UNIT 14 ACCURACY ASSESSMENT

Structure

14.1 INTRODUCTION

In the previous unit, you have learnt about different image classification methods which help us to create thematic maps. We also discussed the advantages and limitations of some of the commonly used classification algorithms. Both supervised and unsupervised classification needs direct or indirect information of the surface characteristics e.g., for unsupervised classification the user must define the classes based on prior information of surface and in case of supervised classification, it is based on training samples from the surface. Quality and quantity of training samples, therefore, have considerable implication on the accuracy of the classified images.

Once you have an interpreted map, the obvious step is that you would want to know how much accurate those outputs are because inaccuracies in outputs will have their bearing on the map's utility and users would have greater confidence in utilising data if its accuracy is high. Hence, assessment of accuracy is a very important part of the interpretation as it not only tells you about quality of maps generated or classified images but also provides you with a benchmark to compare different interpretation and classification methods.

In this unit, you will learn about accuracy assessment and related concepts and methods. We will also discuss the role of sampling size for the purpose of accuracy assessment.

Objectives

After studying this unit, you should be able to:

- define accuracy assessment;
- discuss need for accuracy assessment;
- generate a error matrix for interpreted outputs;
- explain the role of sampling size in accuracy assessment; and
- list measures for accuracy assessment.

14.2 CONCEPT OF ACCURACY ASSESSMENT

Accuracy assessment is the final step in the analysis of remote sensing data which help us to verify how accurate our results are. It is carried out once the interpretation/classification has been completed. Here, we are interested in assessing accuracy of thematic maps or classified images which is known as thematic or classification accuracy. The accuracy is concerned with the correspondence between class label and 'true' class. A 'true' class is defined as what is observed on the ground during field surveys. For example, a class labeled as water on a classified image/map is actually water on the ground.

In order to perform accuracy assessment correctly, we need to compare two sources of information which include:

- interpreted map/classified image derived from the remote sensing data and
- reference map, high resolution images or ground truth data.

Relationship between these two sets of information is commonly expressed in two forms, namely -

- *error matrix* that describes the comparison of these two sources of information and
- *Kappa coefficient* which consists a multivariate measure of agreement between rows and columns of error matrix.

The error matrix and kappa coefficient have been discussed in detail in the sections 14.4 and 14.5, respectively. However, let us first discuss what accuracy assessment is along with its need and sources of errors.

14.2.1 Definition

Accuracy defines correctness and it measures the degree of agreement between a standard that assumed to be correct and a map created from an image. A visually interpreted map or classified image is only said to be highly accurate, when it corresponds closely with the assumed standard.

Accuracy is referred to in many different contexts. In the context of image interpretation, accuracy assessment determines the quality of information derived from remotely sensed data. Assessment can be either qualitative or quantitative. In qualitative assessment, you determine if a map 'looks right' by comparing what you see in the map or image with what you see on the ground. However, quantitative assessments attempt to identify and measure remote sensing based map errors. In such assessments, you compare map data with ground truth data, which is assumed to be 100% correct.

Accuracy of image classification is most often reported as a percentage correct and is represented in terms of consumer's accuracy and producer's accuracy. The *consumer's accuracy* (CA) is computed using the number of correctly classified pixels to the total number of pixels assigned to a particular category. It takes errors of commission into account by telling the consumer that, for all areas identified as category X, a certain percentage are actually correct. The *producer's accuracy* (PA) informs the image analyst of the number of pixels correctly classified in a particular category as a percentage of the total number of pixels actually belonging to that category in the image. Producer's accuracy measures errors of omission.

The term *consumer's accuracy* is used when a classified image is examined from the user's point of view. *Producer's accuracy* is used when same is viewed from analyst's perspective.

14.2.2 Need for Accuracy Assessment

The need for assessing accuracy of a map generated from any remotely sensed product has become a universal requirement and an integral part of any classification project. The user community needs to know accuracy of the classified image data being used. Moreover, different projects have different accuracy requirement and only those classified images which are above a certain level of accuracy can be used. Furthermore, accuracy becomes a critical issue while working in a Geographical Information System (GIS) framework where you use several layers of remotely sensed data. In such cases, it would be very important to know the overall accuracy which is dependent upon knowing the accuracy of each of data layers.

There are a number of reasons why assessment of accuracy is so important. Some of them are given below:

- accuracy assessment allows self-evaluation and to learn from mistakes in the classification process
- it provides quantitative comparison of various methods, algorithms and analysts and
- it also ensures greater reliability of the resulting maps/spatial information to use in decision-making process.

The need for accuracy assessment is emphasised in literature as well as in anecdotal evidence. For example, maps of wetlands from various states of India (e.g., Jammu and Kashmir, Rajasthan, Tamil Nadu, West Bengal) have been made by several central, state and local agencies using techniques that included satellite images, aerial photographs and field data. Simply comparing the various wetland maps would yield little agreement about location, size and extent of these. In the absence of a valid accuracy assessment you may never know which of these maps to use.

A map using remotely sensed or other spatial data cannot be regarded as the final product without taking necessary steps towards assessing accuracy or validity of that map.

A number of methods exist to investigate accuracy/error in spatial data including visual inspection, non-site-specific analysis, generating difference images, error budget analysis and quantitative accuracy assessment.

14.2.3 Sources of Errors

Classification error occurs when a pixel (or feature) belonging to one category is assigned to another category. *Errors of omission* occur when a feature is left out of the category being evaluated. *Errors of commission* occur when a feature is incorrectly included in the category being evaluated. For example, errors of omission are the allotment of errors of barren land on the ground to the agricultural land category on the map. This has caused the removal of an area of real barren land on the ground from the map. In the same way, errors of commission will be the assignment an area of agricultural land on the ground to the barren land on the map. Hence, an error of omission in one category will be counted as an error of commission in another category.

As you know that accuracy assessment is performed by comparing the map produced by remote sensing analysis to a reference map based on a different information source. One might ask why remote sensing analysis is needed if the reference map for comparison already exists. One of the primary purposes of accuracy assessment and error analysis in this case is to permit quantitative comparisons of different interpretations. Classifications done from images acquired at different times, classified by different procedures, or produced by different individuals can be evaluated using a pixel-by-pixel and point-by-point comparison. The results must be considered in the context of the application to determine which is the most correct or most useful for a particular purpose. In order to be compared, both the map to be evaluated, the reference map must be accurately registered geometrically to each other. They must have been classified using same scheme and at the same level of detail. One simple method of comparison is to calculate the total area assigned to each category in both maps and to compare the overall figures. This type of assessment is called *non-site-specific accuracy* (Fig. 14.1a). On the other hand, *site-specific accuracy* is based on a comparison of the two maps at specific locations (i.e. individual pixels in two digital images) (Fig. 14.1b). In this type of comparison, it is obvious that the degree to which pixels in one image spatially align with pixels in the second image contributes to the accuracy assessment result. It is important to note that errors in classification should be distinguished from errors in registration or positioning of boundaries. Another useful form of site-specific accuracy assessment is to compare field or training data at a number of locations within the image, similar to the way spatial accuracy assessment using ground check points is performed for digital orthophotographs and terrain models.

In *non-site-specific accuracy*, for example, two images or maps can be compared only on the basis of total area in each category as shown in Fig. 14.1a. In *site-specific accuracy* two images are compared on the basis of pixel-by-pixel or cell-by-cell as shown in Fig. 14.1b.

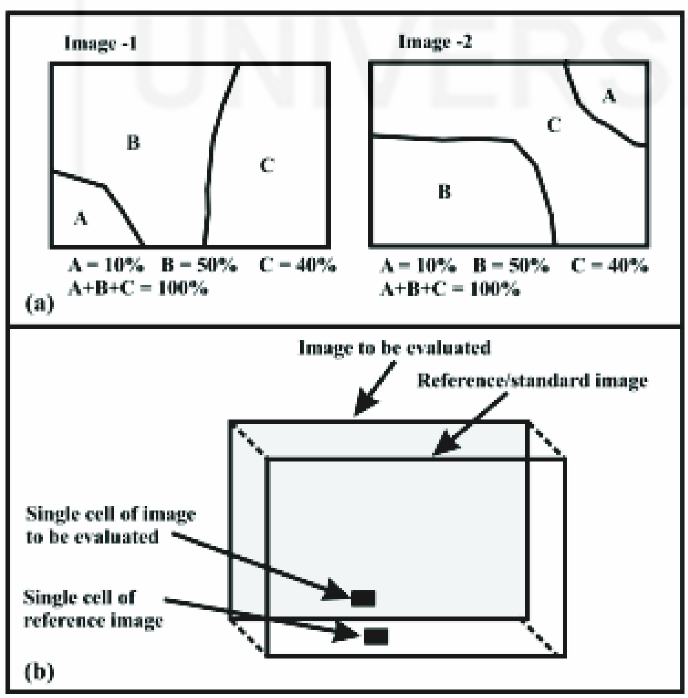

Fig. 14.1: (a) Non-site-specific accuracy in which two images are compared based on their total areas. Note that the area of image 1(i.e. A+B+C) is equal to the area of image 2 (i.e. A+B+C) and (b) site-specific accuracy in which two images are compared on a site-by-site (i.e. cell-by-cell or pixel by pixel) (source: modified from Campbell, 1996)

Check Your Progress I

Spend 5 mins

1) List the prerequisites for accuracy assessment.

....

....

....

....

14.3 CONSIDERATION OF SAMPLING SIZE AND SCHEME

Sample size is an important consideration while assessing the accuracy of remotely sensed data. Collection of sample size needs time and money. Therefore, it must be kept to a minimum. Yet it is critical to maintain a large enough sample size so that analysis performed is statistically valid. In remote sensing literature, many researchers have published equations and guidelines for choosing the appropriate sample size. The majority of them have used an equation based on the binomial distribution or the normal approximation to the binomial probability distribution to compute the required sample size. These techniques are statistically sound for computing the sample size needed to compute the overall accuracy of a classification or the overall accuracy of a single category. The equations are based on the proportion of correctly classified samples (e.g., pixels, clusters or polygons), on some allowable error. However, these techniques were not designed to choose a sample size for creating an error matrix. In the case of an error matrix, it is not simply a matter of correct or incorrect. Given an error matrix with n land cover categories, for a given category there is 1 (one) correct answer and n–1 incorrect answers. Sufficient samples must be acquired to be able to adequately represent this confusion. Therefore, the use of these techniques for determining the sample size for an error matrix is not inappropriate. Instead, the use of the multinomial probability distribution is recommended.

Binomial distribution treats errors for all classes equally, and therefore, can only estimate the accuracy of the map as whole.

Error matrix is the established form for reporting site-specific error. It is also known as *confusion matrix*.

Multinomial distribution investigates the errors associated with individual classes. It is based on the confusion matrix.

Traditional thinking about sampling does not often apply because of the large number of pixels in a remotely sensed image. For example, a 0.5% sample of a single Landsat TM scene can be over 3,00,000 pixels. Most, if not all, assessments should not be performed on a per-pixel basis because of problems with exact single pixel location. Practical considerations more often dictate the sample size selection. A balance between what is statistically sound and what is practically attainable must be found. A generally accepted rule of thumb is to use a minimum of 50 samples for each land class (LC) category in the error matrix. This rule also tends to agree with the results of computing sample size using the multinomial distribution. If the area is especially large or the classification has a large number of LC categories (i.e. more than 12 categories), the minimum number of samples should be increased to 75 to 100 samples per category.

The number of samples for each category can also be weighted based on the relative importance of that category within the objectives of the mapping or on the inherent variability within each of the categories. Sometimes, it is better to concentrate the sampling on the categories of interest and increase their number of samples while reducing the number of samples taken in less important categories. Also, it may be useful to take fewer samples in categories that show little variability such as water or forest plantations and increase sampling in the categories that are more variable such as uneven-aged forests or riparian areas. In summary, the goal is to balance the statistical recommendations to obtain an adequate sample from which to generate an appropriate error matrix within the objectives, time, cost and practical limitations of the mapping project.

Along with sample size, sampling scheme is an important part of any accuracy assessment. Selection of the proper scheme is absolutely critical in generating an error matrix that is representative of the entire classified image. Poor choice in sampling scheme can result in significant biases being introduced into the error matrix that may over or under estimate true accuracy. In addition, the use of proper sampling scheme may be essential depending on the analysis techniques to be applied to the error matrix. Many researchers have expressed opinions about proper sampling scheme to use, including everything from simple random sampling to stratified, systematic and unaligned sampling. Despite all these opinions, very little work has actually been performed in this area. One of the studies carried out on sampling simulations on three geographically diverse areas such as forest, agriculture and rangeland concluded that in all cases simple random sampling and stratified random sampling provided satisfactory results. Despite the desirable statistical properties of simple random sampling, this sampling scheme is not always very practical to apply. Simple random sampling tends to under-sample small but possibly very important areas unless the sample size is significantly increased. For this reason, stratified random sampling is recommended where a minimum number of samples are selected from each stratum (i.e. category). Even stratified random sampling can be somewhat impractical because of having to collect ground information for the accuracy assessment at random locations on the ground.

There are two problems which arise while using random locations:

- location can be very difficult to access and
- they can only be selected after the classification has been performed.

The second condition limits accuracy assessment data of being collected late in the project instead of in conjunction with the training data collection, thereby increasing costs of the project. In addition, in some projects time between project beginning and accuracy assessment may be so long as to cause temporal problems in collecting reference data.

14.4 CALCULATION OF CLASSIFICATION ACCURACY

You have learnt about the theoretical concept of accuracy assessment and sample size and scheme consideration for it. Let us now discuss about the methods to calculate accuracy. As we have discussed in the introductory section, accuracy is often expressed in terms of consumer's and producer's accuracies which is obtained from an error matrix. Now you will read about the error matrix, its generation and interpretation.

14.4.1 Error Matrix

Once a classification exercise has been carried out, there is a need to determine the degree of error in the end product which includes identified categories on the map. Errors are the result of incorrect labeling of the pixels for a category. The most commonly used method of representing the degree of accuracy of a classification is to build a $k \times k$ array, where k represents the number of categories. For example, in Table 14.1, the left hand side of the table is marked with the categories on the standard (i.e. reference) map/data. The top side of same table is marked with the same k categories but these categories represent end product of a created map to be evaluated. The values in the matrix indicate the numbers of pixels. This arrangement establishes a standard form which helps to find site-specific error in the end product and is known as *error matrix*. Error matrix is useful for the determination of overall errors for each category and misclassifications by category, as a result it is also known as *confusion matrix*. The strength of a confusion matrix is that it not only identifies the nature of the classification errors but also their quantities.

An *error matrix* is a square array of rows and columns in which each row and column represents one category/class in the interpreted map. Error matrix is also known as confusion matrix, evaluation matrix, or a contingency table.

Table 14.1: A hypothetical error matrix

		Classification result (i.e. classified image to be evaluated)								
Ground truth (i.e. reference image)		Forest	Bush	Crop	Urban	Open land	Water	Unclassified	**Row total**	**Accuracy (producer's accuracy)**
	Forest	**440**	40	0	0	30	10	10	**520**	**0.83**
	Bush	20	**220**	0	0	40	10	20	**290**	**0.71**
	Crop	10	10	**210**	10	50	10	60	**300**	**0.58**
	Urban	20	0	20	**240**	100	10	40	**390**	**0.56**
	Open land	0	0	10	10	**230**	0	10	**250**	**0.88**
	Water	0	20	0	0	0	**240**	10	**260**	**0.89**
	Column total	**490**	**290**	**240**	**260**	**450**	**280**	**240**	**1580**	
	Reliability (user's accuracy)	**0.90**	**0.76**	**0.88**	**0.92**	**0.51**	**0.86**			

Error matrix is a set array (rows and columns) that can be used to evaluate the degree of correctness of classified image. According to Campbell (1987), it is a method of reporting site-specific error. It is derived from a comparison of two types of maps such as a standard (reference) map and a classified map. It has two-dimensional arrangement in which rows show the reference data and column show the classified data.

14.4.2 Generation of Error Matrix

For generation of the error matrix you require two images namely, classified image (i.e. image under evaluation) and a standard or reference map derived from field survey. Sometimes, high resolution images are also used in the absence of a reference map. By comparing these two data, you can determine exactly how each site on standard/reference data is represented in the classified image. Before making a comparison, the classifier or analyst should make a network of appropriate (i.e. neither very small nor very large) uniform cells that form the units of comparison for site-specific accuracy assessment. Then two images are superimposed either by manually or digitally depending on the availability of the images. Then superimposed images are analysed on the basis of a cell-by-cell in case of manual comparison or pixel-by-pixel in case of digital assessment and tabulated for each cell/pixel the dominant category shown on the standard/reference data and category of the corresponding cell/pixel on the classified image. The classifier also keeps a count of the numbers of cells or pixels in each reference category as they are assigned to categories on the created image (see Table 14.1). Finally, the summation of the tabulation forms the basis for generation of the error matrix.

14.4.3 Interpretation of Error Matrix

Table 14.1 shows an example of an error/confusion matrix based on a classification result. Now let us try to understand how a confusion matrix is composed and how do you calculate accuracy based on the matrix.

You can read about the various components of the confusion matrix outlined below:

- rows correspond to classes in the ground truth map (or test set)
- columns correspond to classes in the classification result
- diagonal elements in the matrix represent the number of correctly classified pixels of each class, i.e. the number of ground truth pixels with a certain class name that actually obtained the same class name during classification. In the example above, 440 pixels of forest in the test set were correctly classified as forest in the classified image
- off-diagonal elements represent misclassified pixels or the classification errors, i.e. the number of ground truth pixels that ended up in another class during classification. In the example above, 40 pixels of forest in the test set were classified as bush in the classified image
  a) off-diagonal row elements represent ground truth pixels of a certain class which were excluded from that class during classification. Such errors are also known as *errors of omission* or *exclusion*. For example, 50 ground truth pixels of crop were excluded from the crop class in the classification and ended up in the open land class
  b) off-diagonal column elements represent ground truth pixels of other classes that were included in a certain classification class. Such errors are also known as *errors of commission* or *inclusion*. For example, 100 ground truth pixels of urban were included in the open land class by the classification and
- numbers in the column *unclassified* represent the ground truth pixels that were found not classified in the classified image.

Accuracy or Producer's Accuracy

Producer's accuracy is defined as the probability that any pixel in that category has been correctly classified. It is the values in column accuracy (producer's accuracy) present the accuracies of the categories in the classified image as shown in Table 14.1. It is calculated as given below:

$$\text{Accuracy (Producer's Accuracy)} = \frac{\text{Total number of correct pixels in a category}}{\text{Total number of pixels of that category derived from the reference data (i.e., row total)}}$$

The water category of Table 14.1, for example, has accuracy 0.89 meaning that approximately 89% of the water ground truth pixels also appear as water pixels in the classified image. This statistics is also known as *errors of commission*.

Accuracy for water category of Table 14.1 can be calculated as given below:
Total number of correct pixels for water = 240.
Total number of pixel in water row
= 0+20+0+0+0+240+10
= 270.
Hence, accuracy for water
= 240/270 = 0.89

The *average accuracy* is calculated as given below:

$$\text{Average Accuracy} = \frac{\text{Sum of all accuracy figures in accuracy column}}{\text{Total number of categories in the test set}}$$

The *average accuracy* of data given in Table 14.1

= (0.83+0.71+0.58+0.56+0.88+0.89) / 6

= 4.428 / 6 = 0.74

= 74.25%

This means average accuracy of the classification shown in Table 14.1 is 74.25% (or 0.74).

Reliability or User's Accuracy

User's accuracy is defined as the probability that a pixel classified on the image actually represents that category on the ground. The figures in row *reliability (user's accuracy)* present the reliability of classes in the classified image (Table 14.1). It is calculated as given below:

$$\text{Reliability (User's Accuracy)} = \frac{\text{Total number of correct pixels in a category}}{\text{Total number of pixels of that category derived from the reference data (i.e., column total)}}$$

The water category of Table 14.1, for example, has reliability 0.86 meaning that approximately 86% of the water pixels in the classified image actually represent water on the ground. This statistics is also called *errors of omission*.

Reliability for water category of Table 14.1 can be calculated as shown below:
Total number of correct pixels for water = 240.
Total number of pixel in water column
= 10+10+10+10+0+240
= 280.
Hence, reliability for water
= 240/280 = 0.86.

The *average reliability* is calculated as given below:

$$\text{Average reliability} = \frac{\text{Sum of all reliability figures in reliability row}}{\text{Total number of categories in the test set}}$$

Average reliability of data given in Table 14.1

= (0.90+0.76+0.88+0.92+0.51+0.86) / 6

= 4.81 / 6 = 0.80

= 80.27%

It indicates average reliability of the classification shown in Table 14.1 as 80.27% (or 0.80).

From the accuracy and reliability values for different classes given in Table 14.1, it can be concluded that the test set classes crop and urban were difficult to classify as many of such test set pixels were excluded from the crop and urban categories, thus the areas of these classes in the classified image are probably underestimated. On the other hand, class open land in the image is not very reliable as many test set pixels of other categories were included in the open land category in the classified image. Thus, the area of open land category in the classified image is probably overestimated.

Overall Accuracy

We have discussed about the individual classes and their accuracies. It is also desirable to calculate a measure of accuracy for the entire image across all classes present in the classified image. The collective accuracy of map for all the classes can be described using *overall accuracy*, which calculates the proportion of pixels correctly classified.

The overall accuracy is calculated as given below:

$$\textbf{Overall accuracy} = \frac{\begin{array}{c}\textbf{Sum of the diagonal elements}\\ \textbf{(as shown in bold letters in Table 14.1)}\end{array}}{\textbf{Total number of accuracy sites (pixels)}}$$

For the sample data presented in Table 14.1, the overall accuracy

= (440+220+210+240+230+240) / (490+290+240+260+450+280)

= 1580 / 2010 = 0.78

= 78%.

It indicates that overall accuracy of the classification shown in Table 14.1 is 78%.

14.4.4 Limitations

Use of confusion matrix for accuracy assessment has become a standard practice in quality assessment of remote sensing products. However, it is not free from limitations due to three crucial assumptions involved in the classification accuracy assessment:

- that the reference data are truly representative of the entire classification, which is quite unlikely
- the reference data and classified image are perfectly co-registered, which is impossible and
- there is no error in the reference data, which again is highly unlikely.

The actual accuracy of our classification is unknown because it is impossible to perfectly assess the true class of every pixel. It is possible to produce a misleading assessment of classification accuracy. Depending on how the reference data are collected, our estimate of accuracy may be either conservative or optimistic. If our estimate is less than the actual classification accuracy, then we have made a conservative estimate. Some of the sources of conservative estimates are:

- errors in reference data
- positional errors and
- minimum mapping unit of reference grid.

- positional errors and
- minimum mapping unit of reference grid.

Similarly, if estimate of accuracy is more than the actual classification accuracy, then we have made an optimistic estimate. Some of the sources of optimistic data estimate are:

- using training data for accuracy assessment
- sampling of reference data not independent of training data sampling and
- sampling from homogeneous groups of pixels.

Therefore, if error matrix is generated by using improper reference data collection methods, then the assessment can be misleading. Sampling methods used for reference data should be reported in detail so that potential users can judge whether there may be significant biases in the classification accuracy assessment.

Check Your Progress II

Spend 5 mins

Study the error matrix shown in Table 14.2. Calculate accuracy and reliability of the forest category.

Table 14.2: Error matrix

	Classification result (i.e. image to be evaluated)			
Ground truth (i.e. reference image)		**Forest**	**Water**	**Urban**
	Forest	**77**	**8**	**0**
	Water	**6**	**84**	**0**
	Urban	**0**	**0**	**74**

Accuracy of the forest class is ....

....

Reliability of the forest class is ....

....

14.5 KAPPA ANALYSIS

You have been introduced above that a commonly cited measure of mapping accuracy is the *overall accuracy* which is the number of correctly classified pixels (sum of major diagonal cells in the error matrix) divided by total number of pixels checked. Though, overall accuracy is a measure of accuracy for the entire image across all classes, it ignores off-diagonal elements (i.e. errors of omission and commission). Further, it is difficult to compare different overall accuracy values if different number of accuracy sites were used.

Two other accuracies such as producer's and consumer's accuracies are traditionally calculated from error matrix. The *producer's accuracy* is a measure of how well a certain area is classified. The *consumer's* or *user's accuracy* is a measurement of reliability of the classification or probability that a pixel on a map actually represents the category on the ground. The class producer's and consumer's errors are illustrated in our example in the above section. All these "naïve" accuracy measures can produce results due to classification of pixels by chance, therefore; do not provide avenues to compare accuracy statistically. This paves way for use of other accuracy assessment methods. In this section, we will discuss another commonly used method known as *Kappa analysis*, in which off-diagonal elements are incorporated as a product of the row and column marginal totals. It is a discrete multivariate technique used to assess classification accuracy from an error matrix. Kappa analysis generates a kappa coefficient or K_{hat} statistics, the values of which range between 0 and 1.

Kappa coefficient (K_{hat}) is a measure of the agreement between two maps taking into account all elements of error matrix. It is defined in terms of error matrix as given below:

$$\boldsymbol{K_{hat} = (Obs - exp)/(1 - Exp)}$$

Where,

Obs = Observed correct, it represents accuracy reported in error matrix (overall accuracy)

Exp = Expected correct, it represents correct classification

14.5.1 Calculation Steps

Kappa coefficient is calculated in the following steps:

Step 1: Construction of error (confusion) matrix (e.g., Table 14.3)

Omission error = 100 – producer's accuracy

Commission error = 100 – user's accuracy

Table 14.3: Error matrix

<table>
<tr><th colspan="7">Classification result (i.e. image to be evaluated)</th></tr>
<tr><th rowspan="6">Ground truth (i.e. reference image)</th><th></th><th>Forest</th><th>Water</th><th>Urban</th><th>Row Marginals</th><th>Commission error</th></tr>
<tr><td>Forest</td><td>28</td><td>14</td><td>15</td><td>57</td><td>51%</td></tr>
<tr><td>Water</td><td>1</td><td>15</td><td>5</td><td>21</td><td>29%</td></tr>
<tr><td>Urban</td><td>1</td><td>1</td><td>20</td><td>22</td><td>9%</td></tr>
<tr><td>Column Marginals</td><td>30</td><td>30</td><td>40</td><td>100</td><td></td></tr>
<tr><td>Omission error</td><td>7%</td><td>50%</td><td>50%</td><td></td><td></td></tr>
</table>

Step 2: Calculation of observed correct

Grand total = Sum of rows and columns

= 28+1+1+14+15+1+15+5+20 = 100

Total correct = Sum of the diagonal = 28+15+20 = 63

Observed correct = Total correct / Grand total = 63 / 100 = 0.63

Overall accuracy = 63%.

Step 3: Calculation of expected correct

Table 14.4: Error matrix showing the products of row and column marginals based on Table 14.3

		Classification result (i.e. image to be evaluated)		
Ground truth (i.e. reference image)		**Forest**	**Water**	**Urban**
	Forest	30x57=**1710**	30x57=1710	40x57=2280
	Water	30x21=630	30x21=**630**	40x21=840
	Urban	30x22=660	30x22=660	40x22=**880**

Grand total = Sum of products of row and column marginals

= 1710+1710+2280+630+630+840+660+660+880

= 10000

Total correct = Sum of products of diagonal

= 1710+630+880

= 3220

Expected correct = Total correct / Grand total

= 3220/10000

= 0.32

Note: For the calculation of expected correct you need to prepare an error matrix showing products of row and column marginals as shown in Table 14.4.

Step 4: Calculation of K_{hat}

Now you have values of observed correct and expected correct.

Observed correct = 0.63

Expected correct = 0.32

As you know that

K_{hat} = (Observed – Expected) / (1 – Expected)

This implies,

$\boldsymbol{K_{hat}}$ = (0.63 – 0.32) / (1 – 0.32)

= 0.31/0.68

= **0.45**.

Kappa coefficient of 0.45 implies that the classification process was avoiding 45% of the error that a completely random classification would generate (Congalton, 1991).

14.5.2 Advantages

One of the advantages of using this method is that you can statistically compare two classification products. For example, two classification maps can be made using different algorithms and you can use the same reference data to verify them. Two K_{hat}s can be derived like K_{hat1}, K_{hat2}. For each K_{hat}, the variance can also be calculated. Kappa coefficient, unlike the overall accuracy, includes errors of omission and commission. Computation of the Kappa

kappa and *average mutual information* (AMI). AMI is based on use of posteriori entropies for one map given that the class identity from the second map allows evaluation of individual class performance. Unlike the percentage correct or Kappa, that measures correctness, the AMI measures consistency between two maps. It provides an alternate viewpoint because it is used to access similarity of maps. For example, it can be used to compare the consistency between maps of the same region that have entirely different themes.

Accuracy assessment is still relatively new and is an evolving area in remote sensing. The effectiveness of different methods and measurement are still being explored and debated.

14.6 SUMMARY

We have studied in this unit about the concepts of accuracy assessments. This can be summarised in the following points:

- Assessing accuracy for each category as well as for the whole image is essential to compare the results of various classification techniques and quality and reliability of the results obtained.
- Accuracy in image classification is affected because of errors of inclusion and errors of exclusion.
- Sampling size is an important consideration for accuracy assessment and sufficient number of samples should be taken for the same.
- Error/confusion matrix can be used for accuracy and reliability assessments.
- Overall accuracy is a measure of accuracy for the whole image across all categories.
- Kappa coefficient is another method for accuracy assessment having a number of advantages over other methods.

14.7 UNIT END QUESTIONS

1) How is accuracy defined?
2) What is a confusion matrix?
3.) Describe errors of omission and errors of commission.

Spend 30 mins

14.8 REFERENCES

- Campbell, J. B., (1987), *Introduction to Remote Sensing*. The Guilford Press, New York.
- Campbell, J. B., (1996), *Introduction to Remote Sensing*. Taylor and Francis, London.
- 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.

14.9 FURTHER/SUGGESTED READING

- Campbell, J. B., (2006), *Introduction to Remote Sensing*. Taylor and Francis, London
- Jensen, J. R., (2004), *Introduction to Digital Image Processing*: *A Remote Sensing Perspective*, Prentice Hall, New Jersey.
- Janssen, L. L. F. and van der Wel, F. J. M., (1994), Accuracy assessment of satellite derived land-cover data: a review, *Photogrammetric Engineering and Remote Sensing*, Vol 60, pp 419-426.

14.10 ANSWERS

Check Your Progress I

Standard (reference) image and classified image data are the basic prerequisites for accuracy assessment.

Check Your Progress II

Accuracy of forest class is 0.9059, which is equal to 90.59%.

Reliability of forest class is 0.9277, which is equal to 92.77%.

Unit End Questions

1) Refer to subsection 14.2.1
2) Refer to subsection 14.4.1
3) Refer to subsections 14.2.3 and 14.4.3

GLOSSARY

Average accuracy: It is the sum of accuracy values in accuracy column which is divided by the number of classes in the test set.

Average reliability: It is the sum of reliability values in reliability column divided by the number of classes in the test set.

Atmospheric correction: It is the correction for the influence of atmosphere on the image by relative or absolute means.

Band radio: It is a narrow section of wavelengths or frequencies in radio broadcasting.

BRDF (**Bi-directional Reflectance Distribution Function**)**:** A function which describes the magnitude of upwelling radiance of the target in terms of illumination angle and the angle of view of the sensor.

Bi-linear interpolation: Involves the estimation of value at a pixel position, the nearest four neighbours around the pixel forming a rectangular plane are selected. An inverse weighted average of the known values based on their distances to the location of the pixel whose value is to be estimated is bi-linear interpolation.

Confusion matrix: It contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using data in the matrix.

Contrast: Ratio between the energy emitted or reflected by an object and its immediate surroundings.

Contrast enhancement: It is an image processing procedure that improves the contrast ratio of images. The original narrow range of digital values is expanded to utilise full range of available digital values.

Contrast ratio: The ratio of reflectances between the brightest and darkest parts of an image.

Contrast stretching: It involves the expanding of a measured range of digital numbers in an image to a larger range to improve contrast of the image and its component parts.

Co-variance: An average product of the differences between the pixels values in each band and the mean of each band. It measures the tendencies of pixel values for the same pixel but in different bands, to vary with each other in relation to the means of their respective bands.

Cubic convolution: A high-order resampling technique in which brightness value of a pixel in a corrected image is interpolated from the brightness values of 16 nearest pixels around the location of the corrected pixel.

Digital image processing: The computer manipulation of digital number values of an image.

Digital number: It is the value assigned to a pixel in a digital image.

Distortions: Are the errors in the remotely sensed image in terms of the pixel shape, position or the recorded value.

Enhancement : A process of enhancing certain features in the image making it more interpretable to the human eye for a particular application.

Geocoding: It is a special case of rectification that includes geographical registration or coding of pixels in an image. Geocoded data are images that have been rectified to a particular map projection and pixel size. The use of standard pixel sizes and coordinates permits convenient overlaying of images from different sensors and maps in a GIS.

Geometric correction : It involves image processing procedure that corrects spatial distortions in an image.

Gray scale: It is a sequence of gray tones ranging from black to white.

Histogram: It is a way of expressing frequency of occurrence of values in a data set within a series of equal ranges or bins, height of each bin representing frequency at which values in data set fall within the chosen range. A cumulative histogram expresses frequency of all values falling within a bin and lower in the range. A smooth curve derived mathematically from a histogram is termed the probability density function.

Histogram equalisation: A process of re-distributing pixel values so that there are approximately the same number of pixels with each value within a range thereby creating a nearly flat histogram for output image.

Hue: It represents the dominant wavelength of a colour.

Image classification: The process of dividing all channels within a multichannel digital remote sensing dataset into discrete surface cover categories or information themes.

Image enhancement: The process of altering the appearance of an image so that the interpreter can extract more information.

Image interpretation: It is a process in which a person extracts information from an image.

Image interpretation key: The characteristic or combination of characteristics that enable an interpreter to identify an object on an image.

Input-to-output mapping: Represents locations of the points in the slave are transferred to the reference by calculating them based on the coefficients of the polynomial transformation.

Intensity: The brightness ranging from black to white.

Mean: Statistical average which is calculated as the sum of a set of values divided by the number of values in the set.

Median: The central value in a set of data such that an equal number of values are greater than and less than the median.

Mode: Represents most commonly occurring value in a set of data. For an image histogram, peak of the curve represents mode.

NDVI (Normalised Difference Vegetation Index): It is a remote sensing way of measuring whether vegetation is alive or dead based on information from visible especially red and near-infrared bands.

Orthorectification: Process of pixel-by-pixel correction of an image for topographic distortion. Every pixel in an orthorectified image appears to view the Earth from directly above, i.e., the image is in an orthographic projection.

Output-to-input mapping: Represents positions in the reference map which are related to its corresponding positions in the slave by calculating them based on the coefficients of the polynomial transformation.

Overall accuracy: Total number of correctly classified pixels (diagonal elements) divided by the total number of test pixels.

Pan-sharpening: It is a computer-enhancement algorithm for improving the resolution of an image. It involves combining the high spatial resolution of a panchromatic image with the colour information from (lower resolution) multispectral data. The result is a colour image that has high resolution of the panchromatic image.

Pixel: A picture element; smallest element of an image that has been electronically coded in an array.

Polynomial: It is a mathematical expression consisting of variables and coefficients.

Producer's accuracy: Fraction of correctly classified pixels with regard to all pixels of that ground truth class.

Rectification: Process of alignment of an image to a map (map projection system). In many cases, the image must also be oriented so that north direction corresponds to the top of the image. It is also known as georeferencing.

Registration: Process of alignment of one image to another image of the same area not necessarily involving a map coordinate system.

Resampling: It is an estimation of pixel values of a rectified image.

Standard deviation: A square root of the variance of a set of values which is used as a measurement of the spread of the values.

Supervised classification: A classification procedure guided by the analyst.

Thematic mapper: An advanced multispectral scanning Earth resources sensor designed to achieve higher image resolution, sharper spectral separation, improved geometric fidelity and greater radiometric accuracy and resolution more than the multispectral sensor. Thematic mapper data are sensed in seven spectral bands simultaneously.

Trade-off: A result of changing one factor in a remote sensing system, there are compensating changes elsewhere in the system; such a compensating change is known as a trade-off.

Training area: A sample of the Earth's surface with known properties; the statistics of the imaged data within the area are used to determine decision boundaries in classification.

Training site: An area of terrain with known properties or characteristics that is used in supervised classification.

Unsupervised classification: The digital information extraction technique in which the computer assigns pixels to categories with no instructions from the operator.

Variance: The measure of central tendency.

ABBREVIATIONS

AMI	:	Average Mutual Information
ARVI	:	Atmospherically Resistant Vegetation Index
AVHRR	:	Advanced Very High Resolution Radiometer
BRDF	:	Bi-directional Reflectance Distribution Function
CA	:	Consumer's Accuracy
DN	:	Digital Number
EM	:	Electromagnetic
EMR	:	Electromagnetic Radiation
ETM$^+$	:	Enhanced Thematic Mapper Plus
EVI	:	Enhanced Vegetation Index
FWHM	:	Full Width at Half-Maximum
GCP	:	Ground Control Point
GIS	:	Geographical Information System
IHS	:	Intensity Hue Saturation
IRS	:	Indian Remote Sensing Satellite Series
ISODATA	:	Iterative Self-Organising Data Analysis Technique
ISRO	:	Indian Space Research Organisation
LC	:	Land Class
LISS	:	Linear Imaging Self-Scanning System
MODIS	:	Moderate Resolution Imaging Spectroradiometer
MSS	:	Multispectral Scanner
MXL	:	Maximum Likelihood
NASA	:	National Aeronautical and Space Administration
NDVI	:	Normalised Difference Vegetation Index
NIR	:	Near Infrared
PA	:	Producer's Accuracy
PCA	:	Principal Component Analysis
RGB	:	Red Green Blue
RMSE	:	Root Mean Square Error
RMS	:	Root Mean Square
SDSS	:	Spatial Decision Support Systems
SPOT	:	Système Pour l'Observation de la Terre
TM	:	Thematic Mapper
UTM	:	Universal Transverse Mercator
VI	:	Vegetation Index