ACCURACY ASSESSMENT
OF
SUB – PIXEL CLASSIFICATION

A DISSERTATION

Submitted in partial fulfilment of the
requirements for the award of the degree
of

MASTER OF TECHNOLOGY

in

CIVIL ENGINEERING
(With Specialization in Geomatics Engineering)

By

RAVINDRA KR. GOYAL

DEPARTMENT OF CIVIL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247667 (INDIA)

JUNE, 2007

CANDIDATE'S DECLARATION

I hereby declare that the work presented in the dissertation titled "**Accuracy Assessment of Sub-pixel Classification**" in partial fulfillment of the requirement for the award of the degree of **Master of Technology** in Civil Engineering with specialization in **Geomatics Engineering**, submitted in the Department of Civil Engineering, Indian Institute of Technology Roorkee, is an authentic record of my own work carried out during a period from October 2006 to June 2007, under the guidance of **Dr. J.K. Ghosh**, Assistant Professor, Department of Civil Engineering, IIT Roorkee, Roorkee, India.

I have not submitted the matter embodied in this dissertation work for the award of any other degree.

Place : Roorkee

Date : 29/June/07

R. Goyal

(Ravindra Kr. Goyal)

CERTIFICATE

This is to certify that the above statement made by the candidate is correct to the best of my knowledge and belief.

J. K. Ghosh

(**Dr. J.K. Ghosh**)

Assistant Professor,

Department of Civil Engineering,

Indian Institute of Technology Roorkee,

Roorkee-247667

Uttarakhand, India

ACKNOWLEDGEMENT

I wish to affirm my earnest acknowledgement and indebtedness to my supervisor Dr. J. K. Ghosh, Assistant Professor, Geomatics Engineering Section, Civil Engineering Department, Indian Institute of Technology, Roorkee, for his copious help and esteemed guidance. He has been a continuous source of inspiration to me. It is beyond my literary capability to express my gratitude for his guidance and making me learning and interpreting things better.

I am also thankful to all faculty members of Geomatics section for their help in numerous ways time to time. I wish to express my deep regards and gratitude to my parents, who are source of encouragement to me. And finally I also owe a lot to my intimate friends for providing me a vital support in discussion about the subject and encouragement whenever I needed.

Date: 29 June, 2007 Ravindra Kr. Goyal

Place: Roorkee

ABSTRACT

Knowledge about land cover is an important input for the modeling of information that can be used for planning of proper utilization of natural resources. The derivation of such information increasingly relies on remote sensing technology due to its ability to acquire measurements of land surfaces at various spatial and temporal scales. One of the major approaches to deriving land cover information from remotely sensed images is classification. Conventionally hard classifiers are used which give the output having every pixel in a single land cover class which is far from the actual scenario on ground as well as loss of information is there. Unlike hard classifiers, sub-pixel (soft) classifiers defer making a definitive judgment about the class membership of any pixel in favor of producing a group of statements about the degree of membership of that pixel in each of the possible classes.

There is a constant endeavor for obtaining more accurate results of classification. Accuracy evaluation of such individual classification technique and mutual comparison of the performance of accuracy assessment methods are key issues of debate and research in the field of remote sensing. Accuracy is itself defined as "the closeness of results of observations, computations, or estimates to the true values or the values accepted as being true' (USGS, 1990). These methods are categorized according to their basic concept like distance, similarity, uncertainty and fuzzy data set. Some latest features like fuzzy correlation coefficient, various entropy measures, fuzzy kappa and new operators in fuzzy error matrix are also discussed. A few of these measures are applied on actual data set with Bayesian and fuzzy classifiers used. Both mixed and pure training data are used for their classification. Their comparative analysis is done through statistical results and graphs.. In both the cases mixed training data provided batter classified image than pure training data. Accuracy results obtained by fuzzy classification were found to be better than Bayesian classifier.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Land cover information has been identified as one of the crucial data components for many aspects of infrastructure planning and development. Remote sensing data are vital source of such information as they are available at various scales and time. Production of such land cover maps can be facilitated by using automated methods for classification. The automated classification of land cover from remotely sensed data forms the basis of producing thematic maps. Conventionally hard classifiers are used which give the output having every pixel in a single land cover class which is far from the actual scenario on ground as well as loss of information is there

Sub-pixel classification techniques avoid the loss of spatial information and generate thematic maps which better represent land cover variations as compared to crisp (hard) classifiers. Mixed pixels are bottom line output here, representing an area on the ground comprising multiple land cover classes and having partial membership grades to all exclusively defined classes. Accuracy evaluation of such individual technique and mutual comparison of their performance are key issue of debate and research in the field of remote sensing. Manipulating accuracy and indexing errors are vital results for users, as they not only reveal the fitness of classified map for specific implications, but also expose propagation of errors in subsequent secondary data.

Why- sub-pixel study so important in the field of remote sensing-

1. Low cost of coarse resolution images, from we can get information near to a high resolution images, with sub pixel level study.
2. Continuous variation in ground properties is reflected.
3. Pixel reflectance value represents mixture ground properties.
4. Various soft classification techniques are available ,which shoud be assessed for their accuracy

In thematic mapping from remotely sensed data, the term accuracy is used typically to express the degree of 'correctness' of a map or classification (Foody, G.M., 2002) and quantified as (dis)agreement of the classified output with reality or ground truth. Error matrix technique of hard classification accuracy assessment has ruled over for recent years but it is hardly suitable for sub pixel classification accuracy assessment. Sub pixel data (soft data) may be of two kinds; Probabilistic and fuzzy set based; depending upon the sum of the membership grades of the pixel. Ground data may also be crisp or soft. Comparisons between these data sets have been done by various methods and still research is going on.

The accuracy assessment methods can be categorized according to their basic concept like distance based, similarity based, uncertainty based and fuzzy data based. Some latest features like fuzzy kappa and, composite and multiplication operator in fuzzy error matrix enhance this domain. These measures are reviewed and summarized along with the tabulation of their mathematical formulation and references.

Measures like entropy was first developed to measure the amount of uncertainty in the information content of the classified data with respect to an ambiguity free pixel based reference data.

Measures of distance were used to find a metric distance between two datasets. Information closeness is a distance measure applicable for two probability distributions and Euclidean or L distances are distance measures for any two distributions, probabilistic or possibilistic.

To find the similarity in class representation between two datasets, similarity measures like correlation coefficient and RMS error were developed for probabilistic datasets, whereas different fuzzy similarity indexes based on fuzzy error matrix were developed to measure the similarity between two fuzzy datasets.

For getting more information about the classification accuracy, conventional error matrix is found to be the best. So, an attempt was made to generate fuzzy error matrix, which also has the capability of providing a number of accuracy measures like those obtained from conventional error matrix. Though error matrix was considered to be the best accuracy assessment tool for hard classification, but fuzzy set based error matrix is not that much popular till now and this is the reason why more techniques are still being developed. Fuzzy functions were another approach to check accuracy when the classified output is hard and reference data is soft. Though, by using this measure different types of errors are evident, but certain disadvantages are associated with this approach as discussed before for which this method does not impress much for accuracy assessment.

It is also observed that though several accuracy measures are continuously being developed in the field of sub-pixel accuracy assessment of classified images, but the literatures do not provide any detailed information about the sampling scheme, sample size, method of obtaining the samples etc. for soft classification, though this is a very important step in classification.

Fuzzy entropy measure and fuzzy correlation coefficient are discussed in detail for the actual ground data. Fuzzy error matrix is also calculated for the same. Process of low pass filter is used to degrade the image and subsequent study is done on the same.

In most of the papers, the reference data is collected from a higher resolution image. Thus the reference data is not truly fuzzy in nature, since the class membership values are class proportions only which sum up to give a value of 1 for each pixel. But fuzzy approach is different from probabilistic approach in the sense that the membership values are obtained from a predefined membership function.

Due to the disadvantages of some of the measures already developed and used in different literatures, there is a scope to develop new methods for accuracy assessment of fuzzy classification that do not suffer from any such disadvantage.

1.2 Objective of the work

The objective of this work is to evaluate different accuracy assessment measures including the novel measures for fuzzy classification accuracy assessment and to make a comparison of their capabilities in different situations. Different accuracy assessment measures have been used to evaluate the accuracy of data set. The measures which give appropriate results are considered to be good accuracy indicator. The results have been analyzed thoroughly to find out the advantages and disadvantages of each of the measures. Depending on the perfection of the results, the best accuracy assessment measures for fuzzy classification accuracy assessment have been recommended.

CHAPTER 2

LITERATURE REVIEW

Extraction of thematic information from satellite images is generally achieved through the application of a conventional classification, which allocates each pixel to a land cover class and thus the whole image gets segmented into a number of classes. This is the easiest way of obtaining a map like representation of the earth. A perfect segmentation is achieved when the class represented in any image matches the class on the actual ground surface for the same area. This is known as the accuracy of classification.

Pixel is the arbitrary spatial unit ,which may represent an area on the ground which comprise one or more than one discrete land cover classes ,accordingly it is termed as hard (pure) and sub (soft/ mixed) pixel. Sub-pixels are abundant in nature. Sub-pixels occur due to higher land cover variation comparative to spatial resolution of the sensor or, two or more ground features fall within an instantaneous field of view of a detector cell. Proportions of the classes in a pixel are represented as the probability distribution or fuzzy set.

While reviewing accuracy aspects of sub-pixel classification, it seems unavoidable to consider about the making of sub-pixel classified map. The concept of sub-pixel has been incorporated in all three stages of classification process. Training and allocation are the two broad spectrum of this computational process.

2.1 Training and Allocation Stages in Sub Pixel Classification

In training, pixels of known class membership in the remote sensing data are characterized and class signatures are derived (Foody, 1996). Most supervised image classification methods need pure pixels for training. Training with mixed pixels may lower the classification accuracy, as the class response derived may not be the actual class response (Foody and Arora, 1996).

However, training becomes complicate when pure pixels are scarce. In these cases, it can be difficult to obtain a sufficiently large number of representative training samples to accurately estimate the spectra of the classes (Lesparre, 2003; Gorte et al., 2003).

The solution for the lack of pure training samples is to be found in the use of mixed pixels to estimate the spectra of pure classes. An advantage of such a fuzzy training method is that more pixels in the image can be used for training, which enables the use of heterogeneous areas for training or the random selection of training pixels. There are two conditions for this method. First, one needs to have estimates of the fractions of the classes in the mixed training samples. Secondly, the spectral values of the mixed pixels should be a linear combination of the spectra of the composing class.

Besides these advantages, however, relatively little attention has been directed to the accommodation of fuzziness caused by mixed pixels in the training stage (Arora,M.K., Foody,G.M.,1996).

Allocation of the pixels in sub-pixels classification methods gives partial membership (value 0 to 1) to all the discrete classes defined. Numerous algorithms are available like fuzzy c-means clustering, maximum likelihood in soft mode, linear mixture modeling, artificial neural network, support vector machine, evidential reasoning classifier, decision tree classification and many more. Data associated with the pixel is soft here. Uncertainty may arise in the data.

Quality of the allocation algorithms reflects through, how accurately they represent the actual scenario on the ground. Without index of accuracy a classification process can not be taken as complete.

2.2 Testing Stage or Accuracy Assessment:

This is typically accomplished by comparing a sample of pixels' classification with some form of reference data. Depending upon the type of classified and ground data accuracy may be categorized in four ways.

1. Hard ground data vs. hard classified data
2. Hard ground data vs. soft classified data
3. Soft ground data vs. hard classified data
4. Soft ground data vs. soft classified data

We are, here, interested in the later three categories. Soft data may itself be of two kinds. Accuracy assessment process takes classified data and reference data as the input. These data sets may be either crisp or soft depending on the availability of resources, requirement of the project, purpose of study, allocation scheme used for the classification etc. Soft data may itself be of two types; fuzzy and probabilistic. Depending upon the input dataset suitable measure is applied to assess the accuracy. The comparison of classified and reference data gives out put which may be represented in presented in various forms like; overall accuracy, users accuracy, producers accuracy, kappa coefficients, percent correct, entropy, distance, similarity ,correlation coefficient .

2.2.1 Reference Data:

Since all the accuracy assessment are done by using sub-pixel information of reference data, so the most important step for assessing accuracy is to obtain accurate reference data. Obtaining true fuzzy reference data is a very difficult task.

Conventionally ground reference data is derived from photogrammetry, field survey, or existing map. In such ground data, the spatial variations evident in reality are commonly obscured (Zhang, J., Foody,G..M., 1998). Other reference data may be higher spatial resolution image than that of the classification being assessed. It may be contemporaneous with the dates of the classifications' source remote sensing imagery. It should possess known (and acceptably high) classification accuracy itself.

2.3.2 Deriving Ground Reference Data in Sub Pixel Format:

1. Degradation Method: - Pure pixels are passed through a low pass filter and results obtained are taken as the combination of various membership values. In the current study this method is followed where PAN image is passed through low pass mean filter to obtain degraded image.

2. Fine Resolution Images Vs. Coarse Resolution: - In most of the cases in the literatures, the reference data has been obtained by using a finer resolution image [Foody, (1996), Foody, (2000), Latifovic and Olthof, (2004), E. Binaghi et al, (1999)]. All the pixels in the reference image are considered to be pure. The membership degree of any class in a pixel in the classified image is simply the proportion of that class in the pixel as obtained from the reference image. The sample pixels chosen from the image should be such that they include all the classes being considered for classification. In this approach

the reference image should be perfectly registered with respect to the classified image, so that no spatial error is accompanied. Without proper registration, the generation of error matrix itself will be erroneous. Problem in this case is that, since class proportions are taken to be the class membership values, the membership always sum up to one for each pixel and thus leads to a probability distribution.

3. Distance Based Interpolation: - Heterogeneities of the pixel's equivalent area are not equally probable (Foody, 1995). Inner parts of the polygon may have 100%probabiity to the class that of the polygon but at the boundary it will decrease, as the influence of surroundings get considered .the changing pattern of class probabilities may be modeled by some quantitative function like interpolation (Wang and Hall, 1996).Distance based interpolation may not be suitable for the features, where probability variation is not continuous. Spatial distribution of points of known probability should also be taken into account.

2.3.3 Sampling Issues:

There are different approaches of sampling, such as simple random sampling, systematic random sampling, stratified random sampling, stratified systematic unaligned sampling, cluster sampling etc.

Though a number of sampling schemes are available but very little information is available in the literature about the sampling schemes used by authors for sub-pixel classification and accuracy assessment. In Woodcock and Gopal, (2000), the sampling sites were randomly selected. Latifovic and Olthof, (2004) have used sampling that was stratified by ecozones. Within each stratum, scenes were selected to closely resemble land cover proportions over the ecozones. Finally chi-square tests were used to verify the sample distribution with actual land cover distribution.

The numbers of samples chosen for accuracy assessment by different authors are available in the literatures, but not much information is available about how this number has been chosen.

2.3.4 Limitations of Accuracy Assessment Process :(Foody, G.M., 2002)

The failure to attain the specific target levels of accuracy is typically taken as the failure of remote sensing as a source of land cover information. Several interrelated problems that limit the quantification of classification accuracy can be listed as:

1. Compatibility of Resolution: – Comparison of two data sets requires same resolution parameters, so that they may be compared cell by cell and the cell represents the same area on the ground.

2. Registration Precision: – The reference image should be perfectly registered with respect to the classified image, so that no spatial error is accompanied. Without proper registration, the generation of error matrix itself will be erroneous. Geo referencing and geo coding of reference data with testing data of classified image is core of their comparison .A little error may propagate as the wrong comparison of land cover features.

3. Non availability of ground data in soft mode: -comparison of sub-pixel classified output may be done with hard reference data but some accuracy measures require it only in soft mode. To obtain soft reference data is the most typical task of the process.

4. Accuracy of the ground or reference data: - It is just an assumption that reference data is accurate representation of reality. In fact, it is another classification which itself may have error in it.

5. Errors should be weighted: - Errors are the misallocation of ground feature to some other class. Some errors are more important or damaging than others (foody, 2002)

Various Measures of Accuracy Assessment of Sub- Pixel Classification are discussed as follows.

2.3 Distance Measures

To derive the distance between two sub-pixel datasets, distance index is obtained for two fuzzy membership distributions or for two probability distributions

2.3.1 Distance Measure for Two Fuzzy Membership Distributions

A number of distance indexes has been proposed by Zwick, R., et al (1987) to measure the distance between two fuzzy datasets which include Euclidean distance, city block distance (L distance), D distance etc. The generalized distance function of these measures is known as Minkowski r-metric and is defined as,

$$d_r(1_\mu,2_\mu)=\left[\sum_{i=1}^{c}\left|1_{\mu_i}-2_{\mu_i}\right|^r\right]^{1/r},r\geq 1 \qquad \text{..............................Eq(2.1)}$$

Where, 1_{μ_i} = membership grade of class i in a pixel of reference image

2_{μ_i} = membership grade of class i in the same pixel of classified image

The cases of r = 1 and 2 were first studied by Kaufmann (1975), i.e the L distance (city block distance) and Euclidean distance respectively. Distance measure d_2^2 or distance D was proposed by Kacprzyk (1976).

All the distance measures like Euclidean distance S (Eq 2.2 and Eq 2.3) and D (Eq 2.4 and Eq2.5) [Kent and Mardia, (1988), Foody (1996), Foody and Arora, (1996)], and L (Eq 2.6 and Eq 2.7) distance [Foody and Arora, (1996)] has been found to provide suitable indexes of accuracy for dealing with fuzzy reference as well as classified data.

The formulations for all the distance measures are provided in appendix A, table 1.

2.3.2 Distance Measure for Two Probability Distributions

For evaluating the metric distance between two probability distributions, measure of cross entropy and information closeness measure had been proposed by Higashi and Klir, (1983), which had been used by Foody, (1995 and 1996) for the accuracy assessment of soft classification. Cross entropy actually refers to the relative entropy between two probability distributions. The definition of cross entropy as given in Eq (2.8) had been first introduced by Kullback, S (1968) as a distance measure between two probability distributions.

One of the measures to express information closeness between two probability distributions is cross-entropy or directed divergence (d). This is to measure the closeness of the probability distribution in each pixel of the ground data $({}^{1}p)$ and that of the classified image $({}^{2}p)$.

$$d({}^{1}p,{}^{2}p) = -\sum_{x} {}^{1}p(x)\log_{2}{}^{2}p(x) + \sum_{x} {}^{1}p(x)\log_{2}{}^{1}p(x) \quad \text{.............Eq(2.8)}$$

The measure of cross-entropy is appropriate for two probability distributions ^{1}p and ^{2}p only when the supports of the two probability distributions are compatible [Higashi and Klir, (1983), Foody, (1996), Chang et al, (1994)]. To overcome this disadvantage, a generalized measure of information closeness (I) had been introduced by Higashi and Klir, (1983) as follows.

$$D({}^{1}p,{}^{2}p) = d({}^{1}p,\frac{{}^{1}p+{}^{2}p}{2}) + d({}^{2}p,\frac{{}^{1}p+{}^{2}p}{2}) \quad \text{...................................Eq(2.9)}$$

Foody (1995 & 1996), Zhang & Kirby, (1997) and Zhang & Foody (1998) have used the measure of information closeness as an accuracy assessment measure for fuzzy reference and classified data. This method is appropriate for two probability distributions, where the summation of probabilities of different classes in a pixel is equal to 1. Thus, to apply these measures to fuzzy ground and classified data, the membership distribution for each pixel is required to be normalized, so that their summation is 1. The measures of information closeness have been applied for the fuzzy output and fuzzy reference data which have been converted to produce new normalized values that sum up to 1 for every pixel.

2.3.3 Discussions on Distance Measure

Distance measures such as D, L and S have been used successfully to find the distance of two sub-pixel datasets either fuzzy or probabilistic in per-pixel basis and for the overall image. But here also no attempt has been made to find the accuracy in terms of distance between the same class representation of the reference and classified data.

Distance measure such as d and I are used to find the distance between two probabilistic datasets. But the logarithmic gain formula suffers from the facts that it provides undefined results for zero class probability or membership values. Thus absence of any class in any pixel would lead to undefined results for cross entropy and consequently information closeness measures.

2.4 Fuzzy Set Based Measures

The concept of fuzzy set was first introduced by Zadeh (1965) for dealing with vagueness in complex systems, and represents a generalization of crisp sets to situations, where the class memberships of single elements cannot be sharply defined. The principle behind fuzzy set theory is that the situation of one class being exactly right and all other classes being equally and exactly wrong often does not exist. Conversely, there is a gradual change from membership to non-membership (Gopal and Woodcock 1994). Thus, in the case of remotely sensed images, rather than assigning individual pixels to just a single class, each pixel may be associated with every class with variable degrees of class membership. (C. Ricotta, 2005)

Several fuzzy set based accuracy assessment measures have been developed by different authors to check the accuracy of fuzzy classification. A brief review of different fuzzy set based measures has been discussed in the following sections.

2.4.1. Fuzzy Error Matrix

The concept of error matrix to fuzzy reference and classified data has been first elaborated by E. Binaghi et al, (1999) as given below.

Let, number of dominant classes present in the reference data is n, i.e., this is denoted by $(r_i)_{i=1,\dots,n}$. Fuzzy classification for n classes is done for the image and the classified data is denoted by $(c_i)_{i=1,\dots\dots,n}$. Number of pixels considered for classification is m.

And, μ_{ri}^{k}= membership grade of i'th class in k'th pixel of reference data

μ_{cj}^{k}= membership grade of j'th class in k'th pixel of classified data

The fuzzy error matrix for each pixel is shown in Table 3.1 and that for overall image in Table 3.2

Any element of i'th row and j'th column of error matrix is obtained as,

a_{ij}^{k}= OPERATOR(μ_{ri}^{k}, μ_{cj}^{k})..Eq(2.10)

Operator may be, for instance, Boolean, multiplication, minimum and composite depending upon the membership of the pixel. Their suitability is shown through table in the next section.

Here, for instance, in Binaghi approach the MIN operator has been used, which indicates the intersection or aggregation of pixel in both the data sets.

The diagonal elements in fuzzy error matrix shown in Table 2.1 represent the membership grades of the classes correctly classified. The off-diagonal elements stand for error of omission or commission, the same as in the conventional classification. The total grades for each row is the membership grade of corresponding class in the classified data and that for each column is the membership grade of corresponding class in the reference data.

TABLE 2.1 FUZZY ERROR MATRIX FOR K'TH PIXEL

Reference data

Classified data

	r_1	r_2	. . . .	r_n
c_1	a_{11}^k	a_{12}^k	. . . .	a_{1n}^k
c_2	a_{21}^k	a_{22}^k	. . . .	a_{2n}^k
.	.	.	. . . .	.
.	.	.	. . . .	.
.	.	.	. . . .	.
c_n	a_{n1}^k	a_{n2}^k	. . . .	a_{nn}^k

TABLE 2.2 FUZZY ERROR MATRIX FOR OVERALL IMAGE

Reference data

Classified data

	R_1	R_2	. . . .	R_n
C_1	A_{11}	A_{12}	. . . .	A_{1n}
C_2	A_{21}	A_{22}	. . . .	A_{2n}
.	.	.	. . . .	.
.	.	.	. . . .	.
.	.	.	. . . .	.
C_n	A_{n1}	A_{n2}	. . . .	A_{nn}

: Where, $A_{ij} = \sum_{k=1}^{m} a_{ij}$

Error matrix has also been used by Shalan et al for evaluating the accuracy of sub-pixel classification. Error matrix had also been generated by Jager and Benz (2000) for sub-pixel data and the approach for generation of error matrix is similar as the approach of Binaghi (1999).Approach of generation of error matrix for the whole image as given in Jager and Benz (2000) is explained below

The measures of accuracy such as overall accuracy, user's accuracy, producer's accuracy etc. are derived from the fuzzy error matrix as follows:

Table 2.3 Accuracy measures in fuzzy error matrix (Binaghi et al., 1999)

Accuracy	Pixel matrix	Image matrix
Overall Accuracy	$\frac{\sum_{i=1}^{n} a_{ii}}{\sum_{i=1}^{n} \mu_{ri}^{k}}$	$\frac{\sum_{i=1}^{n} A_{ii}}{\sum_{k=1}^{m}\sum_{i=1}^{n} (\mu_{ri}^{k})}$
User's accuracy	$\frac{a_{ii}^{k}}{\mu_{ci}^{k}}$	$\frac{A_{ii}}{\sum_{k=1}^{m} \mu_{ci}^{k}}$
Producer's accuracy	$\frac{a_{ii}^{k}}{\mu_{ri}^{k}}$	$\frac{A_{ii}}{\sum_{k=1}^{m} \mu_{ri}^{k}}$

Let, R_i be the set of membership grades of class i in the reference data over all the pixels, i.e.,

$$R_i = (\mu_{ri}^{1}, \mu_{ri}^{2}, \ldots\ldots\ldots\ldots \mu_{ri}^{m})$$

And, let C_j be the set of membership grades of class j in the classified data over all the pixels, i.e.,

$$C_j = (\mu_{cj}^{1}, \mu_{cj}^{2}, \ldots\ldots\ldots\ldots \mu_{cj}^{m})$$

Then each element of error matrix of i'th row and j'th column is obtained as,

$$A_{ij} = R_i \cap C_j = \sum_{k=1}^{m} [\min(\mu_{ri}^{k}, \mu_{cj}^{k})] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Eq (2.11)}$$

The accuracy in this case was judged by using different fuzzy similarity measures. For full similarity, the similarity index is 1 and for no similarity it is 0.

Jager and Benz (2000) have used a hypothetical image to test the accuracy and different fuzzy similarity indexes, such as fuzzy overall accuracy, fuzzy Hellden's accuracy etc have been derived for the classified data.

Advantages:

1. Error matrix is superior to other measures since it is able to represent the individual category measures and the situations of overestimation and underestimation, i.e., more information about accuracy of classification can be obtained from error matrix rather than only a single index of accuracy [Binaghi et al (1999)]

Disadvantages:

1. For every pixel separate class by class matrix is prepared, which consumes a lot memory of computer and increased computation includes time complexity consideration.

2.4.2. Fuzzy Similarity Measures: Fuzzy Neighborhood and Category Vector

Another fuzzy set based approach for measuring the similarity between a set of pixel based classifications has been used by A. Hagen (2002). Fuzzy category and neighborhood vectors have been developed by him for measuring the accuracy or similarity between two hard datasets. If the number of categories present in an image is C, then the fuzzy category vector for a pixel is defined as,

$$V_{cat} = \begin{pmatrix} \mu_{cat,1} \\ \mu_{car,2} \\ \cdot \\ \cdot \\ \cdot \\ \mu_{cat,C} \end{pmatrix} \qquad \text{...Eq(2.12)}$$

where, $\mu_{cat,i} = 1$ for original category and, $0 \leq \mu_{cat,i} \leq 1$ for other categories.

In representing the fuzziness of location the effect of neighborhood is taken into account. The different membership contributions of the neighboring cells are combined by calculating the fuzzy union of all neighboring cells multiplied by their distance based membership.

Fuzzy neighborhood vector for a pixel is thus defined as,

$$V_{nbh} = \begin{pmatrix} \mu_{nbh,1} \\ \mu_{nbh,2} \\ \cdot \\ \cdot \\ \\ \mu_{nbh,c} \end{pmatrix} \qquad \text{.....Eq(2.13)}$$

where, $\mu_{nbh,i} = \left| \mu_{cat,i,1} \times m_1, \mu_{cat,i,2} \times m_2, \ldots\ldots, \mu_{cat,i,N} \times m_N \right|$

$\mu_{cat,i,j}$ = Membership of category 'i' for neighboring pixel 'j' [0 or 1 in A. Hagen, (2000)]

m_j = distance based membership of neighboring cell 'j'

The similarity of the same pixel between two fuzzy classified images has been found by taking the intersection of the fuzzy neighborhood vectors of the same pixel in two images. An intersection that is greater than 0.5 indicates good similarity.

Two way similarity:-

Another measure of similarity as used by Hagen, A., (2002) is "two way similarity", which overcomes the disadvantage of the first method of carrying out the comparison excluding the cell itself (Eq 2.14). In this case, first fuzzy neighborhood vector of a pixel in one image (say, image 1) is compared (intersection) to the crisp vector (formed by taking μ_{cat} – 1 for original category and 0 for other categories in V_{cat}) of the same pixel in the other image (say, image 2). Then the crisp vector of the same pixel in image 1 is compared to the fuzzy neighborhood vector of image 2. Finally, the lower of the two comparison results establishes the similarity at that location.

Limitation: Only cell by cell comparison, not overall image comparison.

Advantage: neighbor hood concept is involved

2.4.3 Fuzzy Functions

There is another fuzzy set based approach for accuracy assessment for fuzzy classification by using different fuzzy functions. This is a fuzzy possibilistic approach for assessing the accuracy. It is different from the abovementioned fuzzy methods in the sense that it is applied when the reference is fuzzy and the classification output is hard [S. Gopal and C. Woodcock, (1994)]. In this method a linguistic scale of reference based on the expert evaluation is constructed, each pixel is then assigned a membership grade corresponding to each class. Finally the accuracy is judged based on the frequency of matches and mismatches, magnitude of error, source of error and nature of error. No single accuracy index is available in this case.

Frequency of error measures the accuracy of the map in terms of the matches and mismatches of the classified data with respect to the reference data. A pixel is assigned to a particular class for which the membership grade is the highest using the 'Max' function (Eq 2.16) or 'Right' function (Eq 2.17) (Appendix A, Table 2). Similarly the magnitude, source and nature of error all are found out by applying different functions as given in (Eq 2.18, 2.19, Eq2.20, Eq 2.21.

In this way, an idea of different types of errors and their distribution can be found out by using these measures. This idea is helpful in formation of error models, which is significant in GIS analysis. Though these measures are applicable for pixel based classified output, an estimate of area can be obtained using these measures [C. Woodcock and S. Gopal, (2000)].

These measures have been used by C. Woodcock and S. Gopal, (2000), De Gloria et al, (2000), Thenkabail, P, S et al (2005), Laba, M, et al (2002) etc. for accuracy assessment of thematic maps. The sample sites were visited by the experts and then the measures of accuracy assessment were applied. But these measures do not provide any overall measure of accuracy of classification. So, there is no option to compare two classifications with respect to a single reference using these methods. Also, the expert evaluation may vary from one expert to other. In such case, there is a need of standardization of results from a number of experts. Finally, the methods have only been used for pixel based classification. But nowadays, newer techniques for sub-pixel level classification are being derived to include most of the information about the different types of land covers present in an image and it is much more difficult to obtain a fuzzy reference data than to obtain a fuzzy classification. So, the methods should be improved so that they can be applied for fuzzy reference as well as fuzzy classification output.

Discussions on Fuzzy set Based measures

Among the fuzzy set based measures only fuzzy error matrix attempts to find the different accuracy indexes such as overall accuracy, user's and producer's accuracy etc. But since fuzzy classification does not consider the spatial accuracy, so the attribute accuracy regarding the user's and producer's accuracy may produce misleading results in many situations.

Development of fuzzy neighborhood and category vector to find the accuracy in terms of similarity is related to pixel based classification. So this approach has nothing to do with fuzzy classification and its accuracy assessment.

Fuzzy functions like max function, right function etc are used to find the accuracy of a sub-pixel classification with respect to a fuzzy reference.

The disadvantage is collection of fuzzy reference data is much more difficult than producing fuzzy classification.

2.5 Uncertainty Measures

According to eminent scientist Albert Einstein "So far as the law of mathematics refer to reality, they are not certain .And so far as they are certain, they do not refer to reality".

Uncertainty is the inability to decide what to do or not, to perform any specific task .In digital image classification uncertainty is categorized in two ways; Ambiguity and Vagueness. The uncertainty in class allocation can be measured in terms of entropy [Foody (1995), Foody (1996), Zhang and Kirby (1997), Zhu, A-Xing, (1997)].Various entropy measures, based on fuzzy and probabilistic data sets, are the tools to mathematically quantify these aspects.

3.2.1 Ambiguity vs. Vagueness

Ambiguity is associated with one to many situation and conflicts of evidences.(Klir and Folger,1988). In remote sensing , this kind of uncertainty arise when , there is confusion regarding allocation of the pixel to a class, as it may be having equal proportion of all the classes . Entropy is a measure of such kind of uncertainty.

Vagueness is associated with the difficulties of making precise distinction. In mapping it is considered as with the problem if locating a sharp dividing line between two continuous classes. Shanon entropy is the generalized tool to quantify such uncertainty.

The entropy that has been used in different literatures for measuring the uncertainty is given in Table 3.4, Eq (2.22). The measure (Eq 2.22) used by Foody is used to find the entropy of a probability distribution [Maselli, M et al (1994)]. For finding the entropy of possibilistic data, the data has been normalized to a probability distribution so that summation of membership grades for each pixel equals to one [Foody (1996)]. The value of entropy is low for pure pixels and high for mixed pixels. If reference data consists of only pure pixels, the entropy value will be zero for the reference data. However, if the classification gives rise to mixed pixels, then it is understood easily that there is some error in classification and that will be reflected with a higher entropy value for the classified image.

Shannon's entropy as given by Eq. (2.23) has been found in S.K. Pal et al (2000), Ghosh, J.K. (1996), De Luca and Termini (1972) etc. but this is not used for accuracy assessment. This measure has been used to measure the uncertainty of fuzzy membership distribution, both normalized and not normalized.

Table2.4 Measures of Uncertainty

Measure	Formulation	Explanation	Reference
Entropy used by Foody	$H(p) = -\sum_{x} p(x)\log_2 p(x)$ ....Eq(2.22)	Measures the uncertainty (in terms of ambiguity) of information content in a probability distribution	1. Foody, 1995 2. Foody, 1996 3. Zhang and Kirby, 1997
	Symbols: x is the class variable ; P(x) represents the membership probability of pixel to class x ; H(p) represents the entropy of pixel		
Shannon Entropy	$H(\mu)=\frac{1}{n\ln(2)}\sum_{i=1}^{n} S_n(\mu(x))$ ....Eq(2.23) Where, $S_n(\mu(x))=-\mu(x)\ln\mu(x)-\{1-\mu(x)\}\ln\{1-\mu(x)\}$	Measures the uncertainty (in terms of vagueness) of information content in a fuzzy membership distribution.	1. Ghosh. J.K, 1996 2. Pal. S.K et al, 2000. 3. Shannon, 1998
	Symbols: x is the class variable ; $\mu(x)$ is the membership of the pixel to class x ; s ($\mu(x)$) is the Shanon parameter.		

2.5.2 Advantages and Disadvantages of Uncertainty Measures

Advantages: –For a probabilistic dataset, i.e., when the class probability values over each pixel sum up to 1, then entropy measure can be used very effectively to compute the information content or uncertainty contained in a dataset.

Disadvantages: - Entropy measure is unsuitable while dealing with fuzzy classifications where summation of membership grades for each pixel does not produce the value of one. Both of the entropy measures, as mentioned in Table3.6 suffer from the common disadvantage that the use of logarithm in the measures restricts their use for only those mixed pixels in which all the classes are present with a membership value greater than 0 and less than 1. But for classes with zero membership value, the measures give rise to undefined solutions. To overcome this limitation, logarithmic gain can be replaced by exponential gain, i.e., $\ln(\mu(x))$ should be replaced by $e^{(1-\mu(x))}$ [Pal, N.R., and Pal. S.K., (1991)]. .The other disadvantage of this measure is that it does not provide any information regarding the accuracy of individual classes in the image.

2.6 Similarity Measure

Similarity measure provides the similarity between two datasets. It is of two types, one dealing with the probabilistic output and the other dealing with fuzzy output. A summary of the similarity measures discussed in this section has been given in Table 3, Appendix A.

2.6.1 Similarity Measures for Probabilistic Data

Different types of similarity measures for probabilistic output have been stated in different literatures, such as correlation coefficient [Foody (2000), Zhang and Foody (1998), Atkinson, P,M, Foody and Cox (1994), Maselli et al (1996)], root mean square error (RMSE) [Foody, (2000), Atkinson, P.M., Martens, K.C., et al], expected sets shared (ESS) [Ricotta, C., (2004)] etc.

2.6.1.1 Pearson's Correlation Coefficient

Pearson's correlation coefficient (CC) (Appendix A, Table 3, Eq 2.24) finds the correlation or similarity among two representations of the same class in both reference and classified data. The value of CC is maximum (1), when there is absolute correlation or no error, and minimum (-1) when there is a totally opposite correlation. Its value is 0 when there is no correlation among the datasets.

$$CC_i = \frac{Cov(1_{\mu i}, 2_{\mu i})}{\sigma_{1_{\mu i}} \times \sigma_{2_{\mu i}}} \quad \text{.......... Eq 2.24}$$

2.6.1.2 RMS Error

Root mean square (Appendix A ,Table 3, Eq 2.25 and Eq 2.26) error also finds the error in any class representation in the classified data with respect to the reference data and the corresponding accuracy in terms of similarity is (1-RMSE). RMSE is 0 if there is no error and it increases as deviation between the representation of the same class in the two datasets increases. But the problem with RMS error is that it is not standardized by any measure of variance. Thus it is large for larger dataset and small for smaller dataset, irrespective of the correlation among datasets. So, correlation coefficient is a better alternative which gives acceptable information about accuracy or similarity.

$$RMSE_j = \sqrt{\frac{\sum_{i=1}^{n}(1_{\mu_{ij}} - 2_{\mu_{ij}})^2}{n-1}} \quad \text{.......... Eq 2.25}$$

$$\frac{\sum_{j=1}^{c} RMSE_j}{c} \quad \text{.......... Eq 2.26}$$

2.6.1.3 Expected Sets Shared (ESS)

ESS is a measure of accuracy assessment developed by C. Ricotta (2004). The term expected sets shared means the number of classes common to the same pixel in both reference and classified data.

Mathematically the expected sets shared for the i'th pixel is obtained as given in Eq(3.27).

$$ESS_i(\mu_A, \mu_B; m) = \sum_{k=1}^{n}(1-(1-\mu_{Ak})^m)(1-(1-\mu_{Bk})^m) \text{ ....Eq(2.27)}$$

where, $\sum_{k=1}^{n}\mu_k = 1$, for each pixel.

μ_A *and* μ_B denote membership vectors of the same pixel (say i'th) in two sample datasets (such as reference and classified),

'n' is the number of class.

Then ESS is the expected number of common classes in the i'th pixel of the two samples. As the value of the parameter 'm' is increased, where 'm' is the sample size, the curve of ESS vs 'm' converges to the number of common classes between the two samples, which have non-zero membership values. Some other measures such as symmetric NESS (Normalized expected sets shared) or asymmetric $NESS_A$ or $NESS_B$ have also been derived. The overall ESS for the whole image is simply the average of the ESS of all the pixels. The pixel based ESS allows checking pair wise similarity and thus helps in finding out the problematic areas in the map.

2.6.2 Similarity Measure for Fuzzy Datasets

The similarity measures for fuzzy output can also be obtained by using different similarity indexes (Eq 3.28) [Townsend (2000); Jager and Benz (2000), Chen, S, (1995)] and fuzzy correlation coefficient [Pal et al, (2000)]. This type of similarity index expresses the similarity between two fuzzy sets. The results express the correlation similarly as for probabilistic correlation; only the formulation is different in this case. However, fuzzy correlation coefficient has not been used by any author to judge the accuracy of fuzzy classification. The formulations of all the similarity measures are given in Appendix A, Table 3.

2.6.3 Discussions on Similarity Measures

Different similarity measures have been developed by different authors to provide information about the similarity between two sub-pixel datasets. But in most cases the similarity indexes have been developed for probabilistic dataset, e.g correlation coefficient, RMS error, expected sets shared etc. So, similarity of fuzzy datasets should not be obtained by using these measures.

2.7 Fuzzy correlation coefficient

There are several sub-pixel accuracy assessment measures proposed by different authors. A review of the same can be found in Ghosh and Mukherjee (2005). It has been observed that most of the measures assess accuracy qualitatively. So, quantitative measures to evaluate accuracy of a fuzzy classification are warrant of the situation.

Statistical correlation coefficient, also known as Pearson's correlation coefficient, has been used as a measure to assess the accuracy of sub-pixel classified images (Foody, 2000) quantitatively. This measure is based on statistical concepts i.e., on the theory of probability. It is based on the assumption that both the variables, whose correlation is required to be found, are normally distributed (Davis, 1986). It uses the mean and standard deviation of the data to find the correlation among datasets. Thus, it is appropriate to restrict the measure in assessment of accuracy of sub-pixel classification based on statistical concepts.

Fuzzy classification gets carried out based on fuzzy concept, more generally on the theory of possibility. Thus, to check the accuracy of fuzzy classification, it is also required to have some method of based on fuzzy concept. Moreover, to evaluate the accuracy of a fuzzy classification, different measures are required such as measure for accuracy assessment of any particular class in any particular pixel, measure for accuracy assessment of any particular pixel consisting of different sub-pixel classes, measure for accuracy assessment of different types of classes (sub-pixels) present in an image or of an image classified into different sub-pixel classes. Thus, there is a need for an accuracy assessment technique compatible to the fuzzy classification.

Pal and Dutta Mazumder (1986) has introduced a measure termed as fuzzy correlation to find the correlation between two (fuzzy) properties of an image (defined by two membership functions). Same measure has been used by Pal & Ghosh (1992) and Pal et al (2000) for segmentation of image.

The objective of this paper it to test the viability of some measures based on fuzzy correlation co-efficient for different accuracy assessments of fuzzy classification.

Measures Based on Fuzzy Correlation Co-efficient

Let $\mu_{r(o,i)}$, $\mu_{c(o,i)}$ be the membership values of reference and classified data any class/object (o) in any pixel (i) in some domain, say Ω. Modified from the definition available in (Pal and Dutta Mazumder, 1986), the different types of measures (fuzzy correlation co-efficient), for assessment of accuracy of fuzzy classification of remote sensing data, can be defined as follows:

__Class based__

(A) One Class in a pixel

Measure for accuracy assessment of a class (o) in any pixel (i), say $C_{o,i}$, can be given by the Equation (2.29) as follows:

$$C_{o,i} = 1 - \frac{4}{[2\mu_{r(o,i)} - 1]^2 + [2\mu_{c(o,i)} - 1]^2}[\mu_{r(o,i)} - \mu_{c(o,i)}]^2 \quad \text{....Eq 2.29}$$

(B) One Class in an image

Measure for accuracy assessment of a class (o) in any image (of n pixels), say C_o, can be given by the Equation (2.30) as follows:

$$C_o = 1 - \frac{4}{X_{r(o)} + X_{c(o)}} \sum_{i=1}^{n} [\mu_{r(o,i)} - \mu_{c(o,i)}]^2 \quad \text{--------- Eq 2.30}$$

Where $X_{r(o)} = \sum_{i=1}^{n} [2\mu_{r(o,i)} - 1]^2$ and $X_{c(o)} = \sum_{i=1}^{n} [2\mu_{c(o,i)} - 1]^2$

Pixel based

Measure for accuracy assessment of any pixel (i) having m types of classes in it, say C_i, can be given by the Equation (2.31) as follows:

$$C_i = 1 - \frac{4}{X_{r(i)} + X_{c(i)}} \sum_{o=1}^{m} [\mu_{r(o,i)} - \mu_{c(o,i)}]^2 \quad \text{---------- ....Eq 2.31}$$

$$X_{r(i)} = \sum_{o=1}^{m} [2\mu_{r(o,i)} - 1]^2 \text{ and } X_{c(i)} = \sum_{i=1}^{m} [2\mu_{c(o,i)} - 1]^2$$

Image based

Measure for accuracy assessment of an image, say C, consisting of n pixels with m classes in each pixel can be given by an Equation (2.32) as follows:

$$C = 1 - \frac{4}{X_{r(o,i)} + X_{c(o,i)}} \sum_{i=1}^{n} \sum_{o=1}^{m} [\mu_{r(o,i)} - \mu_{c(o,i)}]^2 \quad \text{............Eq 2.32}$$

Where $X_{r(i)} = \sum_{i=1}^{n}\sum_{o=1}^{m}[2\mu_{r(o,i)} - 1]^2$ *and* $X_{c(i)} = \sum_{i=1}^{n}\sum_{i=1}^{m}[2\mu_{c(o,i)} - 1]^2$

The properties which these measures possess are given below:

(i) For high or low values of both $\mu_{r(o,i)}$ and $\mu_{c(o,i)}$, measure value is high and designates high degree of accuracy in classification. Particular measure represents the degree of accuracy of that particular category.

(ii) For high values of $\mu_{r(o,i)}$ and low values of $\mu_{c(o,i)}$ or vice versa, measure value is low thus indicates a low degree of accuracy in classification.

(iii) $|$ Measure value$| \leq 1$, for all $\mu_{r(o,i)}$ and $\mu_{c(o,i)}$

(iv) Measure value for $\mu_{r(o,i)}$ and $\mu_{c(o,i)}$ is same for that of $\mu_{c(o,i)}$ and $\mu_{r(o,i)}$.

(v) Measure value for $\mu_{r(o,i)}$ and $\mu_{c(o,i)}$ is same for that of $(1-\mu_{r(o,i)})$ and $(1-\mu_{c(o,i)})$

Thus, the accuracy of fuzzy classification is thus expected to be assessed by using the measures based on fuzzy correlation coefficient.

2.8 .Operator In Error Matrix: new possibilities

Pontius, R.G, 2006 et al. have developed a generalized cross tabulation matrix, where the entry in the cell of matrix is not always traditional, like Boolean operator for hard data set or minimum operator for fuzzy data set. Some new operators are also introduced .Which operator should be used when; such kind of suitability is shown in table 3.4. Mathematical formulation of these operators and their assessment are as follows;

NOTATION: - ***Pnij*** is the entry in row i and column j.

Pni* denote the membership of pixel into class i, value between 0 and 1, both inclusive.

Pn*j denote the membership of pixel into class j, value between 0 and 1, both inclusive.

Boolean operator is for crisp classification, for instance,

$$Pnij = 1 \text{ if } Pni\bullet = Pn \bullet j = 1$$
$$0 \text{ else} \quad \text{.............................eq(2.33)}$$

2.8.1 MINIMUM Operator

Is most frequently used operator because it gives common region of aggregation but suitable only for fuzzy data

$$Pnij = MIN(Pni\bullet, Pn \bullet j) \text{.............................eq(2.34)}$$

2.8.2. MULTIPLICATION Operator

$$Pnij = Pni \bullet \times Pn \bullet j \text{..................................eq(2.34)}$$

2.8.3 COMPOSITE Operator

$$Pnij = (Pni \bullet - Pnii) \times \left[\frac{(Pn \bullet j - Pnjj)}{\sum_{j=1}^{J} (Pn \bullet j - Pnjj)} \right] \text{for } i \neq j \text{................................eq(2.35)}$$

Their utilization in the accuracy assessment depends on the circumstances. New term ontology is defined by the author to categorize the type of pixel.

Table on next page explain relation between the pixel and type of operator to be used .

Table 2.5 Suitability of Operators in Error Matrix (Pontius, R.G. and Cheuk, M. L., 2006)

Operator	Classification	Sum of entries equal to 1	Diagonal matrix for completely identical maps
BOOLEAN	Hard	Yes	Yes
	Each pixel has membership in exactly one class. The concept of location within pixel is irrelevant		
MINIMUM	Soft (Fuzzy data)	No	No
	Each pixel has membership according to fuzzy set theory in order to acknowledge ambiguity. The sum of the class membership can be different than 100%.		
MULTIPLICATION	Soft (Probabilistic data)	Yes	No
	Each pixel has membership in a class according to the probability that a randomly selected point within the pixel belongs to that classes. The concept of location within pixel exists in terms of infinitely small points, whose spatial distribution within the pixel is random.		
COMPOSITE	Soft (Probabilistic data)	Yes	Yes
	Each pixel has membership in a class according to the probability that a randomly selected point within the pixel belongs to that classes. The concept of location within pixel exists in terms of infinitely small points, whose spatial distribution within the pixel is random.		

Advantages:-

1. They have also considered the concept of location of the ground feature within the pixel.

2. Concept of weight is also considered, accordingly there are three common reasons why scientist would want to weight some pixel differently than other

A. Each pixel may represent substantially different amount of the area of the earth surface.

(Global or regional level maps)

B. Weight zero masks, pixel that are out side area of study and non zero weight to the pixels in analysis; in accuracy assessment zero weight can be assigned to the pixel where ground information has not been collected.

C. When coarse pixels are aggregated from a different no. of fine pixels and study areas is not a perfect square then assign weight to the pixels.

$$P+ij = \quad ij = \frac{\sum_{n=1}^{N} (W_n \times Pnij)}{\sum_{n=1}^{N} W_n} \quad \text{.....eq(2.36)}$$

Weights are included here to make standard entry.

2.9 Fuzzy kappa - new possibility

The similarity measures discussed in the earlier section represent similarity by a cell by cell comparison method. But it is often required to find the similarity of overall images. In such cases just taking the average similarity obtained over all the pixels do not serve the purpose since "the expected value of similarity would be strongly influenced by the number of categories in the map and also by the numerical distribution of the cells over those categories"[A. Hagen, (2003]. In such cases Kappa statistics has been introduced.

The Fuzzy Kappa is calculated in the same manner as the (crisp) Kappa, as shown in equation (3.12).

$$\text{Fuzzy Kappa} = K_{Fuzzy} = \frac{S-E}{1-E} \quad \text{.....Eq(2.37)}$$

where S is the overall similarity and E is the expected overall similarity. They are calculated as.

Expected overall similarity :

$$E = E(S) = \frac{\sum_{l=1}^{n} E(S_l)}{n} \text{ ....Eq(2.38)}$$

where $$E(S_l) = \sum_{i=1}^{z} (Pl,ixl,i) \text{ ....Eq(2.39)}$$

In other words, Pi is the probability distribution of the outcome of the similarity values which are in the vector Xi, and thus the expected local similarity can be calculated as the sum product of probability and similarity (see equation (3.14):

Equation (3.13) calculates the expected similarity as the average expected similarity over all cells:

Overall similarity:

$$S = \frac{\sum_{i=1}^{n} S_l}{n} \text{ ....eq(2.40)}$$

Equation (3.15) calculates the overall similarity of the cell, S_l by taking the minimum similarity of mapA to the category found in mapB at that location and vice versa:

' l ' is cell parameter variable .

$$S_l = \min(sim^A_{l,m^b_l}, sim^B_{l,m^A_L}) \text{ ....eq(2.41)}$$

Disadvantage

1. The calculation detailed in this paper can be time-consuming.

Advantage

1. It involves the neighborhoods impact or proximity relation.
2. Fuzziness of location and fuzziness of category, both are considered.

CONCLUSIONS

Measures like entropy was first developed to measure the amount of uncertainty in the information content of the classified data with respect to an ambiguity free pixel based reference data. Measures of distance were used to find a metric distance between two datasets. Information closeness is a distance measure applicable for two probability distributions and Euclidean or L distances are distance measures for any two distributions, probabilistic or possibilistic. To find the similarity in class representation between two datasets, similarity measures like correlation coefficient and RMS error were developed for probabilistic datasets, whereas different fuzzy similarity indexes based on fuzzy error matrix were developed to measure the similarity between two fuzzy datasets.

For getting more information about the classification accuracy, conventional error matrix is found to be the best. So, an attempt was made to generate fuzzy error matrix, which also has the capability of providing a number of accuracy measures like those obtained from conventional error matrix. Though error matrix was considered to be the best accuracy assessment tool for hard classification, but fuzzy set based error matrix is not that much popular till now and this is the reason why more techniques are still being developed. Fuzzy functions were another approach to check accuracy when the classified output is hard and reference data is soft. Though, by using this measure different types of errors are evident, but certain disadvantages are associated with this approach as discussed before for which this method does not impress much for accuracy assessment. Some recently developed statistical techniques of accuracy assessment are ESS (soft reference and classified data) and sub-pixel fractional error matrix (hard classified data and soft reference data) technique. But these measures do not take into account the fuzziness of dataset.

It is also observed that though several accuracy measures are continuously being developed in the field of sub-pixel accuracy assessment of classified images, but the literatures do not provide any detailed information about the sampling scheme, sample size, method of obtaining the samples etc. for soft classification, though this is a very important step in classification. In most of the papers, the reference data is collected from a higher resolution image. Thus the reference data is not truly fuzzy in nature, since the class membership values are class proportions only which sum up to give a value of 1 for each pixel. But fuzzy approach is different from probabilistic approach in the sense that the membership values are obtained from a predefined membership function. Weight and proximity relation of the pixel are also open for new research.

Due to the disadvantages of some of the measures already developed and used in different literatures, there is a scope to develop new methods for accuracy assessment of fuzzy classification that do not suffer from any such disadvantage.

CHAPTER 3

DATA SET AND METHODOLOGY

3.1 Data set

The study area is located in the North Cacher Hills district of Assam (INDIA) lying between the latitudes 25° 00' N to 25° 15' N and longitudes 92° 45' E to 93° 00' E covering an area of about700 square kilometers. The area is represented by Survey of India topographic sheet number 83 C/16 (at1:50,000 scale).An IRS (Indian Remote Sensing) 1D PAN Image (Path number 112 and Row number 054) of the region is being used for this study. The original image was resampled at 6 meter resolution (Figure 3.2) using nearest neighborhood sampling scheme. Image was geographically registered with topographic map (83 C/16) using poly conic projection. An image of 2048 pixel was cut from the original image and similar area was being collected from scanned topographic map (Figure 3.1).

An IRS (Indian remote sensing) satellite LISS 3 pan image of Assam region (map id 83 c16) is used for this study. Image was geographically registered with registered LISS image by poly conic projection.

3.1.1 Geographical distribution of the land cover features

The area is rich in vegetation, mainly covered with forest on hilly terrain. The area is drained by the Jatinga River. In the valley thin water streams are available. In the river region sandy area and agriculture land is found. Bare soil is also available which includes Urban and jhoom cultivation features in this study.

3.2 Methodology

Maximum likelihood (MLC) crisp classification of the PAN image has been used as reference data. In order to classify the PAN image, training data for five dominant classes present in the study area were extracted from a VDU through interactive software. Before collection of samples of data, locations of some training fields are marked on the hard copy of a FCC (IRS 1C LISS III) of the scene by visual interpretation corroborated through topographic map as base data and finally through field reconnaissance survey. Training samples consisting of 3325, 2525, 1843, 1764 and 2719 pixels were considered respectively for water, forest, agriculture, sandy area and bare soil land cover types. The statistics of training samples are as given in Table 3.1.The training samples (their statistics) were then used to classify re sampled

PAN image (Figure 3.2) by Maximum likelihood classifier (MLC). The accuracy of the classified image (Figure 3.3), at the training locations, was found to be 100 percent. The MLC classified image together with the topographic map was then used as reference data and was assumed to be 100% accurate. For each pixel in these reference data at 6m resolution, the land covers are known and considered to be one of the five classes (water, forest, agriculture, sandy area and bare soil) and assumed pure. The original re sampled PAN image was than degraded by using 3 by 3 low pass mean filter.

The PAN degraded image was then classified using MLC soft classifier and FUZZY classifier using pure (Table 3.2) and mixed training data (Table 3.3). The fragmented images from all the classifiers for five different land cover classes are as shown in Figure 3.5. The class compositions of the land cover classes at reference locations were then found from the classified fragmented images. These class compositions of the Fragmented classified images represent the fuzzy classified data (testing data) to be used for accuracy assessment. Thus, from the degraded image (Figure 3.4) mix training samples were being collected using MLC classified image as the reference image. The statistics of training samples are as given in Table 3.3.

Reference samples, consisting of 100 pixels for all the five broad land cover classes, were collected from the degraded PAN image using random sampling technique. The testing sites were mostly selected at the boundaries of the polygon features in order to get mixed pixels. Statistics of the reference samples are as given in Table 3.4. Using the reference data, actual class proportions of reference pixels in the degraded image were then computed using 3 X 3 windows in MLC classified image (Figure 3.2). These class compositions of the degraded PAN image represent the sub-pixel reference data to be used for accuracy assessment.

In order to collect testing samples for accuracy assessment of fuzzy classification, MLC soft classifier (Wang, 1990) and FUZZY classifier (Key et al, 1989) were being used in supervised mode.

While collecting training data from pan image some different land cover classes which can be identified to be separate with the help of topographic map as well as google earth images used as ancillary data they may give almost same range of DN values depending upon the surrounding effect. Mixed training data are given in appendix B. Statistics of mix training data was calculated by fuzzy mean and fuzzy standard deviation.

PAN image (Figure 3.2) by Maximum likelihood classifier (MLC). The accuracy of the classified image (Figure 3.3), at the training locations, was found to be 100 percent. The MLC classified image together with the topographic map was then used as reference data and was assumed to be 100% accurate. For each pixel in these reference data at 6m resolution, the land covers are known and considered to be one of the five classes (water, forest, agriculture, sandy area and bare soil) and assumed pure. The original re sampled PAN image was than degraded by using 3 by 3 low pass mean filter.

The PAN degraded image was then classified using MLC soft classifier and FUZZY classifier using pure (Table 3.2) and mixed training data (Table 3.3). The fragmented images from all the classifiers for five different land cover classes are as shown in Figure 3.5. The class compositions of the land cover classes at reference locations were then found from the classified fragmented images. These class compositions of the Fragmented classified images represent the fuzzy classified data (testing data) to be used for accuracy assessment. Thus, from the degraded image (Figure 3.4) mix training samples were being collected using MLC classified image as the reference image. The statistics of training samples are as given in Table 3.3.

Reference samples, consisting of 100 pixels for all the five broad land cover classes, were collected from the degraded PAN image using random sampling technique. The testing sites were mostly selected at the boundaries of the polygon features in order to get mixed pixels. Statistics of the reference samples are as given in Table 3.4. Using the reference data, actual class proportions of reference pixels in the degraded image were then computed using 3 X 3 windows in MLC classified image (Figure 3.2). These class compositions of the degraded PAN image represent the sub-pixel reference data to be used for accuracy assessment.

In order to collect testing samples for accuracy assessment of fuzzy classification, MLC soft classifier (Wang, 1990) and FUZZY classifier (Key et al, 1989) were being used in supervised mode.

While collecting training data from pan image some different land cover classes which can be identified to be separate with the help of topographic map as well as google earth images used as ancillary data they may give almost same range of DN values depending upon the surrounding effect. Mixed training data are given in appendix B. Statistics of mix training data was calculated by fuzzy mean and fuzzy standard deviation.

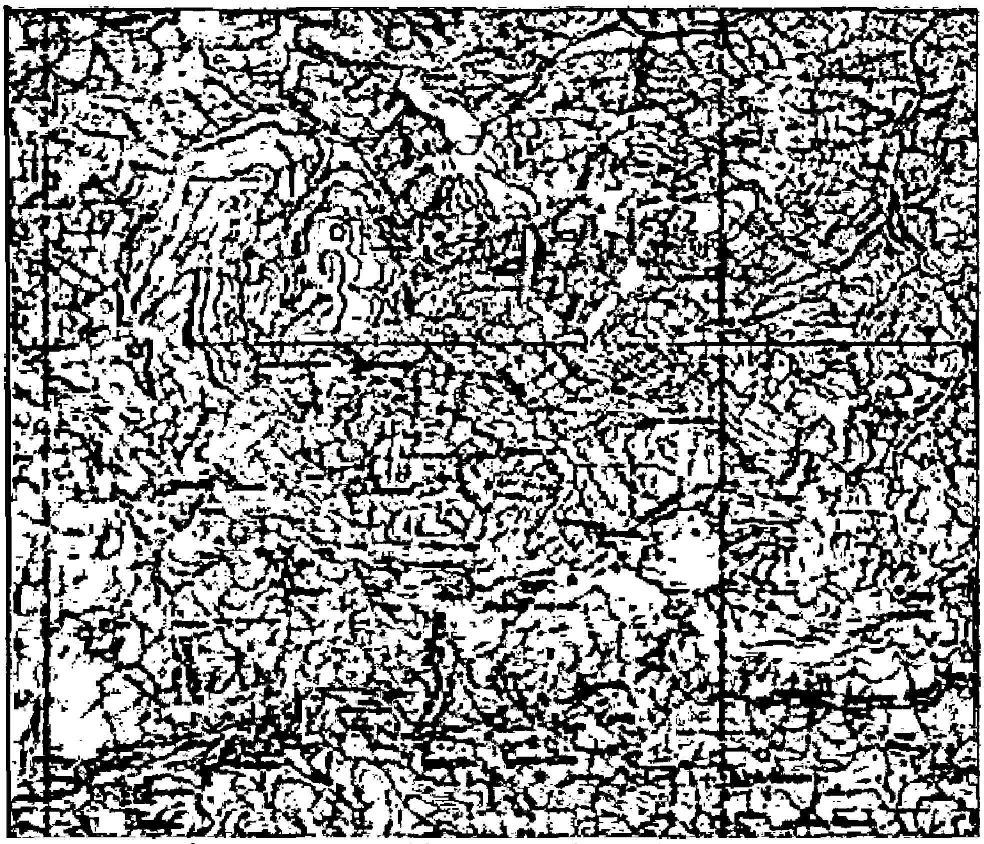


Fig3.1: Topographic map of the region (map id 83C16)

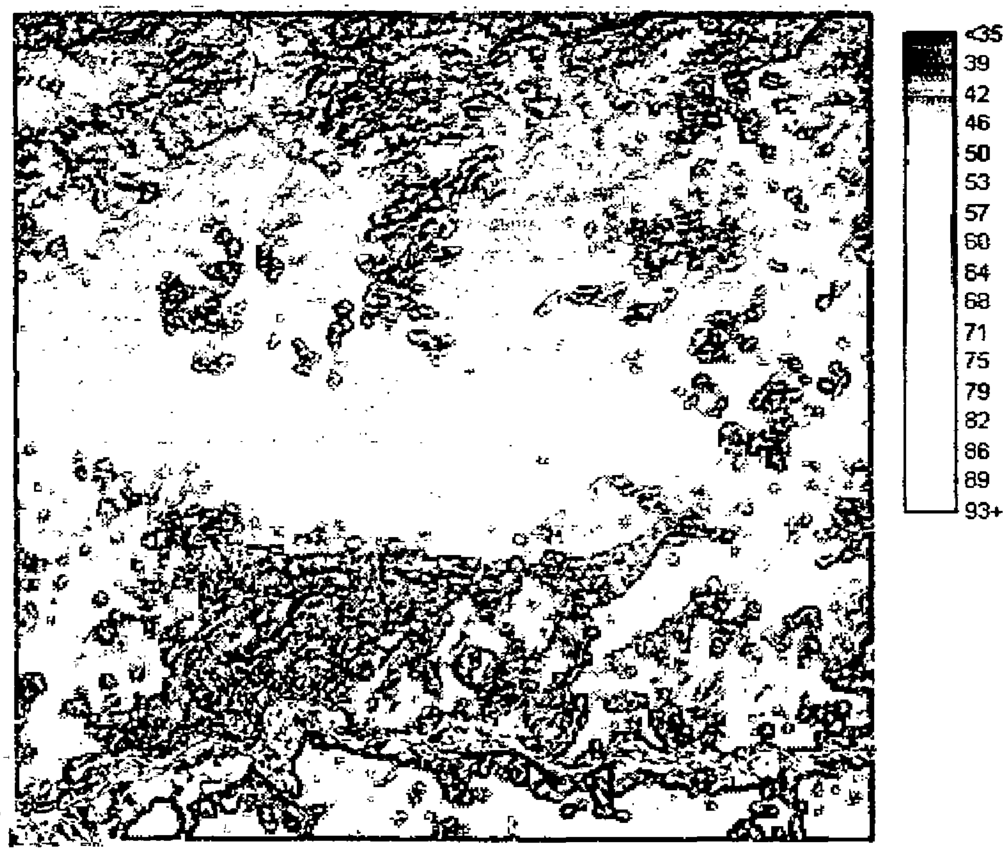


Fig 3.2: PAN image (cut of 2048 size), map number 83C16

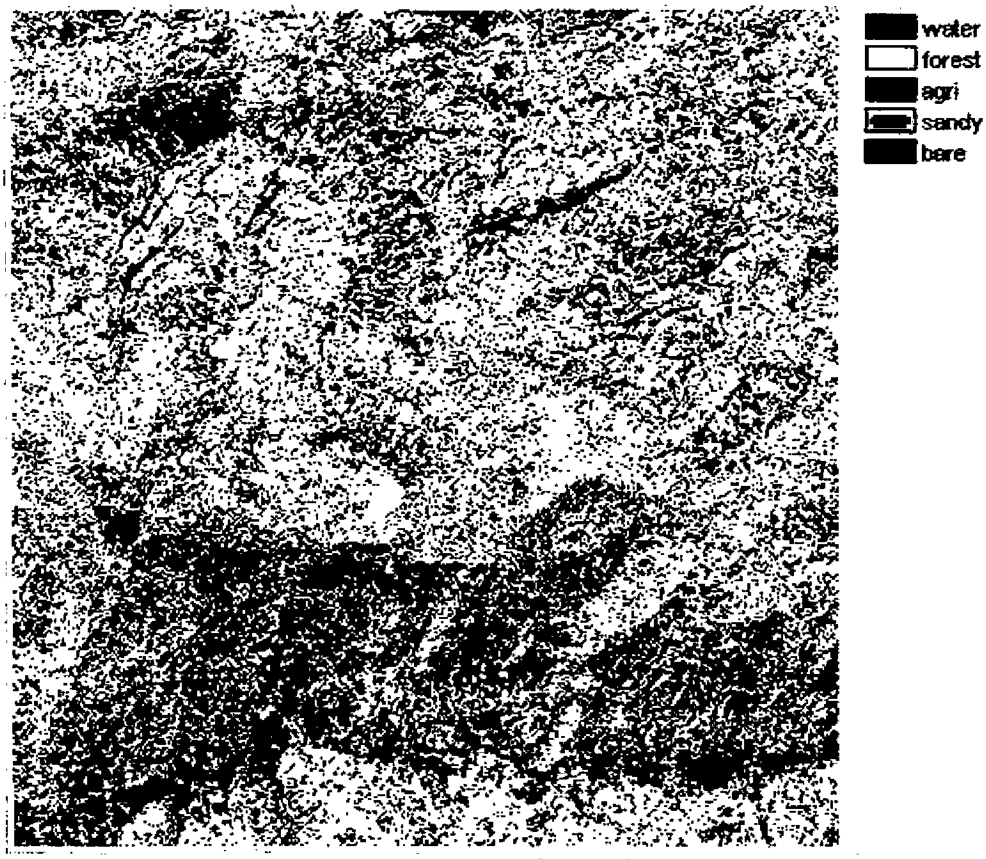


Fig 3.3: hard MLC classification of PAN image shown in fig 3.1

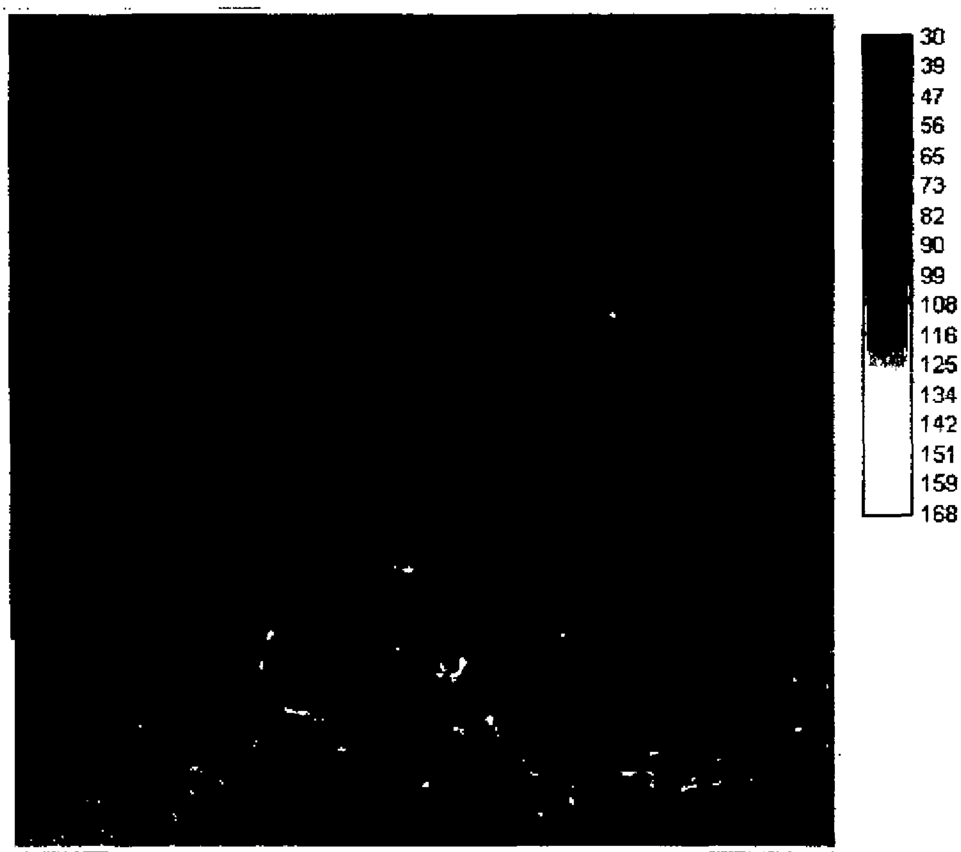


Fig3.4: degraded pan image by 3 by 3 mean filter

class	No of training pixel	Mi n	max	mean	Standard deviation
Water	3325	29	48	34.6	4.8
Forest	2525	43	60	51.3	6.4
Agriculture	1843	91	141	99.4	18.4
Sandy area	1764	73	89	80.4	8.5
Bare soil	2719	61	76	67.5	6.1

Table 3.1 Statistics of pure training data from PAN image

class	No of training pixel	Mi n	max	mean	Standard deviation
Water	3436	32	43	34.81	2.48
Forest	2924	47	57	51.58	2.6
Agriculture	1881	88	125	99.16	10.9
Sandy area	1991	73	87	79.9	6.7
Bare soil	2811	62	73	67.4	3.4

Table 3.2: statistics of pure training data to classify degraded pan image

Class	Number of pixel	Mi n	Max	mean	Standard deviation
Water	100	37	77	43.08	7.49
Forest	150	39	89	52.01	9.72
Agriculture	100	57	94	85.72	8.30
Sandy area	150	40	95	79.29	9.37
Bare soil	100	43	88	67.14	10.15

Table 3.3 Statistics of mixed training data from degraded PAN image

The range of min to max of training data decides about the standard deviation. In case of mixed training data fuzzy standard deviation is higher in comparison to pure training data so the sub-pixel classified output is having more mixed pixel in comparative to pure training data.

In pure training data I have taken exclusive classes without any overlapping the is the reason in soft classified image also pixels are inclined towards purity their membership is tending to 0 or 1 but in mixed training data is highly overlapped that's why classified membership of pixels are varying from entire range between 0 to 1.

3.3. Classifier used

3.3.1 Bayesian classifier

Unlike hard classifiers, soft classifiers defer making a definitive judgment about the class membership of any pixel in favor of producing a group of statements about the degree of membership of that pixel in each of the possible classes. Like traditional supervised classification procedures, each uses training site information for the purpose of classifying each image pixel. However, unlike traditional hard classifiers, the output is not a single classified land cover map, but rather, a set of images (one per class) that expresses (in the case of Bayesian classifier) for each pixel the probability that it belongs to each class.

BAYCLASS is closely related to the MAXLIKE hard classifier available with IDRISI image processing software used in this study, in that it computes the posterior probability of belonging to each considered class according to Bays' Theorem.

The variance/covariance matrix derived from training site data is that which allows one to assess the multivariate conditional probability. This quantity is then modified by the prior probability of the hypothesis being true and then normalized by the sum of such considerations over all classes. This latter step is important in that it makes the assumption that the classes considered are the only classes that are possible as interpretations for the pixel under consideration. Thus even weak support for a specific interpretation may appear to be strong if it is the strongest of the possible choices given.

When no knowledge exists about the prior probabilities with which each class can occur, then equal prior probabilities should be used (the default). While having reasonable knowledge of the expected proportional area of each class over the image as a whole, we can choose another option (specify a prior probability value for each signature). Thus if you expect that 42% of the area is under a given cover type, the a priori probability of that class is 0.42. The third option is to enter prior probabilities as a separate real number image (with values between 0-1) for each class. This allows you to incorporate spatial predictive models into your determination of prior probabilities. For example, one may decide that the prior probability of an area known in the past to be forest changing to residential is highly likely near to roads and highly unlikely far away from roads. This can be expressed quite easily since each pixel is given a separate prior probability value using this approach. As always, the sum of probabilities for each pixel must be 1.0. The final option allows you to specify either a uniform value or an image of probabilities. In all cases except equal probabilities, prior probabilities are specified in the second dialog screen.

BAYCLASS is a confident classifier. It assumes that the only possible interpretation of a pixel is one of those classes for which training site data have been provided. It therefore admits to no ignorance. As a result, lack of evidence for an alternative hypothesis constitutes support for the hypotheses that remain. In this context, a pixel for which reflectance data only very weakly supports a particular class is treated as unequivocally belonging to that class (p = 1.0) if no support exists for any other interpretation.

The prime motivation for the use of BAYCLASS is sub-pixel classification -- i.e., to determine the extent to which mixed pixels exist in the image, and their relative proportions. In the context of mixture analysis, the probabilities of BAYCLASS are interpreted directly as statements of proportional representation. Thus if a pixel has posterior probabilities of belonging to deciduous and conifer of 0.68 and 0.32 respectively, this would be interpreted as evidence that the pixel contains 68% deciduous species and 32% conifers. Note, however, that this requires several important assumptions to be true. First, it requires that the assumption that the classes for which training site data have been provided are exhaustive (i.e., that there are no other possible interpretations for that pixel). Second, it assumes that the conditional probability distributions do not overlap in the case of pure pixels. In practice, these conditions may be difficult to meet.

A typical remote sensing data clustering is the maximum likelihood supervised procedure. It consists of the estimation of a suitable mixture of distributions, based on training samples only, and in the subsequent pixel-by pixel classification, performed by maximizing the likelihood ratio. In this way all the information on the parameters of the distributions, contained in the unsurveyed samples, is lost.

3.3.2 Fuzzy classifier

Fuzzy classification is a generalized approach which takes into account uncertainty of classification in terms of both ambiguity and vagueness leading to sub-pixel classification of land cover types. In this approach each pixel is provided with a membership grade for each of the considered classes. The membership grades are determined on the basis of predefined membership functions. The membership values lie between 1 and 0 (included) for each class in each pixel in which 1 implies full membership, 0 implies no membership and values in between represent sub-pixel membership of a class. Membership function is different for each class. Thus, by defining certain membership function of each class, the membership values of each pixel in all classes are determined for classification of data. Finally the accuracy of classification required to be checked.

However the accuracy of the representation provided by a fuzzy classification is difficult to evaluate. Conventional measures of classification cannot be used. The accuracy of a classification may be indicated by the way in which the strength of class membership is partitioned between the classes and how closely this represents the partitioning of class membership on the ground. Both thematic and spatial accuracy should be high for a classification so that information from the classified image can be used with a high confidence by the users of the classified map.

Fuzzy set membership is based on the standard distance of each pixel to the mean reflectance on each band for a signature. To accommodate quality of training signatures and width of classes, the user inputs the Z-score (standard deviation units) at which fuzzy set membership decreases to zero.

It requires entering the Z-score at which fuzzy membership decreases to zero. The Z-score for 0 fuzzy set memberships can be decided by two parameters: quality of your signature, and width of each class. If the signature is pure and the class width is small, a small Z-score should be selected. If the signature is mixed and the class width is large, a large Z-score should be selected.

The fuzzy set membership is calculated based on standardized Euclidean distance from the mean of the signature, using a sigmoid membership function (see FUZZY). The underlying logic is that the mean of a signature represents the ideal point for the class, where fuzzy set membership is 1. When distance increases, fuzzy set membership decreases, until it reaches the user-defined Z-score distance where fuzzy set membership decreases to 0.

The un-normalized procedure assumes that the fuzzy set membership for each class is derived independently, and incomplete information or overlapping signatures may exist. The sum of values for a pixel for all class images may be other than 1 in this case. The normalized procedure assumes that full information is achieved and signatures do not overlap, thus the fuzzy set membership for a pixel for all class images sums to 1.

The software used for both the above classification methods is IDRISI KILIMANZARO.

Both above classifiers are used for pure as well as mixed training data, soIn the report, now onwards these are named as

Classifier 1 Bayesian classification with pure training data

Classifier 2 fuzzy classification with pure training data

Classifier 3 Bayesian classification with mixed training data

Classifier 4 fuzzy classification with mixed training data

Collection of reference data were done with 3 by 3 window in pan degraded image .for each class 100 dominating pixels were collected , statistics of same is as shown in table below.

Class	Number of pixels	Mi n	Max	Mean	Standard deviation
Water	100	38	71	44.61	4.8
Forest	100	42	89	54.12	7.36
Agriculture	100	44	98	84.99	9.21
Sandy area	100	50	91	79.16	8.89
Bare soil	100	39	89	65.95	8.12

Table 3.4 Statistics of Reference sample from degraded PAN image

Fragmented images are shown on next page.

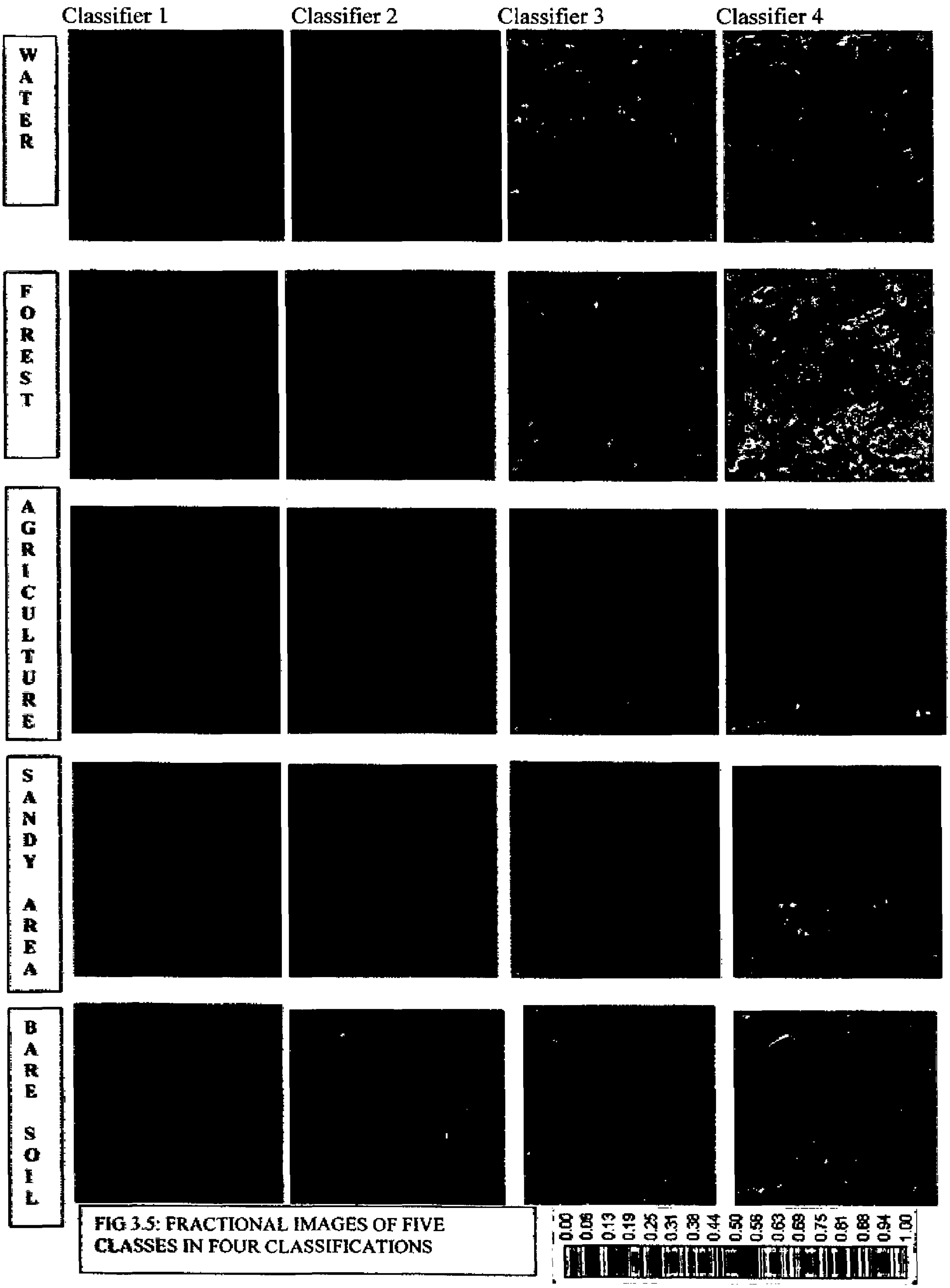


FIG 3.5: FRACTIONAL IMAGES OF FIVE CLASSES IN FOUR CLASSIFICATIONS

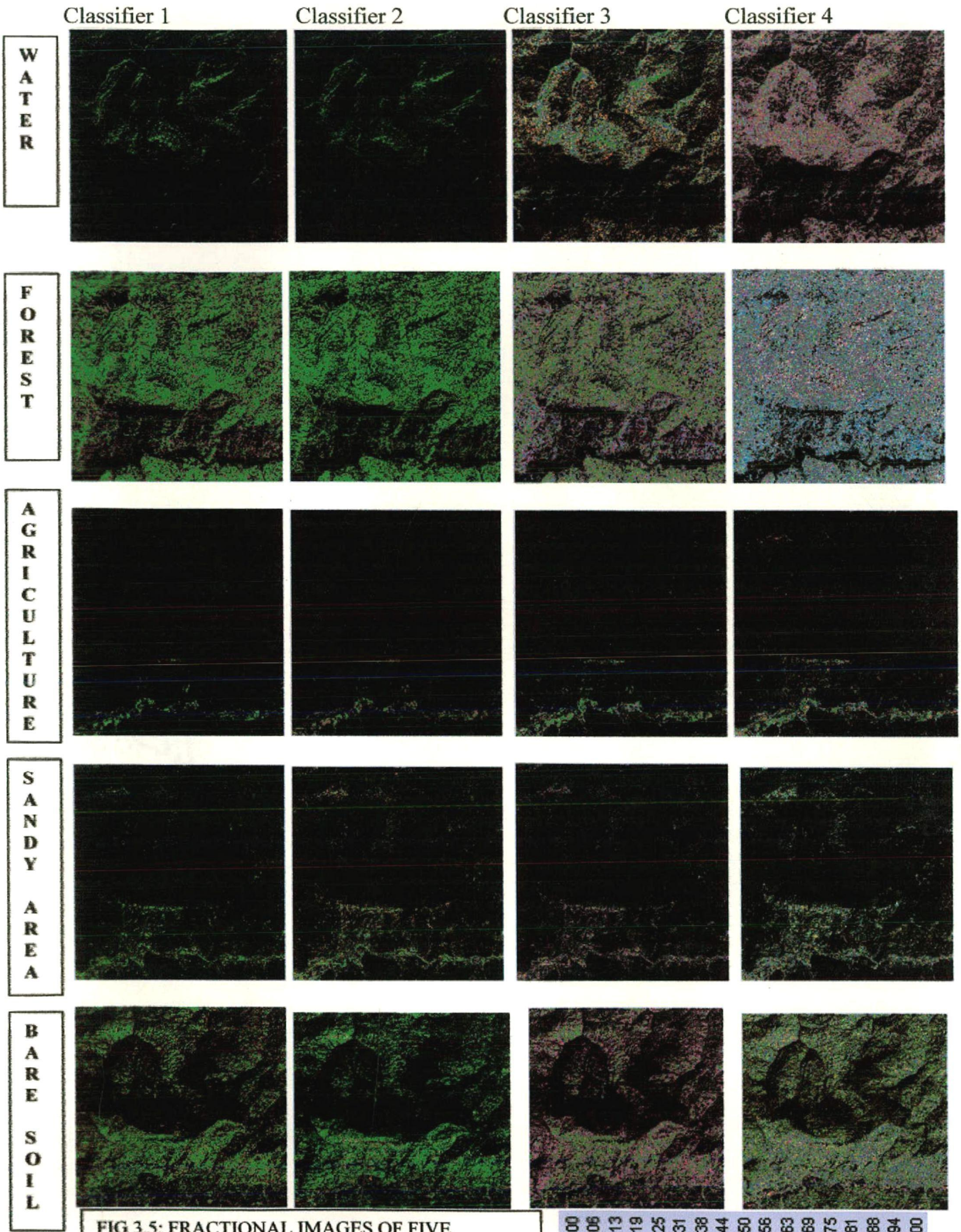


FIG 3.5: FRACTIONAL IMAGES OF FIVE CLASSES IN FOUR CLASSIFICATIONS

CHAPTER 4

RESULTS AND ANALYSIS OF ACCURACY ASSESSMENT METHODS

On the basis of data set in the previous chapter a few measures have been applied which are listed below. Formulation and theoretical background of these have been covered in chapter on literature review.

1. fuzzy correlation coefficient
2. entropy measures
3. fuzzy error matrix

Results obtained are presented in tabular and graphical format.

4.1 Fuzzy correlation coefficient

Four different accuracy assessment of fuzzy classification can be carried out making use fuzzy correlation coefficient (F-cr). These are to find accuracy assessment of (i) a single class in a single pixel,(ii) a single class in all the reference samples taken together, (iii) a pixel (considering all the classes in the same pixel) and (iv) an image (for all the classes in all pixels). A comparison with the statistical correlation coefficient can be carried out for the accuracy assessment of type (ii) only as it is not possible to find out statistical correlation for other types. The estimated values for the different types of accuracy assessment for the test data (consisting of reference and their corresponding classified data) were as explained below.

4.1.1. Class based Accuracy assessment

The correlation of the proportion of different land cover classes in a pixel with those present in reference data were estimated.

4.1.1.1 Accuracy assessment of one class in a single pixel

Under this, fuzzy correlation between the proportion of a particular type of land cover present in a classified data and that present in the corresponding pixel in reference data were being estimated.

Table 4.1 a sample of fuzzy correlation coefficient between Reference and Classified data

ID	Reference Membership value	Classified Membership value	F(cr)	ID	Reference Membership value	Classified Membership value	F(cr)
1	0.11	0.00	0.96	17	0.55	0.17	-0.33
2		0.40	0.46	18		0.43	-0.97
3		0.66	-0.72	19		0.72	0.46
4		0.90	-0.99	20		0.82	0.33
5	0.22	0.00	0.85	21	0.66	0.01	-0.61
6		0.12	0.95	22		0.27	-0.95
7		0.68	-0.90	23		0.72	0.95
8		0.85	-0.97	24		0.98	0.61
9	0.33	0.01	0.61	25	0.77	0.0	-0.84
10		0.27	0.95	26		0.43	-0.46
11		0.73	-0.96	27		0.79	0.99
12		0.97	-0.61	28		0.96	0.88
13	0.44	0.01	0.22	29	0.88	0.00	-0.96
14		0.49	0.48	30		0.46	-0.18
15		0.51	-0.48	31		0.67	0.74
16		0.73	-0.46	32		0.96	0.98

A sample of calculated fuzzy correlation coefficients for some classified and reference data are shown in Table 4.1

It has been found that if both the values of the reference and classified data lie on one side of 0.5i.e., if the degree of belongingness of both the reference and classified data in any particular class is either low (less than 0.5) or high (greater than 0.5), their correlation has been found to be positive. For example, if the membership values are 0.88 and 0.67, fuzzy correlation was found to be 0.75.

And in other cases where values lie one in each side of 0.5 i.e., if the degree of belongingness of one of the reference or classified data is either low (< or high and the other having high or low respectively, their correlation coefficient has been found to be negative. For example, is the membership values are 0.88 and 0.46, fuzzy correlation was found to be -0.18.

Thus, the sign of correlation coefficient provides information bout the type of degree of belongingness of the component cover class present in the reference and classified data i.e., whether the type of component cover class present in the reference and that represented by classified data are same or different.

It has also been found that if the sub-pixel composition of a particular class in the reference and classified data are close to each other their fuzzy correlation Coefficients are high. Thus, it has been found that lesser the membership values, higher is the correlation Coefficient, approaching towards 1. Further, the variation has been found to be non-linear approximating second order polynomial. For example, is the membership values are 0.11 and 0.00, fuzzy correlation was found to be 0.97 and for 0.11 and 0.40, the correlation was found to be 0.46. This observation is valid if both the values of the reference and classified data are present on the same side of 0.5. This is attributed to fact that 0.5 represents a situation having maximum vagueness thus designating the class around 0.5 is most uncertain. The magnitude of the negative correlation is high when data are away from 0.5 i.e., one of them near 0 and the other is near 1.0. This is due to the fact nearer to 1 or 0 signifies their certainty of belongingness or not belongingness to that particular class is high. Thus, higher negative value of correlation coefficient signifies higher possibility of their belongingness to different classes.

Further, the magnitude of the negative correlation is low when both the data are near but opposite to 0.5. This is due to the fact nearer to value of 0.5 higher is the vagueness that in these cases both reference and classified data are having higher vagueness i.e., the certainty of belongingness to a particular class is very low and as the values are in the opposite side of 0.5,possibility of their belongingness lie in different classes. A graphical plot for the fuzzy correlation coefficient for water class present in the 100 testing pixels for both the classifiers are as shown in figure 4.1.

4.1.1.2 Accuracy assessment of one class in an image

In this method of assessment, the fuzzy correlation of one land cover present in all the testing pixels has been taken into consideration. Fuzzy correlation between the proportion of a particular land cover class present in all classified testing pixels and that present in the corresponding pixels in reference data consisting of 500 pixels were being estimated using Equation 2. The fuzzy correlation coefficients (F-cr) for different land cover classes for two different classifiers are shown in Table 5. It can be found that for all the classes trend with respect to a classifier is same i.e., fuzzy correlation coefficient for different land cover classes from one classifier is higher from that that of other. Further, the statistical correlation coefficients

(Table 6) were estimated and found to have the same trend as that of fuzzy correlation coefficients.

Table 4.2 Fuzzy Correlation coefficient for class based testing pixels

Classes →	Water	Forest	Agriculture	Sand	Bare Soil
BAYMIX	0.84	0.73	0.80	0.80	0.78
FUZMIX	0.94	0.88	0.89	0.91	0.84

Table 4.3 Statistical Correlation coefficient for class based testing pixels

Classes →	Water	Forest	Agriculture	Sand	Bare Soil
BAYMIX	0.86	0.47	0.72	0.53	0.47
FUZMIX	0.90	0.74	0.77	0.78	0.57

4.1.2 Pixel based Accuracy assessment

Further, fuzzy correlation was estimated to assess accuracy of classification of a pixel considering all the classes together. A sample of estimated values using Equation 3 is as shown in Table 7. In this, the correlation coefficient of a pixel considering the membership values of all land cover classes (W-Water, F-forest, A-Agriculture, S-Sandy area, B-Bare soil) present in a single pixel were estimated for same set of reference data but different classified data as provided by classifier 1 and classifier 2. It has been found that for all pixels the trend with respect to a classifier is same i.e., fuzzy correlation coefficient for all pixels estimated for one classifier is higher than that from other.

Table 4.4 Fuzzy Correlation coefficient for pixel based testing pixels

Classes	Water	Forest	Agriculture	Sand	Bare Soil
BAYMIX	0.79	0.76	0.81	0.81	0.86
FUZMIX	0.92	0.88	0.91	0.91	0.91

4.1.3 Image based Accuracy assessment

Finally, fuzzy correlation can be estimated to assess the accuracy of classification of an image considering all the classes present in all the pixels in the image together by using Equation 4. The same was being estimated for all the 500 testing data. The fuzzy correlation coefficients are found to be 0.79and 0.89 for Classifier 1 and Classifier 2 respectively.

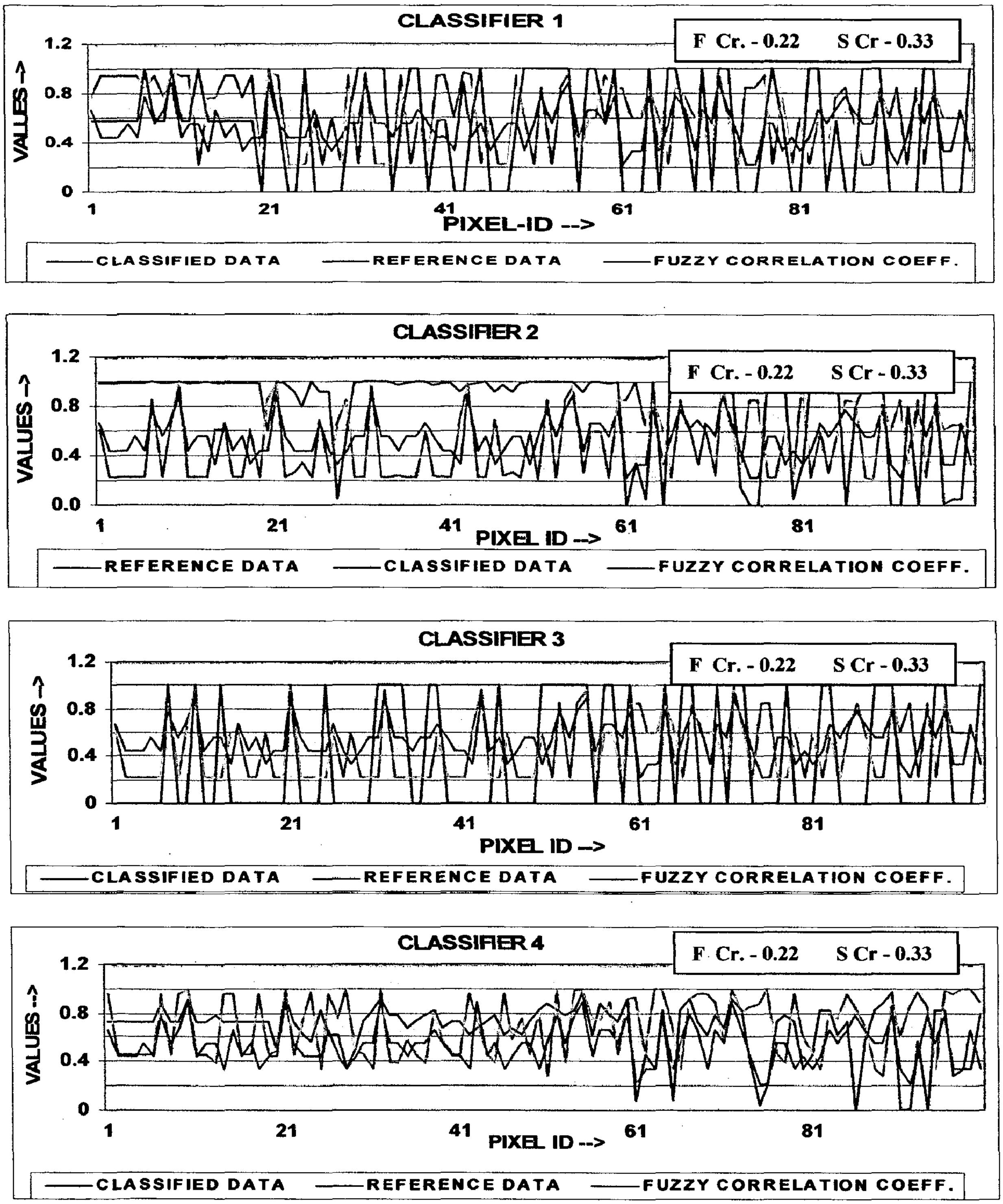
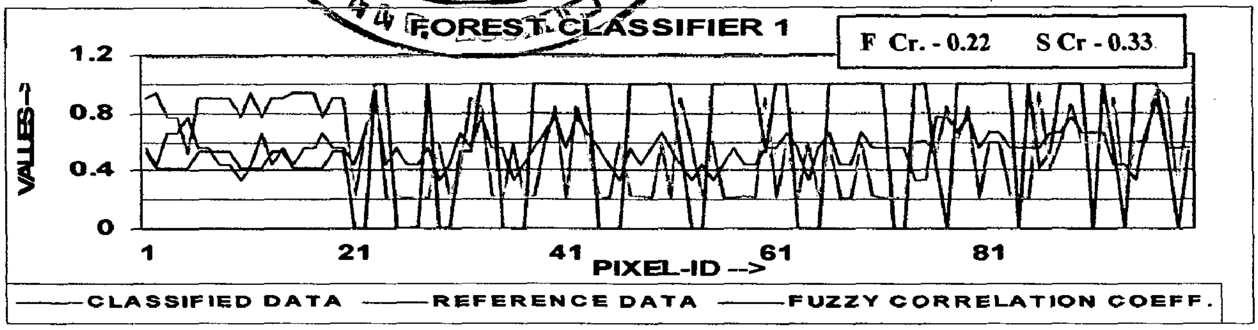
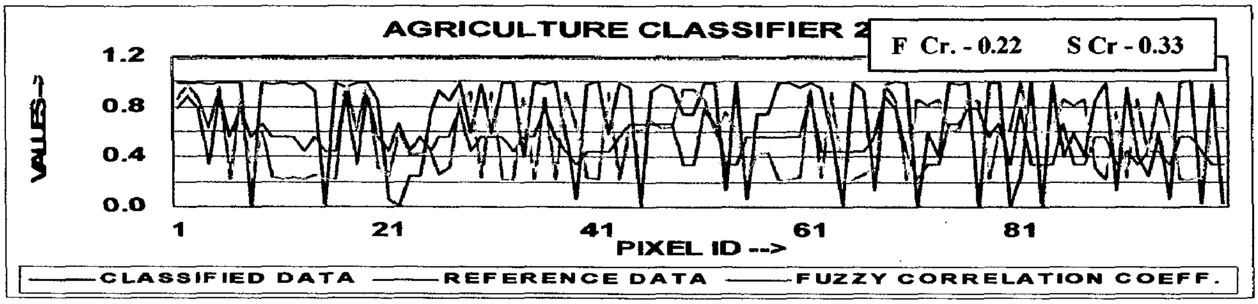
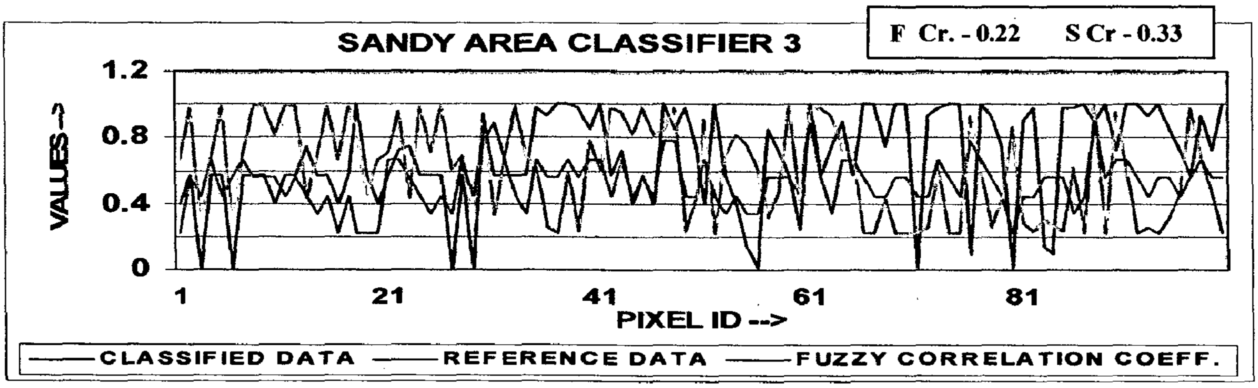
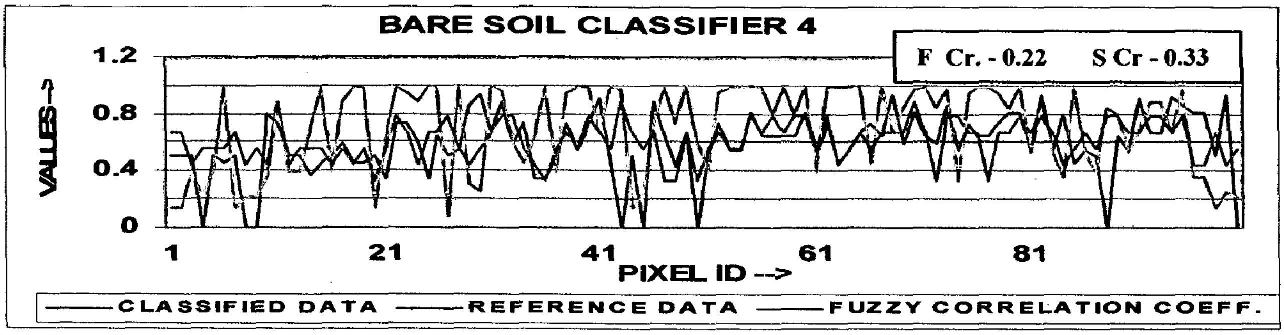


Fig 4.1 graphs showing variation of fuzzy correlation coefficient


FOREST CLASSIFIER 1
F Cr. - 0.22 S Cr - 0.33
1.2
0.8
0.4
0
VALUES-->
1
21
41
61
81
PIXEL-ID -->
CLASSIFIED DATA
REFERENCE DATA
FUZZY CORRELATION COEFF.


AGRICULTURE CLASSIFIER 2
F Cr. - 0.22 S Cr - 0.33
1.2
0.8
0.4
0.0
VALUES-->
1
21
41
61
81
PIXEL ID -->
CLASSIFIED DATA
REFERENCE DATA
FUZZY CORRELATION COEFF.


SANDY AREA CLASSIFIER 3
F Cr. - 0.22 S Cr - 0.33
1.2
0.8
0.4
0
VALUES-->
1
21
41
61
81
PIXEL ID -->
CLASSIFIED DATA
REFERENCE DATA
FUZZY CORRELATION COEFF.


BARE SOIL CLASSIFIER 4
F Cr. - 0.22 S Cr - 0.33
1.2
0.8
0.4
0
VALUES-->
1
21
41
61
81
PIXEL ID -->
CLASSIFIED DATA
REFERENCE DATA
FUZZY CORRELATION COEFF.

Classifier	Fuzzy cor. coefficient
BAYMIX	0.79
FUZMIX	0.89

Table 4.5 Fuzzy Correlation coefficient for overall image

6. Conclusion

To evaluate a fuzzy classification, accuracy of different sub-pixel classes (pixel based as well as image based), accuracy of individual pixel and accuracy of whole image are required to be assessed. A set of measures based on fuzzy correlation coefficient are being applied towards evaluation of outputs from two different fuzzy classifiers, in this study. It has been found that the measures evaluate four types accuracy of fuzzy classification qualitatively and also provide other quantitative information. Thus, it can be concluded that the proposed measures can be used to assess different types of accuracy associated with fuzzy classification.

4.2 Entropy measures

Kaufmann entropy – according to the formulation given and theory entropy measures are also found to be a good measure for the accuracy assessment. it can also be measured in the same four ways .Here are the table showing various combination of ref and classified data by which variation of Kaufmann entropy can be interpreted.

Higher values of entropy show good classification comparatively .it is related to the uncertainty of the classification. As the fractional values become higher entropy become higher. Low entropy means pixel are towards 0 or 1 value. Tables on the next page reflect the same result. Fuzzy classifier is having higher entropy value so it can be considered as better classifier as well as entropy measure is a good indicator of accuracy.

For pixel based ,Kaufmann become similar to foody entropy.

TABLE 4.6 IMAGE BASED RESULTS

CLASSIFIER	REFERENCE DATA					CLASSIFIED DATA				
	W	F	A	S	B	W	F	A	S	B
FUZZY	26.26	27.25	27.68	27.29	26.17	23.75	24.70	25.10	26.20	24.93
BAYSIAN						17.54	19.25	19.20	21.05	20.49

Table 4. 7 Samples of Kaufmann entropy for pixels

						Classified data												
	Reference data					BAYESIAN MIX					FUZZY MIX							
No.	W	F	A	S	B	W	F	A	S	B	W	F	A	S	B	KE-REF	KE-POOR	KE-GOOD
1	0.667	0.333	0.000	0.000	0.000	0.987	0.013	0.000	0.000	0.000	0.725	0.275	0.000	0.000	0.000	1.57	1.03	1.52
2	0.333	0.667	0.000	0.000	0.000	0.987	0.013	0.000	0.000	0.000	0.725	0.275	0.000	0.000	0.000	1.57	1.03	1.52
3	0.444	0.556	0.000	0.000	0.000	0.967	0.033	0.000	0.000	0.000	0.674	0.328	0.000	0.000	0.000	1.64	1.01	1.57
4	0.444	0.222	0.000	0.000	0.333	0.150	0.850	0.000	0.000	0.000	0.400	0.600	0.000	0.000	0.000	1.90	1.30	1.62
5	0.000	0.556	0.000	0.000	0.444	0.000	0.995	0.000	0.000	0.005	0.008	0.798	0.000	0.000	0.195	1.64	1.00	1.43
6	0.000	0.333	0.000	0.222	0.444	0.000	0.005	0.000	0.000	0.995	0.000	0.192	0.000	0.000	0.808	1.90	1.01	1.40
7	0.111	0.444	0.000	0.111	0.333	0.000	0.978	0.000	0.000	0.022	0.000	0.725	0.000	0.000	0.276	1.96	1.05	1.52
8	0.000	0.000	0.333	0.667	0.000	0.000	0.000	0.978	0.022	0.000	0.000	0.000	0.705	0.295	0.000	1.57	1.05	1.54
9	0.000	0.000	0.667	0.333	0.000	0.000	0.000	0.000	0.015	0.985	0.000	0.000	0.000	0.266	0.734	1.57	1.04	1.51
10	0.000	0.000	0.333	0.222	0.444	0.000	0.000	0.994	0.006	0.000	0.000	0.000	0.794	0.206	0.000	1.90	1.01	1.43

4.3 DISTANCE BASED MEASURES

S, D and L are some familiar measures usually used for accuracy assessment of fuzzy classification (Foody, G.M. 1996, 2000). These measures provide a distinct trend about the accuracy associated with sub-pixel classification.

Classifier	Bayesian pure	Fuzzy pure	Bayesian mix	Fuzzy mix
S distance	0.074	0.072	0.075	0.0335
D distance	0.532	0.527	0.544	0.330
L distance	0.811	0.803	0.824	0.512

Analysis – the variation of distance value gives the accuracy difference in different classifications .lower is the distance, better is the classification.
Here it can be observed that mix training data are giving better classification than pure.

4.4 FUZZY ERROR MATRIX

Fuzzy error matrix is calculated on the basis of min value between reference and classified data of a particular pixel. Formulas used are described in literature review.
This matrix is calculated for fuzzy classification using mixed training data.

Reference / classified	Water	forest	agriculture	Sandy area	Bare soil	Total
Water	65.699	39.651	1.140	1.173	3.645	111.3072
forest	61.306	79.504	4.600	8.225	41.647	195.282
agriculture	0.895	1.629	62.899	47.024	10.230	122.677
Sandy area	2.058	7.274	60.163	77.563	41.570	188.628
Bare soil	7.708	41.309	25.473	49.321	86.526	210.337
total	137.666	169.3667	154.2745	183.3056	183.6183	828.2312

As the sum of the diagonal values is 372.190 and sum of the membership values of all pixel in all classes in reference data is 499.889 so

Overall accuracy = 372.19/499.889= 0.744

Class	Water	forest	agriculture	Sandy area	Bare soil
Sum – classified data	95.68	96.58	99.24	95.00	113.50
Sum- reference data	72.00	120.00	74.55	108.22	125.11

User's accuracy

Diagonal values are divided by the sum of the membership values of the entire pixel in classified map for a particular class.

Class	Water	forest	agriculture	Sandy area	Bare soil
User's accuracy	0.68	0.82	0.63	0.81	0.76

Producer's accuracy

Diagonal values are divided by the sum of the membership values of the entire pixel in reference map for a particular class.

Class	Water	forest	agriculture	Sandy area	Bare soil
Producer's accuracy	0.91	0.66	0.84	0.71	0.69

REFERENCES

1. **Binaghi, E, Brivio, A.P, Ghezzi, P, Rampini, A,,** 1999, "*A fuzzy set-based accuracy assessment of soft classification*", Pattern Recognition Letters, 20, pp.935-948.

2. **Cochran, W.G.**, 1972, *Sampling Techniques,* 2nd edition, New Delhi, Wiley Eastern Private Limited.

3. **Congalton, R.G.,** 1991, "*A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data*", Remote Sensing of Environment, Vol. 37, pp. 35-46.

4. **Foody, G.M.,** 1995, "*Cross-entropy for the evaluation of the accuracy of a fuzzy land cover classification with fuzzy ground data*", ISPRS Journal of Photogrammetry and Remote Sensing, Vol:50, No:5, pp.2-12.

5. **Foody, G.M.,** 1996, "*Approaches for the production and evaluation of fuzzy land cover classifications from remotely sensed data*", International Journal of Remote Sensing, Vol.17, No.7, 1317-1340.

6. **Foody, G.M. and Arora, M.K.,** 1996, "*Incorporating mixed pixels in the training, allocation and testing stages of supervised classifications*", Pattern Recognition Letters, Vol:17, pp.1389-1398.

7. **Foody, G.M.,** 2000, "*Estimation of sub-pixel land cover composition in the presence of untrained classes*", Computers and Geosciences, Vol:26, pp.469-478.

8. **Foody, G.M. and Cox, D.P.,** 1994, "*Sub-pixel Land Cover Composition Estimation Using a Linear Mixture Model and Fuzzy Membership Functions*", International Journal of Remote Sensing, Vol:15, pp-619-631.

9. **Foody, G.M.,** 2002, "*Status of land cover classification accuracy assessment*", Remote Sensing of Environment, Vol.80, pp. 185-201.

10. **Gopal, S. and Woodcock, C.,** 1994, "*Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy Sets*", Photogrammetric Engineering and Remote Sensing, Vol.60, No.2, pp. 181-188.

11. **Gopal, S. and Woodcock, C.E.,** 2000, "*Fuzzy set theory and thematic maps: accuracy assessment and area estimation*", International Journal of Geographical Information Science, Vol:14, No:2, pp.153-172.

12. **Hagen, A.** 2003, "*Fuzzy set approach to assessing similarity of categorical maps*", International Journal of Geographical Information Science, vol:17, no:3, 235-249.

13. **Hagen A. et al.,** 2005, "Further developments of a fuzzy set map comparison approach" , International Journal of Geographical Information Science, vol:19, no:7, 769-785.

14. **Higashi, M. and Klir., G.J.,** 1983, "*On the notion of distance representing information closeness: possibility and probability distributions*", International Journal of General Systems, Vol:9, pp.103-115.

15. **Jager. G. and Benz, U.,** 2000, "*Measures of Classification Accuracy Based on Fuzzy Similarity", IEEE Transactions on Geoscience and Remote Sensing*", Vol.38, No.3, pp. 1462-1467.

16. **Janssen, L.L.F. and Vanderwel F.J.M.,** 1994, "*Accuracy Assessment of Satellite Derived Land Cover Data: A Review*", Photogrammetric Engineering and Remote Sensing, Vol:60, No:4, pp:419-426.

17. **Jensen., John. R**., 1986, "*Introductory Digital Image Processing*", Prentice Hall International Limited,

18. **Lesparre, Gorte et al.,** 2003 "*using mixed pixels for the training of a maximum likelihood classification*" . The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 34, Part XXX

19 **Latifovic, R., and Olthof, I.,** 2004, "*Accuracy assessment using sub-pixel fractional error matrices of global land cover products derived from satellite data*", Remote Sensing of Environment, Vol:90, pp.153-165.

20. **Lillesand, T.M. and Keifer, R.W.,** 1994, "*Remote Sensing and Image Interpretation*", 3rd edition, John Wiley and Sons.

21. **Leeuw,J. de et al.,**2006. 'Comparing *accuracy assessments to infer superiority of mage classification methods",* International Journal of Remote Sensing, Vol. 27, No. 1, 223–232

22. **Pal, S.K., Ghosh, A. and Uma Shankar, B.,** 2000, "*Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation*", International Journal of Remote Sensing, Vol.21, No.11, pp.2269-2300.

23. **Pontius, R.G. and Cheuk, M. L.,** 2006. "A generalized cross tabulation matrix to compare soft-classified maps at multiple resolution", International Journal of Geographical Information Science, vol:20, no:1, 1-30

24. **Ricotta, C.,** 2004, "*Evaluating the classification accuracy of fuzzy thematic maps with a simple parametric measure*", International Journal of Remote Sensing, Vol:25, No:11, pp. 2169-2176.

26. **Shalan, M.A., Arora, M.K. and Elgy, J.**, "*CASCAM: Crisp and Soft Classification Accuracy Assessment Software*", www.geocomputation.org/2003/Papers/Shalan_Paper.pdf

27. **Stehman., S.V. and Czaplewski, R.L.**, 1998, "*Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles*", Remote Sensing of Environment, Vol:64, pp.331-344.

28. **Townsend, P.A.**, 2000, "*A Quantitative Fuzzy Approach to Assess Mapped Vegetation Classifications for Ecological Applications*", Remote Sensing of Environment, Vol.72, pp.253-267.

29. **Zhang. J. and Foody, G.M.**, 1998, "*A fuzzy classification of sub-urban land cover from remotely sensed imagery*", International Journal of Remote Sensing, Vol:19, No.14, pp.2721-2738.

30. **Zwick, R., Carlstein., E and Budescu**, D.V., 1987, "*Measures of Similarity Among Fuzzy Concepts: A Comparative Analysis*", International Journal of Approximate Reasoning, Vol:1, pp.221-242.

RELATED PUBLICATION COMMUNICATED BY THE AUTHOR

1. Accuracy Assessment for Sub-pixel Thematic Classification of Satellite Data: A Brief Review, International Journal of Remote Sensing, Ghosh J.K , Ravindra Goyal and K. Mukherjee (2006). (To be revised).

2. Measures Based on Fuzzy Correlation Coefficient for Accuracy assessment of Fuzzy Classification, International Journal of Remote Sensing, Dr. Jayanta Kumar Ghosh, Ravindra Goyal and Kriti Mukherjee (communicated)

APPENDIX A

Table 1 Distance Measures of Accuracy Assessment

Measure		Image/ Pixel based	Formulation	Explanation	References
Euclidean Distance (fuzzy data based)	S	Pixel	$S_j = \sum_{i=1}^{c} ({}^1\mu_{ij} - {}^2\mu_{ij})^2 / c$ ....Eq(2.2)	c = no. of classes ; μi= membership value of i th pixel; Finds the distance between two pixels.	1. Kent and Mardia 1988 2. Foody, 1996 3. Foody and Arora 1996
	$\overline{S}$	Image	$\frac{\sum_{j=1}^{n} S_j}{n}$ ....Eq(2.3)	n = no. of pixels; Finds the distance between two classifications or two sub-pixel level information.	
	D	Pixel	$D_j = \sqrt{\sum_{i=1}^{c} ({}^1\mu_{ij} - {}^2\mu_{ij})^2}$ ....Eq(2.4)	c = no. of classes; μi=membership value of i th pixel ; Finds the distance between two pixels.	
	$\overline{D}$	Image	$\frac{\sum_{j=1}^{n} D_j}{n}$ ....Eq(2.5)	n = no. of pixels; Finds the distance between two classifications or two sub-pixel level information.	
L Distance (fuzzy data based)	L	Pixel	$L_j = \sum_{i=1}^{c} \left\| {}^1\mu_{ij} - {}^2\mu_{ij} \right\|$ ....Eq(2.6)	c = no. of classes; μi=membership value of i th pixel; Finds the distance between two pixels.	1. Foody and Arora 1996
	$\overline{L}$	Image	$\frac{\sum_{j=1}^{n} L_j}{n}$ ....Eq(2.7)	n = no. of pixels; Finds the distance between two classifications or two sub-pixel level information.	
Information Closeness (probabilistic data based)	d	Pixel	$d(1p,2p) = -\sum_x {}^1p(x)\log_2 {}^2p(x) + \sum_x {}^1p(x)\log_2 {}^1p(x)$ ....Eq(2.8)	Measures the directed divergence between two probability distributions or normalized fuzzy distributions, which are compatible to each other and provides an index of accuracy. x is the class variable ; P(x) represents the membership probability of pixel in class x	1. Foody, 1995 2. Foody, 1996 3. Zhang and Kirby 1997 1. Higashi and Klir 1983
	I	Pixel	$D({}^1p,{}^2p) = d({}^1p, \frac{{}^1p+{}^2p}{2}) + d({}^2p, \frac{{}^1p+{}^2p}{2})$ where d is from last eq. ....Eq(2.9)	Measures the distance between the information content of any two probability distributions or normalized fuzzy. P represents same pixel in two different images.	

Table 2 Fuzzy set based measures of accuracy assessment

Measure	Formulation	Explanation	Reference
Fuzzy neighborhod and category vector	$S(V_A,V_B)=\left\lfloor\left(\mu_{Ai},\mu_{Bi}\right\vert_{\min}\right)_{i=1,\dots c}\Big\rfloor_{\max}$ ....Eq(2.13)	Similarity of two fuzzy datasets can be compared using these methods. For the first two cases, if result is more than 0.5, it indicates a good similarity. K_{Fuzzy} provides an overall similarity measure for the whole image.	1. Townsend, 2000 2. Jager and Benz, 2000 3. A. Hagen, 2003
	$S_{twoway}(V_A,V_B)=\left\vert S(V_A,V_{crispB}),S(V_{crispA},V_B)\right\vert_{\min}$ ....Eq(2.14)	Symbols: μ represents membership V represents vector S is the similarity	
Fuzzy kappa	Fuzzy Kappa = $K_{Fuzzy}=\frac{S-E}{1-E}$ $E=E(S)=\frac{\sum_{i=1}^{n}E(S_i)}{n}$ $S=\frac{\sum_{i=1}^{n}S_i}{n}$ eq (2.15)	Similarity parameters above give cell by cell similarity only. To have overall image similarity, kappa is generated. fuzzy kappa include uncertainty of location and uncertainty of class definition both .	Hagen, A., 2005
		Symbol: S is the overall similarity E is expected overall similarity l is location parameter	
Fuzzy functions	Frequency of Matches and Mismatches — Max function $MAX(x,c)=\{1 \quad if \quad \mu_c(x)\geq\mu_{c'}(x) \quad for \quad all \quad c'\in C$ .....Eq(2.16) $=\{0 \quad otherwise$	MAX(x,c) is 1 if the category for a pixel as obtained from the classified output has the highest membership grade to belong to the pixel according to expert evaluation.	1. Gopal and Woodcock, 1994 2. Woodcock and Gopal, 2000
	Right function $RIGHT(x,c)=\begin{cases}1, & if \quad \mu_c(x)\geq\tau \\ 0 & otherwise\end{cases}$ ......Eq(2.17) from the classified data for a pixel and the maximum membership grade assigned to the pixel among all other classes.	a sight belongs to a category 'c', as obtained from classification, if $\mu_c(x)$ from expert evaluation is greater than a predetermined membership grade τ . The more the number of matches compared to mismatches i.e., more the result is 1, the less is the error	

Table 2 continued

Measure		Formulation	Explanation		Reference
	Source Of Error	membership function for every pixel, $\lambda(x)=\{c \mid c\in C \text{ and } \mu_c(x)\ge\tau\}$ ....Eq(2.18)	Each of the pixels is allocated to one or more number of classes. Accordingly there will be single membership, double membership etc. The cells for which double or more membership occur are said to be the source of error.		1. Gopal and Woodcock, 1994 2. Woodcock and Gopal, 2000
			Symbols	τ is the threshold value defined same as in match and mismatch. $\mu_c(x)$ is the membership of pixel x to the class c. C is the set of all discrete defined classes.	
	Magnitude of Error	$\Delta x=\mu_{C(x)}(x)-\max\limits_{c\ne C(X)}\mu_c(x)$ ....Eq(2.19)	Magnitude of error implies the seriousness of error in classification. This is the difference in the membership grade between the class obtained.		
			Symbols	Where, C(x) is the class obtained from the classified	
	Nature of Error	Confusion function: $\zeta(x)=\{c/c\in C \,\&\, \mu_c(x)>\mu_{c(x)}(x)\}$ ....Eq(2.20) Ambiguity function: $\eta(x)=\{c/c\in C \,\&\, \mu_c(x)=\mu_{c(x)}(x)\}$ ....Eq(2.21)	Nature of error means the confusion and ambiguity in the class assignment. Confusion function gives the set of the pixels whose membership grade for the class assigned by the expert is more than that for any other class as obtained from classification. Ambiguity function gives the set of the pixels whose membership grade for the class assigned by the expert is equal to that for any other class as obtained from classification. Number of pixels in each class which are in confusion and ambiguous can be found using these functions.		
			symbols	Same as above .	

Table 3 Similarity Measures of Accuracy Assessment

Measure	Image/ Pixel Based	Formulation	Explanation	Refere
Correlation Coefficient (probabilistic data)	Image	$CC_i = \frac{Cov(1_{\mu i}, 2_{\mu i})}{\sigma_{1_{\mu_i}} \times \sigma_{2_{\mu_i}}}$ ....Eq(2.24)	Finds the correlation of the probability distributions of the same class representation in two images. 0<CC≤1, for positive correlation;-1≤CC<0, for negative correlation;CC=0, for no correlation.	1. Foody 2000
RMSE (probabilistic data)	Image	$RMSE_j = \sqrt{\frac{\sum_{i=1}^{n}(1_{\mu_{ij}} - 2_{\mu_{ij}})^2}{n-1}}$ ....Eq(2.25)	n = no. of pixels; Finds the distance between the representations of the same class in two sub-pixel level image datasets. i is the pixel variable(value 1 to n) j is the class for which RMSE is being calculated	1. Foody 2000
		$\frac{\sum_{j=1}^{c} RMSE_j}{c}$ ....Eq(2.26)	c = no. of classes; Finds the overall RMS error of a sub-pixel classification with respect to soft reference data. j is the class variable (value 1 to c)	
Expected Sets Shared (probabilistic data)	Pixel	$ESS_i(\mu_A, \mu_B; m) = \sum_{k=1}^{n}(1-(1-\mu_{Ak})^m)(1-(1-\mu_{Bk})^m)$ ....Eq(2.27)	μ_A = membership distribution for a pixel in the reference data; μ_B = membership distribution of the corresponding pixel in the classified data; n = number of classes; m = size of the random sample drawn from the population ESS finds the number of classes common to the same pixel of both reference and classified image	1. C. Ric 2004
Similarity Index (fuzzy data)	Image	$SI_{ir} = 1 - CC_{ir}$; $CC_{ir} = \frac{\sum_k \lvert IV_{ki} - IV_{kr} \rvert}{\sum_k (IV_{ki} + IV_{kr})}$ ......Eq(2.28)	Finds the similarity between two fuzzy membership distributions. CC_{ir} is the dissimilarity index between classified data 'r' and reference data 'i'. IV_{ki} and IV_{kr} are the importance values of the k'th species at site 'i' and 'r' respectively.	1. Towi 2000

APPENDIX B MIX TRAINING DATA SAMPLES FOR PAN DEGRADED IMAGE ALONGWITH COORDINATES

ID	ROW	COL	MAP X	MAP Y	DN	WATER	FOREST	AGRI	SANDY	BARE
1	7	3	-4546	23575	58	0	0.56	0	0	0.44
2	26	6	-4432	23593	58	0	0.67	0	0	0.33
3	73	6	-4150	23593	56	0	0.67	0	0	0.33
4	30	9	-4408	23611	61	0.22		0	0	0
5	11	10	-4522	23617	59	0.11	0.67	0	0	0.22
6	27	14	-4426	23641	52	0.22	0.67	0	0	0.11
7	72	15	-4156	23647	56	0	0.78	0	0	0.22
8	56	17	-4252	23659	53	0.11	0.67	0	0	0.22
9	25	18	-4438	23665	48	0.11	0.89	0	0	0
10	6	19	-4552	23671	58	0	0.67	0	0.11	0.22
11	41	21	-4342	23683	68	0	0.22	0	0.33	0.44
12	51	24	-4282	23701	54	0.11	0.67	0	0	0.22
13	60	24	-4228	23701	60	0	0.67	0	0	0.33
14	10	25	-4528	23707	61	0	0.56	0	0.11	0.33
15	31	25	-4402	23707	60	0	0.44	0	0	0.56
16	14	28	-4504	23725	48	0	0.44	0	0.11	0.44
17	74	28	-4144	23725	47	0.33	0.67	0	0	0
18	33	29	-4390	23731	60	0	0.56	0	0	0.44
19	78	31	-4120	23743	53	0	0.78	0	0	0.22
20	82	31	-4096	23743	53	0	0.89	0	0	0.11
21	24	33	-4444	23755	46	0.33	0.67	0	0	0
22	269	49	-2974	23851	47	0.22	0.78	0	0	0
23	291	49	-2842	23851	46	0.22	0.78	0	0	0
24	266	54	-2992	23881	45	0.22	0.78	0	0	0
25	276	55	-2932	23887	48	0.33	0.67	0	0	0
26	281	59	-2902	23911	45	0.11	0.89	0	0	0
27	289	60	-2854	23917	46	0.11	0.89	0	0	0
28	264	62	-3004	23929	45	0.33	0.67	0	0	0
29	283	66	-2890	23953	45	0.22	0.78	0	0	0
30	263	68	-3010	23965	45	0.22	0.78	0	0	0

32	269	75	-2974	24007	44	0.33	0.67	0	0	0
33	870	81	632	24043	49	0.11	0.89	0	0	0
34	874	81	656	24043	47	0.56	0.44	0	0	0
35	879	82	686	24049	48	0.44	0.56	0	0	0
36	900	83	812	24055	45	0.44	0.56	0	0	0
37	1437	95	4034	24127	45	0.33	0.67	0	0	0
38	1449	96	4106	24133	43	0.33	0.67	0	0	0
39	1483	96	4310	24133	43	0.44	0.56	0	0	0
40	1468	98	4220	24145	43	0.22	0.78	0	0	0
41	1484	111	4316	24223	42	0.44	0.56	0	0	0
42	1461	119	4178	24271	44	0.33	0.67	0	0	0
43	1467	119	4214	24271	54	0.33	0.67	0	0	0
44	1479	121	4286	24283	44	0.11	0.89	0	0	0
45	1444	124	4076	24301	46	0.22	0.78	0	0	0
46	1439	124	4046	24301	44	0.33	0.67	0	0	0
47	855	180	542	24637	50	0.22	0.78	0	0	0
48	813	181	290	24643	48	0.11	0.89	0	0	0
49	835	181	422	24643	46	0.11	0.89	0	0	0
50	827	188	374	24685	45	0.22	0.78	0	0	0
51	818	189	320	24691	43	0.22	0.78	0	0	0
52	808	189	260	24691	43	0.11	0.89	0	0	0
53	854	189	536	24691	56	0	0.78	0	0	0.22
54	816	191	308	24703	49	0.11	0.89	0	0	0
55	807	195	254	24727	51	0.22	0.78	0	0	0
56	844	195	476	24727	42	0.56	0.44	0	0	0
57	847	196	494	24733	39	0.56	0.44	0	0	0
58	834	200	416	24757	47	0.44	0.56	0	0	0
59	834	204	416	24781	44	0.22	0.78	0	0	0
60	855	205	542	24787	45	0.11	0.89	0	0	0
61	848	210	500	24817	46	0.11	0.89	0	0	0
62	328	217	-2620	24859	43	0.56	0.44	0	0	0
63	344	218	-2524	24865	46	0.33	0.67	0	0	0
64	320	220	-2668	24877	46	0.22	0.78	0	0	0
65	315	221	-2698	24883	45	0.33	0.67	0	0	0

66	342	226	-2536	24913	43	0.33	0.67	0	0	0
67	312	228	-2716	24925	51	0.22	0.67	0	0	0.11
68	342	231	-2536	24943	53	0	0.89	0	0	0.11
69	319	235	-2674	24967	43	0.44	0.56	0	0	0
70	1633	235	5210	24967	52	0.11	0.56	0	0	0.33
71	1616	238	5108	24985	45	0.44	0.56	0	0	0
72	1618	242	5120	25009	49	0.11	0.89	0	0	0
73	327	242	-2626	25009	44	0.44	0.56	0	0	0
74	1632	243	5204	25015	59	0	0.56	0	0	0.44
75	336	243	-2572	25015	57	0	0.78	0	0	0.22
76	1623	245	5150	25027	57	0	0.67	0	0	0.33
77	1610	248	5072	25045	46	0.33	0.67	0	0	0
78	1634	249	5216	25051	46	0.56	0.44	0	0	0
79	332	251	-2596	25063	46	0.33	0.67	0	0	0
80	1621	251	5138	25063	46	0.11	0.56	0	0	0.33
81	312	252	-2716	25069	44	0.44	0.56	0	0	0
82	1607	253	5054	25075	44	0.33	0.67	0	0	0
83	1618	254	5120	25081	58	0	0.67	0	0	0.33
84	321	255	-2662	25087	48	0.22	0.67	0	0	0.11
85	328	257	-2620	25099	57	0.22	0.44	0	0.11	0.22
86	1626	257	5168	25099	58	0	0.67	0	0	0.33
87	320	259	-2668	25111	50	0.22	0.67	0	0	0.11
88	334	260	-2584	25117	58	0	0.67	0	0	0.33
89	342	263	-2536	25135	57	0	0.56	0	0	0.44
90	318	264	-2680	25141	64	0	0.67	0	0	0.33
91	346	267	-2512	25159	59	0	0.56	0	0	0.44
92	385	268	-2278	25165	58	0.11	0.44	0	0.11	0.33
93	338	270	-2560	25177	48	0.22	0.78	0	0	0
94	319	270	-2674	25177	65	0.44	0.56	0	0	0
95	311	271	-2722	25183	56	0	0.78	0	0	0.22
96	326	274	-2632	25201	55	0.11	0.67	0	0	0.22
97	330	276	-2608	25213	51	0.11	0.78	0	0	0.11
98	322	278	-2656	25225	45	0.33	0.67	0	0	0
99	510	351	-1528	25663	57	0	0.67	0	0	0.33

100	488	352	-1660	25669	58	0	0.56	0	0	0.44
101	474	354	-1744	25681	61	0	0.56	0	0	0.44
102	502	355	-1576	25687	60	0	0.56	0	0	0.44
103	508	362	-1540	25729	50	0.11	0.78	0	0	0.11
104	464	372	-1804	25789	60	0	0.56	0	0.22	0.22
105	515	377	-1498	25819	60	0	0.44	0	0.22	0.33
106	477	379	-1726	25831	65	0	0.67	0	0	0.33
107	372	387	-2356	25879	60	0	0.44	0	0	0.56
108	488	388	-1660	25885	61	0	0.44	0	0.22	0.33
109	507	389	-1546	25891	64	0	0.44	0	0.11	0.44
110	358	390	-2440	25897	65	0	0.44	0	0.33	0.22
111	373	391	-2350	25903	49	0.33	0.56	0	0	0.11
112	366	391	-2392	25903	43	0.33	0.67	0	0	0
113	355	394	-2458	25921	46	0.33	0.67	0	0	0
114	373	395	-2350	25927	46	0.33	0.67	0	0	0
115	357	399	-2446	25951	58	0	0.56	0	0	0.44
116	373	401	-2350	25963	61	0	0.56	0	0.11	0.33
117	366	401	-2392	25963	60	0	0.56	0	0.11	0.33
118	364	407	-2404	25999	51	0	0.78	0	0	0.22
119	358	407	-2440	25999	57	0.11	0.78	0	0	0.11
120	375	408	-2338	26005	62	0	0.67	0	0.11	0.22
121	639	431	-754	26143	56	0	0.78	0	0	0.22
122	653	432	-670	26149	62	0	0.33	0	0.11	0.56
123	656	434	-652	26161	58	0	0.78	0	0	0.22
124	639	434	-754	26161	62	0	0.56	0	0.33	0.11
125	654	437	-664	26179	57	0	0.78	0	0	0.22
126	633	437	-790	26179	50	0.11	0.78	0	0	0.11
127	658	449	-640	26251	81	0	0.44	0	0.22	0.33
128	664	452	-604	26269	65	0	0.44	0	0.22	0.33
129	643	453	-730	26275	57	0	0.67	0	0.11	0.22
130	641	469	-742	26371	63	0	0.44	0	0.11	0.44
131	1265	474	3002	26401	89	0	0.67	0	0.11	0.22
132	1276	474	3068	26401	47	0.22	0.78	0	0	0
133	1257	474	2954	26401	50	0	0.67	0	0	0.33

134	1264	479	2996	26431	59	0	0.67	0	0	0.33
135	1253	479	2930	26431	63	0	0.56	0	0.33	0.11
136	1273	481	3050	26443	60	0	0.56	0	0	0.44
137	1263	486	2990	26473	65	0	0.33	0	0.22	0.44
138	1268	493	3020	26515	59	0	0.56	0	0	0.44
139	1278	498	3080	26545	51	0.11	0.56	0	0	0.33
140	1266	506	3008	26593	58	0	0.78	0	0	0.22
141	1597	563	4994	26935	53	0	0.89	0	0	0.11
142	1623	565	5150	26947	56	0	0.78	0	0	0.22
143	1607	569	5054	26971	58	0	0.67	0	0.11	0.22
144	1610	569	5072	26971	58	0	0.67	0	0	0.33
145	1632	570	5204	26977	47	0.44	0.56	0	0	0
146	1602	577	5024	27019	46	0.22	0.78	0	0	0
147	1626	581	5168	27043	48	0.33	0.67	0	0	0
148	1608	581	5060	27043	48	0.22	0.78	0	0	0
149	1597	585	4994	27067	57	0	0.78	0	0	0.22
150	1631	587	5198	27079	58	0	0.78	0	0	0.22
151	1682	787	5504	28279	43	0.44	0.56	0	0	0
152	1695	798	5582	28345	43	0.44	0.56	0	0	0
153	1694	802	5576	28369	44	0.44	0.56	0	0	0
154	1684	808	5516	28405	42	0.44	0.56	0	0	0
155	1668	813	5420	28435	43	0.44	0.56	0	0	0
156	251	839	-3082	28591	45	0.44	0.56	0	0	0
157	255	840	-3058	28597	45	0.33	0.67	0	0	0
158	248	843	-3100	28615	43	0.33	0.67	0	0	0
159	245	843	-3118	28615	41	0.78	0.22	0	0	0
160	1369	846	3626	28633	45	0.44	0.56	0	0	0
161	1380	846	3692	28633	45	0.56	0.44	0	0	0
162	1386	847	3728	28639	44	0.44	0.56	0	0	0
163	247	847	-3106	28639	43	0.44	0.56	0	0	0
164	255	848	-3058	28645	46	0.33	0.56	0	0	0.11
165	243	848	-3130	28645	43	0.44	0.56	0	0	0
166	256	851	-3052	28663	45	0.56	0.44	0	0	0
167	1386	852	3728	28669	42	0.56	0.44	0	0	0