

Development of a Simulated Annealing-assisted System for Land-use/Land-cover Classification

Cynthia F. van der Wiele (Corresponding author)

Center for Earth Observation, College of Natural Resources, North Carolina State University

Campus Box 7106, Raleigh 27695, North Carolina, USA

Tel.: 1-919-294-4492 E-mail: cynthiavanderwiele@gmail.com

Siamak Khorram

Department of Environmental Science, Policy and Management

University of California, Berkeley, 94720, USA

&

North Carolina State University, Raleigh, NC 27695-7106 (S.K.), USA

E-mail: khorrham@berkeley.edu

Hui Yuan

ERDAS Inc. Beijing, China

E-mail: Hui.Yuan@erdas.com

Received: November 28, 2011 Accepted: December 6, 2011 Published: February 1, 2012

doi:10.5539/esr.v1n1p2 URL: <http://dx.doi.org/10.5539/esr.v1n1p2>

Abstract

Local minima limitations in unsupervised approaches using K-means is still problematic in producing accurate land use/land cover classifications. In response, we developed algorithms of Simulated Annealing (SA) systems based on K-means. We hypothesized that SA-based systems can reduce the likelihood of converging on a local minimum. Two automated SA-based classification systems were developed and applied to a Landsat TM data: a single SA-based (S-SA) system and an integrated SA-based (I-SA) system, which reduces computational intensity. We hypothesized that the I-SA system could produce more efficient classifications than the S-SA system. Kappa statistical analysis on the resulting error matrices demonstrated that the SA-based system significantly improved the classification accuracy over that of the K-means algorithm when appropriate parameters combined as a cooling schedule were chosen. The knowledge and insights gained can facilitate the incorporation of SA random search procedures into other approaches that are similarly limited by local minimum problems to improve accuracy.

Keywords: Simulated annealing (SA), Land use/land cover, Classification systems, Landsat TM

1. Introduction

Remotely-sensed satellite images have long been used to generate land cover inventories. Accurately classifying these images into reliable land cover types is imperative. Many classification methods have been investigated and applied (Belward *et al.*, 1990; Benediktsson *et al.*, 1990; Bischof *et al.*, 1992; Foody, 1995; Khorram *et al.*, 2011; Ratle *et al.*, 2010; Sergi *et al.*, 1995; Serpico and Roli, 1995; Zheng *et al.*, 2008). Classification methods generally comprise two categories—supervised and unsupervised. In a supervised classification, *a priori* knowledge about each land cover type is used to support collecting training samples for each class. Training samples must be reliable and of a sufficient quantity to estimate the characteristics of classes with adequate accuracy. Frequently, however, training sample collection can be problematic due to insufficient or unreliable information available and complicated or heterogeneous landscape conditions. Alternatively, unsupervised classification requires minimal information and therefore becomes necessary when reliable information and adequate knowledge about the study

area are not available. Unsupervised classification has operational advantages over supervised methods in terms of reduced interaction time by the producer; however, it offers less direct control over the resulting classes (Thomson *et al.*, 1998).

Unsupervised classification relies on “clustering” in the spectral domain (Bandyopadhyay, Maulik and Mukhopadhyay, 2007; Maulik and Bandyopadhyay, 2003). Clustering is a generic labeling procedure designed to find natural groupings, or clusters, in multidimensional data based on measured or perceived similarities among spectral patterns (Han and Kamber, 2006; Maulik and Bandyopadhyay, 2002). Similarly, the spectral space between a pair of data samples can be determined using various distance metrics such as the Euclidean distance and then assigned to the same class (Rui and Wunsch, 2005). A desired classification based on clustering could be obtained by iteratively minimizing the total Euclidean distance between pattern vectors and their associated cluster center vector. Once clustering is completed, analysts must relate the resulting clusters to the land-cover classes on the ground (Jensen, 1996).

While a variety of clustering algorithms have been used for remotely sensed classification applications (Bedawi and Kamel, 2009; Belward *et al.*, 1990; Chen *et al.*, 2010; Guneriussen *et al.*, 1997; Palubinskas, 1999; Remund *et al.*, 1997; Thomson *et al.*, 1998; Yang *et al.*, 2010), *K*-means has been one of the most widely-used because it is computationally economical (Celik, 2009; Javed *et al.*, 2008; Tou and Gonzales, 1977; Zheng *et al.*, 2008). However, this algorithm has a serious deficiency; *K*-means is sensitive to user-defined starting conditions and when initialized with different methods, will result in different clusters. *K*-means may converge to one of the possible local minima and not necessarily to the global minimum (Javed *et al.*, 2008). As a consequence, 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 solutions to this particular classification problem (Hegarat-Masclé *et al.*, 1996; Klein and Dubes, 1989; Wang and Wang, 2009). Since many remote sensing classification applications are intricate and their distance functions may not be convex, the local minima issue often becomes more challenging. This motivated us to develop an unsupervised approach that could overcome the local minima problem and closely approximate the global or near-global optimal solution.

Simulated Annealing (SA) originated from an analogy between the physical annealing process of solids and large combinatorial optimization problems (Černý, 1985; Kirkpatrick *et al.*, 1983). Geman and Geman (1984) established that SA has the potential to find or approximate the global or near-global optimal in a combinatorial optimization problem. Although SA often requires greater computational time, the improved network could self-adapt to choose momentum parameters according to annealing temperature, thus enabling the network to escape from local minimum spots and converge stably. We hypothesized that a SA-based classification system could reduce the limitation of local minima and thereby improve the accuracy for land cover classification. All unsupervised classifiers have the advantage of being computationally more efficient, thus applicable to large land areas. This investigation explored the development of algorithms and application of SA in unsupervised classifiers for more accurate land cover classification with specific objectives to:

- 1) Investigate the applicability of two SA systems for land cover classifications using remotely sensed Landsat Thematic Mapper (TM) multispectral data;
- 2) Study the effects of controlling parameters in these SA systems on classification accuracy and efficiency and;
- 3) Evaluate and compare the resultant classifications from these SA-based algorithms versus the classical unsupervised *K*-means algorithm.

2. Simulated Annealing as Adapted for Clustering Problems

2.1 Clustering by minimizing the distance function

In most clustering problems, one begins with N patterns which are partitioned into K clusters. Algorithms that minimize the sum of squared Euclidean distance between patterns and cluster centers have many theoretical and practical advantages (Tou and Gonzales, 1977). This minimization can be expressed as the generalized least-squared error function $J(V)$:

$$J(V) = \sum_{j=1}^N d_{ij}^2 \quad (1)$$

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; $2 \leq k < N$,

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

$$v_i = \frac{\sum_{j=1}^{N_i} y_j}{N_i} \quad (2)$$

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 $\| \cdot \|$ 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 given 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 actual applications, the objective, d_{ij}^2 , is minimized for each pattern, y_j .

Based on Tou and Gonzales' (1977) minimization objective described above, the K -means algorithm (Figure 1) consists of the following steps:

- 1) Select an initial cluster configuration arbitrarily. Choose K initial cluster centers v_1, v_2, \dots, v_k randomly.
- 2) Repeat:
 - a) Distribute the pixels $\{Y\}$ among the K cluster domains, using the relation, $y \in S_i$ if $\|y - v_i\| < \|y - v_j\|$ for all $i, j = 1, 2, \dots, k, i \neq j$, where S_i denotes the set of pixels whose cluster center is v_i ;
 - b) Calculate the new cluster centers such that the sum of the squared distance $J(V)$ from all pixels in S_i to the new cluster center is minimized;
- 3) Repeat 2) until convergence conditions are satisfied, which could be defined as no change in cluster centers or as changes below a selected threshold.

The K -means clustering algorithm is computationally efficient and produces good results if the clusters are compact, well-separated in the feature space, and hyperspherical in shape when the Euclidean distance is used (Jain *et al.*, 2000). However, these conditions may not be satisfied in many complex remote sensing applications. Previous investigations have demonstrated that the classical K -means algorithm have not produced satisfactory classifications primarily due to the limitation of the local minimum (Bratsolis, 2009; Hegarat-Masclé *et al.*, 1996; Kolonko and Tran, 1997; Selim and Alsultan, 1991).

2.2 Simulated Annealing: basic principles

Simulated annealing (SA), first described by Metropolis *et al.* (1953), and independently introduced by Kirkpatrick *et al.* (1983), originated from an analogy between the physical annealing processes of solids and large combinatorial optimization problems (Černý, 1985). In the physical annealing process, if the temperature decreases sufficiently slowly, the ground state, (i.e., the minimal energy state of particles of a solid that are arranged in a highly structured lattice), can be found. The physical annealing process has been simulated in combinatorial optimization problems to search for a globally optimal configuration in order to minimize a predefined cost function. Clustering, or unsupervised classification, is one type of combinatorial optimization problem. A number of different clustering approaches have been used to perform LU/LC classifications using remotely sensed data, but were found to suffer from the local minimum problem (Hegarat-Masclé *et al.*, 1996; Klein and Dubes, 1989; Kolonko and Tran, 1997). SA has long been successfully used in various clustering applications (Brown and Huntley, 1992; Khorram *et al.*, 2011; McErlean *et al.*, 1990; Selim and Alsultan, 1991), making it worthwhile research to develop and investigate SA-based classification systems for LU/LC classification using remotely sensed data due to being able to overcome the local minimum problem (Dai and Khorram, 1999; Geman and Geman, 1984; Pang, Chen and Chen, 2006; Yuan, Van Der Wiele, and Khorram, 2009).

Mathematically, SA can be modeled by means of a Markov chain (Feller, 1950). The basic procedure involves a cooling schedule, in which the temperature parameter T , starts out sufficiently high and is gradually lowered in a given schedule to minimize the energy or cost function associated with a specific problem formulation. At each temperature T , a small, randomly generated perturbation is repetitively applied until the system reaches thermal equilibrium. Then the algorithm moves to the next temperature in the given schedule. The rule of accepting a

perturbation is based upon the *Metropolis criterion* (Metropolis *et al.*, 1953): 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* (Geman and Geman, 1984; Laarhoven, 1988).

2.3 Adaptation of simulated annealing for clustering issues

To adapt SA to clustering problems, the energy function is defined by Equation 1. We started with a high controlling temperature denoted as T_0 and an arbitrary initial clustering assignment. For a given randomly selected pattern, y_j , we reassigned 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 , were respectively computed. We accepted the new assignment either with a distance decrease $\Delta d_{mn} < 0$, 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 was considered as a “transition”. The adapted SA algorithm (Figure 2) is denoted as the single SA-based (S-SA) algorithm in the following sections.

2.4 Definition of parameters in simulated annealing

The SA algorithm is quite straightforward to implement. However, the quality of the outcome depends on the convergence of the algorithm, which is governed by the cooling schedule. Cooling schedule parameters should be 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 can be difficult as the “best” choice will be different for each application. The investigation for an adequate cooling schedule is still a work in progress and has long been a predominant theme (Aarts and Korst, 1989; Aarts and Laarhoven, 1987; Chainate *et al.*, 2008; Collins *et al.*, 1987; Laarhoven, 1988; Prajapati *et al.*, 2009). Accordingly, the following parameters were specified and chosen for their relative simplicity and expedient computation:

- *The initial value of the temperature, T_0 .* 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 must decrease sufficiently slowly. Convergence in probability to the global minimum has only been proven for the logarithmic annealing schedule (Geman and Geman, 1984): $T_n = \frac{1}{\ln n}$. Although this inverse-log schedule guarantees the global convergence for a wide class of combinatorial problems, it is far too slow for almost any practical application. The 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 (Aarts and Korst, 1989; Burkard and Rendl, 1984; Hegarat-Masclé and Olivier, 1996; Laarhoven, 1988; Lundy and Mees, 1986).
- *The final value of the temperature T_{final} .* This value is used to determine when to discontinue the annealing process. In our study, we set T_{final} as a number approaching zero. Via an appropriate cooling schedule, while T_n is approaching T_{final} , a configuration with minimal energy should be targeted.
- *A finite number of accepted transitions at each temperature T_n .* This factor determines whether the thermal equilibrium at each temperature, a critical 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 an image is scanned 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 during 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 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 together determine the length of the homogeneous Markov chain at each T_n .

2.5 Integration of the SA and K-means for clustering refinement

Although straightforward to implement, SA is computationally demanding, which has been one of the primary barriers to its widespread application in remote sensing applications using large data sets. It is therefore necessary to investigate ways of implementing SA more efficiently. We believe that if the initial clustering is as close to the optimal clustering as possible, then the search for the global optimum will be both expeditious 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 more rapid cooling schedule. Although K-means may not be able to find the global optimum for a specified problem, the resulting clustering solution 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 more efficiently. To some degree, the I-SA system keeps the random search capability better than SA; therefore reducing the chance of committing to undesirable local minima.

3. Applications to Land-use/Land-cover Classification

3.1 Description of data sets and classification scheme

We developed three unsupervised classification systems 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 (Bai *et al.*, 2009; Dai and Khorram, 1999; Khorram *et al.*, 1987; Vogel and Strohbach, 2009). The TM image, shown in Figure 3, covers an area of 2,856.15 ha. The study area was located near the Raleigh-Durham International Airport in North Carolina [USA]—an area of familiarity, less complexity, and nearby proximity to the researchers. The TM image shown in Figure 4 is 7,831.89 ha and is located in Georgia [USA] near Macon. For reference convenience in the following sections, the first TM image will be denoted as TM-1 and the second one as 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 an additional 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, it was easy to determine that five land cover classes were dominant: urban, mixed forest, evergreen forest, grassland/agriculture, and water. The TM-2 image was much more heterogeneous than TM-1. 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 (Reflected Green Band), band3 (Reflected 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 selection of these bands was due to the reflective properties of the features and the land use land cover characteristics of the study areas.

The classification applications using TM-1 and TM-2 were examined to evaluate the wide applicability and potential of the SA-based systems and their reliability in applications with varying classification complexities. In order to evaluate the performances, the resulting classifications were first evaluated by its mathematical performance index $J(V)$. Based on the aforementioned clustering formulation, our minimization objective was to minimize the sum of squared Euclidean distance $J(V)$. The resulting classified map with lowest $J(V)$ was more accurate than others. In order to confirm the more accurate results being produced by the lowest $J(V)$, the classifications results were selected from both algorithms and tested against the reference data that were manually derived from aerial photos, the original TM images, and other ancillary data.

3.2 Analysis of the selection and roles of parameters in SA

A large number of experiments using various parameter candidates were conducted by applying the S-SA system to image TM-1 in order to assess the role of the aforementioned parameters. 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 run 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 was allowed to change among the selected range while all other parameters were fixed. For example, in Figure 5, each data group represented 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 Figures 5–8; the studied parameter versus computational time (in seconds), are depicted in Figures 9 through 12.

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 assure all transitions would be accepted at the beginning of clustering, while the lowest T_0 candidate (10) can only accept a small number of transitions. The decrement factor μ was used to decrease the temperature in the cooling schedule by multiplying it with the current temperature. The SA implementation required that temperature decreases sufficiently slowly but there was no criterion to specify precisely how slow the decrease ought 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 was no clear evidence from Figure 5 to indicate that the magnitudes of T_0 made a significant difference in this clustering application in terms of reducing $J(V)$. Lower initial values like 10, 100 may also be applicable. This might attributed to selected T_0 candidates being too high for this particular application. From Figure 6, we observed that comparatively the slowest decrement factor μ (0.99) had the most stable performance, but there was 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 was combined 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 were essential to the convergence of SA. The GP was 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 would be accepted at each temperature. From the behaviors of the five groups of data in Figure 7, it was found that there were no significant differences among the four IET candidates in all groups except the Group I. Figure 8 illustrates 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 were studied as well using the same graph method. Figure 9 to Figure 12 illustrates that 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 were the factors that escalated 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, 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 to 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. We recommend the future investigators expand on this optimization, because we ran out of the scope of this project and left this further optimization to those who are interested in the follow up of this research.

4. Classification Results and Analysis

The classification results of the three clustering algorithms using two TM images: TM-1 and TM-2 are presented in this section. A number of experiments were performed using either different initializations in *K*-means or

different parameter combinations in the two SA-based algorithms. The classification result with the lowest $J(V)$ from each algorithm was selected for further accuracy assessment.

The classified maps generated from TM-1 are shown in Figure 13. The SA-related parameters and their performance in terms of $J(V)$ and computational time are shown in Table 1. 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 computationally-efficient algorithm. To validate the three classified maps in terms of land-cover/land-use classification accuracy, 253 sample points were randomly selected and visually interpreted based on the reference data derived from false composite TM-1 imagery and aerial photos. The resulting error matrices are presented in Tables 2 through 4.

Among the three classified maps, the highest overall accuracy (91.3%) was obtained by the I-SA algorithm. The Kappa analysis was used to statistically determine if one of the derived error matrices was statistically significantly different than another (Bishop *et al.*, 1975). The KHAT values shown in Table 5 demonstrate that there are strong agreements between the classification results and the reference data for all of these three error matrices. From the Z-tests shown in Table 6, it can be concluded 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 classification systems.

Figure 14 depicts the classified maps generated using TM-2. Their SA-related parameters and the performance indexes are presented in Table 7. To conduct the accuracy assessments, 299 sample points were randomly selected from TM-2 and visually interpreted. The resulting error matrices are presented in Tables 8 through 10. The highest overall accuracy, 75.59%, was obtained by the S-SA algorithm. The Kappa analysis shown in Tables 11 and 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.

By analyzing the results, the classification performances of the three clustering algorithms using TM-1 and TM-2 are clearly different in several aspects. First, the classification accuracy in TM-1 was relatively high and satisfactory while the classification accuracy in TM-2 was moderate. Second, the I-SA algorithm in TM-1 had the best classification performance while its performance in TM-2 was poor with the S-SA algorithm producing the best classification. Third, the resulting SA-related classification improvement in TM-2 was more significant than that in TM-1. We believe that these differences may be due to differences in the complexity of data structures and distributions between TM-1 and TM-2. Visually, we observed that the landscape condition in TM-1 was less complex than TM-2 case. Five typical land cover classes can be easily differentiated. From the distribution plots shown in Figure 15, it is further demonstrated that the five classes in TM-1 are spectrally compact and well-separated. The comparatively homogeneous study area reduced the classification complexity and contributes to achieving relatively high classification accuracy in TM-1 classification. The landscape conditions in TM-2 were more complex than TM-1 because the study area was more heterogeneous and contained more LU/LC classes. The spectral complexity in TM-2 was demonstrated as a more scattered distribution among the seven classes shown in Figure 16.

The increase in spectral complexity explains why the accuracy of resulting classification of TM-2 was lower than that of TM-1. In a more complex classification case like TM-2, the classifier often suffers more local minima than a less complex 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 not performed as well in TM-2. In a complex 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.

In both cases—TM-1 and TM-2—the I-SA or the S-SA improved the classification as compared to *K*-means. This verified our assumption that the SA-based algorithms are able to overcome the local minimum limitations of *K*-means and improve the classification accuracy. Furthermore, *K*-Means algorithm is too restrictive to be optimized. 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 complex classification applications.

5. Discussions 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 were investigated and adapted to solve the local minimum problem in unsupervised LU/LC classification, thereby improving classification accuracy. Two

SA-based unsupervised classification systems were developed. Their applicability and implementations for land use and land cover classifications 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 were found to be essential for the success of the SA-based systems: the complexity of the application and 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. Decision boundaries are often difficult to find in many applications because the search is often complex 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 less complex applications. Our experimental analysis suggests that, for simpler applications, the I-SA system is a good option because it is likely to produce better classification more efficiently. However, for more intricate applications, (e.g., either when more land classes are desired or the landscape conditions are very heterogeneous), the I-SA system may not produce improved results, in which case the S-SA system should then be used.

Theoretically, SA has been proven to be able to provide the global optimal solution in clustering problems albeit with a long run time. In practice, the globally optimizing nature of SA can be taken advantage of in order to improve the classification accuracy and reduce the uncertainties for land cover classification. This can be accomplished with appropriate selections of initial temperature and cooling schedule. K -means can be implemented efficiently and produce satisfactory results if classified patterns contain well-separated clusters such as hyper-spectral data. The performance of SA-based systems is sensitive to the choices of controlling parameters. If appropriate parameters are chosen, the resultant classification can be improved significantly. The effective implementation of the SA-based systems in classification problems is still a substantial issue for which there is no definitive analytical solution to date. Based on our findings, we reached the following conclusions:

- 1) The selection of each individual parameter in the cooling schedule may not be as critical. It may be more important to select an optimal parameter combination for accurate and efficient classification using SA. The speed of the cooling schedule should be slowed when 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 less complex applications. As application complexity increases, T_0 must also be increased.
- 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 recommended 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. That is, greater numbers of IET and lower GP result in a longer Markov Chain. There is a trade-off between large decrements μ and small IET and higher GP. One could either choose smaller μ with greater IET and lower GP, or larger μ with smaller IET and larger GP.

6. Recommendations for Future Research

As we indicated in part of our previous findings (Khorram *et al.*, 2011), SA-based systems have been studied for their potential for improving classification accuracies in LU/LC classification applications using 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 schemes as well as using many other remotely sensed data. Datasets such as SPOT should be tested for similar applications. More importantly, we believe that application of SA classification techniques could be very useful when applied to high resolution satellite data such as IKONAS and Quickbird. There may be merit in testing the applicability of SA to hyper-spectral data. Additional classification applications would be valuable to provide more evidence 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.

Finally, it is recommended that multiple study sites with different complexities and landscape characteristics during two dates in a year (growing season and leaf off season), could eliminate some of the issues we encountered such as the complexity of the landscape and the seasonal variability of the vegetative cover.

References

- Aarts, E. H. L., & Korst, J. (1989). *Simulated annealing and Boltzmann machines: a stochastic approach to combinatorial optimization and neural computing*. Hoboken, NJ: John Wiley & Sons, 272 pp.
- Aarts, E. H. L., & van Laarhoven, P. J. M. (1987). Simulated annealing: a pedestrian review of the theory and some applications. In P. A. Devijver, J. Kittler (Eds.), *Pattern Recognition, theory and applications*. (Proceedings NATO Advanced Study Institute, Spa-Balmoral, Belgium, June 9-20, 1986), (NATO ASI Series, Series F: *Computer and Systems Sciences*, 30, 179-192). Berlin: Springer.
- Anderson, J. R., Hardy, E. E., Roach, J. T., *et al.* (1976). A land use and land cover classification system for use with remote sensor data. *U.S. Geological Survey Professional Paper*, 964, 1-28. Washington, D.C.: USGS.
- Bai, M., Lui, H., Huang, W., *et al.* (2009). Automated farmland extraction from multi-temporal Landsat TM data based on artificial neural network. *Geoinformatics, 2009 17th International Conference on*, 12-14 August 2009, Fairfax, VA, USA. Pp. 1-4.
- Bandyopadhyay, S., Pal, S. K., & Murthy, C. A. (1998). Simulated annealing based pattern classification. *Information Sciences*, 109(1-4), 165-184.
- Bedawi, S. M., & Kamel, M. S. (2009). A comparative study of clustering methods for urban areas segmentation from high resolution remote sensing image. *Intelligent Systems Design and Applications, ISDA '09. Ninth International Conference on*, 30 November-2 December 2009, Pisa, Italy. Pp.169-174.
- Belward, A. S., Taylor, J. C., Stuttard, M. J., *et al.* (1990). An unsupervised approach to the classification of semi-natural vegetation from Landsat Thematic Mapper data. *Int J Remote Sens.*, 11, 429-445. <http://dx.doi.org/10.1080/01431169008955031>
- Benediktsson, J. A., Swain, P. H., & Ersoy, O. K. (1990). Neural network approaches versus statistical methods in classification of multi-source remote sensing data. *IEEE Trans Geosci Remote Sens.*, 28, 540-551. <http://dx.doi.org/10.1109/TGRS.1990.572944>
- Bischof, H., Schneider, W., & Pinz, A. J. (1992). Multispectral classification of Landsat images using neural networks. *IEEE Trans Geosci Remote Sens.*, 30, 482-490. <http://dx.doi.org/10.1109/36.142926>
- Bishop, Y., Fienberg, S., & Holland, P. (1975). *Discrete multivariate analysis – theory and Practice*. Cambridge, MA: MIT Press pp.575.
- Bratsolis, E. (2009). Unsupervised segmentation of agricultural regions using TerraSAR-X images. *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*, 12-17 July 2009, Cape Town, South Africa, pp. 3, III-416-III-419.
- Brown, D. E., & Huntley, C. L. (1992). A practical application of simulated annealing to clustering. *Pattern recognition.*, 25(4), 401-412. [http://dx.doi.org/10.1016/0031-3203\(92\)90088-Z](http://dx.doi.org/10.1016/0031-3203(92)90088-Z)
- Burkard, R. E., & Rendl, F. (1984). A thermodynamically motivated simulation procedure for combinatorial optimization problems. *European Journal of Oper. Res.*, 17(2), 169-174. [http://dx.doi.org/10.1016/0377-2217\(84\)90231-5](http://dx.doi.org/10.1016/0377-2217(84)90231-5)
- Celik, T. (2009). Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *IEEE Geoscience and Remote Sensing Letters*, 6(4), 772-776. <http://dx.doi.org/10.1109/LGRS.2009.2025059>
- Černý, V. (1985). Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. *J Optim Theory Appl.*, 45, 45 –51.
- Chen, Y., Bruzzone, L., Fengyue, S., *et al.* (2010). A fuzzy-statistics-based affinity propagation technique for clustering in multispectral images. *IEEE Trans Geosci Remote Sens.*, 48, 2647-2659. <http://dx.doi.org/10.1109/TGRS.2010.2040035>
- Dai, X., & Khorram, S. (1999). Data fusion using artificial neural networks: a case study on multitemporal change analysis. *Comput., Environ., and Urban Systems.*, 23, 19-31. [http://dx.doi.org/10.1016/S0198-9715\(98\)00051-9](http://dx.doi.org/10.1016/S0198-9715(98)00051-9)

- Foody, G. M., McCulloch, M. B., & Yates, W. B. (1995). Classification of remotely sensed data by an artificial neural network: issues related to training data characteristics. *Photogramm Eng Remote Sensing*, 61, 391-401.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell.*, 6, 721-741. <http://dx.doi.org/10.1109/TPAMI.1984.4767596>
- Granville, V., Krivanek, M., & Rasson, J-P. (1994). Simulated annealing: a proof of convergence. *IEEE Trans Pattern Anal Mach Intell.*, 16, 652-656. <http://dx.doi.org/10.1109/34.295910>
- Guneriusen, T., Johnsen, H., Solberg, R., *et al.* (1997). Snow monitoring using EMISAR and ERS-1 data within the European multi-sensor airborne campaign EMAC-95. *Geosc. and Remote Sens.*, 2, 631-633.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (2nd ed.). Amsterdam: Elsevier.
- Hegarat-Masclé, L., Vidal-Madjar, S. D., & Olivier, P. (1996). Applications of simulated annealing to SAR image and classification problems. *Int J Remote Sens.*, 17, 1761-1776. <http://dx.doi.org/10.1080/01431169608948738>
- Javed, K., Babri, H. A., & Saeed, M. (2008). The behavior of k-means: an empirical study. *Electrical Engineering, ICEE 2008, Second International Conference on*, 25-26 March 2008, Lahore, UK, pp 1-6.
- Jenson, J. R. (1996). *Introductory digital image processing*. Englewood Cliffs, NJ: Prentice-Hall.
- Khorram, S., Brockhaus, J. A., & Cheshire, H. M. (1987). Comparison of Landsat MSS and TM data for urban land-use classification. *IEEE Trans Geosci Remote Sens.*, 25, 238-243. <http://dx.doi.org/10.1109/TGRS.1987.289823>
- Khorram, S., Van Der Wiele, C. F., & Yuan, H. (2011). Development of a modified neural network-based land cover classification system using automated data selector and multi-resolution remotely sensed data. *Geocarto International*, 26(6), 435-457. <http://dx.doi.org/10.1080/10106049.2011.600462>
- Kirkpatrick, S., Gelatt, C. D., Jr., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220, 671-688. <http://dx.doi.org/10.1126/science.220.4598.671>
- Klein, R. W., & Dubes, R. C. (1989). Experiments in projection and clustering by annealing. *Pattern Recognition*, 22, 213-220. [http://dx.doi.org/10.1016/0031-3203\(89\)90067-8](http://dx.doi.org/10.1016/0031-3203(89)90067-8)
- Kolonko, M., & Tran, M. T. (1997). Convergence of simulated annealing with feedback Temperature schedules. *Probability in the Engineering and Informational Sciences*, 11, 279-304. <http://dx.doi.org/10.1017/S0269964800004836>
- Leong, H. W., Wong, D. F., & Liu, C. L. (1985). A simulated-annealing channel router. *Proc. ICCAD*. Santa Clara, CA. 226-229.
- Locatelli, M. (1996). Convergence properties of simulated annealing for continuous global optimization. *J Appl Prob Stat.*, 33, 1127-1140. <http://dx.doi.org/10.2307/3214991>
- Lundy, M., & Mees, A. (1986). Convergence of an annealing algorithm. *Math. Prog.*, 34, 111-124. <http://dx.doi.org/10.1007/BF01582166>
- Maulik, U., & Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. *IEEE Trans Pattern Anal Mach Intell.*, 24, 1650-1654. <http://dx.doi.org/10.1109/TPAMI.2002.1114856>
- Maulik, U., & Bandyopadhyay, S. (2003). Fuzzy partitioning using a real-coded variable-length genetic algorithm for pixel classification. *IEEE Trans Geosci Remote Sens.*, 41, 1075-1081. <http://dx.doi.org/10.1109/TGRS.2003.810924>
- McErlean, F. J., Bell, D. A., & McClean, S. I. (1990). The use of simulated annealing for clustering data in databases. *Information Systems*, 15, 233-245. [http://dx.doi.org/10.1016/0306-4379\(90\)90037-P](http://dx.doi.org/10.1016/0306-4379(90)90037-P)
- Metropolis, N. (1953). Equation of state calculations by fast computing machines. *J Chem Phys.*, 21, 1087-1092. <http://dx.doi.org/10.1063/1.1699114>
- Palubinskas, G., Datcu, M., & Pac, R. (1999). Clustering algorithms for large sets of heterogeneous remote sensing data. *Geoscience and Remote Sensing Symposium, 1999, IGARSS '99 Proceedings, IEEE 1999 International*, 28 June-02 July 1999, Hamburg, Germany. 3, 1591-1593.
- Pang, X., Chen L., & Chen, W. (2006). Application of neural network based on simulated annealing to classification of remote sensing image. *Proceedings of the 6th World Congress on Intelligent Control and Automation* 21-23 June 2006, Dalian, China. 2874-2877.

- Ratle, F., Camps-Valls, G., & Weston, J. (2010). Semisupervised neural networks for efficient hyperspectral image classification. *IEEE Trans Geosci Remote Sens.*, 48(5), 2271-2282. <http://dx.doi.org/10.1109/TGRS.2009.2037898>
- Remund, Q. P., Long, D. G., & Drinkwater, M. R. (1998). Polar sea-ice classification using enhanced resolution NSCAT data. *Geoscience and Remote Sensing Symposium Proceedings, 1998. IGARSS '98. 1998 IEEE International.*, 06-10 July, 1998, Seattle, WA, USA. 4, 1976-1978.
- Rose, K. (1998). Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 86(11), 2210-2239. <http://dx.doi.org/10.1109/5.726788>
- Rui, X., & Wunsch, D., II. (2005). Survey of Clustering Algorithms. *IEEE Trans Neural Networks*, 16, 645-678. <http://dx.doi.org/10.1109/TNN.2005.845141>
- Selim, S. Z., & Alsultan, K. (1991). A simulated annealing algorithm for the clustering problem. *Pattern Recognition*, 24, 1003-1008. [http://dx.doi.org/10.1016/0031-3203\(91\)90097-O](http://dx.doi.org/10.1016/0031-3203(91)90097-O)
- Sergi, R., Solaiman, B., Mouchot, M. C., *et al.* (1995). LandsatTM images classification using principal components analysis and neural networks. *Geoscience and Remote Sensing Symposium, Proceedings of IGARSS 1995*, 1-14 July 1995, Firenze, Italy. *Quantitative remote sensing for science and applications*, 3, 1927-1929.
- Serpico, S. B., & Roli, F. (1995). Classification of multisensor remote-sensing images by structured neural networks. *IEEE Trans Geosci Remote Sens.*, 33, 562-578. <http://dx.doi.org/10.1109/36.387573>
- Thomson, A. G., Fuller, R. M., & Eastwoods, J. A. (1998). Supervised versus unsupervised methods for classification of coasts and river corridors from airborne remote sensing. *Int J Remote Sens.*, 18, 3423-3431. <http://dx.doi.org/10.1080/014311698214091>
- Tou, J. T., & Gonzales, R. C. (1977). *Pattern Recognition Principles*. Reading, MA: Addison- Wesley Publishing Co.
- Van Laarhoven, P. J. M. (1988). *Theoretical and computational aspects of simulated annealing*. Center for mathematics and computer science, Amsterdam, Netherlands. P. 33.
- Vogel, M., & Strohbach, M. (2009). Monitoring of savanna degradation in Namibia using Landsat TM/ETM+ data. *Geoscience and Remote Sensing Symposium, IEEE International, IGARSS '09, 13-17 July 2009, Cape Town, South Africa*, III-931-III-934.
- Wang, S., & Wang, X. (2009). Local search algorithm for K-means clustering based on minimum sub-cluster size. *Pattern Recognition. CCPR 2009. Chinese Conference on*, 4-6 November 2009, Nanjing, China. 1-4.
- White, S. R. (1984). Concepts of scale in simulated annealing. *Proc. IEEE ICCD, Port Chester, NY*, pp. 646-651.
- Yang, W., Zeng, Z., & Zhang, S. (2010). Application of combining watershed and fast clustering method in image segmentation. *Computer Modeling and Simulation, ICCMS, 2nd International Conference on*, 22-24 January 2010, Sanya, China. 3, 170-174.
- Yuan, H., Van Der Wiele, C. F., & Khorram, S. (2009). An automated artificial neural network system for land use/land cover classification from Landsat TM imagery. *Remote Sensing*, 1, 243-265. <http://dx.doi.org/10.3390/rs1030243>
- Zheng, J., Cui, Z., Lui, A., *et al.* (2008). A K-means remote sensing image classification method based on AdaBoost. *Natural Computation, ICNC 2008, Fourth International Conference on*, 18-20 October 2008, Jinan, China. 4, 27-32.

Table 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 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. 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 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 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 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			

Table 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 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										
Classified Image		C1	C2	C3	C4	C5	C6	C7	Classified Totals	Users' Accuracy
	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 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										
Classified Image		C1	C2	C3	C4	C5	C6	C7	Classified Totals	Users' Accuracy
	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 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 = $199/299 = 66.56\%$										

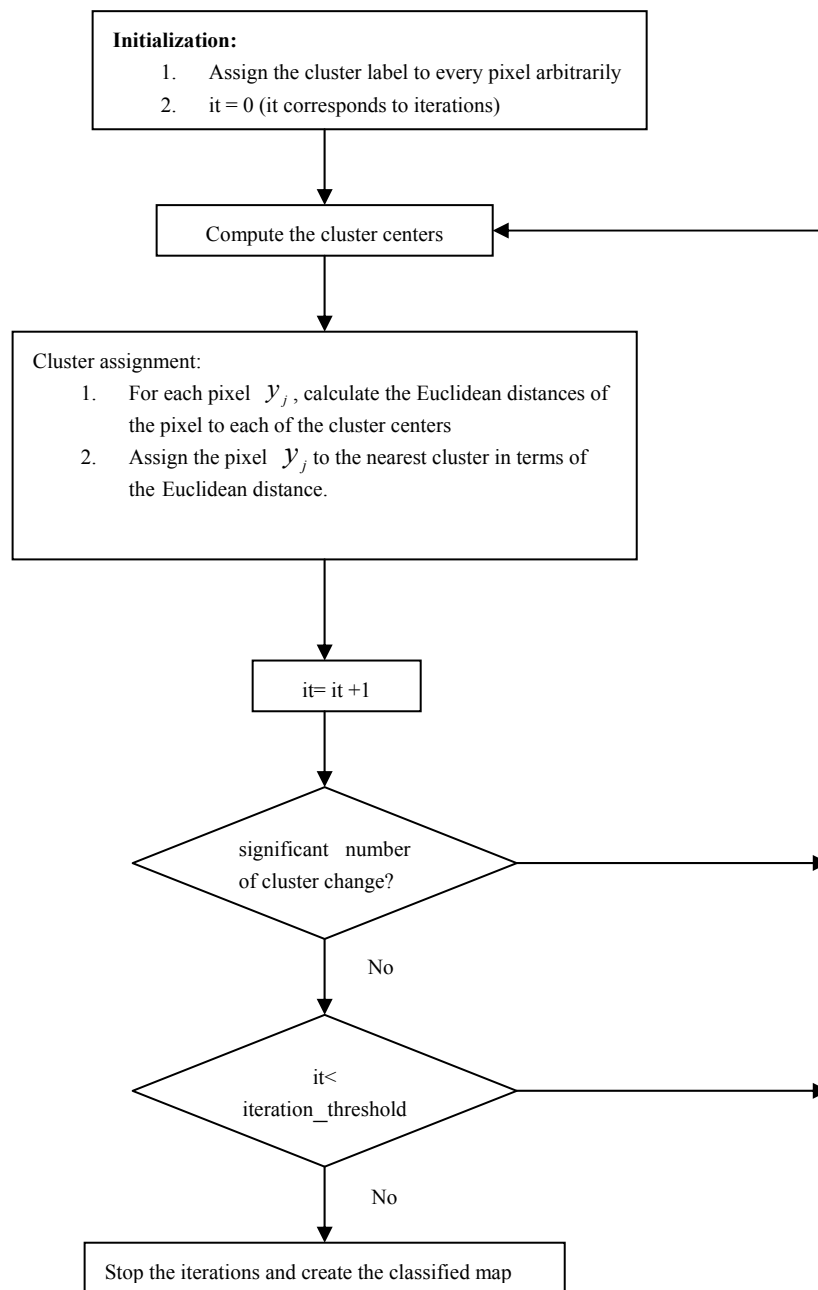


Figure 1. The flowchart of K-means clustering algorithm

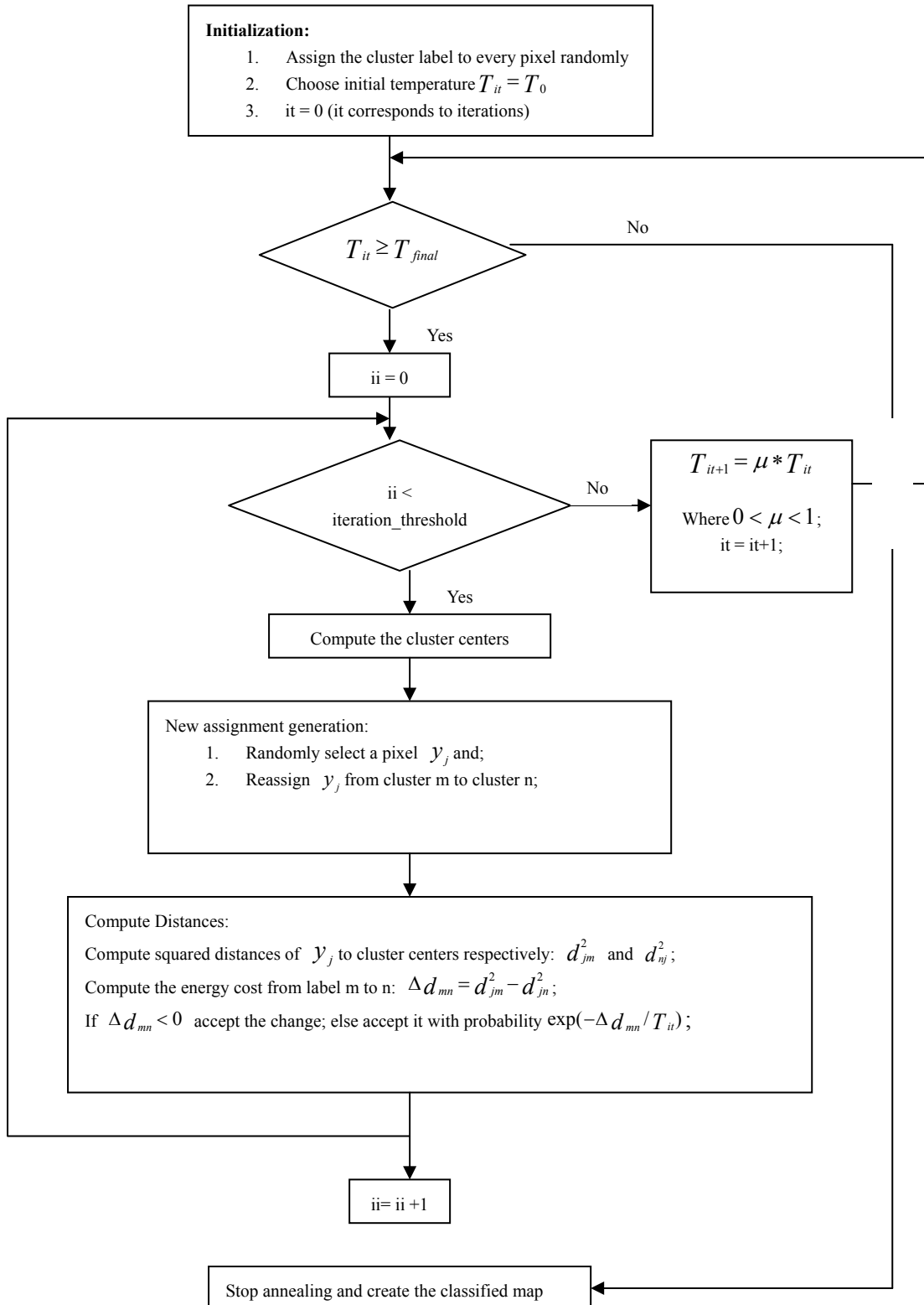


Figure 2. The adapted single SA-based clustering algorithm

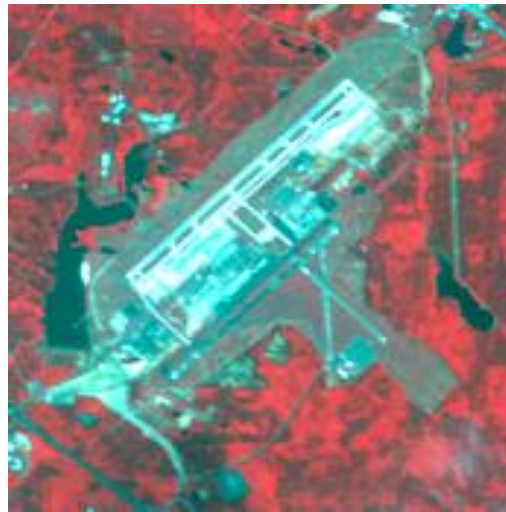


Figure 3. Original Landsat TM-1



Figure 4. Original Landsat TM-2

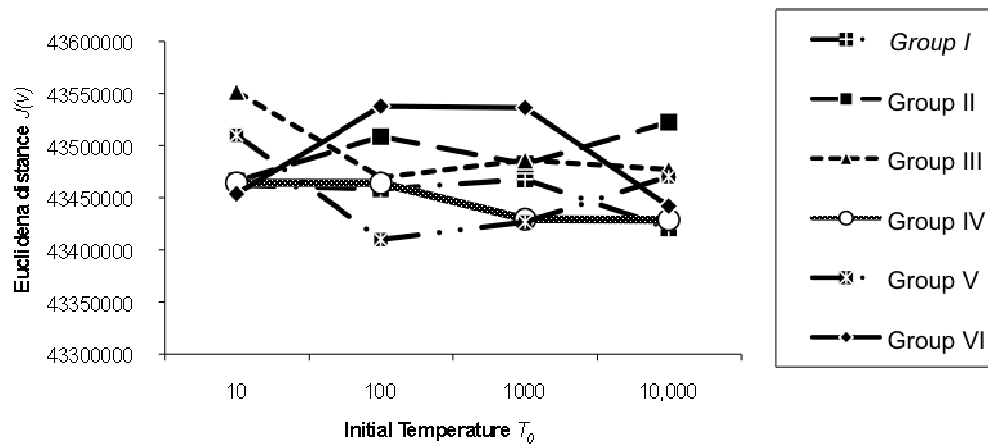


Figure 5. Effect of T_0 on the magnitude of $J(V)$

Group I: $\mu = 0.99$, IET = 20, GP = 0.90

Group II: $\mu = 0.90$, IET = 20, GP = 0.90

Group III: $\mu = 0.85$, IET = 20, GP = 0.90

Group IV: $\mu = 0.80$, IET = 20, GP = 0.90

Group V: $\mu = 0.90$, IET = 20, GP = 0.85

Group VI: $\mu = 0.90$, IET = 40, GP = 0.90

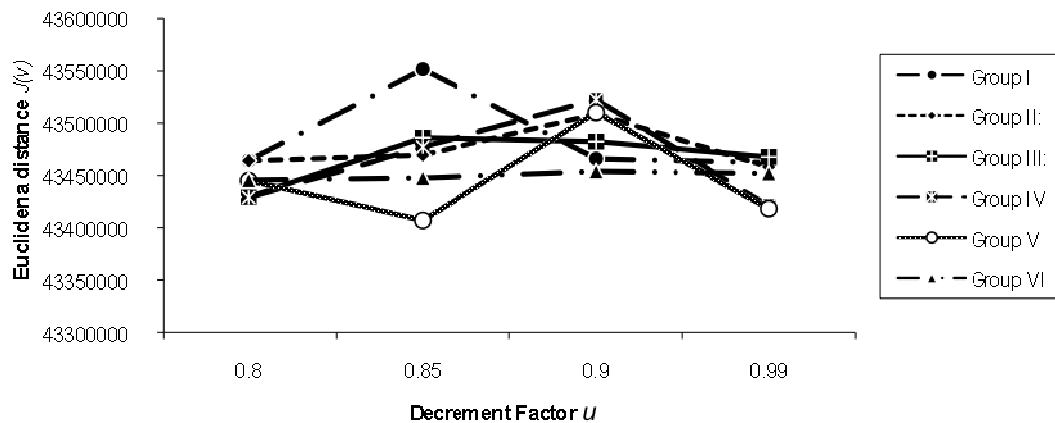


Figure 6. Effect of μ on the magnitude of $J(V)$

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

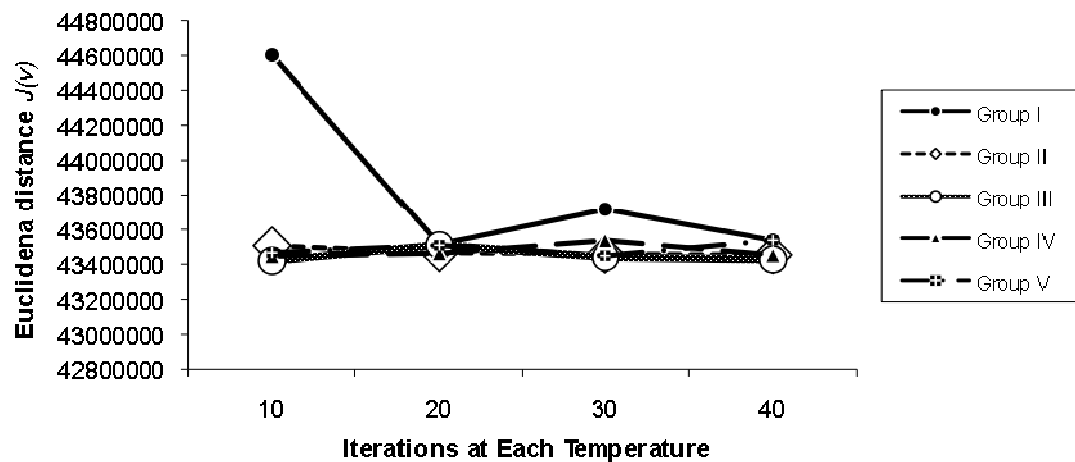


Figure 7. Effect of the IET on magnitude of $J(V)$

Group I: $T_0 = 10$, $\mu = 0.90$, GP = 0.95

Group II: $T_0 = 10$, $\mu = 0.90$, GP = 0.90

Group III: $T_0 = 10$, $\mu = 0.90$, GP = 0.85

Group IV: $T_0 = 10$, $\mu = 0.99$, GP = 0.90

Group V: $T_0 = 100$, $\mu = 0.90$, GP = 0.90

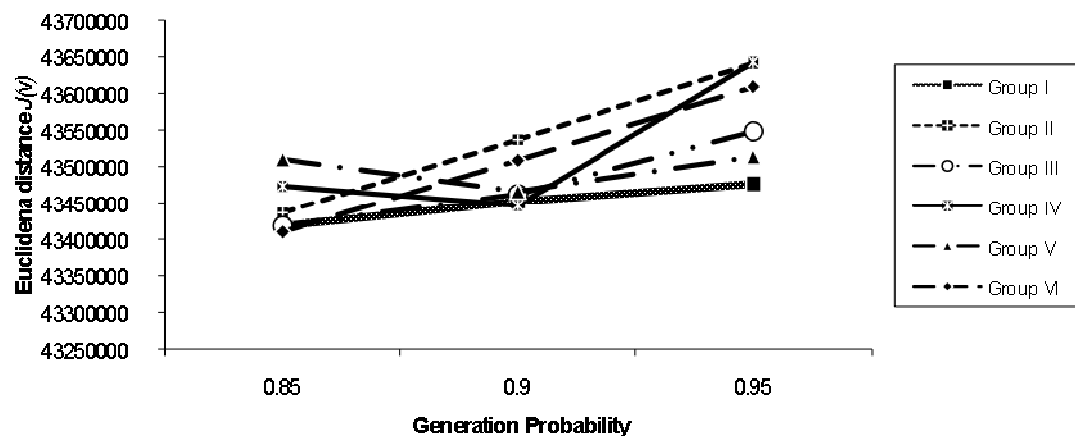


Figure 8. Effect of the GP on magnitude of $J(V)$

Group I: $T_0 = 10$, $\mu = 0.99$, IET = 40

Group II: $T_0 = 10$, $\mu = 0.99$, IET = 30

Group III: $T_0 = 10$, $\mu = 0.99$, IET = 20

Group IV: $T_0 = 10$, $\mu = 0.99$, IET = 10

Group V: $T_0 = 10$, $\mu = 0.90$, IET = 20

Group VI: $T_0 = 100$, $\mu = 0.90$, IET = 20

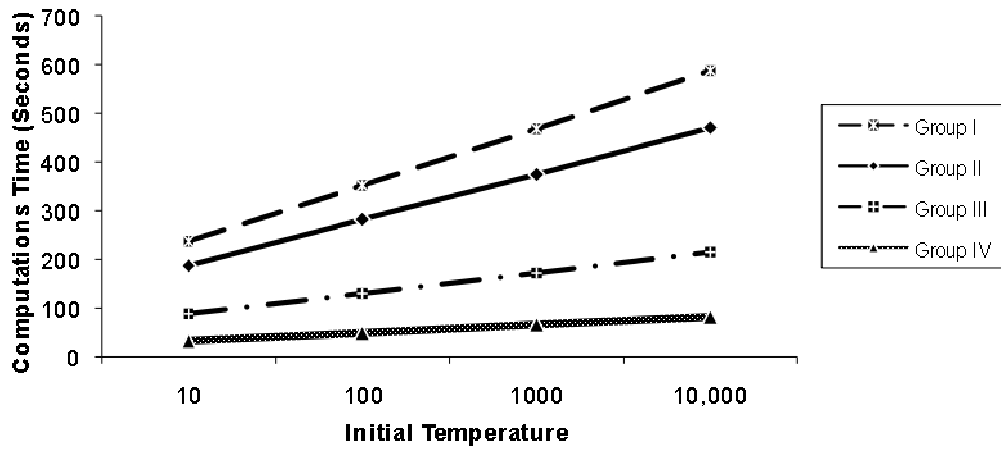


Figure 9. Effect of T_0 on the Computational Time

Group I: $\mu = 0.90$, IET = 40, GP = 0.95

Group II: $\mu = 0.90$, IET = 30, GP = 0.90

Group III: $\mu = 0.85$, IET = 20, GP = 0.85

Group IV: $\mu = 0.80$, IET = 10, GP = 0.85

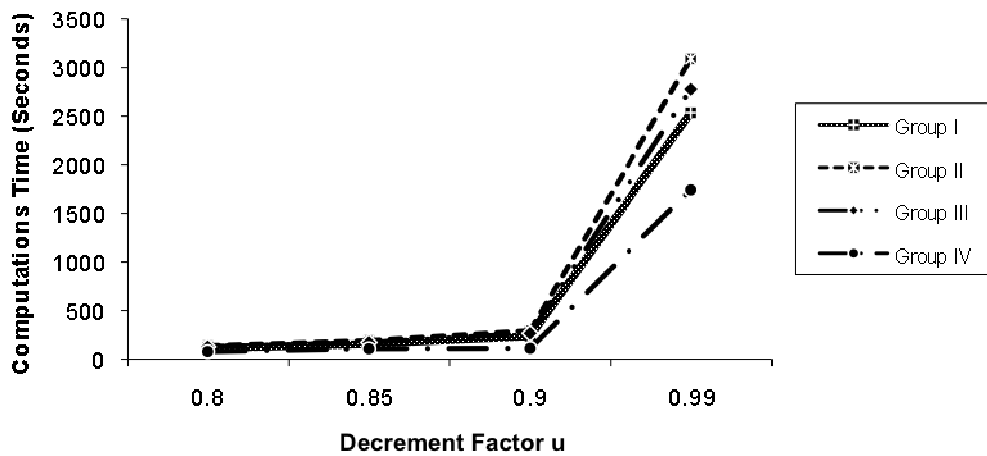


Figure 10. Effect of μ on the Computational Time

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

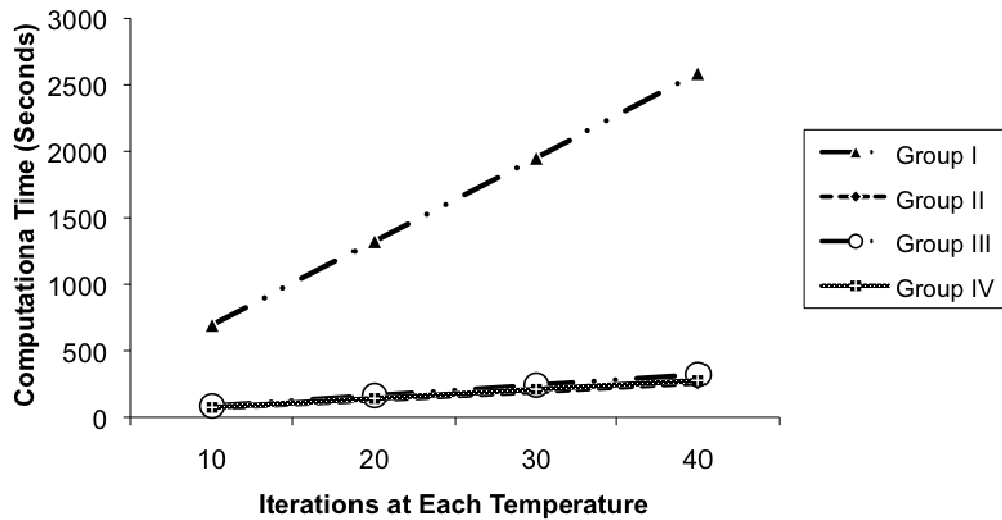


Figure 11. Effect of the IET on the Computational Time

Group I: $T_0 = 10$, $\mu = 0.99$, GP = 0.90

Group II: $T_0 = 100$, $\mu = 0.85$, GP = 0.85

Group III: $T_0 = 1000$, $\mu = 0.85$, GP = 0.90

Group IV: $T_0 = 10,000$, $\mu = 0.80$, GP = 0.95

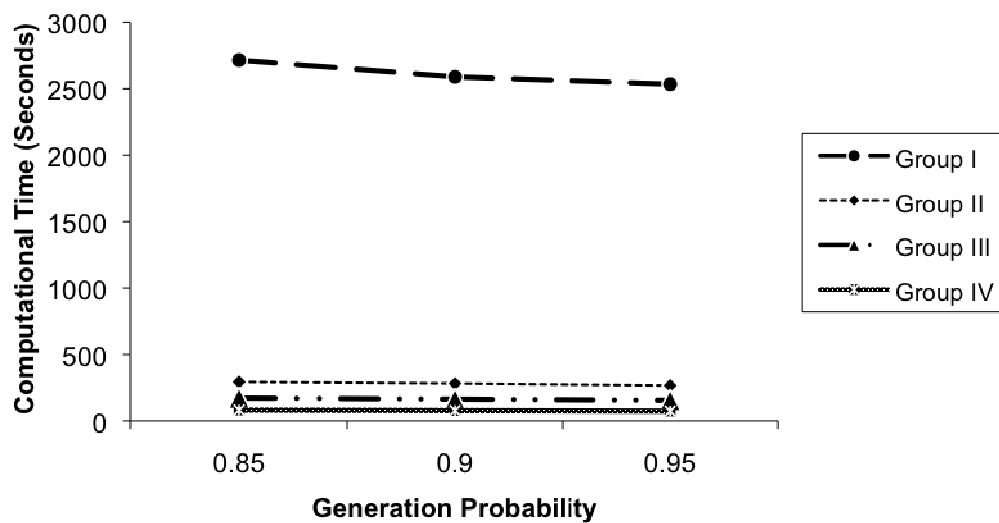


Figure 12. Effect of the GP on the Computational Time

Group I: $T_0 = 10$, $\mu = 0.99$, IET = 40

Group II: $T_0 = 100$, $\mu = 0.90$, IET = 30

Group III: $T_0 = 1000$, $\mu = 0.85$, IET = 20

Group IV: $T_0 = 10,000$, $\mu = 0.80$, IET = 10

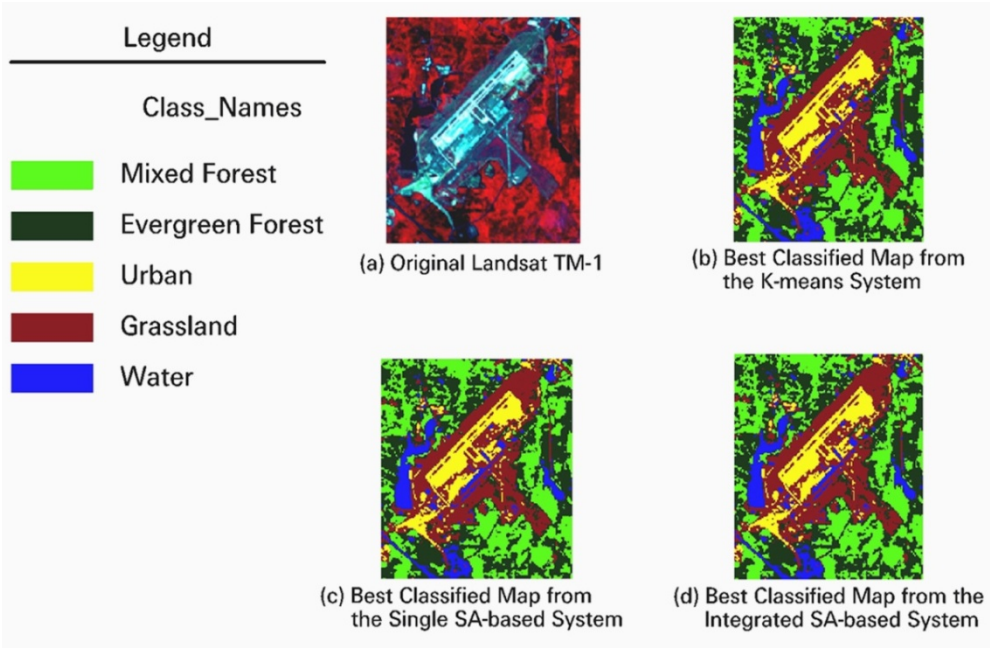


Figure 13. The original image TM-1 and the classified maps

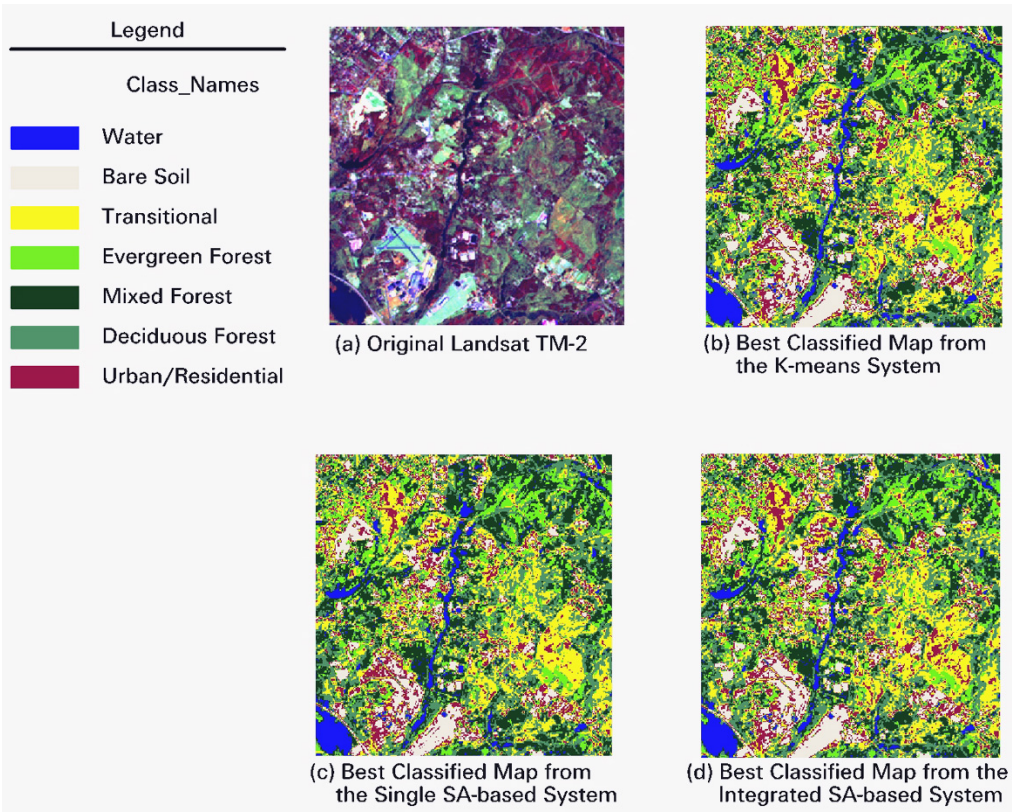
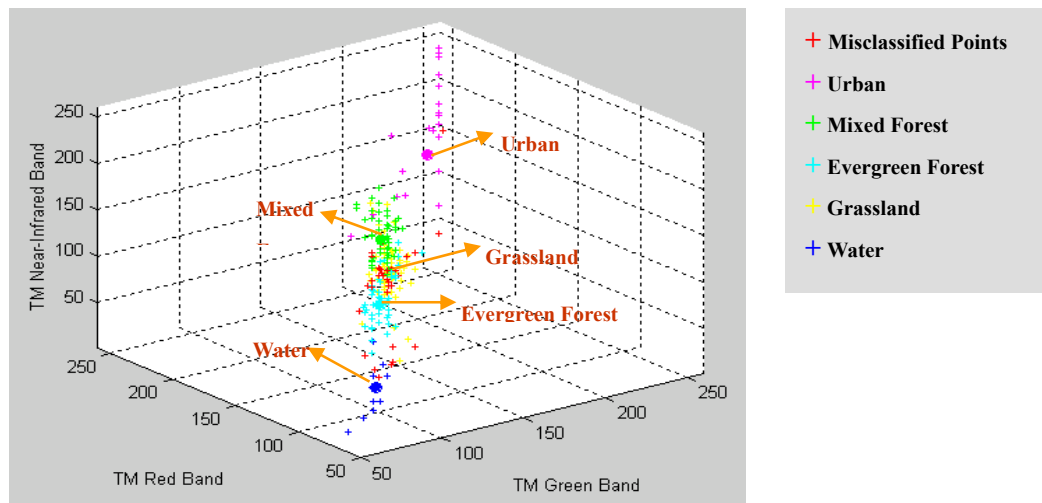
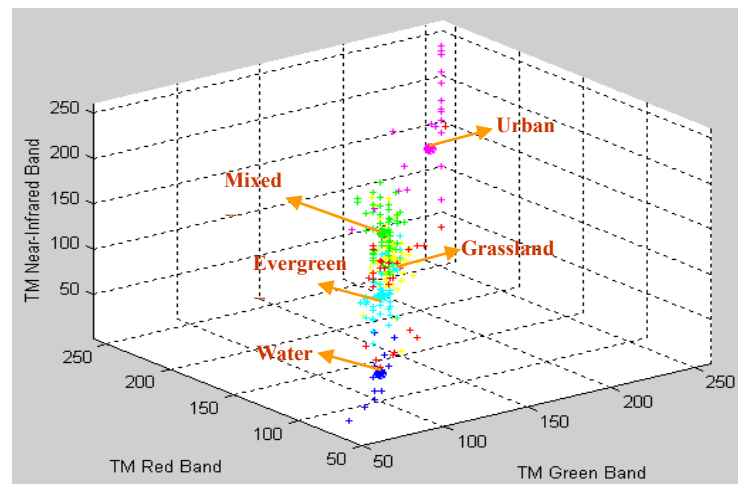


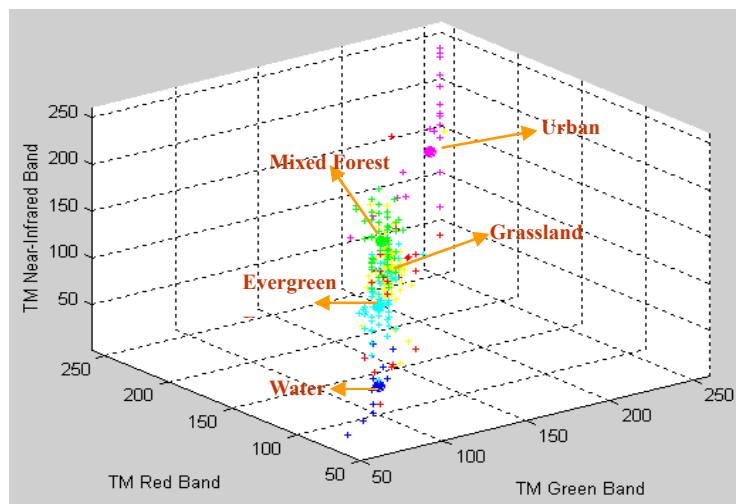
Figure 14. The original image TM-2 and the classified maps



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

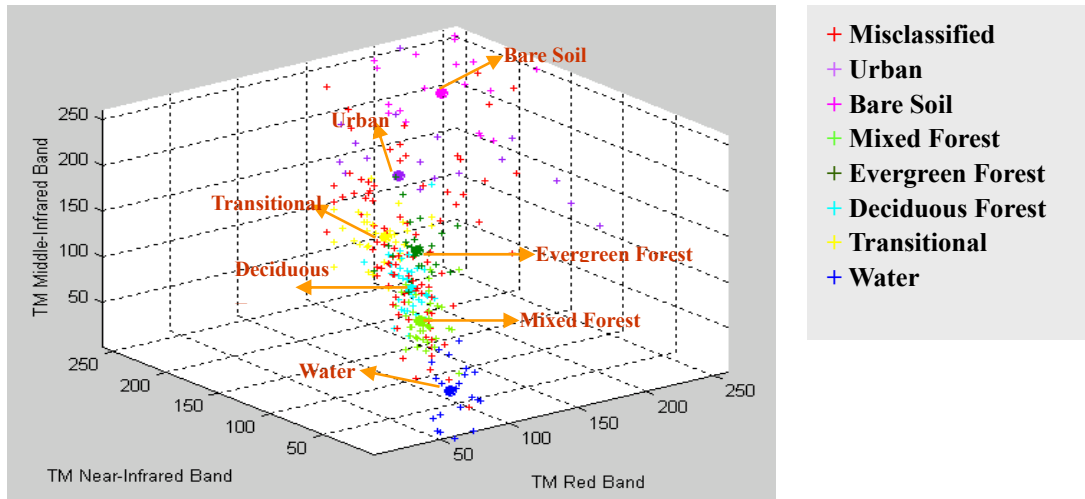


(b) Scattered Distribution of 253 points among the five classes from the Single SA-based Algorithm

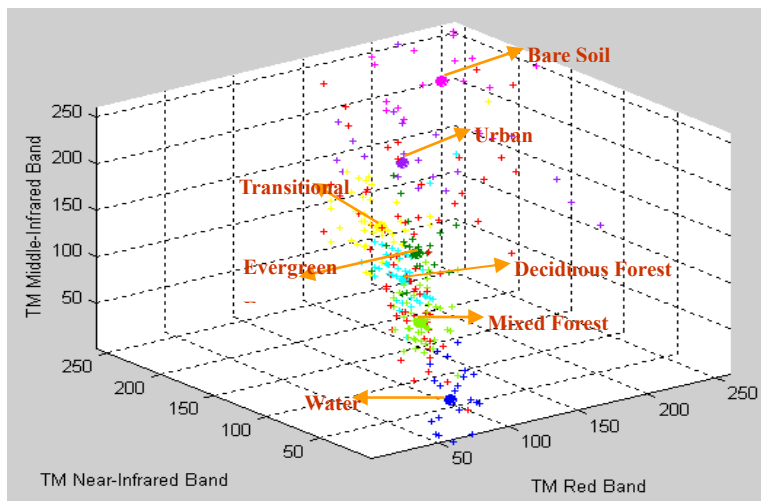


(c) Scattered Distribution of 253 points among the five classes from the Integrated SA-based Algorithm

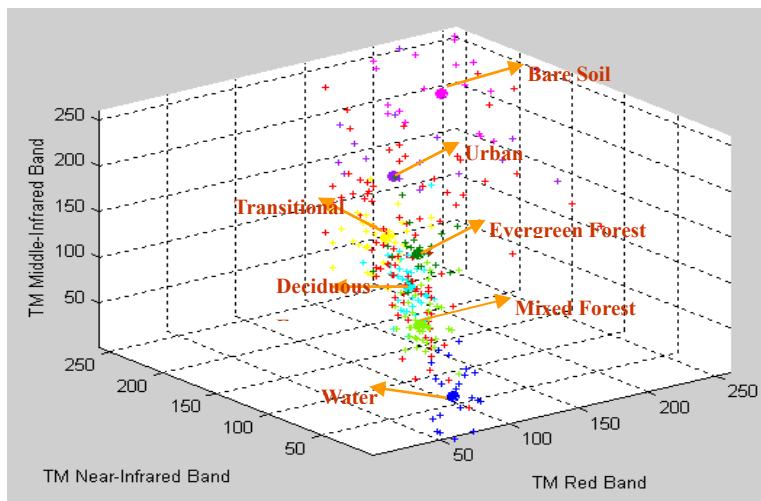
Figure 15. Scatter plots of the 253 reference points from TM-1 and their class labels



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