IOSR Journal of Research & Method in Education (IOSR-JRME)
e-ISSN: 2320–7388,p-ISSN: 2320–737X Volume 5, Issue 2 Ver. II (Mar - Apr. 2015), PP 29-37
www.iosrjournals.org

Study the Changes in Al-ahwaz Marshal using Principal Component Analysis and Classification Technique

Alyaa Hussein Ali[1], Maysaa Raba Naeemah[2], Alaa Noori Mazher [3], Sabah Noori Mazhir [4,]
[1,2,] *(Physical Department / College of Science for women University of Baghdad, Iraq)*
[3] *(Computer Department / University of Technology, Iraq)*
[4] *(Physical Department / College of Science for women University of Baghdad, Iraq)*

Abstract: *In this paper, six bands data sets taken from the landsat-7 ETM (Enhanced Thematic Mapper).To determine the changes which occur in Alahwaz marshal during 2000 and 2013. the principal component analysis has been used to detect the changes which occur during the years(2000 and 2013). The unsupervised classification process which is the modified (K-mean) is used to classify the changes occur in the marshal region and the area surrounding it. This type of classification depend on the color and distance between the classes, a hybrid methods which are the principal component analysis with the K-mean which has been implemented to produced the best change detection regions.*
Keywords: *PCA, K-mean, statistical parameters.*

I. Introduction

Remote sensing can be defined as the collection and interpretation of information about an object without being in physical contact with the object. Aircraft and satellites are the common platforms for remote sensing of the earth and its natural resources.

A sensor is a device that measures and records electromagnetic energy. Sensors can be divided into two groups. Passive sensors depend on an external source of energy, usually the sun. The most common passive sensor is the photographic camera. Active sensors have their own source of energy, an example would be a radar gun. These sensors send out a signal and measure the amount reflected back. Active sensors are more controlled because they do not depend upon varying illumination conditions [1]. Remote sensing can generally used to study the change detection which occur during different time. There are different methods which can be used to monitoring the changes. The most common one is the K-L transformation or which is also known as principal component analysis also, the classification process gives good indication about the changes which occurs during the years. The best unsupervised classification method is the K-mean. Change Detection Methods: Change detection methods were used to detect and identify the temporal changes in the scene so that the selection of the appropriate method takes on considerable significance for detecting the changes. The technique of K-L transformation or PCA produces good estimates of change. Most studies using PCA for change detection tried to attach physical interpretation, it represents the most empirical approach to mapping change [2]. The K-means algorithm is one of the simplest and most efficient unsupervised learning algorithms to solve clustering problems in image segmentation.

II. Principal Component Analysis

Principal Component Analysis (PCA) is the general name for a technique which uses mathematical principles to transforms a number of possibly correlated variables into a smaller number of variables called principal components[3]. Principal Component Analysis (PCA) is the transformation of the multivariate data to a new set of components where data variation can be expressed by a first few components. PCA achieves this by removing the redundancy in the data set. This redundancy is quantified by the correlation of the variables. Hence, PCA transforms a correlated set of data to an uncorrelated set[4][5].

The mean position of the pixels in the space is defined by the expected value of the pixel vector x, according to equation (2)

$$m = \varepsilon\{x\} = \frac{1}{k}\sum_{k=1}^{k} x_k \qquad \ldots\ldots\ldots (1)$$

Where (m) is the mean pixel vector and the x_k are the individual pixel vectors of total number k; ε is the expectation operator. While the mean vector is useful to define the average or expected position of the pixels in multispectral vector space, it is the value to have available means by which their scatter or spread is described. This is the role of the covariance matrix which is defined as covariance matrix [6].

The covariance matrix takes the form of correlation matrix whose elements represented the covariance between the images. The diagonal elements of covariance matrix are the variance of each element.

$$\Sigma_x = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \dots & c_{1j} \\ c_{21} & c_{22} & c_{23} & \dots & c_{2j} \\ c_{31} & c_{32} & c_{33} & \dots & c_{3j} \\ \dots & \dots & \dots & \dots & \dots \\ c_{i1} & c_{i2} & c_{i3} & \dots & c_{ij} \end{bmatrix} \quad \dots\dots\dots (2)$$

The eigenvector and eigenvalue of the covariance matrix denoted by a_iandλ_i respectively where i= 1, 2, 3……N×N, where N is the dimension of the image.

The transformation matrix consists of the eigenvector of the covariance matrix.

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1N} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2N} \\ a_{31} & \dots & \dots & \dots & a_{3N} \\ \dots & \dots & \dots & \dots & \dots \\ a_{N1} & a_{N2} & a_{N3} & \dots & a_{N^2} \end{bmatrix} \quad \dots\dots\dots (3)$$

Where A is unity matrix such that $A^{-1} = A^T$

To compute the principal component the covariance matrix should be diagonalized.

$$\mathrm{A}\,\Sigma_x\,A^T = \begin{bmatrix} \lambda & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda_{N^2} \end{bmatrix} \quad \dots\dots\dots (4)$$

$$\lambda_1 \dots\dots\dots \lambda_N$$

The eigenvalue are which are uncorrelated **[7][2]** .

PCA has traditionally been used in remote sensing as a means of data compaction. For a typical multispectral image band set, it is common to find that the first two or three components are able to explain virtually all of the original variability in reflectance values. Later components thus tend to be dominated by noise effects. By rejecting these later components, the volume of data is reduced with no appreciable loss of information. Given that the later components are dominated by noise, it is also possible to use PCA as a noise removal technique [8].

III. Study Area

Maysan is one of Iraq's eighteen provinces, it lies in southeastern section of Iraq, within the precipitation valley. The total area of the province is 16072 square kilometers representing 3.7% of the total area of Iraq. The total population in the province of Maysan is about 920,315 people living and resettle differentiated between rural and urban, and it is comprised of 6 districts and 9 regions .This province is bordered from the north by the Waist Province, from the east by Iran, from the west by Dhi Qar province, and from the south by Basra province figure (1). Regarding its astronomical location, it lies between latitudes 35"15 – 32"45 north and longitudes 46"30 – 47"30 east [9] .

Figure (1) Maysan Province

The geographical nature of the province is divided between land and water, where the marshlands form more than 40% of the total area extending between its southern and southeastern parts as well as the southwestern part. In addition to that, Degla River flows in the middle of the province, with its numerous branches such as Al-Magar River, al-Mashrah, and Al-Amara, in addition to many branches originating from the eastern side of the province. As for the land areas of the province, they are characterized by being flat in most part, except for some of the eastern areas bordering Iran, which are characterized by the presence of hills that are considered an extension to the Hamreen Mountain Series.

With respect to the features of the surface, it includes several forms, the most important of which are the Degla River banks and its high streams, the low riverbed areas, as well as the marshlands and swamp areas that were dried up during the previous regime, but efforts now continue to rehabilitate them once again. There are also sand dune hills on the western and northeastern sides, some are fixed and others mobile. There is also a series of high hills that extends along the international borders with the Islamic Republic of Iran, and whose height exceeds (135m) above sea level in some areas. Ripples of valleys lie at the bases of those hills. The province climate also the annual rate of maximum temperature is 31.5 degrees centigrade, and the minimum temperature is 17.1 degrees centigrade, With respect to rain, the total annual rainfall is 161mm, falling during the period from October till March, and rarely falls during the remaining months of the year . and various types of winds that blow over Maysan[9] .

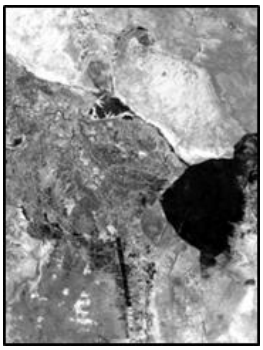
IV. Methodology

Our search focused on detecting the changes occur during (2000 and 2013). Six bands have been taken for the south of Iraq. Table(1) represented the bands information, the first method which is used to detect the changes is the PCA, figure(2) shows the six bands images for southeastern section of Iraq for two different time, figure(3) shows the PCA of the images shown in figure(2).

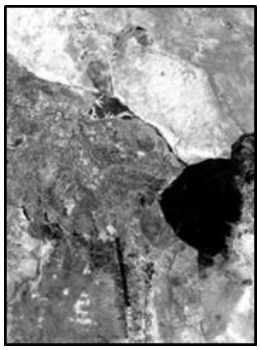
Table (1) application of the Landsat-7(Enhanced Thematic Mapper) [10].

TM Bands

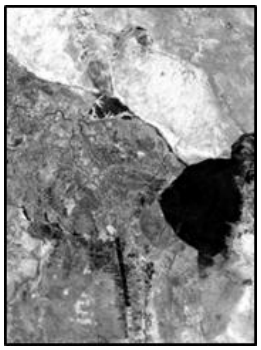
Channel	Wavelength Range (μm)	Application
TM 1	0.45 - 0.52 (blue)	soil/vegetation discrimination; bathymetry/coastal mapping; cultural/urban feature identification
TM 2	0.52 - 0.60 (green)	green vegetation mapping (measures reflectance peak); cultural/urban feature identification
TM 3	0.63 - 0.69 (red)	vegetated vs. non-vegetated and plant species discrimination (plant chlorophyll absorption); cultural/urban feature identification
TM 4	0.76 - 0.90 (near IR)	identification of plant/vegetation types, health, and biomass content; water body delineation; soil moisture
TM 5	1.55 - 1.75 (short wave IR)	sensitive to moisture in soil and vegetation; discriminating snow and cloud-covered areas
TM 6	10.4 - 12.5 (thermal IR)	vegetation stress and soil moisture discrimination related to thermal radiation; thermal mapping (urban, water)
TM 7	2.08 - 2.35 (short wave IR)	discrimination of mineral and rock types; sensitive to vegetation moisture content



Band 1of study area 2000



Band 2of study area 2000

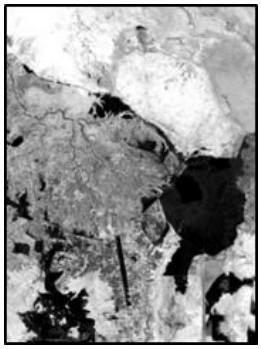


Band 3of study area 2000

Band 4of study area 2000

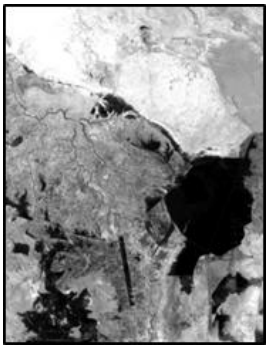
Band 5of study area 2000

Band 7of study area 2000



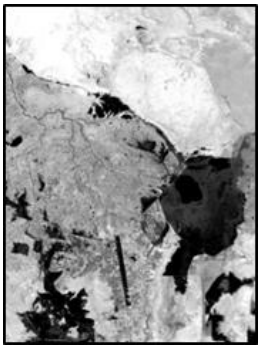
Band 1of study area 2013

Band 2of study area 2013

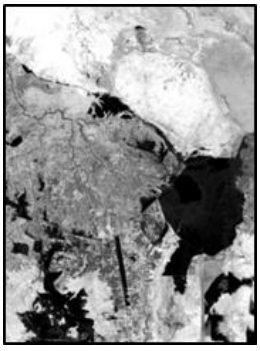


Band 3of study area 2013

Band 4of study area 2013

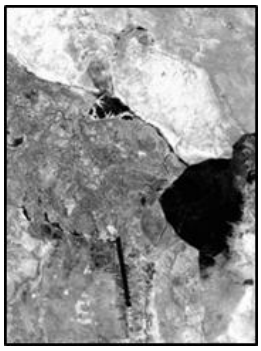


Band 5of study area 2013

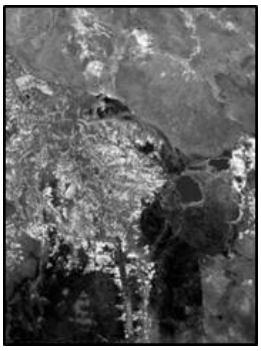


Band 7of study area 2013

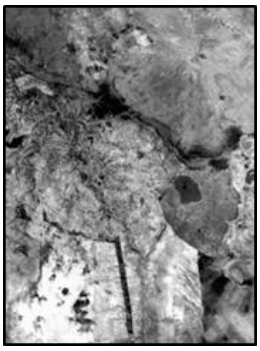
Figure (2) shows the six bands images for southeastern section of Iraq (2000, 2013).



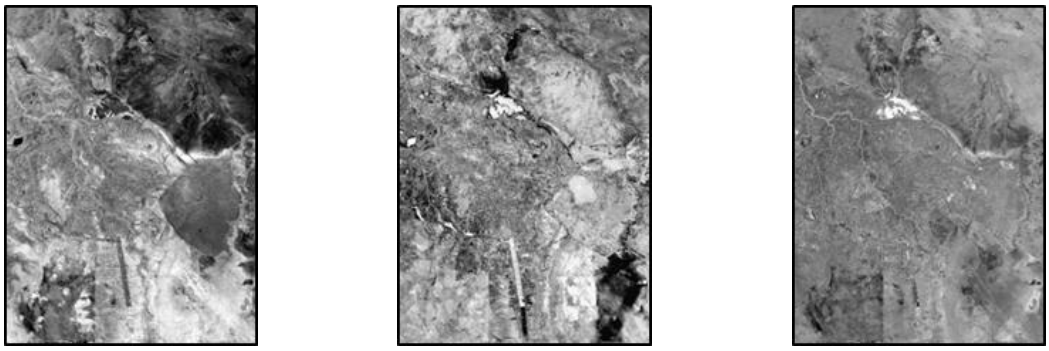
PCA 1(2000)



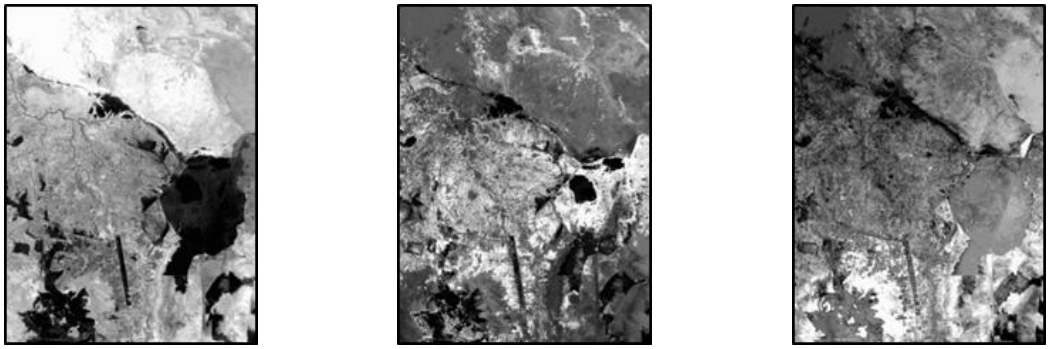
PCA 2(2000)



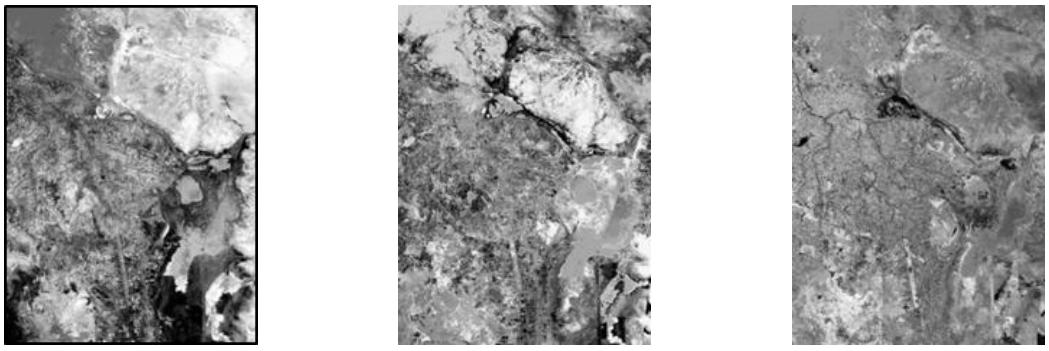
PCA 3(2000)



PCA 4(2000) PCA 5(2000) PCA 6(2000)



PCA 1(2013) PCA 2(2013) PCA 3(2013)



PCA 4(2013) PCA 5(2013) PCA 6(2013)

Figure (3) Show the principal component images for study area years (2000-2013).

V. Modified K-Mean

In our study the modified k-mean clustering which depend on the color and distance of the classes has been used. The classes are grouped depend on the color and distance.

$$p = \sum_{j=1}^{k} \sum_{i \in S_j} distance\left(x_i^{(j)}, c_j\right) \quad \ldots\ldots\ldots (5)$$

Where distance is measured between points $x_i^{(j)}$ and the cluster center c_j, S_j is set containing elements of cluster j and k is number of clusters. Algorithm consists of following steps [10] :

1. In a set of N points , corresponding to image pixels , choose k points as initial cluster centers (centroids)c_j
2. Assign each point to nearest clusterS_j based on its distance from cluster center c_j.
3. For each clusterS_j compute a meanμ_j of each cluster and set the mean as new cluster center $(c_j = \mu_j)$.
4. Repeat the steps 2 and 3 until the centroids no longer move.

VI. Image Classification 2000 And 2013

The classification method which applied in the current search is unsupervised classification techniques which is K-Mean classification technique[11]. The study area have colored using the density slicing process in order to use the modified classification process. Which is the modified K-means depend on the distance and on the color properties for each class. So, the study area could be classified into five classes. These classes represent five major features in the study area (wet vegetation, dry vegetation, wet soil, dry soil, and water) as illustrated in the following figures (4a,b,c).

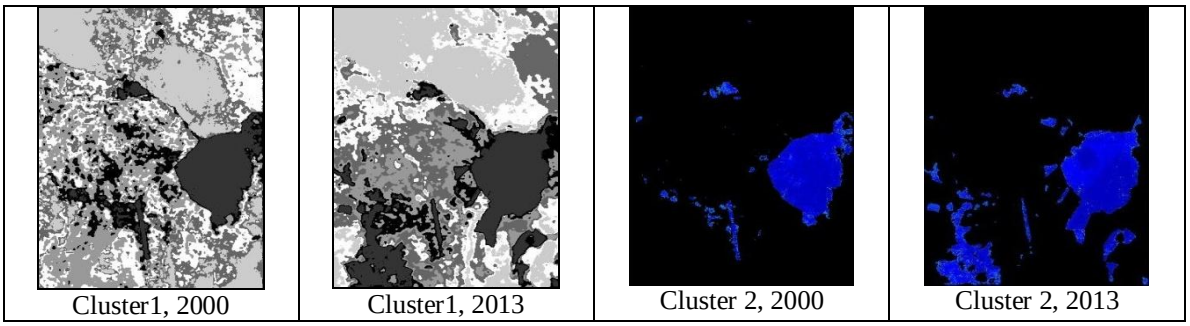


Figure (4a) Shows the K-mean clusters for the Band 1 in (2000 and 2013).

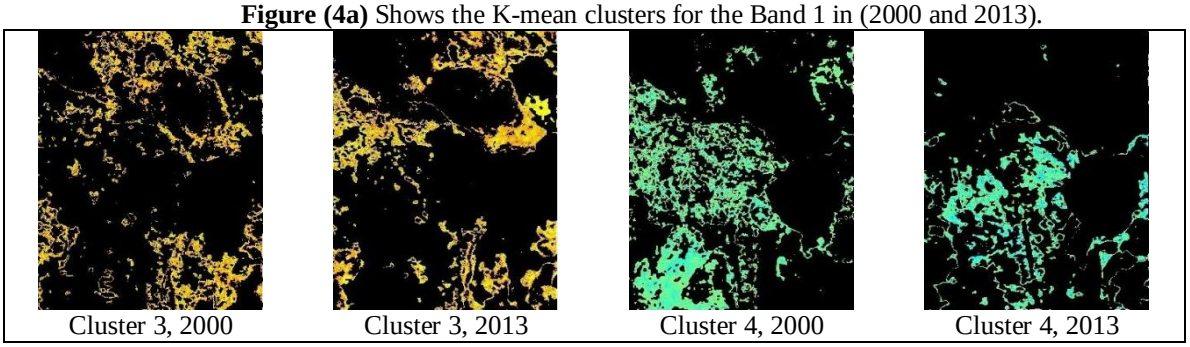


Figure (4b) Shows the K-mean clusters for the Band 1 in (2000 and 2013).

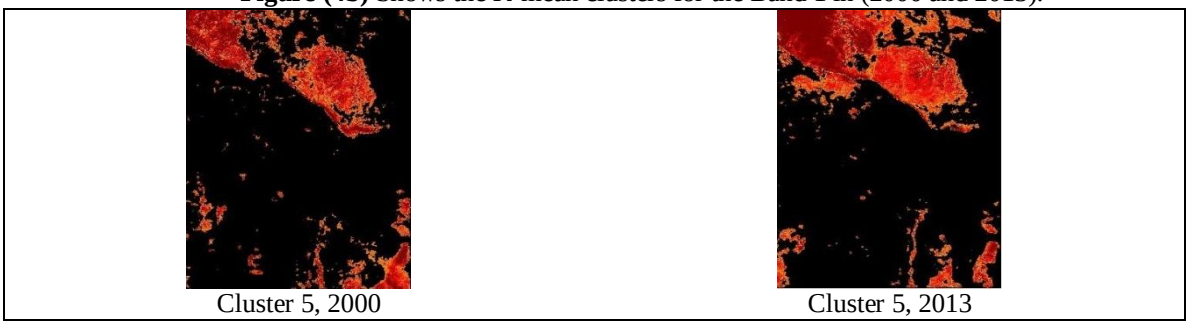


Figure (4c) Shows the K-mean clusters for the Band 1 in (2000 and 2013).

Table (2) represented the area for each class of Band 1 in both 2000 and 2013 year. One band is taken just to show how the changes can occur in this long time. This can be clearly shown from the classes and from figure (5). Also, table (2) shows the changes in the area for each class for the time different from 2000 to 2013.

The table (2) shows the statistical properties of the classification of study area.

Class	Area 2000	Area2013
Cluster1	87880	88049
Cluster2	21401	36690
Cluster3	62256	58040
Cluster4	61935	48935
Cluster5	52014	52441

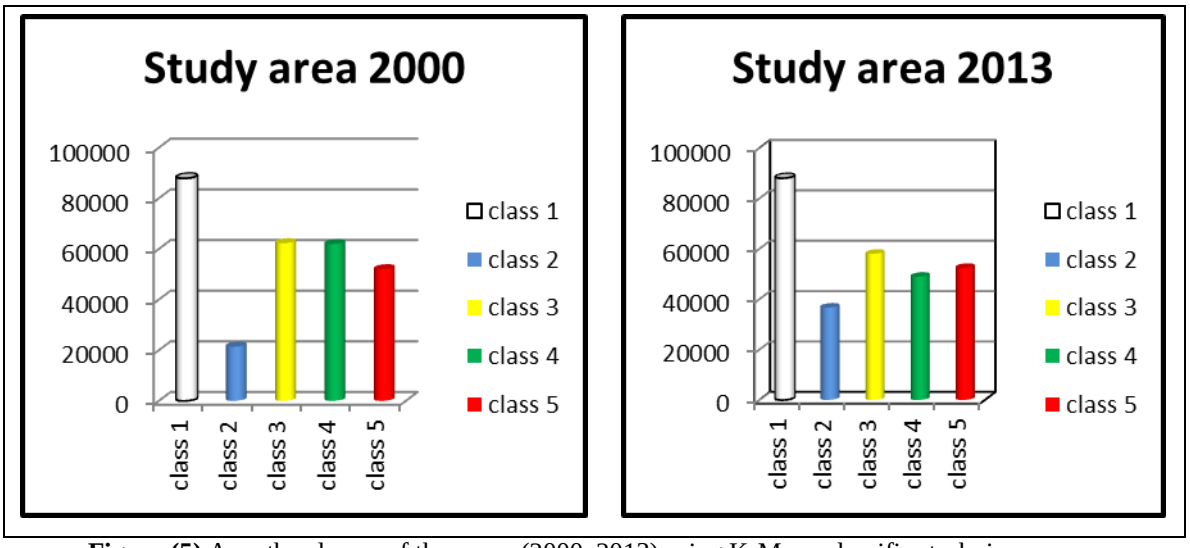


Figure (5) Area the classes of the scene (2000, 2013) using K-Mean classifier technique.

VII. PCA 2000 And 2013 Classification

The principal component (PCA) for the images (bands) in the defend time (2000, 2013) has been calculated, the result of this PC used with K-mean classification method to separated the changes in the area during these years into classes. this has been obtained by using five classes as shown in figure (6,7,8) in the present work just PC1 has been taken to classify the changes in the study area since PC1 contain the most information in the used Bands.

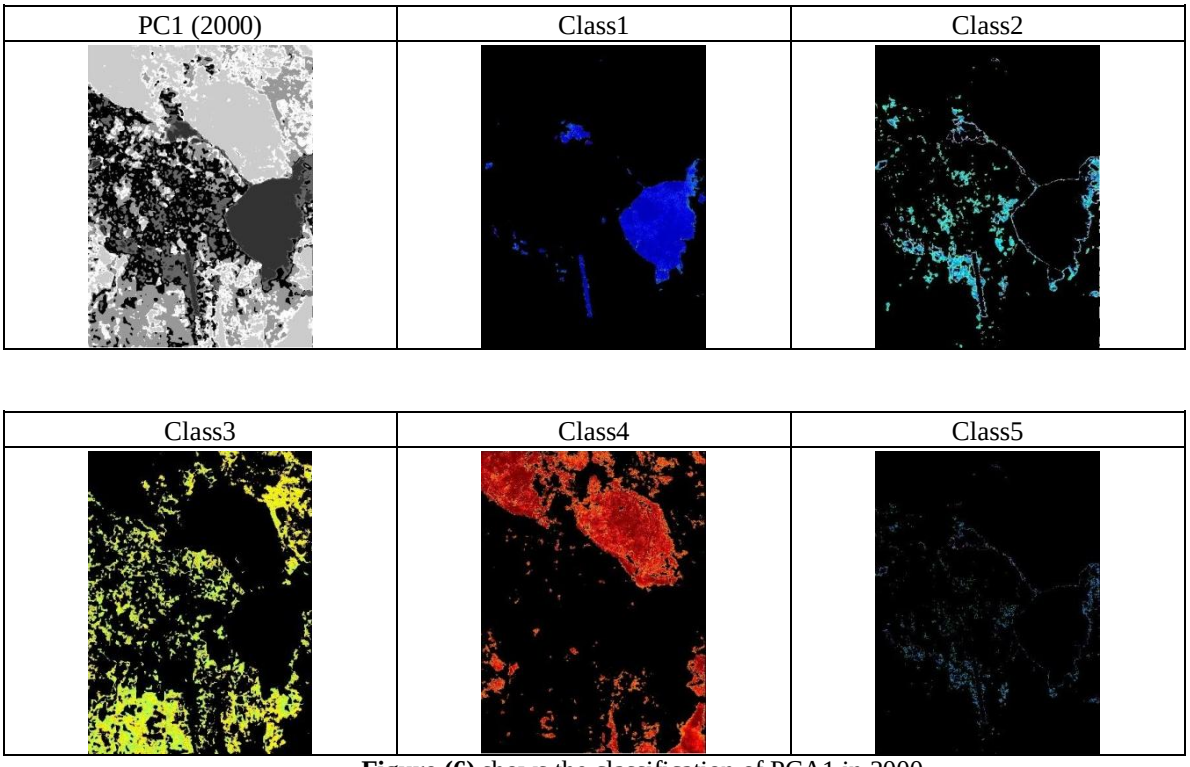


Figure (6) shows the classification of PCA1 in 2000

Table (3) ,(4) shows the statistical properties for the classes of PC1 from the value of area , mean , correlation and stander deviation . One can see that the classes for each year's (2013, 2000) different from each other some classes increase, other classes decreases. These tables show the changes occur during these years. This can clearly appear in the Figure (6) and (8). Figure(7) represent the graphical distribution of the changed area.

The table (3) shows the statistical properties for classification of PCA1 in 2000

Class	Area	mean	Correlation	Stdev
PC1	86107	0.5397	0.9691	0.3276
Class1	20093	0.0136	0.9629	0.0424
Class2	39492	0.0499	0.9099	0.1471
Class3	62951	0.194	0.9455	0.3118
Class4	56174	0.0920	0.9063	0.1533
Class5	9687	0.0121	0.8123	0.0417

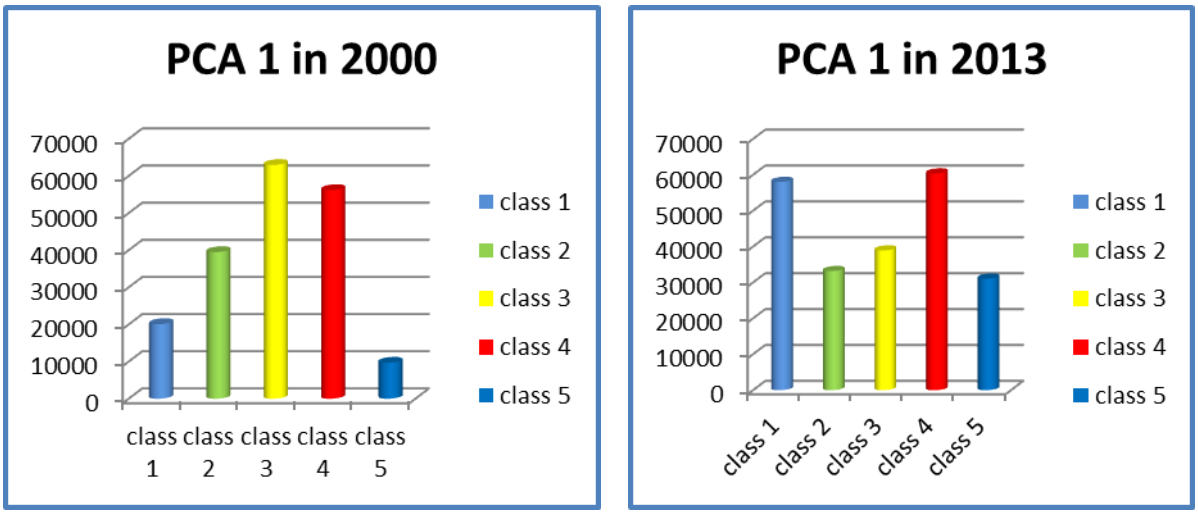


Figure (7) Area of classes of PCA1 in(2000, 2013) using K-Mean classifier technique.

The table (4) shows the statistical properties of the classification of PCA 1 in 2013

Class PC1 (2013)	Area	mean	Correlation	Stdev
PC1	70771	0. 4349	0.9773	0.3717
Class1	58035	0. 1247	0.9437	0.1742
Class2	33207	0. 0320	0.7611	0.0822
Class3	38934	0. 0543	0.9136	0.1546
Class4	60391	0. 1930	0.9595	0.3042
Class5	31080	0. 0174	0.9639	0.0377

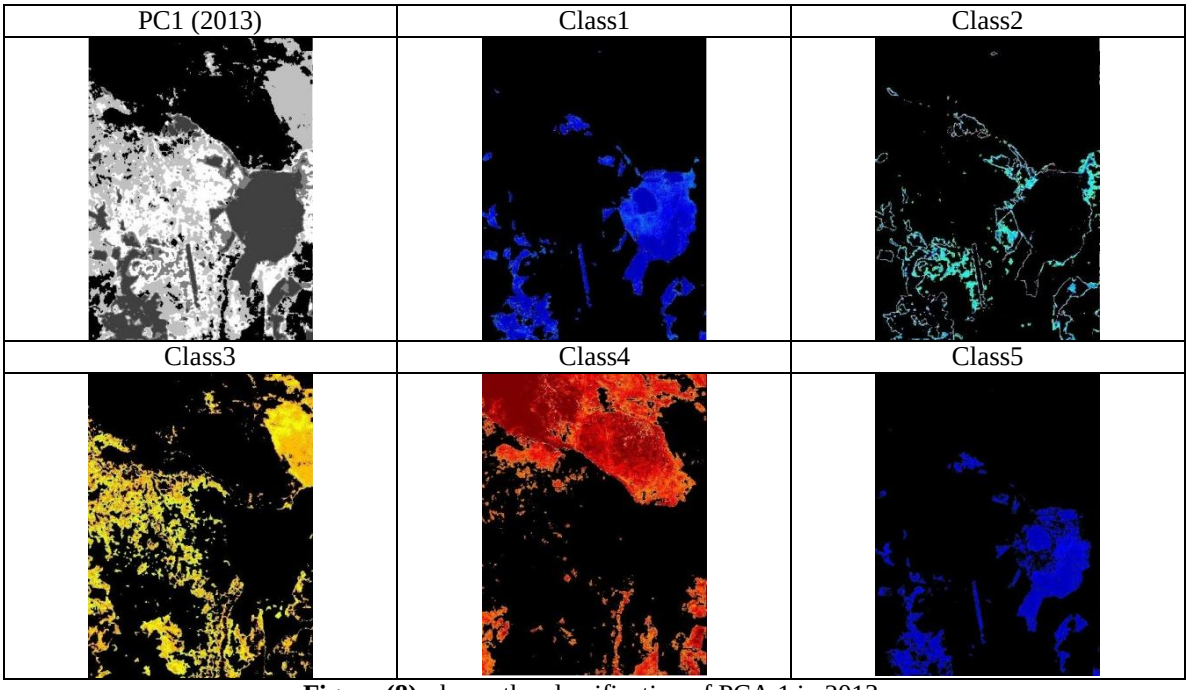


Figure (8) shows the classification of PCA 1 in 2013.

VIII. Conclusion

this search aim to show the changes that occur during 2000and 2013, the idea behind this search is how the areas and the water become different, some areas changed also, the Alahwaz marshal has changed in structure this is shown clearly in the classes and the statistical features. The PCA of the six bands have been taken but only PC1 has implemented in the K-mean since the first PCA contain the most information, the PCA for 2000 and 2013 shows clearly the changes.

References

[1]. Sanderson ,R 'Introduction to remote sensing 'New mexico state university. Introduction to Remote Sensing and Image Processing.
[2]. Ali, H.A. (1999). Studying the Flooded Area by Principal Component Analysis of Multi-Temporal Landsat Thematic Mapper Data, College of Science ,Dep. Physics , Al-Nahrain University, Baghdad, Iraq. (MSc, thesis).
[3]. Richardson, M. Principal Component Analysis. Report on May 2009, (http://people.maths.ox.ac.uk/richardsonm/SignalProcPCA.pdf).
[4]. Russ, J. C. The Image Processing Handbook (Third Edition). CRC Press LLC (1999).
[5]. Ilsever, M., Unsalan, C. Two-Dimensional Change Detection Methods, Remote Sensing Applications. pp.71. Springer, London (2012).
[6]. http://www.springer.com/series/10028.
[7]. Richards , J.A, Jia, X. (2006). Remote Sensing Digital Image Analysis: An Introduction, 4th edition : Springer – Verlag, New York, 439p.
[8]. Gonzalez, R. C. and Woods, R. E. (2008). Digital Image Processing, 3rd ed., Prentice Hall, Upper Saddle River, NJ.
[9]. Eastman, J. R. Idrisi Andes Guide to GIS and Image Processing, , Clark University, Worcester, 2006.
[10]. Republic of Iraq Ministry of Planning. Maysan Provincial Development Strategy. Iraq Information Portal, Web.sep,2011.
[11]. Canada Centre for Remote Sensing. 2007. Tutorial: Fundamentals of Remote Sensing. http://www.ccrs.nrcan.gc.ca/resource/tutor/fundam/chapter2/08_e.php.
[12]. loay. A. George, Laith. A. Al-ani and Alyaa H. Ali. June-2014."Cloud Height Classification Using Texture Analysis of Meteosat Images". Journal of Baghdad for Scince.Vol.11, No.2, pp652-659.