

Determination of Flood Inundated Areas Using RS Techniques in the Western Black Sea Region of Turkey

A. Ünal ŞORMAN

Middle East Technical University,

Civil Engineering Department, 06531 Ankara-TURKEY

Veysel DOĞANOĞLU

Middle East Technical University,

Civil Engineering Department, 06531 Ankara-TURKEY

Received 05.07.2000

Abstract

The remote sensing and geographic information systems are supplying new tools for flood inundation studies. RS/GIS techniques are used for the determination of the flood extent in the 20-21 May 1998 flood in the Western Black Sea region of Turkey. Two Advanced Very High Resolution Radiometer (NOAA/AVHRR) satellite images are processed in this study. The images are geocoded, and radiometrically and geometrically corrected. Unsupervised and supervised classification techniques are applied to the pre-processed images respectively in order to determine the flooded areas. The classified images are then compared for change detection. Detailed statistical analyses are performed with the results of the classifications. Ground truth observations for accuracy assessment of the classification are prepared as GIS layers after digitizing the land use maps, river networks, city locations and gaging stations. Various overlay analyses and spatial queries are performed for the interpretation of the classified image.

Key Words: RS/GIS, Flood, NOAA/AVHRR, Image Classification

Batı Karadeniz Bölgesinde UA Teknikleriyle Su Basma Alanlarının Belirlenmesi

Özet

Uzaktan Algılama ve Coğrafi Bilgi Sistemleri (UA/CBS) su basma alanlarının belirlenmesi çalışmaları için yeni araçlar sunmaktadır. 20-21 Mayıs 1998 tarihlerinde Batı Karadeniz Bölgesinde meydana gelen taşkında etkilenen alanların belirlenmesinde UA/CBS teknikleri kullanılmıştır. Bu çalışmada iki adet Gelişmiş Çok Yüksek Çözünürlüklü Radyometre (NOAA/AVHRR) uydu görüntüsü işlenmiştir. Görüntüler koordinatlandırılıp geometrik ve radyometrik açıdan düzeltiltiler. Su baskını olan alanın belirlenmesi için ham görüntüleme kontrollü ve kontrolsüz sınıflandırma teknikleri uygulanmıştır. Sınıflandırılmış görüntüler fark tesbiti için kıyaslanmıştır. Sınıflandırılmaların sonuçları üzerine ayrıntılı istatistiksel analizler yapılmıştır. Sınıflandırmanın hassasiyetini tespit edebilmek için toprak kullanım haritaları, nehir ağları, şehir ve akım rasat istasyon yerleşimleri sayısallaştırılmış ve CBS altlıkları olarak saklanmıştır. Bu altlıklar uydu görüntüleriyle üstüste bindirilerek ve alansal sorgulamalar yapılarak sınıflandırma yorumlanmıştır.

Anahtar Sözcükler: UA/CBS, Taşkın, NOAA/AVHRR, Görüntü Sınıflandırması

Introduction

The study area, known as the Filyos and Bartın river basins, affected by the recent flood in May 1998, is located in the Western Black Sea region in Turkey. The area is gaged by the stream runoff stations, and precipitation measuring stations which are operated

by both the State Hydraulic Works (DSİ) and the Electrical Power Resources Survey and Development Administration (EİEİ) for runoff and the state meteorological organization (DMI) for climatic records.

The location map of the region, with the operating measuring stations in this region, is shown in Figure 1.

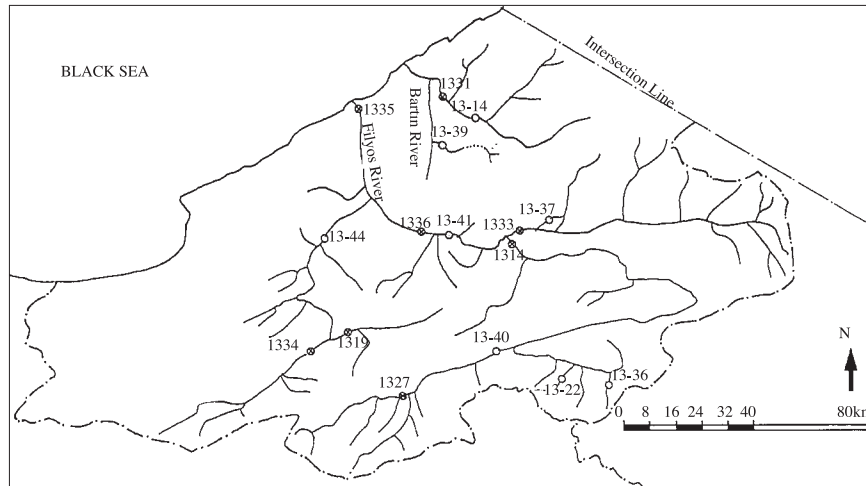


Figure 1. Map of the Western Black Sea Region of Anatolia in Turkey

The objective of this study is to make an overall evaluation of the results obtained from supervised classification of NOAA/AVHRR images, so that the areas within the study area inundated by the May 1998 flood can be identified using RS/GIS techniques. Two sets of cloud free satellite scenes before and after the major flood which took place on 20-21 May 1998 in the Western Black Sea region are fully processed.

The supervised classification is followed selecting the training samples and statistical evaluation of the characteristics of these samples. The signature parameters, and the distances between the signatures and the error matrices are determined in order to distinguish the various land use categories affected by flooding.

Classification of Satellite Images

General information

An image is comprised of picture elements (pixels) having a digital or numerical record of radiance measurements. These measurements are taken by imaging sensors in platforms (satellites or aircraft)

as they scan the ground below and to either side of the platform. As the platform moves, an image of the Earth's surface is built up by all the pixels in each scan line (Mather, 1989).

Since these images are digital (numerical in nature), they can be interpreted more effectively by computers.

Classification is such a numerical processing of the image data (radiance measurements of all the pixels in all bands of the sensor), made in order to associate each pixel in the image to a class that describes a real world object. For instance, a pixel having radiance measurements of (10, 10, 200, 280, 290) in the five bands of the NOAA14 image is assigned to its most likely class by either of two techniques of classification.

Classification techniques

Supervised classification In supervised classification, the number of the separable patterns that exist in the image area is known or at least can be estimated. Ideal patterns are formed by estimating the statistical properties of the values of each of these separable patterns.

This can be done by choosing training samples

that contain these separable patterns. For example, training samples can be taken from the separable parts of the image that are known to be ‘vegetation’ or ‘clear water’. The radiance values of each individual pixel then can be compared with each of these ideal patterns (training samples) and that pixel will be labeled as belonging to the class represented by the most similar pattern.

There are different techniques of supervised classification that the remote sensing softwares support such as Maximum Likelihood, Minimum Distance, Parallelepiped and Mahalanobis Distance techniques. In the present study we used the Minimum Distance Technique, which is based on the minimum distance decision rule (also called spectral distance) that calculates the spectral distance between the measurement vector for the candidate pixel and the mean vector for each sample and assigns the candidate pixel to the class having the minimum spectral distance (Mather, 1989).

Unsupervised classification In unsupervised classification, as the name implies, there is no control of the user in selecting the separable patterns. In this classification, an initial set of arbitrarily chosen patterns is used to allocate and reallocate the individual pixels by using some decision rule or classifier. At each step, the initial set of patterns is changed, and the procedure is repeated until an adequate convergence is obtained.

An initial unsupervised classification can be used as a pre-study of a more efficient supervised classification in refining the spectral classes present in the image. In other words, the results of the unsupervised classification can be helpful in selecting the training samples for a supervised classification.

In either of these methods, the result of the classification is a digital thematic map showing the geographical distribution of a ‘class’ such as vegetation type or water quality.

Procedure of Classification Process

Two NOAA/AVHRR-14 satellite images of the Western Black Sea region were obtained in order to determine the extent of areas that were affected by the flood that occurred in the region on 21-22 May, 1998, by remote sensing techniques. The images were scanned on May 7th, 1998 (before the flood), and on May 25th, 1998 (after the flood), and this time difference allowed an interpretation of the effects of the flood (change detection).

Although the temporal resolution of NOAA-14

images is 12 hours (in other words, the same location is scanned by the satellite every 12 hours) and it is possible to receive NOAA images every day, these images are greatly affected by clouds and cloud-free images should be selected for a classification procedure. So after an investigation of the available images by checking their quick-looks, the images of aforementioned dates are selected as cloud-free images.

The procedure steps are presented in a flowchart, shown in Figure 2. The satellite images were received in a raw image and zipped data format. They were unzipped and processed with the software ‘Quorum to Level1B’ in order to convert this raw image data to Level1B format, so that remote sensing software can be applied for processing.

Radiometric and atmospheric calibrations were applied first to the images to correct the deficiencies or flaws that could result from the imaging sensor in the platform (satellite) or from the atmosphere. Then, a geometric correction was applied to georeference the images so that the images could be integrated with Geographic Information System (GIS) data.

After this preprocessing, some enhancement techniques were applied to the geometrically corrected image for a better visual appearance and effective interpretation of the image. The images were now ready for the classification starting with selecting the training samples.

For the accuracy assessment of the classified images, the results of the classification should be compared with reference data of the area studied. In a site trip involving the first writer and the technical staff of EIE, the coordinates of the stream gaging stations in geographical and Universal Transverse Mercator (UTM) projections were determined by using Global Positioning System (GPS) equipment. The reference data is prepared by digitizing with GIS software in three different data types:

- Point Data - The coordinates of cities and stream gaging stations
- Line Data - The river network, city boundaries
- Polygon Data - The land use of a part of the Bartın Basin

These data are saved as different layers for overlay analysis using them and the classified thematic image for the determination of the flooded areas.

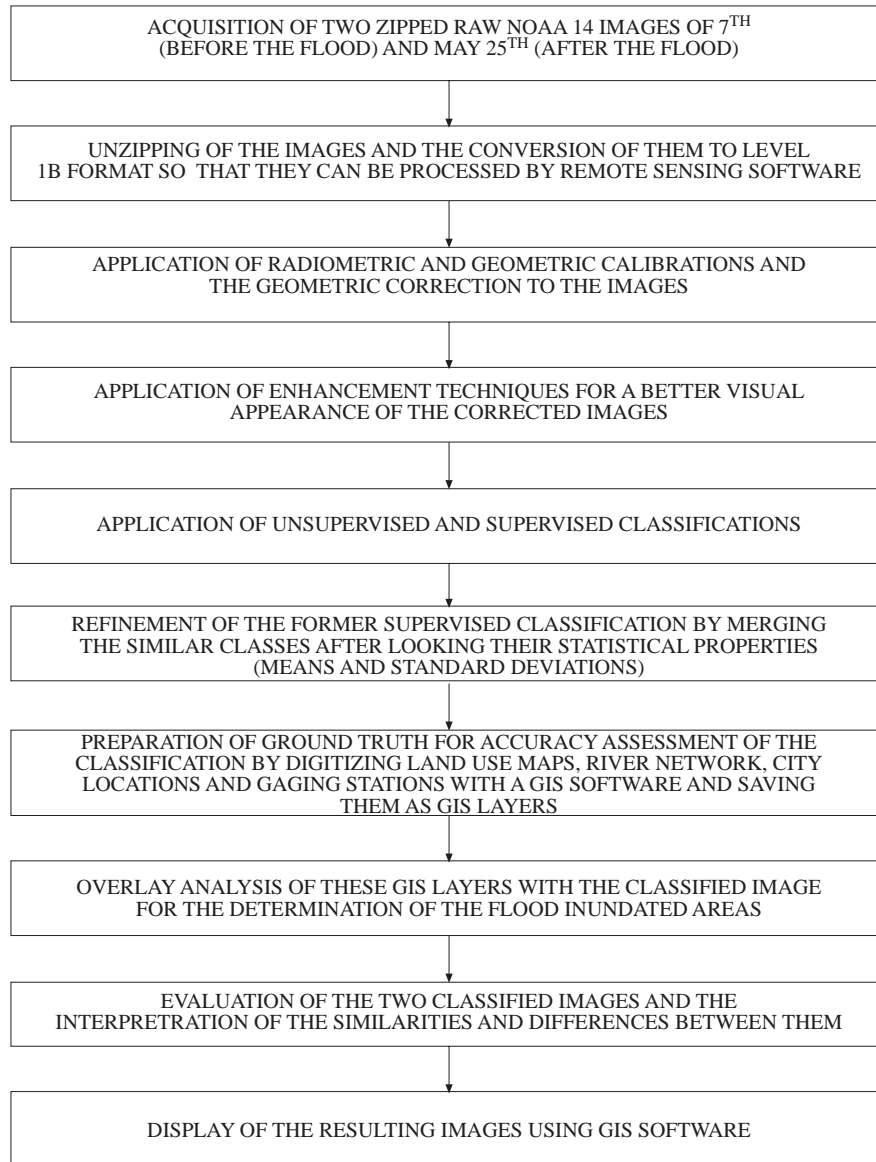


Figure 2. The Flowchart of the Application of Remote Sensing Techniques for the Determination of the Flood Inundated Areas in the western Black Sea Region

Selection of the training samples

Training samples, selected pixels that are extracted from the image to be classified, are used to estimate certain statistical characteristics of the classes to which the pixels are to be allocated. These statistical characteristics vary from one technique of supervised classification to another but generally the mean value, variance, and extreme values are those which need to be estimated from the training samples.

In the beginning, the number and location of each

separable pattern on the image at least should be estimated, although some of them are clearly distinct such as ‘sea’ or ‘clouds’, some visual enhancements like linear stretch or histogram equalization were applied to interpret the image more effectively. A few unsupervised classifications were also done to see the locations of possible earth object classes to take the training samples.

The spectral surface profile option of the software is also used in the investigation of the different spectral patterns in the image. A spectral surface profile

of an image is a plot of radiance values of the pixels in the image in any of the bands selected in a 3D space. The X and Y axes define the spatial location of that pixel on the image and the Z axis defines the radiance measured in each pixel (ERDAS, 1998). This profile should not be mixed with the topography of the image area.

An example of this profile (from the image taken after the flood) is given in Figure 3 for May 25, 1998, which was taken along the Filyos river, the possible flooded area, in band 1. In this plot, several different

patterns can be recognized: the high radiance values at the front side of the area (to be defined as clouds after classification, rows 40-50, columns 10-20), the flat portion having nearly the same radiance values (to be defined as Forest later, rows 30-50, columns 5-20), the high ridge at the left upper corner of the area (to be defined as the sediments in the sea) and finally the small ridge (river network starting from row 10, column 10 and splitting into two branches at row 40, column 10) etc.

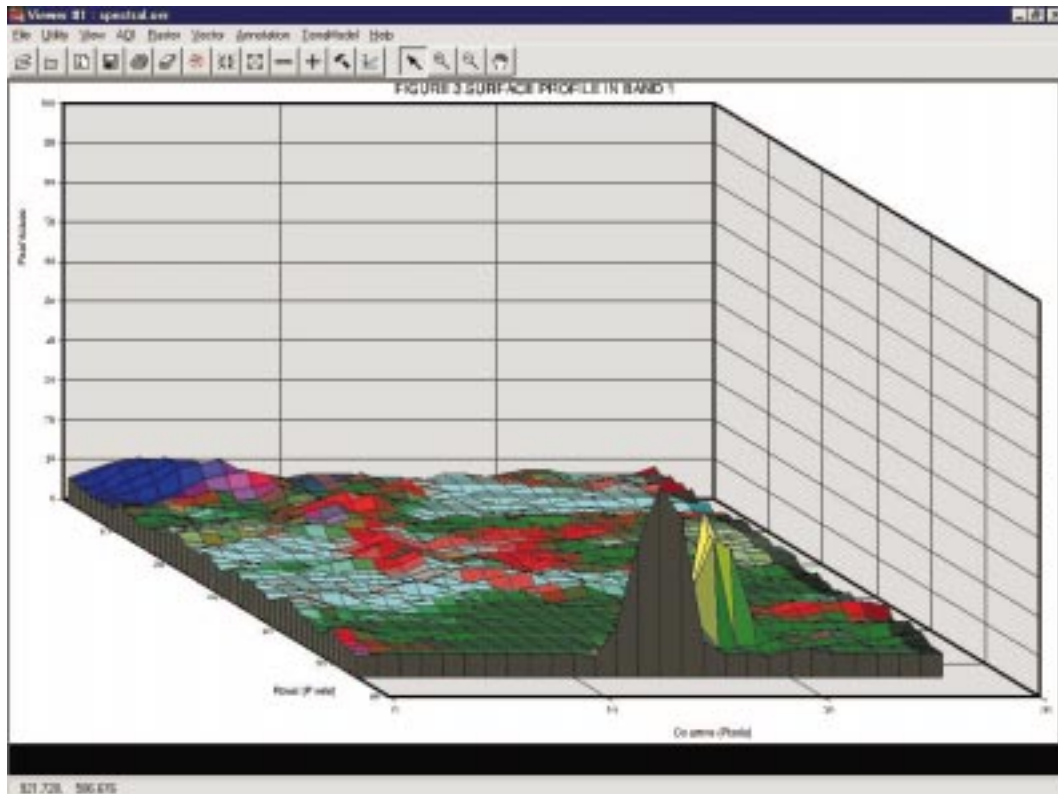


Figure 3. Surface Profile for Band 1

The final set of training samples, or signatures, consists of 10 samples (Table 1), the minimum distance supervised classification is applied, and 10 different earth objects were identified on the resulting classified image shown in Figure 4. The result of the classification will be discussed in later sections; the training samples should be examined first by some tools of the software to measure their classification capability.

Statistical evaluation of the training samples

As mentioned above, the statistical characteristics of the training areas are used to allocate the pixels of the image to the class which they belong to. These statistical characteristics can also be used to evaluate the efficiency of the supervised classification. The values shown in Table 2 are examples of such statistical values and can be viewed separately for an evaluation of all the signatures.

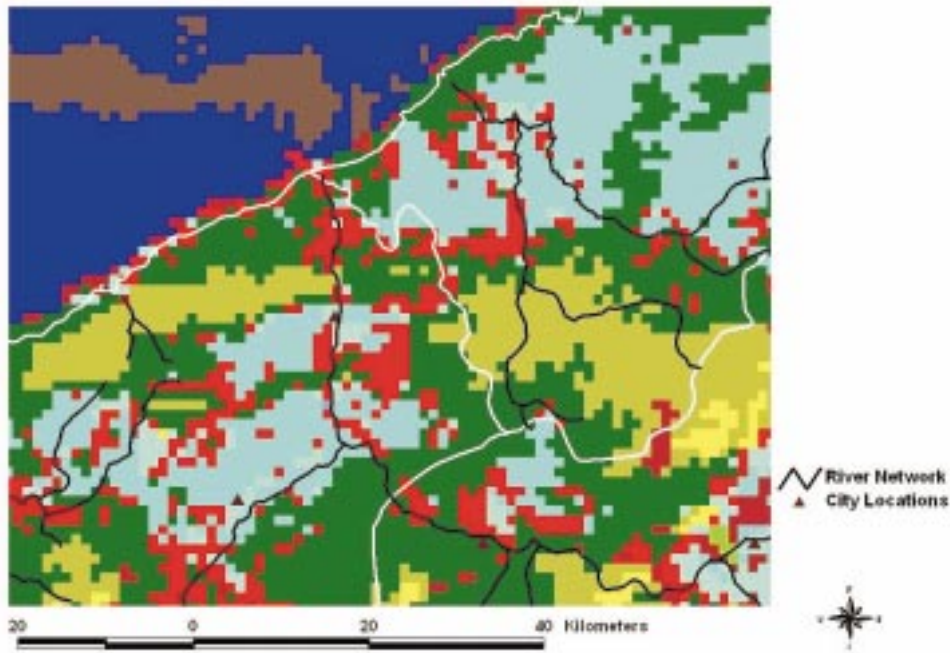


Figure 4. The Classified Thematic Image of May 25th (Bartın and Filyos Basins)

Table 1. Signature Listing

SIGNATURE	EARTH OBJECT	COLOR IN THE IMAGE
SEA	SEA	BLUE
SEDIMENT	SEDIMENT IN THE SEA	DARK BROWN
LAND1	A LAND USE TYPE	SIENNA
LAND2	A LAND USE TYPE	CYAN
LAND3	A LAND USE TYPE	PURPLE
CLOUD1	DENSE CLOUDS	YELLOW
CLOUD2	CLOUDS	YELLOW
FOREST	FOREST	DARK GREEN
SHADOW	PROBABLE SHADOW OF A CLOUD	GREEN
FLOOD	FLOODED AREAS	RED

Table 2. Statistical Values of Signature Flood

Layer	Minimum	Maximum	Mean	Standard Dev.
1	7.00	13.00	9.134	1.291
2	12.00	20.00	16.216	1.933
3	293.00	305.00	300.351	2.208
4	285.00	297.00	293.575	1.917
5	283.00	295.00	291.381	1.934

Moreover, these characteristics can also be used in creating graphs of signature parameters so that one can compare signatures. The graph appears as sets of ellipses that are based on the mean and stan-

dard deviation of one signature. For one band pair, by comparing these ellipses and examining the overlaps between the ellipses, similar groups of pixels can be extracted. If the ellipses do not overlap, then it

is understood that each signature represents distinct pixels in the two bands of the band pair that is desirable for a classification. If they overlap, a modification of the signature set may be needed (ERDAS, 1998). Such an analysis was applied to the signatures of the classified image. Then the ellipses are plotted for bands 2 and 5 using two standard deviations from the mean. Due to the apparent distinction of ellipses of classes ‘cloud’ and ‘Land2’ the plot is

limited for the other 8 classes. As can be seen from the plot (Figure 5), the ellipses of ‘shadow’, ‘sea’, ‘sediment’ and ‘land 1-3’ do not overlap and this distinction underlines the fact that the signature sets of these classes were well selected. However, there are overlaps in the ellipses of classes ‘flood’, ‘forest’ and ‘land 2’ stating that the signature sets of these three classes do not represent distinct pixels. This problem will be addressed in the next evaluations of signatures.

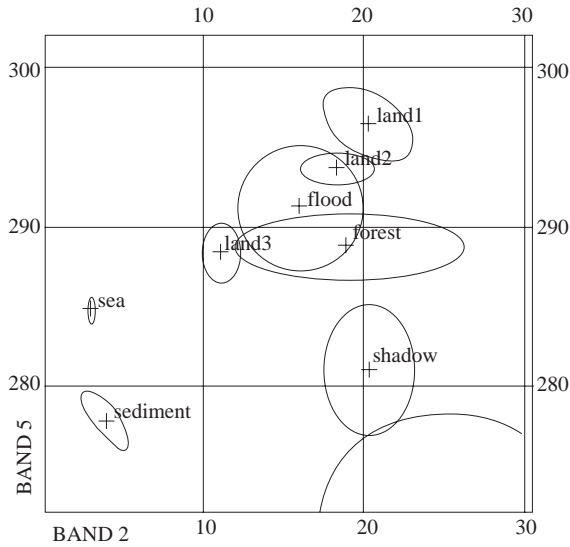


Figure 5. The Scatter Diagram of Bands 2 vs. Band 5

The statistical characteristics are also used in the computation of the statistical distance between signatures that can be used again to compare how distinct the signatures are from one another. Euclidian spectral distance is used for our image, and the separability matrix given below in Table 3 is obtained. The minimum distance (5) is between the flood and land type 2 as shown in the last column of Table 3, which is in fact expected because, as can be seen from the classified image, this land type surrounds the rivers in that region (flooded areas) and the reflectance values of these two classes will not be very different. Another small distance (8) is observed between ‘forest’ and ‘flood’ meaning that these two classes are not so separable (see overlapping discussion above). The maximum distance (91) is between ‘sea’ and ‘cloud1’, and the distances between the other classes and ‘cloud1’ are also very high. This is expected since the radiance values of clouds are very different from those of the other classes as stated above in the scatter diagram discussion.

The separability matrix is helpful in refining the training samples. The training samples having small separabilities can be merged together to form one signature and the number of classified earth objects can be decreased.

Another application of a statistical evaluation is the error matrix of the training areas. This utility allows one to see how many pixels in each training sample were assigned to each class. It is usually expected that the pixels of a training sample would be classified to the class they train. However, the pixels of the training sample weight the statistics of the signature, and it is rare that they are so homogenous that every pixel becomes assigned to the expected class (ERDAS, 1998).

In the error matrix shown in Table 4, it is seen that all the pixels in the training areas of ‘sea’, ‘sediment’, ‘Land2’, and ‘Land3’ are assigned to their classes. This verifies the homogeneity of these training areas. However, the other areas do not have such a homogeneity, and some pixels in them are assigned

to the other classes. But the percentage of these pixels is small enough when compared to the pixels that are allocated to their expected classes. The problem in the signatures of classes ‘flood’, ‘land2’ and ‘forest’ again can be seen here: 79.84% of the training area

selected for representing the ‘flood’ class is assigned to its class whereas 11.63% and 3.88% of them are assigned to the class ‘land2’ and to the class ‘forest’ respectively.

Table 3. Separability Matrix

	Sea	Sediment	Land1	Land2	Land3	Cloud1	Cloud2	Forest	Shadow	Flood
Sea	0	10	30	24	11	91	43	19	19	21
Sediment		0	37	31	18	86	38	24	19	28
Land1			0	8	20	90	47	16	25	10
Land2				0	13	90	46	9	20	5
Land3					0	89	41	8	14	11
Cloud1						0	53	85	78	88
Cloud2							0	40	30	42
Forest								0	11	8
Shadow									0	17
Flood										0

Table 4. Error Matrix of the Training Areas in Percentages

	Sea	Sediment	Land1	Land2	Land3	Cloud1	Cloud2	Forest	Shadow	Flood
Sea	100									
Sediment		100					8.33			
Land1			90							
Land2			10	100						11.63
Land3					100			6.38		2.33
Cloud1						85.71				
Cloud2						14.29	91.67			
Forest								80.85	2.04	3.88
Shadow									97.96	2.33
Flood								12.77		79.84

Overlaying of the Classified Images on Land-use Maps

The classified thematic image of the May 25th image is given in Figure 4 with the GIS layers (river network, stream gaging stations and city boundaries) prepared as reference data. In the image, it is seen that the sea (blue) is clearly identified (the borders of the class ‘sea’ are consistent with the coastal zone of the Black Sea). The observed red pixels along the rivers in the region (Filyos and Bartın) are the flood inundated pixels in the region. This fact is clearly seen when the river network is overlaid on the image as in Figure 4 and also in the query of land use map for alluvial zones in Bartın with the classified image

in Figure 6a. These pixels are not observed on the classified image taken before the flood occurrence. The dark green pixels are allocated to the class ‘forest’ and this allocation is verified with an overlay analysis with the digitized land use maps of the region, a query is done for forested area having slope more than 20% and the resulting polygons are consistent with the class ‘forest’ as shown in Figure 6b. More evidence concerning this allocation comes from the results of a Principal Component Analysis (PCA) that will be described in the next sections. The cyan pixels represent a different land type (land2) and the light green pixels are believed to be a shadow of a cloud or a unrecoverable deficiency in the image since they are not observed in the May 7th image.

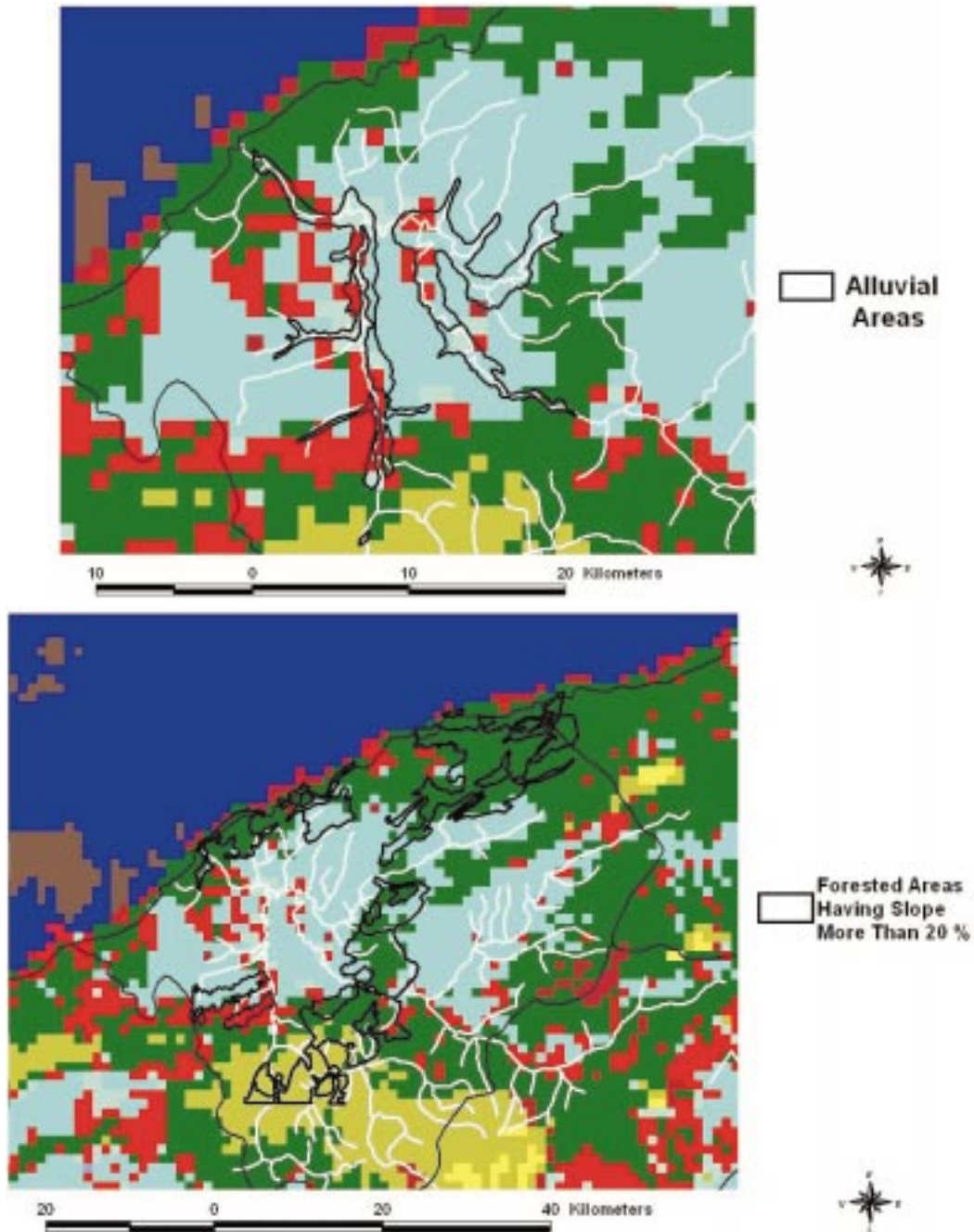


Figure 6. The Classified Image and The Land Use Map of Bartın Basin

After the above overlay analyses, the classification results given in Table 5 seemed to be a successful and were found to be consistent with the ground truth. There were some problems with the signatures of classes 'flood', 'forest' and 'land2' as mentioned in

the preceding discussions, and these problems will not allow a perfect classification with any classification techniques. Five percent of the total 60000 pixels remained unclassified after the classification with the minimum distance method.

Table 5. Classification Results

CLASS	Number of Pixels	Percentage (%)
SEA	22604	37.67
FOREST	7886	13.14
FLOOD	6617	11.03
CLOUD2	4708	7.85
SHADOW	4442	7.40
LAND2	4404	7.34
LAND1	2435	4.06
LAND3	1931	3.22
CLOUD1	1567	2.61
SEDIMENT	666	1.11
UNCLASSIFIED	2740	4.57

Decision Tree Classifiers

A decision tree classifier is an alternative classification technique. Hypothetical spectral plots of a range of values for each training sample (signature) in desired bands are used. The midpoint of the range is the mean value of that sample in that band, and the minimum and maximum values are three standard deviations above and below the mean value (Lillesand and Kiefer, 1994). For our image, the spectral plots of 8 of 10 classes were drawn since

the other two classes are cloud types that are clearly distinct and easily separable from the others (see discussions above about separability and overlapping).

In band 3, it is seen that ‘sea’ and ‘sediment’ are separable from the others and are extracted as two different classes by masking them (assigning zero value to those chosen pixels). The remaining 6 classes in level 1 are also separated through the steps given below by repeating this process recursively at each layer Figure 7.

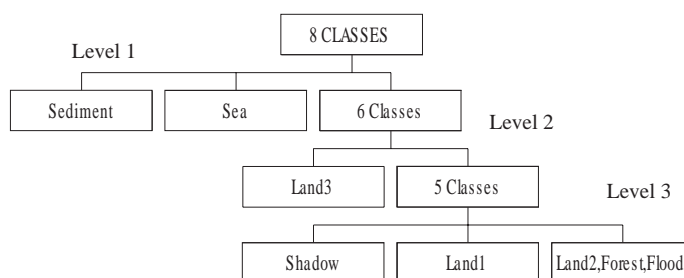


Figure 7. Flowchart of the Decision Tree Classification

The procedure is time consuming and a satisfactory result is not guaranteed. The three classes (Land2, Forest, Flood) could not be extracted as different classes with this method. For these three overlapping classes a more elaborate method of supervised classification could be suggested.

Principal Component Analysis

Principal component analysis (PCA) is a multivariate statistical technique which selects eigenvector loadings of variables that each extracted component is uncorrelated with.

The statistical variance in multispectral images is related to the spectral response of the various Earth

objects. When the five bands of NOAA/AVHRR images are treated as variables and subject to the transformation, the ordering of the principal components is influenced both by the spatial distribution of the various surface materials and by the image statistics (Mather, 1989).

The examination of the PCA eigenvector loading enables one to decide which component images will contain information directly related to the spectral signatures. Whether a targeted surface type is highlighted by dark or light pixels in the relevant principal component image can also be predicted. The magnitude and sign of eigenvector loadings give an indication of which spectral properties of landcover

(vegetation, rocks, forest) are responsible for the statistical variance.

This transformation analysis was applied to the five bands of the May 25th 1998 NOAA image and the following results, as shown in Table 6, are ob-

tained. The percentage of the eigenvalues indicate the % of variance explained in each vector. For instance, the first principal component (PC1) given in the first row indicates 75.55% variance.

Table 6. The PCA Results

Input Bands	NOAA Band1	NOAA Band2	NOAA Band3	NOAA Band4	NOAA Band5
Band Means	11.488	15.061	295.34	286.62	285.18
SD of Bands	12.56	12.44	7.2	7.409	6.855

Eigenvector Matrix

	Band 1	Band 2	Band 3	Band 4	Band 5	Eigenvalues	%
PC1	0.65815	-0.13088	0.06441	0.736722	-0.05292	352.1804	75.546
PC2	0.62114	0.387175	-0.52066	-0.43893	0.023144	94.83563	20.343
PC3	0.216094	0.537067	0.797623	-0.16449	0.040022	12.36032	2.651
PC4	-0.2675	0.547826	-0.19853	0.30301	-0.70505	6.49473	1.393
PC5	-0.25054	0.494379	-0.22169	0.38172	0.705664	0.310369	0.067

PC2 indicates the difference between the visible and infrared (IR) regions. The signs of visible band eigenvector loadings are positive, whereas they are negative in IR regions. So the landcover having high reflectance in visible regions (for example clouds) was darker and materials having high reflectance in IR regions (Land1-3, forest) was brighter in the PC2 image.

PC3 indicates strongly vegetated pixels as dark pixels, because one of the most dominant loadings in PC3 is band 2, in which vegetation has the highest reflectance in the electromagnetic spectrum and it was observed that the dark pixels on this component are nearly the same as those of the ‘forest’ area when the two images are overlaid, as it should be.

Conclusions

The overlaying of processed images using various GIS techniques with a digitized land use map by sev-

eral spatial queries is discussed. Detailed analyses are performed with these results of supervised and unsupervised classifications. The percentages of the classified pixels are determined with ten (10) different land classes but some difficulties are encountered with the signature of three classes between flooded areas, forest and land use type 2. Ground truth observations are used to check the accuracy of the classification techniques.

Overall, the analysis followed in this research appears to be a successful one, which is consistent with the ground truth records. Due to the scale factor resulting from the large spatial resolution (1.1 km * 1.1 km) of the NOAA images, more refined analysis is needed with a fine spatial resolution of satellite images (LANDSAT, SPOT, etc.) in further studies. The procedure can be followed in a similar way in future work in order to determine the flood inundated areas and help to avoid loss of life and property.

References

ERDAS, Imagine Online User Manual, 1998 Lillesand, T.M. and Kiefer. R.W, Remote Sensing and Image Interpretation, John Wiley & Sons, 1994.

Mather, P., Computer Processing of Remotely-Sensed Images, John Wiley & Sons, 1989.