

ACCURACY OF FOREST AND TREE COVER ASSESSMENT

6.1 Introduction

In remote sensing technology, accuracy is assessed to give a reliability level to the results obtained after interpreting the satellite data and in wall to wall approach the same is determined by error matrix. On the other hand, in a sampling estimation it is only the precision which is determined by calculating standard error of the estimates and not the accuracy.

In the current report, three important statistics namely; accuracy of forest cover mapping and precision of tree cover and growing stock estimates have been calculated. It is to be reiterated that assessment of forest cover of the country has been carried out by interpretation of the satellite data using wall to wall approach, whereas estimation of tree cover and growing stock have been done by field inventory using sampling based approach. Therefore, separate methodologies have been followed for quantifying the accuracy of forest cover assessment and precision of tree cover and growing stocks estimates.

6.2 Methodology for Accuracy Assessment of Forest Cover

In remote sensing technology the errors may occur due to many reasons. The radiometric errors arise due to random variations in the functioning of the sensor or by the intervening atmosphere between the terrain and remote sensing system, i.e., the radiant flux reflected by the terrain may not truly resemble the energy recorded by the sensor. The geometric errors occur due to variations in altitude, velocity of sensor platform, panoramic distortions, earth curvature,

atmospheric refraction, relief, displacement, etc. There are procedures to minimize these errors but these cannot be totally eliminated. Remote sensing system also has limitations on account of spatial, spectral, temporal and radiometric resolutions. Besides this, errors arise due to interpretation and classification of cloud or shadow effects, or seasonal variation in the canopy of deciduous trees or bushy and agricultural vegetation getting mixed with forest crop, human errors etc. While interpreting the remote sensing data, all these errors influence the accuracy of the assessment.

In forest cover assessment 'accuracy assessment' describes as to how correctly the satellite imageries have been interpreted to match the position on the ground. This is done by comparing the interpreted data of selected areas with 'reference dataset' also called 'the ground truth'. There is no standard or universal method of assessing the accuracy. It can be done through a carefully designed field verification of the selected sampled areas or by using very high resolution satellite images or both.

Ideally, the sampled areas randomly selected for verification should be distributed over the whole assessment area. However, there are certain difficulties in this approach. Firstly, it is difficult to do field verification of all such areas some of which may not be easily accessible and it would require massive manpower, time and cost; and secondly, there is a time lag of about 1 to 2 years between the date of satellite data used and the ground truth period. To overcome the difficulties, FSI has designed its own methodology under which the recent field inventory data of NFI available with FSI are used. In addition, high resolution

satellite data (5.8m) are also used for verification.

In the current report, out of the 15,197 sample plot data of NFI collected during 2006-2008 by FSI, a total of 4,302 points were selected randomly for accuracy assessment, both from forest and non-forest areas. Care was taken to see that about 100 sample points fell within each of the 3 classes of the forest cover and in each physiographic zone. Since the geo-coordinates of the sample points were known, points with corresponding coordinates were selected on the classified forest cover map. The ground truth data from NFI for all the points were recorded giving land use class at each point in about 1 ha area, and compared with the classified image to prepare the error matrix. Higher resolution satellite data (5.8 m) for such sample points was used to supplement ground truth.

The error or confusion matrix is an array of numbers arranged in rows (generally, map classification) and columns (generally, ground truth) wherein, the numbers of rows and columns are equal, representing different classes (dense forest, open forest, etc). The randomly selected sampling units are equal in extent to the minimum mappable area of forest cover in this study. Any entry made along the major diagonal of the error matrix implies agreements and the non-diagonal elements indicate disagreements or 'wrong' classification. The percentage of matching ('correctly classified') sampling units (i.e. sum of all diagonal elements) out of the total considered sampling units in the error matrix provides a measure of 'overall accuracy'. Similarly, accuracy of each class can be measured by calculating the

percentage of correctly classified sampling units (diagonal element) compared to the total sampling units in that class in that row or column.

6.3 Accuracy of Forest Cover: Findings

The error matrix has been prepared and presented in Table 6.3.1. The first diagonal element at column 1 and row 1 i.e. 197 indicates the correct classification of VDF on 197 out of 202 sample points. The off-diagonal points show misclassification in the respective classes.

The error matrix reveals that out of 4,302 sampling points where observations were made, classification of 3,962 sampling points (the sum of the elements along the main diagonal of the matrix) was found correct. The 'overall accuracy' of classification therefore, works out to be 92.10% which would be termed 'high' by the general accepted norms.

Further, a simplified error matrix can be created by grouping land use classes broadly into two classes 'forest' and 'non - forest'. This is done by combining VDF, MDF and OF into one class viz. 'forest' and scrub and non-forest classes into 'non-forest'. The simplified error matrix is given in Table 6.3.2.

The simplified error matrix reveals that interpretation of 4,118 sample points was found correct, giving an overall accuracy of 95.72%.

Besides the overall accuracy, the accuracy of individual classes has also been determined by

Table 6.3.1: Error matrix

Classification Classes	Ground truth (based on field inventory data)						User's Accuracy(%)
	VDF	MDF	OF	Scrub	NF	Total	
VDF	197	4	0	0	1	202	97.52
MDF	7	1,350	63	10	22	1,452	92.98
OF	3	42	925	9	34	1,013	91.31
SCRUB	0	7	3	168	9	187	89.84
NF	3	40	55	28	1,322	1,388	91.30
Total	210	1,443	1,044	215	1,388	4,302	
Producer's Accuracy (%)	93.81	93.56	88.43	78.14	95.24		
Overall Accuracy	92.10%						
Overall Kappa Statistic	0.89						

Table 6.3.2: Simplified error matrix

Classification Classes	Ground truth (based on field inventory data)			Users's Accuracy(%)
	Forest	Non-Forest	Total	
Forest	2,591	76	2,667	97.15
Non-Forest	108	1,527	1,635	93.39
Total	2,699	1,603	4,302	
Producer's Accuracy (%)	96.00	95.26		
Overall Accuracy	95.72%			
Overall Kappa Statistic	0.91			

calculating 'producer's accuracy' and 'user's accuracy'.

The producer's accuracy measures how correctly any specific land-use class has been classified. It is derived by dividing the number of correct sampling points in one class divided by the total number of points derived from reference data. It includes the error of omission which refers to the proportion of observed features on the ground that is not classified in the map. The more is the error of omission, the lower is the producer's accuracy.

The user's accuracy on the other hand is a measure of the reliability of the map. It informs the user how well the map represents what is really on the ground. It can be obtained by dividing the correct classified units in a class by the total number of units that were classified in that class. One class in the map can have two types of classes on the ground. The 'right' class, which refers to the same land cover class in the map and on the ground; and 'wrong' classes, which show a different land cover on the ground than that depicted on the map. The latter classes are referred to as the errors of commission. The more is the error of commission, lower is the user's accuracy.

From Table 6.3.1 it is found that the producer's accuracy for VDF, MDF, OF, Scrub and Non-forest classes are 93.81, 93.56, 88.43, 78.14 and 95.24% respectively. Similarly, user's accuracy for these classes is 97.52, 92.98, 91.31, 89.84 and 91.30% respectively. These levels of accuracy are well above the acceptable norms.

Similarly Table 6.3.2 depicts simplified error matrix showing higher accuracy levels. The producer's accuracy for forest and non-forest classes is found to be 96.00 and 95.26 % respectively while user's

accuracy for these classes is 97.15 and 93.39% respectively.

To further authenticate the results of accuracy, the Kappa analysis, which is a multivariate technique, provides a statistic known as K_{HAT} . This coefficient gives a measure of overall agreement of matrix. In contrast to the overall accuracy (the ratio of the sum of diagonal values to total number of sampling units in the matrix), the Kappa coefficient (K_{HAT}) also takes into account the non-diagonal elements. This statistic usually ranges between 0 and 1 and is used to indicate whether the correct values of the error matrix are due to true agreement or due to chance agreement. Any classification having Kappa coefficient more than 0.6 is considered as statistically sound. K_{HAT} calculated from the error matrix given at Table 6.3.1 is 0.89, which indicates that an observed classification is 89% better than the one resulting from chance.

6.4 Precision of Tree Cover Estimates

As already mentioned in the previous chapters the estimation of tree cover is done through a separate exercise where sample survey data collected during NFI specially for TOF are used. The exercise is not linked to the forest cover mapping. Therefore, instead of accuracy, precision has been calculated.

The detailed methodology of NFI has already been explained in the previous chapters. To state briefly, the country has been stratified into 14 physiographic zones and districts are randomly selected from each zone as first stage unit. Assessment of TOF is done separately for rural & urban areas of the districts. For rural areas, TOF resources are identified and stratified into block, linear and scattered strata with

the help of high resolution satellite data (5.8m).

For urban areas, UFS blocks are taken as sampling units. Optimum numbers of sampling points are selected from each stratum of rural areas and UFS blocks from urban areas for field inventory. The total tree cover for a selected district is obtained by aggregating tree cover of all components. The tree cover thus obtained for the districts are used to estimate the tree cover for the physiographic zone by using ratio method of estimation. Further, aggregation of the data of physiographic zones provides the national level estimate. Since physiographic zone is the stratum of the NFI the precision of estimate is adjudged at physiographic zone level by using necessary formula. After combining estimates at physiographic zone, the national level estimate have been derived following stratified sampling formulae.

Table 6.4.1 gives the precision of estimates (SE%) for each physiographic zone. The overall

Table 6.4.1: Physiographic zone wise precision of estimates

S. No.	Physiographic zone	S.E.%
1	Western Himalayas	3.21
2	Eastern Himalayas	19.61
3	North East	10.74
4	Northern Plains	9.92
5	Eastern Plains	10.79
6	Western Plains	8.28
7	Central Highlands	10.13
8	North Deccan	7.58
9	East Deccan	9.84
10	South Deccan	12.45
11	Western Ghats	6.37
12	Eastern Ghats	6.78
13	West Coast	6.42
14	East Coast	9.64
	National level	2.54%

precision of the estimate at the country level is 2.54 %.

