CASCAM: Crisp and Soft Classification Accuracy Measurement Software

Mohammad A. Shalan[1], Manoj K. Arora[2] and John Elgy[1]

[1]School of Engineering and Applied Sciences, Aston University, Birmingham, UK
[2]Department of Civil Engineering, IIT Roorkee, India

Abstract

No image classification is complete until an assessment of accuracy has been performed. Accuracy serves as the basis for the analysis of errors, which may creep in during the classification process due to complex interactions between the spatial structure of landscape, sensor resolutions and classification algorithms. Indeed much research has been conducted on devising a number of accuracy measures for both crisp and soft classifications, they have not been implemented in commercial image processing software. The aim of this paper is to provide the details and the utility of an indigenous software for accuracy assessment of both crisp and soft classification of remote sensing data.

1. Introduction

Classification is the fundamental image processing task to extract information from remote sensing data. Both crisp and soft classifications may be performed. In a crisp classification, each image pixel is assumed pure and is classified to one class. Often, particularly in coarse spatial resolution images, the pixels may be mixed containing two or more classes. Soft classifications that assign multiple class memberships to a pixel may be appropriate for images dominated by mixed pixels. Both supervised and unsupervised approaches may be adopted. Generally, supervised classification is adopted that involves three stages; training, allocation and testing.

Whether the goal is to produce a crisp or a soft classification, the assessment of classification accuracy in the testing stage is a critical step as it allows a degree of confidence to be attached to the classification. In essence, the assessment of accuracy involves comparison of classified image with reference or ground data (also known as ground truth). The reference data may be gathered from field surveys, existing maps, aerial photographs and any datasets at a higher resolution than the image being classified. The accuracy of crisp classification may be determined using conventional error matrix based measures such as the overall accuracy, user’s and producer’s accuracy, kappa coefficient of agreement etc. (Congalton, 1991). A number of other measures, for example, Tau coefficient (Ma and Redmond, 1995), and classification success index (Koukoulas and Blackburn, 2001) have also been proposed, though used sparingly. To evaluate the accuracy of soft classification, the classified outputs are often hardened so that the error matrix based measures may be used. Degrading soft classification to crisp results in the loss of information contained in the soft outputs thereby hampering its proper evaluation. Moreover, the reference data are also not always error-free and may contain uncertainties, and therefore may be treated as soft or fuzzy (Foody, 1995). Hence, alternative accuracy measures that may appropriately include the fuzziness in the classification outputs and/or reference data have been proposed. These include root mean square error, correlation coefficient (Foody and Cox, 1994), entropy (Maselli et al. 1994) and distance based measures (Foody, 1996), fuzzy set (Gopal and Woodcock, 1994) and fuzzy error matrix based measures (Binaghi et al. 1999). The growth of so many accuracy measures both for crisp and soft classification clearly indicates the current research potential of classification accuracy assessment procedures as no single measure may universally be adopted. Depending upon the nature of classification outputs, uncertainties in reference data and the quality of information desired by the end user, it may be necessary to adopt not one but a combination of accuracy measures.

Despite much of the research being conducted on classification accuracy assessment and its importance, the current image processing software are limited in providing sufficient accuracy information to the user. For example, the well known and the most widely used software namely ERDAS Imagine, ENVI and IDRISI, contain accuracy assessment modules that can report only a few crisp accuracy measures namely overall, producer's and user's accuracy and kappa coefficient. No other competitive accuracy measures have been included, which may appeal to a user in cases where the assumptions regarding the current accuracy measures are not met by the dataset. Also, there is no provision for assessing the accuracy of soft classification. Only IDRISI has a measure called classification uncertainty that specifies the quality of classification on per pixel basis, and thus may not be treated as a measure to indicate the accuracy of whole classification. Thus, to evaluate the accuracy of soft classification, the users either have to depend on other statistical and mathematical software, where import/export of the data from one package to another may be a tedious task, or they may have to code their own software. Further, to critically examine the usefulness of a particular accuracy measure vis a vis other measures, dedicated software for classification accuracy assessment needs to be developed.

The aim of this paper is to introduce an accuracy assessment software specifically written to evaluate the quality of both crisp and soft classifications of remote sensing data. The implementation of the software will be demonstrated through a soft land cover classification, the results of which have recently been accepted for publication in IJRS Letters (Shalan et al., 2003).

2. Details of Software

The software has been written in Matlab and has been named as Crisp And Soft Classification Accuracy Measurement (CASCAM). MATLAB® language is a high-performance language for technical computing and particularly well-suited to designing predictive mathematical models and developing application-specific algorithms, for more details the reader may visit the MathWorks web site (www.mathworks.com). To facilitate different operations, a number of Matlab functions and graphical user interface resources have been used to develop this software. The minimum requirement to run this software is Windows '95' operating system.

3. Data File Management

The input and output (I/O) images are read in standard ASCII text files with extension either .asc or .txt. This format corresponds to the ERDAS Imagine ASCII file format that

consists of pixels in each row with sequential columns indicating location of pixels (X and Y coordinates in any measurement unit) and their intensity values in various bands numbered as b1, b2, etc. For classification output files in ASCII format, columns representing bands will indicate class memberships of each pixel (i.e., class identity in crisp classification and class proportions in soft classification). Sample data files for input images and classification outputs (crisp and soft) are given in Tables 1a, 1b and 1c.
The images can also be imported and exported in three standard graphics format namely JPG, TIFF and BMP. However, these images need to be converted into ASCII format before these can be used for further processing in this software. All the files can be read from and saved to the appropriate data directories.

Table 1a Input image file format

X	Y	Band1	Band2	Band3	…..
44541	2962276	119	85	89	…..
44566	2962276	116	86	92	…..
44591	2962276	113	82	95	…..
44616	2962276	109	76	96	…..
44641	2962276	109	73	96	…..

Table 1b Crisp classification output file format

X	Y	Class Identity
44541	2962276	2
44566	2962276	2
44591	2962276	1
44616	2962276	2
44641	2962276	5

Table 1c Fuzzy classification output file format

X	Y	Class1	Class2	Class3	Class4	Class5
44541	2962276	0.292	0.048	0.426	0.090	0.144
44566	2962276	0.315	0.056	0.363	0.103	0.163
44591	2962276	0.366	0.067	0.244	0.126	0.199
44616	2962276	0.389	0.076	0.139	0.153	0.244
44641	2962276	0.316	0.101	0.108	0.198	0.277

4. Salient Features of CASCAM

The software CASCAM consists of five basic modules namely display, training, classification, testing and accuracy measurement modules (Table 2). The pictorial layout of menu bar, various popup menus and button bars are shown in Figure 1. Each of these modules is now described in detail.

Table 2 Five basic modules of CASCAM software

Module	Popup Menu
Display	Display input images
	Display crisp classified images
	Display soft classified images
Training	Generate training data
	Plot training data
	Display histogram
Classification	Maximum likelihood classifier
	Fuzzy C-means classifier
Testing	Generate testing data from classification
	Generate crisp/soft reference image
Accuracy Measurement	Crisp accuracy measures
	Fuzzy accuracy measures

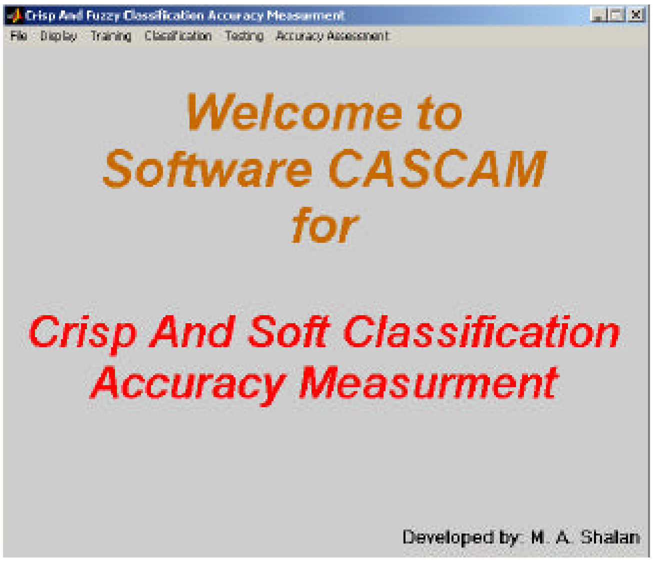

Figure 1, Main Menu of Software CASCAM

4.1 Display module

This module has been written to view input and output images stored as ASCII, JPG, TIFF and BMP files. A panchromatic input image will be displayed as a grayscale image representing various shades of gray ranging from black to white. The multi-spectral input image may be displayed as a single band grayscale image or as a False Color Composite (FCC) by representing any three bands in three primary colours – blue, green and red. To display a crisp classification output, the user has the option of assigning a particular colour to each class from the in-built colour pallet. Soft classification outputs in the form of fraction or proportion images are displayed as grayscale images with bright areas denoting higher class proportion and vice versa. The number of fraction images is equal to the number of classes to be mapped.

4.2 Training module

Training is the first stage of a supervised classification process. This module will allow the user to interactively define training areas for each class on the displayed image. The areas may be marked both as polygon boundaries and on per pixel basis. The spatial location of these training areas may also be represented graphically in a plot (Figure 2). The marked training areas are used to extract image information and are stored as an ASCII file on pixel by pixel basis for each class sequentially (Table 3).

Table 3 Training data file format

X	Y	B1	B2	B3	…..	Class ID
45016	2962276	102	77	83	…..	1
45841	2961976	103	70	67	…..	1
46241	2961976	102	76	82	…..	2
46291	2961976	101	75	80	…..	2
46341	2961976	101	75	81	…..	2
46316	2961951	102	75	81	…..	3
44816	2961926	103	71	67	…..	3

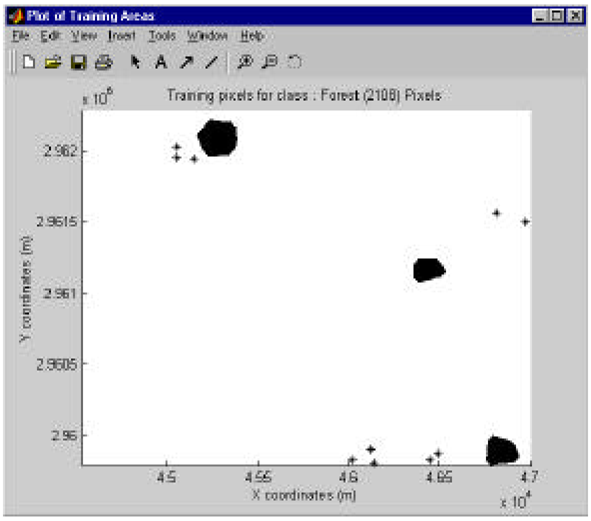

Figure 2, Sample Plot for Training Areas

Often and in particular, for statistical classifiers, it is necessary to examine the quality of training areas of a class by examining the histogram. True multi modal classes should be broken down into a number of pure spectral signature classes. A uni-modal histogram indicates good quality training data of a class (i.e., the pixels defining the training areas of that class are relatively pure). Therefore, the provision for the display of histograms of classes in each band has also been made in this module. Figure 3 shows a sample plot of the histogram for training areas of a class in one of the bands.

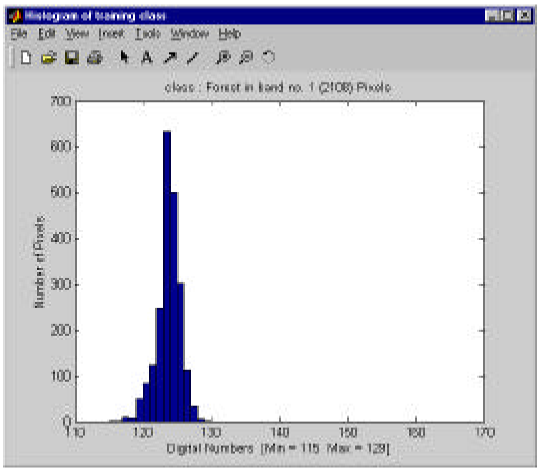

Figure 3, Histogram for a Class in Band 1

The user also has the option of directly inputting the training data file in ASCII format created in any other software. In this case, the user will be prompted to provide the number of classes to be mapped and the total number of training pixels in each class.

4.3 Classification module

The focus of CASCAM is on accuracy measurement. Therefore, for demonstration purposes, the formulations of two classifiers with markedly different characteristics, a probabilistic - the maximum likelihood classifier (MLC) (Mather, 1999) and a distribution free - the fuzzy-c means classifier (FCM) (Bezdek, 1984), have been incorporated in this module. The former is highly dependent on the data distributional assumptions. However, these assumptions are often not met and, therefore, an alternative non-parametric classifier such as the FCM may be advantageous to produce crisp and soft classifications.

MLC is the most widely used classifier in remote sensing community. In majority of studies, this classifier has been used as a crisp classifier. However, the output of an MLC in the form of a posteriori class probabilities may be related to the actual class proportions for each pixel on ground thereby providing soft classification. To run MLC in this software, the file containing apriori class probabilities of each pixel may also be supplied. If the file is unavailable, equal apriori class probabilities are considered by default.

The FCM is based on an iterative clustering algorithm that may be employed to partition pixels of image into class proportions. It is essentially an unsupervised clustering algorithm. However, here it has also been implemented in the supervised mode (Wang, 1990). The user has the option of selecting different parameters pertinent to this classifier. In the supervised mode, cluster means are computed from the training areas

created in the training module. The soft classification outputs from this classifier are represented in a class membership matrix, which can be hardened, if desired, to produce a crisp classification where a pixel is assigned a class having the highest membership value.
The outputs of crisp and soft classifications from these classifiers are stored in ASCII files as per the format specified in Tables 1b and 1c respectively, and may be used subsequently in testing and accuracy measurement modules. These files can directly be accessed in the display module to display crisp and soft classification images.

4.4 Testing module

Testing stage is the last stage of supervised classification, where accuracy is assessed. An appropriate sample of testing pixels with known class identity (crisp reference data) or known class proportions (soft reference data) is selected. Among a number of sampling schemes (Congalton, 1988), simple random sampling has been used here to generate the testing datasets. The testing data sample size is to be provided by the user.
If we assume that the uncertainty of classification is due to a mixture of land cover classes within a pixel then the creation of soft reference data is based on deriving class proportions from an existing map at a finer resolution than the image being classified. For instance, in the working example here, a land cover map produced from IRS PAN image at 5m spatial resolution, has been used as reference data to assess the accuracy of classification produced from IRS LISS image at 25m spatial resolution. Thus, a pixel of LISS image corresponds to an even number of 5m pixels (in this case 25 pixels) to facilitate in determining class proportions that sum to one for each pixel. These class proportions of pixels are called as soft reference data. The soft reference data are hardened to produce crisp reference data for accuracy assessment of crisp classification.
The testing pixels, either obtained from crisp or soft reference data, are stored in an ASCII file according to the format specified in Tables 1b and 1c respectively. This file is later used in the accuracy measurement module.

4.5 Accuracy measurement module

The accuracy of crisp and soft classifications is determined in this module. A number of crisp and soft accuracy measures have been incorporated.
For crisp classification accuracy assessment (Figure 4), first an error matrix is generated from the testing data set. The user has to supply two files: crisp classified image obtained in classification module and crisp reference data obtained in testing module. Alternatively, an error matrix generated from other sources may also be provided. The elements of the error matrix are used to derive a number of accuracy measures, which have been divided into three groups in this module:

i) Percent correct measures
ii) Kappa coefficients
iii) Tau coefficients

The formulations of all the accuracy measures considered under these groupings are given in Table 4. Further details on these formulations can be found in the respective references cited in this table.

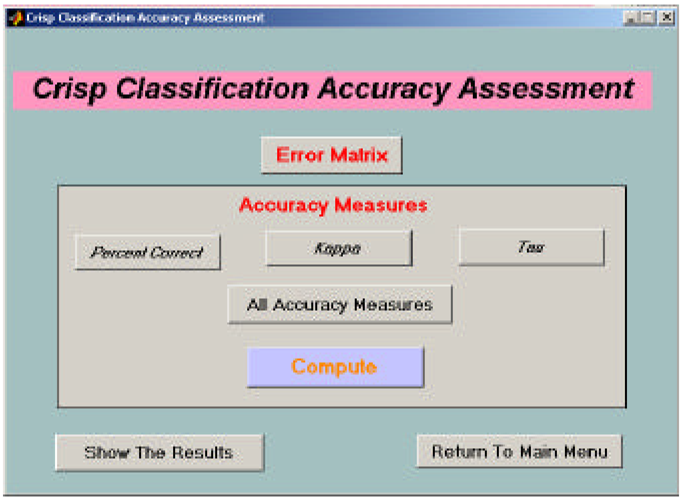

Figure 4, Menu for Crisp Classification Accuracy Assessment

In the first group, five accuracy measures – overall, user's and producer's accuracy, (Story and Congalton, 1986), average and combined accuracy (Fung and LeDrew, 1988) have been incorporated. While overall, average and combined accuracy signify the quality of whole classification, user's and producer's indicate the quality of individual class. Although, overall accuracy may be biased towards the class with a large number of testing samples, average accuracy computed from user's and producer's perspective may be biased towards the class having a small number of samples (Fung and LeDrew, 1988). Combined accuracy may be used to reduce the biases of overall and average accuracy. Producer's and user's accuracy are used to indicate accuracy of individual classes. Producer's accuracy is so aptly called, since the producer of the classification is interested in knowing how well the samples from the reference data can be mapped using remotely sensed image. In contrast, user's accuracy indicates the probability that a sample from the classification represents an actual class on reference data (Story and Congalton, 1986).

However, accuracy measures in the percent correct group do not take into account the agreement between the data sets (i.e., classified output and reference data) that arises due to chance alone. Thus, these measures tend to overestimate the classification accuracy (Ma and Redmond, 1995). The kappa coefficient of agreement has the ability to control the chance agreement that is the result of the misclassifications represented by the off-diagonal elements of the error matrix. Thus, the second group of accuracy measures in this module is formed to consist four measures from kappa family - kappa coefficient of agreement (Congalton et al. 1983), weighted kappa (Rosenfield and Fitzpatrick-Lins, 1986) and conditional kappa (user's and producer's perspective) (Rosenfield and Fitzpatrick-Lins, 1986). When some classes have more confusion than others, weighted kappa may be implemented since it does not treat all the misclassifications (disagreements) equally and tends to give more weight to the confusions that are more serious than others (Cohen, 1968; Hubert, 1978). In this software, a weight matrix has to be provided by the user. To determine the accuracy of individual classes, a conditional kappa may be computed.

Table 4: Crisp classification accuracy measures

Measure	Base Reference	Formulation	Definition of terms
Overall accuracy	Story and Congalton (1986)	$\frac{1}{N}\sum_{i=1}^{c} n_{ii}$	N is total number of testing pixels.
User's accuracy	Story and Congalton (1986)	n_{ii}/N_i	n_{ii} is the number of samples correctly classified.
Producer's accuracy	Story and Congalton (1986)	n_{ii}/M_i	N_i is the row total for class i.
Average accuracy (User's)	Fung and LeDrew (1988)	$\frac{1}{c}\sum_{i=1}^{c}\frac{n_{ii}}{N_i}$	M_i is the column total for class i.
Average accuracy (Producer's)	Fung and LeDrew (1988)	$\frac{1}{c}\sum_{i=1}^{c}\frac{n_{ii}}{M_i}$	$P_o = \frac{1}{N}\sum_{i=1}^{c} n_{ii}$ is the observed proportion of agreement
Combined accuracy (User's)	Fung and LeDrew (1988)	$\frac{1}{2}\left[OA + AA_u\right]$	$P_e = \frac{1}{N^2}\sum_{i=1}^{c} N_i M_i$ is the expected chance agreement
Combined accuracy (Producer's)	Fung and LeDrew (1988)	$\frac{1}{2}\left[OA + AA_P\right]$	v_{ij} is the weight P_{Oij} is the observed cell proportion
Kappa coefficient of agreement	Congalton *et al.* (1983)	$\frac{P_o - P_e}{1 - P_e}$	Pe_{ij} is the expected cell proportion
Weighted Kappa	Rosenfield and Fitzpatrick-Lins (1986)	$1 - \frac{\sum v_{ij} P_{oij}}{\sum v_{ij} P_{cij}}$	$Po_{(i+)}$ is the observed agreement according to user's approach
Conditional Kappa (User's)	Rosenfield and Fitzpatrick-Lins (1986)	$\frac{P_{o(i+)} - P_{e(i+)}}{1 - P_{e(i+)}}$	$Pe_{(i+)}$ is the agreement expected by chance for i^{th} row
Conditional Kappa (Producer's)	Rosenfield and Fitzpatrick-Lins (1986	$\frac{P_{o(+i)} - P_{e(+i)}}{1 - P_{e(+i)}}$	$Po_{(+i)}$ is the observed agreement according to producer's approach
Tau coefficient (equal probability)	Ma and Redmond (1995)	$\frac{P_o - \frac{1}{c}}{1 - \frac{1}{c}}$	$Pe_{(+i)}$ is the agreement expected by chance for i^{th} column
Tau coefficient (unequal probability)	Ma and Redmond (1995)	$\frac{P_o - P_r}{1 - P_r}$	$P_r = \frac{1}{N}\sum_{i=1}^{c} n_{i+} x_i$
Conditional Tau (User's)	Naesset (1996)	$\frac{P_{o(i+)} - P_i}{1 - P_i}$	
Conditional Tau (Producer's)	Naesset (1996)	$\frac{P_{o(+i)} - P_i}{1 - P_i}$	x_i is the unequal priori probability of class membership
			P_i is the a priori probability of class membership

Nevertheless, as argued by Foody (1992) and later supported by Ma and Redmond (1995), the kappa family may overrate the chance agreement that may result into an underestimation of accuracy. Therefore, alternatives to kappa coefficients such as Tau coefficient have been proposed. These form the third group of accuracy measures in this module. The critical difference between the two coefficients is that Tau is based on a priori probabilities of class membership whereas kappa uses the a posteriori probabilities. The a priori probabilities for Tau coefficients may be equal or unequal and need to be supplied by the user in this software. A conditional Tau coefficient is used to indicate the accuracy of an individual class (Naesset, 1996).
The error matrix based measures inherently assume that each pixel is associated with only one class in the crisp classification and only one class in the reference data. Use of these measures to assess the accuracy of soft classification may therefore under or over estimate their accuracy, as the soft classification outputs have to be degraded to adhere to this assumption. For evaluation of soft classification (Figure 5), at the first instance, entropy may be used as a measure to indicate the uncertainty in the classification (Table 5). Entropy shows how the strength of class membership (i.e., soft outputs) is partitioned between the classes for each pixel (Foody, 1995). The entropy for a pixel is maximised when the pixel has equal class memberships for all the classes. Conversely, its value is minimum, when the pixel is entirely allocated to one class. It, thus, shows the degree to which a classification output is soft (i.e., uncertain) or crisp. The accuracy of soft classification outputs is determined by comparing these with soft reference data, as generated in testing module for a sample of testing samples. A number of accuracy measures may be used. For simplicity, these measures have been divided into three groups in this module,

i) Measures of closeness
ii) Measures based on fuzzy error matrix
iii) Correlation coefficient.

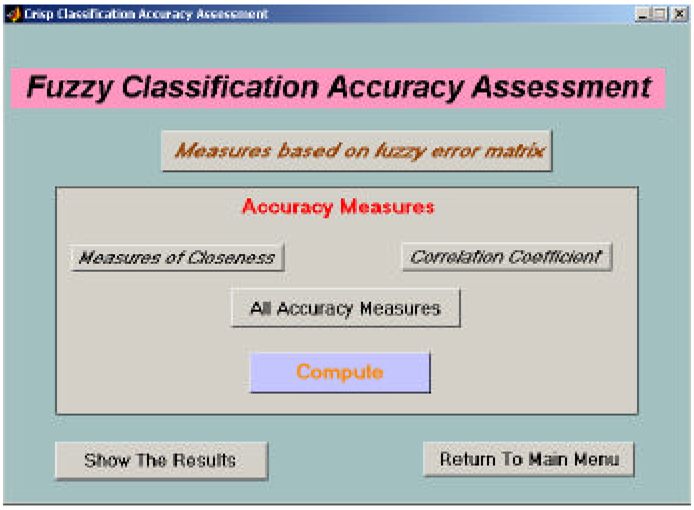

Figure 5, Menu for Fuzzy Classification Accuracy Assessment

The formulations of all the accuracy measures considered under these groupings are given in Table 5. Further details on these formulations can be found in the respective references cited in this table.

The first group includes cross entropy, L1 and Euclidean distances and generalized measure of information closeness. These measures estimate the separation of two data sets based on the relative extent or proportion of each class in the pixel (Foody and Arora, 1996). Lower the values of these measures, higher is the accuracy of classification. The distance measures and cross entropy may be applicable when there is compatibility between the probability distributions of the classified outputs and reference data. On the other hand, the generalized measure of information closeness may be used even if the probability distributions of the two datasets are not compatible (Foody, 1996).

Recently, Binaghi et al. (1999) have proposed the concept of fuzzy error matrix, which can be generated on the lines of conventional error matrix for crisp classification. The class proportions of a pixel in soft classification output are compared with the class proportions of that pixel in the soft reference data (created in the testing module). Fuzzy set 'min' operator is used to find minimum of the two values, which is recorded in the appropriate column, and summed over all the pixels to generate the fuzzy error matrix. The elements of this matrix are used to compute overall, user's and producer's accuracy, which have the same interpretation as described earlier. From the point of view of standardizing the accuracy assessment procedures for both crisp and soft classification, these measures appear more logical to be used in assessing the quality of remotely sensed derived classifications. The correlation coefficients may also be used to indicate the accuracy of individual classes and have been defined in the third group of accuracy measures. The higher the correlation coefficient better is the accuracy of an individual class.

The user has the option of computing the crisp and soft accuracy measures individually or all in one go. The error matrix, the fuzzy error matrix and the values of the selected accuracy measures can also be saved to a text file.

Table 5: Soft classification accuracy measures

Measure	Base Reference	Formulation	Definition of terms
Entropy (H)	Maselli *et al.* (1994)	$-\sum_{i=1}^{c}({}^{2}p_i)\log_2({}^{2}p_i)$	${}^{1}p_i$ is the proportion of i^{th} class in a pixel from the fuzzy reference data. ${}^{2}p_i$ is the proportion of i^{th} class in a pixel from the fuzzy classification.
Euclidean distance (E)	Kent and Mardia (1988), Foody (1996)	$\sum_{i=1}^{c}({}^{1}p_i - {}^{2}p_i)^2 \big/ c$	
L_1 (City Block) distance	Foody and Arora (1996)	$\sum_{i=1}^{c}\left\|{}^{1}p_i - {}^{2}p_i\right\| \big/ c$	
Cross-entropy or Direct divergence (D)	Foody (1995)	$D({}^{1}p,{}^{2}p) = -\sum_{i=1}^{c}({}^{1}p_i)\log_2({}^{2}p_i) + \sum_{i=1}^{c}({}^{1}p_i)\log_2({}^{1}p_i)$	${}^{1}p$ is the probability distribution of fuzzy reference data. ${}^{2}p$ is the probability distribution of fuzzy classification output.
Measure of information closeness (I)	Foody (1996)	$I({}^{1}p,{}^{2}p) = D\left({}^{1}p, \frac{{}^{1}p+{}^{2}p}{2}\right) + D\left({}^{2}p, \frac{{}^{1}p+{}^{2}p}{2}\right)$	
Correlation coefficients (R)	Foody and Cox (1994), Maselli *et al.* (1996)	$\frac{\mathrm{Cov}({}^{1}p_i,{}^{2}p_i)}{\mathrm{Std}({}^{1}p_i)\mathrm{Std}({}^{2}p_i)}$	$\mathrm{Cov}({}^{1}p_i,{}^{2}p_i)$ is the covariance between the two datasets. $\mathrm{Std}({}^{1}p_i)$ and $\mathrm{Std}({}^{2}p_i)$ are the standard deviations of the respective datasets

5. A Working Example

To demonstrate the software, a case study on accuracy assessment of soft classification from IRS 1C remote sensing data is briefly presented here. More details can be found in Shalan et al. (2003). IRS 1C LISS image (Figure 6) was used as primary image to produce soft classification. Five dominant land cover classes in the region namely agriculture, forest, grassland, urban and sandy areas were considered. An MLC derived crisp classification of the PAN image into five land cover classes was used as reference data (Figure 7). The LISS image was registered to PAN image derived land cover classification to an accuracy of 1/3rd of a pixel, using first order polynomial transformation and nearest neighbourhood resampling. The registered LISS and PAN images were resampled to 25 m and 5 m respectively such that a LISS pixel corresponds to an even number of PAN pixels (in this case 25 pixels) to facilitate in generating soft reference data in the form of class proportions.

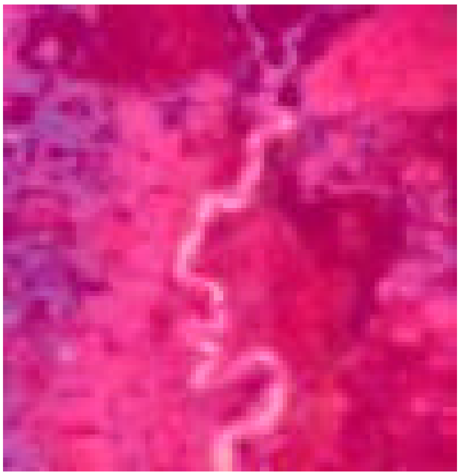

Figure 6 IRS 1C LISS III FCC (Red: band 4, Blue: band 2, Green: band 1)

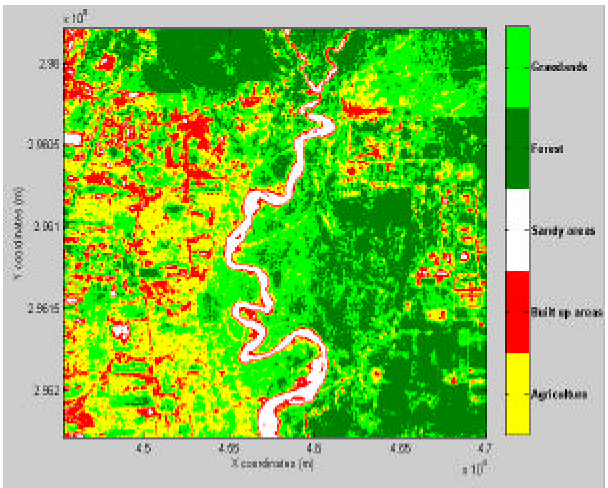

Figure 7 Classified PAN Image used as Reference Data

The two classifiers implemented in classification module of this software, MLC and FCM, were used to perform soft classification. For effectual comparison with MLC, the supervised version of FCM was applied here. In the formulation of FCM, a weighting factor m that describes the degree of fuzziness has to be provided. Here, m was set to 2.0, motivated by the study conducted by Foody (1996), where this value was found to produce the most accurate fuzzy classification. The training data file was created using training module of this software and consisted of randomly selected 997, 286, 279, 1596 and 805 pixels for agriculture, urban area, sandy area, forest and grassland respectively. In the testing module, a total of 650 testing pixels from the entire image were randomly selected for accuracy assessment. The accuracy of soft classification was evaluated using cross entropy, Euclidean distance, and correlation coefficients as defined in the accuracy measurement module of the software.

Entropy was used to examine the degree of uncertainty in the soft classification outputs (Table 6). For a five-class problem, the maximum value of entropy is 0.690. From table, it can be seen that the average entropy values (computed over all the pixels) for the soft classifications produced from both the classifiers, are very close to the maximum entropy value. This clearly illustrates the presence of class mixtures (or uncertainties) in the dataset. In Table 6, the lower values of cross entropy and Euclidean distance for MLC demonstrate that this classifier has produced more accurate classifications than FCM for the dataset considered. From correlation coefficients, it can be seen that the class sandy area has been classified as the most accurate class by both the classifiers, as this class was very well spectrally separable from the other classes.

Table 6 Accuracy of fuzzy classifications produced from MLC and FCM

Accuracy Measure (Average)	MLC	FCM
Entropy	0.526	0.565
Cross-entropy	0.262	0.287
Euclidean Distance	0.057	0.060

Table 7 Correlation coefficients of classes from soft classifications

Class	MLC	FCM
Agriculture	0.590	0.495
Urban area	0.507	0.626
Sandy area	0.854	0.860
Forest	0.708	0.583
Grassland	0.402	0.366

To inspect soft classifications visually, fraction images portraying the spatial distribution of five land cover classes were produced using the display module of the software (Figure 8). From this figure also, it can be observed that for all the classes, MLC has generally predicted closer relationship of class proportions with the reference data than the FCM.

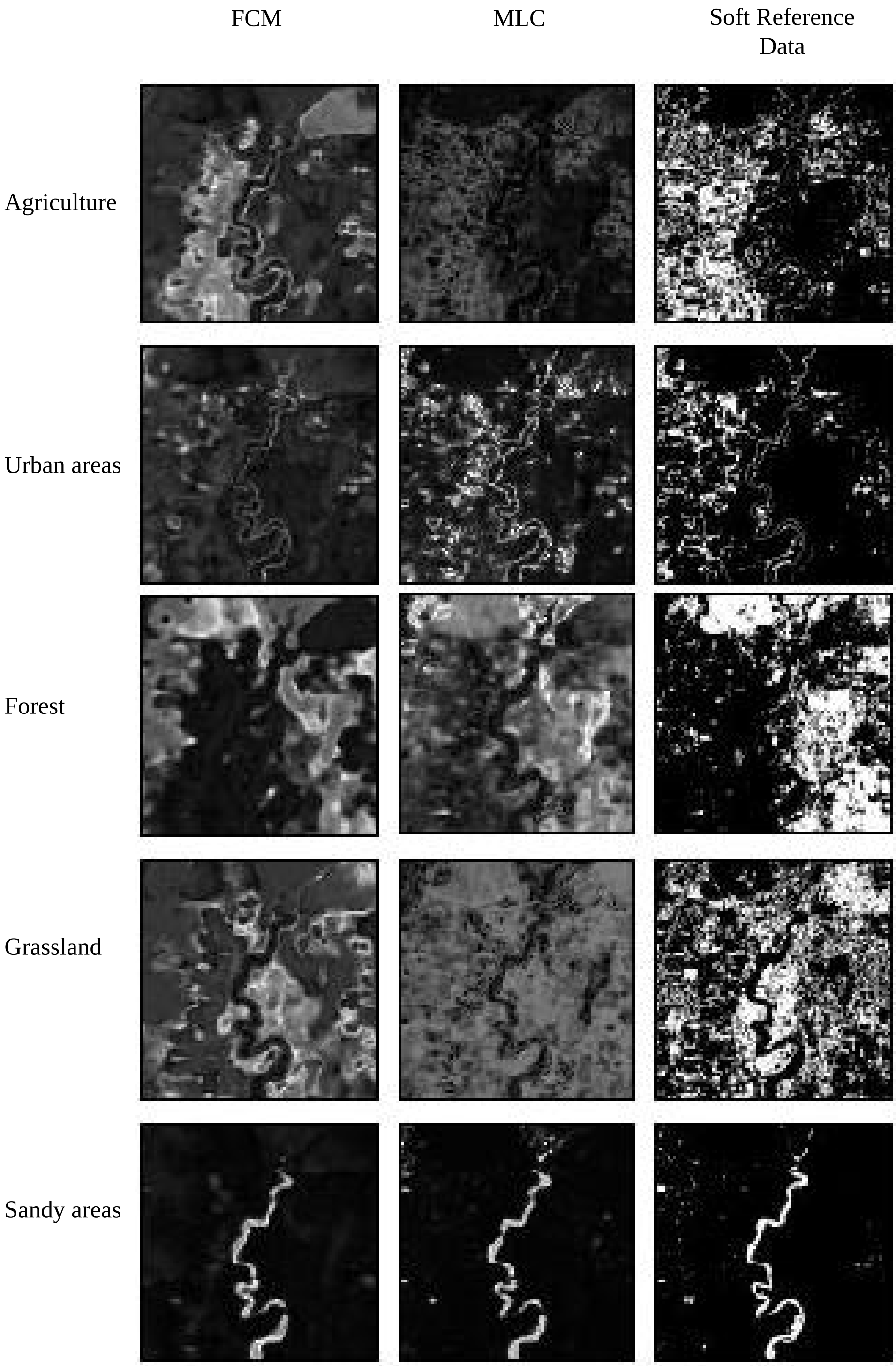

Figure 8: Fraction Images from MLC and FCM Compared with Soft Reference Data

6. Summary

Classification accuracy assessment is an important step of image classification process. A number of accuracy measures for both crisp and soft classifications have been proposed. No measure has been universally adopted. Often, a combination of accuracy measures may have to be used to describe the quality of classification completely. However, the current image processing software lack the provision of various accuracy measures. In this paper, details of accuracy measurement software named as CASCAM, written exclusively for the accuracy assessment of crisp and soft classification assessment of remote sensing data have been provided. A range of accuracy measures has been incorporated. The software, developed on Matlab platform, is interactive and user-friendly. The capabilities of the software have been demonstrated through a case study on soft land cover classification from IRS 1C LISS remote sensing data.

7. Acknowledgments

This software was developed when M. A. Slalan was a post-graduate at IIT Roorkee, India, on a scholarship from Ministry of High Education (MHE), Syria, under the scheme of Indian Council for Cultural Relation (ICCR). This paper has been written when M. A. Shalan was at Aston University, UK, under a scholarship from the MHE.

8. References

Bezdek, J.C., Ehrlich, R., and Full W., 1984, “FCM: The Fuzzy c-Means Clustering Algorithm”, *Computers and Geosciences*, Vol. 10, pp. 191-203.

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

Cohen, J. 1968, Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit, *Psychological Bulletin*, (70): 213-220.

Congalton, R. G. 1988, A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 54,593-600.

Congalton, R.G., Oderwald, R. G., and Mead, R. A. 1983, Assessing Landsat Classification Accuracy using Discrete Multivariate Analysis Statistical Techniques, *Photogrammetric Engineering and Remote Sensing*, Vol. 49, pp. 1671-1678.

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.

Foody, G. M., 1992, On compensation for chance agreement in image classification accuracy assessment, *Photogrammetric Engineering & Remote Sensing*, 58, 1459-1460.

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*, 50, 2-12.

Foody, G. M., 1996, Approaches for the production and evaluation of fuzzy land cover classifications from remotely sensed data, *International Journal of Remote Sensing*, 17, 1317-1340.

Foody, G. M., and Arora, M. K., 1996, Incorporating mixed pixel in the training, allocation and testing stages of supervised classification, *Pattern Recognition Letters*, 17, 1389-1398.

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*, 15, 619-631.

Fung, T., and LeDrew, E., 1988, The Determination of Optimal Threshold Levels for Change Detection using Various Accuracy Indices, *Photogrammetric Engineering & Remote Sensing*, Vo. 54, pp. 1449-1454.

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.

Hubert J.L, 1978, A general formula for the variance of Cohen's weighted kappa, *Psychology Bulletin*, 85, 183-184.

Jensen, J.R., 1986, Introductory Digital Image Processing, (*New York*, *Prentice Hall*)

Kent, J. T., and Mardia, K. V., 1988, Spatial Classification Using Fuzzy Membership Models, *I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, pp. 917-924.

Koukoulas, S., and Blackburn, G.A., 2001, Introducing New Indices for Accuracy Evaluation of Classified Images Representing Semi-Natural Woodland Environments, *Photogrammetric Engineering & Remote Sensing*, 67, 499-510.

Ma, Z., and Redmond, R.L., 1995, Tau coefficients for accuracy assessment of classification of remote sensing data, *Photogrammetric Engineering & Remote Sensing*, 61, 435-439

Maselli, F., Rodolf, A., and Conese, C., 1996, Fuzzy Classification of Spatially Degraded Thematic Mapper Data for the Estimation of Sub-pixel Components, *International Journal of Remote Sensing*, 17, pp. 537-551.

Maselli, F., Conese, C., and Petkov, L., 1994, Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications, *ISPRS Journal of Photogrammetry and Remote Sensing*, 49(2): 13-20

Mather, P. M., 1999, Computer Processing of Remote Sensing Data, Second Edition, *Chichester*, *John Wiley*.

Naesset, E., 1996, Conditional tau coefficient for assessment of producer's accuracy of classified remotely sensed data, *ISPRS Journal of Photogrammetry and Remote sensing*, 51, pp. 91-98

Rosenfield, G. H., and Fitzpatrick-Lins, 1986, A coefficient of agreement as a measure of thematic classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 52, 223-227

Shalan, M. A., Arora, M. K., and Ghosh, S. K. 2003, Evaluation of fuzzy classifications from IRS 1C LISS III data, *International Journal of Remote Sensing Letters* (In Press).

Story, M., and Congalton, R. G., 1986, Accuracy assessment: A user's perspective, *Photogrammetric Engineering & Remote Sensing*, 52, 397-399.

Wang, F., 1990, Fuzzy supervised classification of remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, 28, 194-201.