**Component-I(A) - Personal Details**

| Role | Name | Affiliation |
|---|---|---|
| Principal Investigator | **Prof.MasoodAhsanSiddiqui** | Department of Geography, JamiaMilliaIslamia, New Delhi |
| Paper Coordinator, if any | **Dr. Mahaveer Punia** | BITS, Jaipur |
| Content Writer/Author (CW) | **Dr. Mahaveer Punia** | BISR, Jaipur |
| Content Reviewer (CR) | **Prof.MasoodAhsanSiddiqui** | Department of Geography, JamiaMilliaIslamia, New Delhi |
| Language Editor (LE) | | |

**Component-I (B) - Description of Module**

| Items | Description of Module |
|---|---|
| Subject Name | Geography |
| Paper Name | **Remote Sensing, GIS, GPS** |
| Module Name/Title | **Supervised Classification** |
| Module Id | RS/GIS 07 |
| Pre-requisites | |
| Objectives | |
| Keywords | |

# Supervised Classification

## Introduction:

The purpose of Image classification is to categorize all pixels in a digital image into different land use / land cover classes. Depending on the interaction between computer and interpreter during classification process, there are two types of classification. These two main categories used to achieve classified output are called Supervised and Unsupervised Classification techniques. Out of the two major methods of image classification, supervised classification is generally chosen when analyst have good knowledge of the area. In supervised classification, analyst select representative samples for each land cover class. The software then uses these "training sites" and applies them to the entire image. Supervised classification uses the spectral signature defined in the training set. There are many classification algorithms used in supervised classification such as maximum likelihood and minimum-distance classification.

Supervised classification is based on the idea that a user can select sample pixels in an image that are representative of specific classes and then direct the image processing software to use these training sites as references for the classification of all other pixels in the image. Training sites are selected based on the knowledge of the user. The user also sets the bounds for how similar other pixels must be to group them together. These bounds are often set based on the spectral characteristics of the training area, plus or minus a certain increment. The user also designates the number of classes that the image is classified into. Many analysts use a combination of supervised and unsupervised classification processes to develop final output analysis and classified maps. The schemetic of the steps used in supervised classification are given below (Fig 1)

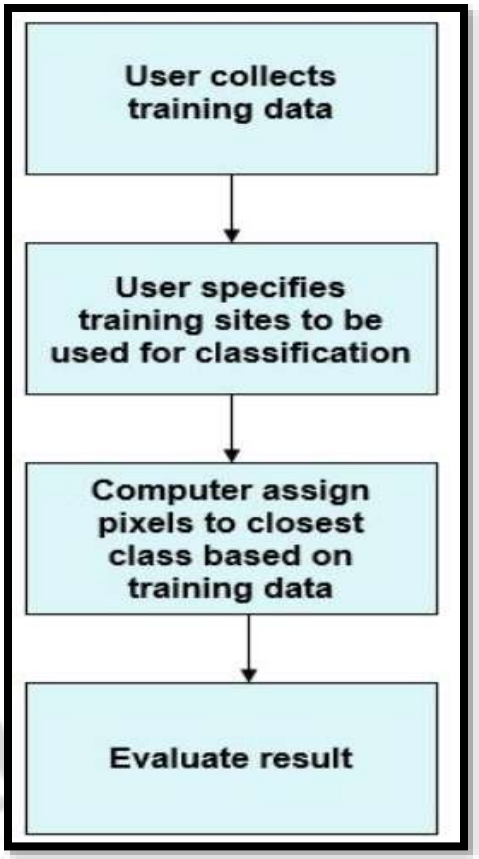


Fig.1 Steps of supervised classification

1. Training data
2. Feature selection
3. Selection of appropriate classification algorithm
4. Post classification smoothening
5. Accuracy assessment

**1. Training Data:**

Training fields are areas of known identity delineated on the digital image, usually by specifying the corner points of a rectangular or polygonal area using line and column numbers within the coordinate system of the digital image. The analyst must, of course, know the correct class for each area. Usually the analyst begins by assembling maps and

aerial photographs of the area to be classified. Specific training areas are identified for each informational category following the key characteristics of training area. The objective is to identify a set of pixels that accurately represents spectral variation present within each information region.

❖ **Key Characteristic of training area:**

**a) Shape:** Shapes of training areas are not important provided that shape does not prohibit accurate delineating and positioning of outlines of regions on digital images. Usually it is easiest to define rectangular or polygonal areas; as such shapes minimize the number of vertices that must be specified.

**b) Location:** Location is important as each informational category should be represented by several training areas positioned throughout the image. Training areas must be positioned in locations that favour accurate and convenient transfer of their outlines from maps and aerial photographs to the digital image. As the training data are to represent variation within the image, they must not be clustered in favoured regions of the image, which may not typify conditions encountered throughout the image as a whole.

**c) Number:** The optimum number of training areas depends on number of categories to be mapped, their diversity, and the resources that can be devoted to delineating training areas. Each information category, or each spectral subclass, should be represented by a number (perhaps 5 to 10 at minimum) of training areas to ensure that spectral properties of each category are represented.

**d) Placement:** Traning areas should be placed in image in a manner that permits convenient and accurate location with respect to distinctive features such as water, or boundary between distinctive features on image. They should be distributed throughout the image so that they provide the basis for representation of diversity present within the scene.

**e)** Uniformity: Perhaps the most important property of a good training area is its uniformity, homogeneity. Data within each training area should exhibit a unimodal frequency distribution for each spectral band to be used.

❖ **Evaluating Signatures:**

There are tests to perform that can help determine whether the signature data are a true representation of the pixels to be classified for each class. One can evaluate signatures that were created either from supervised or unsupervised training. There are number of methods for evaluation of the signatures:

i. Graphical method
ii. Signature saperability
iii. Divergence
iv. Transformed divergence

**i)** **Graphical Method:** Draw and view ellipse diagrams and scatter plots of data file values for every pair of bands.

ii) **Signature Saperability:** It is a statistical measure of distance between two signatures. Saperability can be calculated for any combination of bands that are used in classification. This method calculate Euclidian distance as below:

$$D = \sqrt{\sum_{i=1}^{n} (d_i - e_i)^2}$$

Where,

D=spectral distance

n=number of band

i=a particular band

$d_i$=data file value of pixel d in band *i*

$e_i$=data file value of pixel *e* in band *i*

**iii) Divergence:**

$$D_{ij} = \frac{1}{2}tr((C_i - C_j)(C_i^{-1} - C_j^{-1})) + \frac{1}{2}tr((C_i^{-1} - C_j^{-1})(\mu_i - \mu_j)(\mu_i - \mu_j)^T)$$

Where,

i,and j=the two signatures being compared

$C_i$=covariance matrix of signature i

$\mu_i$=mean vector of signature i

tr=trace function

T=transposition function

**iv) Transformed Divergence:**

$$D_{ij} = \frac{1}{2}tr((C_i - C_j)(C_i^{-1} - C_j^{-1})) + \frac{1}{2}tr((C_i^{-1} - C_j^{-1})(\mu_i - \mu_j)(\mu_i - \mu_j)^T)$$

$$TD_{ij} = 2000\left(1 - \exp\left(\frac{-D_{ij}}{8}\right)\right)$$

Where,

i,and j=the two signatures being compared

$C_i$=covariance matrix of signature i

$\mu_i$=mean vector of signature i

tr=trace function

T=transposition function

The scale of the divergence values can range from 0 to 2,000. As a general rule, if the result is greater than 1,900, then the classes can be separated. Between 1,700 and 1,900, the separation is fairly good. Below 1,700, the separation is poor.

**Selection of appropriate classification algorithm**

Various supervised classification algorithms may be used to assign an unknown pixel to one of a number of classes. The choice of a particular classifier or decision rule depends on the nature of

the input data and the desired output. Parametric classification algorithm assumes that the observed measurement Xe for reach class in each spectral band during the training phase of the supervised classification are Gaussian in nature i.e. they are normally distributed. Non parametric classification algorithm makes no such assumptions. Among the most frequently used classification algorithms are the parallelepiped, minimum distance, and maximum likelihood decision rules. An example of data samples is given below in figure 2

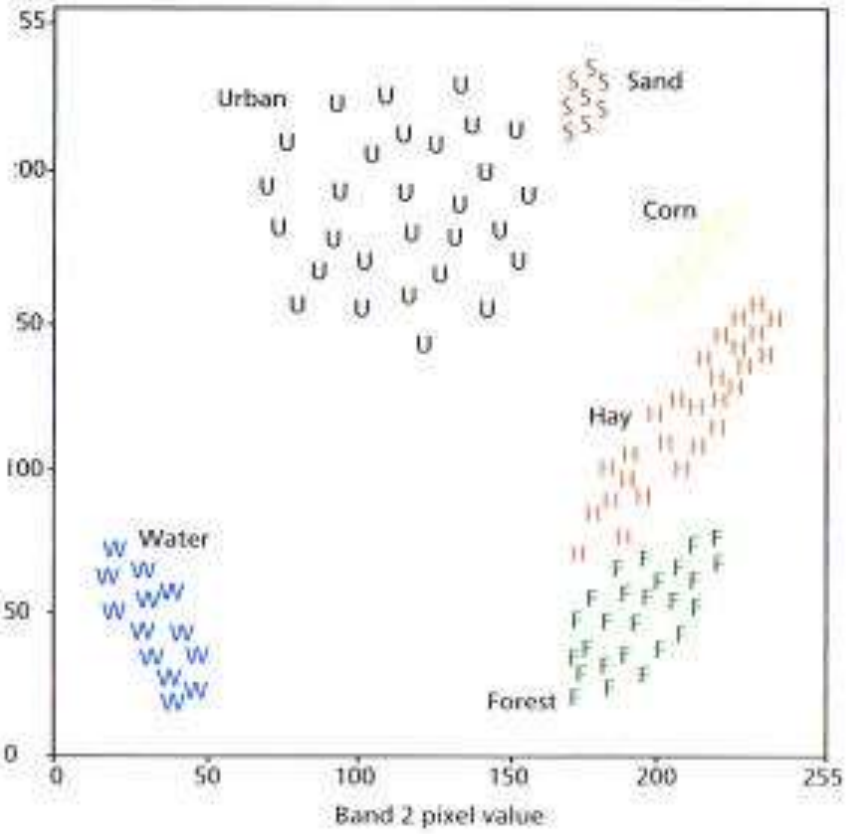


Fig.2 The training samples

**Parallelepiped Classification Algorithm:**

This is a widely used decision rule based on simple Boolean "and/or" logic. Training data in n spectral bands are used in performing the classification. Brightness values from each pixel of the multispectral imagery are used to produce an n-dimensional mean vector.

$Mc = (\mu_{ck1}, \mu_{c2}, \mu_{c3}, \ldots \mu_{cn})$ with mck being the mean value of the training data obtained for class c in band k out of m possible classes., Sck is the standard deviation of the training data of class c of band k out of m possible classes.

Using a one-standard deviation threshold, a parallelepiped algorithm decides BVijk is in class c if, and only if,

$\mu_{ck} - S_{ck} \leq BVijk \leq \mu_{ck} + S_{ck}$

where,

c = 1, 2, 3, ...., m, number of classes

k = 1, 2, 3, ...., n, number of bands

Therefore, if the low and high decision boundaries are defined as

$L_{ck} = \mu_{ck} - S_{ck}$

And

$H_{ck} = \mu_{ck} + + S_{ck}$

The parallelepiped algorithm becomes

$L_{ck} \leq BCijk \leq H_{ck}$

These decision boundaries form an n-dimensional parallelepiped in feature space. If the pixel value lies above the lower threshold and below the high threshold for all n bands evaluated, it is assigned to an unclassified category. Although it is only possible to analyze visually up to three dimensions, but it is possible to create an n-dimensional parallelepiped for classification purposes.

The parallelepiped algorithm is a computationally efficient method of classifying remote sensor data. Unfortunately, because some parallelepipeds overlap, it is possible that an unknown candidate pixel might satisfy the criteria of more than one class. In such cases it is usually assigned to the first class for which it meets all criteria. A more elegant solution is to take this pixel that can be assigned to more than one class and use a minimum distance to means decision rule to assign it to just one class. Below figure 3 shows parallelepiped algorithm

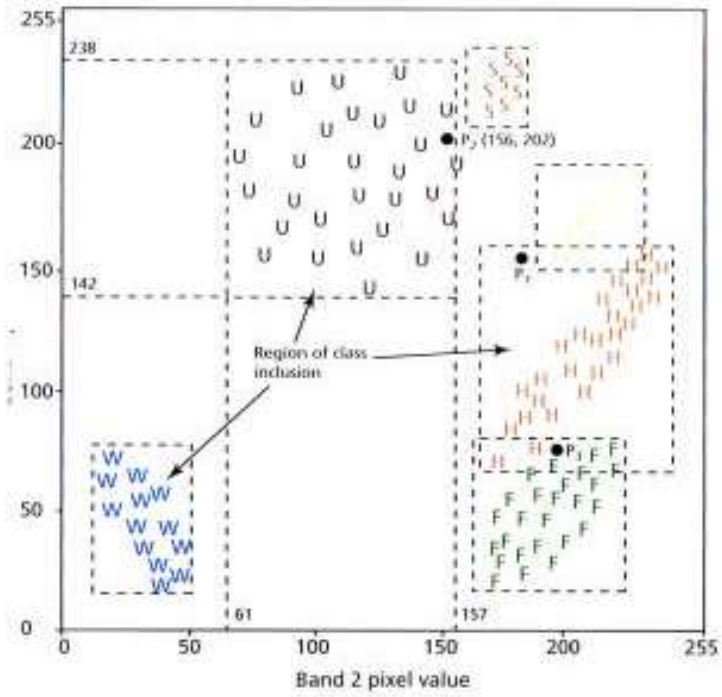


Fig. 3 Parallelepiped algorithm

**Minimum Distance to Mean Classification Algorithm:**

This decision rule is computationally simple and commonly used. When used properly it can result in classification accuracy comparable to other more computationally intensive algorithms, such as the maximum likelihood algorithm. Like the parallelepiped algorithm, it requires that the user provide the mean vectors for each class in each band mck from the training data. To perform a minimum distance classification, a program must calculate the distance to each mean vector, mck from each unknown pixel (BVijk). It is possible to calculate this distance using Euclidean distance based on the Pythagorean theorem (Figure 4).

The computation of the Euclidean distance from point to the mean of Class-1 measured in band relies on the equation:

Dist = SQRT (BVijk -mck ) 2 + (BVijl -mcl) 2

Where, mck and mcl represent the mean vectors for class c measured in bands k and l.

Many minimum-distance algorithms analyst specify a distance or threshold from the class means beyond which a pixel will not be assigned to a category even though it is nearest to the mean of that category.

When more than two bands are evaluated in a classification, it is possible to extend the logic of computing the distance between just two points in n space using the equation

$$Dab = \sum_{I=1}^{n} (ai - bi)^2$$

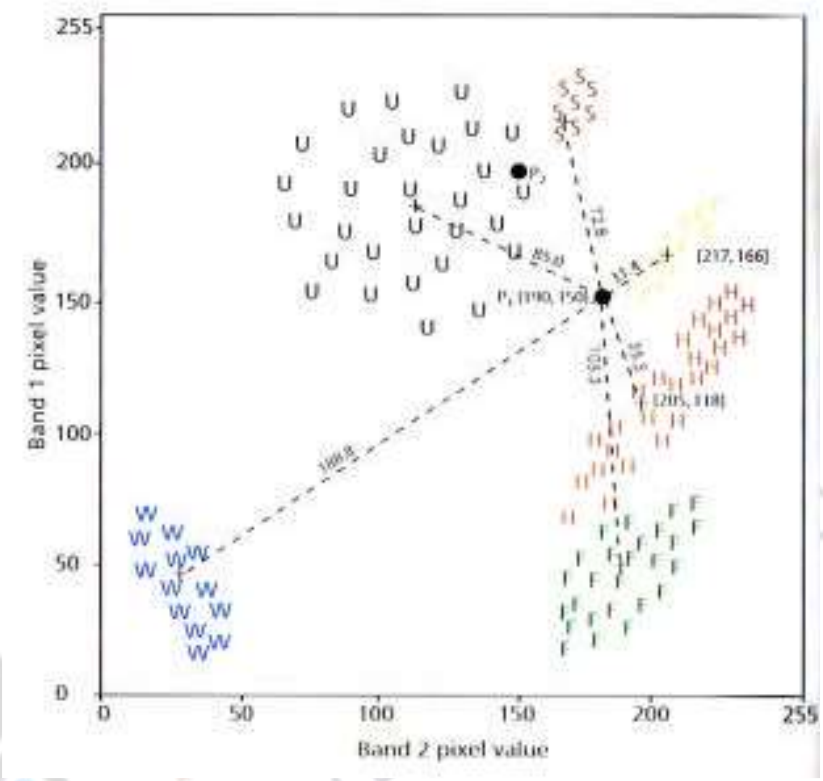


Fig. 4 Minimum Distance to mean algorithm

**Maximum Likelihood Classification Algorithm: (Fig 5)**

The maximum likelihood decision rule assigns each pixel having pattern measurements or features X to the class c whose units are most probable or likely to have given rise to feature vector x. It assumes that the training data statistics for each class in each band are normally distributed, that is, Gaussian. In other words, training data with bi-or trimodal histograms in a single band are not ideal. In such cases, the individual modes probably represent individual classes that should be trained upon individually and labeled as separate classes. This would then produce unimodal, Gaussian training class statistics that would fulfill the normal distribution requirement.

Maximum likelihood classification makes use of the statistics including the mean measurement vector, divergence, standard deviation and probability. Mean of class c (Mc) for each class and

the covariance matrix of class c for bands k through l, Vc. The decision rule applied to the unknown measurement vector X is.

Decide X is in class c if, and only if,

$pc^3$ pi, where i = 1, 2, 3, ..., m possible classes

and

$pc = \{-0.5\log_e[\det(Vc)]\} - [0.5(X-Mc)^T Vc^{-1} (X-Mc)]$

and det(Vc) is the determinant of the covariance matrix Vc. Therefore, to classify the measurement vector X of an unknown pixel into a class, the maximum likelihood decision rule computes the value pc for each class. Then it assigns the pixel to the class that has the largest (or maximum) value.

Now let us consider the computations required. In the first pass, p1 is computed, with V1 and M1 being the covariance matrix and mean vectors for class 1. Next p2 is computed using V2 and M2 This continues for all m classes. The pixel or measurement vector X is assigned to the class that produces the largest or maximum pc. The measurement vector X used in each step of the calculation consists of n elements (the number of bands being analyzed). For example, if all six bands were being analyzed, each unknown pixel would have a measurement vector X of

$$X = \begin{matrix} BVi,j,l \\ BV\ ij,2 \\ BVij,3 \\ BVij,4 \\ BVij,5 \\ BVij,6 \end{matrix}$$

The Bayes's decision rule is identical to the maximum likelihood decision rule that it does not assume that each class has equal probabilities (equal probability contours are shown in Figure 6). A priori probabilities have been used successfully as a way of incorporating the effects of relief and other terrain characteristics in improving classification accuracy. The maximum likelihood and Bayes's classification require many more computations per pixel than either the parallelepiped or minimum distance classification algorithms. They do not always produce superior results.

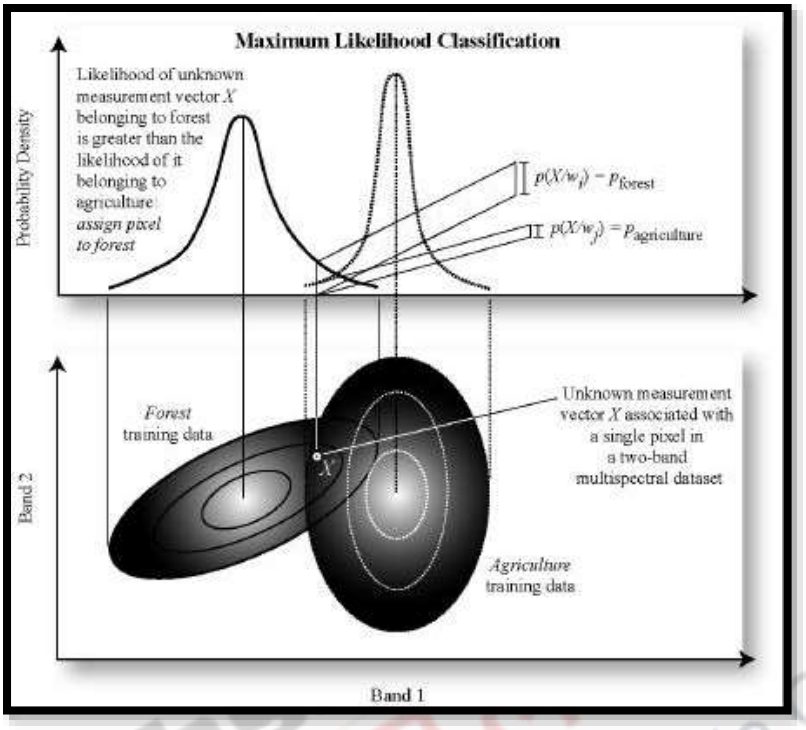


Fig.5 Maximum Likelihood algorithm

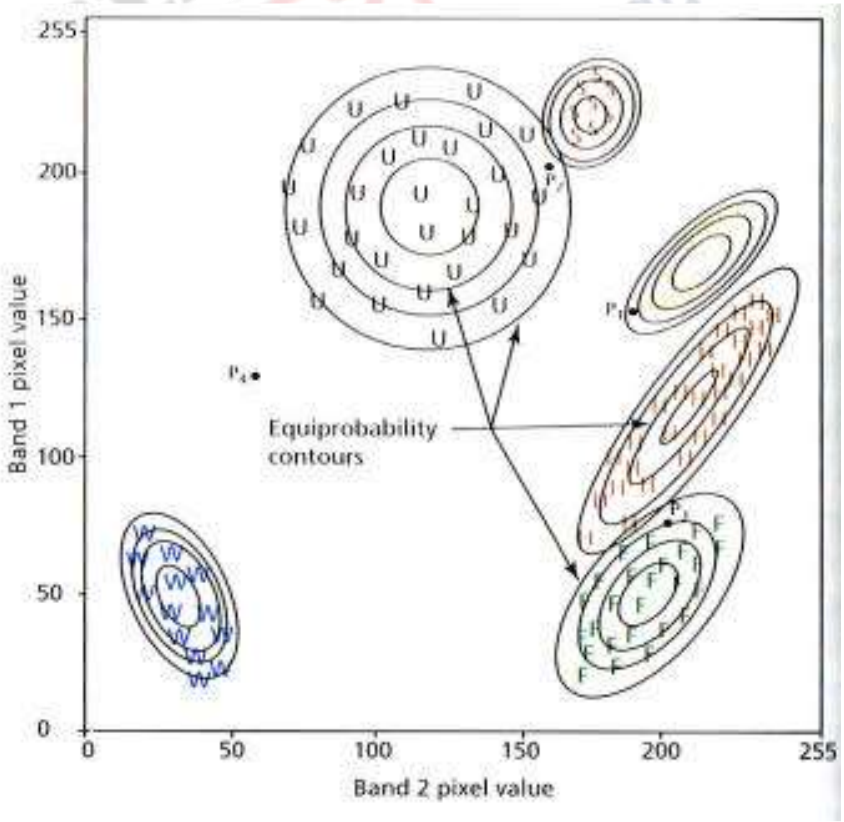


Fig.6 Equal probability contours

**Classification Accuracy Assessment:**

Quantitatively assessing classification accuracy requires the collection of some in situ data or a priori knowledge about some parts of the terrain, which can then be compared with the remote sensing derived classification map. Thus to asses classification accuracy it is necessary to compare two classification maps 1) the remote sensing derived map, and 2) assumed true map. The assumed true map may be derived from in situ investigation or quite often from the interpretation of remotely sensed data obtained at a larger scale or higher resolution.

**Overall Classification Map Accuracy Assessment**

To determine the overall accuracy of a remotely sensed classified map it is necessary to ascertain whether the map meets or exceeds some predetermined classification accuracy criteria. Overall accuracy assessment evaluates the agreement between the two maps in total area or each category. They usually do not evaluate construction errors that occur in the various categories.

**Site Specific Classification Map Accuracy Assessment**

This type of error analysis compares the accuracy of the remote sensing derived classification map pixel by pixel with the assumed true land use map. First, it is possible to conduct a site-specific error evaluation based only on the training pixels used to train the classifier in a supervised classification. This simply means that those pixel locations i, j used to train the classifier are carefully evaluated on both the classified map from remote sensing data products and the assumed true map. If training samples are distributed randomly throughout the study area, this evaluation may consider representative of the study area. If they act biased by the analyst a prior knowledge of where certain land cover types exist in the scene. Because of this bias, the classification accuracy for pixels found within the training sites are generally higher than for the remainder of the map because these are the data locations that were used to train the classifier.

Conversely if others test locations in the study area are identified and correctly labeled prior to classification and if these are not used in the training of the classification algorithm they can be used to evaluate the accuracy of the classified map. This procedure generally yields a more credible classification accuracy assessment. However additional ground truth is required for these test site coupled with problem of determining how many pixels are necessary in each test

site class. Also the method of identifying the location of the test sites prior to classification is important since many statistical tests require that locations be randomly selected (e .g using a random number generator for the identification off unbiased row and column coordinates) so that the analyst does not bias their selection.

Once the Criterion for objectively identifying the location of specific pixels to be compared is determined, it is necessary to identify the class assigned to each pixel in both the remote sensing derived map and the assumed true map. These data are tabulated and reported in a contingency table( error matrix), where overall classification accuracy and misclassification between categories are identified.

It takes the form of an m x m matrix, where m is the number of classes under investigation. The rows in the matrix represent the assumed true classes, while the columns are associated with the remote sensing derived land use. The entries in the contingency table represent the raw number of pixels encountered in each condition; however, they may be expressed as percentages, if these number becomes too large. One of the most important characteristics of such matrices is their ability to summarize errors of omission and commission. These procedures allow quantitative evaluation of the classification accuracy. Their proper use enhances the credibility, of using remote sensing derived land use information.

**Classification error matrix:**

One of the most common means of expressing classification accuracy is the preparation of classification error matrix sometimes called confusion or a contingency table. Error matrices compare on a category-by-category basis, the relationship between known reference data and the corresponding results of an automated classification. Such matrices are square, with the number of rows and columns equal to the number of categories whose classification accuracy is being assessed. Figure 7 is an error matrix that an image analyst has prepared to determine how well a Classification has categorized a representative subset of pixels used in the training process of a supervised classification. This matrix stems from classifying the sampled training set pixels and listing the known cover types used for training (columns) versus the Pixels actually classified into each land cover category by the classifier (rows).

| | | *W* | *S* | *F* | *U* | *C* | *H* | *Row Total* |
|---|---|---|---|---|---|---|---|---|
| **W** | **480** | 0 | 5 | 0 | 0 | 0 | 485 | |
| **S** | 0 | **52** | 0 | 20 | 0 | 0 | 72 | |
| **F** | 0 | 0 | **313** | 40 | 0 | 0 | 353 | |
| **U** | 0 | 16 | 0 | **126** | 0 | 0 | 142 | |
| **C** | 0 | 0 | 0 | 38 | **342** | 79 | 459 | |
| **H** | 0 | 0 | 38 | 24 | 60 | **359** | 481 | |
| **Column Total** | 480 | 68 | 356 | 248 | 402 | 438 | **1992** | |

Fig. 7 Table error matrix

| **Producer's Accuracy** | **Users Accuracy** |
|---|---|
| W= 480/480 = 100% | W= 480/485 = 99% |
| S = 052/068 = 16% | S = 052/072 = 72% |
| F = 313/356 = 88% | F = 313/352 = 87% |
| U = 126,241l = 51% | U = 126/147 = 99% |
| C = 342/402 = 85% | C = 342/459 = 74% |
| H = 359/438 = 82% | H =359/481= 75% |

**Overall accuracy = (480 + 52 + 313+ 126+ 342 +359)/1992= 84%**

W=water; S=sand, F=forest; U=Urban; C=corn; H=hay

An error matrix expresses several characteristics about classification performance. For example, one can study the various classification errors of omission (exclusion) and commission (inclusion). Note in Fig 7 the training set pixels that are classified into the proper land cover categories are located along the major diagonal of the error matrix (running from upper left to lower right). All non-diagonal elements of the matrix represent errors of omission or commission. Omission errors correspond to non-diagonal column elements (e.g. 16 pixels that

should have classified as “sand” were omitted from that category). Commission errors are represented by non-diagonal row elements (e.g. 38 urban pixels plus 79 hay pixels were improperly included in the corn category).

Several other measures for e.g. the overall accuracy of classification can be computed from the error matrix. It is determined by dividing the total number correctly classified pixels (sum of elements along the major diagonal) by the total number of reference pixels. Likewise, the accuracy’s of individual categories can be calculated by dividing the number of correctly classified pixels in each category by either the total number of pixels in the corresponding rows or column. Producers accuracy which indicates how well the training sets pixels of a given cover type are classified can be determined by dividing the number of correctly classified pixels in each category by number of training sets used for that category (column total). Whereas the Users accuracy is computed by dividing the number of correctly classified pixels in each category by the total number of pixels that were classified in that category (row total). This figure is a measure of commission error and indicates the probability that a pixel classified into a given category actually represent that category on ground.

Note the error matrix in the table indicates an overall accuracy of 84%. However producers accuracy range from just 51%(urban) to 100% (water) and users accuracy range from 72%(sand) to 99% (water). This error matrix is based on training data. If the results are good it indicates that the training samples are spectrally separable and the classification works well in the training areas. This aids in the training set refinement process, but indicates little about classifier performance elsewhere in the scene.

**Kappa Coefficient:**

Discrete multivariate techniques have been used to statistically evaluate the accuracy of remote sensing derived maps and error matrices since 1983 and are widely adopted. These techniques are appropriate as the remotely sensed data are discrete rather than continuous and are also binomially or multinomially distributed rather than normally distributed. Kappa analysis yields a Khat statistic that is the measure of agreement of accuracy. The Khat statistic is computed as:

$$\mathbf{Khat = N”\Sigma^{r} xii - (\Sigma^{r} xi+ * x+i) / N^{2} - \Sigma^{r} (xi+ * x+i)^{r}}$$

Where r is the number of rows in the matrix xii is the number of observations in row i and column i, and xi+ and x+i are the marginal totals for the row i and column i respectively and N is the total number of observations.