ACCURACY ASSESSMENT AS A TOOL OF VALIDATION OF REMOTELY SENSED DATA OF LAND USE/LAND COVER OF NATIONAL CAPITAL TERRITORY OF DELHI, INDIA

Sandesh Yadav\textsuperscript{1*} and Shams Perwaiz\textsuperscript{2}
\textsuperscript{1}Freelance Researcher, New Delhi, India
\textsuperscript{2}Teacher, Sr. High School, Jokihat, Araria (Bihar) India
Corresponding author’s Email: sandesh\_official@yahoo.in

Abstract: Accuracy assessment is a tool of validation of remotely sensed data and other spatial information and analyses the ‘goodness factor’ of the maps derived from satellite images. If we talk about the remotely sensed data of land use/land cover then accuracy assessment helps in analyzing the spatial details of particular landscape which in lieu serve the various purposes of policy making and administrative planning. The data related to land use/land cover is of utmost importance and helps in environmental modelling specifically models dealing with climate change at all levels viz. micro level, meso level and macro level. The present research study aims to analyze the accuracy assessment of remotely sensed data of land use/land cover of national capital territory of Delhi, India. Three different sets of Landsat images for the years 1990, 2000 and 2015 have been evaluated by level of overall accuracy, producer’s accuracy, consumer’s accuracy and kappa coefficient.

Key words: Accuracy Assessment, Validation, Land Use/Land Cover, Kappa Coefficient.

Introduction
Accuracy assessment or validation of spatial data facilitates the decision making process. The estimates based on accuracy assessment are more accurate, efficient and effective. Mapping related to different issues viz. forestation/deforestation, wetlands and estimation of land use/land cover will be inappropriate and incomplete without accuracy assessment or validation. This investigates validity of map with all possible error. The present study of land use/land cover of NCT of Delhi has been discussed in the light of quantitative accuracy assessment.

Study Area
The study looks at the city of Delhi, national capital of India (Figure 1) due to its high degree of urbanization, dominant grey infrastructure and high density of population. It is located between the latitude 28°24´17´´ N and 28°53´00´´ N and longitude 76°50´2´´ E and 77°20´37´´ E. It has an area of about 1483 km\textsuperscript{2} with maximum length of 51.90 km and maximum width of 48.48 km. The study employs Landsat-4, 5, TM for 7 November 1990; Landsat-7 ETM+ for 16 November 2000 and Landsat-8 OLI for 9 November 2015 (see Table 1). All images were obtained from the archives of United States Geological Survey (USGS). The Three scenes fell within the path 147 and row 40 of the WRS-2 (Worldwide Reference System) from which the data for the location under the study could be extracted. All bands 1-5 and 7 have spatial resolution of 30 m and the thermal infrared band (band 6) has a spatial resolution of 60m for Landsat 7 and 120m for Landsat 5. For Landsat OLI, thermal infrared band (10, 11) has a spatial resolution of 100m.

Objective of the Study
The objectives of the present study are as follows:
1. To identify the land use/land cover classes (1990-2015) through the Anderson’s satellite image classification in the National Capital Territory of Delhi, India.
2. To analyze the accuracy assessment of land use/land cover classification (1990-2015) of National Capital Territory of Delhi, India.
Methodology
The present research study involves steps viz. pre-image processing, Image processing, post-image processing, and generation of LU/LC map and accuracy assessment. The description/definition of land use/land cover terms has been adopted from Manual of National wide land use/Land Cover Mapping using multi-temporal Satellite data of NRSC (2006). Work of Goodchild (1989) on accuracy of space derived data and Janssen (1904) on accuracy of land use/land cover data, are also considered to derive results.
Data Set and Data Sources
In the present research study, the land surface biophysical types retrieved from Landsat thematic mapper, enhanced thematic mapper and operational land imager images of Delhi for 1990, 2000 and 2015 were analyzed. Table 1 shows the details of satellite images used in the present research study. The satellite images of the National Capital Territory of Delhi for three different periods that is 1990, 2000, 2015 were obtained from the United States Geological Survey (USGS) and National Remote Sensing Centre (NRSC), Hyderabad, India.

Table 01: Landsat Meta-Data for the Study Area

| City  | Path/Row | Satellite Platform | Band | Date Acquired | Resolution |
|-------|----------|--------------------|------|---------------|------------|
| Delhi | 147/40   | Landsat-4,5 TM     | 1-5 & 7, 6 | 07/11/1990 | 30m/120m |
|       |          | Landsat-7 ETM+     | 1-5 & 7, 8, 6 | 16/11/2000 | 30m/60m |
|       |          | Landsat-8 OLI      | 1-9, 10, 11 | 09/11/2015 | 30m/100m |

Source: United States Geological Survey, 2016

Tool of Analysis
For accuracy assessment a discrete multivariate technique was used called KAPPA Index (Cohen, 1960). The result of performing a KAPPA analysis is a KHAT statistic (an estimate of KAPPA), which is another measure of agreement or accuracy. The KHAT statistic is computed as:

\[ \Delta = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} + x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{i+} + x_{+i})} \]

Where, \( r \) is the number of rows in the matrix, \( x_{ii} \) is the number of observations in row \( i \) and column \( i \), \( x_{i+} \) and \( x_{+i} \) are the marginal totals of row \( i \) and column \( i \), respectively, and \( N \) is the total number of observations (Bishop et. al., 1975). Congalton et al., 1993 originally proposed the use of this statistic for remotely sensed data to clear up the confusion caused by a topological error. KHAT equation assumes a multinomial sampling model and that the variance is derived using the delta method. Cohen’s (1960) kappa index is used to measure the accuracy. Hudson (1987) worked on correct formation of kappa coefficient. The first mention of a kappa-like statistic is attributed to Galton (1892) and Stehman (1995).

Software and Platforms
The following software and platforms are involved in the present research study:

1. ArcGIS ver. 10.1
2. Statistical Package for the Social Sciences (SPSS).
3. Office 365.

Results and Discussion
Image Interpretation
Pre-image processing is the first stage in dealing with spatial data and involves steps like layer stacking, band combination, geometric correction and radiometric correction. Here, geometric correction involves removal of aberrations of image while on the other hand; radiometric corrections involve atmospheric correction, noise correction and shadow correction. Image processing begins with the defining of training sites (AOI) and these training sites were selected on the basis of clearly identified area in satellite image through the on-screen digitized features. This digitization process is followed by the extraction of signatures in order to obtain information of each pixel on the image. The final step in image processing is the supervised classification of image to obtain different classes of land use/land cover. During this process, non-parametric rule was used and image was classified into six class’s viz. Water body, vegetation, crop land, fallow land, river bed and built-up. Post-image processing step involves thematic editing using the
re-coding step of supervised classification and map was generated through the layout processing in ArcGIS 10.4.

**Figure 03: Land Use and Land Cover, 1990, 2000 and 2015, NCT of Delhi**

![Image of land use and land cover maps for 1990, 2000, and 2015]

Source: Landsat Imagery, 7th November 1990, 16th November 2000 and 9th November 2015.

**Land Use/Land Cover Changes during the Period 1990-2015**

Before carrying out interpretation of maps, we need to understand the colour scheme adopted in the maps (Fig. 3) to understand the changes in land use/land cover during the period 1990-2015 in the study area. As per the colour scheme, water body and vegetation cover are depicted by their natural colors that is blue and green respectively while built-up and cropland are depicted by red and yellow colour respectively. The other categories viz. Fallow land and barren land are represented by pink and magenta colour respectively. The prevalence of red colour on the LULC map of 1990 shows the dominance of built-up with variations in built-up density at the core and peripheral regions of the study area. At the core, northern part, north-eastern part, eastern part, south-eastern part, southern part, south-eastern part, southern part, and north-western part are regions of high built-up density while on the other hand, the rarefied concentration of the built-up exist in the peripheral areas of northern part, south-eastern part, south-western part, western part and north-western part of the study area. The LULC category of cropland has its limits in peripheral areas of northern part, south-western part and north-western part of the study area. The vegetation cover was present in the core and peripheral areas of the eastern part and south-eastern part of the study area. The fallow land has been seen in the peripheral areas of northern part, south-western part and north-western part of the study area. The LULC map for the year 2000 confirms an increase in the density of built-up both at the core and periphery of the study area. During the period 1990-2000, the built-up expansion took place in the south-eastern part, south-western part, western part and north-western part of the periphery while the built-up became denser at the core of the study area. The LULC map for the year 2015 shows shrinkage in areal expansion of cropland, vegetation cover and fallow land due to built-up expansion. During the period 2000-2015, the cropland experienced slight decrease in area in the peripheral parts of south-western part, western part and north-western part of the study area. On the other hand, the vegetation cover got reduced in area drastically with huge losses at the periphery of the study area.

The study area has an area of 1,48,413.0 hectare (Table 3). The categories occupying major share include built-up area (42.4 percent), vegetation (30.4 percent), and cropland (17.8 percent) while the minor share occupying categories include fallow land (7.4 percent), water bodies (1.3 percent) and riverbed (0.8 percent). During the period 1990-2000, built-up experienced tremendous increase from 42.4 percent (1990) to 49.7 percent (2000). These categories include vegetation [30.4 percent (1990) and 25.1 (2000)], crop land [17.8 percent (1990) and 16.9 percent (2000)], fallow land [7.4 percent (1990) and 6.4 percent (2000)], river bed [0.8 percent
(1990) and 0.7 percent (2000)) and water bodies [1.3 percent (1999) and 1.2 percent (2000)]. During the period 2000-2015, the similar trend was observed in land use/land cover changes. The built-up showed drastic increase in its area while the other categories (except river bed) followed similar declining trend as in the year 2000. Consequently, the built-up increased from 49.7 percent (2000) to 58.2 percent (2015) and the categories which experienced decline in area includes water bodies [1.2 percent (2000) and 1.1 percent (2015)], vegetation [25.1 percent (2000) and 21.3 percent (2015)], fallow land [6.4 percent (2000) and 3.6 percent (2015)], riverbed [0.7 percent (2000 and 2015)] did not show any change.

Table 03: Trend and Magnitude of Land Use and Land Cover, 1990-2015, NCT of Delhi

| LULC Classes  | 1990 | 2000 | 2015 | 1990-2000 | 2000-2015 |
|---------------|------|------|------|-----------|-----------|
|                | Hectares | Percent | Hectares | Percent | Hectares | Percent | Percent | Percent |
| Water Bodies   | 1899.4 | 1.3  | 1819.8 | 1.2    | 1583  | 1.1    | -0.05   | -0.15   |
| Vegetation     | 26402.8 | 30.4 | 37225.5 | 25.1   | 31675.3 | 21.3   | -5.31   | -3.73   |
| Cropland       | 45118.8 | 17.8 | 25012.4 | 16.9   | 22468.7 | 15.1   | -0.936  | -1.71   |
| Fallow land    | 10980.9 | 7.4  | 9519.6 | 6.4    | 5308.0 | 3.6    | -0.989  | -2.83   |
| Riverbed       | 1114.4  | 0.8  | 1016.2 | 0.7    | 1001.9 | 0.68   | -0.066  | -0.009  |
| Built-up       | 62897.3 | 42.4 | 73819.5 | 49.7   | 86375.6 | 58.2   | +7.539  | +8.46   |

Source: Calculated and compiled by the scholar from Landsat imagery of 1990, 2000 and 2015

Accuracy Assessment

A confusion matrix scheme is used for quantitative accuracy assessment of image classification for all the years. The verification and original classification pixels are along x-axis and y-axis respectively in the resultant matrix. Producer’s accuracy and User’s accuracy, has been calculated for each class.

Table 04: Accuracy Assessment Land Use and Land Cover Classification, 1990

| LULC Classes | Water body | Vegetation | Crop land | Fallow land | Riverbed | Built-up | Row total | User’s Accuracy |
|--------------|------------|------------|-----------|-------------|----------|----------|-----------|----------------|
| Water body   | 45         | 0          | 1         | 2           | 1        | 1        | 50        | 90.00          |
| Vegetation   | 1          | 30         | 2         | 1           | 2        | 0        | 36        | 83.33          |
| Cropland     | 2          | 0          | 38        | 0           | 0        | 0        | 40        | 95.00          |
| Fallow land  | 1          | 1          | 0         | 28          | 4        | 2        | 36        | 77.78          |
| Riverbed     | 1          | 0          | 1         | 1           | 38       | 0        | 41        | 92.68          |
| Built-up     | 0          | 0          | 0         | 1           | 1        | 45       | 47        | 95.74          |
| Column total | 50         | 31         | 42        | 33          | 46       | 48       | 250       |                |
| Producer’s Accuracy | 90.00 | 96.77 | 90.48 | 84.85 | 82.61 | 93.75 | 89.60 | Overall Accuracy |

Source: Calculated and Compiled by the Scholar from Landsat images of 1990, 2000 and 2015

For the year 1990 (Table 4), The producer’s accuracy, 1990, for water body, vegetation, crop land, fallow land, riverbed, built up is 90.00 percent, 96.77 percent, 90.48 percent, 84.85 percent, 82.61 percent, 93.75 percent respectively. The user’s accuracy for water body, vegetation, and crop land and fallow land, riverbed, built up is 90.00 percent, 83.33 percent, 95.00 percent, 77.78 percent, 92.68 percent, and 95.74 percent respectively. For the year 2000 (Table 5), the producer’s accuracy, 2000, for water body, vegetation, crop land, fallow land, riverbed, built up is 90.00 percent, 93.75 percent, 86.36 percent, 83.78 percent, 80.85, 86.54 percent respectively. The user’s accuracy for water body, vegetation, and crop land, and fallow land, riverbed, built up is 90.00 percent, 81.08 percent, 86.36 percent, 77.50 percent, 88.37 percent, and 93.75 percent respectively.

For the year 2015 (Table 6), The producer’s accuracy, 2015, for water body, vegetation, crop land, fallow land, riverbed, built up is 89.58 percent, 97.22 percent, 90.48 percent, 87.50 percent, 84.00 percent, 93.75 percent respectively. The user’s accuracy for water body,
vegetation, and crop land, and fallow land, riverbed, built up is 89.58 percent, 85.37 percent, 95.00 percent, 81.40 percent, 93.33 percent, and 95.74 percent respectively. The classified images of 1990, 2000 and 2015 have an overall accuracy of 89.6 percent, 86.64 percent, 90.15 percent respectively. On the other hand, the Kappa statistics of the classified images of 1990, 2000, 2015 were calculated as 0.89, 0.86 and 0.91 respectively (Table 7).

### Table 05: Accuracy Assessment Land Use and Land Cover Classification, 2000

| LULC Classes | Water body | Vegetation | Crop land | Fallow land | Riverbed | Built up | Row Total | User’s Accuracy |
|--------------|------------|------------|-----------|-------------|----------|----------|-----------|----------------|
| Water body   | 45         | 0          | 1         | 2           | 1        | 1        | 1         | 50             | 90.00          |
| Vegetation   | 1          | 30         | 2         | 1           | 2        | 1        | 1         | 37             | 81.08          |
| Cropland     | 2          | 1          | 38        | 1           | 1        | 1        | 1         | 44             | 86.36          |
| Fallow land  | 1          | 1          | 1         | 31          | 4        | 2        | 2         | 40             | 77.50          |
| Riverbed     | 1          | 0          | 1         | 1           | 38       | 2        | 3         | 43             | 88.37          |
| Built-up     | 0          | 0          | 1         | 1           | 1        | 45       | 48        | 93.75          |
| Column Total | 50         | 32         | 44        | 37          | 47       | 52       | 262       |                |
| Producer’s Accuracy | 90.00 | 93.75 | 86.36 | 83.78 | 80.85 | 86.54 | 86.64 | Overall Accuracy |

Source: Calculated and Compiled by the Scholar from Landsat images of 1990, 2000 and 2015

### Table 06: Accuracy Assessment Land Use and Land Cover Classification, 2015

| LULC Classes | Water body | Vegetation | Crop land | Fallow land | Riverbed | Built up | Row Total | User’s Accuracy |
|--------------|------------|------------|-----------|-------------|----------|----------|-----------|----------------|
| Water body   | 43         | 0          | 1         | 2           | 1        | 1        | 48        | 89.58          |
| Vegetation   | 1          | 35         | 2         | 1           | 0        | 0        | 41        | 85.37          |
| Cropland     | 2          | 0          | 38        | 0           | 2        | 0        | 40        | 95.00          |
| Fallow land  | 1          | 1          | 1         | 35          | 4        | 2        | 43        | 81.40          |
| Riverbed     | 1          | 0          | 1         | 42          | 0        | 45       | 47        | 95.74          |
| Built Up     | 0          | 0          | 0         | 1           | 1        | 45       | 47        | 95.74          |
| Column Total | 48         | 36         | 42        | 40          | 50       | 48       | 264       |                |
| Producer’s Accuracy | 89.58 | 97.22 | 90.48 | 87.50 | 84.00 | 93.75 | 90.15 | Overall Accuracy |

Source: Calculated and Compiled by the Scholar from Landsat images of 1990, 2000 and 2015

### Table 07: Summary of Accuracy (percent) and Kappa Statistics

| LULC Classes | 1990 Accuracy | 2000 Accuracy | 2015 Accuracy |
|--------------|---------------|---------------|---------------|
|               | Producer’s    | User’s        | Producer’s    | User’s        | Producer’s | User’s |
| Water body   | 90.00         | 90.00         | 90.00         | 90.00         | 89.58      | 89.58 |
| Vegetation   | 96.77         | 83.33         | 93.75         | 81.08         | 97.22      | 85.37 |
| Cropland     | 90.48         | 95.00         | 86.36         | 86.36         | 90.48      | 95.00 |
| Fallow land  | 84.85         | 77.78         | 83.85         | 77.50         | 87.50      | 81.40 |
| Riverbed     | 82.61         | 92.68         | 80.85         | 88.37         | 84.00      | 93.33 |
| Built-Up     | 93.75         | 95.74         | 86.54         | 93.75         | 93.75      | 95.74 |
| Overall Accuracy | 89.60 | 86.64 | 90.15 | |
| Kappa Statistics | 0.89  | 0.86  | 0.91  | |

Source: Calculated and Compiled by the Scholar from the Landsat images of 1990, 2000 and 2015

### Conclusion

Accuracy assessment or validation justifies the ‘goodness factor’ of maps and facilitate the descriptive and analytical evaluation of the spatial data. Irrespective of user, whether novice or expert, accuracy assessment need to be made necessary and sufficient part while carrying out interpretation of spatial data.
References
1. Bishop, L.R., et.al. (1975). The Quantitative Measurement of Foam Stability, Journal of Institute of Brewing.
2. Cohen J (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37–46.
3. Congalton, R.G. (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. Remote Sensing of Environment, 37, 35-46. https://doi.org/10.1016/0034-4257(91)90048-B
4. Goodchild M, Gopal S (Eds) (1989). The accuracy of spatial databases. (Taylor and Francis: New York) 290 pp.
5. Hudson W, Ramm C (1987). Correct formulation of the kappa coefficient of agreement. Photogrammetric Engineering and Remote Sensing. 53, 421–422.
6. Janssen L, van der Well F (1994). Accuracy assessment of satellite derived land-cover data: A review. Photogrammetric Engineering and Remote Sensing 60, 419–426.
7. NRSA (2006). Manual of National wide Land Use/Land cover Mapping using multi-temporal Satellite Data, National Remote Sensing Agency, Hyderabad.
8. Stehman, S. V. (1995). Thematic map accuracy assessment from the perspective of finite population sampling. International Journal of Remote Sensing, 16, 589 – 593.