

2.2.3. Fitness criterion

The fitness function chosen in this study is a wellknown spectral angle mapper based nearest neighbor classifier (SAM). It is responsible for calculating fitness scores of the chromosome population during the evolution process (see Fig. 3). This means that the evolution is guided by the classification accuracy reported by SAM. Chromosomes (i.e. a subset of spectral bands) that possess higher classification accuracy are likely to have more chance to mate and produce young chromosomes that possess lower classification accuracy. The reader is recommended to consult Kruse et al. (1993) and Keshava (2004) for additional details on SAM.

3. Experiments and results

3.1. Initializing the genetic search algorithm

GA was initialized with the following parameters: population size=1000, crossover rate=100%, and mutation rate=1%. The maximum number of generations was 500. The fitness function (i.e. SAM) was trained with half of the mangrove spectra of Table 1 (15 spectra per class), and the other half was used for calculating online fitness progress

3.2. Choosing an appropriate chromosome size

Since the genetic algorithm in use was an unconstrained combinatorial optimization search (i.e. search without any constraint or penalty on the size of a chromosome), preliminary runs of GA had to be carried out to look for an appropriate chromosome size (i.e. chromosome size = the number of genes in a chromosome) that maintained high class separability (i.e. classification accuracy). The 80% level of classification accuracy was chosen as a threshold in this study as it was appropriate for separating very similar spectra of 16 mangrove species (USGS level III or IV (Anderson et al., 1976)). As a result, it was found that a minimum chromosome size that could maintain class separability above the chosen threshold was four. A comparison between the performance of three different chromosome sizes (i.e. chromosome size=2, 3, and 4) was illustrated in Fig. 5.

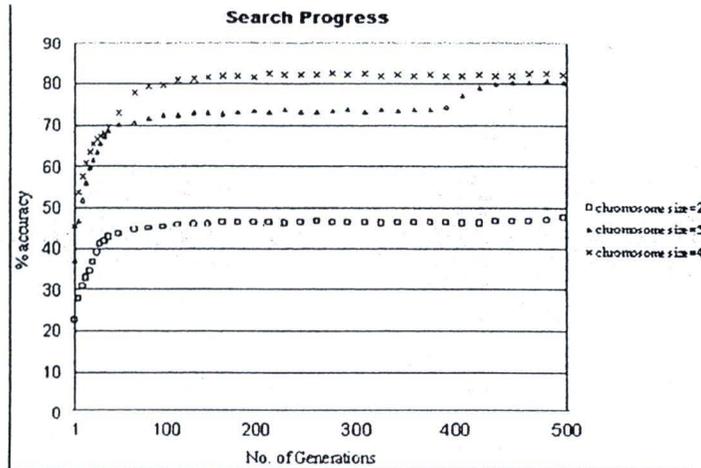


Fig. 5. A comparison between the performance of three different chromosome sizes.

3.3. Running the genetic search algorithm

GA with chromosome size four was repeatedly run 30 times to check the consistency of the results. The spectra were randomly rotated at the start of every run (i.e. data rotation) to avoid the bias. The real-time progress was plotted for each run in Fig. 6. The highest fitness score of each run was marked with a cross. Overall, the genetic algorithm quickly reached an averaged fitness score level of 80% at about the 100th generation.

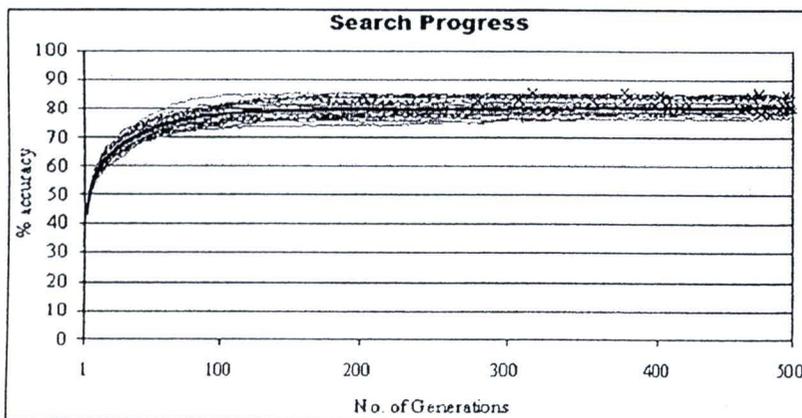


Fig. 6. The real-time progress of 30 runs (gray lines) with their peaks (crosses), mean (black line), and standard deviation limits (dashed lines).

An example of the evolution process of a single run was illustrated in Fig. 7 to give an impression of how GA worked. The horizontal axis represented band labels (or genes) from B1 to B2151. The vertical axis of Fig. 7 is the frequency of gene types found in the population. In general, the gene distribution pattern converged from originally 2151 types of genes at the 1st generation (Fig. 7a) to only a few kinds of genes at the 500th generation (Fig. 7f). The convergence quickly happened as

early as the 100th generation (Fig. 7b) where most genes were already extinct. This convergent evolution from Fig. 7a to b directly connected to Fig. 6 where the majority of the progress lines leveled off at the 100th generation as the convergence happened. Genes that dominated the evolution were individually texted in the plots. In this example, at the last generation (the 500th generation), the gene pool was dominated by the following genes: B369, B915, B1050, B1262, and B1297.

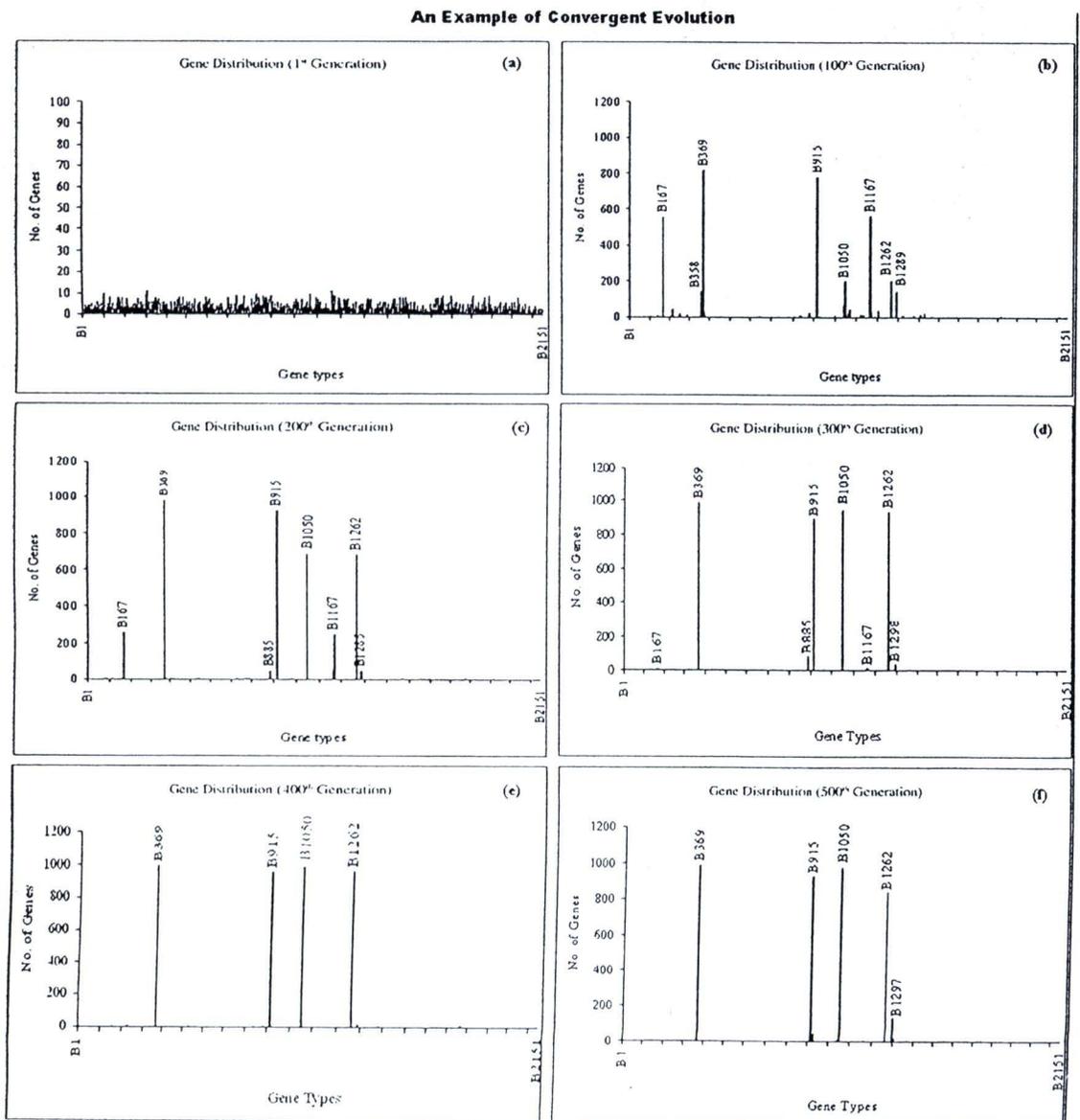


Fig. 7. This is an example of the convergence of gene distribution patterns from (a) the 1st generation to (f) the 500th generation. The horizontal axis represents band labels from B1 to B2151. The vertical axis of each plot is the frequency (count) of gene types found in the population.

The winning chromosomes from every run were reported in Table 2 along with their fitness scores (i.e. SAM classification accuracy). The best of all were chromosome No.2 and No.10. Both possessed an 86% level of classification accuracy. Then, all the genes of the 30 winning chromosomes (i.e. 120 genes in total) are grouped by minimizing their variance. The results were illustrated in a plot against a mangrove reflectance (Fig. 8). It was found that the genes (spectral bands) can be grouped at 6 different spectral positions (mean \pm standard deviation): visible area (21 genes at 513 \pm 19 nm); red edge (15 genes at 717 \pm 16 nm); nearinfrared region (9 genes at 1263 \pm 23 nm); infrared slope (44 genes at 1385 \pm 27 nm); mid-infrared absorption pitch (5 genes at 1489 \pm 21 nm), and mid-infrared peak (26 genes at 1669 \pm 25 nm).

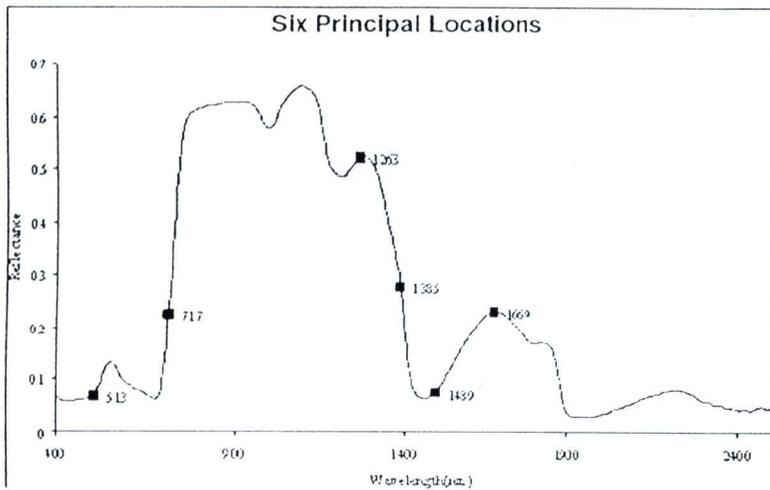


Fig. 8. After all the genes of the 30 winning chromosomes reported in Table 2 are grouped by minimizing their variance, it was found that the genes (spectral bands) can be grouped at 6 different spectral positions.

Table 2
30 winning chromosomes with their encapsulated genes

Chromosome no.	Genes (nanometer)				Fitness scores (%)
1	523	1358	1385	1710	83
2	518	1381	1393	1685	86
3	517	708	1436	1639	82
4	523	1372	1418	1671	81
5	524	1375	1496	1681	82
6	521	1378	1398	1665	79
7	549	1333	1390	1681	81
8	679	1316	1389	1673	79
9	535	1384	1506	1667	85
10	716	1246	1409	1607	86
11	722	758	1392	1436	80
12	534	1364	1385	1685	84
13	528	1363	1408	1661	81
14	725	1264	1402	1682	79
15	533	1369	1507	1660	79
16	548	1335	1458	1644	80
17	546	714	1403	1626	82
18	515	1380	1409	1674	78
19	593	1388	1480	1667	80
20	711	1234	1381	1699	83
21	536	717	1230	1397	85
22	526	725	1253	1395	83
23	495	705	1355	1398	81
24	726	1282	1381	1692	83
25	532	1350	1435	1676	79
26	523	1368	1418	1671	85
27	717	1290	1389	1668	80
28	717	1276	1405	1721	82
29	713	1289	1393	1658	78
30	528	1337	1386	1628	81

3.4. Testing the key hypothesis

The key hypothesis of this study was tested to see whether the results of band selection done by GA were meaningful. Specifically, the results of GA were statistically compared against the results of random selection using t-tests. The Jeffries–Matusita (J–M) distance was chosen as an evaluation tool (Richards, 1993). For each of the 30 winning chromosomes in Table 2, its 4 encapsulated spectral bands were used for calculating J–M distances between all mangrove classes. The averaged J–M distances of the 30 winning chromosomes were demonstrated in Table 3a. Next, the J–M distances were calculated for 30 sets of randomly generated band combinations, and their averaged J–M distances were reported in Table 3b. Subsequently, the t-test results between the two cases were demonstrated in Table 3c in terms of p-values. It was clear that the class separability of band combinations selected by the genetic algorithm was significantly higher than the class separability of randomly selected band combinations with a 95% level of confidence ($\alpha=0.05$), as most of the p-values (94/120 \approx 78%) in Table 3c were ≤ 0.05 .



Table 3

A statistical comparison between the class separability of band combinations selected by the genetic algorithm and the class separability of band combinations selected by chance (please see Table 1 for the class information): (a) genetic search result; (b) random search result; and (c) p-values

Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
<i>(a)</i>																
1																
2	2.00															
3	2.00	2.00														
4	1.99	1.74	2.00													
5	1.98	2.00	2.00	1.87												
6	2.00	2.00	2.00	2.00	1.84											
7	1.98	2.00	2.00	1.98	2.00	2.00										
8	2.00	1.95	2.00	1.90	2.00	2.00	1.99									
9	2.00	2.00	2.00	2.00	2.00	1.99	2.00	2.00								
10	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00							
11	2.00	1.97	2.00	1.87	1.95	2.00	1.97	1.99	2.00	2.00						
12	2.00	2.00	1.99	2.00	2.00	1.99	2.00	2.00	2.00	2.00	2.00					
13	1.97	2.00	1.94	2.00	1.99	2.00	2.00	2.00	2.00	2.00	2.00	2.00				
14	1.99	2.00	1.96	2.00	1.97	1.96	2.00	2.00	2.00	2.00	2.00	1.98	1.97			
15	1.98	2.00	2.00	1.99	1.82	1.97	2.00	2.00	2.00	2.00	1.98	2.00	1.97	1.98		
16	1.91	2.00	2.00	1.96	1.99	2.00	1.98	2.00	2.00	2.00	2.00	2.00	1.99	2.00	1.98	
<i>(b)</i>																
1																
2	1.99															
3	1.84	2.00														
4	1.89	1.41	1.98													
5	1.73	1.98	1.97	1.77												
6	1.56	1.99	1.91	1.95	1.59											
7	1.88	1.86	1.99	1.55	1.64	1.88										
8	1.99	1.64	2.00	1.73	1.98	1.98	1.88									
9	1.96	1.99	1.98	1.96	1.95	1.92	1.98	1.97								
10	1.95	2.00	1.95	1.99	1.98	1.98	1.98	2.06	1.94							
11	1.81	1.97	1.97	1.68	1.67	1.92	1.67	1.98	1.95	1.96						
12	1.82	1.98	1.95	1.91	1.90	1.82	1.88	1.97	1.92	1.92	1.86					
13	1.75	2.00	1.22	1.98	1.97	1.83	1.99	2.00	1.97	1.94	1.96	1.96				
14	1.51	1.99	1.68	1.97	1.79	1.46	1.93	2.00	1.94	1.91	1.93	1.81	1.55			
15	1.79	2.00	1.98	1.94	1.75	1.77	1.97	2.00	1.93	1.99	1.81	1.86	1.96	1.82		
16	1.58	1.82	1.95	1.36	1.47	1.78	1.38	1.90	1.97	1.97	1.51	1.77	1.94	1.83	1.77	
<i>(c)</i>																
1																
2	0.01															
3	0.00	0.11														
4	0.00	0.00	0.05													
5	0.00	0.03	0.03	0.02												
6	0.00	0.06	0.00	0.02	0.00											
7	0.00	0.00	0.02	0.00	0.00	0.00										
8	0.06	0.00	0.16	0.00	0.12	0.13	0.01									
9	0.00	0.07	0.03	0.05	0.05	0.01	0.10	0.04								
10	0.04	0.10	0.04	0.06	0.08	0.02	0.10	0.15	0.01							
11	0.00	0.50	0.03	0.00	0.00	0.00	0.00	0.06	0.02	0.06						
12	0.00	0.09	0.02	0.04	0.02	0.00	0.01	0.14	0.00	0.01	0.01					
13	0.00	0.07	0.00	0.05	0.03	0.00	0.04	0.08	0.10	0.02	0.02	0.04				
14	0.00	0.11	0.00	0.02	0.00	0.00	0.02	0.13	0.00	0.02	0.00	0.00	0.00			
15	0.00	0.03	0.02	0.01	0.12	0.00	0.01	0.03	0.01	0.12	0.00	0.00	0.22	0.00		
16	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.01	0.00	0.00	

สำนักงานคณะกรรมการวิจัยแห่งชาติ
 ห้องสมุดงานวิจัย
 วันที่... 2... 1... พ.ย. 2555.....
 เลขทะเบียน..... B-47392.....
 เลขเรียกหนังสือ.....