image processing. *Englewood Cliffs, NJ: Pretentice Hall,* 1996.

Kanellopoulos, I., and G.G. Wilkinson, "Strategies and best practice for neural network image classification," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 711 – 725, 1997.

Khorram, S., *et al.*, "Comparison of Landsat MSS and TM data for urban land-use classification", *IEEE Transactions on Geoscience and Remote Sensing*, vol.25, no. 2, pp.238 – 243, 1987.

Kirkpatrick, S., C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671-688, 1983.

Klein, R. W., and R. C. Dubes, "Experiments in projection and clustering by annealing," *Pattern Recognition*, vol. 22, no. 2, pp. 213-220, 1989.

Kohonen, T., "Self-orgazning formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, pp. 56 – 69, 1982.

Laarhoven, P.J.M., Theoretical and computational aspects of simulated annealing. *Amsterdam, Netherlands: Center for mathematics and computer science,* 1988.

Lee, J., *et al.*, "A neural network approach to cloud classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, pp.846 – 855, 1990.

Lippmann, R. P, "Pattern recognition using neural networks," *IEEE Communications Magazine*, pp. 47 – 64, 1989.

Lippmann, R. P., "An introduction to computing with neural networks," *IEEE ASSP Magazine*, vol. 4, no. 2, pp. 4 – 22, 1987.

Paola, J. D., and R. A. Schowengerdt, "The effect of neural-network structure on a multispectral land-use/land-cover classification," *Photogrametric Engineering and Remote Sensing*, vol. 63, pp. 535 – 544, 1997.

Principe, J.C., N. R. Euliano, and W. C. Lefebvre. Neural and adaptive systems: fundamentals through simulations. *John Wiley & Sons, Inc.*, pp. 100 – 222, 1999.

Richards, J. A., Remote sensing digital image analysis. *Springer-Verlag*, 1986.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, Parallel Distributed Processing. *Cambridge, MA: MIT Press*, 1986.

Schowengerdt, R. A., Techniques for image processing and classification in remote Sensing. *New York: Academic Press*, pp. 129 – 214,1983.

Serpico, S. B., and F. Roli, "Classification of multisensor remote-sensing images by structured neural networks," IEEE Trans. Geosci. And Rem. Sens., vol. 33, no. 3, pp. 562-578, 1995.

Solberg, A.H. S., *et al.*, "Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images*", IEEE Transactions on Geoscience* and Remote Sensing, vol. 32, no. 4, pp.768 – 778, 1994.

Swain, P. H., and S. M. Davis, Remote sensing: the quantitative approach. *New York: McGraw-Hill*, pp. 136 – 188,1978.

Thomson, A.G., R.M. Fuller, and J.A. Eastwoods, "Supervised versus unsupervised methods for classification of coasts and river corridors from airborne remote sensing," *International Journal of Remote Sensing*, vol. 18, pp. 3423 – 3431, 1998.

Ulaby, F.T., *et al.*, "Crop classification using airborne RADAR and Landsat data", I*EEE*

Transactions on Geoscience and Remote Sensing, vol. 20, pp.518 – 528, 1982.

Zhuang, X., B. A. Engel, M. F. Baumgardner, and P. H. Swain, "Improving classification of crop residues using digital land ownership data and Landsat TM imagery," *Photogrammetic Engineering & Remote Sensing*, vol. 57, no. 11, pp. 1487 – 1492, 1991.

5. #Paper 3 Development of a two-stage neural network-based land cover mapping system using multisource remotely sensed data

ABSTRACT --

Inclusion of complementary information from additional data sources, or "data fusion", can improve classification accuracy. Artificial Neural Network (ANN) approaches have been shown to have great potential to fuse multiple source data sets. However, there is a potential problem with ANN networks. Their performance is very sensitive to the quality and quantity of collected training data sets. Training data sets are traditionally collected by visual interpretation of the scene and manually selecting the sits for each class. This manual collecting process in a data fusion application may become problematic due to the complex nature of multiple source fusion. The increase of input dimensions in a fusion application requires increasing the amount of training data considerably. The typical spectral signatures of classes are compounded by factors such as class changes between fused images and misregistration error, which make visual interpretation of the scene problematic. To improve the quality of training sets, it is essential to develop techniques to assist the manual training data collection process. In this study, we have developed a two-stage neural network based multisource classification of Landsat Thematic Mapper (TM) and multispectral Satellite Pour l'Observation de la Terre (SPOT) images to derive detailed Land-Use/Land-Cover (LU/LC) categories. In the first stage of this classification scheme, the Kohonen's Self-

Organizing Mapping (KSOM) neural network was adapted as an Automated Data Selector (ADS) to replace the manual process of training data selection to help improve the quality of selected training sets. In the second stage, the Multilayer Perceptron (MLP) network was trained using the automatically extracted training data sets and then used to perform the final classification.

Two hypotheses were tested in this study: 1) the ADS adapted from the KSOM network could provide adequate and reliable training data sets and thus improve the classification performance of the MLP classification, and 2) the fusion of the two Landsat TM and SPOT images using the ANN approaches could increase the classification accuracy of the derived land categories. Experimental results have shown that the ADS adapted from the KSOM network effectively improved the quality of training data and significantly increased the overall classification accuracy and efficiency. We conclude from this study that the fusion of multiple remotely sensed data can have very good classification performance if appropriate training data sets are collected. This two-stage classification scheme proves to be a viable classification scheme design especially in classification applications where it is difficult to collect sufficient and reliable training data sets.

5.1 Introduction

During recent years advances in space and computer technologies have made accessible large amounts and a variety of remotely sensed data about the Earth and its ecological environment. This offers great potential to fuse various data sources to address

a variety of environmental problems. The fusion of complementary data can help reduce the uncertainty and imprecision of the single source data and improve the classification accuracy (Hegarat-Mascle *et al.* 1997; Luo & Kay 1989). To fully take advantage of these multiple source data, it is mandatory to develop effective data fusion techniques to combine complimentary characteristics of various source data.

Artificial Neural networks (ANNs) are fundamentally network systems simulating the working process in human mind, in which a large number of small and interconnected processors or neurons work together to solve difficult classification and optimization problems. ANN approaches have a significant advantage over statistical classification methods in that they are non-parametric and require little *a priori* knowledge of the distribution model of input data (Benediktsson *et al.* 1990). The distribution-free attribute of ANN approaches makes them attractive approaches for multisource classification because "statistical distributions of remotely sensed data may vary with different classes as well as with data sources" (Paola & Schowengerdt 1997). ANN approaches have been heavily studied for the classification of multiple data sets (Dai & Khorram 1999; Wan & Fraser 1999; Bruzzone *et al.* 1997; Serpico & Roli 1995; Zhuang *et al.* 1991; Benediktsson *et al.* 1990). These studies strongly suggest that ANNs are able to fuse a variety of data sets including remotely sensed multispectral imagery and ancillary GIS data, especially when accurate estimation of statistical distribution of multiple source data is difficult or impossible.

5.1.1 Problem statement

Bendiktsson *et al.* (1990) have investigated multisource classification using neural network approaches. They demonstrate that a three-layer Multilayer Perceptron (MLP) network using the Backpropagation (BP) training algorithm is sufficient for complex image classification problems. However, many other researchers (Zhou 1999; Foody *et al.* 1995; Zhuang *et al.* 1991) point out that there is a potential problem with supervised MLP networks. Their classification performance is very sensitive to the quality and quantity of training data selected for each class. The influence of training data sets is very likely to be extended to the network generalization stage and affect the final classification accuracy. Therefore, to obtain a satisfactory classification, the training data must provide a representative description of each category as complete as possible and the amount of training data for each category should be sufficient (Bischof *et al.* 1992). Principe *et al.* (1999) suggest that the size of the training set increases proportionally along with the increases of the spectral variability of desired classes, the number of associated weights, and the desired classification accuracy.

For supervised remote sensing applications, training data sets are traditionally collected by an analyst to manually identify a number of representative sites for each category based on their typical spectral and spatial signatures. Using the manual collecting technique, the quality of training data sets is limited by: 1) the complexity of the application (e.g., the desired class level and landscape conditions on the ground), and 2) the quality of acquired images (e.g., spectral and spatial resolutions). Other disadvantages of the manual collecting process are: 1) laborious human work is involved,

2) uncertainty and inconsistency may be introduced from the manual process, and 3) the quality of the collected training data sets highly depends on an expert analyst with excellent image interpretation skills.

In a data fusion application, the manual collecting process often becomes more challenging. To fuse multiple remotely sensed images, we co-register and stack multiple source images to create a fused image that consists of spectral bands from each of the remotely sensed images. The stacking of spectral bands from multiple source images inevitably results in the increase of input dimensions in a fusion application. The increase of input dimensions certainly requires considerably increasing the amount of training data for each land category. However, landscape conditions can be so complicated as to make it impossible to collect large amounts of appropriate training data for each category by a manual collecting process. In addition to the required increase of training data sets, the co-registration and stacking of multiple image data sets may change the spectral characteristics in the fused image. The typical spectral and spatial signatures of some land categories may be changed if there are significant class changes occurring during the time interval between the acquired multiple images and misregistration errors. All these adverse factors may make visual interpretation of the scene problematic. Thus, it is critical to develop techniques to assist the manual training data collection. Ideally, we require an Automated Data Selector (ADS) that is able to automatically extract sufficient and reliable training data sets.

5.1.2 A proposed two-stage neural network based multisource classification

The aforementioned training data collection problem has been recognized in many ANN fusion studies (Zhou 1999; Foody *et al.* 1995; Zhuang *et al.* 1991; Benediktsson *et al.* 1990). Hybrid classification approaches have been investigated in recent years to alleviate the burden of manual training data collection (Goncalves *et al.* 1998; Nogami *et al.* 1997; Yoshida and Omatu 1994). The common practice in these studies is to use the clustering capability of unsupervised KSOM networks to assist the manual training data collection. However, no automated training data selection is observed from any of them. In this study, we propose to develop a two-stage neural network based multisource classification, in which the training data sets are collected by an ADS.

In Chapter 4, we have developed an automated two-module ANN classification system. The availability of the unsupervised Kohonen's Self-Organizing Mapping (KSOM) and supervised MLP networks in the ANN system offers great flexibility to develop hybrid classification schemes needed for complex classification applications. From the study described in Chapter 4, we have shown that the unsupervised KSOM network module can produce a feature map consisting of a number of spectral clusters that may not always be visible to the naked eye. Based on this, we adapted the unsupervised KSOM network to play the role of ADS in this study. Using the ADS adapted from the KSOM network and the ANN classification system, a two-stage neural network based multisource classification of Landsat Thematic Mapper (TM) and multispectral SPOT images was performed. In the first stage of this multisource classification, the ADS replaced the traditional manual training data collection to select

training data sets for supervised MLP classification. In the second stage, the MLP network was trained using the automatically selected training data sets and then used to perform the final classification. *Based on this two-stage neural network based multisource classification, two hypotheses were tested: 1) the ADS adapted from the KSOM network could provide adequate and reliable training data sets and thus improve the classification performance of the MLP classification, and 2) the fusion of the two Landsat TM and SPOT images using the ANN approaches could increase the classification accuracy of the derived land categories.*

The two neural network models including KSOM and MLP network models are briefly introduced in section 5.2 and followed by an elaborate description of the associated procedures to adapt the KSOM network to the ADS. The two-stage multisource classification of Landsat TM and SPOT images is presented in section 5.3 where related implementing details are provided and the classification results are discussed. General conclusions from the study and future work are summarized and given in section 5.4.

5.2 Neural Network Methodologies and the Automated Data Selector

KSOM and MLP networks are two of the most popular network models in remote sensing classification applications (Bezdek *et al.* 1992; Benediktsson *et al.* 1990). In sections 5.2.1 and 5.2.2, we briefly introduce these two network models and their training algorithms. In section 5.2.3, the ADS adapted from the KSOM network is presented.

5.2.1 Kohonen's Self-Organizing Mapping (KSOM) neural network model

KSOM network, proposed by Kohonen (1982), is a relatively simple clustering tool to group dimensionally complex data into clusters by its competitive learning. KSOM network is basically a competitive two-layer network, consisting of an input layer with a number of input nodes, an output or competitive layer with a number of competitive nodes arranged by rows and columns, and a set of connection weights linking input nodes with each competitive node. The number of input nodes normally equals the number of input dimensions while the number of competitive nodes is contingent upon the number of spectrally separable clusters. When an input pattern is presented to the KSOM network, the node in the competitive layer with the closest weight vector to the input vector in terms of Euclidean distance is declared as the "winning" node. During the training procedure only this winning node and the nodes in the restricted neighborhood update their associated weight vectors. Due to this spatial neighborhood property, KSOM network can preserve the topological relationships that exist within the original input data, by which the clusters with similar spectral signatures in image classification will be assigned to the neighboring nodes in the competitive layer. The goal of the KSOM network training is described as minimization of the sum of squared Euclidean distances of the input vector from the associated weight vector of its winning node. The mean square error J (MSE) can be used as a classification performance criterion given by

$$J = \frac{1}{2N}\sum_{i=1}^{N} D_i^2 \qquad \text{--- (5.1)}$$

where N is the number of training patterns. D_i^2 denotes the Euclidean distance between the network input vector and the associated weight vector of its winning node. Detailed implementation of KSOM network can be found in Chapter 4.

5.2.2 Multilayer Perceptron (MLP) neural network model

A typical MLP network consists of one input layer, one or more hidden layers and one output layer. All nodes in different layers are connected by associated weights. Generally, the number of input nodes equals the input dimension and the number of output nodes equals to the number of desired classes, i.e., each input node in the input layer corresponds to an spectral band and each output node in the output layer to one of the desired classes. MLP networks are usually implemented using the BP training algorithm (Rumelhart & McClelland 1986). The training process requires a set of training patterns with inputs and corresponding desired outputs. For each input pattern presented to the MLP network, the current network output of the input pattern is calculated using the current weights. In the next step, the error or difference between the network output and the desired output will be backpropagated to adjust the weights between layers to move the network output closer to the desired output. By iteratively presenting each training pattern to the MLP network, the inherent input-output relationships can be approximated with a general objective of minimizing the total error between the network outputs and desired outputs for the entire set of training patterns. This minimization objective can be denoted by the mean square error J (MSE) given by

$$J = \frac{1}{2N}\sum_{i=1}^{N}\varepsilon_i^2 \qquad \text{--- (5.2)}$$

where N is the number of training patterns. ε_i^2 denotes the Euclidean distance between the network output of the input pattern and its desired output. Once this training process is completed, the MLP network is used to classify new patterns beyond the training patterns.

5.2.3 The adapted automated data selector

In the automated ANN classification system described in Chapter 4, each classification module is composed of three sub-modules: the pattern conversion sub-module, the network training sub-module, and the network generalization sub-module. The pattern conversion sub-module is used to extract a certain number of training and testing patterns for each class. For the supervised MLP classification, the class label for each pattern must be provided. The network training sub-module is to perform the network training. Once a well-trained network is generated from the network training process, the network generalization sub-module is run to generalize the entire image and produce a land cover map using this well-trained network.

The development of the ADS using the ANN system was mainly performed on the KSOM network generalization sub-module. The original KSOM network generalization sub-module shown in Figure 5.1 requires the user to provide five parameters: 1) the input image file to be classified, 2) a trained KSOM network, 3) the number of desired classes, 4) a file name for the resultant classified map, and 5) a file

name for the output pattern file. The output pattern file produced from the KSOM network generalization sub-module is to provide training pattern files for supervised MLP network. To adapt the unsupervised KSOM network to play the ADS, we extended the functions of the KSOM generalization sub-module that is shown in Figure 5.2. Three additional parameters are required in Figure 5.2: 1) the number of training pixels per class, 2) the generation probability, and 3) the distance threshold.

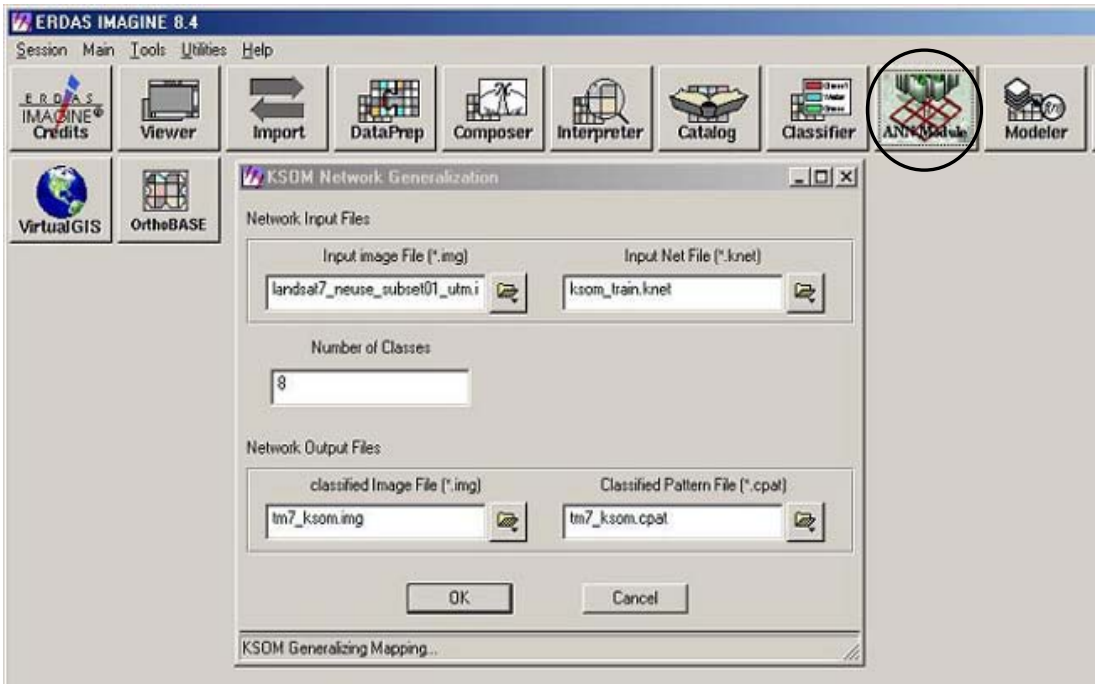


Figure 5. 1 The original KSOM Generalization Sub-module Interface of the ANN System

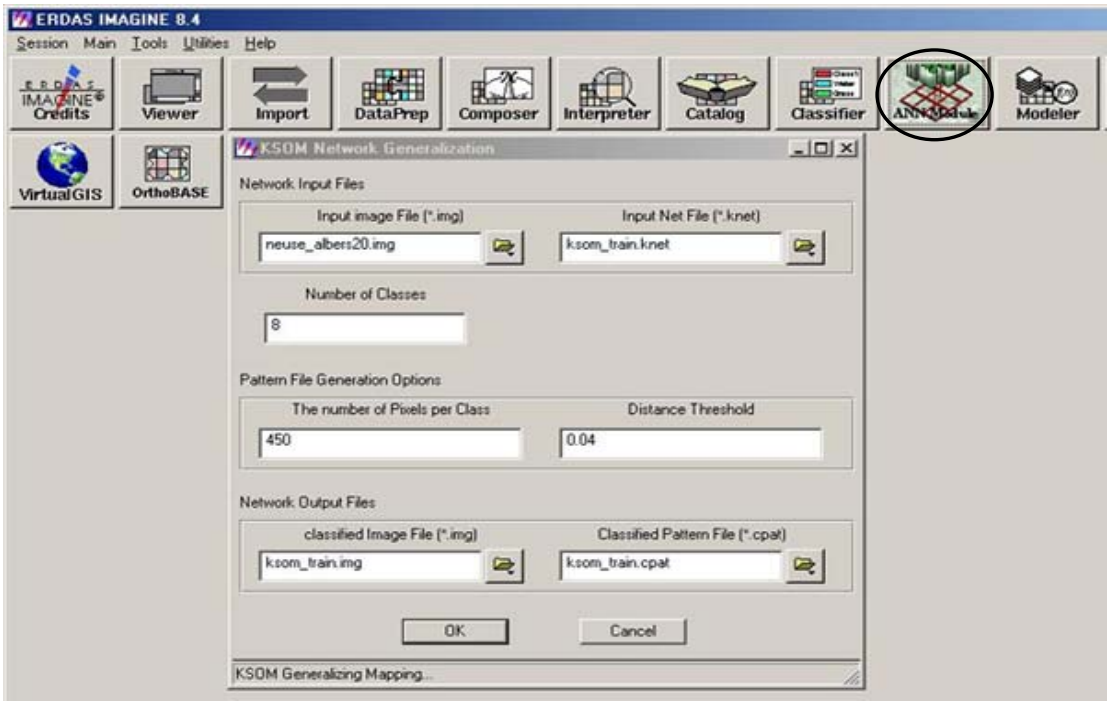


Figure 5. 2 The Modified Interface of the KSOM Generalization Sub-module

The implementing procedures of the ADS are described as follows:

1) Initially, set the number of training set for all classes to zero and start to generalize the pixels on the image one by one.
2) For each pixel on the image, its class label, for instance, *K*, is calculated. If the size of the training set in the class *K* is lesser than the defined number of training pixel per class, we go to the next step 3). Otherwise, go to the step 5).
3) For each pixel, a uniform probability is drawn. If the drawn uniform probability is larger than the defined generation probability, the pixel is considered as a potential training candidate pixel and we go to the next step 4) to further examine this pixel. Otherwise, go back to the step 5).
4) For the potential training candidate pixel, the Euclidean distance of the pixel from the associated weight vector of its winning node is calculated. If its Euclidean distance is lower than the defined distance threshold, this pixel is selected and added into its training set. The number of the training set for the class *K* is incremented by 1.
5) Repeat the steps from 2) to 4) until all the pixels on the image have been examined or the certain number of training pixels for all classes is obtained.

The number of training pixels per class determines the size of the training sets. The generation probability determines the distribution of the selected training pixels on the entire image in a way that that the larger the generation probability, the more scattered the selected training data distribute on the original image. The distance

threshold is used to determine the quality of the selected training pixels based on the assumption that the larger the distance threshold, the more likely boundary pixels are to be chosen. The MSE from each run of the KSOM generalization can be obtained using Equation 5.1 to provide assistance in determining an optimal distance threshold.

5.3 A Two-stage Neural Network based Multisource Classification

Using the ANN system and the ADS adapted from the KSOM network, we performed a two-stage classification application using Landsat Thematic Mapper (TM) and multispectral SPOT images. The flowchart of the two-stage ANN classification is shown in Figure 5.3. In section 5.3.1, the selected study area is introduced and the desired categories are defined. In section 5.3.2, image pre-processing procedures are described. The implementing details of the two-stage neural network based multisource classification are presented in section 5.3.3. The resulting classifications are compared and discussed in section 5.3.4.

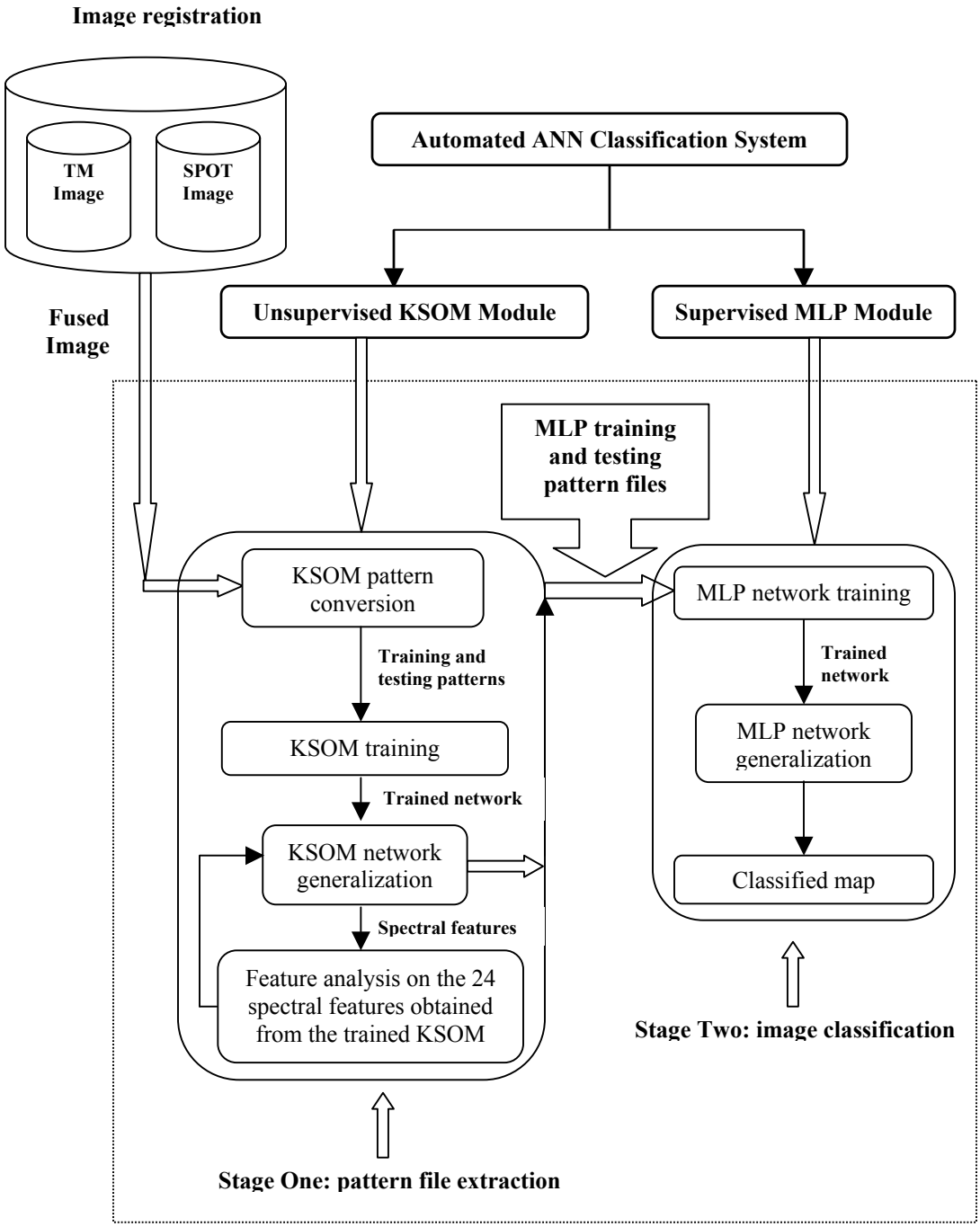


Figure 5. 3 The flowchart of the two-stage ANN classification

5.3.1 Study area and class definition

The lower Neuse Basin area in the North Carolina State was selected to perform the two-stage fusion application in this study. The study area covers a region consisting of four counties including Craven, Jones, Pamlico, and onslow. The study area is chosen because there are a variety of urban, agricultural, hydrologic thematic features in the coastal plain. In this two-stage fusion study, two remotely sensed images were acquired: 1) a Landsat TM image of the study area acquired in September 1999, and 2) a multispectral SPOT image acquired in September 1998. The Landsat TM image has a spatial resolution of 30m and spectral resolution of 6 optical bands and one thermal band, while the multispectral SPOT has a higher spatial resolution of 20m with lower spectral resolution of 4 optical bands. Via the fusion of these two images, the differentiating capability of LU/LC classes is enhanced because both the spatial and spectral resolutions in the fused image are improved as compared with the single TM and SPOT image.

By visually interpreting both the original images and assisted by the high spatial resolution Digital Color Infrared (CIR) Digital Orthophoto Quarter Quad (DOQQ) data acquired between January and March 1998, eight land categories were defined in Table 5.1. The categories were defined to reflect the classification capability improvement of certain categories in this fusion application. Based on the category definitions, the complementary temporal information from the two image data sets is properly combined to improve the classification accuracy of the derived categories if the selected training data sets can provide such information. For example, the fallow areas on either of the

images are assigned to the agricultural category instead of the bare category. The sites in the earliest stage of crop growth on either of the images are also assigned to the agricultural category instead of the transitional category.

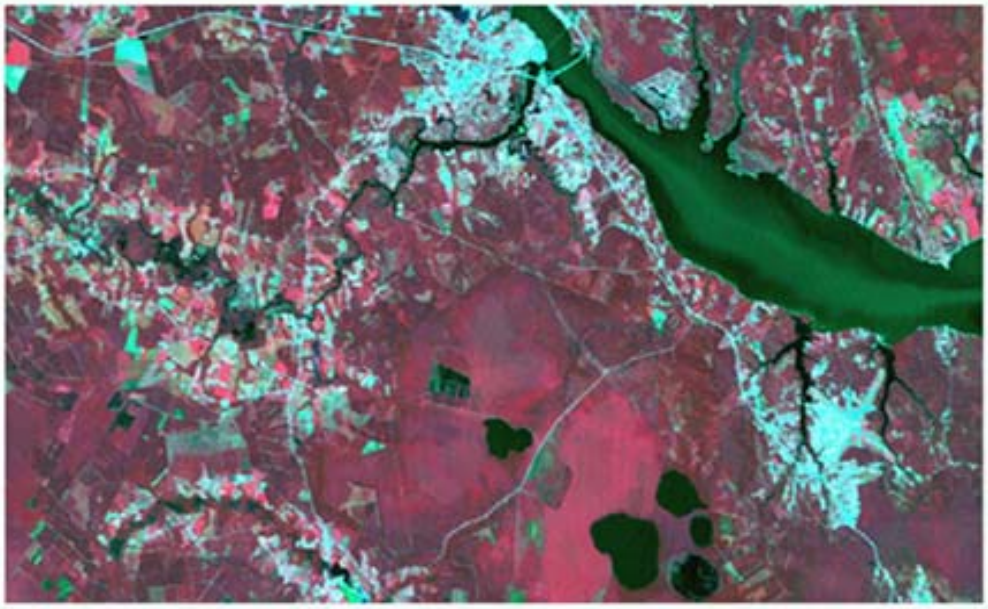
Table 5. 1 Classification scheme and category definition

Class Number	Class Name	Class Definition
1	Urban	Commercial/Industrial/Residential/transportation
2	Forest	Natural Forested Upland including evergreen, deciduous, and mixed forests
3	Agricultural field	Planted crop fields, fallow agricultural fields, or hay/ pasture.
4	Other Grassland	Vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes, such as parks and gold courses.
5	Bare Ground	Bare construction sites, rock, and sand. Must be bare ground on both images.
6	Transitional Area	Areas dynamically changing from one land cover to another, mainly consisting of freshly cleared forestlands as well as areas in the earliest stages of forest or crop growth. Must be within the transitional stage on both images.
7	Woody Wetland	Areas of forested or shrubland vegetation where soil or substrate is periodically saturated with or covered with water.
8	Water	All areas of open water

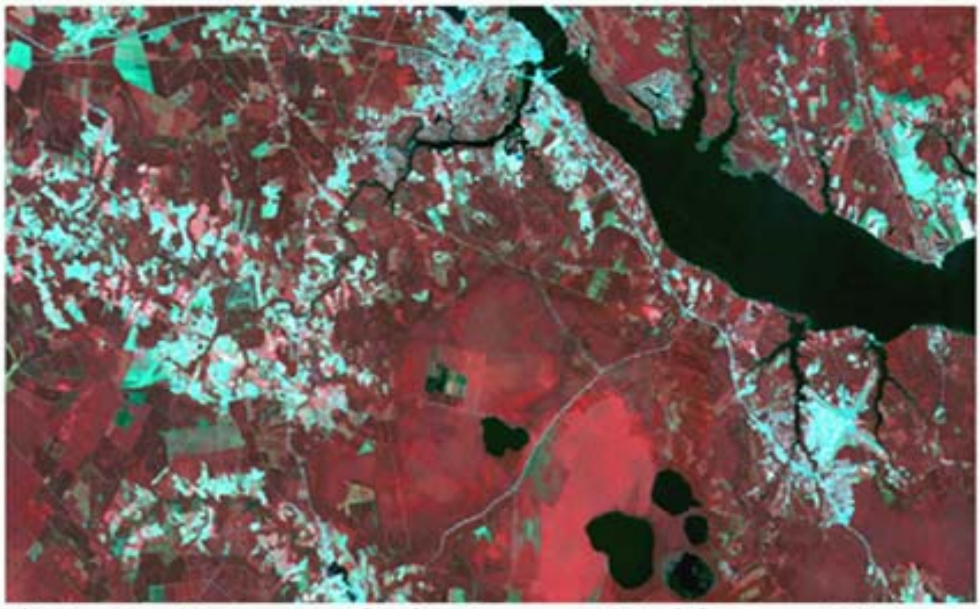
5.3.2 Image preprocessing

All the image pre-processing procedures were done using ERDAS IMAGINE image geometric correction and layer stacking tools (ERDAS 1994). Before classification, the Landsat TM image was co-registered with the SPOT image to the same common reference system. The co-registered Landsat TM image is shown in Figure 5.4(a) and the multispectral SPOT image is shown in Figure 5.4(b). After the co-registration, the two images were stacked together to produce a fused image with the spectral resolution of eight optical bands and the spatial resolution of 20m. In the fused

image, we included at least one spectral band from each of the available electromagnetic spectral bands in both the images. Among the eight spectral bands used in this fusion application, four spectral bands were from the Landsat TM image and other four spectral bands from the SPOT image. The four spectral bands from the Landsat TM image were TM_band 1 (blue band), TM_band 4 (near-infrared band), TM_Band5 (first middle-infrared band), and TM_Band7 (second middle-infrared band). The four spectral bands from the SPOT image were SPOT_Band1 (green band), SPOT_Band2 (red band), SPOT_Band3 (near-infrared band), and SPOT_Band4 (middle-infrared band). Inclusion of four infrared bands is believed to able to enhance the capability of differentiating various vegetation covers.



(a) The Landsat TM image with 20m spatial resolution and Albers Conical Equal Area projection.



(b) The Multipsectral SPOT image with 20m spatial resolution and Albers Conical Equal Area projection.

Figure 5. 4 Two multispectral images that are used in the fusion application

5.3.4 Two-stage multisource classification

In the first stage of the two-stage multisource classification, we used the KSOM network to classify the fused image into a number of spectral clusters and then matched them to one of the eight categories we had defined. To train a KSOM network, training and testing data set are needed. Due to the unsupervised nature of KSOM networks, the requirements of the training data sets for the KSOM network training are much less restrictive than the supervised MLP network. In the KSOM training, a number of image subsets were first extracted from the fused image using ERDAS IMAGINE image subset tools and then input into the pattern generation sub-module of KSOM module. Two patterns files were produced including a training pattern file with 3600 training patterns and a testing pattern file with 480 testing patterns. These two pattern files were used to train the KSOM network.

The optimal KSOM network training was obtained when 24 output nodes were arranged as 2 rows and 12 columns. The KSOM training parameters were set as follows: i) the training iterations (5000), ii) the initial learning rate (0.2), iii) the final learning rate (0.01), iv) the initial neighborhood size (12), and v) the neighborhood decrement interval iterations (250). The well-trained KSOM network from the optimal KSOM training was used to generalize the entire image into a classified map with 24 spectral classes. By visual interpretation of these 24 classes, we matched each of them to one of the defined eight land categories. Once the spectral feature analysis using the KSOM network was

done, the ADS was applied to generate training and testing patterns files for the supervised MLP classification in the second stage.

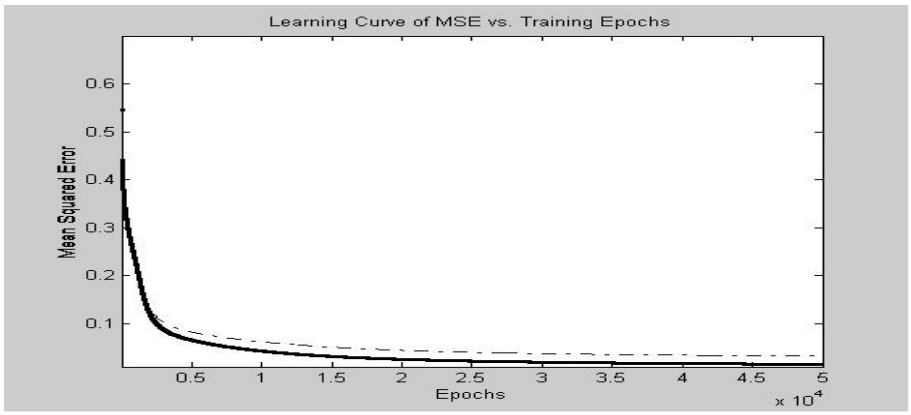
In this study, two training pattern files were created from the ADS: 1) the first training pattern file, denoted as train1_0.04_0.6, was created with the number of pixel per class (450), the distance threshold (0.04), and the generation probability (0.6), and 2) the second training pattern file, denoted as train2_0.09_0.95, was created with the number of pixel per class (450), the distance threshold (0.09), and the generation probability (0.95). Similarly, a testing pattern file composing of a total 480 pixels with 60 pixels per class was created with the distance threshold (0.09) and the generation probability (0.95). We suggest using a high generation probability to generate testing patterns because we postulate that the testing patterns that are distributed widely on the fused image can provide a better representation of categories.

In the second stage of the two-stage multisource classification, we trained the MLP network using the automatically selected training and testing patterns from the first stage. By experimentation, the MLP training using the training pattern file train1_0.04_0.6 was found to have better generalization performance than that using the training pattern file train2_0.09_0.95. The optimal MLP network training using the training pattern file train1_0.04_0.6 was obtained when 20 hidden nodes were used. The MLP training parameters were set as follows: i) the training epochs (50,000), ii) the initial learning rate (1.5), iii) the final learning rate (0.01), and iv) the momentum (0.8). The learning curve of the MSE vs. training epoch from this MLP training is plotted in

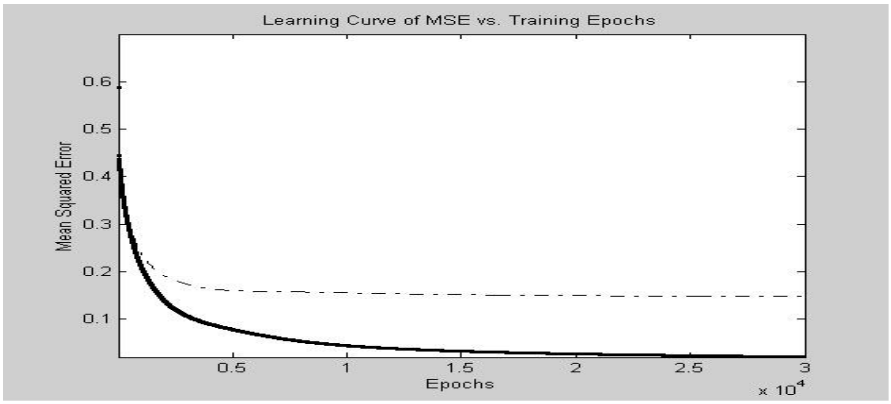
Figure 5.5(a). The classified map from the two-stage classification is shown in Figure 5.6(a).

In addition to the two-stage multisource classification, we have also performed two one-stage MLP classifications: 1) a one-stage MLP multisource classification, and 2) a one-stage MLP-TM classification. In the one-stage MLP multisource classification, we manually collected 3600 training pixels with 450 pixels per class and 480 testing pixels with 60 pixels per class from the fused image. The training and testing patterns manually collected were directly applied to train a supervised MLP network. The well-trained MLP network was then used to classify the fused image of the Landsat TM and SPOT images into a classified map with the eight categories. The learning curve of the MSE vs. training epoch from this MLP training is plotted in Figure 5.5(b). The resulting classified map from the one-stage MLP multisource classification is shown in Figure 5.6(b). In the one-stage MLP-TM classification, we manually collected 3360 training pixels with 420 training pixels per class and 360 testing pixels with 45 testing pixels per class from the single source TM image instead of the fused image. The training and testing patterns manually collected from the single source TM image were applied to train a new supervised MLP network. The well-trained MLP network was then used to classify the Landsat TM image into a classified map with the eight categories. The resulting learning curve from the one-stage MLP-TM classification is plotted in Figure 5.5(c) and the resulted classified map is shown in Figure 5.6(c). The comparison between the one-stage MLP multisource classification and the two-stage multisource classification is used to test the effectiveness of the ADS. By comparing the one-stage MLP multisource

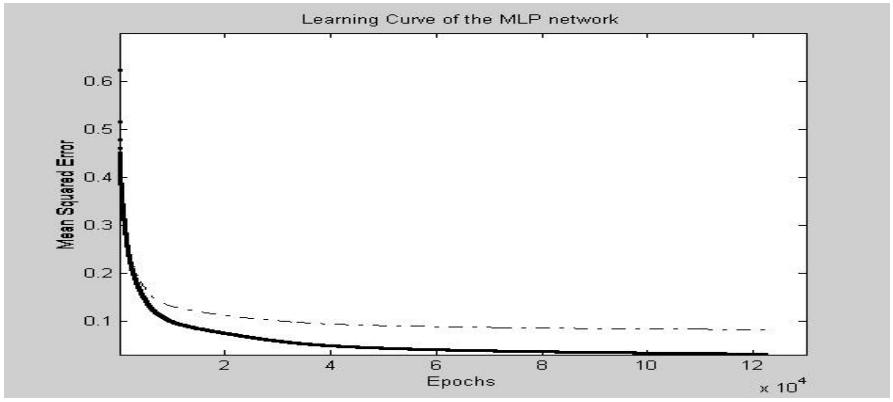
classification with the one-stage MLP-TM classification, the effects of the fusion of the two images are tested.



(a) The learning curve of the MLP training in the two-stage multisource classification using the training and testing data sets generated by the ADS



(b) The learning curve of the MLP training in the one-stage MLP multisource classification using the manually selected training and testing data sets



(c) The learning curve of the MLP training in the one-stage MLP-TM classification using the manually selected training and testing data sets

Figure 5. 5 Learning curves of the MLP training

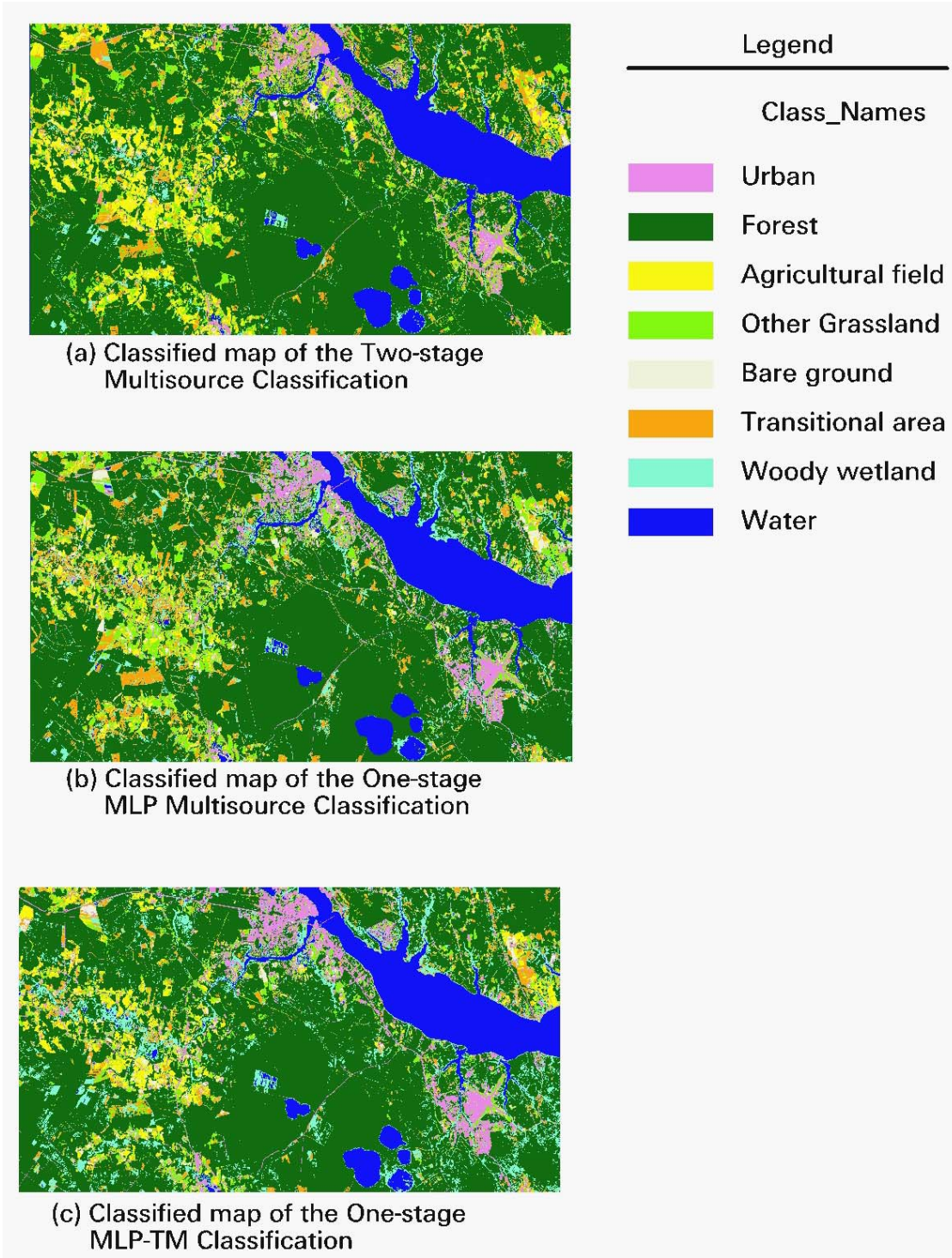


Figure 5. 6 Classified maps of the study area

5.3.4 Classification results and discussion

To quantitatively evaluate the classified maps, we used a stratified sampling method to select a total 480 reference points with 60 points per class from Figure 5.6(a). The 480 reference points were used to assess the classification accuracy for the three classified maps shown in Figure 5.6(a-c). By visual interpreting of the 480 points, three error matrices for the three classified maps were generated and shown from Table 5.2 to Table 5.4. From the three tables, we observe that the two-stage multisource classification has the highest accuracy (81.88%), approximately 18.8% higher than the one-stage MLP multisource classification and 10.2% higher than the one-stage MLP-TM classification. The Kappa statistic and analysis shown in Table 5.5 and Table 5.6 demonstrate that the two-stage multisource classification is significantly superior to the two one-stage MLP classifications. By comparing Table 5.2 and Table 5.4, we see that the two-stage multisource classification has obtained significant overall and individual classification accuracy increases, especially in the two categories of bare ground and transitional area. This is because in the one-stage MLP-TM classification, the single TM image is unable to differentiate fallow sites from the bare category and separate agricultural sites in the early stage of crop growth from the transitional category. In the two-stage multisource classification, the complementary temporal information from both the TM and SPOT images is combined in the fused image and thus the differentiating capabilities in these categories are significantly improved.

Table 5. 2 Error matrix on the classified map of the two-stage multisource classification

		Reference Data									
Classified Image		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Classified Totals**	**Users' Accuracy**
	1	**51**		2	1	4				58	87.9%
	2	1	**72**		1		1	3		78	92.3%
	3	3	2	**65**	8	2	1			81	80.3%
	4	3	7	6	**39**		5			60	65.0%
	5			7	1	**22**				30	73.3%
	6	4	2	6	8	1	**43**	2		67	64.2%
	7	1	2				1	**40**	1	45	88.9%
	8								**61**	61	100.0%
	Reference Totals	63	85	86	58	29	51	45	62	**480**	
	Producers' Accuracy	81.0%	84.7%	75.6%	67.2%	75.9%	84.3%	88.9%	98.4%		
		Overall Accuracy: 393/480 = 81.88%									

Table 5. 3 Error matrix on the classified map of the one-stage MLP multisource classification

		Reference Data									
Classified Image		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Classified Totals**	**Users' Accuracy**
	1	**52**		4	8	6				70	74.3%
	2		**66**		1		2	6		75	88.0%
	3		2	**36**	17	4				59	61.0%
	4	2	1	32	**15**	1	3			54	27.8%
	5	7		9	1	**17**	7			41	41.5%
	6		8	5	14		**23**	3		53	43.4%
	7	2	8		2		16	**33**	2	63	52.4%
	8					1		3	**61**	65	93.9%
	Reference Totals	63	85	86	58	29	51	45	62	**480**	
	Producers' Accuracy	82.5%	77.6%	41.9%	25.9%	58.6%	45.1%	73.3%	98.4%		
		Overall Accuracy: 303/480 = 63.13%									

Table 5. 4 Error matrix on the classified map of the one-stage MLP-TM classification

		Reference Data									
Classified Image		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Classified Totals**	**Users' Accuracy**
	1	**58**	1	5	5	13	3			85	68.2%
	2	1	**68**	2	6		7			84	81.0%
	3			**43**	6		1			50	86.0%
	4	1	1	11	**30**	1	7			51	58.2%
	5	2		4	1	**15**	3			26	57.7%
	6	1	2	11	10	1	**26**	2		53	49.1%
	7		14	1		1	7	**43**	2	68	63.2%
	8							2	**61**	63	96.8%
	Reference Totals	63	86	77	58	31	54	47	63	**480**	
	Producers' Accuracy	92.1%	79.1%	55.8%	51.7%	48.4%	48.2%	91.5%	96.8%		
		Overall Accuracy: 344/480 = 71.67%									

Table 5. 5 Individual Kappa Analysis on the error matrices Table 5.2, Table5.3, and Table 5.4

	Table 5.2	Table 5.3	Table5.4
Khat	0.79	0.58	0.68
Kappa variance	0.00041	0.00063	0.00055
Z-Value	39.0	23.7	29.2

Table 5. 6 Kappa analysis results for the comparisons of Table 5.2, Table5.3, and Table 5.4

	Table 5.2	Table 5.3	Table 5.4
Table 5.2	0	6.66	3.13
Table 5.3		0	3.45
Table 5.4			0

Furthermore, we surprisingly observe that the one-stage MLP-TM classification using thee fused image is significantly better than the one-stage MLP multisource

classification using the single TM image. We postulate that this is because the training data sets manually collected in the one-stage MLP multisource classification fail to provide sufficient representation of each land category. Generally, a broad land category with certain spectral variability on a remotely sensed image may consist of several spectral features. In a data fusion application, there may arise many unexpected spectral features for some categories due to the fusion of multiple images. These spectral features may not be easily identifiable to the naked eye. Due to these unidentifiable spectral features, the manual data collecting process that was used to collect the training data sets in the one-stage MLP multisource classification failed to provide a complete representation of every class. It is this incomplete representation that made the classification performance of the one-stage MLP fusion inferior to that of the one-stage MLP-TM classification in this case study.

By combining our analyses on the three classifications, we conclude that fusion of the two images can increase the accuracy of the desired land categories if adequate and reliable training data sets are collected. The ADS adapted from the KSOM network is shown to be able to select high quality training data sets in this fusion application and thus improve the resulting classification accuracy. This is further confirmed by the learning curves plotted in Figure 5.5(a-c). Among the three classifications, the MLP training in the two-stage multisource classification had the best generalization performance in terms of minimizing the MSEs of both the training and testing data sets that were calculated using Equation 5.2.

In a typical data fusion application, there are two important factors that may have significant effects on the resulting classification performance: 1) the effect of class changes occurring during the time interval between the images, and 2) the effect of misregistration error. For ease of implementation, "we normally assume that the images are acquired at the same time, or no class changes have occurred between the acquisition dates" (Serpico & Roli 1995). However, this ideal condition is not always satisfied in practice. In the fusion application, our visual interpretation of the Landsat TM and SPOT images have shown that there were class changes between the two images on many sites. Although the desired land categories have been defined to account for a certain of class changes, e.g., the class change between planted crops and fallow areas, other types of significant class changes occurred during the period could cause a great deal of interpretation confusion in accuracy assessment and underestimate the classification results. To study the effect of the class changes between the two images, we dropped the 40 reference points with significant class changes from the 480 reference points. The resultant error matrix using the remaining 440 reference points is given in Table 5.7.

In addition to the class changes between the images, misregistration error is unavoidable when multiple image data sets with different spatial resolutions and projection systems are co-registered. Dai and Khorram (1998) emphasize that the misregistration error has significant effects on the accuracy of multisource analysis such as data fusion and change detection. Significant misregistration errors may render the fused data useless for classification, or even result in poorer classification accuracy than that using single remote sensing data. Although we have taken measures to reduce the

misregistration error during the image pre-processing, it is impossible to reduce the misregistration error to zero. To study the effect of the misregistration error, we dropped 9 reference points that were found incorrectly registered from the 480 reference points. Most of these incorrectly registered points were located on the boundary between two different land categories. The resultant error matrix using the remaining 471 reference points is given in Table 5.8. To show the classification accuracy that could be obtained without significant class changes and misregistration errors, we dropped the 48 reference points either with significant class changes or misregistration errors from the 480 reference points. The resultant error matrix is shown in Table 5.9. Based on the Kappa analyses shown in Table 5.10 and Table 5.11, we see that in this data fusion application the effect of the time changes is significant and the effect of the geometrical co-registration of the two images is insignificant.

Table 5. 7 Error matrix on the classified result of the fused image by dropping 40 reference points with time change from Table 5.2

		Reference Data									
Classified Image		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Classified Totals**	**Users' Accuracy**
	1	**51**				2				53	96.2%
	2	1	**71**		1			3		76	93.4%
	3	2	2	**54**	6	2	1			67	80.6%
	4	3	6	5	**38**		4			56	67.9%
	5					**22**				22	100.0%
	6	4	2	3	7	1	**42**	2		61	68.9%
	7	1	2					**40**	1	44	90.9%
	8								**61**	61	100.0%
	Reference Totals	62	83	62	52	27	47	45	62	**440**	
	Producers' Accuracy	82.3%	86.6%	87.1%	73.1%	81.5%	89.4%	88.9%	98.4%		
		Overall Accuracy: 379/440 = 86.14%									

Table 5. 8 Error matrix on the classified result of the fused image by dropping 9 reference points with significant misregistration error from Table 5.2

		Reference Data									
Classified Image		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Classified Totals**	**Users' Accuracy**
	1	**51**		2	1	4				58	87.9%
	2	1	**72**		1		1	2		77	93.5%
	3	3	1	**64**	8	2	1			79	81.0%
	4	2	6	6	**39**		5			58	67.2%
	5			7	1	**22**				30	73.3%
	6	4	2	6	6	1	**43**	2		65	66.1%
	7		2				1	**40**		43	93.0%
	8								**61**	61	100.0%
	Reference Totals	61	83	85	56	29	51	44	61	**471**	
	Producers' Accuracy	83.6%	86.6%	75.3%	69.6%	75.9%	84.3%	90.9%	100.0%		
		Overall Accuracy: 393/471 = 83.23%									

Table 5. 9 Error matrix on the classified result of the fused image by dropping 48 reference points either with class changes or with significant misregistration error from Table 5.2

		Reference Data									
Classified Image		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Classified Totals**	**Users' Accuracy**
	1	**51**				2				53	96.2%
	2	1	**71**		1			2		75	94.7%
	3	2	1	**54**	6	2	1			66	81.8%
	4	2	5	5	**38**		4			54	70.4%
	5					**22**				22	100.0%
	6	4	1	3	5	1	**42**	2		59	71.2%
	7		2					**40**		42	95.2%
	8								**61**	61	100.0%
	Reference Totals	60	80	62	50	27	47	44	61	**432**	
	Producers' Accuracy	85.0%	88.8%	87.1%	76.0%	81.5%	89.4%	90.9%	100.0%		
		Overall Accuracy: 379/432 = 87.73%									

Table 5. 10 Individual Kappa Analysis on Table 5.2, Table5.7, Table 5.8, and Table 5.9

	Table 5.2	Table 5.7	Table 5.8	Table 5.9
Khat	0.79	0.86	0.81	0.86
Kappa variance	0.00041	0.00036	0.00040	0.00033
Z-Value	39.0	45.4	40.6	47.2

Table 5. 11 Individual Kappa Analysis on the effects of the time change and the image registration on the classification of the fused image

	Table 5.2	Table 5.7	Table 5.8	Table 5.9
Table 5.2	0	2.58	0.55	2.48
Table 5.7		0	2.04	0.15
Table 5.8			0	1.93
Table 5.9				0

5.4 Summary and Conclusions

ANN approaches have an advantage over statistical classification methods in that they are non-parametric and require little *a priori* knowledge of the distribution model of input data. It is the distribution-free nature of ANN approaches that makes them attractive approaches for multisource remotely sensing image classification. However, there is a potential problem with ANN networks. They are very sensitive to the quality and quantity of collected training data sets. A well-trained neural network may fail to generalize well for those land categories that are poorly represented in training data sets. The task of manually selecting a sufficient and reliable training data set for each class is usually very laborious, and may become very problematic especially in a data fusion application.

To address the training data collection problem, we have proposed a two-stage neural network based multisource classification of Landsat TM and multispectral SPOT images in this paper. In the first stage, the unsupervised KSOM network was performed to play the ADS to automatically select training and testing data sets for supervised MLP classification. In the second stage, the MLP network was trained by the automatically selected training sets and then used to conduct the final classification. In addition to the two-stage multisource classification, two one-stage MLP classifications were also performed. Among the three classifications, the two-stage multisource classification obtained the best classification accuracy and efficiency and the one-stage MLP-TM classification had a better classification accuracy than that of the one-stage MLP multisource classification. To validate the fusion of the two images, two important fusion

factors have been investigated along with this fusion application. It has been shown that the effect of the class changes during the time interval between the two acquired images is significant and the effect of the geometrical co-registration of the two images is insignificant.

Based on our analysis of the experimental results from the three classifications, four major conclusions were drawn from this study as follows:

1) The ADS was shown to significantly improve the quality of training sets and thus the resulting classification accuracy. The ADS adapted from the KSOM network played an essential role in the two-stage multisource classification. Without using the ADS, the performance of the multisource classification using the manual data collecting technique was inferior to that of the single TM image classification in this case study.
2) The fusion of multiple data sources using the ANN classification approaches has improved the classification accuracy of the derived land cover mapping when adequate and reliable training data sets were used.
3) The two-stage classification scheme has been shown to be an effective hybrid classification design, which could be applied to classification applications that have difficult ties to collect sufficient and reliable training data sets.
4) This two-stage neural network based fusion study verified the fusion capability of the ANN classification approaches and the applicability of the automated ANN classification system in multisource classification.

For future work, more study is needed to verify the robustness of the ANN system using various remotely sensed data. The data collecting performance of optical sensor is sensitive to rain, fog, hail, smoke, and most importantly, clouds. Fortunately, Radio Detecting and Ranging (RADAR) sensor can overcome these restrictions. Fusion of RADAR and optical satellite data has the potential of taking advantage of the complementary characteristics in both data sets and improve the resulting classification accuracy. Furthermore, ancillary attribute data, such as elevation, slope, aspect, and texture, contains a lot of spatial information that is useful for LU/LC classifications, but may not be easily extracted from remotely sensed imageries. Inclusion of such ancillary data into LU/LC classification applications would be very useful in improving classification accuracy. ANN approaches are known to be independent of the distribution model of the input data, which provides great potential to fuse a variety of data sets. The two-stage neural network classification scheme used in this study has been shown to be viable and well-suited for land cover mapping in the fusion of Landsat TM and SPOT images. It is thus necessary to further study the suitability of the ANN system in fusion of various types of spatial data including remotely sensed optical, RADAR, and ancillary attribute data to derive the most accurate classification.

References

Benediktsson, J.A., P. H. Swain, and O. K. Ersoy, “Neural network approaches

versus statistical methods in classification of multi-source remote sensing data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 28, pp. 540-551, 1990.

Benediktsson, J.A., P. H. Swain, "A method of statistical multisource classification with a mechanism to weight the influence of the data sources," *IEEE Proceedings. IGARSS 1989*, pp. 517 – 520, 1989.

Bezdek, J. C., *et al.* "Fuzzy Kohonen clustering networks," *IEEE International Conference on Fuzzy Systems*, pp. 1035 – 1043, 1992.

Bischof, H., W. Schneider, and A. J. Pinz, " Multispectral classification of landsat images using neural networks," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 30, no. 3, pp. 482 – 490, 1992.

Bruzzone, L., C. Conese, F. Maselli, and F. Roli, "Multisource classification of complex rural areas by statistical and neural network approaches," *Photogrametric Engineering and Remote Sensing*, vol. 63, pp. 523 – 533, 1997.

Dai, X. L., and S. Khorram, "Data fusion using artificial neural networks: a case study on multitemporal change analysis," *Computers, Environments, and Urban Systems*, vol. 23, pp. 19 – 31, 1999.

Dai, X. L., and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 36, no. 5, pp. 1566 - 1577, 1998.

Foody, G. M., "The significance of border training patterns in classification by a

feedforward neural network using back propagation learning," *International Journal of Remote Sensing*, vol. 20, no. 18, pp. 3549 – 3562, 1999.

Foody, G.M., M.B. McCulloch, and W.B. Yates, "The effect of training set size and composition on artificial neural network classification," *International Journal of Remote Sensing*, vol. 16, pp. 1707 – 1723, 1995.

Franklin, S. E. "Terrain analysis from Digital patterns in geomorphometry and Landsat MSS spectral response", *Photogrammetic Engineering & Remote Sensing*, vol. 53, no. 1, pp. 59 – 65, 1987.

Goncalves, M. L., *et al.*, "A neural architecture for the classification of remote sensing imagery with advanced learning algorithms," *Proceedings of IEEE Signal Processing Society Workshop*, pp. 577 – 586, 1998.

Hegarat-Mascle, S. L., *et al.*, "Application of Dempster-Shafter Evidence Theory to unsupervised classification in multisource remote sensing," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 1018-1031, 1997.

Kohonen, T., "Self-orgazning formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, pp. 56 – 69, 1982.

Lee, T., J. A. Richards, and P. H. Swain, "Probablistic and evidential approaches for multisource data analysis," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 283 – 293, 1987.

Luo, R. C., and M. G. Kay, "Multisensor integration and fusion in intelligent systems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, pp. 901 – 931, 1989.

Murai, H., and S. Omatu, "Remote sensing image analysis using a neural network and knowledge-based processing," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 811 – 828, 1997.

Nogami, Y., *et al.*, "Remote sensing data analysis by Kohonen feature map and competitive learning," *IEEE SMC'97*, vol. 1, pp. 524 – 529, 1997.

Paola, J. D., and R. A. Schowengerdt, "The effect of neural-network structure on a multispectral land-use/land-cover classification," *Photogrametric Engineering and Remote Sensing*, vol. 63, pp. 535 – 544, 1997.

Principe, J. C., N. R. Euliano, and W. C. Lefebvre, Neural and adaptive systems: fundamentals through simulations. *John Wiley & Sons, Inc.*, pp. 656, 1999.

Rumelhart, D. E., and J. L. McClelland, Parallel distributed computing. *PDP research group, Cambridge, MA: MIT Press*, vol. 3, 1986.

Serpico, S. B., and L. Bruzzone, "Training of neural networks for classification of imbalanced remote sensing data," *IEEE Proceedings. IGARSS 1997*, pp. 1202 – 1204, 1997.

Serpico, S. B., and F. Roli, "Classification of multisensor remote-sensing images by structured neural networks," *IEEE Transactions on Geoscience* and Remote Sensing, vol. 33, no. 3, pp. 562-578, 1995.

Solberg, A. H. S., *et al.*, "Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images*", IEEE Transactions on Geoscience* and Remote Sensing, vol. 32, no. 4, pp.768 – 778, 1994.

Strahler, A.H., *et al.*, "Improving forest cover classification accuracy from Landsat

by incorporating topographic information," *Proc., 12th International Symposium on Remote Sensing of the Environment*, pp. 927 – 942, 1978.

Wan, W. J., and D. Fraser, "Multisource data fusion with multiple self-organizing maps," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, pp.1344 – 1349, 1999.

Yoshida, T., and S. Omatu, "Nueral network approach to land cover mapping," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 32, pp.1103 - 1109, 1994.

Zhou, W. Y., "Verification of the nonparametric characteristics of backpropagation neural networks for image classification," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 37, pp.771 - 779, 1999.

Zhuang, X., B. A. Engel, M. F. Baumgardner, and P. H. Swain, "Improving classification of crop residues using digital land ownership data and Landsat TM imagery," *Photogrammetic Engineering & Remote Sensing*, vol. 57, no. 11, pp. 1487 – 1492, 1991.

6. Summary and General Conclusions

In this research, we have investigated, developed, implemented, and evaluated advanced Land-Use/Land-Cover (LU/LC) classification systems using multispectral and multisource remotely sensed data. Three interconnected studies have been conducted: 1) development and evaluation of LU/LC classification systems using Simulated Annealing (SA), 2) development of an automated Artificial Neural Network (ANN) classification system for remote sensing based land cover mapping, and 3) development of a two-stage multisource classification of Landsat Thematic Mapper (TM) and Satellite Pour l'Observation de la Terre (SPOT) images. The major contributions from each of the studies are summarized from sections 6.1 to 6.3. A general summary is presented in section 6.4 and some future work is discussed in section 6.5.

6.1 Study One

In the first study of this research, we have developed two SA-based classification systems and investigated their suitability for LU/LC mapping using Landsat TM images. The first system developed in this study was the Single SA-based (S-SA) system and the second system was the Integrated SA-based (I-SA) system. Our experimental results have shown that the SA-based systems helped overcome the local minimum problem in *K*-means and thus significantly improved the classification accuracy as compared with that of the *K*-means algorithm. Two factors were found to be essential for the success of the SA-based systems: the complexity of the application and the selection of controlling

parameters. We suggest that, for simple applications, the I-SA system is a good choice because I-SA can produce a better classification more efficiently. However, for complex applications, the S-SA system should be used. Our study suggests that the selection of each single parameter in the cooling schedule of SA may not be so critical, and that it is more important to choose an optimal parameter combination to obtain accurate and efficient classification. This study is the first time that SA has been intensively investigated using satellite images for LU/LC mapping applications.

6.2 Study Two

In the second study of this research, we developed an automated two-module ANN classification system consisting of a supervised Multilayer Perceptron (MLP) network module and an unsupervised Kohonens's Self-Organizing Mapping (KSOM) network module. In the KSOM network module, we incorporated SA global searching procedures into the standard KSOM network to reduce the likelihood to be trapped by local minima. To test the applicability of the ANN system, we have applied the system to perform LU/LC classifications of a Landsat TM image using the supervised MLP network classification and two unsupervised KSOM network classifications. Experimental results have shown that among the three classifications, the supervised MLP network had the best classification performance. The KSOM-SA network refined by SA has improved the classification performance as compared with the standard KSOM network. Based on our analysis of the experimental results, we suggest that in complex LU/LC mapping applications supervised MLP networks should be used, and

that unsupervised KSOM networks can be used to assist in selecting training data for supervised classification.

In addition to providing a case study to verify the superior classification capability of the ANN approaches in image classifications using remotely sensed data, our unique contribution in this study is the development of the automated two-module ANN system. Although ANN classification approaches have been shown to be well-suited for a variety of remotely sensed classifications in numerous studies (Zhang & Foody 2001; Dai & Khorram 1999; Kanellopoulos & Wilkinson 1997; Paola & Schowengerdt 1997; Foody 1995; Heermann & Khazenie 1992; Benediktsson *et al.* 1990), their uses are still limited to those neural network professionals. This is because there is no automated ANN classification system provided by any commercial remote sensing software package. In this study, we have successfully developed and added the automated ANN classification system into the very popular commercial remote sensing software package ERDAS IMAGINE. Our case study has demonstrated that the ANN system is able to fulfill the tasks of network training pattern creation, network training, and network generalization. The successful fulfillments lead us to the conclusion that the ANN system is a fully automated and user-friendly classification system and can be easily operated by a remote sensing expert who may not have strong backgrounds in neural networks.

6.3 Study Three

In the third study of this research, we adapted the automated ANN system to perform a two-stage neural network based multisource classification of Landsat TM and

multispectral SPOT images. Our experimental results and analysis demonstrate that the fusion of multiple data sources using the ANN classification approaches is able to improve classification accuracy in LU/LC classification applications if adequate and reliable training data sets are used. However, the traditional manual training data collection may become incapable of collecting such training data sets in data fusion applications. The training data collection problem has been recognized in many ANN fusion studies (Zhou 1999; Foody *et al.* 1995; Foody 1999; Zhuang *et al.* 1991; Benediktsson *et al.* 1990). Some hybrid classification approaches have been investigated in recent years to assist the manual training data collection (Goncalves *et al.* 1998; Nogami *et al.* 1997; Yoshida and Omatu 1994). However, no automated training data selection is observed from any of them. To fill the gap, we created an Automated Data Selector (ADS) based on the unsupervised KSOM module in the ANN system to replace the manual training data collection in the two-stage multisource classification. The ADS has been shown to significantly improve the quality of training sets and thus the overall classification accuracy. The role of the ADS is essential in this two-stage multisource classification. Without using the ADS, the multisource classification accuracy of the Landsat TM and SPOT images using the manual training data collection technique is even inferior to the single source TM image classification. This study strongly suggests that this two-stage classification scheme is a valuable hybrid classification design that may be applied to classification applications when it is difficult to collect sufficient and reliable training data sets.

6.4 General Summary

These three individual studies act cooperatively in this research. The first study on SA-based classification systems provides working knowledge and experience to incorporate SA global searching procedures into classification approaches that suffer from the same local minimum problem as *K*-means. Using the knowledge obtained from the first study, we successfully refine the standard KSOM learning algorithm with SA in the second study. Our classification results from both the first and second studies demonstrate that SA-based systems overcome the local minimum problem and thus improve the resulting classification accuracy. Furthermore, the development of the two-module ANN classification system presented in the second study makes the third study possible. The unsupervised KSOM and supervised MLP network modules provided by the ANN classification system can be implemented independently (e.g., in the second study) or complementarily (e.g., in the third study). This offers great flexibility to develop the hybrid classification scheme presented in the third study. Furthermore, the development of the two-stage multisource classification presented in the third study verifies the flexibility and wide usability of the ANN classification system. In summary, these three studies strongly demonstrate that the automated SA and ANN classification systems are feasible and robust for accurate LU/LC mapping using remotely sensed images.

In this research, we found that, in addition to the classification approaches, two factors had effects on the resulting LU/LC classification accuracy: 1) the complexity of

the application including the desired class level and the landscape conditions, and 2) the quality of acquired images including the spectral and spatial resolutions. Generally to say, the increase of the application complexity requires the increase of the spectral and spatial resolution of the acquired images. Remotely sensed data with high spectral and spatial resolutions may provide sufficient information to derive more accurate thematic land information in complex classification applications.

6.5 Recommended Future work

Base on the three studies we have conducted in this research, the future work can be extended in the following directions:

- SA-based classification systems have been studied for their potential for improving classification accuracies in LU/LC classification applications using Landsat TM images. This is the first time that SA has been intensively investigated for LU/LC mapping using Landsat TM images. To test the general applicability and robustness of SA-based classification systems for LU/LC mapping applications using remotely sensed data, it is necessary to apply SA-based classification systems for various LU/LC classification applications using many other remotely sensed data other than Landsat TM data, such as SPOT. Additional classification applications would be valuable to provide more evidences to demonstrate the potential of SA-based classification systems to improve classification accuracy as compared to many traditional classification approaches suffering from the local minimum problem.

- In this research, we have developed an automated two-module ANN classification system consisting of two neural network modules: a supervised MLP module and an unsupervised KSOM module. In addition to these two neural network models, many other ANN approaches, e.g., ART network (Carpenter *et al.* 1997) and Radial Basis Function (RBF) network (Bishop 1995), have been shown to be suitable in remote sensing classification. It would be worthwhile to add these network modules into the ANN classification system. Multiple ANN classification alternatives will offer more flexibility to develop effective hybrid classification schemes so that an optimal classification can be obtained.
- The ANN system has been shown to able to fuse the Landsat TM and SPOT data for LU/LC classification. Other than multiple optical sensor data, there are many other data sets that may provide complimentary information for LU/LC classification. For example, optical sensor is sensitive to rain, fog, hail, smoke, and clouds. Radio Detecting and Ranging (RADAR) sensor can overcome these restrictions. Fusion of RADAR and optical satellite data has the potential of combining the complementary characteristics in both data sets and thus improve classification accuracy. In addition, ancillary attribute data, such as elevation, slope, aspect, and texture, provides a lot of spatial information that is useful for LU/LC classifications, but may not be easily derived from remotely sensed imageries. Inclusion of such ancillary data into LU/LC classification applications would be very helpful in improving classification accuracy. An ideal and powerful fusion system is expected to be able to fuse a variety of spatial data sets.

ANN approaches are known to be independent of the distribution model of the input data, which provides great potential to fuse a variety of data sets. It is thus necessary to study the applicability of the ANN system in fusion of various types of spatial data including remotely sensed optical, RADAR, and ancillary attribute data such that the most accurate LU/LC classification is derived.

Bibliography

Aarts, E. H. L., and P. J. M. Van Laarhoven, "Simulated annealing: a pedestrian review of the theory and some applications", *Pattern recognition theory and applications,* pp.179-192, 1987.

Aarts, E., and J. Korst, Simulated annealing and Boltzmann machines: a stochastic approach to combinatorial optimization and neural computing. *John Wiley & Sons,* 1989.

Anderson, J. R., *et al.*, "A land use and land cover classification system for use with remotely sensed data," *Washington, DC: U.S. Geological Survey Professional Paper 964*, pp. 28, 1976.

Ahern, F.J., *et al.*, "Simultaneous microwave and optical wavelength observations of agricultural targets", *Canadian Journal of Remote Sensing*, vol. 4, no. 2, pp. 127 – 142, 1978.

Amar, F., *et al.*, "Comparison of neural network algorithms for remote sensing applications," *IEEE Proceedings. IGARSS 1995*, vol. 1, pp. 694 – 696, 1995.

Aragon, C. R., *et al.*, "Optimization by simulated annealing: an experimental evaluation," *working paper*, 1984.

Atkinson, P., and A. R. L. Tatnall, "Neural networks in remote sensing," *International Journal of Remote Sensing*, vol. 18, pp. 699 – 709, 1997.

Babu, G. P., "Self-organizing neural networks for spatial data," *Pattern Recognition Letters*, vol. 18, pp. 133 –142, 1997.

Ball, G. H., and D. J. Hall, "ISODATA: a novel method data analysis and pattern classification," *Technical report, Stanford Research Institute, Menlo Park, California, USA*, 1965.

Bandyopadhyay, S., S. K. Pal, and C. A. Murthy, "Simulated annealing based pattern classification," *Journal of Information Sciences*, vol. 109, pp. 165 – 184, 1998.

Baraldi, A. and F. Parmiggiani, "A neural network for unsupervised categorization of multivalued input patterns: an application to satellite image clustering," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 33, pp. 305 - 316, 1995.

Belward, A. S., *et al.*, "An unsupervised approach to the classification of semi-natural vegetation from Landsat Thematic Mapper data," *International Journal of Remote Sensing*, vol.11, pp.429 – 445, 1990.

Benediktsson, J. A., and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 37, pp. 1367 - 1377, 1999.

Benediktsson, J. A., and J. R. Sveinsson, "Feature extraction for multisource data classification with artificial neural networks," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 727-740, 1997.

Benediktsson, J.A., *et al.*, "Parallel consensual neural networks," *IEEE Transaction on Neural Networks*, vol. 8, pp. 54 - 64, 1997.

Benediktsson, J.A., P. H. Swain, and O. K. Ersoy, "Neural network approaches

versus statistical methods in classification of multi-source remote sensing data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 28, pp. 540-551, 1990.

Benediktsson, J.A., P. H. Swain, "A method of statistical multisource classification with a mechanism to weight the influence of the data sources," *IEEE Proceedings. IGARSS 1989*, pp. 517 – 520, 1989.

Bezdek, J. C., *et al.* "Fuzzy Kohonen clustering networks," *IEEE International Conference on Fuzzy Systems*, pp. 1035 – 1043, 1992.

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

Bilbo, G. L., *et al.*, "Mean field annealing: a formalism for constructing GNC-like algorithms," *IEEE Transaction on Neural Networks*, vol. 3, no. 1, pp. 131 – 138, 1992.

Bischof, H., and A. Leonardis, "Finding optimal neural networks for land use classification," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 36, no.1, pp. 337 – 341, 1998.

Bischof, H., W. Schneider, and A. J. Pinz, " Multispectral classification of landsat images using neural networks," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 30, no. 3, pp. 482 – 490, 1992.

Bishop, C. M, Radial basis functions. In *Neural networks for pattern recognition*. *Oxford, New York: Clarendon Press*, 1995.

Bishop, Y., S. Fienberg, and P. Holland, Discrete multivariate analysis – theory and

Practice. *MIT Press: Cambridge, Massachusetts*, pp.575, 1975.

Bolstad, P.V, and T. M. Lillesand, "Rapid maximum likelihood classification," *Photogrammetic Engineering & Remote Sensing*, vol. 57, pp. 64 – 74, 1991.

Bonomi, E., and J. L. Lutton, "The N-city traveling salesman problem: statistical mechanics and Metropolis algorithm," *SIAM Review*, vol.26, pp. 551 – 568, 1984.

Brisco, B., "Multi-dimensional approaches to the RADAR remote sensing of land cover", *Ph.D. thesis, Department of Geography, University of Kansas, Lawrence, KS*, 1985.

Brisco, B., and R.J. Brown, "Multidate SAR/TM synergism for crop classification in western Canada", *Photogrammetic Engineering & Remote Sensing*, vol. 61, no. 8, pp. 1009 – 1014, 1995.

Brown, D. E., and C. L. Huntley, "A practical application of simulated annealing to clustering," *Pattern recognition*, vol. 25,no. 4, pp. 401-412, 1992.

Brown, J. F., *et al.*, "Using multisource data in global land-cover characterization: concepts, requirments, and methods", *Photogrammetic Engineering & Remote Sensing*, vol. 59, pp. 977-987, 1993.

Bruzzone, L., D. F. Prieto, and S. B. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1350 – 1359, 1999.

Bruzzone, L., C. Conese, F. Maselli, and F. Roli, "Multisource classification of

complex rural areas by statistical and neural network approaches," *Photogrametric Engineering and Remote Sensing*, vol. 63, pp. 523 – 533, 1997.

Bruzzone, L., and D. F. Prieto, "unsupervised retraining of a maximum likelihood classifier of multitemporal remote sensing images," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 456 – 460, 2001.

Burkard, R.E., and F. Rendl, "A thermodynamically motivated simulation procedure for combinatorial optimization problems," *European Journal of Oper. Res.*, vol.17, pp. 169 – 174, 1984.

Carpenter, G.A., M.N. Gjaja, S. Gopal, and C.E. Woodcock, "ART neural networks for remote sensing: vegetation classification from Landsat TM and terrain data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 35, no. 2, pp. 308-325, 1997.

Cerny, V., "Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm," *Journal of Optimization Theory and Applications*, vol. 45, pp. 45 –51, 1985.

Chen, Z., *et al.*, "Texture segmentation based on Wavelet and Kohonen network for remotely sensed images," *IEEE SMC'99 Conference Proceedings*, vol. 6, pp. 816 – 821, 1999.

Cihlar, J., *et al.*, "Land Cover classification with AVHRR multispectral composites in northern environments", *Remote sensing of Environment*, vol.58, pp.36 – 51,1996.

Cihlar, J., *et al.*, "Classification by progressive generalization: a new automated

methodology for remote sensing multichannel data," *International Journal of Remote Sensing*, vol. 19, no. 14, pp. 2685 - 2704, 1998.

Couloigner, *et al.*, "Benefit of the future SPOT-5 and of data fusion to urban roads mapping", *International Journal of Remote Sensing*, vol. 19, no. 8, pp. 1519 – 1532, 1998.

Dai, X. L., and S. Khorram, "Data fusion using artificial neural networks: a case study on multitemporal change analysis," *Computers, Environments, and Urban Systems*, vol. 23, pp. 19 – 31, 1999.

Dai, X. L., and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 36, no. 5, pp. 1566 - 1577, 1998.

Dai, X. L., and S. Khorram, "Development of a new automated land cover change detection system from remotely sensed imagery based on artificial neural networks," *IEEE Proceedings. IGARSS 1997*, pp. 1029 – 1031, 1997.

Dubes, R., and A. K. Jain, "Cluster methodologies in exploratory data analysis," Advances in Computing, vol. 19, pp. 113 – 228, 1980.

Dubes, R., and A. K. Jain, "Cluster techniques: the user's dilemma," *Pattern Recognition*, vol. 8, pp. 247 – 260, 1976.

Duda, R. O., and P. E. Hart. Pattern classification and scene analysis. *New York: Wiley-Interscience*. 1973.

Ediriwickrema, J., and S. Khorram, "Hierarchical maximum-likelihood classification for improved accuracies," *IEEE Transaction on Geoscience and Remote Sensing*,

vol. 35, no. 4, pp. 810- 816, 1997.

Elachi, C.T., "Spaceborne RADAR Remote sensing: applications and techniques", *IEEE Press*, New York, pp.255, 1988.

Eppler, W. G., "An improved version of the table look up algorithm for pattern recognition," in *Proceedings, of 9th International Symposium on Remote Sensing of Environment*, Ann Arbor, MI, pp. 793 – 812, 1974.

ERDAS, ERDAS field guide. Atlanta, GA: ERDAS, Inc., 1994.

Ersoy, O. K., and D. Hong, "Parallel self-organizing hierarchical neural networks," *IEEE Transaction on Industrial Electronics*, vol. 40, no. 2, pp. 218 – 227, 1993.

Evans, D.L., *et al.*, "Mapping forest distributions with AVHRR data", *World Resource Rev.,* v.(5), pp. 66-71, 1993.

Fitch, J. P., *et al.*, "Ship wave detection procedure using conjugate gradient trained neural network," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 29, 1991.

Foody, G. M., "The significance of border training patterns in classification by a feedforward neural network using back propagation learning," *International Journal of Remote Sensing*, vol. 20, no. 18, pp. 3549 – 3562, 1999.

Foody, G. M., and M. K. Arora, "An evaluation of some factors affecting the accuracy of classification by an artificial neural network," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 799-810, 1997.

Foody, G. M., "Approaches for the production and evaluation of fuzzy land cover classifications from remotely sensed data," *International Journal of Remote Sensing*, vol. 17, pp. 1317 – 1340, 1996.

Foody, G.M., M.B. McCulloch, and W.B. Yates, "The effect of training set size and composition on artificial neural network classification," *International Journal of Remote Sensing*, vol. 16, pp. 1707 – 1723, 1995.

Foody, G. M., et al., "Derivation and application of probabilities of class membership form the maximum-likelihood classification," *Photogrammetic Engineering & Remote Sensing*, vol. 58, pp. 1335 – 1341, 1992.

Frank, T. D., "Mapping dominant vegetation communities in the Colorado Rocky Mountain front range with Landsat Thematic Mapper and digital terrain data," *Photogrammetic Engineering & Remote Sensing*, vol.54, no.12, pp.1727 – 1734, 1988.

Franklin, S. E., and B. A. Wilson, "A three-stage classifier for remote sensing of mountain environments," *Photogrammetic Engineering & Remote Sensing*, vol.58, no. 4, pp. 449 – 454, 1992.

Franklin, S. E. "Terrain analysis from Digital patterns in geomorphometry and Landsat MSS spectral response", *Photogrammetic Engineering & Remote Sensing*, vol. 53, no. 1, pp. 59 – 65, 1987.

Fukunaga, K., Introduction to statistical pattern recognition. 2nd. *New York: Academic*, 1990.

Geman, S., and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Iintelligence*,, vol. 6, pp. 721 – 741, 1984.

Goncalves, M. L., *et al.*, "A neural architecture for the classification of remote sensing

imagery with advanced learning algorithms," *Proceedings of IEEE Signal Processing Society Workshop*, pp. 577 – 586, 1998.

Gong P. *et al*., "An assessment of some factors influencing multispectral land-cover classification", *Photogrammetic Engineering & Remote Sensing*, vol.56, no. 5, pp. 597 – 603, 1990.

Granville, V., M. Krivanek, and Jean-Paul Rasson, "Simulated annealing: a proof of convergence," *IEEE Transactions on Pattern Analysis and Machine Iintelligence*, vol. 16, no. 6, pp. 652-656, 1994.

Guneriussen, T., *et al.*, "Snow monitoring using EMISAR and ERS-1 data within the European multi-sensor airborne campaign EMAC-95," *the 1997 IEEE International Geoscience and Remote Sensing Symposiums*, vol.2, pp. 631 –633,1997.

Haack, B.N., *et al.*, "Assessment of Landsat MSS and TM data for urban and near-urban landcover digital classification", *Remote Sensing of Environment*, vol. 21, no. 2, pp. 201 – 213, 1987.

Harrison, A.R., and T.R. Richards, "Multispectral classification of urban land use using SPOT HRV data", *Digest – International Geoscience and Remote Sensing Symposium, Edinburgh, U.K.*, pp. 205 – 206, 1988.

Heermann, P. D., and N. Khazenie, "Classification of multispectral remote sensing data using a back-propagation neural network," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 30, pp.81-88, 1992.

Hegarat-Mascle, S. L., *et al.*, "Application of Dempster-Shafter Evidence Theory

to unsupervised classification in multisource remote sensing," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 1018-1031, 1997.

Hegarat-Mascle, S. L., S. D. Vidal-Madjar, and P. Olivier, "Applications of simulated annealing to SAR image and classification problems," *International Journal of Remote Sensing*, vol. 17, no. 9, pp. 1761-1776, 1996.

Hodgson, M. E., "Reducing the computational requirements of the minimum-distance classifier," *Remote Sensing of Environment*, vol. 25, pp. 117 – 128, 1988.

Hopfield, J. J., and D. W. Tank, "Neural computation of decisions in optimization problems," *Biol. Cybern.*, vol. 52, pp. 141 – 152, 1985.

Huang, W. M., and R. P. Lippman, "Neural net and traditional classifiers," *Neural Info. Processing Syst., D. Anderson, ed., New York: American Institute of Physics*, pp. 387 – 396, 1988.

Jaakkola, S., *et al.*, "Applicability of SPOT for forest inventory, mapping and change monitoring*", SPOT –1 First In-flight results (Toulouse Cepadues-Editions)*, pp. 259 – 264, 1987.

Jain, A. K., R. P. W. Dulin, and J. C. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22, no. 1, pp. 4 – 37, 2000.

Jain, A. K., Fundamentals of digital image processing. *Englewood Cliffs, New Jersy: Prentice-Hall,* pp. 414 – 421, 1989.

Jain, A.K., and R. Dubes, Algorithms for Clustering Data. *Englewood Cliffs, New Jersy: Prentice-Hall*, pp. 92 – 117,1988.

Jenson, J. R., Introductory digital image processing. *Englewood Cliffs, NJ: Pretentice Hall,* 1996.

Kanellopoulos, I., and G.G. Wilkinson, "Strategies and best practice for neural network image classification," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 711 – 725, 1997.

Kavzoglu, T., and P. M. Mather, "Pruning artificial neural networks: an example using land cover classification of multi-sensor images," *International Journal of Remote Sensing*, vol. 20, no. 14, pp. 2787 – 2803, 1999.

Key, J., *et al.*, "Classification of merged AVHRR and SMMR arctic data with neural networks," *Photogrammetic Engineering & Remote Sensing*, vol.55, pp. 1331 – 1338, 1989.

Khorram, S., *et al.*, "Comparison of Landsat MSS and TM data for urban land-use classification", *IEEE Transactions on Geoscience and Remote Sensing*, vol.25, no. 2, pp.238 – 243, 1987.

Kiema, J. B. K, "Texture analysis and data fusion in the extraction of topographic objects from satellite imagery," *International Journal of Remote Sensing*, vol. 23, no. 4, pp. 767 – 776, 2002.

Kirkpatrick, S., C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671-688, 1983.

Klein, R. W., and R. C. Dubes, "Experiments in projection and clustering by annealing," *Pattern Recognition*, vol. 22, no. 2, pp. 213-220, 1989.

Kohonen, T., "Self-orgazning formation of topologically correct feature maps," *Biol.*

Cybern., vol. 43, pp. 56 – 69, 1982.

Kolonko, M., and M. T. Tran, "Convergence of simulated annealing with feedback temperature schedules," *Probability in the Engineering and Informational Sciences*, vol. 11, pp. 279-304, 1997.

Kux, H. J. H., *et al.*, "Evaluation of RADARSAT for land use and land cover dynamics in the Southern Brazilian Amazon State of Acre", *Canadian Journal of Remote Sensing*, vol. 24, n. 4, pp. 350 –359, 1998.

Laarhoven, P.J.M., Theoretical and computational aspects of simulated annealing. *Amsterdam, Netherlands: Center for mathematics and computer science,* 1988.

Laba, M., S. D. Smith, and S. D. Degloria, "Landsat-based land cover mapping in the lower Yuna River watershed in the Dominican Republic," *International Journal of Remote Sensing*, vol. 18, no. 14, pp. 3011 – 3025, 1997.

Lark, R. M, "A reappraisal of unsupervised classification, I: correspondence between spectral and conceptual classes," *International Journal of Remote Sensing*, vol. 16, no. 8, pp. 1425 – 1443, 1995.

Lee, J., *et al.*, "A neural network approach to cloud classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, pp.846 – 855, 1990.

Lee, L., and D. A. Landgrebe, "Analyzing high-dimensional mutltispectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, no. 4, pp. 792 – 800, 1993.

Lee, T., J. A. Richards, and P. H. Swain, "Probablistic and evidential approaches for

multisource data analysis," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 283 – 293, 1987.

Leong, H.W., D.F. Wong, and C. L. Liu, "A simulated-annealing channel router," *Proc. ICCAD*, Santa Clara, pp. 226 – 229, 1985.

Lillesand, T. M., and R. F. Keifer, Remote sensing and image interpretation. 4th. *NewYork: Wiley*, 2000.

Lin, C. T., Y. C. Lee, and H. C. Pu, "Satellite sensor image classification using cascaded architecture of neural fuzzy network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 1033 – 1043, 2000.

Lippmann, R. P., An introduction to computing with neural networks. *IEEE Acoust. Speech Signal Process.*, pp. 4 – 22, 1987.

Lippmann, R. P, "Pattern recognition using neural networks," *IEEE Communications Magazine*, pp. 47 – 64, 1989.

Locatellli, M., "Convergence properties of simulated annealing for continuous global optimization," *Journal of Applied Probability*, vol. 33, pp. 1127-1140, 1996.

Loveland, T. R., *et al.*, "Development of landcover characteristics database for the conterminous U.S.", *Photogrammetic Engineering & Remote Sensing*, v.(57), pp. 1453-1463, 1991.

Lundy, M., and A. Mees, "Convergence of an annealing algorithm," *Math. Prog.*, vol. 34, pp. 111 – 124, 1986.

Luo, R. C., and M. G. Kay, "Multisensor integration and fusion in intelligent systems,"

IEEE Transactions on Systems, Man, and Cybernetics, vol. 19, pp. 901 – 931, 1989.

Malingreau, J. P., *et al.*, "AVHRR for monitoring tropical deforestation", *International Journal of Remote Sensing*, vol.10, pp. 855 – 867, 1989.

Mao, J., K. Mohiuddin, and A. K. Jain, "Parsimonious network design and feature selection through node pruning," *Proc. 12th International Conference on Pattern Recognition*, pp. 622 –624, 1994.

Marth, P. M., "A computationally-efficient maximum-likelihood classifier employing prior probabilities for remotely sensed data," *International Journal of Remote Sensing*, vol. 6, no. 2, pp. 369 – 376, 1985.

Maselli, F., C. Conese, L. Petkov, and R. Resti, "Inclusion of priori probabilities derived from nonparametric process into the maximum-likelihood classifier," *Photogrammetic Engineering & Remote Sensing*, vol. 58, pp. 201 – 207, 1992.

Mausel, P.W., W.J., Kamber, and J. K. Lee, "Optimum band selection for supervised classification of multispectral data," *Photogrammetic Engineering & Remote Sensing*, vol.56, no.1, pp.55 – 60, 1990.

McErlean, F. J., D. A. Bell, and S.I. McClean, "The use of simulated annealing for clustering data in databases," *Information Systems*, vol. 15, no. 2, pp. 233-245, 1990.

Metropolis, N., *et al.*, "Equation of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087-1092, 1953.

Murai, H., and S. Omatu, "Remote sensing image analysis using a neural network and

knowledge-based processing," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 811 – 828, 1997.

Nogami, Y., *et al.*, "Remote sensing data analysis by Kohonen feature map and competitive learning," *IEEE SMC'97*, vol. 1, pp. 524 – 529, 1997.

Nour, M. A., and G. R. Madey, "Heuristic and optimization approaches to extending the Kohonen slef-organizing algorithm," *European Journal of Operational Research*, vol. 93, pp. 428 – 448, 1996.

Quegan, S., *et al.*, "Combining unsupervised and knowledge-based method in large-scale forest classification," *IEEE Proceedings. IGARSS 2000*, vol. 1, pp. 426 – 428, 2000.

Palylyk, C. L, and P. H., Crown, "Application of clustering to Landsat MSS digital data for Peatland inventory," Canadian Journal of Remote Sensing, vol. 10, pp. 201 – 208, 1984.

Paola, J. D., and R. A. Schowengerdt, "The effect of neural-network structure on a multispectral land-use/land-cover classification," *Photogrametric Engineering and Remote Sensing*, vol. 63, pp. 535 – 544, 1997.

Paola, J. D., and R. A. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, pp. 981 – 996, 1995.

Paris, J. F., and H. H. Kwong, "Characterization of vegetation with combined Thematic Mapper (TM) and shuttle imaging RADAR (SIR-B) image data",

Photogrammetic Engineering & Remote Sensing, vol. 54, no. 8, pp. 1187 – 1193, 1988.

Pierce, L., *et al.*, "An automated unsupervised/supervised classification methodology," *IEEE Proceedings. IGARSS 1998*, vol. 4, pp. 1781 – 1783, 1998.

Pohl, C., and J. L. Van Genderen, "Multisensor image fusion in remote sensing: concepts, methods, and applications," *International Journal of Remote Sensing*, vol. 19, no. 5, pp. 823 – 854, 1998.

Principe, J. C., N. R. Euliano, and W. C. Lefebvre, Neural and adaptive systems: fundamentals through simulations. *John Wiley & Sons, Inc.*, pp. 656, 1999.

Remund, Q. P., D. G. Long, and M. R. Drinkwater, "Polar sea-ice classification using enhanced resolution NSCAT data," ," *IEEE Proceedings. IGARSS 1998*, vol. 4, pp. 1976 - 1978, 1998.

Richards, J. A., Remote sensing digital image analysis. *Springer-Verlag*, 1986.

Richards, J. A., D. A. Landgrebe, and P. H. Swain, "A means for utilizing ancillary information in multispectral classification," *Remote Sensing of Environment*, vol. 12, pp. 463 – 477, 1982.

Rosin, P. L., and F. Fierens, "Improving neural network generalization," *IEEE Proceedings. IGARSS 1995*, pp. 1255 – 1257, 1995.

Rose, K., "Deterministic annealing for clustering, compression, classification, regression, and related optimization problem," *IEEE Proceeding*, vol. 86, pp. 2210 – 2239, 1998.

Rumelhart, D. E., and J. L. McClelland, Parallel distributed computing. *PDP research*

group, Cambridge, MA: MIT Press, vol. 3, 1986.

Running, S. W., *et al.*, " A remote sensing based vegetation classification logic for global land cover analysis", *Remote sensing environment*, v.(50), pp.39-48, 1995.

Sader, S. A., D. Ahl, W. S. Liou, "Accuracy of Landsat-TM and GIS rule –based methods for forest wetland classification in Marine," *Remote Sensing of Environment*, vol.53, pp. 133 – 144, 1995.

Schalkoff, R. J., Pattern recognition: statistical, structural, and neural approaches. *John Wiley & Sons, Inc.*, 1992.

Schowengerdt, R. A., Techniques for image processing and classification in remote Sensing. *New York: Academic Press*, pp. 129 – 214,1983.

Selim, S. Z., and K. Alsultan, "A simulated annealing algorithm for the clustering problem," *Pattern Recognition*, vol. 24, no. 10, pp.1003-1008, 1991.

Selim, S. Z., and M.A. Ismail, "K-means-type algorithms: a generalized convergence theorem and characterization of local optimality," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 81 - 87, 1984.

Serpico, S. B., and L. Bruzzone, "Training of neural networks for classification of imbalanced remote sensing data," *IEEE Proceedings. IGARSS 1997*, pp. 1202 – 1204, 1997.

Serpico, S. B., and F. Roli, "Classification of multisensor remote-sensing images by structured neural networks," *IEEE Transactions on Geoscience* and Remote Sensing, vol. 33, no. 3, pp. 562-578, 1995.

Shimabukuro, *et al.*, "Land cover classification from RADARSAT data of the

Tapajos National forest, Brazil", *Canadian Journal of Remote Sensing*, vol. 24, no. 4, pp. 393 – 401, 1998.

Shlien, S., and A. Smith, "A rapid method to generate spectral theme classification of Landsat imagery," *Remote Sensing of Environment*, vol. 4, pp. 67 – 77, 1975/1977.

Snyder, W., *et al.*, "Segmentation brain images using mean field annealing," *Inform. Process. Med. Imag., Proc. 12th Int. Conf. IPMI, Wye, UK,* pp. 218 – 226, 1991.

Solberg, A. H. S., T. Text, and A. K. Jain, "A Markov random field model for classification of multisource satellite imagery,"*IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, pp. 100 – 112, 1996.

Solberg, A. H. S., *et al.*, "Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images*", IEEE Transactions on Geoscience* and Remote Sensing, vol. 32, no. 4, pp.768 – 778, 1994.

Strahler, A.H., "The use of prior probabilities in maximum likelihood classification of remotely sensed data," *Remote Sensing of the Environment*, vol. 10, pp. 135 – 163, 1980.

Strahler, A.H., *et al.*, "Improving forest cover classification accuracy from Landsat by incorporating topographic information," *Proc., 12th International Symposium on Remote Sensing of the Environment*, pp. 927 – 942, 1978.

Swain, P. H., and S. M. Davis, Remote sensing: the quantitative approach. *New York: McGraw-Hill*, pp. 136 – 188,1978.

Thomson, A.G., R.M. Fuller, and J.A. Eastwoods, "Supervised versus unsupervised

methods for classification of coasts and river corridors from airborne remote sensing," *International Journal of Remote Sensing*, vol. 18, pp. 3423 – 3431, 1998.

Toda, M., R. Kubo, and N. Saito, Statistical physics. *Springer-Verlag, Berlin*, 1983.

Townshend, J. R., *et al.*, "The 1km resolution global data set: needs of the international geosphere bioshpere programme," *International Journal of Remote Sensing*, vol. 15, pp. 3417 - 3441, 1994.

Townshend, J. R., "Global data sets for land applications from the advanced very high resolution radiometer: introduction", *International Journal of Remote Sensing*, vol. 15, pp.3319-3332, 1994.

Townshend, J.R., and C.O. Justice, "Analysis of the dynamics of African vegetation using the normalized difference vegetation index", *International Journal of Remote Sensing*, vol.7, pp. 1435 – 1445, 1986.

Tou, J. T., and R. C. Gonzales, Pattern recognition principles, *Addison- Wesley Publishing Company,* 1977.

Tzeng, Y. C., K. S. Chen, *et al.*, "A fuzzy neural network to SAR image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, pp. 096 – 1102, 1994.

Tzeng, Y. C., and K. S. Chen, "A fuzzy neural network to SAR image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, pp.301 – 307, 1998.

Ulaby, F.T., *et al.*, "Crop classification using airborne RADAR and Landsat data", *IEEE*

Transactions on Geoscience and Remote Sensing, vol. 20, pp.518 – 528, 1982.

Wald, L., "A European proposal for terms of reference in data fusion," *International Archives of Photogrammetry and Remote Sensing*, vol. 32, Part 4, Commission VII, pp. 651 – 654, 1998.

Wan, W. J., and D. Fraser, "Multisource data fusion with multiple self-organizing maps," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, pp.1344 – 1349, 1999.

Webos, P., "Beyond regression: new tools for prediction and analysis in the Behavioral sciences," *Ph.D. dissertation, Appl. Math., Harvard University,* 1974.

White, S. R., "Concepts of scale in simulated annealing," *Proc. ICCD*, Port Chester, pp. 646 – 651, 1984.

Xiao, R. R., R. Wilson, and R. Carande, "Neural network classification with IFSAR and multispectral data fusion," *IEEE Proceedings. IGARSS 1998*, vol. 3, pp. 1327 – 1329, 1998.

Yoshida, T., and S. Omatu, "Nueral network approach to land cover mapping," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 32, pp.1103 - 1109, 1994.

Yuan, H., S. Khorram, and X. L. Dai, "Applications of simulated annealing minimization technique for unsupervised classification of remotely sensed data," *IEEE Proceedings. IGARSS 1999*, vol. 1, pp. 134 – 136, 1999.

Zhang, J., and G. M. Foody, "Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: statistical and artificial neural network approaches," *International Journal of Remote Sensing*, vol. 22, no. 4,

pp. 615 – 628, 2001.

Zhang, Z. X. *et al.*, "Automatic registration of multi-source imagery based on global image matching," *Photogrammetic Engineering & Remote Sensing*, vol. 66, no. 5, pp. 625 – 629, 2000.

Zhou, W. Y., "Verification of the nonparametric characteristics of backpropagation neural networks for image classification," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 37, pp.771 - 779, 1999.

Zhuang, X., B. A. Engel, M. F. Baumgardner, and P. H. Swain, "Improving classification of crop residues using digital land ownership data and Landsat TM imagery," *Photogrammetic Engineering & Remote Sensing*, vol. 57, no. 11, pp. 1487 – 1492, 1991.

APPENDIX: Glossary

ADS

An automated data selector tool that is developed to automatically select training data for *supervised classification*.

AVHRR

Advanced Very High Resolution Radiometer (AVHRR), a satellite system launched by the National Oceanic and Atmospheric Administration (NOAA)

a priori knowledge

Some *known* knowledge of the scene with respect to the identity and location of desired land categories used to support collecting training samples for each class.

Accuracy assessment

The process of quantifying the quality (accuracy) of a *classification*

Analyst

A person who performs the analyses including the *accuracy assessment*

ANN

Artificial neural network approaches simulating the working process in human mind and using a large number of small and interconnected processors or neurons work together to solve difficult classification and optimization problems

BP

Backpropagation, a typical learning algorithm for supervised *MLP* neural network

Band (image)

A sensitivity of an image in a range of the electromagnetic spectrum

Boltzmann distribution

A statistical probability distribution to characterize the *physical annealing* process of solids

Class

A category of interest for a particular application

Classification

A map composed of *thematic classes*

Class center

A typical representative for each *class*

Classified map

See "*classification*"

Classifier

A mathematical algorithm that assigns each part of an image to the appropriate *thematic class*

Clustering approaches

See *unsupervised classification*

Combinatorial optimization problem

A kind of problem to search for an optimal solution to minimize a *cost function*

Confidence interval

A statistical estimation of a parameter that provides a variance around the estimation

CG

Conjugated gradient, an advanced learning algorithm for supervised *MLP* neural network

Convex

A simple function with only one minimum

Cost function

A function defined to represent the cost of finding a solution in a particular problem

Cooling schedule

A method of determining a certain number of parameters to approximate an optimal solution in *SA*

Data layer

A particular image or a *band* of an image or other data types

Data Fusion

A technique that includes complementary information from multiple data sources

DOQQ

Digital Orthorectified Quarter Quadrangle, a USGS product based on 1:40,000 scale aerial photos with a spatial resolution of one-meter

Distance function

A function defined in unsupervised classification to indicate how close between the *class centers* and *pattern*s

Error matrix

An accuracy assessment matrix having information on categorical and overall accuracy of a thematic map as compared to reference data

ERDAS

A commercial remote sensing software package typically used for the analyses of image form data

Euclidean distance

A *distance function* without consideration of correlation

Gaussian distribution

A statistical distribution that is defined by the mean and standard deviation

Ground truth

See *reference data*

IKONOS

I-KOH-NOHS, Greek "Image". A satellite system launched to provide very high spatial resolution image data by Space Imaging, Inc.

Interpreter

A person who interprets image and *reference data* to determine the correct *classes* for the *reference sites*

ISODATA

Iterative Self-Organizing Data Analysis Technique, an unsupervised classification technique

Kappa

A discrete multivariate technique that tests whether one *error matrix* is significantly different from another

KHAT

A discrete multivariate technique used in accuracy assessment for statistically determining if one error matrix is significantly different than another

K-means

An unsupervised classification technique to naturally group high-dimensional data into clusters based on a distance function

KSOM

An unsupervised Kohonen's Self-Organizing neural network

Landsat Thematic Mapper

A satellite system launched by NASA

LU/LC

Land Use and Land Cover

Markov chain

A mathematical description of a sequence of trials where the result of each trial only depends on the result of the previous trial

Maximum likelihood

A classification technique that bases its decisions on Gaussian probability of class membership

Metropolis criterion

A rule for accepting a new solution in Simulated Annealing

Minimum distance

A classification technique that bases on the proximity of class centers and patterns

Misregistration error

An error incurred when multiple images are co-registered

MLP

A supervised Multilayer Perceptron neural network that usually uses single or multilayer perceptrons to estimate the inherent input-output relationship via a supervised *BP* training algorithm.

Multispectral

An image or sensor that has more than one spectral *band*

Multisource classification

A *classification* using multiple data sources

Neuse River Basin

One of the major hydrologic units in North Carolina

Normal distribution

See *Gaussian distribution*

Parallelpiped

A classification technique based on a simple Boolean “and/or” logic decision rule

Pattern

An object to be classified

Pixel

The smallest unit of measure of an image, or picture element

Physical annealing

A process in which the *temperature* is gradually lowering and a solid structure with the minimal energy is obtained

Point

A *sampling unit* that is a single image *pixel*

RADAR

Radio Detecting and Ranging, a form of data used in remote sensing

RADARSAT

Radio Detecting and Ranging Satellite, a Synthetic Aperture Radar (SAR) based satellite system launched by the Canadian government

Reference data

The agreed-upon correct class labels or alternatively acceptably agreed data for the *reference sites*

Reference site

See *site*

SA

Simulated Annealing, a generic *combinatorial optimization* method that is developed based on the analogy of the *physical annealing* process of solids

Sampling scheme

A statistically valid method of placing reference sites throughout the *study area* for *accuracy assessment*

Sampling units

The individual samples sites used to collect *reference data*

Site

Relating to specific sample sites on the ground rather than aggregate areas

SPOT

Satellit Pour l'Observation de la Terre, French "Satellite for Observation", a satellite system launched by the French government

Study area

A particular area of interest

Supervised classification

A *classification* that is trained using *training samples*

Temperature

A control parameter in *Simulated Annealing* based on the analogy with the *physical annealing* of solid

Thematic

Composed of *classes* that represent ground phenomena of interest

Thermal equilibrium

A stable state with the minimal energy at a given *temperature* in the *physical annealing*

Training data

Class samples used to train a *classifier*

Training sample

See *training data*

Unsupervised classification

A *classification* that finds natural groupings in multidimensional data based on measured or perceived similarities among patterns