5. Results and analysis

5.1 General

This chapter shows the results obtained for the individual objectives, based on the results obtained the analysis has been carried out. The chapter is divided into six sections, each section contains the results obtained for the specific objective.

5.2 Results and analysis

1. **Objective:** To demonstrate a comparative assessment of discrepancy in the hydrological behaviour of the DEMs in terms of terrain representation at the catchment scale.

   i. Visual Comparison:

   Streams comparison

   When the delineated streams are overlaid over high-resolution imagery, the Cartosat-generated network is much closer to the actual river network followed by the SRTM-derived drainage network as shown in Figure 5.1. It has been observed that the drainage network delineated by ASTER is highly misleading. Moreover, sinks around the actual river have considerably contributed to the deviation of ASTER DEM- and SRTM DEM-derived streams.

   Watershed comparison

   The area enclosed by the watershed generated by SRTM and ASTER is comparatively much larger than that generated by Cartosat as shown in Figures 5.2 and 5.3. The area of the watershed delineated by Cartosat is 1289.4 km^2, whereas the area of the watershed delineated by ASTER is 1624.8 km^2 (26% larger). Moreover, the SRTM-based watershed area is 2026.3 km^2, which is 56% larger than the Cartosat boundary. The perimeter of the watershed delineated by Cartosat is 279.44 km, whereas the perimeter of the watershed delineated by ASTER is 315.4 km (50.26% larger). Moreover, the SRTM-based watershed perimeter is 294.9 km, which is 40.5% larger than the Cartosat watershed perimeter.

   Ridge line inspection over relief map

   The relief map used in the Cartographic relief depiction shows the shape of the terrain in a realistic fashion and also demonstrates the three-dimensional surface that is illuminated from a point light source. Moreover, the watersheds overlaid over the relief map and satellite imagery show that the watersheds delineated by ASTER and SRTM could not follow the ridgeline and hence they have encompassed the Dhadhar river in them. As shown in Figure 5.4, the highlighted yellow circles show the locations where the ASTER watershed and the SRTM watershed encompass the Dhadhar river. Clearly, it can also be observed that the Cartosat-derived boundary follows the actual ridgeline. In the flow-direction process, a depressionless

DEM is considered to be the desired input. An erroneous flow-direction raster may be resulted in the presence of sinks. Moreover, there may be legitimate sinks in the data in some of the cases. By taking into consideration the flow networks associated with each type of elevation data, the cause of the difference in the watershed boundaries can be found. It has been observed that the flow networks generated from the ASTER- and SRTM-based DEM had several errors. Map algebra was used to find where the Fill tool had filled the sinks in order to determine the cause of the errors in the streams network. As shown in Figure 5.5 (a)–(c), it was found that the errors in the stream network occurred where filling of sinks along the actual river have had occurred. The deviation of ASTER DEM- and SRTM DEM-derived streams from the actual stream show a large number of sinks around the actual river have considerably contributed to the deviation of the generated stream network. Such error indicates that there were probably residual and artifactual anomalies that most certainly degraded the overall accuracy of ASTER and SRTM DEMs. As a result of underestimating the elevation at certain points, pits, and depressions are considered false in the Fill method as mentioned above. Therefore, the depressions are filled, and thus raising the elevation until it reaches the lower neighbour. As a result, the larger the number of continuously affected pixels, the more the result of the flow-direction assignment is affected. Figure 5.5 (a)–(c) shows that ASTER data contain a large number of depressions or pits followed by the SRTM data, whereas the Cartosat data contain the least amount of depressions or pits.

Slope Gradients Classes

Moreover, six classes of slope were established for a better understanding of terrain as shown in Table 5.1. As shown in Figure 5.6 and Table 5.1, the slope values were classified according to the Brazilian Agricultural Research Corporation standards. According to the SRTM and Cartosat, the result showed that the maximum area in watershed belongs to flat relief with a declivity value (in %) between 0 and 2.99. According to the ASTER, however, the maximum area belongs to smooth relief with a declivity value (in %) between 3 and 7.99.

Table 5.1: Slope classification as per Brazilian agricultural research corporation standards.

Declivity (%)	Relief classes	ASTER AREA(Km^2)	Cartosat AREA(Km^2)	SRTM AREA(Km^2)
0 – 2.99	Flat	252.58	687.16	638.32
3 –7.99	Smooth	677.55	481.20	600.01
8 – 19.99	Corrugated	335.48	89.73	38.94
20 – 44.99	Heavily Corrugated	16.07	20.05	5.15
45 – 74.99	Mountainous	2.18	4.78	1.56
< 75	Steepest	1.39	2.34	1.26

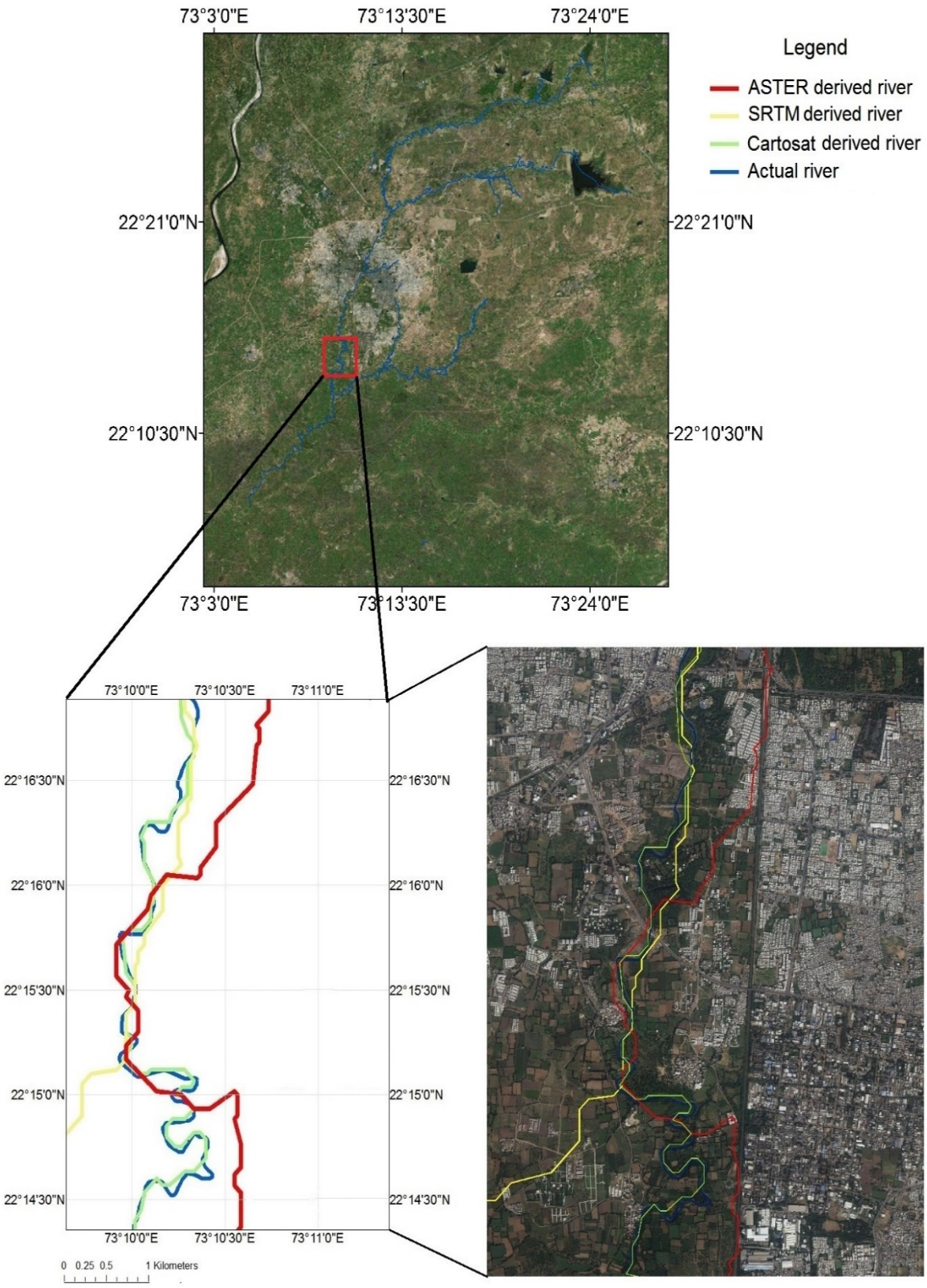


Figure 5.1: Map showing ASTER, SRTM, and Cartosat derived river deviation from the actual river.

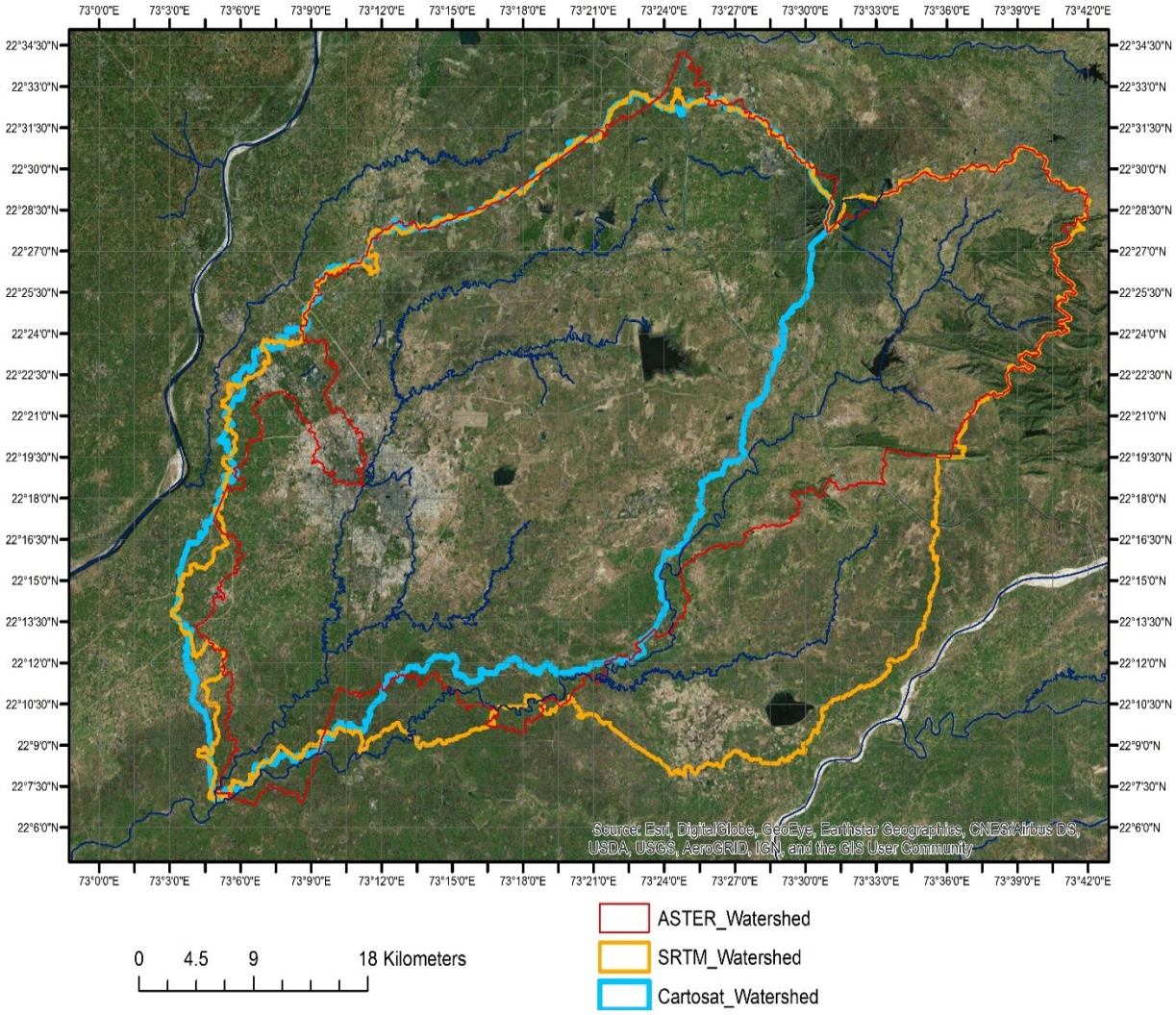


Figure 5.2: Delineated watershed boundaries from ASTER, SRTM and Cartosat data.

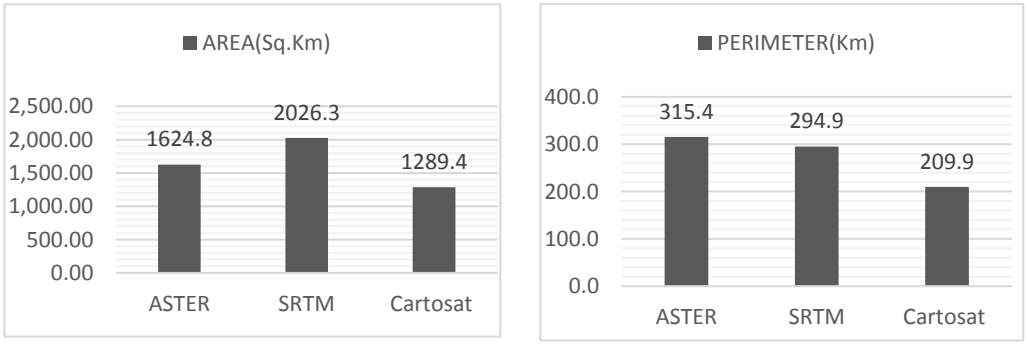


Figure 5.3: Area and perimeter of delineated watersheds.

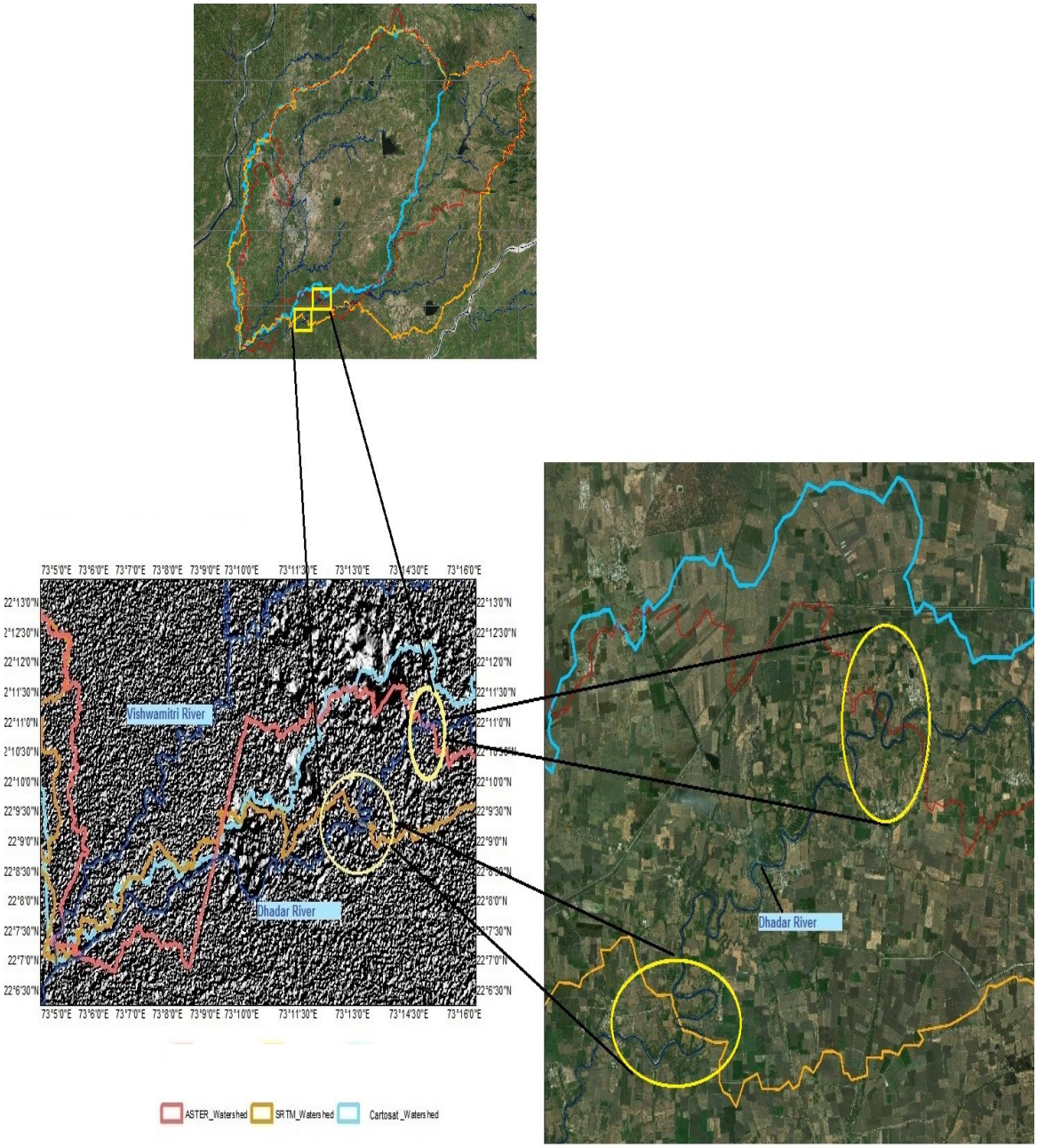


Figure 5.4: Visual Inspection of Watersheds derived from ASTER, SRTM, and Cartosat over the Shaded Relief map and Satellite imagery.

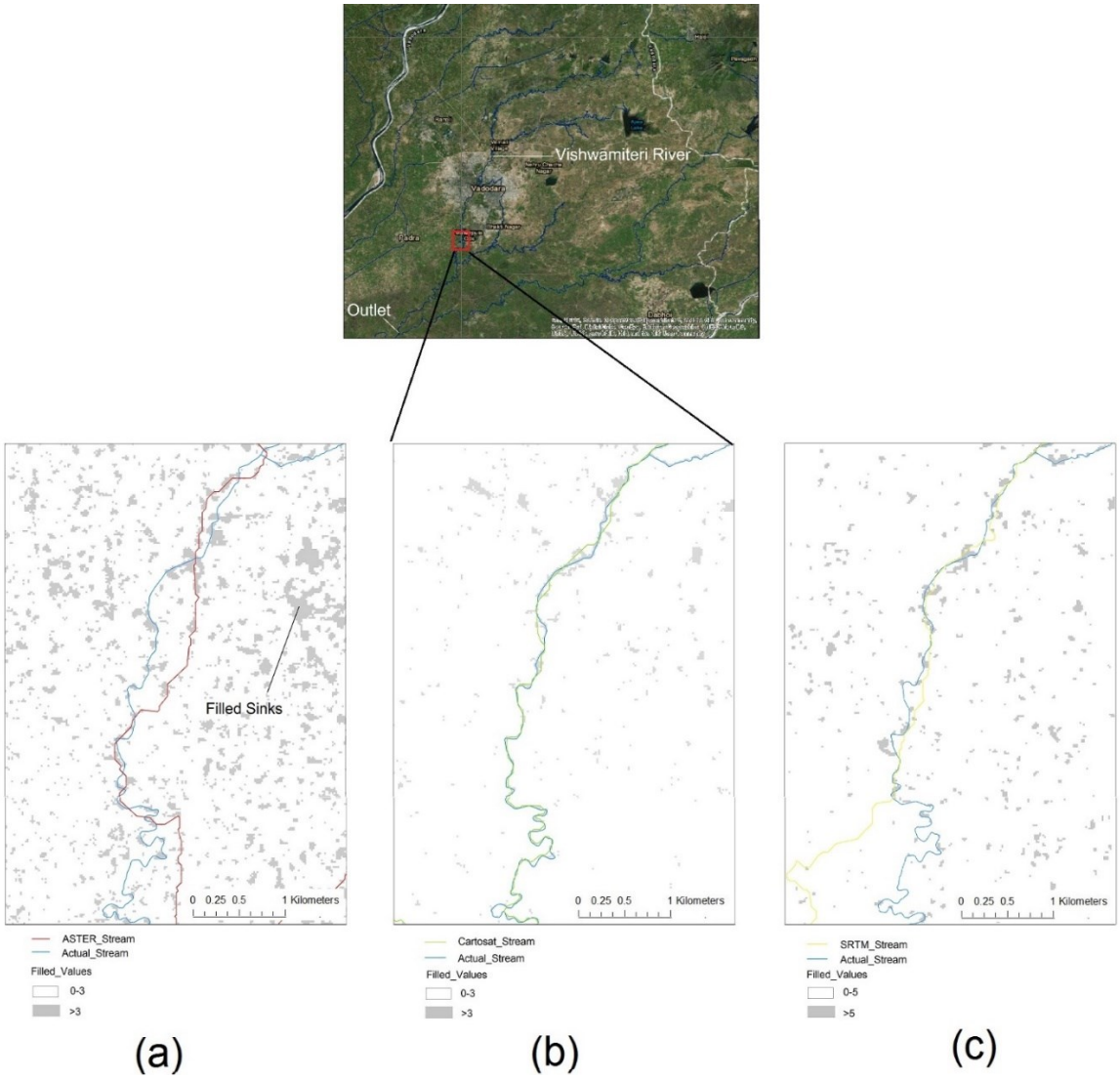


Figure 5.5: (a) Filled sinks in ASTER (b) Filled sinks in Cartosat (c) Filled sinks in SRTM.

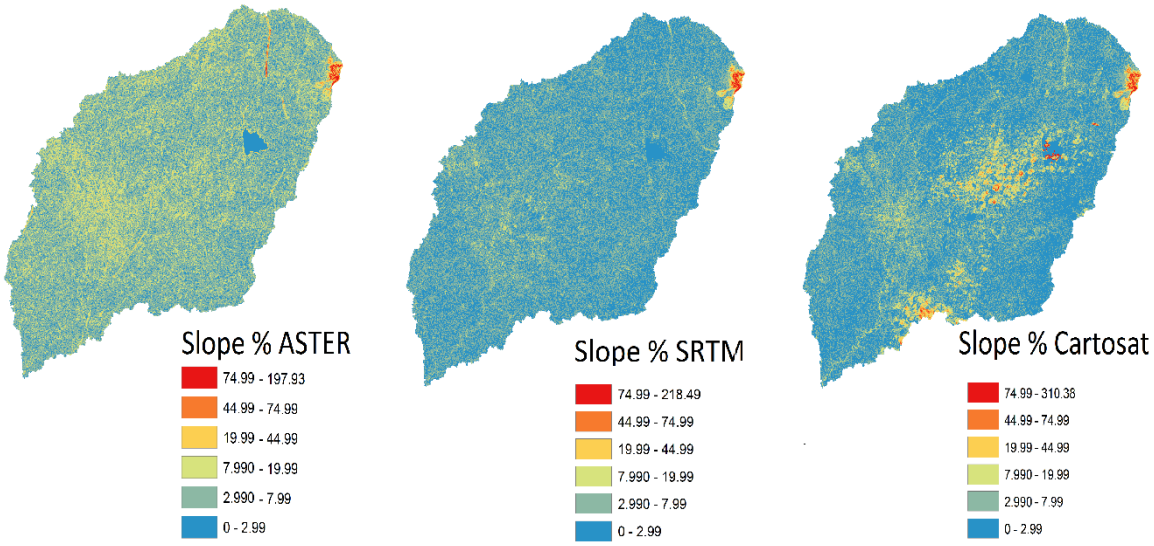


Figure 5.6: Slope Classes of ASTER, SRTM and Cartosat.

ii. Statistical Comparison:

In order to provide evidence of the statistical significance of the results, the amount of data (274476 pixels) from all three DEMs is used. The accuracy and quality of the reference data should be at least one order better as compared to the data to be evaluated. Based on the above observations of the DEMs, Cartosat DEM exhibited better results, hence taken as reference data for further analysis. The absolute difference between the mean value of SRTM and Cartosat is 1.55 m, and of ASTER and Cartosat is found to be 5.38 m as shown in Table 5.2. The standard deviation showed that the Cartosat dataset was less spread out as compared to the other two datasets, which was also confirmed by the difference between the upper and lower quartiles (interquartile range). For the samples, the positive value of skewness showed that the distributions of the data were positively skewed or skewed right, i.e., the right tail of the distribution was longer as compared to the left tail of the distribution. For ASTER, the value of kurtosis was 0.02, which showed that the distribution is leptokurtic, i.e., its tails are longer and fatter. For SRTM and Cartosat, the negative value of kurtosis showed that the distribution is platykurtic, i.e., its tails are shorter and thinner.

Table 5.2: Statistics for ASTER, SRTM, and Cartosat derived elevation.

Statistic		ASTER	SRTM	Cartosat
Mean (meters)		45.82	52.75	51.2
95% Confidence Interval for Mean	Lower Bound (meters)	45.78	52.72	51.18
	Upper Bound (meters)	45.85	52.78	51.23
5% Trimmed Mean (meters)		45.52	52.39	50.87
Median (meters)		45	51	50
Variance (meters)		73.78	64.12	47.5
Std. Deviation (meters)		8.59	8.01	6.89
Minimum (meters)		13	34	18
Maximum (meters)		83	81	90
Range (meters)		70	47 m	72
Interquartile Range (meters)		12	12	10
Skewness (Dimensionless)		0.52	0.64	0.64
Kurtosis (Dimensionless)		0.02	-0.24	-0.07

In addition, the normality test is conducted over the datasets. The Kolmogorov–Smirnov (K–S) sample test is a nonparametric test with the null hypothesis, which demonstrates that the data were obtained from a normal distribution as shown in Table 5.3. The criteria used to reject or accept the null hypothesis is that if the *p*-value is smaller than the significance level $\alpha = 0.05$, the null hypothesis is rejected. Moreover, there are sufficient pieces of evidence that the data are not normally distributed. For instance, the null hypothesis was rejected (the significance level is 0.05) in all the above three cases. The *p*-value (0.00) of less than 0.05 showed that there

is not enough evidence to prove that the data are normal. The histograms of the ASTER, SRTM, and Cartosat sample data are shown in Figure 5.7 (a) to (c), respectively. The normal quantile-quantile (Q–Q) plot and the detrended normal Q–Q plot are also drawn to support or refute the claim of normality. The quantile-quantile plot is shown in Figure 5.7 (d)–(f), which compares the observed quantiles of the data with those of the normally distributed data. The observed quantiles of the data are depicted as circles, whereas the quantiles of data that we would expect to see if the data were normally distributed are depicted as a solid line. The data are approximately normally distributed, if the points are on or close to the line. Similarly, the sample data are not normally distributed if the points are not clustered on the 45° line or they, in fact, follow a curve. Moreover, the detrended normal Q–Q plot Figure 5.7 (g)–(i) provides the same information as the normal Q–Q plot, but in a different way. In the detrended plot, the horizontal line at the origin represents the quantiles if the data were normally distributed, whereas the dots represent the magnitude and direction of deviation in the observed quantiles. Each dot is calculated by subtracting the expected quantile from the observed quantile. The related-sample Wilcoxon signed rank test is conducted (significance level 0.05) by using the null hypothesis: the median of differences between Cartosat and SRTM equals 0 and the decision of the test is to retain the null hypothesis as the p-value (1.00) was more than 0.05. Likewise, the same test was also conducted by using the null hypothesis: the median of differences between Cartosat and ASTER equals to 0 and the decision of the test is to reject the null hypothesis as the p-value (0.00) was less than 0.05. The results also showed that 9.12% and 65.65% data in ASTER and SRTM, respectively, have more elevation than the Cartosat data, the 86.69% and 23.34% data in ASTER and SRTM, respectively, have less elevation than the Cartosat data, and 4.19% and 11.01% data in ASTER and SRTM, respectively, have the same elevation as the Cartosat data. Figures 5.8 (a) and (b) and Table 5.4 demonstrate the corresponding results.

In order to assess the level of correlation between the DEMs, the correlation scatter-plots were drawn as shown in Figure 5.9 (a) and (b). It was difficult to create a scatter plot from each pixel in a DEM as each DEM contains over a million pixels. However, a total of 274,476 pixels were used for the analysis. As shown in Table 5.5, the correlation coefficient of 0.83, 0.94, and 0.85 was obtained by Pearson's correlation analysis between the ASTER and Cartosat, SRTM and Cartosat, and ASTER and SRTM, respectively (correlation is found to be significant at the level of 0.01). For instance, the correlation value of 0.94 indicates a strong positive linear correlation between SRTM and Cartosat. Similarly, the simple linear regression analysis is demonstrated by means of scatter-plots. In this case, the analysis of the determination coefficient (R^2) of the regression line shows that the Cartosat DEM is considered adequate for describing the ASTER

DEM by 68.9% and the SRTM DEM by 87.9%. Considering Cartosat as the reference DEM, the RMSE calculated was used for evaluating the vertical accuracy of the ASTER DEM and the SRTM DEM. For ASTER the calculated RMSE is found to be 7.21 m, whereas for SRTM it is 3.24 m. Error maps for ASTER and SRTM were produced by subtracting their elevation values from the respective Cartosat values as shown in Figure 5.10 (a) and (c). For the ASTER error map, the mean value and standard deviation were found to be 5.39 m and 4.79 m, respectively. The corresponding frequency histograms indicate that the ASTER elevation data were highly underestimated as shown in Figure 5.10 (b). For the SRTM error map, the mean value and standard deviation were found to be –1.55 m and 2.85 m, respectively. The corresponding frequency histograms indicate that the SRTM elevation data were slightly overestimated as shown in Figure 5.10 (d).

Table 5.3: Normality test results.

Null Hypothesis	Test	Sig.	Decision
The distribution of ASTER elevation is normal with mean 45.82 and standard deviation of 8.590.	One-Sample Kolmogorov-Smirnov Test	.000[1]	Reject the null hypothesis.
The distribution of SRTM elevation is normal with mean 52.75 and standard deviation of 8.008.	One-Sample Kolmogorov-Smirnov Test	.000[1]	Reject the null hypothesis.
The distribution of Cartosat elevation is normal with mean 51 and standard deviation of 6.892.	One-Sample Kolmogorov-Smirnov Test	.000[1]	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Table 5.4: Wilcoxon signed ranks test assuming the median of differences between Cartosat and ASTER, and Cartosat and SRTM equals 0.

Wilcoxon Signed Ranks Test	Cartosat - ASTER	Cartosat - SRTM
Negative Ranks	25033[a]	180203[d]
Positive Ranks	237924[b]	64039[e]
Ties	11519[c]	30234[f]
Total	274476	274476

a. Cartosat< ASTER
b. Cartosat > ASTER
c. Cartosat = ASTER
d. Cartosat < SRTM
e. Cartosat > SRTM
f. Cartosat = SRTM

Table 5.5: Correlation coefficients for ASTER, SRTM, and Cartosat derived elevation.

	ASTER	*SRTM*	*Cartosat*
ASTER	1.00		
SRTM	0.85	1.00	
Cartosat	0.83	0.94	1.00

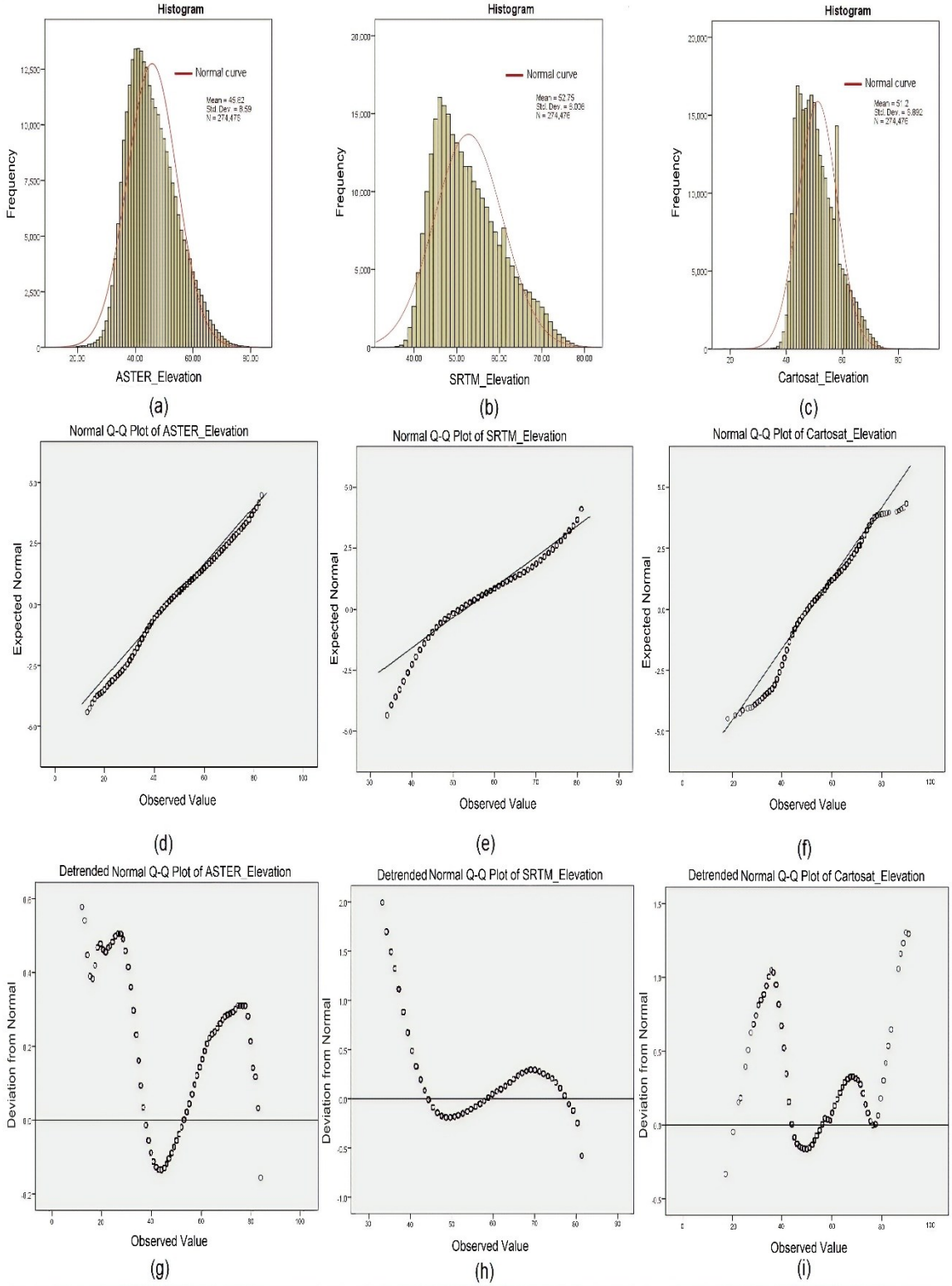


Figure 5.7: (a) Histogram of ASTER data (b) Histogram of SRTM data (c) histogram of Cartosat data (d) Quantile-Quantile Plot of ASTER data (e) Quantile-Quantile Plot of SRTM data (f) Quantile-Quantile Plot of Cartosat data (g) Detrended Normal Q-Q Plot of ASTER data (h) Detrended Normal Q-Q Plot of SRTM data (I) Detrended Normal Q-Q Plot of Cartosat data.

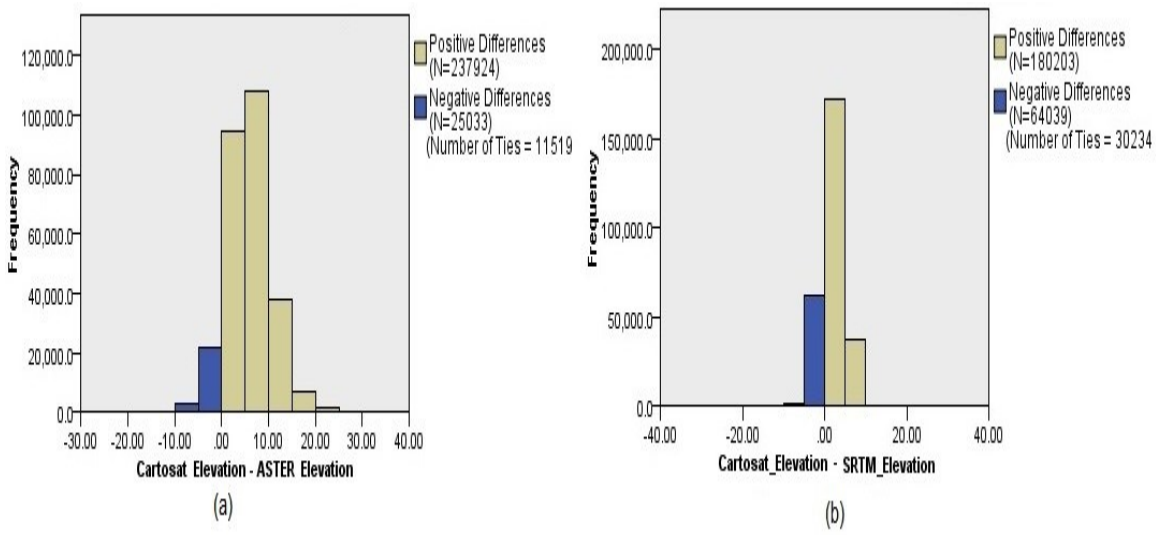


Figure 5.8: (a) Related-Samples Wilcoxon Signed Rank Test for ASTER (b) Related-Samples Wilcoxon Signed Rank Test for SRTM.

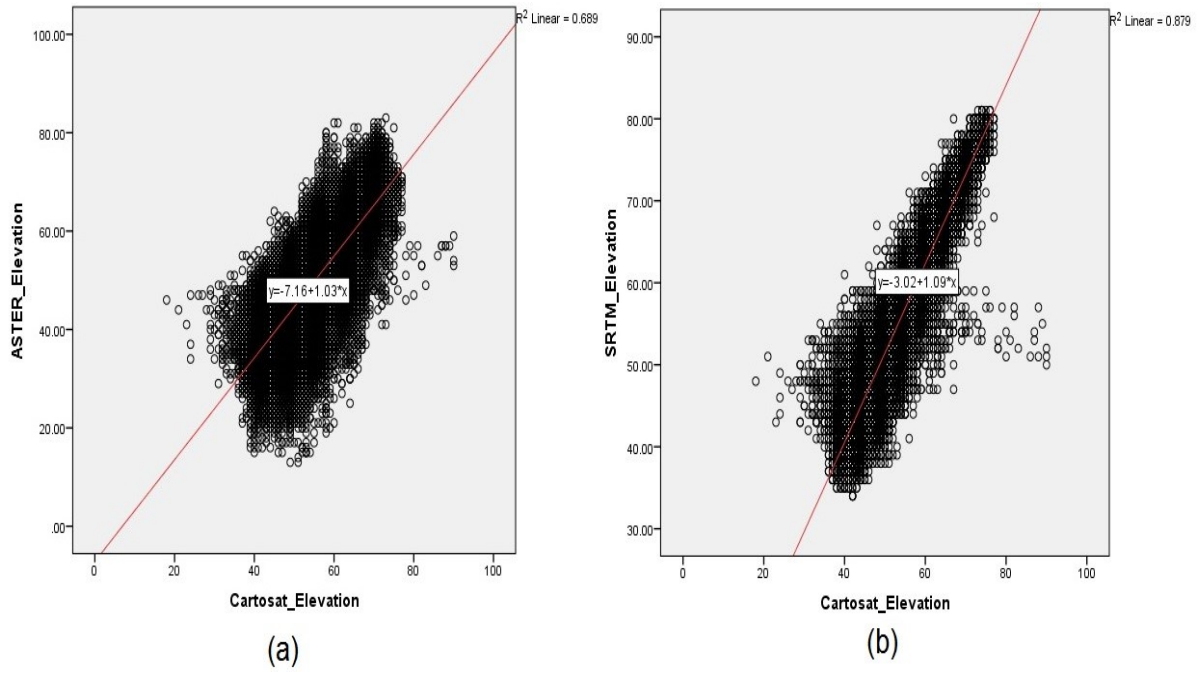
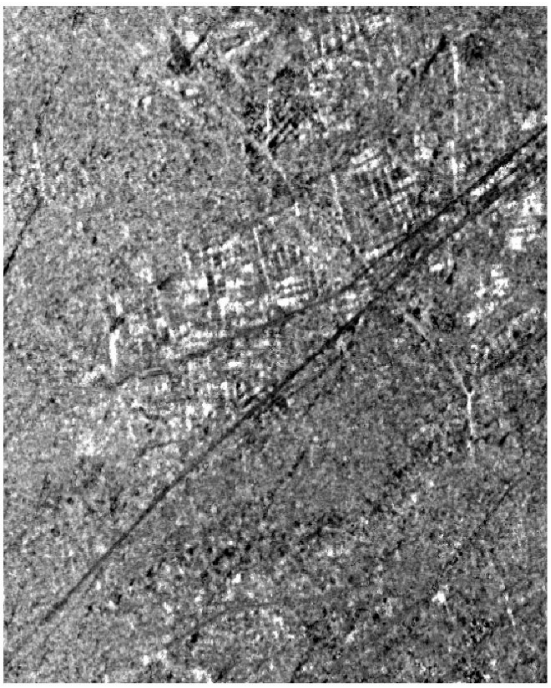
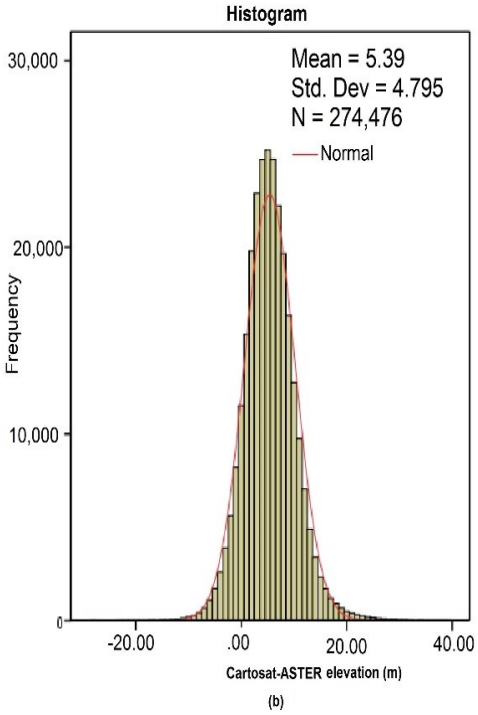


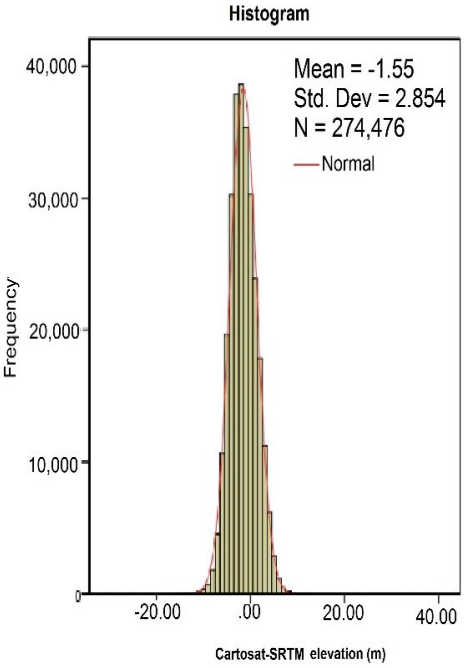
Figure 5.9: (a) Scatterplot showing linear regression of ASTER derived elevation vs Cartosat derived elevation (b) Scatterplot showing linear regression of SRTM derived elevation vs Cartosat derived elevation.





(b)

(a)



(d)

(c)

Figure 5.10: a) Error map for ASTER (b) Histogram of error map of ASTER (c) Error map for SRTM (d) Histogram of error map of SRTM.

2. **Objective:** To develop an approach to analyze Sentinel–2 satellite data using traditional and principal component analysis based approaches to create land use and land cover map, which is a prerequisite for developing the curve number.

   Principal component analysis was used for data compression of Sentinel-2 multispectral data to statistically maximise the amount of information from the original data (bands) to a smaller number of components, called principal components. The first few principal components possess most of the variability of the data. The first principal component band derived from the first eigenvector had the maximum amount of the total variance of the Sentinel–2 data set containing bands 2–8, 8A, 11 and 12. The first three principal component bands accounted for 97.92% of eigenvalues. The variances percentage of the remaining principal component bands decreased in order of the corresponding eigenvalues. The variance of the PCA bands 4 to 10 is small and mostly had noises; thus the bands were removed from the analysis. The eigenvalues and accumulative percentage of eigenvalues in PCA bands 2–8, 8A, 11 and 12 are given in Table 5.7.

   A loading plot shows how strongly each band influences a principal component. It ranges from -1 to 1. Loadings close to -1 or 1 indicate that the band strongly influences the component. Loadings close to 0 indicate that the band has a weak influence on the component. Figure 5.11 shows the loading plot of the bands. It can be seen that bands 6-8A have large positive loadings on component 1. While, bands 2-5 and 11-12 have large positive loadings on component 2. Factor loading shows how much each band has contributed to the factor. Rescaled factor loadings of the principal components from the original (raw) image data are shown in Table 5.8 using component matrix. Bands 6-8A (Vegetation Red Edge and near infrared bands are highly correlated, Table 5.6) loaded highly in the first principal component, bands 2-5 and 11-12 (highly correlated), loaded highly (positively) in principal component 2 and Band 2-5 marginally loaded onto principal component 3.

   Principal component 4-10 can be termed as noise components since no factor loading is prominent. It has been observed that deciduous plants have a sharp order-of-magnitude increase in leaf reflectance between approximately 700 to 750 nm wavelength and healthy vegetation reflects highly in near infrared band. Principal component 1 can be called as healthy vegetation component as it has the highest factor loading of 0.99 from band 7 and band 8A. However, principal component 2 and principal component 3 can't be generalised as loading is scattered across the spectrum. PCA transformed the correlated Sentinel–2 dataset into a substantially smaller set of uncorrelated variables representing most of the information present in the original dataset. Figure 5.12 (a–c) shows the PCA bands derived from the Sentinel–2 data

and Figure 5.12 (d–f) shows the frequency distribution of corresponding principal component bands. The total range (maximum value - minimum value) of PCA band 3 is greater than the PCA band 1 and PCA band 2. However, most of the pixels fall in a small range around the mean of 6486.13, which shows information loss in PCA band 3.

The frequency distribution reveal that the variance of the first principal component is the highest, followed by the second and then by the third. The calculated values of variance for PCA bands 1, 2 and 3 were 395879.55, 284229.30, and 45254.05, respectively. The image produced from PCA band 1 data resembles original image and it contains most of the pertinent information inherent to the scene due to high variance.

Adjacent bands in a multispectral remotely sensed image are often highly correlated and often convey almost the same information about the object. A high correlation meaning thereby that the bands are not statistically independent. A low degree of correlation was observed among the PCA bands 1, 2, and 3 (Figure 5.13 (a–c)). The correlation values between PCA bands 1 and 2, 1 and 3, and 3 and 2 were 0.01, –0.01, and 0.25, respectively. Non-structured appearance of the scatter plots and low correlation values confirm that there is no relationship among the PCA bands. High Correlation in stack 1 shows (Table 5.6) that there is redundancy of information and if this redundancy can be reduced, then information can be compressed. The correlation between the bands which exist in the original data has disappeared in the principal components. So, the PCA was able to reduce correlation significantly. The components are new uncorrelated bands created by linearly combining original data, keeping as much information as possible from the original data. In this way, a classification using the first few principal components can be better than one performed by means of the original dataset. If we compare the PCA band 1, PCA band 2 and PCA band 3 images, PCA band 1 has higher brightness values for the pixels than PCA band 2 and PCA band 3, and PCA band 1 shows a higher degree of contrast than PCA band 2 and PCA band 3.

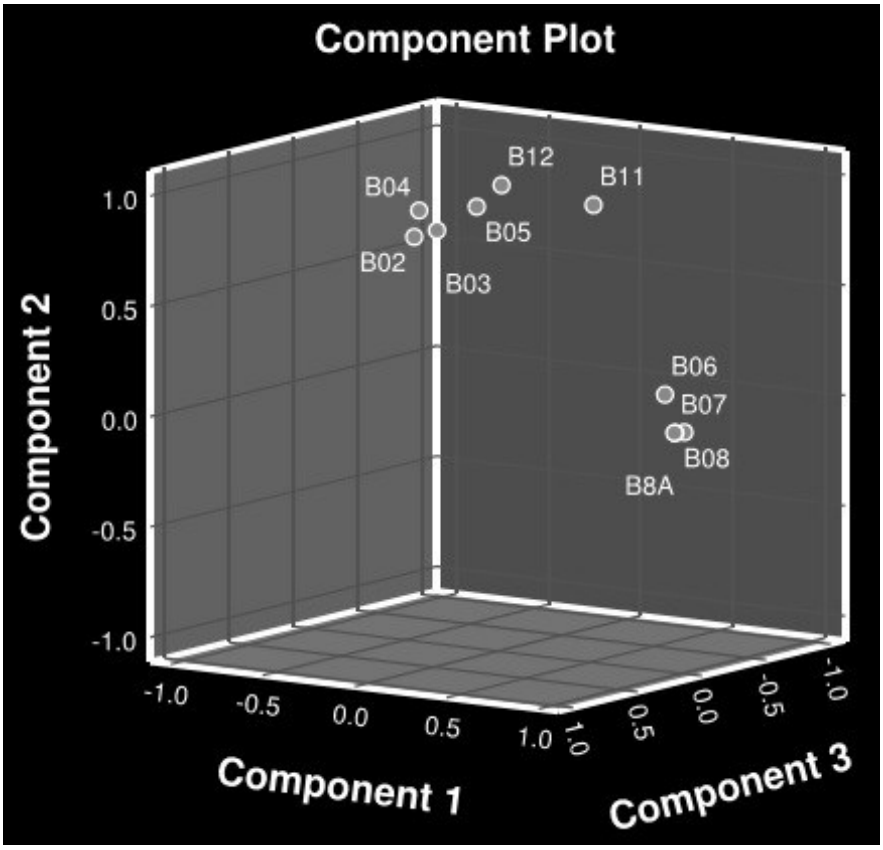


Figure 5.11: Loading plot of PCA component 1, 2 and 3.

Table 5.6: Covariance (Correlation) matrix of sentinel–2 bands.

Sentinel -2 Bands	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	Band 8	Band 8A	Band 11	Band 12
Band 2	40426.8 (1.00)	43694.4 (0.98)	63465.1 (0.95)	45236.1 (0.88)	6071.1 (0.08)	-8485.5 (-0.08)	-9932.3 (-0.10)	-14377.9 (-0.12)	42030.1 (0.53)	55946.5 (0.68)
Band 3	43694.4 (0.98)	49362.7 (1.00)	70250.5 (0.95)	52331.9 (0.92)	15490.6 (0.19)	1389.2 (0.01)	360.7 (0.00)	-3597.5 (-0.03)	52797.9 (0.60)	63910.7 (0.71)
Band 4	63465.1 (0.95)	70250.5 (0.95)	111007.2 (1.00)	78097.1 (0.92)	-1973.4 (-0.01)	-32521.7 (-0.19)	-35057.0 (-0.20)	-42296.8 (-0.22)	84124.7 (0.64)	111431.2 (0.82)
Band 5	45236.1 (0.88)	52331.9 (0.92)	78097.1 (0.92)	65216.5 (1.00)	23153.8 (0.24)	6455.0 (0.05)	3933.8 (0.03)	3102.1 (0.02)	77195.9 (0.76)	86156.7 (0.83)
Band 6	6071.1 (0.08)	15490.6 (0.19)	-1973.4 (-0.01)	23153.8 (0.24)	141140.0 (1.00)	189019.7 (0.97)	179641.9 (0.93)	209046.3 (0.96)	55332.7 (0.37)	-1619.9 (-0.01)
Band 7	-8485.5 (-0.08)	1389.2 (0.01)	-32521.7 (-0.19)	6455.0 (0.05)	189019.7 (0.97)	270417.6 (1.00)	256318.7 (0.96)	298626.5 (0.99)	47352.7 (0.23)	-34947.6 (-0.16)
Band 8	-9932.3 (-0.09)	360.7 (0.01)	-35057.0 (-0.20)	3933.8 (-0.20)	179641.9 (0.93)	256318.7 (0.96)	265932.9 (1.00)	286099.6 (0.96)	47354.6 (0.23)	-33365.4 (-0.16)
Band 8A	-14377.9 (-	--3597.5 (-0.03)	-42296.8 (-0.22)	3102.1 (-0.22)	209046.3 (0.96)	298626.5 (0.99)	286099.6 (0.96)	336385.7 (1.00)	58667.2 (0.26)	-37889.3 (-0.16)
Band 11	42030.1 (0.53)	52797.9 (0.60)	84124.7 (0.64)	77195.9 (0.64)	55332.7 (0.37)	47352.7 (0.23)	47354.6 (0.23)	58667.2 (0.25)	157985.6 (1.00)	142013.0 (0.88)
Band 12	55946.5 (0.69)	63910.7 (0.71)	111431.2 (0.83)	86156.7 (0.83)	-1619.9 (-0.01)	-34947.6 (-0.16)	-33365.4 (-0.16)	-37889.3 (-0.16)	142013.0 (0.88)	166130.1 (1.00)

Table 5.7: Total Variance Explained.

	Extraction Sums of Squared Loadings		
	Raw		
Number of Principal Component Bands	Eigen Value	Percent of Eigen Values (%)	Accumulative (%) of Eigen Values
1	1005312.99	62.68	62.68
2	494281.26	30.82	93.49
3	70972.23	4.42	97.92
4	16477.60	1.03	98.94
5	6189.42	0.39	99.33
6	5111.35	0.32	99.65
7	1953.55	0.12	99.77
8	1663.12	0.10	99.87
9	1542.72	0.10	99.97
10	500.84	0.03	100.00

Table 5.8: Factor loading component Matrix.

	Component Matrix Raw (Rescaled)									
Sentinel -2 Bands	Principal Components									
	1	2	3	4	5	6	7	8	9	10
Band 2	-18.35 (-.09)	167.20 (.83)	103.98 (.52)	5.49 (.03)	-12.10 (-.06)	17.20 (.09)	18.63 (.09)	18.41 (.09)	0.70 (.00)	12.77 (.06)
Band 3	1.76 (.01)	192.55 (.87)	107.16 (.48)	8.57 (.04)	1.93 (.01)	15.36 (.07)	12.13 (.05)	5.46 (.02)	-0.73 (.00)	-17.60 (-.08)
Band 4	-64.03 (-.19)	304.89 (.92)	111.57 (.33)	-1.53 (.00)	-21.95 (-.07)	10.56 (.03)	-17.89 (-.05)	-23.79 (-.07)	-2.44 (-.01)	3.50 (.01)
Band 5	13.67 (.05)	240.77 (.94)	60.90 (.24)	-9.04 (-.04)	43.60 (.17)	-25.95 (-.10)	-20.07 (-.08)	14.97 (.06)	8.16 (.03)	1.06 (.00)
Band 6	364.50 (.97)	60.92 (.16)	32.36 (.09)	-33.49 (-.09)	35.78 (.10)	-15.52 (-.04)	18.50 (.05)	-14.92 (-.04)	-17.56 (-.05)	2.69 (.01)
Band 7	516.58 (.99)	-7.86 (-.02)	23.99 (.05)	-36.34 (-.07)	-22.32 (-.04)	-18.02 (-.03)	7.19 (.01)	-4.80 (-.01)	26.65 (.05)	-0.84 (.00)
Band 8	503.62 (.98)	-11.11 (-.02)	6.98 (.01)	109.94 (.21)	5.31 (.01)	-1.64 (.00)	-2.38 (.00)	-2.14 (.00)	-0.06 (.00)	1.12 (.00)
Band 8A	576.95 (.99)	-13.39 (-.02)	-17.32 (-.03)	-39.76 (-.07)	-18.38 (-.03)	18.95 (.03)	-17.01 (-.03)	14.46 (.02)	-15.98 (-.03)	-1.44 (.00)
Band 11	105.01 (.26)	351.58 (.88)	-145.95 (-.37)	-6.05 (-.02)	22.02 (.06)	37.34 (.09)	3.38 (.01)	-4.17 (-.01)	10.18 (.03)	1.00 (.00)
Band 12	-58.71 (-.14)	388.12 (.95)	-96.16 (-.24)	11.97 (.03)	-32.04 (-.08)	-38.38 (-.09)	5.98 (.01)	5.22 (.01)	-9.56 (-.02)	-1.55 (.00)

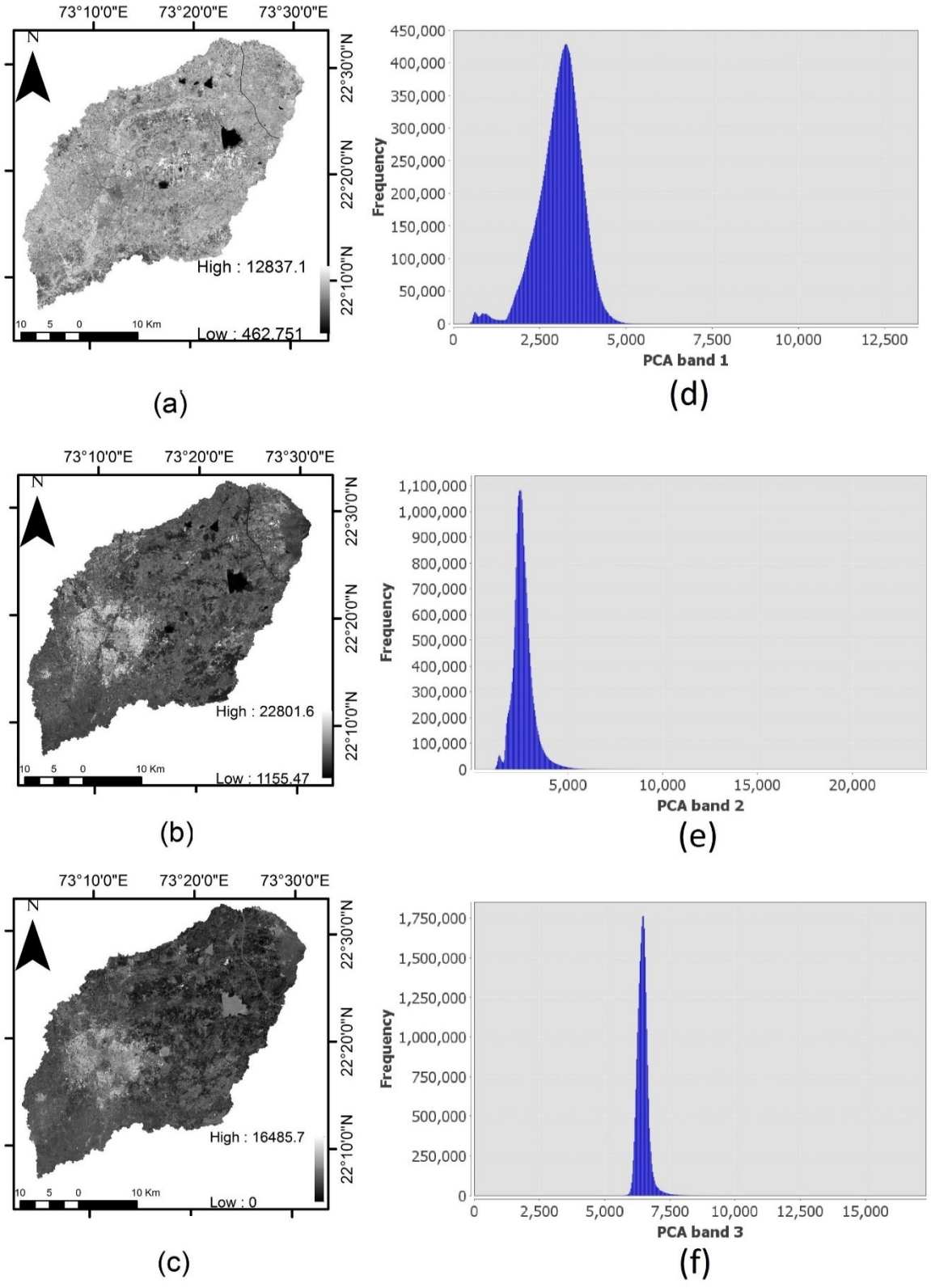


Figure 5.12: Visual comparison of the principal component bands derived from the Sentinel–2 data (a) PCA band 1 (b) PCA band 2 (c) PCA band 3 and (d) Frequency distribution of principal component band 1 (e) Frequency distribution of principal component band 2 (f) Frequency distribution of principal component band 3.

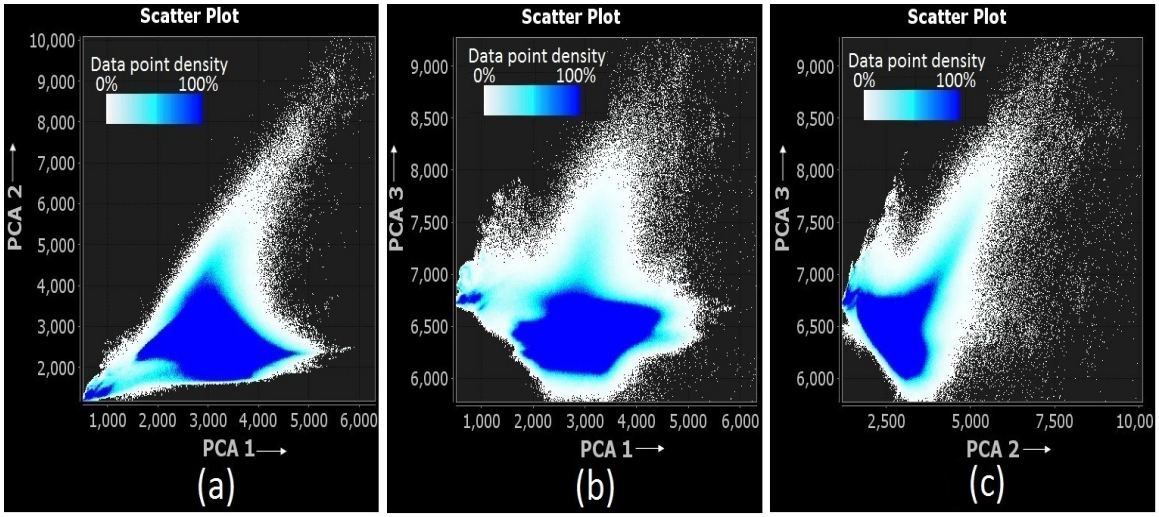


Figure 5.13: Non-structured appearance of the scatter plots leads to the conclusion that there is very low relationship among the PCA bands (a) Scatter plot between PCA band 1 and 2 (b) Scatter plot between PCA band 1 and 3 (c) Scatter plot between PCA band 2 and 3.

Land use and land cover classes were selected based on knowledge about the specific study area. Seven major land use and land cover classes were identified, viz., Water, Built-up, Mixed forest, Cultivated land, Barren land, Fallow land with Vertisols dominance, and Fallow land with Inceptisols dominance. The classifications were conducted using Sentinel–2 original bands (traditional approach) and PCA-based approach. Prediction performances of the three classifier algorithms, MLE, RF, and SVM, were evaluated to reveal the efficiency of two different land use and land cover classification approaches with training data of less than 1000 pixels per class. Training data for each land use and land cover class were collected as a group of pixels. Stratified random sampling was used to obtain the testing data. The classification performed on the original Sentinel–2 bands led to an unacceptable outcome with a classification overall accuracy of 22% for MLE. However, 60% and 64% classification overall accuracy was achieved with RF and SVM classifiers, respectively. The classification results of RF and SVM are acceptable as the training data were limited to less than 1000 pixel per class. In the PCA-based classification approach, the same training polygons were used to avoid the optimistic bias in classification. PCA based approach significantly improved the overall classification accuracy of all the three classifiers. The overall classification accuracy varied considerably among the classifiers. The overall classification accuracy of MLE classifier was increased from 22% to 41% (19% increase) in the PCA based approach. The overall accuracy RF classifier was increased by 10% reaching 70%, whereas SVM classifier outperformed both the classifiers with 76% overall accuracy (increased by 12%). Spectral response curve is the curve showing the variation of reflectance or emittance (in terms of Digital numbers) of a material with respect to wavelengths. Figure 5.14

(b) and 5.15 (b) show the variation of responses of the land use and land cover classes in Stack 1 and Stack 2 respectively. Classes having similar responses are hard to separate. It was also observed that as the spectral distance (or separability) of the classes Water and Built-up in Stack 1 and Stack 2 (Figure 5.14 (b) and 5.15 (b)) in relation to other classes was more, that is why the user's accuracy (UA) and producer's accuracy (PA) for water and Built-up classes are high, for both traditional and PCA-based classification approaches. Kappa coefficient showed a similar trend as that of the overall accuracy. Overall accuracy is calculated by the proportion of the correctly classified pixels to the total number of pixels. Highest Kappa coefficient values were obtained with SVM in PCA based approach followed by the RF. Comparative visual analysis of both the approaches with all the three classifiers are shown in Figure 5.16 (a–f) and 5.17, and the results of accuracy assessment are given in Table 5.9.

Table 5.9: Accuracy assessment of traditional and PCA based approach for LULC classification using MLE, RF, and SVM classifiers.

		Sentinel-2 Bands						PCA Bands					
		MLE		RF		SVM		MLE		RF		SVM	
LULC	Sub classes of LULC	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
Water	Water	0.78	1.00	1.00	1.00	1.00	1.00	0.78	1.00	1.00	1.00	1.00	1.00
Cultivated land	Cultivated land Crop 1	0.00	0.00	0.60	0.60	0.80	0.57	0.60	0.33	0.60	0.60	0.80	0.57
	Cultivated land Crop 2	0.00	0.00	0.80	0.47	0.60	0.55	0.60	0.60	0.90	0.64	0.60	0.67
	Sparsely vegetated	0.00	0.00	0.47	0.50	0.53	0.56	0.41	0.78	0.65	0.69	0.88	0.88
Barren land	Barren land	0.00	0.00	0.50	0.25	0.50	0.27	0.33	0.40	0.50	0.27	0.50	0.43
Fallow land (Vertisols dominance)	Fallow land 1 Vertisols dominance	0.00	0.00	0.60	0.75	0.70	0.64	0.00	0.00	0.30	1.00	0.60	0.86
	Fallow land 2 Vertisols dominance	0.33	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00	1.00	0.60
Fallow land (Inseptisol dominance)	Fallow land 1 Inseptisol dominance	0.00	0.00	0.75	0.75	0.75	0.75	0.00	0.00	0.75	0.75	1.00	0.67
	Fallow land 2 Inseptisol dominance	1.00	0.14	0.13	1.00	0.50	1.00	0.00	0.00	0.63	1.00	0.75	1.00
Mixed forest	Mixed forest	0.00	0.00	0.17	0.50	0.17	0.50	0.00	0.00	0.17	0.25	0.33	0.25
Builtup	Builtup	0.00	0.00	1.00	1.00	1.00	1.00	0.17	1.00	1.00	0.86	1.00	0.86
	Mixed builtup 1	0.14	0.08	0.86	0.60	0.71	0.71	1.00	0.19	0.86	1.00	1.00	1.00
	Mixed builtup 2	0.56	0.24	0.33	0.43	0.44	0.44	0.89	0.35	0.89	0.62	0.56	1.00
	Kappa coefficient	0.15		0.56		0.60		0.35		0.67		0.74	
	Overall Accuracy (%)	22.00		60.00		64.00		41.00		70.00		76.00	

Note: UA-User's accuracy, PA-Producer's accuracy, MLE- Maximum likelihood estimation, RF- Random forest tree, SVM- Support vector machine

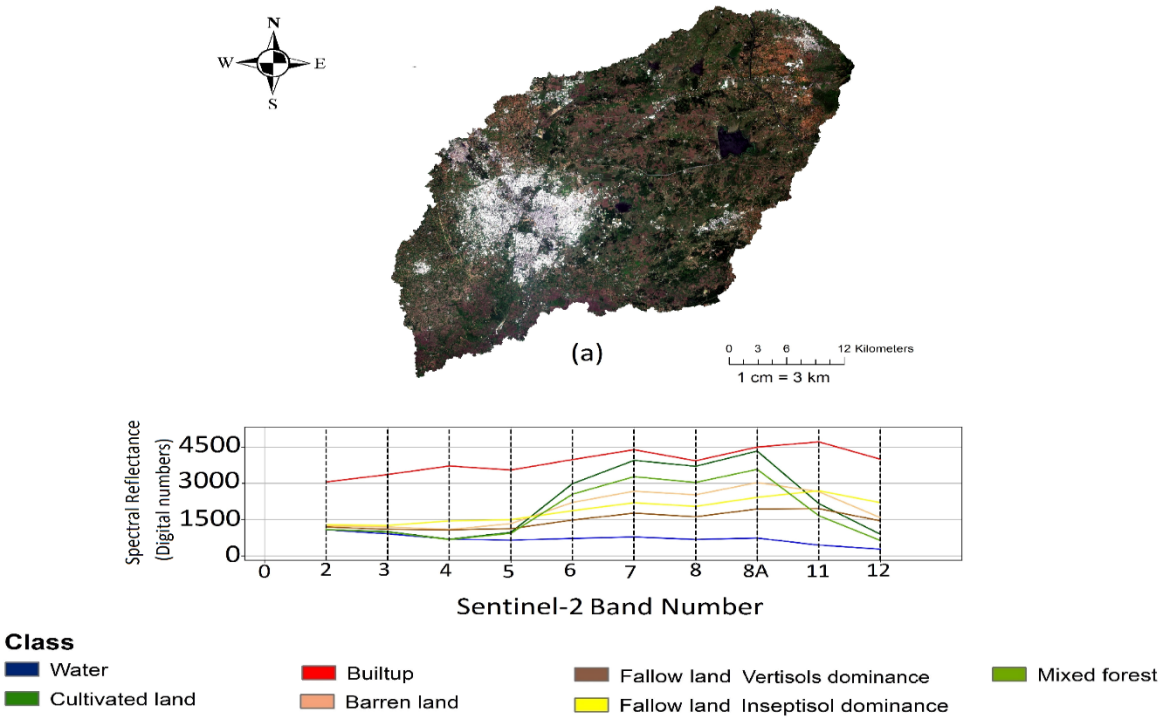


Figure 5.14: (a) Sentinel–2 bands visualized in natural colors composite (Stack 1) (b) Spectral response curves of LULC classes in Sentinel–2 bands.

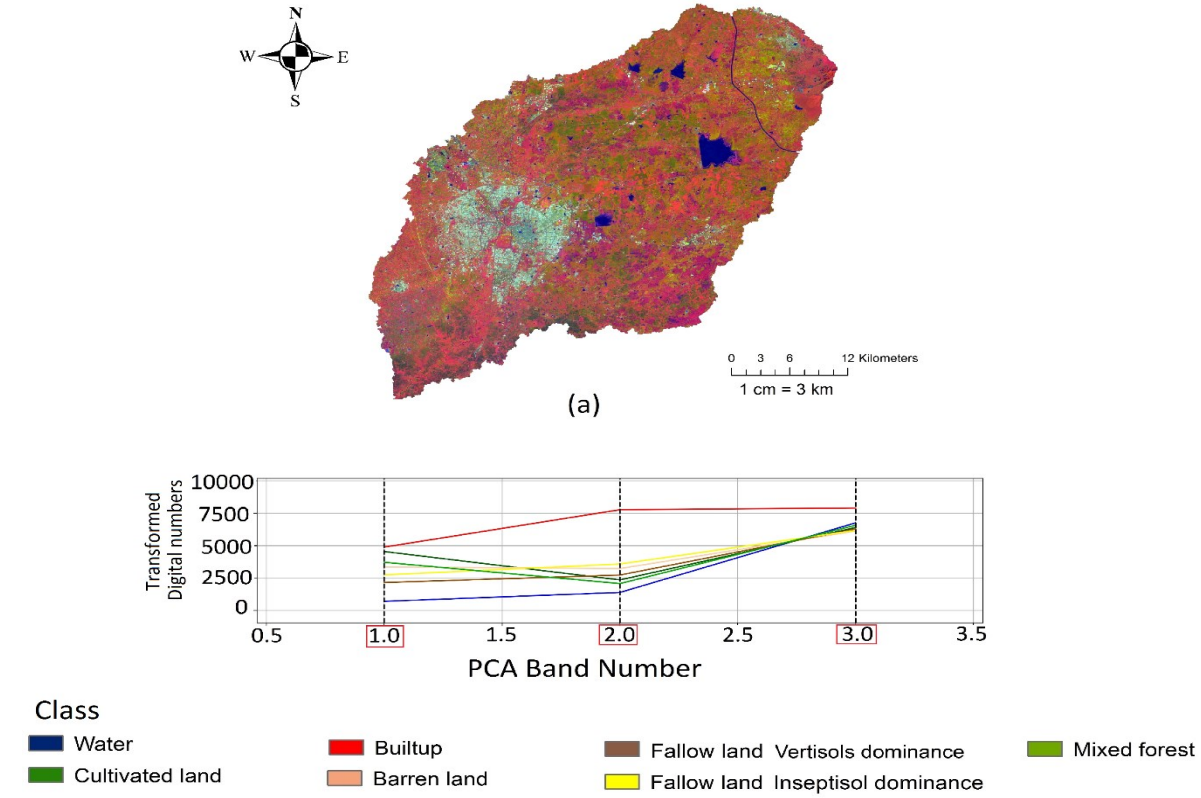


Figure 5.15: (a) The first three principal components visualized in false colors composite (Stack2) (b) Response curves of LULC classes in PCA bands.

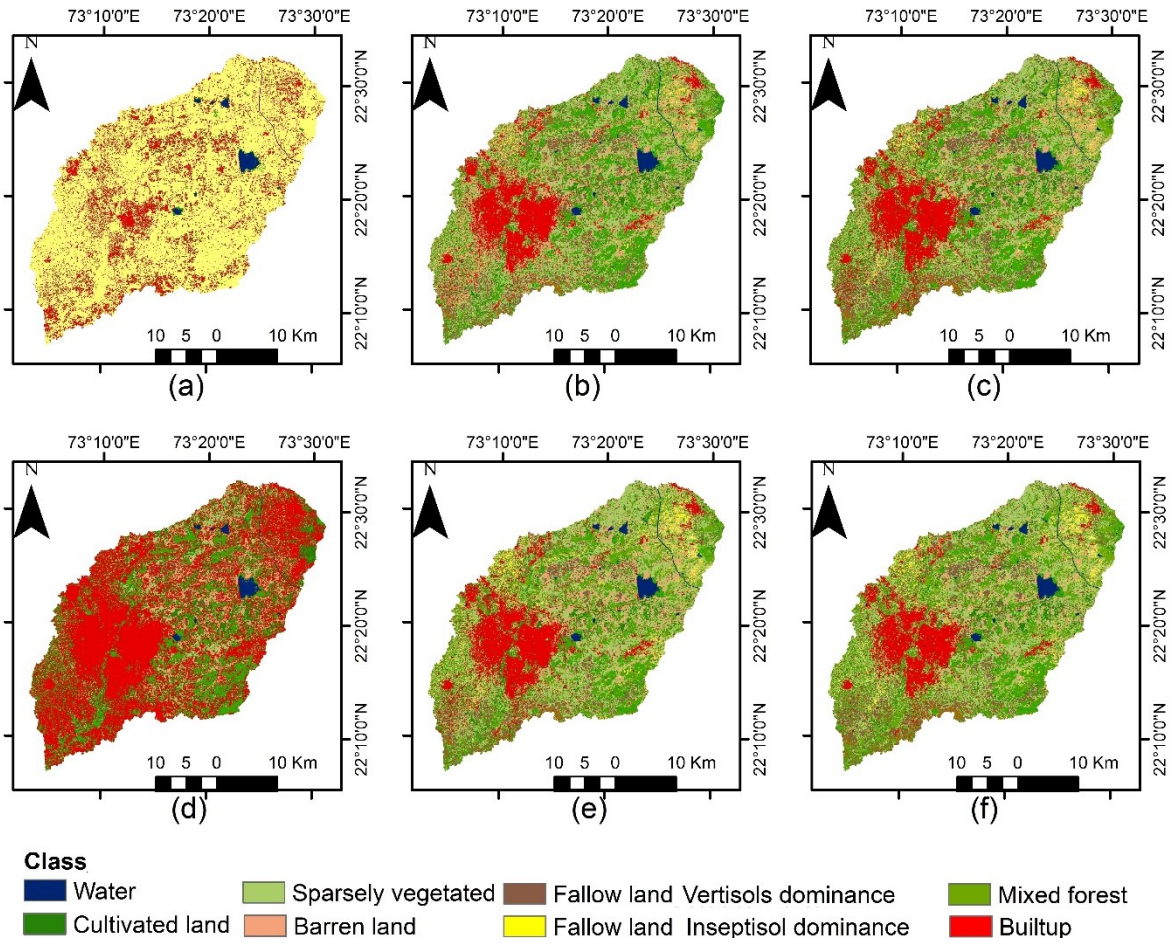


Figure 5.16: Comparative visual analysis of classification using traditional approach in (a) MLE (b) RF (c) SVM classifiers and PCA based approach in (d) MLE (e) RF (f) SVM classifiers.

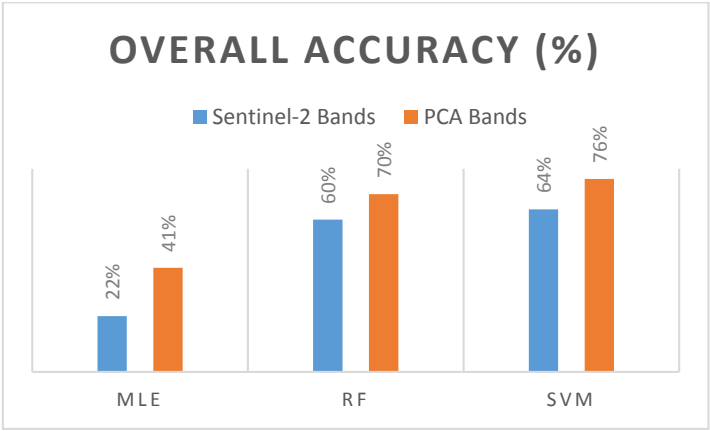


Figure 5.17: Classification accuracies obtained by the parametric (MLE) and non-parametric (SVM and RF) classifiers for traditional and PCA based approach.

The observed difference in classification accuracies obtained by the parametric (MLE) and non-parametric (SVM and RF) classifiers are similar to the other reported studies (Yang et al., (2011); Huang et al., (2002)). Nonetheless, little difference was also observed between the prediction performances of the SVM and RF machine learning algorithms. The land use and land cover

classification results showed some diversity due to the low number of training data, yet the global performance of both the classifiers were very similar for both the approaches. Hec-GeoHMS was used to create the Curve Number for the antecedent moisture condition II, generated using traditional approach and PCA based approach. Figure 5.18 shows the Curve Number II maps generated by both the approaches and statistics is shown in Table 5.10. It is evident from the results that land use and land cover map influence the Curve Number Map significantly.

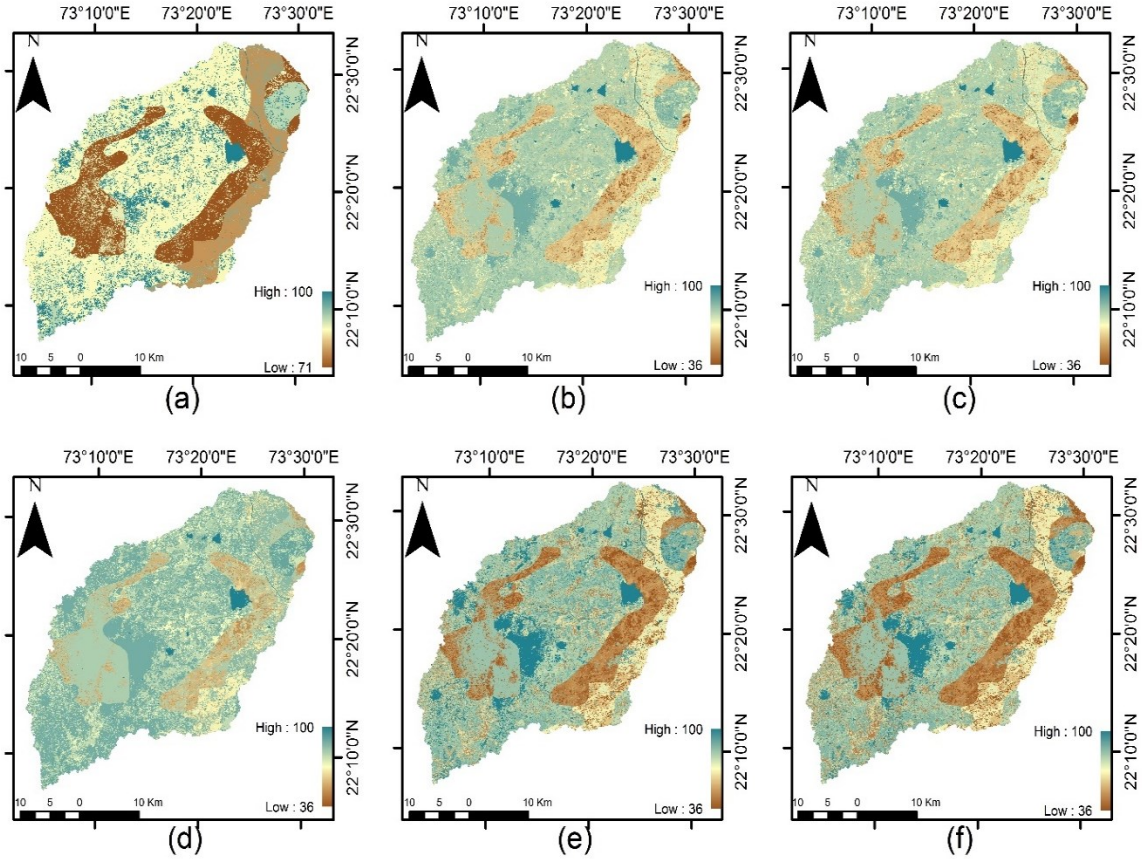


Figure 5.18: Curve number for antecedent moisture condition II generated from traditional approach (a) MLE (b) RF (c) SVM classifiers and PCA based approach in (d) MLE (e) RF (f) SVM classifiers.

Table 5.10: Resulted Curve Number II statistics using traditional and PCA based approach classification.

	Sentinel-2 Bands			PCA Bands		
	MLE	RF	SVM	MLE	RF	SVM
Minimum	71	36	36	36	36	36
Maximum	100	100	100	100	100	100
Mean	85.75	84.38	84.04	87.85	83.51	82.39
Median	87	88	88	89	87	88
Standard deviation	6.53	9.11	9.54	7.36	10.75	11.56

For the further analysis Curve Number II map generated with the help of the SVM classified (Figure 5.18 (F)) land use and land cover map using PCA based approach was used because of its highest accuracy. The calculated Curve Number is also termed as CN II for AMC II (Antecedent Moisture Condition II). The calculated Curve Number are adjusted to dry moisture conditions (called as AMC I) and high moisture conditions (called as AMC III) by using adjusting factors (Eq. (5.1 and 5.2)) (Subramanya, (2013)). The CN I, CN II and CN III values varied from 19 to 100, 36 to 100 and 56 to 100, respectively for the study area (Figure 5.19). Lower numbers indicate low runoff potential while larger numbers indicate an increased runoff potential.

AMC-I $$CN_I = \frac{CN_{II}}{2.281 - 0.01281\ CN_{II}} \quad (5.1)$$

AMC-II $$CN_{III} = \frac{CN_{II}}{0.427 - 0.00573\ CN_{II}} \quad (5.2)$$

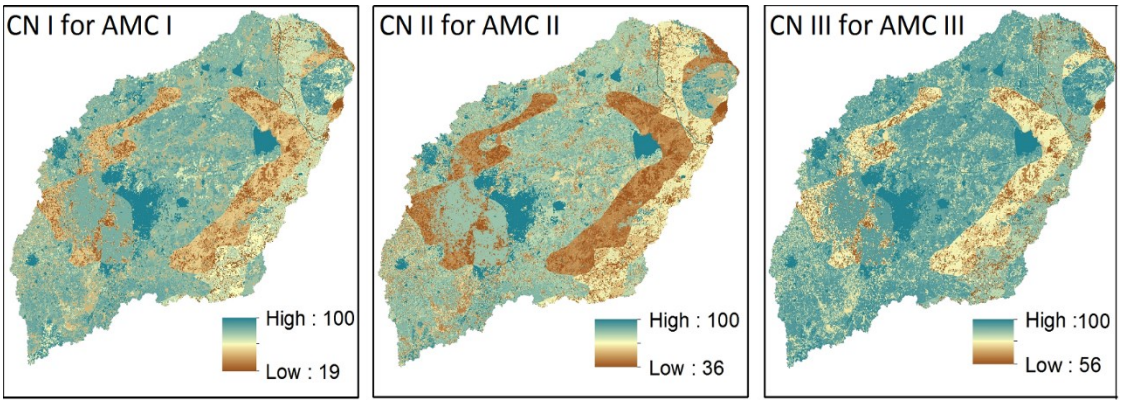


Figure 5.19: Curve number maps for antecedent moisture condition I, II and II generated from SVM classified LULC using PCA based approach.

3. **Objective:** To perform Morphometrical analysis of Vishwamitri watershed and prioritization of sub-watersheds for assessing the flood influencing characteristics of the sub-watersheds of the Vishwamitri watershed.

Watershed morphometry reveals lumped or semi-distributed watershed features. Watershed hydrology is highly influenced by its morphometry. Runoff potential is directly related to a variety of morphometrical parameters including drainage density, drainage frequency, mean bifurcation ratio, drainage texture, and elongation ratio (i.e., the greater the values of these parameters, the greater the watershed's runoff potential and vice versa). Morphometric parameters were directly calculated from the Cartosat-1 30m DEM by using Arc-hydro tools. Morphometry of Vishwamitri watershed and sub-watersheds (Figure 5.20), and its hydrological importance are discussed in detail below.

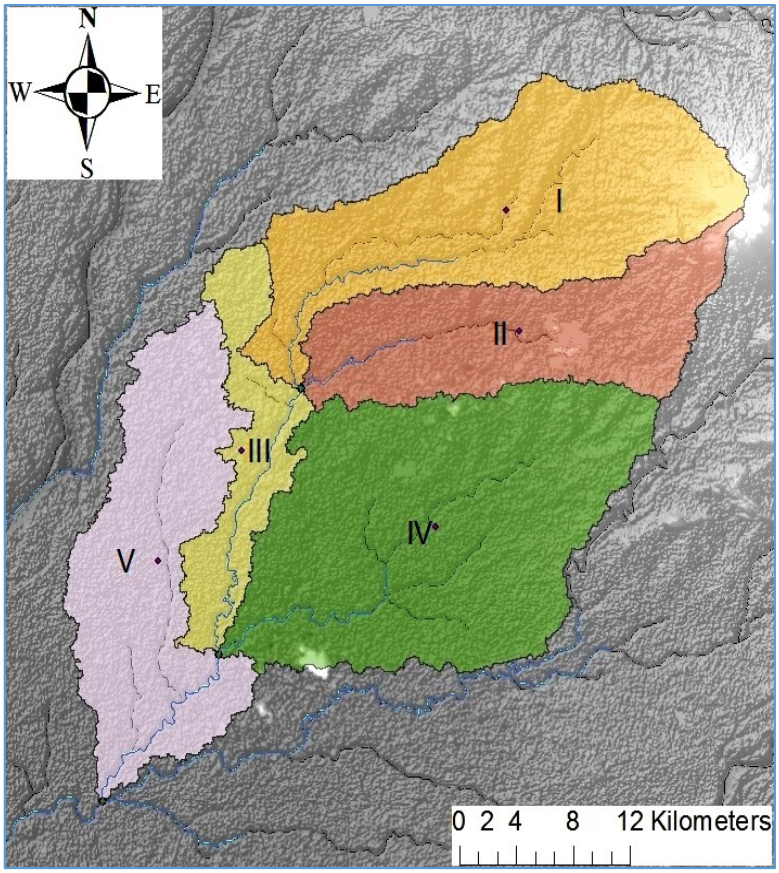


Figure 5.20: Sub watersheds of Vishwamitri watershed.

I. Basic parameters:

The basin length (L_b) is the longest length of the watershed from the head waters to the point of confluence. The basin length determines the shape of the watershed. High basin length indicates an elongated watershed. The computed L_b for the Vishwamitri watershed using Arc-hydro tool is 66.232 km. Another significant parameter is the area of the watershed *(A)*, the computed Vishwamitri watershed area is 1289.39 km^2. Perimeter of watershed (P) is the outer boundary of the watershed that enclosed its area. The computed perimeter for the Vishwamitri watershed is 279.44 km. Basin length, Area and Perimeter of the watershed are used as a watershed shape and size indicator.

II. Linear, aerial, and relief morphometric parameters:

Stream order (S_u): Stream ordering is a method of assigning a numeric order to links in a stream network. It is defined as a measure of the stream's position in the hierarchy of tributaries. There are different systems available for ordering streams. Based on the Strahler, (1964) system of stream ordering, the watershed has been designated as a fifth order watershed in Figure 5.21. The stream order increases when streams of the same order intersect. Therefore, the intersection of two first-order links will create a second-order link, the intersection of two

second-order links will create a third-order link, and so on. The intersection of two links of different orders, however, will not result in an increase in order. For example, the intersection of a first-order and second-order link will not create a third-order link but will retain the order of the highest ordered link (Omran et al., (2016)). The highest frequency is seen in the first-order streams in this study. (Table 5.11). The research area's hilly section has a higher number of first-order streams, indicating that the topography is dense and the bedrock lithology is compacted.

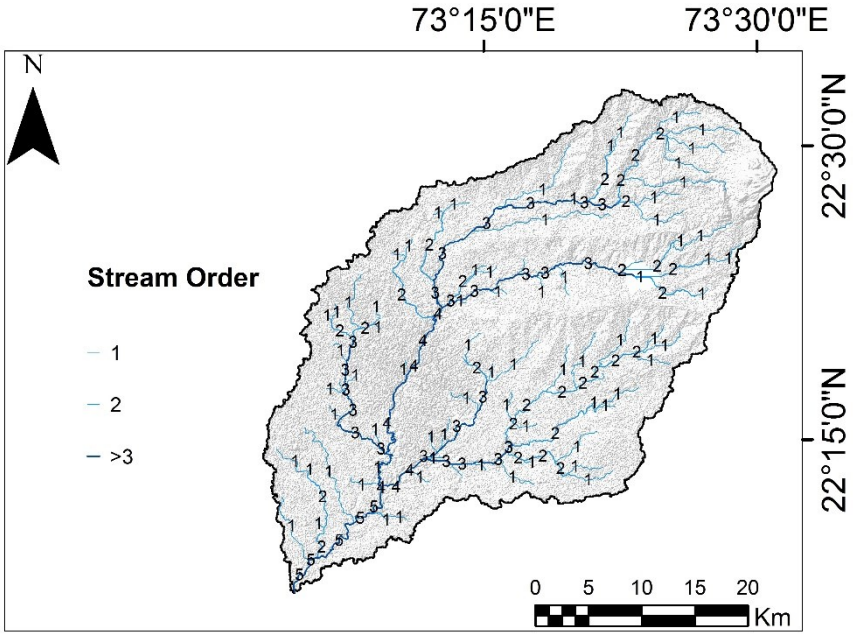


Figure 5.21: Drainage network map showing stream order of the Vishwamitri watershed.

Stream Length (L_u): The total length of individual stream segments of each order is the stream length of that order. Generally, the total length of stream segments is the maximum in first-order streams and decreases with an increase in the stream order (Pande & Moharir, (2017)). Streams of relatively short length represent areas of steep slopes and finer texture, while longer stream lengths are generally indicative of low gradients. Total stream length of each stream order is tabulated in Table 5.11.

Table 5.11: Computed stream order, stream number, stream length and bifurcation ratio.

Stream order (S_u)	Stream number (N_u)	Stream Length (Km) (L_u)	Bifurcation Ratio R_b
1	85	294.5	2.1
2	39	124.7	1.3
3	30	88.1	3
4	10	33.0	2
5	5	16.7	
Total	169	557.1	Mean 2.1

Bifurcation Ratio (R_b): The bifurcation ratio is the ratio of the number of stream segments of given order N_u to the number of streams in the next higher order (N_u+1). R_b is an important parameter to affect peak of the runoff hydrograph. High R_b values indicate instantaneous discharge and possibility of flash flooding during extended rainy hours (Rakesh et al., (2000)). However, R_b does not precisely remain constant between stream orders because of variations in watershed geometry, lithology, and tectonics. Flat terrain has a low R_b value, whereas mountainous or highly dissected terrain has values from 3 to 5 (Horton, (1945); Strahler, (1957)). In the present study, mean bifurcation ratio (Rbm) for the overall watershed is 2.12. The low R_b value for Vishwamitri watershed suggests delayed hydrograph peak. The lower value of R_b is an indicator of the watersheds, which have undergone fewer or less structural disruptions and the drainage pattern has not been distorted because of structural disturbances (Nag, (1998)). The higher value of R_b indicates highly dissected terrain, mature topography with a higher degree of drainage integration, and higher discharge potential (Eze & Efiong, (2010)). In particular, high R_b value of sub-watershed SW I indicates an early hydrograph peak with high potential for flash flooding during storm events amongst all the sub-watersheds. It is usual to use the weighted mean R_b value to characterize a watershed using more representative values in situations when the values of R_b differ for sequential stream orders (Moges & Bhole, (2015)). For this reason, the weighted mean R_b of the study watershed was calculated as follow.

$$Weighted\ mean\ bifurcation\ ratio\ (WRB) = \frac{R_{b1}N(u_1) + R_{b2}N(u_2) + R_{b3}N(u_3) + R_{b1}N(u_4)}{total\ number\ of\ stream\ segments} \quad 5.3$$

Where,

R_{b1} = *bifurcation ratio between 1st and 2nd order*
R_{b2} = *bifurcation ratio between 2nd and 3rd order*
R_{b3} = *bifurcation ratio between 3rd and order*
R_{b4} = *bifurcation ratio between 4th and 5th order*
$N(u_1)$ = *total number of streams involved in R_{b1} computation*
$N(u_2)$ = *total number of streams involved in R_{b2} computation*
$N(u_3)$ = *total number of streams involved in R_{b3} computation*
$N(u_4)$ = *total number of streams involved in R_{b4} computation*

The weighted mean bifurcation ratio (WRB) for the watershed of Vishwamitri is 3 and indicates that geological structures (tectonic activity) exert very low influence on the pattern of streams.

Drainage density (D_d): Drainage density of the watershed is calculated by dividing the total length of streams of all orders by the drainage area of the watershed to indicate the closeness of spaces between channels. In other words, it provides the quantitative value for the average length of all channels for the whole watershed. The measurement of drainage density provides a numerical measurement of landscape dissection and runoff potential (Reddy et al., (2004)). The D_d of Vishwamitri watershed is 0.43 km/km^2. The D_d of sub-watersheds range from 0.35-0.5. Drainage density has been classified with the following value ranges (km/km^2), i.e., very

coarse (<2), coarse (2-4), moderate (4-6), fine (6-8), and very fine (>8) (Sukristiyanti et al., (2018)). A high value of D_d indicates a relatively high density of streams, high runoff, a quick stream response, and consequently, a low infiltration rate. By contrast, low drainage density of a watershed implies low runoff and takes longer time to peak. Low class of D_d shows a poorly drained watershed with a slow hydrologic response. In addition to this, low class of D_d has a resistant permeable subsurface material, dense vegetation cover and low relief.

Drainage frequency (F_s): Drainage frequency is defined as the total number of streams of all orders per unit area. The result (Table 5.12) shows that F_s is maximum in sub-watershed SW III (0.19/km^2), followed by SW II and SW V (0.15/km^2), SW IV (0.14/km^2) and SW I (0.10/km^2). The discharge from SW I takes longer to peak because of low runoff rates due to lesser number of streams. Overall, the results of F_s reflect early peak discharge for sub-watersheds in order of their decreasing drainage frequency value, resulting in flash floods. F_s for Vishwamitri watershed is 0.13/km^2.

Drainage Texture (R_t): Drainage texture is defined as the total number of stream segments of all orders in a river watershed to the perimeter of the watershed. According to Smith, (1950), drainage texture has been classified into very coarse (<2), coarse (2–4), moderate (4–6), fine (6–8), and very fine (>8). According to this classification, Vishwamitri watershed has very coarse drainage texture (0.6 km^{-1}). Also, The R_t value for Vishwamitri sub-watersheds ranges from 0.14-0.35. The result (Table 5.12) shows that R_t is maximum in sub-watershed SW IV (0.35/km), followed by SW II (0.22/km), and SW V (0.21/km), SW I(0.20/km) and SW III (0.13/km). Hydrologically very coarse texture watersheds have large basin lag time periods (Altaf et al., (2013)).

Relief ratio (R_r): The relief ratio is called the maximum relief of the horizontal distance parallel to the main drainage line along the longest dimension of the watershed. It is a good indicator of the intensity of water flows from a catchment slope. It is the measurement of the overall steepness of a watershed. The high R_r implies shorter lag time and the watershed attains higher peak discharge and flow velocities. With increasing relief, steeper hill slopes and higher stream gradients, the time of concentration of runoff decreases, thereby increasing flood peaks (Bhatt & Ahmed, 2014). The R_r for Vishwamitri watershed is 0.01, indicating overall nearly flat terrain or lower slope values. The R_r values for sub-watersheds range between 0.00-0.02. The SW III, SW IV and SW V have 0 R_r indicating flat terrain with longer basin length and their influence on flood is much less. While, sub-watersheds SW I and SW II have relatively high values of R_r and contribute more water in a short period of time and cause floods in the lower region of the watershed.

Ruggedness number (R_n): The ruggedness number is expressed as the product of watershed relief and drainage density. High R_n occurs in those watersheds which have steep and long slopes and fine texture, thus, is highly susceptible to erosion and increased peak discharge. Slope is another important indicator of runoff, which provides general representation of relief ruggedness within the watershed. The calculated R_n value of Vishwamitri watershed is 0.32. The low R_n value of Vishwamitri watershed due to low relief and lesser degree of terrain complexity, causing less water flow. In the upper Vishwamitri watershed, SW I and SW II have relatively high R_n values, indicating that they have high relief, fine texture, and possibilities of high surface flow (Table 5.12). Moreover, these sub- watersheds are susceptible to erosion and producing increased peak discharge. The SW III, SW IV and SW V have the lowest R_n values because of low relief and lesser degree of terrain complexity causing less water flow.

Form factor (F_f): Form factor is the ratio of the area of the watershed and square of the watershed length. F_f represents the shape or outlines of a watershed and is useful in predicting the flow intensity of a catchment and has a direct link to peak discharge. High F_f values occur in watersheds having potential to produce high peak flows in short duration and low F_f values are vice versa. For a perfectly circular watershed, the form factor value would always be greater than 0.78. The smaller the value of form factor, more elongated will be the watershed. The low F_f value of 0.29 of Vishwamitri watershed reveals that the shape of the watershed is elongated, it has less side flow for shorter duration and high main flow for longer duration. F_f of sub-watersheds of the Vishwamitri watershed is given the Table 5.12. The F_f values for sub-watersheds range between 0.1-0.3, and indicate elongated shape of sub-watersheds.

Circularity ratio (R_c): Circularity ratio is the ratio between the area of a watershed to the area of the circle having the same circumference as the perimeter of the watershed. The R_c values can attain a maximum of 1.0 where the outline of the watershed is approaching near circularity. A numerically low R_c indicates an elongated shape, while higher values are an expression of approach to near circularity. Elongated watersheds are characterised by longer lag times and lower peak discharge. In the study area, the overall R_c value of Vishwamitri watershed is 0.21 and, for sub-watersheds it ranges from 0.07-0.2. The R_c values suggest the elongated shape of the Vishwamitri watershed and its sub watersheds.

Elongation ratio (R_e): It is defined as the ratio of diameter of a circle with the same area as that of the watershed to the maximum basin length. The R_e values vary from 1 for circular watersheds and 0 for elongated watersheds. High R_e values occur for circular watersheds, considered as highly hazardous, because they yield peak flow in a short period of time compared to low R_e in elongated watersheds (Masoud, (2016)). These values can be grouped into three categories,

namely, circular (>0.9), oval (0.9–0.8), less elongated 0.8–0.7) and elongated (<0.7) (Lama & Malti, (2019)). The overall R_e value of Vishwamitri watershed is 0.61 and, for sub-watersheds it ranges from 0.37-0.62.

Length of overland flow (L_g): Length of overland flow is the length of water over the ground before it gets concentrated into certain stream channels. There are three classes of L_g i.e., low value (< 0.2), moderate value (0.2 – 0.3), and high value (>0.3). The low L_g value shows high relief and short flow paths, which are more susceptible to flash flooding. Meanwhile, a high L_g value implies gentle slopes and long paths of flow. L_g value for overall watershed is 1.43 and, for sub-watersheds it ranges from 1-1.43. The SW I has the lowest value of L_g, which means it is more susceptible to flash flooding.

III. Hypsometry Analysis:

The hypsometry and the hypsometric integral (HI) are used in classical conceptual geomorphometric models of landscape evolution as follows: i) for HI above 0.60 the area is considered young; ii) for HI ranging between 0.35 - 0.60 the area is in a steady state balance or mature phase and iii) HI below 0.35 characterizes a Monadnock phase in landscape evolution. Vishwamitri watershed is certainly indicative of a marked old stage in the basin's evolution (Figure 5.22), meaning that the watershed has reached the equilibrium in the longitudinal profiles of the river. This is further attested by the very low hypsometric integral (HI = 0.04). Low value of HI occurs in terrain characterized by isolated relief features standing above extensive level surfaces (Pike & Wilson, 1971).

The hypsometric analysis is important to analyze as an erosional process directly affects the morphometric of the watershed. Based on the hypsometric analysis the Vishwamitri watershed is stable or in the old stage of the erosional process, the computed flood influencing parameters and calculated compound value of the watershed will hold true until there is a major structural disturbance that occurred due to tectonic activity. If the HI value were high, the computed flood influencing parameters and calculated compound value of the Vishwamitri watershed would need to be calculated again to compensate for the changes that occurred due to the erosional process of the young watershed.

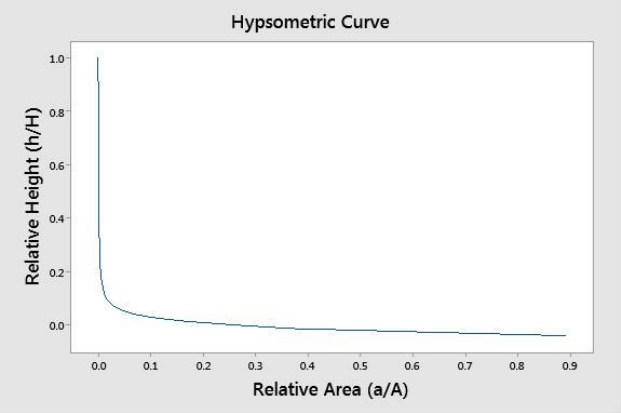


Figure 5.22: Hypsometry curve of Vishwamitri watershed showing Monadnock phase in landscape evolution.

IV. Compound value and weightage:

Single or limited parameters cannot present a comprehensive picture of the flood hazard potential of any sub-watershed, and hence, each of the linear, aerial, and relief morphometric parameters along with curve number is taken into consideration for assessing the flood influencing characteristics of the five sub-watersheds of the Vishwamitri watershed (Table 5.12), as these parameters have a direct but variable relationship with flood runoff. Therefore, influencing value or rank (highest weightage 5 and least 1) is given to each sub-watershed based on the nature of the selected parameter (Table 5.13). Prioritization was achieved through the allocation of weights to the individual indicators contributing to flood runoff and a compound value (C_v) was calculated for final prioritization. C_v is derived by calculating the average of ranks assigned to the individual parameters. The sub-watershed with highest C_v contributes most to flood runoff and as a result needs highest priority for flood mitigation measures, whereas the sub-watershed with lowest C_v is contributing least to flood runoff thereby has low priority. Thus an index of high, medium and low priority was produced. Based on the integration of each flood influencing parameter and calculated compound value, the SW I and IV areas of Vishwamitri watershed have been categorized into high priority, SW II and V into moderate priority, and SW III into low priority. In order to mitigate floods, it is proposed that there is a significant need to create a flood spill channel that can take up to one-third of the total flow of the Vishwamitri river. Moreover, to prevent floods in the downstream agricultural areas and settlements, an additional reservoir must be created in SW I. Along with this, mitigation measures such as, check dams, nala bunds, gully plug, bundhis (local name in India), percolation tanks, etc. can be constructed in a planned and systematic manner in SW I, II and IV to create water buffers within the catchment, which will help reducing vulnerability to seasonal variations in rainfall. Nala

bunds and percolation tanks are structures built across or closer to nalas (streams) to increase water percolation, increase the moisture regime of the soil, and restrict silt flow.

Table 5.12: Flood influencing characteristics of the five sub-watersheds of the Vishwamitri watershed.

Sub-watershed ID	**I**	**II**	**III**	**IV**	**V**
Form factor (F_f)	0.166	0.154	0.108	0.304	0.208
Circulatory ratio (R_c)s	0.159	0.141	0.067	0.199	0.106
Elongation ratio (R_e)	0.459	0.442	0.371	0.622	0.515
Drainage density (D_d)	0.500	0.403	0.497	0.418	0.349
Drainage texture (R_t)	0.203	0.228	0.139	0.353	0.217
Relief ratio (R_r)	0.017	0.020	0.001	0.002	0.001
Ruggedness number (R_n)	0.362	0.293	0.015	0.024	0.011
Weightd bifurcation ratio (WRB)	3.643	2.119	3.155	3.213	3.205
Length of overland flow (L_g)	1.000	1.240	1.005	1.195	1.431
Drainage frequency (Fs)	0.102	0.150	0.189	0.138	0.151
Curve number (CN)	81.972	83.060	81.197	81.846	83.607

Table 5.13: Prioritization of sub-watersheds based on compound value.

Sub-watershed ID	F_f	R_c	R_e	D_d	R_t	R_r	R_n	WRB	L_g	F_s	CN	sum	C_v
I	3	4	3	5	2	4	5	5	5	1	3	40	3.64
II	2	3	2	2	4	5	4	1	2	3	4	32	2.91
III	1	1	1	4	1	2	2	2	4	5	1	24	2.18
IV	5	5	5	3	5	3	3	4	3	2	2	40	3.64
V	4	2	4	1	3	1	1	3	1	4	5	29	2.64

4. **Objective**: To identify potential runoff storage zones based on the various physical characteristics of the Vishwamitri watershed using a GIS-based conceptual framework that combines through analytic hierarchy process using multi criteria decision-making method.

Rainfall analysis

Based on past 56 years (1961 to 2016) rainfall data analysis of rain gauge stations across watershed, it was determined that the SPI indicates extremely dry years 1.8 % of the time, moderately dry years 10.7 % of the time, moderately wet years 3.6 % of the time, near normal years 73.2 % of the time, very wet years 7.1 % of the time and extremely wet years 3.6 % of the time. Classification of annual rainfall of the study area based on SPI is shown in Table 5.14. The result therefore suggests that the overall drought events between these years were not severe. The precipitation analysis suggests that the water shortage in the region may be overcome by identification of suitable sites for water storage. The computed average annual rainfall for

Vadodara, Savali, Padra and Waghodia rain gauge stations are 928.96 mm, 891.56 mm, 883.85 mm and 854.53 mm respectively. The seasonal distribution of the precipitation in the study area varies and falls mostly as rain in monsoon season (June to September). Thiessen polygon of rain gauge stations are shown in Figure 5.23. Promoting rain water harvesting in areas receiving less than 100 mm/year or more than 1000 mm/year of rains is not recommended (Kahinda et al., (2008)). Water based activity are not feasible in areas that receive less than 100 mm/year of rain, also, there is no incentive to implement rain water harvesting schemes in areas with annual rains in excess of 1000 mm/year. Rainfall analysis of Vishawmitri watershed shows potential to carry out water based activity in the area.

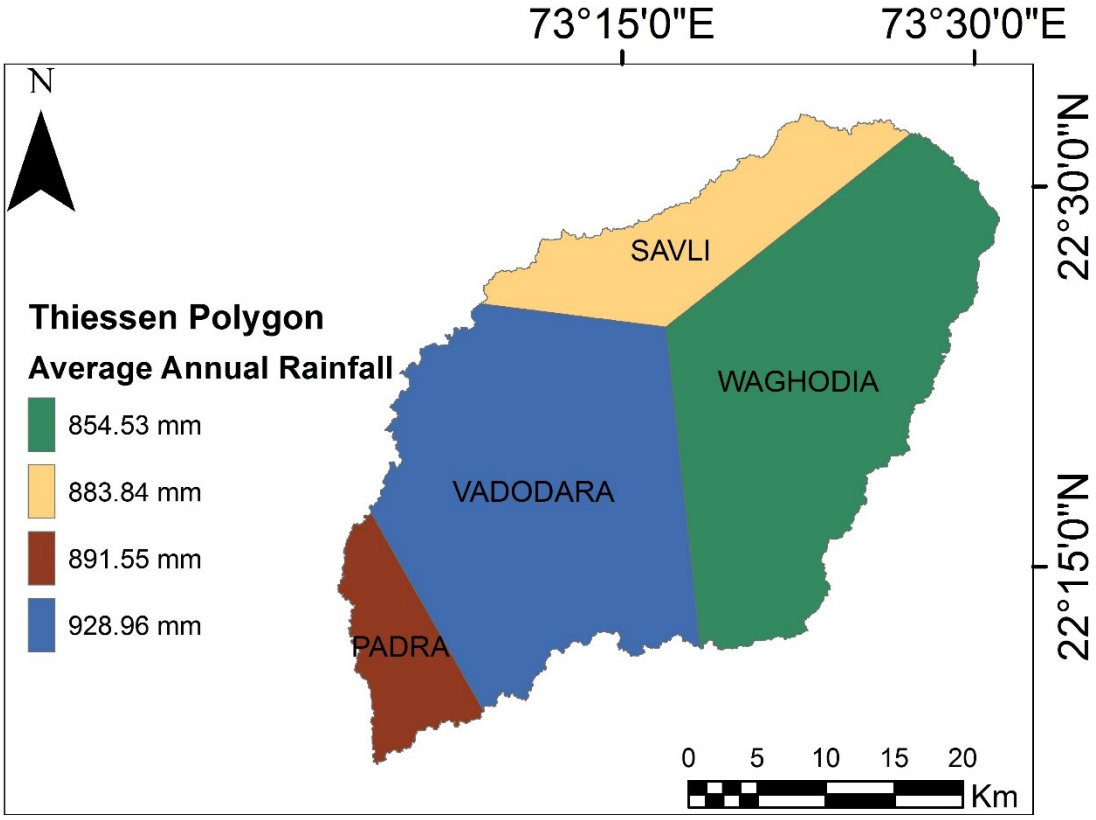


Figure 5.23: Thiessen polygon of rain gauge stations for Vishwamitri watershed.

Table 5.14: Classification of annual rainfall based on SPI.

Classification	**Years**
Dry extreme years	2008
Moderately dry years	1972, 1974, 1986, 1987, 1999,2000
Moderately wet years	1983, 1994
Near normal years	1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1971, 1973, 1975, 1977, 1978, 1979, 1980, 1981, 1982, 1984, 1985, 1988, 1989, 1990, 1991, 1992, 1993, 1995, 1996, 1997, 1998, 2001, 2002, 2003, 2004, 2007, 2009, 2011, 2012, 2014, 2015, 2016
Very wet years	1976, 2006, 2010, 2013
Wet extreme years	1970, 2005

Topography wetness index (TWI)

High values of the TWI are found in converging and flat areas and are expected to have much water accumulation and low slope. In contrast, steep locations and diverging areas receive a small index value and have relatively lower water accumulation. Consequently, the index is a relative measure of the hydrological conditions of a given location in the landscape. Figure 5.24 shows the calculated TWI for the Vishwamitri watershed.

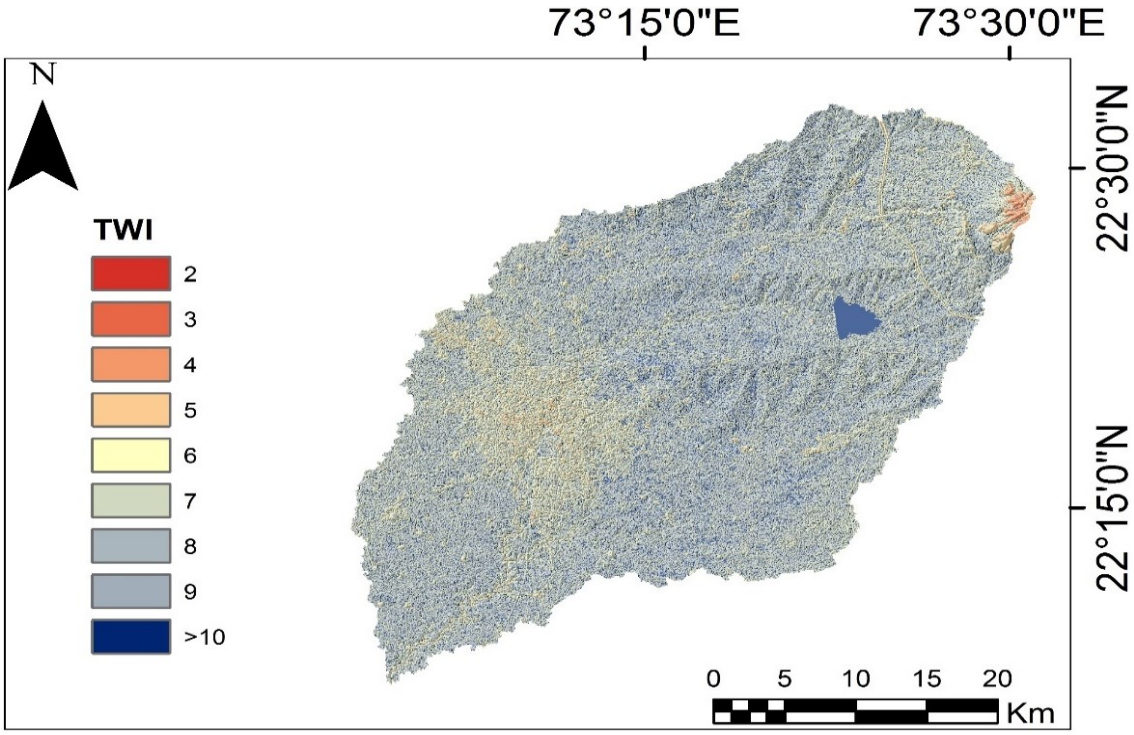


Figure 5.24: Spatial variation of topography wetness index across Vishwamitri watershed.

Generation of slope map using Topography Position Index (TPI)

Positive TPI values indicates that the target point location is higher than the average of its surroundings, as defined by the neighbourhood (ridges). TPI for the study area is shown in Figure 5.25 and Figure 5.26 shows conversion of elevation values to TPI along the cross section (shown in red colour in TPI map). Negative TPI values represent locations that are lower than their surroundings (valleys). TPI values near zero are either flat areas (where the slope is near zero) or areas of constant slope (where the slope of the point is significantly greater than zero). Slope classification was done as suggested by Weiss, (2001) and Jenness, (2006). Slope plays a significant role in the amounts of runoff and sedimentation, the speed of water flow and the amount of material required to construct a dyke (the required height)(Ammar et al., (2016)). Results shows (Table 5.15 and Figure 5.27) that maximum area belongs to flat class (35.45%), flat areas are never strictly horizontal, also, flat slopes lead to a decrease in the surface runoff

velocity, which results in a longer period of time for the runoff to drain; there are gentle slopes in a seemingly flat area. Ponds are suitable for small flat areas with slopes 5%, 0.15 % belongs to middle slope, nala bunds are suitable on moderate slopes of 5–10%, 12.49 % area belongs to upper slope, terracing is suitable for steeper slopes of 5–30%. Ridges and upper slope together forms 30.79 % of the area, they indicate least potential for rainwater harvesting because higher sloping land is inappropriate for constructing water storage structures. Valley and lower slope together constitutes for 33.58 % of the area, small dams or check dams like structures are preferable on such sites.

Table 5.15: Slope classifications using TPI as basis of landform classification for the study area.

Class	Description Breakpoints	Area (km^2)
Valley	TPI ≤ -4.6	255
Lower Slope	- 4.6< TPI ≤ -2.3	178
Flat Slope	-2.3 < TPI < 2.3, Slope ≤ 5°	457
Middle Slope	--2.3 < TPI < 2.3, Slope > 5°	2
Upper Slope	2.3 < TPI ≤ 4.6	161
Ridge	TPI > 4.6	236

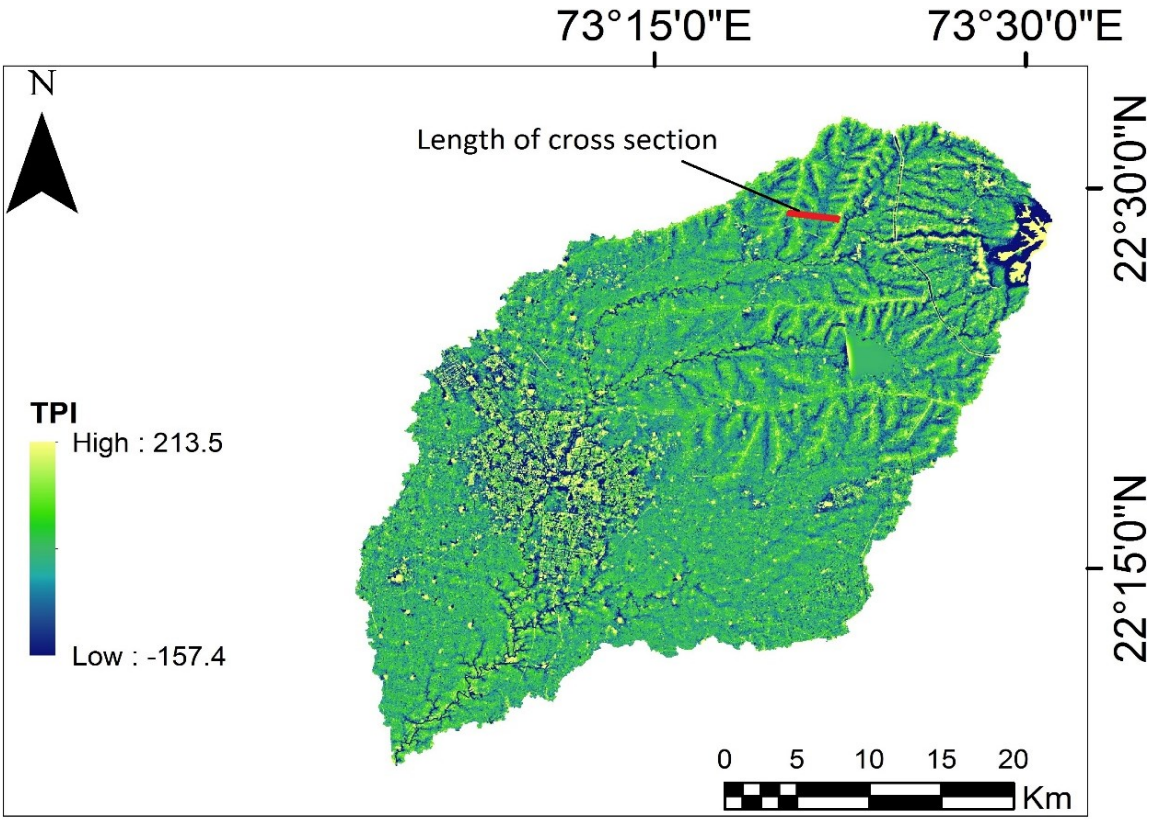


Figure 5.25: Variation of topography position index across the Vishwamitri watershed.

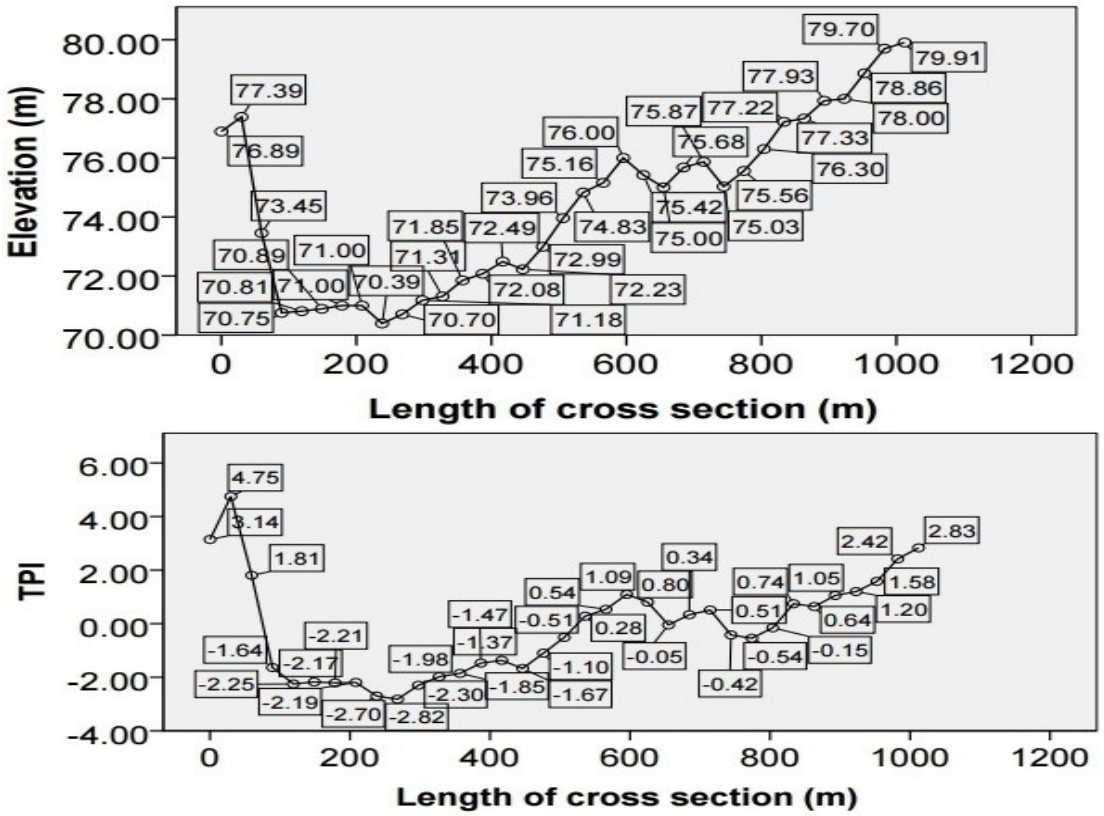


Figure 5.26: Comparison between the original DEM and TPI along the cross-sectional profile (red line).

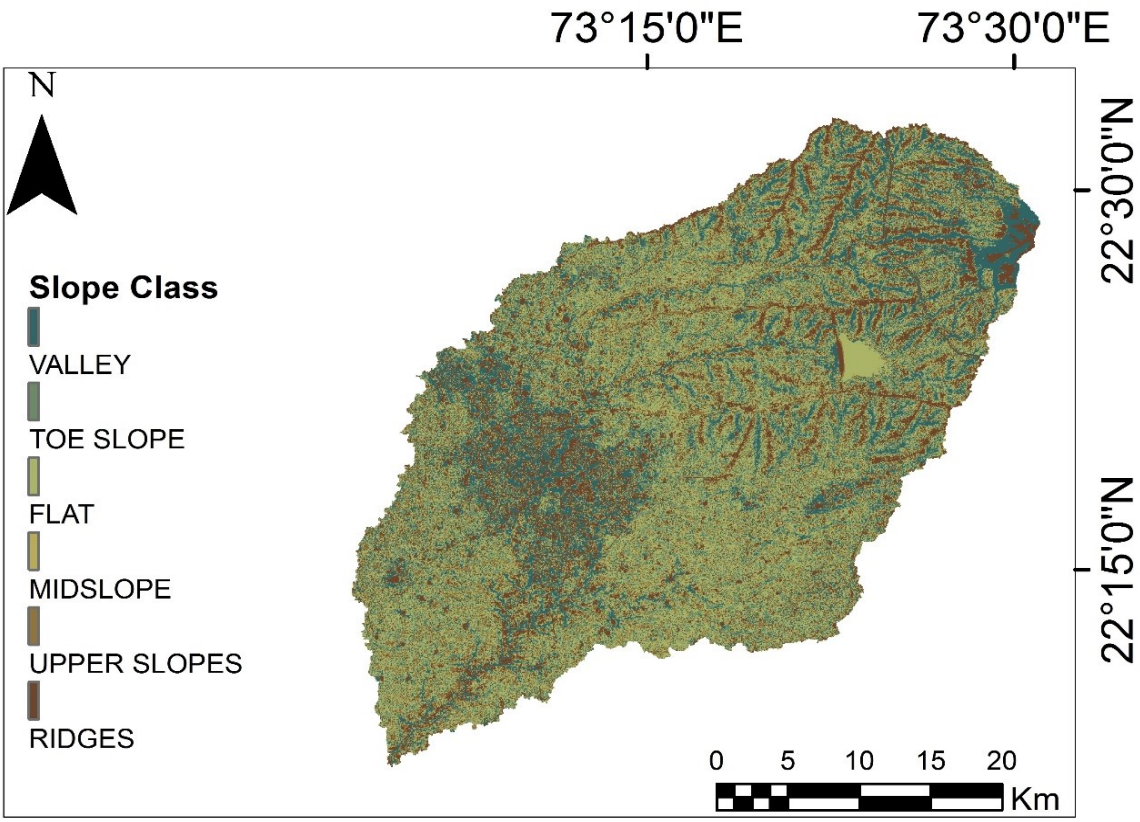


Figure 5.27: Resulted Slope map of the study area using TPI as basis of landform classification.

Land use/land cover

The land use/land cover map of the study area is shown in Figure 5.28, which reveals that there are seven major types of land use/land cover namely Waterbodies, Builtup, Mixed forest, Cultivated land, Barren land, Fallow land with vertisols dominance and Fallow land with inceptisols dominance. Major portion of the study area (about 35%) is Agricultural land (cultivated land, fallow land with vertisols dominance and fallow land with inseptisol dominance) followed by Sparsely vegetated (19%), Mixed forest (14%), Builtup (12%), Barren land (5%) and Water bodies (2%). Land-use classes such as barren land and sparsely vegetated land are generally recommended for water storage zones/structures.

Soil texture

According to the classification based on soil texture, seven types of soils are found in the Vishwamitri watershed (Figure 5.29). Typic Ustifluvents and Fluventic Haplustepts correspond to HSG group A, Udic Haplustepts corresponds to HSG group B, Chromic Haplusterts and Typic Haplustepts correspond to HSG group C, and Lithic Haplustepts and Vertic Haplustepts corresponds to HSG group D. The most dominating soil, Chromic Haplusterts (HSG group C), covers 48.94% of the total watershed area. HSG-A has the lowest runoff potential (typically contains more than 90% sand and less than 10% clay), HSG-B has moderately low runoff potential (typically contains between 10 to 20% clay and 50 to 90% sand), HSG-C has moderately high runoff potential (typically contains between 20 to 40% clay and less than 50% sand) and HSG-D has high runoff potential (typically contains more than 40% clay and less than 50% sand)(Ross et al., (2018)).

Stream order

Suitable zones/sites for surface water storage structures, all the1st, 2nd and 3rd order streams were extracted from the drainage network map and a stream-order buffer map was developed with a buffer distance of 50 m on both sides of the streams. It can be seen from the drainage network map (Figure 5.21) that the study area has a fifth order drainage network, with a good drainage network in the eastern portion. Length of the first order streams in the study area is nearly 294.5 km (about 52.9% of total length of drainage). The second and third order streams have also considerable drainage lengths, 124.7 km (22.4%) and 88.1 km (15.8%), respectively. The fourth order stream contributes to drainage with a length of 33.0 km, which accounts for 5.9% of the total drainage length. Mainly, the Vishwamitri River is a fifth order stream having a drainage length of 16.7 km (3.0%).

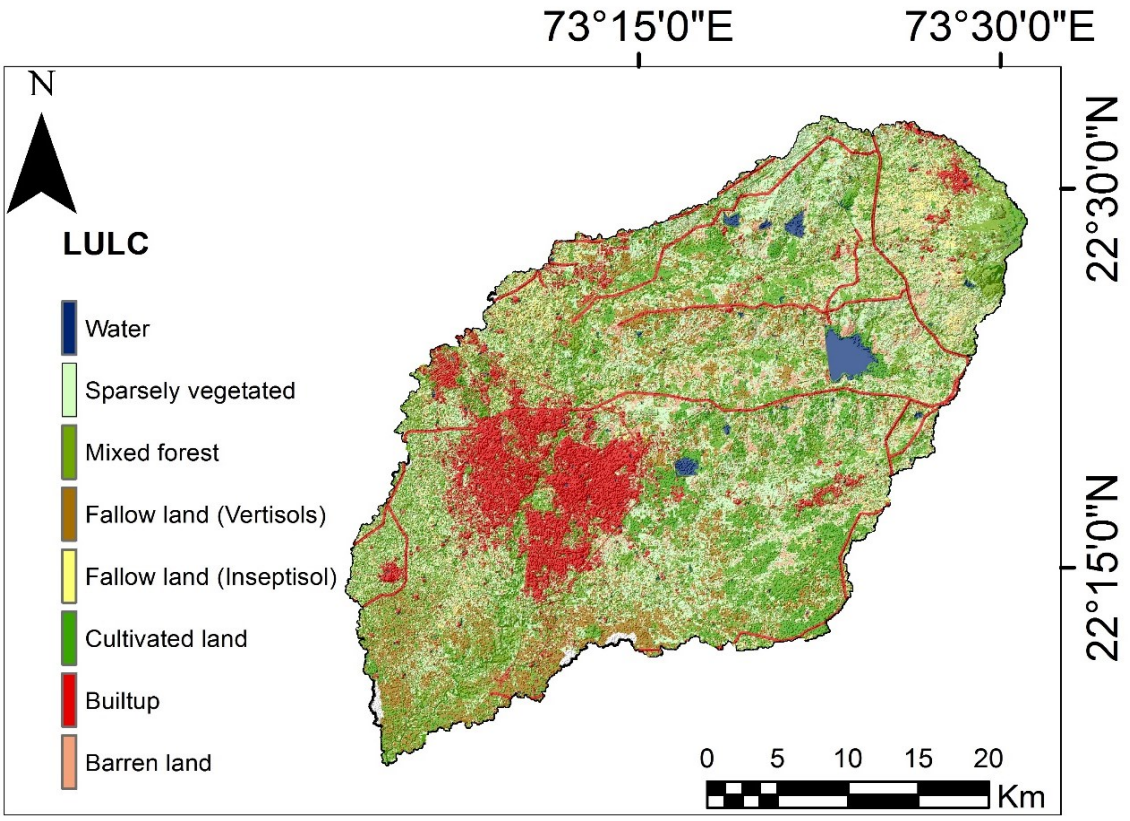


Figure 5.28: Land use/land cover map of the Vishwamitri watershed.

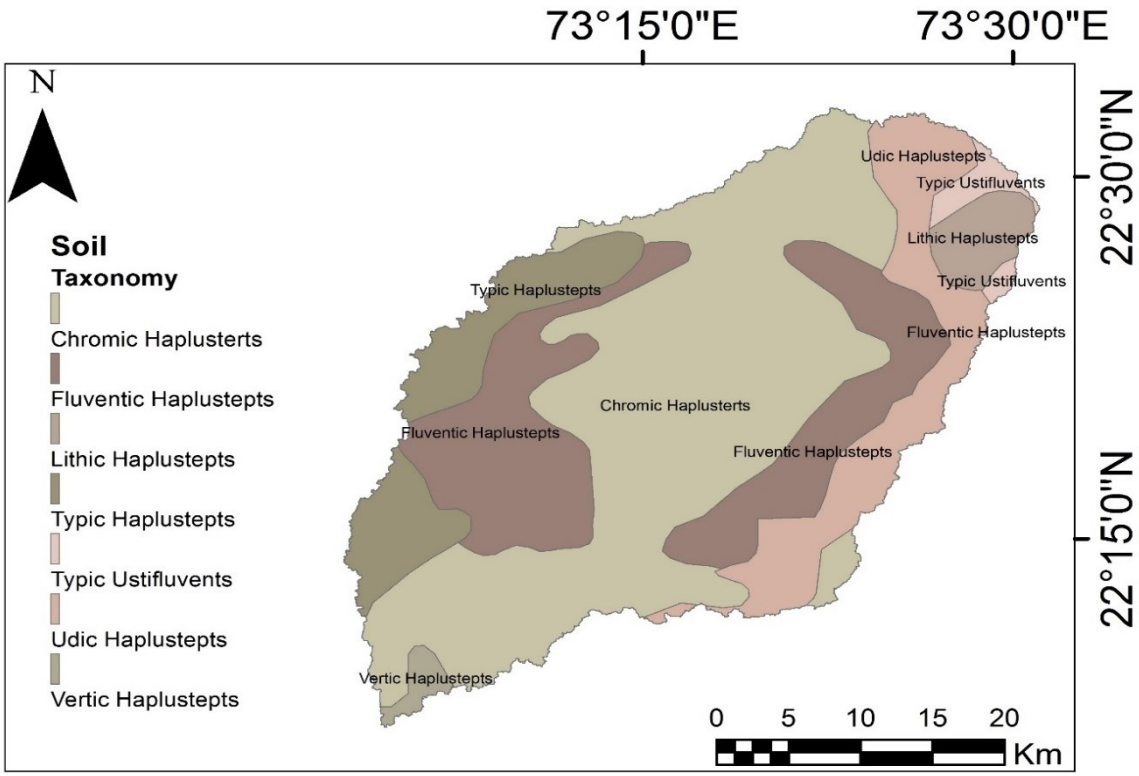


Figure 5.29: Soil data based on soil texture collected from (NBSS & LUP).

Curve number (CN)

Curve Number considers the relationship between land use/land cover and hydrologic soil group, which together make up the curve number.The Curve Number value varied from 36 to 100 for the study area (Figure 5.30) with mean value of 82.39 and standard deviation of 11.56, lower numbers indicate low runoff potential while larger numbers indicate an increased runoff potential. A CN value of 100 represents a condition of zero potential maximum retention that suggests an impermeable catchment having maximum runoff-generation capability. On the other hand, a CN value of 0 suggests an infinitely abstracting catchment having zero runoff-generation capability (Jha et al., (2014)). With this approach, the suitable locations for surface water storage zones/structures were located in areas with the highest capacity for runoff generation and nearby to existing drainage lines.

Height above nearest drainage (HAND)

Low-lying land adjacent to streams is more susceptible to be flooded than higher land. HAND values for the study area varies from 0 to 749 m. HAND raster was prepared for the 4^{th} and 5^{th} order streams of Vishwamitri watershed as they are highly susceptible to flooding. On suitability scale, lower values were assigned for HAND values ranging from 0−2 meters as these areas are more prone to flooding and it is not recommended to build water storage structure on such zones. Also, lower values were assigned on suitability scale for extremely high HAND values (greater than 48 meters) because higher HAND value means point moves away from the river. Figure 5.31 shows the Height above nearest drainage (HAND) map of the study area.

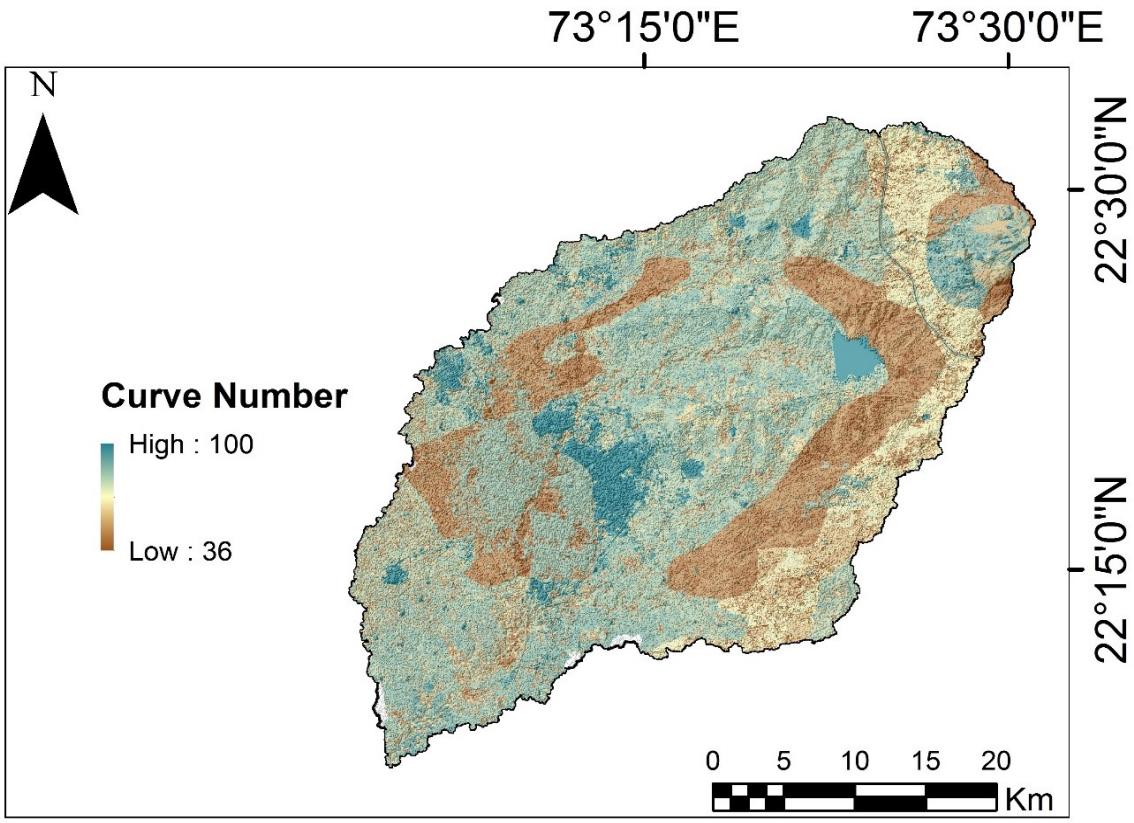


Figure 5.30: Variation of curve number across the Vishwamitri watershed for AMC II condition.

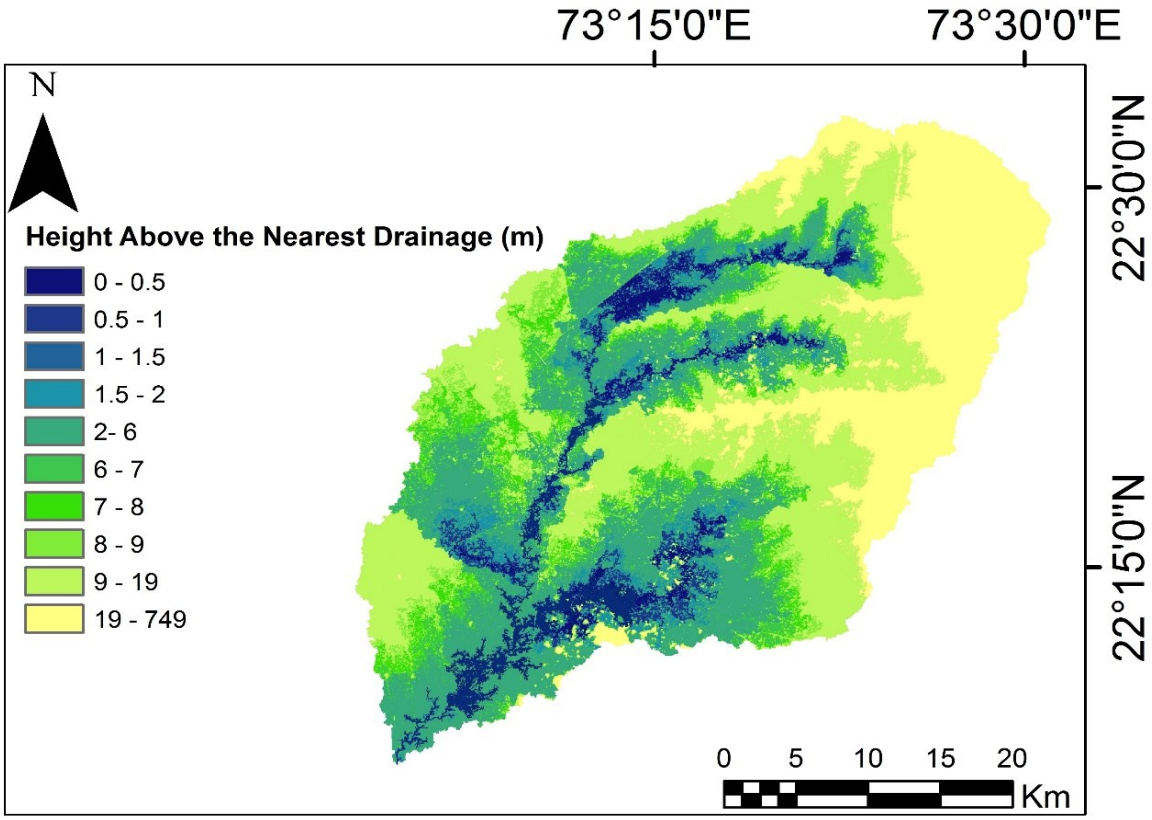


Figure 5.31: Height above nearest drainage (HAND) map of the Vishwamitri watershed.

Determining criteria weights using AHP

AHP provided a systematic approach to conduct MCDM. To derive suitability maps for potential runoff storage zones, the criteria maps have to be related to the result of the AHP. The AHP pair-wise matrix for the criteria used in this study is presented in Table 5.16. For all the five spatial layers the relative importance is derived. Thus, a common scale (0 % to 100 %) is obtained from AHP procedure. As seen in Table 5.16, the most important criterion for decision-making is Slope (22.50%), followed by LULC (21.20%), Curve number (14.80%), HAND (14.50%), Stream order (14.40%) and TWI (12.60%). Calculated principal eigenvector is 6.47, which is computed with the square reciprocal matrix of pairwise comparisons between criteria. Since the AHP may have inconsistencies in establishing the values for the pairwise comparison matrix, it is important to calculate this level of inconsistency using the consistency ratio (CR) (Rincón et al., (2018)). The CR of the pair-wise matrix is 7.5% (which is less than 10 %) and thus the judgments made and compiled in the pair-wise matrix of Table 5.16 are acceptable. This implies that the comparisons were performed with good judgment, weightage for each criterion is suitable to weighted overlay.

Table 5.16: Resulting weights for the criteria based on pairwise comparisons.

	Slope Based on TPI	TWI	LULC	Curve number	Stream order	HAND	Priority	Rank
Slope Based on TPI	1	2	0.5	2	2	2	22.50%	1
TWI	0.5	1	1	1	1	0.5	12.60%	6
LULC	2	1	1	1	2	1	21.20%	2
Curve number	0.5	1	1	1	0.5	2	14.80%	3
Stream order	0.5	1	0.5	2	1	1	14.40%	5
HAND	0.5	2	1	0.5	1	1	14.50%	4

Weighted Overlay Process (WOP) within GIS

Potential runoff storage zones of the study area (Figure 5.32) was generated by integrating the thematic layers of slope, LULC, curve number, HAND, stream order and TWI using weighted

overlay process (WOP) within GIS. Resulted raster was classified into four classes namely (a) Not suitable (b) Marginally Suitable (c) Moderately Suitable (d) Optimally Suitable. Result shows that 17 % of the area is optimally suitable, 33.2% of the area is moderately suitable, 33.1 % of the area is marginally suitable and 18.7% of the area is not suitable for water storage zones/structures. Sixteen suitable sites on such zones (optimally suitable class) have also been identified for water storage structures, as shown in Figure 5.33. Criteria of selection of these sites are: first, proximity of the sites to the agricultural fields. Second, sites should be on unused or barren land. Third, narrow cross-section of the valley with high shoulders to minimise the amount of construction material needed for building the small dams or check dams, nala bunds, gully plug and bundhis. Results are also confirmed by the already built water storage structures in derived potential runoff storage zones which are in optimally suitable class (Figure 5.34).

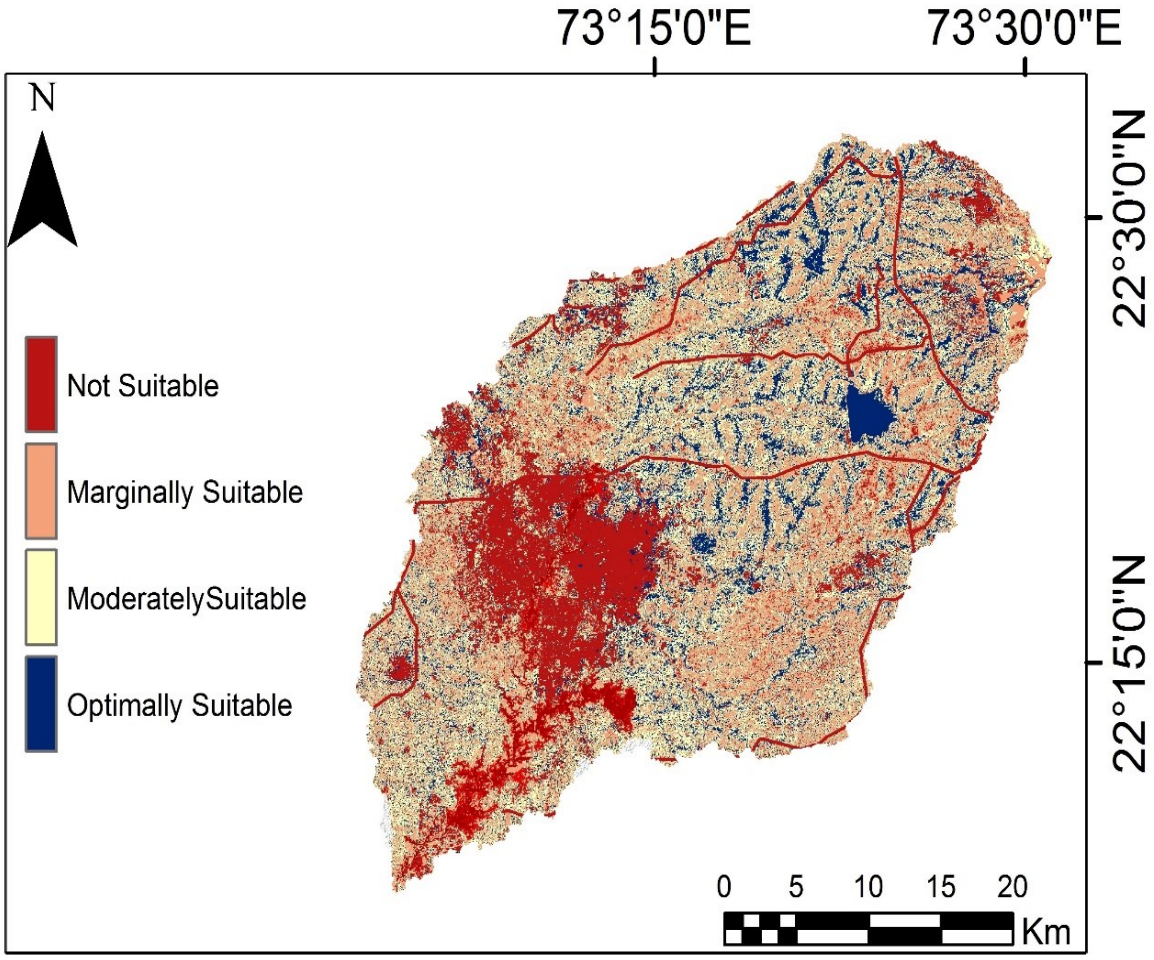


Figure 5.32: Potential runoff storage zones of the study area.

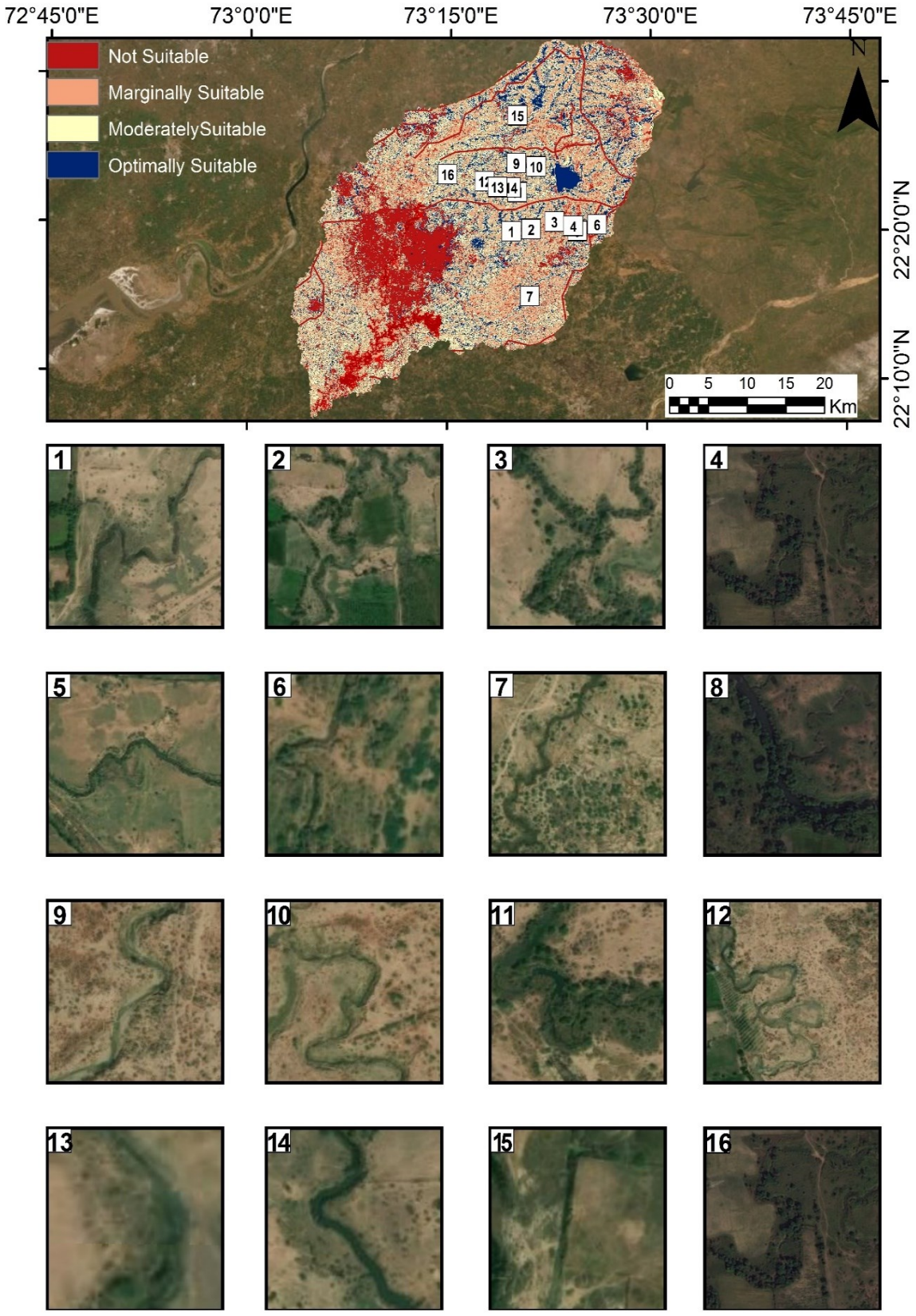


Figure 5.33: Identified sites for water storage structures on potential runoff storage zones.

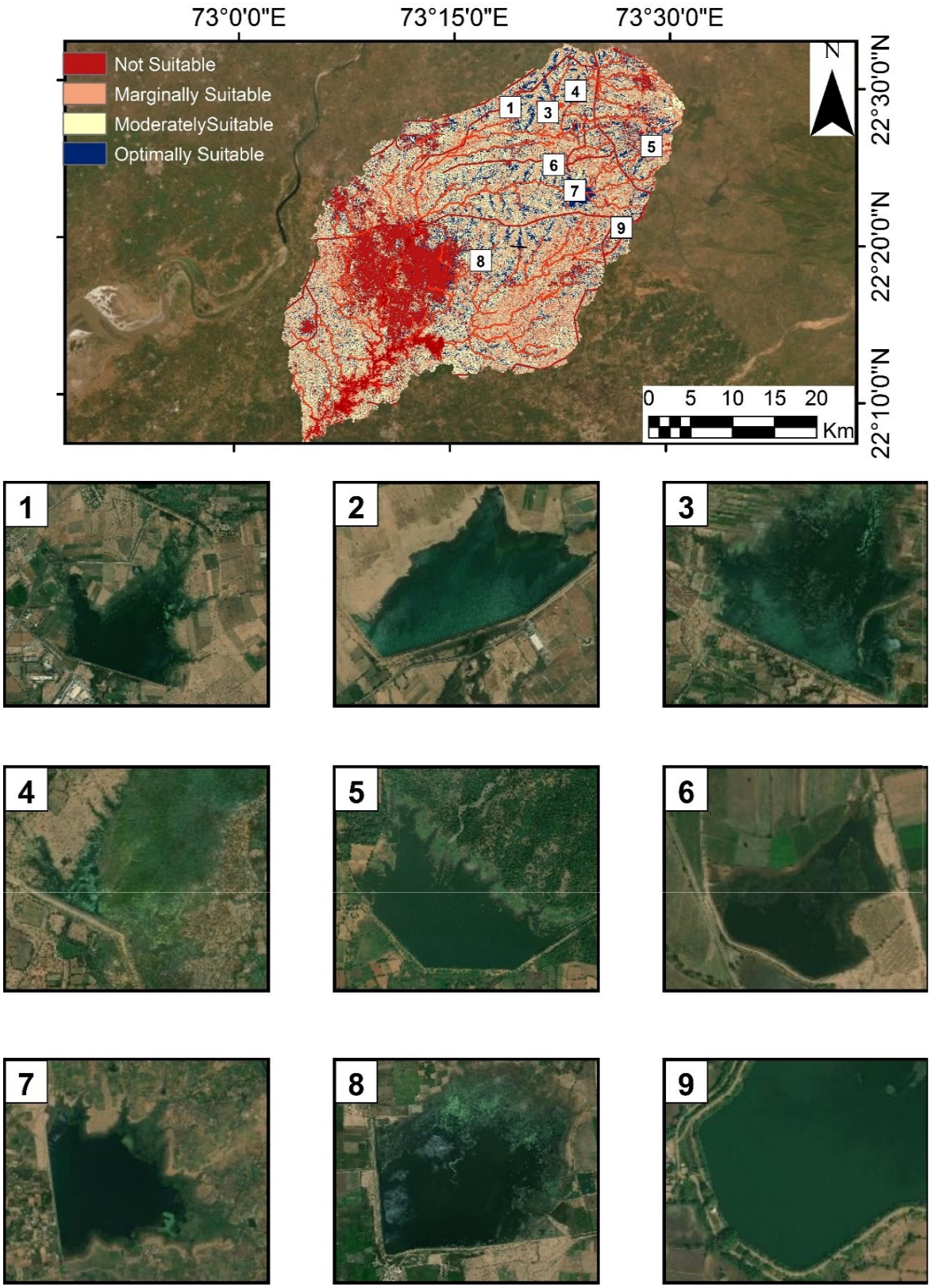


Figure 5.34: Already built water storage structures on the derived potential runoff storage zones.

5. **Objective**: To develop an approach for operational flood extent mapping using Synthetic Aperture Radar (SAR) and preparation of flood inundation map for data scarce region using 2D flow modelling using rain on grid model.

A. Speckle filtering

The performance efficiency of filters in speckle suppression, feature preservation, and preventing the loss of meaningful data was evaluated using Mean square error (MSE), Equivalent Number of Looks (ENL), Speckle suppression index (SSI), Speckle Mean Preservation Index (SMPI), examination of mean, standard deviation, and also by close visual assessment.

The MSE value indicates the amount of error present in a filtered image. It allows the comparison of the pixel values of a filtered image to the degraded image before filtering. It is observed that for study area Kerala (window size 3×3) Frost, Gamma map, and Lee sigma filters showed very low MSE, indicating their effectiveness in feature preservation. Median filter showed the highest MSE values indicating poor performance in terms of feature preservation. Boxcar and Lee filters performed moderately. For Kerala (window size 5×5) and Assam (window size 3×3 and 5×5), Frost and Lee sigma showed low MSE, Median filter showed the highest MSE and moderate performance was observed for Boxcar, Gamma map and Lee (Figure 5.35). The high value of MSE depicts a greater difference between the original test image and despeckled image which concludes the significant speckle reduction but at the cost of feature loss.

Lower SSI and SMPI values (Figures 5.36 and 5.37) were obtained for Median filter (3×3 and 5×5) in both the study areas indicating high efficiency in speckle suppression. Moderate values were obtained for Gamma map, Lee and Boxcar indicating medium proficiency in the reduction of speckle. None of the filters produced value 0 and 1. A value of 0, indicates complete mean preservation and 100 % noise reduction. Whereas, a value of 1 indicates 0 % speckle reduction. It was observed that as the window size increased from 3×3 to 5×5, the SSI values reduced for all the filters, indicating superior speckle suppression with an increase in the size of the moving window. The lowest SSI values, as well as SMPI values, corresponds to Median filter (3×3 and 5×5) indicating high efficiency in speckle suppression.

For both the study areas Median filter (3×3 and 5×5) had high ENL values (Figure 5.38) indicating a higher efficiency in smoothing speckle noise over homogeneous areas, which shows enhanced capacity to distinguish the distinct gray levels within the image. For constant flat areas where the sample variance is null, ENL becomes ∞, this will repute highly blurred data as excellent. According to ENL, SSI and SMPI values, Gamma map (3×3 and 5×5), Lee (3×3 and 5×5) and Boxcar (3×3 and 5×5) filters showed moderate performance for both the study areas.

The application of filters should ideally not bring about any change in the mean of target image while it should reduce the standard deviation. The filter, Median (3×3 and 5×5) for both the study areas was most effective in increasing the standard deviation but also changed the mean value considerably thereby implying that the filters reduced speckle but also caused considerable loss of meaningful data. The median filter is better than the Boxcar filter in terms of preserving the edges between two different features, but it does not preserve single pixel-wide features, which will be altered if speckle noise is present. Median filter preserves the texture information very well for small window size (3×3) but does not retain the mean value at an acceptable level. Since the median is less sensitive than the mean to extreme values (outliers), those extreme values are more effectively removed. On the contrary, Boxcar (3×3 and 5×5) for both the study areas, made the least change in the mean, while reducing the standard deviation moderately. However, Lee filter (3×3 and 5×5) provided a fair balance by reducing the standard deviation without drastically affecting the mean (Figures 5.39 and 5.40; Tables 5.17 and 5.18).

The Median filter performed well in terms of ENL, SSI, and SMPI values; however, its performance in terms of speckle reduction and feature preservation was far inferior compared to the Boxcar and Lee filter. However, since the objective of the speckle filtering was to use these SAR images for inundation mapping, the performance of the filters on a water body in terms of reduction of standard deviation while preserving the mean of the original image was the most important. The performance of Boxcar and Lee filters was far better in feature preservation in the filtered images followed by Gamma map. Although quantitative measures are often employed to compare different speckle suppression filters, it has been noted by Raouf & Lichtenegger, (1997) and others that visual inspection probably provides the best assessment of the performance of a speckle filter. Visual assessment is an easy and efficient way to investigate both the capability of a filter to suppress speckles and its effectiveness in preserving image details. Lee et al., (1994) stated that, in general, filters using small windows (such as 3 × 3) preserve texture information better. Visual examination was, therefore, carried out and it was observed that the filters which reduced speckles effectively also resulted in considerable loss of meaningful data (Figures 5.41 (a)-(m) and 5.42 (a)-(m)). Lee sigma, and Gamma map clearly resulted in the loss of edges and details. It was difficult to grade Boxcar, Frost and Lee filter visually, since the variation was not perceivable. However, it was observed that point scatters were over filtered, transformed to spread targets and sharp edges were generally blurred in Boxcar filter (Figure 5.43).

Hence, Lee filter was chosen for further analysis as it had low MSE, SSI and SMPI values compared with frost filter, and a higher percentage change in standard deviation compared to the Boxcar and frost filter for most of the cases.

The raw SAR data in VH and VV polarization acquired during the crisis events on 21 August 2018 and 14 Jul 2019 for study areas Kerala and Assam are shown in Figure 5.44 (a) and (b), and Figure 5.45 (a) and (b), respectively. The multi-looked, calibrated, filtered (Lee) data of the VH and VV are shown in Figure 5.44 (c) and (d), and Figure 5.45 (c) and (d), respectively. Multi-looked, calibrated, filtered (Lee) SAR data were not projected on the map coordinates of each pixel. The pixel was in the original coordinate position of data (rows/columns) in the field of ground range. In the orthorectified imageries, each of the pixels that were corrected and projected using the Range-Doppler terrain correction appeared at the actual position. The Range-Doppler terrain corrected pixels and their respective histograms for Kerala and Assam regions are shown in Figures 5.44 (e)–(h) and 5.45 (e)–(h), respectively.

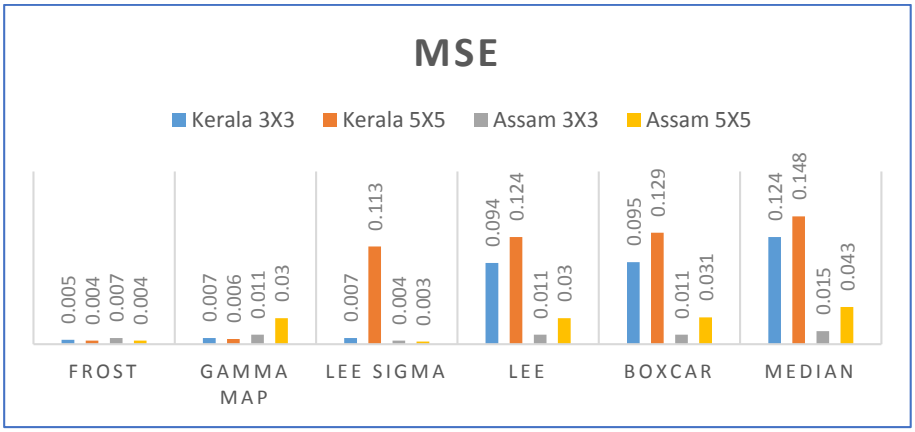


Figure 5.35: Mean Square Error.

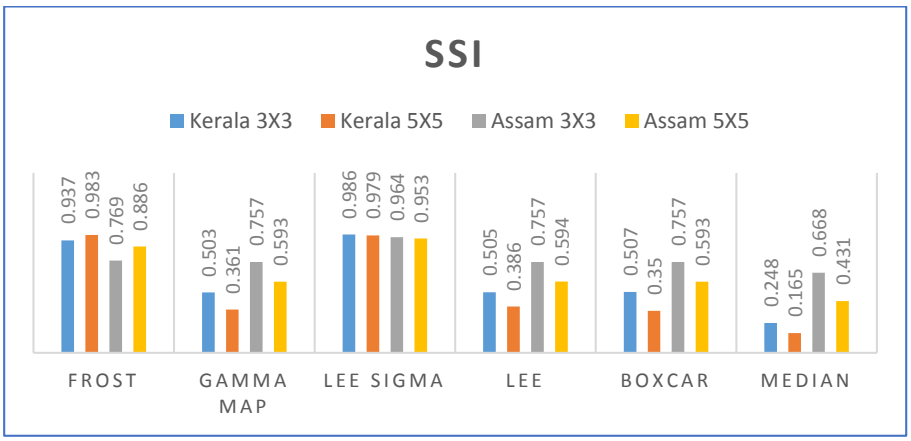


Figure 5.36: Speckle Suppression Index.

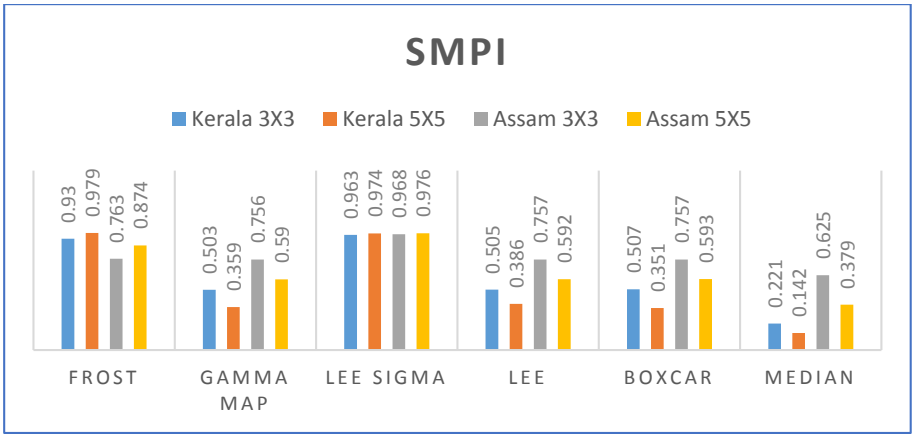


Figure 5.37: Speckle Mean Preservation Index.

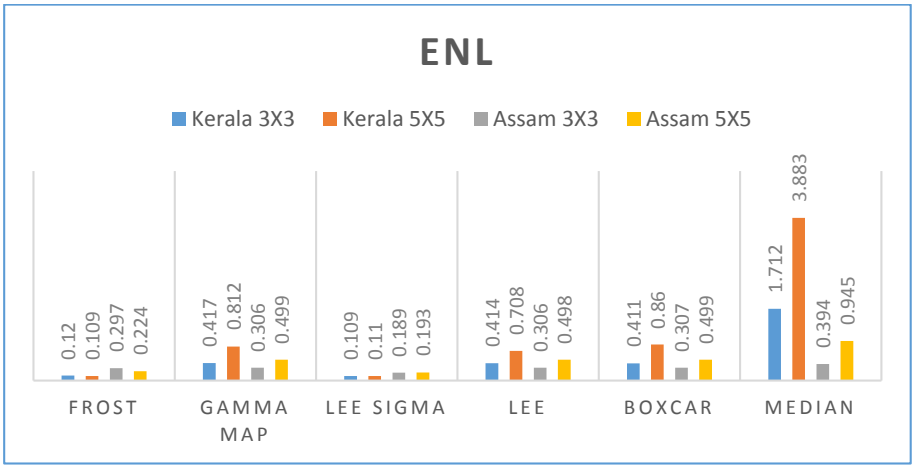


Figure 5.38: Equivalent Number of Looks.

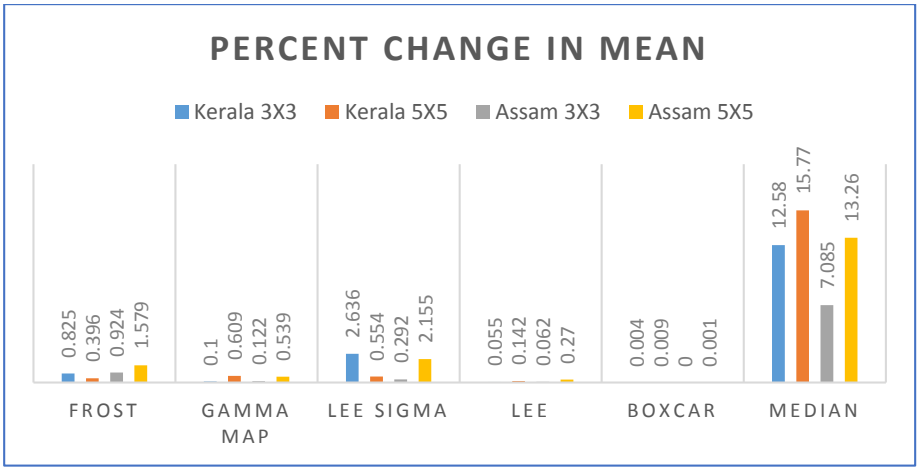


Figure 5.39: Percent change in mean.

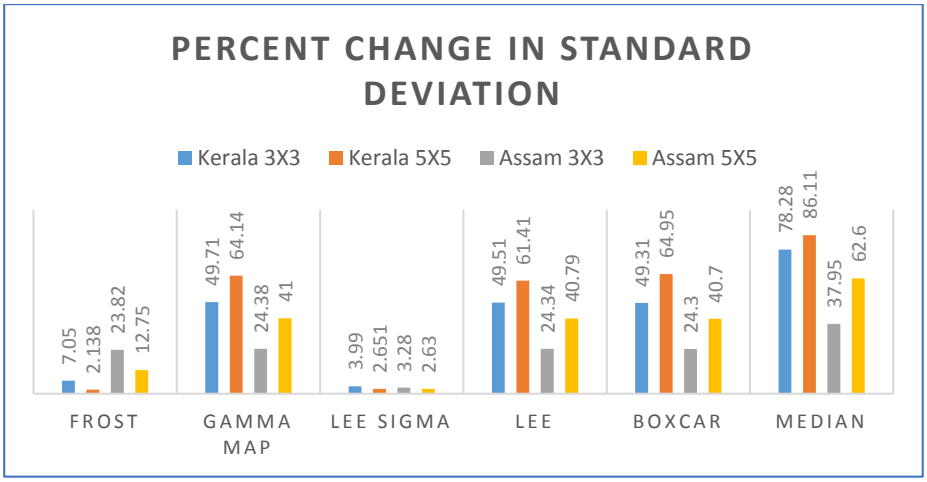


Figure 5.40: Percent change in standard deviation.

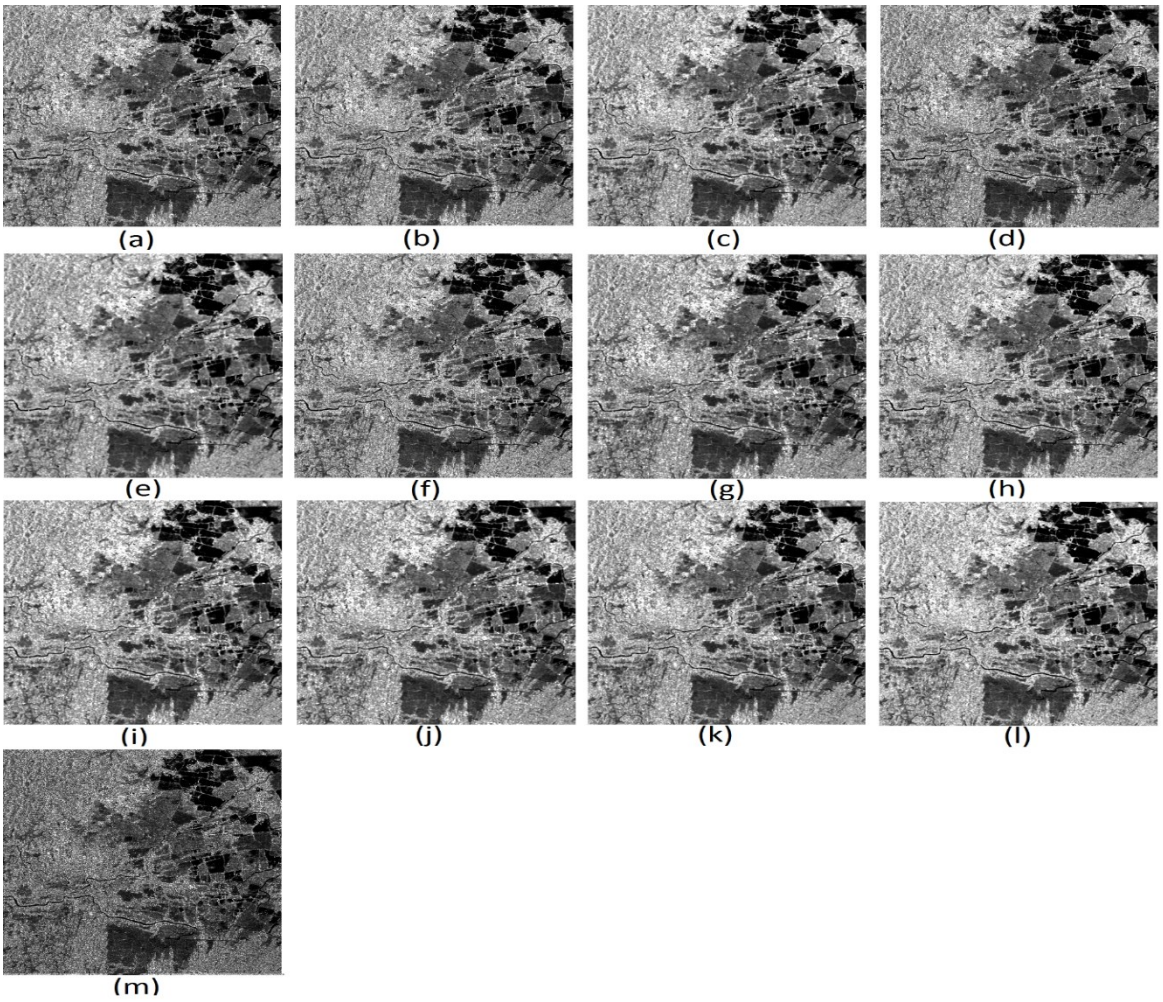


Figure 5.41: Visual comparison of de-noising methods on VV polarization over Kerala test data (a) Boxcar 3×3 (b) Frost 3×3 (c) Gamma map 3×3 (d) Lee 3×3 (e) Lee sigma 3×3 (f) Median 3×3 (g) Boxcar 5×5 (h) Frost 5×5 (i) Gamma map 5×5 (j) Lee 5×5 (k) Lee sigma 5×5 (l) Median 5×5 and (m) Unfiltered test image in VV polarization.

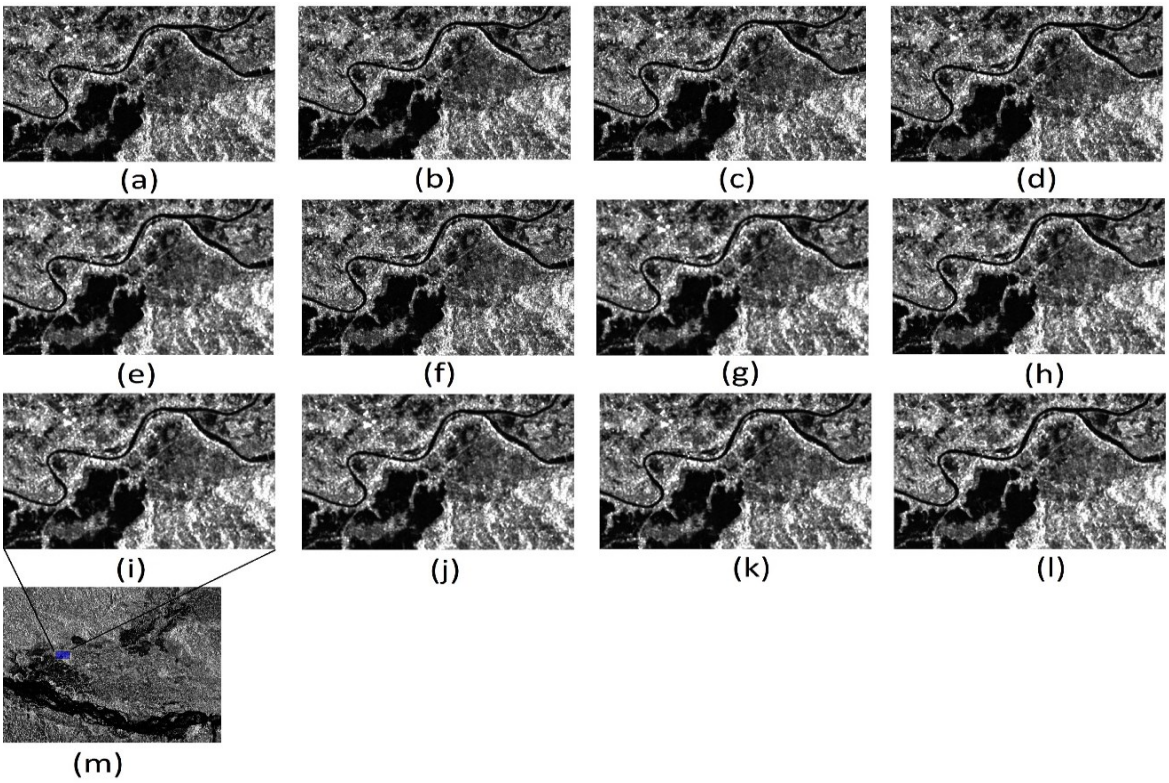


Figure 5.42: Visual comparison of de-noising methods on VV polarization over Assam test data (a) Boxcar 3×3 (b) Frost 3×3 (c) Gamma map 3×3 (d) Lee 3×3 (e) Lee sigma 3×3 (f) Median 3×3 (g) Boxcar 5×5 (h) Frost 5×5 (i) Gamma map 5×5 (j) Lee 5×5 (k) Lee sigma 5×5 (l) Median 5×5 and (m) Unfiltered test image in VV polarization.

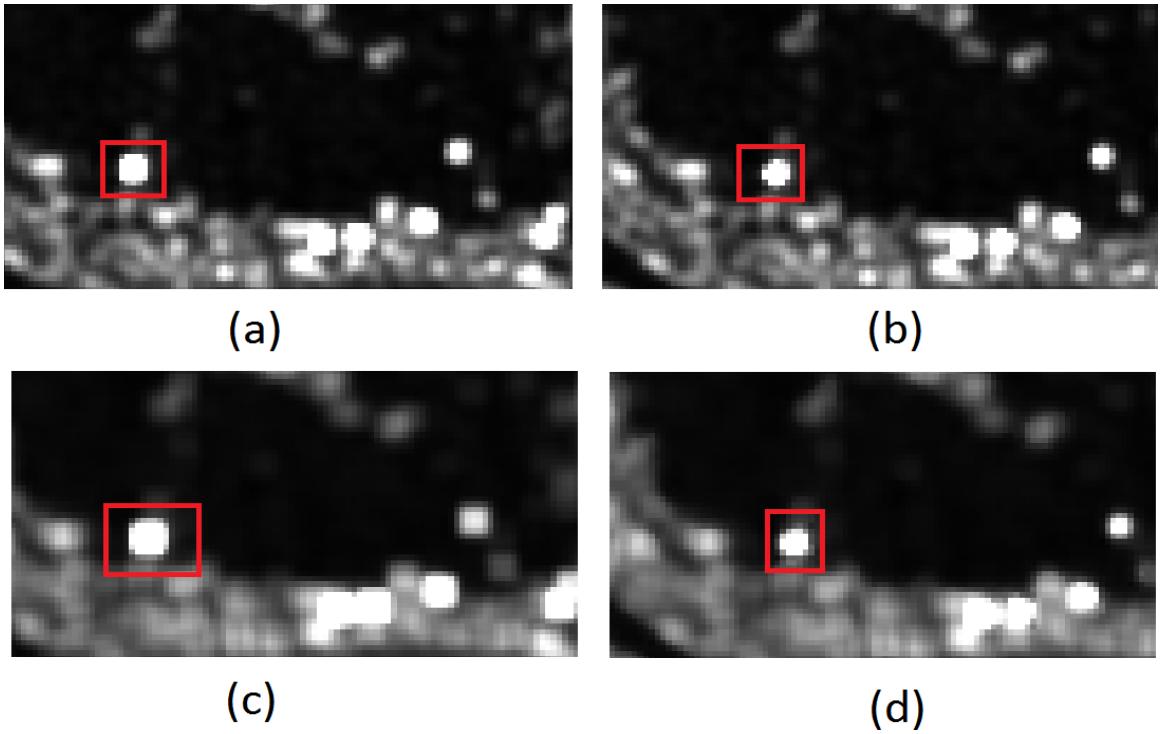


Figure 5.43: Point target scattering on (a) Boxcar 3×3 (b) Lee 3×3 (c) Boxcar 5×5 and (d) Lee 5×5.

Table 5.17: Quantitative evaluation of the filters over Kerala.

	Kerala											
	MSE		SSI		SMPI		ENL		Percent Change – Mean		Percent Change – Standard Deviation	
Filters	3×3	5×5	3×3	5×5	3×3	5×5	3×3	5×5	3×3	5×5	3×3	5×5
Frost	0.005	0.004	0.937	0.983	0.930	0.979	0.120	0.109	0.825	0.396	7.05	2.138
Gamma map	0.007	0.006	0.503	0.361	0.503	0.359	0.417	0.812	0.100	0.609	49.71	64.14
Lee sigma	0.007	0.113	0.986	0.979	0.963	0.974	0.109	0.110	2.636	0.554	3.99	2.651
Lee	0.094	0.124	0.505	0.386	0.505	0.386	0.414	0.708	0.055	0.142	49.51	61.41
Boxcar	0.095	0.129	0.507	0.350	0.507	0.351	0.411	0.860	0.004	0.009	49.31	64.95
Median	0.124	0.148	0.248	0.165	0.221	0.142	1.712	3.883	12.58	15.77	78.28	86.11

Table 5.18: Quantitative evaluation of the filters over Assam.

	Assam											
	MSE		SSI		SMPI		ENL		Percent Change – Mean		Percent Change – Standard Deviation	
Filters	3×3	5×5	3×3	5×5	3×3	5×5	3×3	5×5	3×3	5×5	3×3	5×5
Frost	0.007	0.004	0.769	0.886	0.763	0.874	0.297	0.224	0.924	1.579	23.82	12.75
Gamma map	0.011	0.030	0.757	0.593	0.756	0.590	0.306	0.499	0.122	0.539	24.38	41.00
Lee sigma	0.004	0.003	0.964	0.953	0.968	0.976	0.189	0.193	0.292	2.155	3.28	2.63
Lee	0.011	0.030	0.757	0.594	0.757	0.592	0.306	0.498	0.062	0.270	24.34	40.79
Boxcar	0.011	0.031	0.757	0.593	0.757	0.593	0.307	0.499	0.000	0.001	24.30	40.70
Median	0.015	0.043	0.668	0.431	0.625	0.379	0.394	0.945	7.085	13.26	37.95	62.60

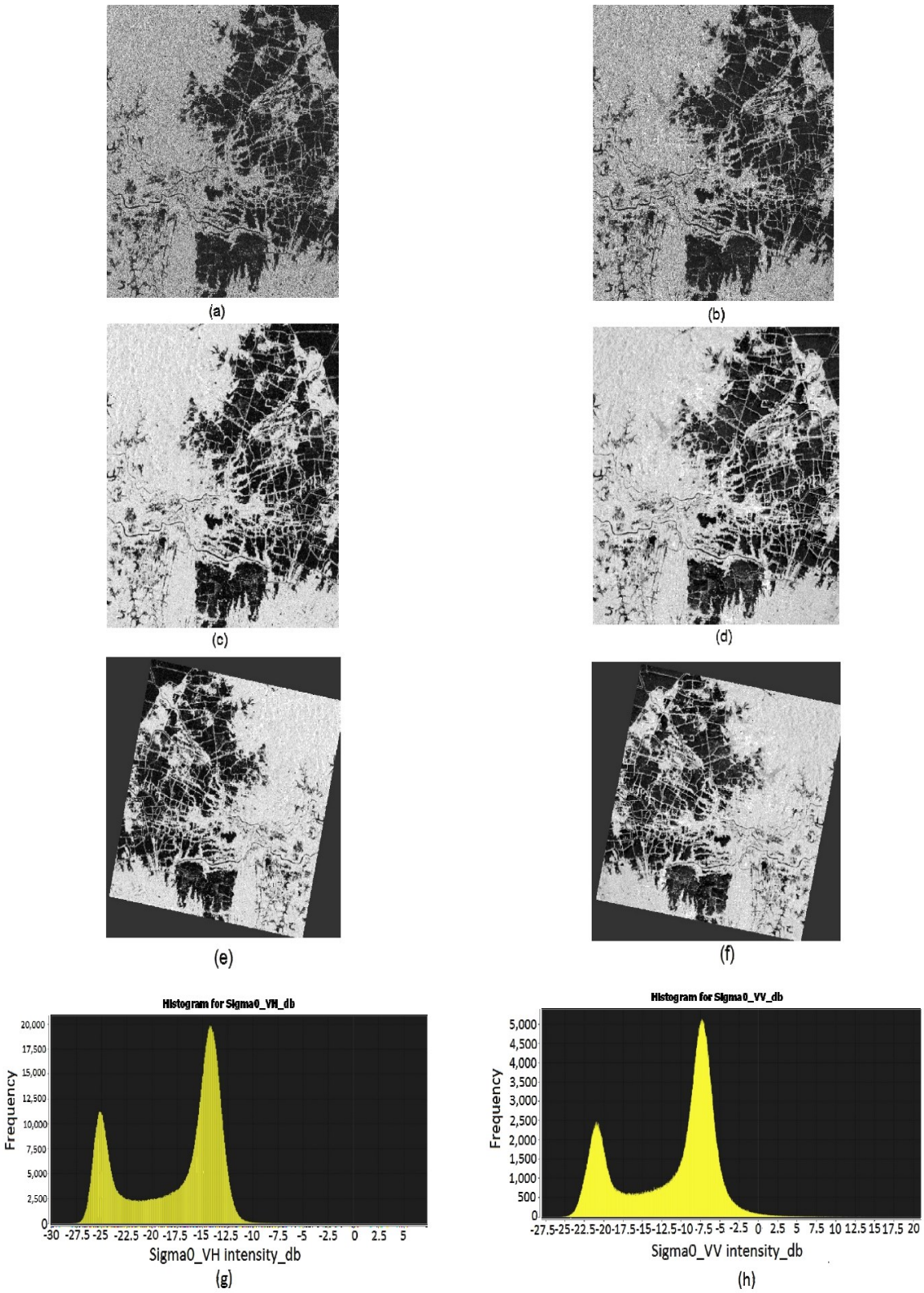


Figure 5.44: Kerala (a) Raw VH amplitude data; (b) Raw VV amplitude data; (c) Multi-looked, calibrated, Filtered (Lee) VH data in dB; (d) Multi-looked, calibrated, Filtered (Lee) VV data in dB; (e) The Range-Doppler terrain corrected VH data; (f) The Range-Doppler terrain corrected VV data; (g) Histogram for Sigma0 VH in dB, and (h) Histogram for Sigma0 VV in dB.

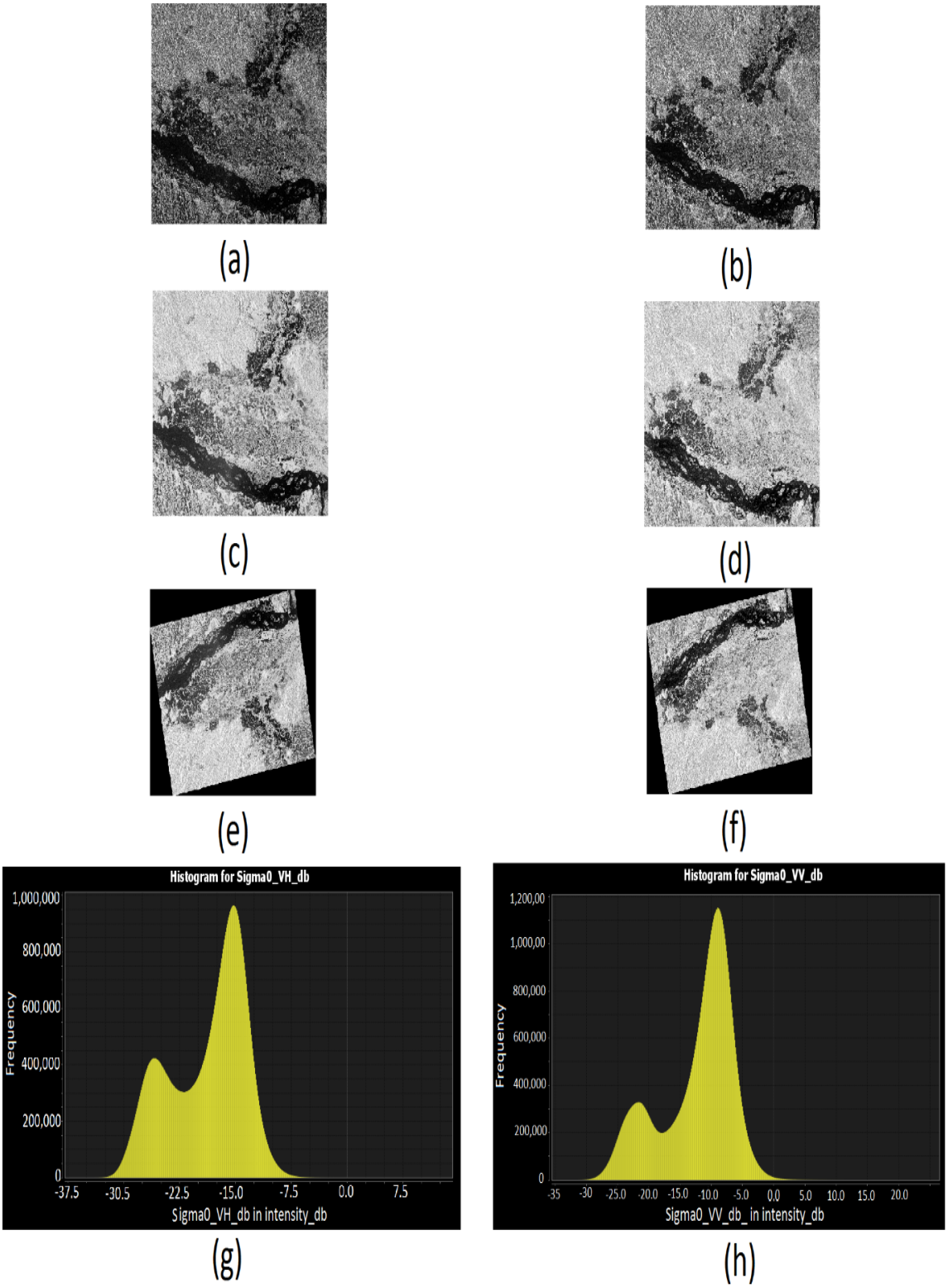


Figure 5.45: Assam (a) Raw VH amplitude data; (b) Raw VV amplitude data; (c) Multi-looked, calibrated, Filtered (Lee) VH data in dB; (d) Multi-looked, calibrated, Filtered (Lee) VV data in dB; (e) The Range-Doppler terrain corrected VH data; (f) The Range-Doppler terrain corrected VV data; (g) Histogram for Sigma0 VH in dB, and (h) Histogram for Sigma0 VV in dB.

Table 5.19 and 5.21, and Figure 5.46 and 5.47 show the comparison of classification results of random forest classifier and support vector machine classifier for VV and VH polarization. The training data were kept the same for both the classifiers to avoid optimistic bias in the classification. For the study area Kerala, the random forest classifier exhibited maximum overall accuracy of 88.80% with the kappa coefficient value of 0.72. Both the classifiers obtained better accuracy results in VV polarization compared to the VH polarization. The least overall accuracy of 82.60% and a kappa coefficient value of 0.63 were observed with random forest in VH polarization, which was followed by the support vector machine in VV polarization. RF achieved higher classification accuracy than SVM by about 5% in VV polarization. However, both the classifiers produced comparable overall accuracies in VH polarization (SVM achieved higher classification accuracy than RF by about 1%). The NDWI calculated for the cloud-free extent is shown in Figure 5.48 (d). The inundated area in the calculated NDWI over the cloud-free extent is 73.88%, which is 41.78 km^2. However, it has also been observed (Table 5.20) that the inundated area using random forest classification on filtered VV data over the cloud-free extent is 71.18%, which is 40.25 Km^2. For the study area Assam, the SVM classifier exhibited maximum overall accuracy of 92% with the kappa coefficient value of 0.81. Both the classifiers obtained better accuracy results in VH polarization compared to the VV polarization. The least overall accuracy of 83.60% and a kappa coefficient value of 0.65 were observed with random forest in VV polarization, which was followed by the RF in VH polarization. SVM achieved higher classification accuracy than RF by about 5.38% in VH polarization. The NDWI calculated for the cloud-free extent is shown in Figure 5.49 (d). The inundated area in the calculated NDWI over the cloud-free extent is 74.09%, which is 491.47 km^2. However, it has also been observed (Table 5.22) that the inundated area using SVM classification on filtered VH data over the cloud-free extent is 62.76%, which is 416.99Km^2. To calculate inundation for the entire scene, threshold value of -10.96 and -19.58 was selected for Kerala and Assam region respectively after analysing the histograms. The calculated inundated area (Figure 5.50) with thresholding technique for Kerala was found 204 km^2 (2% more than classified VV polarised data using RF algorithm). Similarly, the calculated inundated area with thresholding technique for Assam was found 3368.90 km^2 (23.46% more than classified VH polarised data using SVM algorithm). For Assam region the variation is very large, a single threshold did not hold well as large swath of a SAR image suffers from environment heterogeneity caused by wind-roughening and satellite framework parameters.

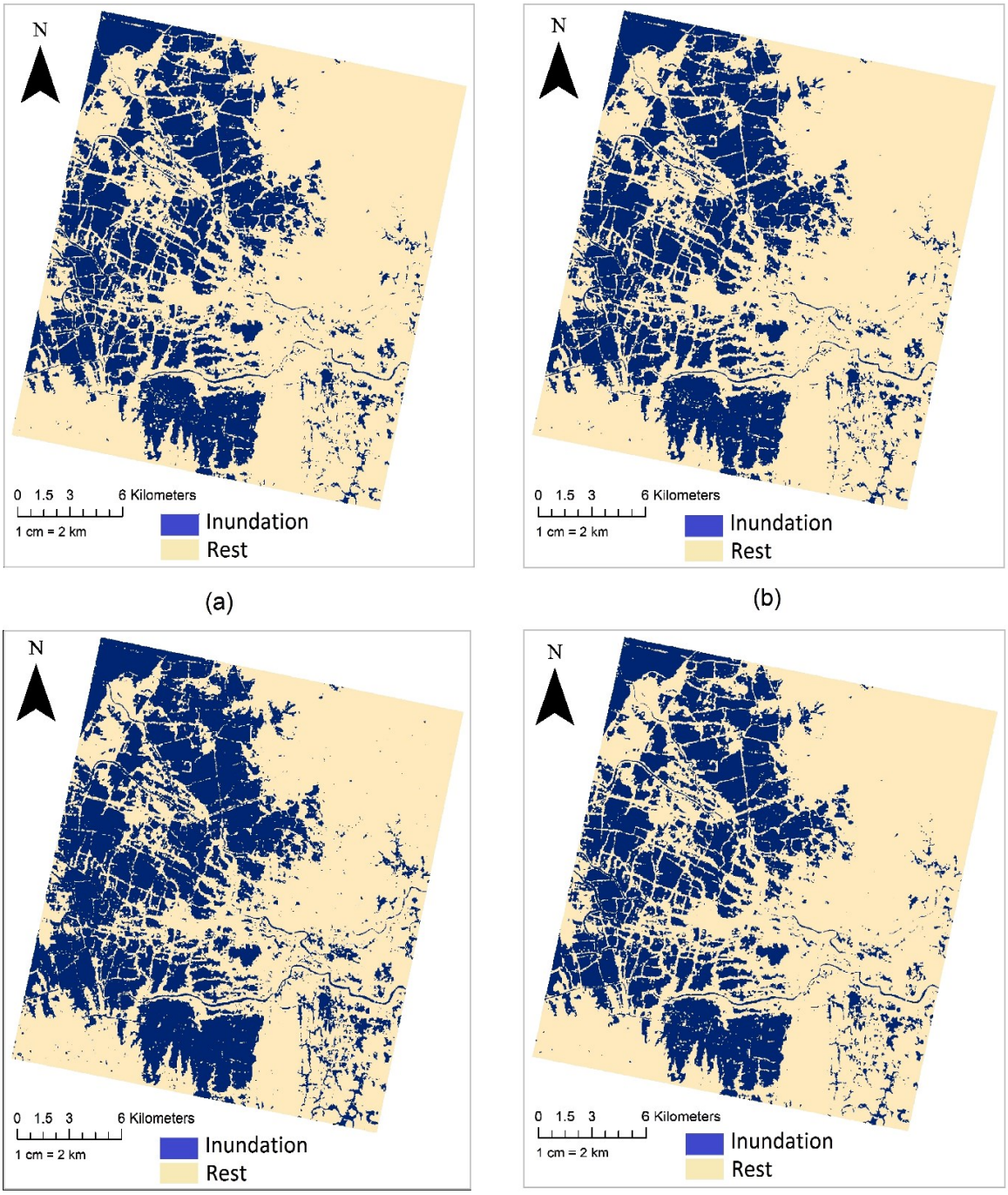


Figure 5.46: Kerala (a) Random forest tree classification on filtered VH; (b) Support vector machine classification on VH; (c) Random forest tree classification on filtered VV; (d) Support vector machine classification on VV.

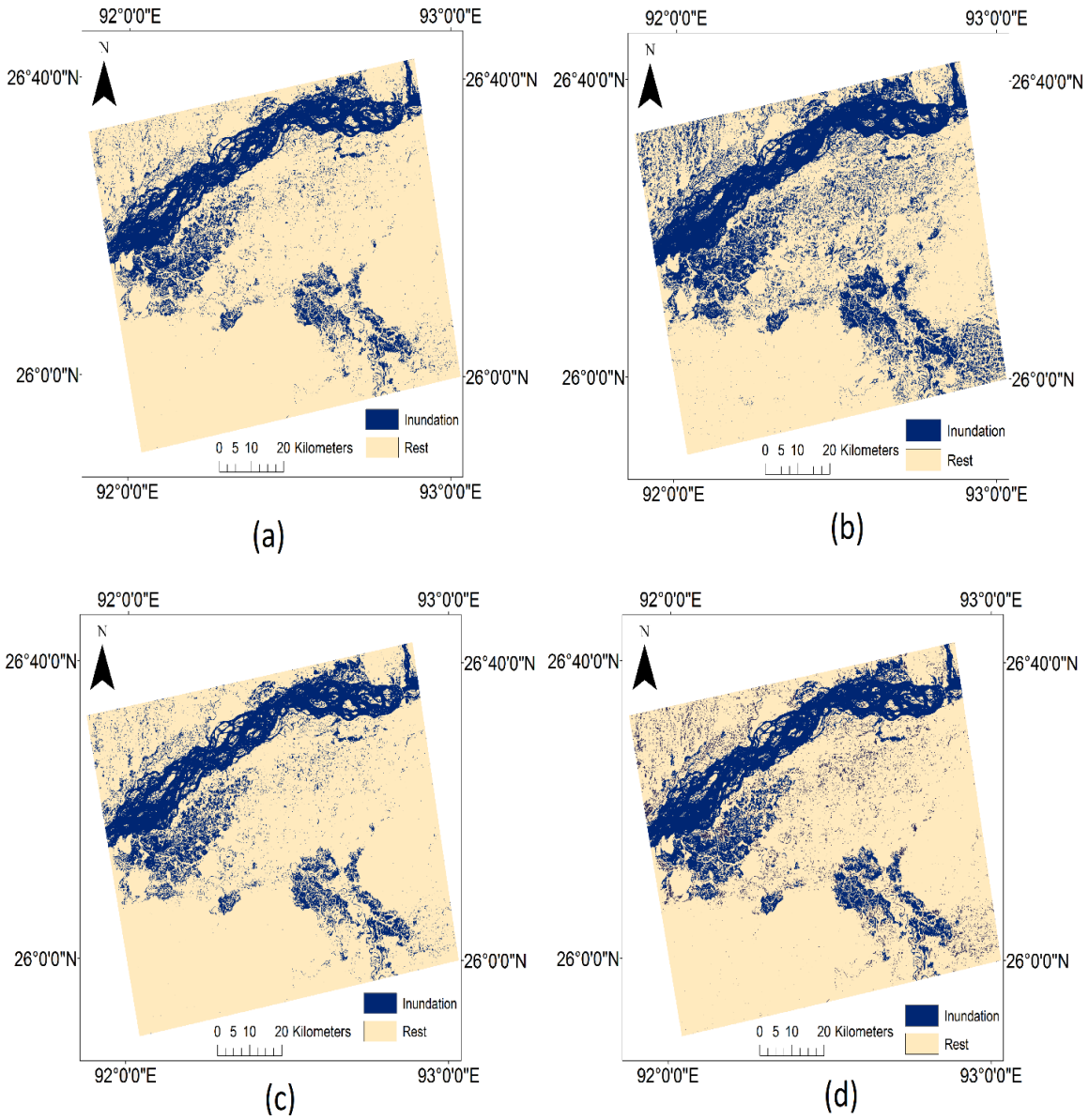


Figure 5.47: Assam (a) Random forest tree classification on filtered VH; (b) Support vector machine classification on VH; (c) Random forest tree classification on filtered VV; (d) Support vector machine classification on VV.

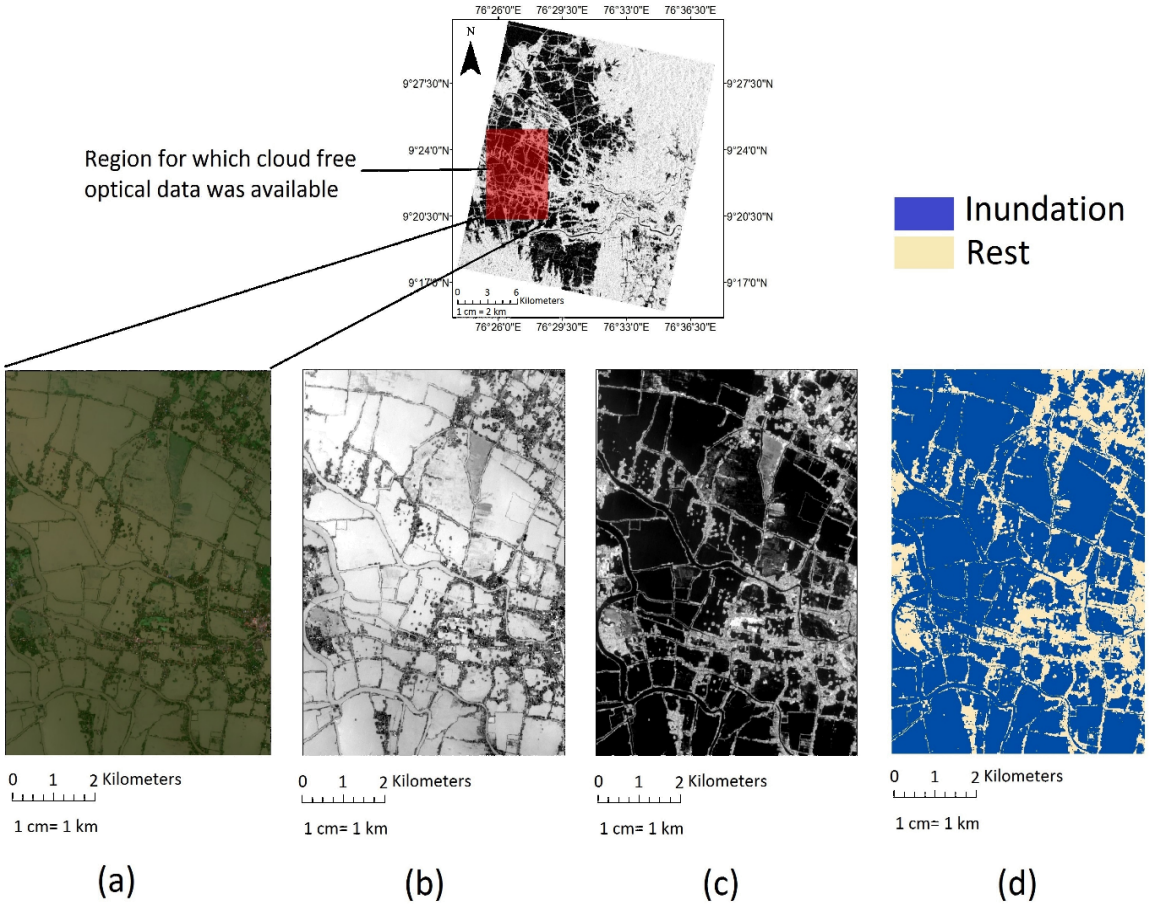


Figure 5.48: (a) Natural colour composite R-B04 G-B03 B-B02, (b) Green band (c) Near-infrared (d) Calculated NDWI over cloud-free extent for Kerala region.

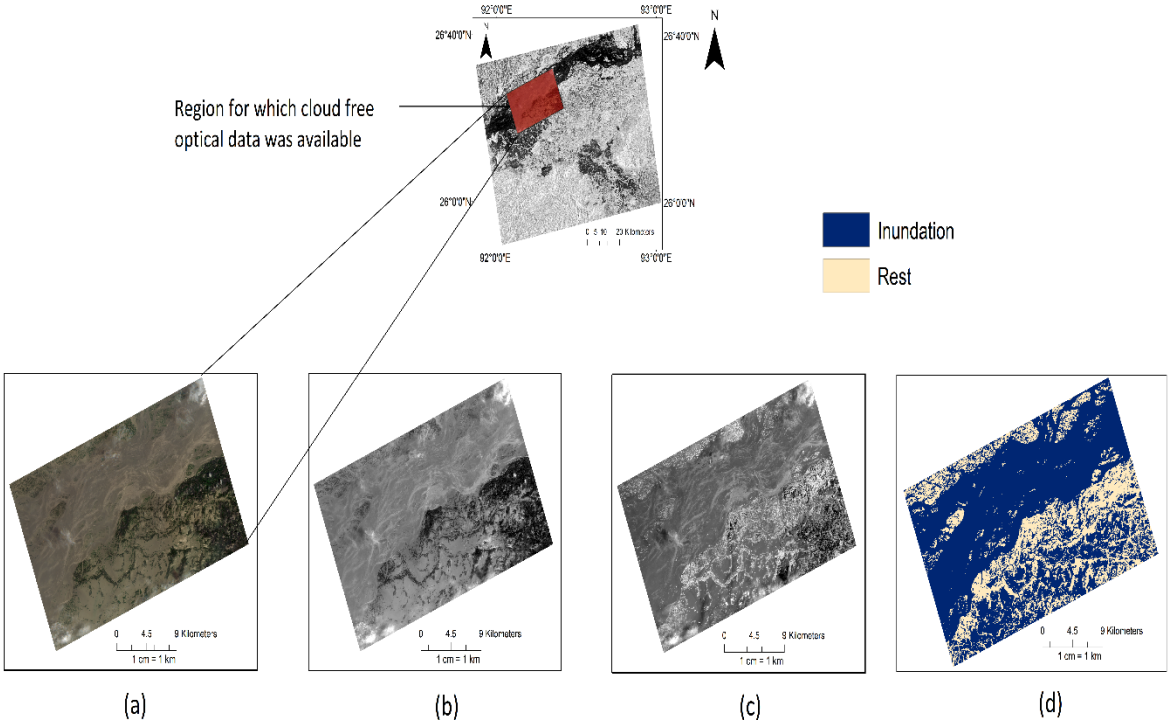


Figure 5.49: (a) Natural colour composite R-B04 G-B03 B-B02, (b) Green band (c) Near-infrared (d) Calculated NDWI over cloud-free extent for Assam region.

Table 5.19: Comparison of user's accuracy (UA), producer's accuracy (PA), overall accuracy (%), and kappa coefficient using random forest tree and support vector machine algorithms for VV and VH polarization over Kerala region.

	VV Polarization				VH Polarization			
	RF		SVM		RF		SVM	
	PA	UA	PA	UA	PA	UA	PA	UA
Inundation	0.89	0.96	0.79	0.99	0.78	0.99	0.79	0.99
Rest	0.88	0.72	0.98	0.60	0.98	0.58	0.98	0.60
Kappa coefficient	0.72		0.64		0.61		0.63	
Overall Accuracy (%)	88.80%		83.80%		82.60%		83.60%	

Note: UA - User's accuracy, PA - Producer's accuracy VV - Vertical-Vertical, VH - Vertical-Horizontal, RF - Random Forest, SVM -Support Vector Machine

Table 5.20: Inundated area statistics of RF and SVM over cloud-free optical data for Kerala region.

	VV Polarization		VH Polarization		NDWI
	RF	SVM	RF	SVM	-
Inundated area (Km^2)	40.25	34.82	34.17	35.19	41.78
Rest (Km^2)	16.29	21.72	22.38	21.36	14.77

Table 5.21: Comparison of user's accuracy (UA), producer's accuracy (PA), overall accuracy (%), and kappa coefficient using random forest tree and support vector machine algorithms for VV and VH polarization over Assam region.

	VV Polarization				VH Polarization			
	RF		SVM		RF		SVM	
	PA	UA	PA	UA	PA	UA	PA	UA
Inundation	0.77	1.0	0.82	1.0	0.81	0.99	0.89	0.99
Rest	1.0	0.62	1	0.69	0.99	0.67	0.99	0.77
Kappa coefficient	0.65		0.72		0.70		0.81	
Overall Accuracy (%)	83.60%		87.60%		86.60%		92.00%	

Note: UA - User's accuracy, PA - Producer's accuracy VV - Vertical-Vertical, VH - Vertical-Horizontal, RF - Random Forest, SVM -Support Vector Machine

Table 5.22: Inundated area statistics of RF and SVM over cloud-free optical data for Assam region.

	VV Polarization		VH Polarization		NDWI
	RF	SVM	RF	SVM	-
Inundated area (Km^2)	363.92	389.21	387.62	416.99	491.47
Rest (Km^2)	300.41	275.12	276.70	247.34	171.86

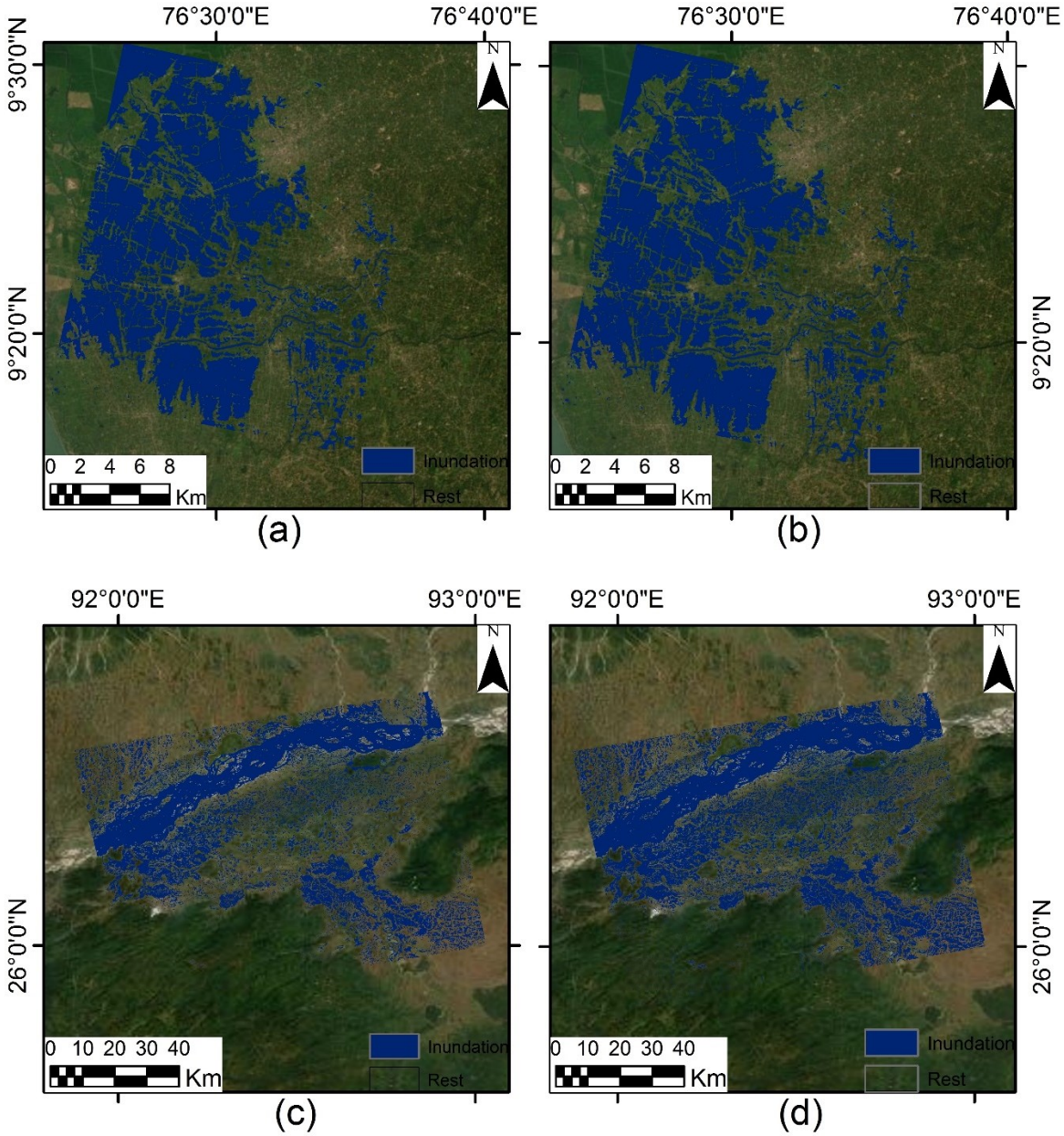


Figure 5.50: (a) Random forest tree classification on filtered VV data (Kerala) (b) Classified filtered VV data with threshold value of -10.96 (Kerala) (c) Support vector machine classification on filtered VH data (Assam) (d) Classified filtered VH data with threshold value of -19.58 (Assam).

B. 2D hydraulic modelling for inundation:

HEC-RAS 5.0.7 only supports a single representative rainfall record. Consequently, a single representative daily rainfall record was created. Thiessen polygon of rain gauge stations are shown in Figure 5.23. For M stations, the average precipitation $\bar{P}$ is calculated as

$$\bar{P} = \sum_{i=1}^{M} P_i \frac{A_i}{A} \qquad 5.4$$

The ratio $\frac{A_i}{A}$ is called weightage factor for each station. Precipitation data, thiessen weightage factor and calculation of thiessen mean for storm event occurred on 30-07-2019 to 03-08-2019 are shown in Tables 5.23 and 5.24.

Table 5.23: Precipitation data and Thiessen weightage factor.

Rain gauge station	Thiessen polygon area (Km^2)	Thiessen weightage factor(fraction of total area)	Station reading (mm) on 30-07-2019	Station reading (mm) on 31-07-2019	Station reading (mm) on 01-08-2019	Station reading (mm) on 02-08-2019	Station reading (mm) on 03-08-2019
Padra	92	0.07	6	78	8	18	27
Savli	178	0.14	14	36	80	4	88
Vadodara	504	0.39	37	499	32	34	101
Waghodia	512	0.40	8	126	12	22	65

Table 5.24: Calculation of Thiessen mean.

Rain gauge station	Weighted station rainfall (mm) on 30-07-2019	Weighted station rainfall (mm) on 31-07-2019	Weighted station rainfall (mm) on 01-08-2019	Weighted station rainfall (mm) on 02-08-2019	Weighted station rainfall (mm) on 03-08-2019
Padra	0.43	5.57	0.57	1.29	1.93
Savli	1.94	4.98	11.07	0.55	12.17
Vadodara	14.52	195.77	12.55	13.34	39.62
Waghodia	3.18	50.14	4.78	8.75	25.87
Thiessen mean of the station readings (P) in mm	20.06	256.46	28.97	23.93	79.59

Low CN values indicate low runoff potential while larger numbers indicate an increased runoff potential. The calculated curve number is also termed as CN II for AMC II (Antecedent Moisture Condition II). The calculated curve number can be adjusted to dry moisture conditions (called as AMC I) and high moisture conditions (called as AMC III) by using adjusting factors. The calculated weighted curve number of the Vishwamitri watershed for AMC I, AMC II and AMC III are 68.99, 84.04 and 92.50, respectively. Empirical equations of daily runoff for Vishwamitri watershed for AMC I, AMC II and AMC III conditions are:

$$Q(\text{mm}) = \frac{(P - 22.83)^2}{(P + 114.16 - 22.83)} \text{ for AMC I} \quad 5.5$$

$$Q(\text{mm}) = \frac{(P - 9.64)^2}{(P + 48.23 - 9.64)} \text{ for AMC II} \quad 5.6$$

$$Q(\text{mm}) = \frac{(P - 4.11)^2}{(P + 20.59 - 4.11)} \text{ for AMC III} \quad 5.7$$

Estimated Daily runoff for the period 30-07-2019 to 03-08-2019 using weighted CN is given in Table 5.25. Variation of curve number across the Vishwamitri watershed for AMC I, AMC II and AMC III conditions are shown in Figure 5.19.

Table 5.25: Estimated Daily runoff for the period 30-07-2019 to 03-08-2019 using weighted CN.

Date	Cumulative time in Hrs	Δt in Hrs	Incremental rainfall in mm	Maximum potential retention (S) in mm	Initial abstraction $I_a = \lambda S$	Daily runoff (Q) in mm
30-07-2019	24	24	20.06	48.24	9.65	1.85
31-07-2019	48	24	256.46	48.24	9.65	206.46
01-08-2019	72	24	28.97	20.60	4.12	13.59
02-08-2019	96	24	23.93	20.60	4.12	9.71
03-08-2019	120	24	79.59	20.60	4.12	59.29

ESTIMATION OF FLOOD FREQUENCY AT KALAGHODA BY GUMBEL'S METHOD

Average of past flood data ($\bar{X}$)

$\bar{X}$ = Summation of peak flood/ number of peak flood

= 4715/18

= 261.94 Cumecs

Standard deviation (σ):

$$\sigma = \sqrt{\frac{\sum(X - \bar{X})^2}{N - 1}} \qquad 5.8$$

Where, N = number of peak flood of past year

$$= \sqrt{\frac{975256.94}{17}}$$

$=$ 239.51 Cumecs

For return period T=15 years

Y_T=reduced variable, a function of T and is given by

$$= -\left[ln.\, ln\frac{T}{T-1}\right] \qquad 5.9$$

$$= -\left[ln.\, ln\frac{15}{15-1}\right]$$

$=$ 2.67

The value of reduced mean (Y_T) and reduced standard deviation (S_N) is taken from the Table 5.26 and Table 5.27 respectively.

Table 5.26: Reduced Mean $\bar{Y}_n$ in Gumbel's Extreme Value Distribution.

N	0	1	2	3	4	5	6	7	8	9
10	0.4952	0.4996	0.5035	0.5070	0.5100	0.5128	0.5157	0.5181	0.5202	0.5220
20	0.5236	0.5252	0.5268	0.5283	0.5296	0.5309	0.5320	0.5332	0.5343	0.5353
30	0.5362	0.5371	0.5380	0.5388	0.5396	0.5402	0.5410	0.5418	0.5424	0.5430
40	0.5436	0.5442	0.5448	0.5453	0.5458	0.5463	0.5468	0.5473	0.5477	0.5481
50	0.5485	0.5489	0.5493	0.5497	0.5501	0.5504	0.5508	0.5511	0.5515	0.5518
60	0.5521	0.5524	0.5527	0.5530	0.5533	0.5535	0.5538	0.5540	0.5543	0.5545
70	0.5548	0.5550	0.5552	0.5555	0.5557	0.5559	0.5561	0.5563	0.5565	0.5567
80	0.5569	0.5570	0.5572	0.5574	0.5576	0.5578	0.5580	0.5581	0.5583	0.5585
90	0.5586	0.5587	0.5589	0.5591	0.5592	0.5593	0.5595	0.5596	0.5598	0.5599
100	0.5600									

Table 5.27: Reduced Standard Deviation S_n in Gumbel's Extreme Value Distribution.

N	0	1	2	3	4	5	6	7	8	9
10	0.9496	0.9676	0.9833	0.9971	1.0095	1.0206	1.0316	1.0411	1.0493	1.0565
20	1.0628	1.0696	1.0754	1.0811	1.0864	1.0915	1.0961	1.1004	1.1047	1.1086
30	1.1124	1.1159	1.1193	1.1226	1.1255	1.1285	1.1313	1.1339	1.1363	1.1388
40	1.1413	1.1436	1.1458	1.1480	1.1499	1.1519	1.1538	1.1557	1.1574	1.1590
50	1.1607	1.1623	1.1638	1.1658	1.1667	1.1681	1.1696	1.1708	1.1721	1.1734
60	1.1747	1.1759	1.1770	1.1782	1.1793	1.1803	1.1814	1.1824	1.1834	1.1844
70	1.1854	1.1863	1.1873	1.1881	1.1890	1.1898	1.1906	1.1915	1.1923	1.1930
80	1.1938	1.1945	1.1953	1.1959	1.1967	1.1973	1.1980	1.1987	1.1994	1.2001
90	1.2007	1.2013	1.2020	1.2026	1.2032	1.2038	1.2044	1.2049	1.2055	1.2060
100	1.2065									

$\bar{Y}_n = 0.5202$

$S_n = 1.0493$

Frequency Factor (K)

$$K = \frac{Y_T - \overline{Y_n}}{S_n} \quad (5.10)$$

$$K = \frac{2.67 - 0.5202}{1.0493}$$

K=2.049

Magnitude of flood (X_T)

$$X_T = \bar{X} + K\sigma \quad (5.11)$$

$$X_{15} = 261.94 + 2.049(239.51)$$

$$X_{15} = 752.69 \text{ cumecs}$$

Due to the unavailability of the cloud free optical data or Synthetic Aperture Radar data and discharge data at the outlet of Vishwamitri watershed for the storm event. The validation of the inundation map was done by field visits for the Vadodara city. The simulation shows good performance when comparing the simulation results with actual field data. Simulated peak flood for the storm event at Kalaghoda Bridge is shown in Figure 5.51. The simulated result shows the peak flood of 884.729 m^3/sec at Kalaghoda Bridge on 01-08-2019. Flood frequency analysis of observed flood peak data from (1996 to 2013) Table 5.28 at Kalaghoda Bridge using Gumbel's

method shows that the simulated peak discharge has return period close to 25 years (872.69 m^3/sec).

Table 5.28: Observed peak flow data at Kalaghoda Bridge.

Sr. No.	Year	Peak Flow Rate in Cumecs (X)
1	1996	265
2	1997	210
3	1998	425
4	1999	50
5	2000	0
6	2001	160
7	2002	85
8	2003	455
9	2004	200
10	2005	805
11	2006	660
12	2007	85
13	2008	75
14	2009	70
15	2010	500
16	2011	75
17	2012	75
18	2013	520

The magnitude of flood for different return periods are calculated and shown in Table 5.29. Ward numbers 2, 5, 6, 8 were severely affected by the flood and percentage of area inundated in these wards varies form 35.69% to 39.86% (Table 5.30 and Figure 5.52), and the average depth of inundation in these wards range from 1.66 m–2.66 m. Ward numbers 4, 7, 10, 12 were moderately affected by the flood and percentage of area inundated in these wards varies from 16.34% to 21.92%, and the average depth of inundation in these wards range from 1.85m–2.75m. Ward numbers 1, 3, 9, 11 were marginally affected by the flood and percentage of area inundated in these wards varies form 0.56% to 3.54%, and the average depth of inundation in these wards range from 0.62m–1.61m. Flood hazard assessment, by means of flood mapping and identification of flood risk areas, is a crucial element in the formulation of any flood management strategy. In order to generate categories of flood hazards, the water depth for each flood extent was classified according to Japanese criteria by the Ministry of Land Infrastructure and Transportation (MLIT) (Table 5.31) (Quiroga et al., (2016)). The criteria suggest five categories of flood hazards: H1—very low hazard (water depth < 0.5 m); H2—low hazard (water depth between 0.5–1 m); H3—medium hazard (water depth between 1–2 m); H4—high hazard (water depth between 2–5 m); H5—extreme hazard (water depth > 5 m) (Figure 5.53). According to flood hazards categories (Table 5.31), 55.65% of total flood extent are located in the very low hazard class (H1) followed by H4—high hazard class (17.73% of total flood extent), H3—medium hazard class (14.69% of total flood extent), H2—low hazard class (7.26% of total flood extent), H5—extreme hazard class (4.67% of total flood extent).

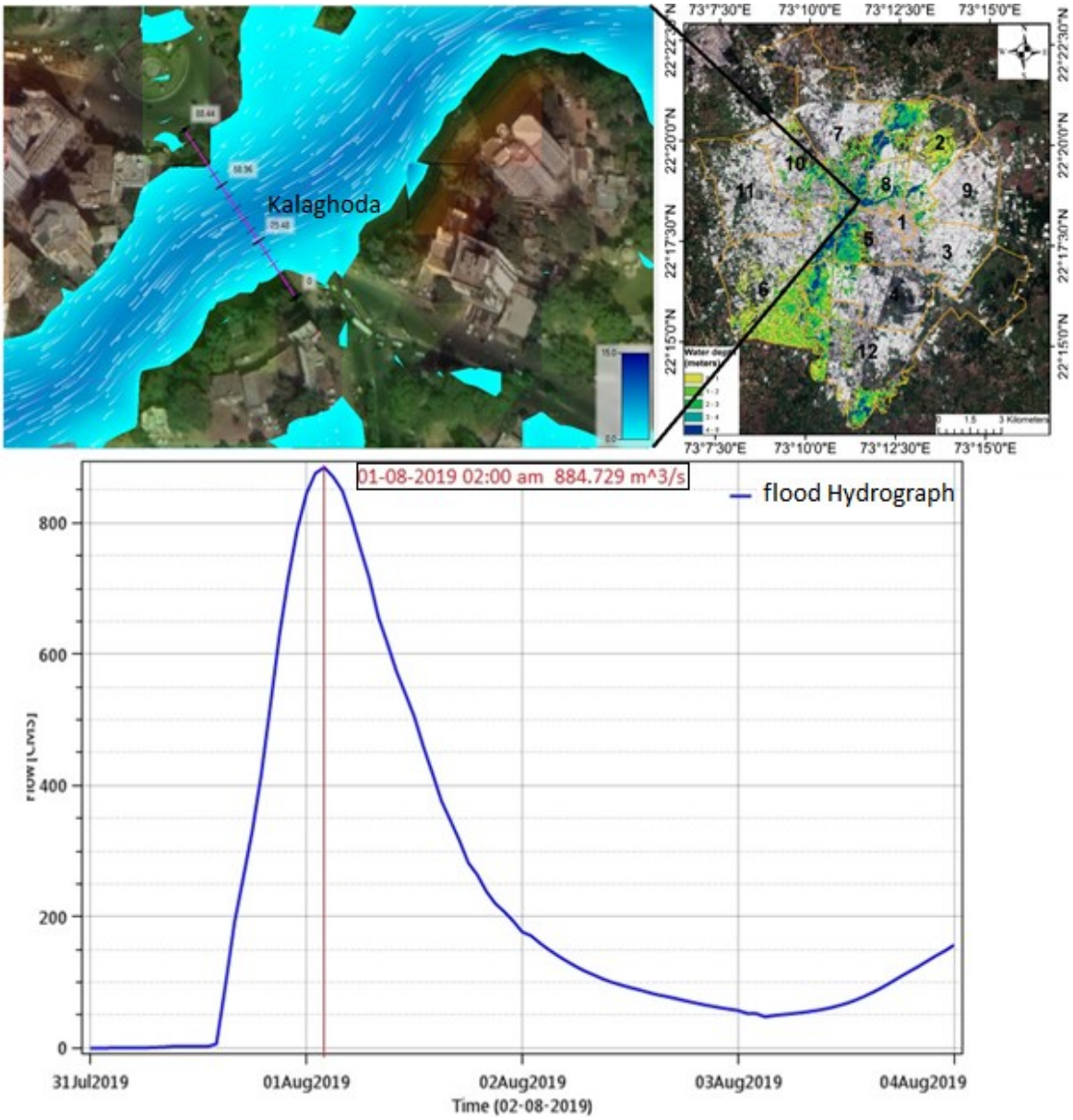


Figure 5.51: Simulated peak flood for the storm event at Kalaghoda Bridge.

Table 5.29: Magnitude of flood for different return period calculated using Gumbel's flood frequency method.

Return Period (years)	Flood Discharge (Cumecs)
15	752.69
25	872.69
35	951.73
50	1033.16

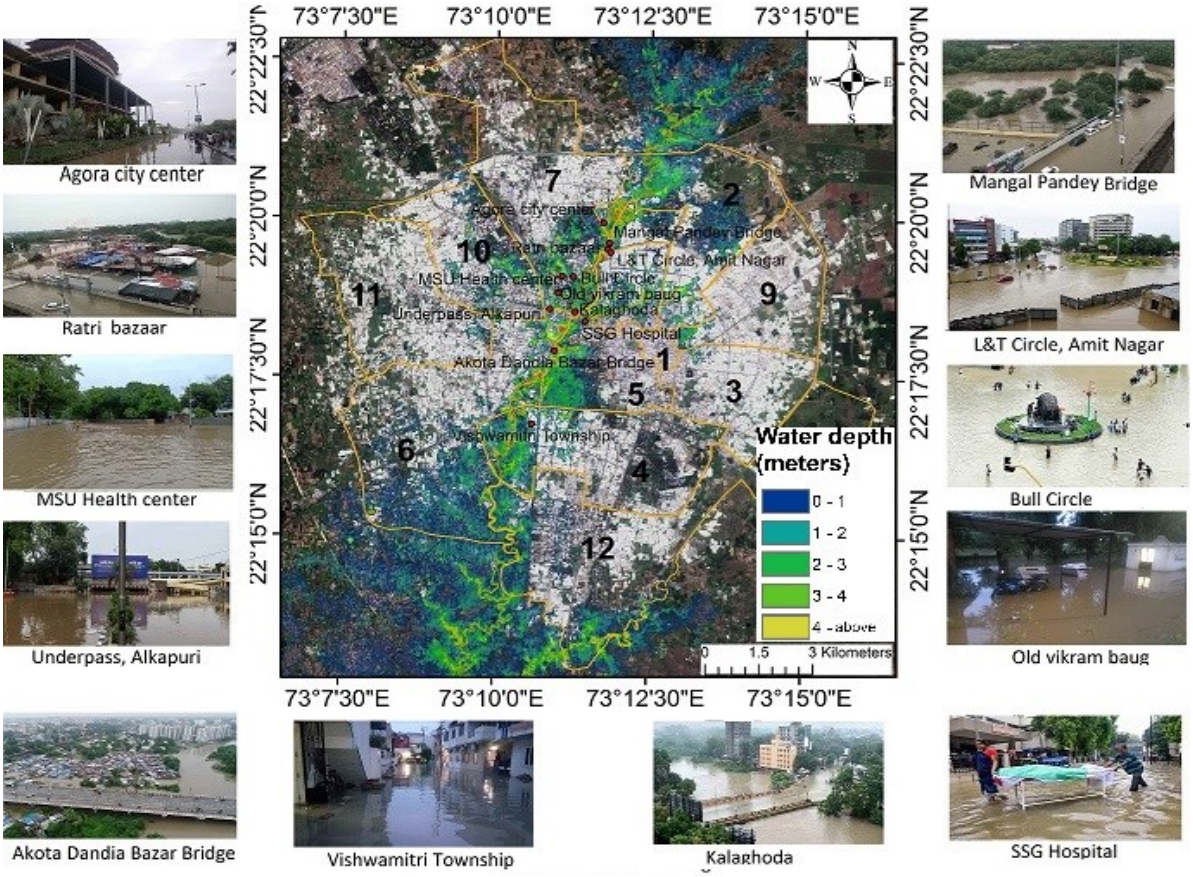


Figure 5.52: Inundation map of Vadodara city with sites visit.

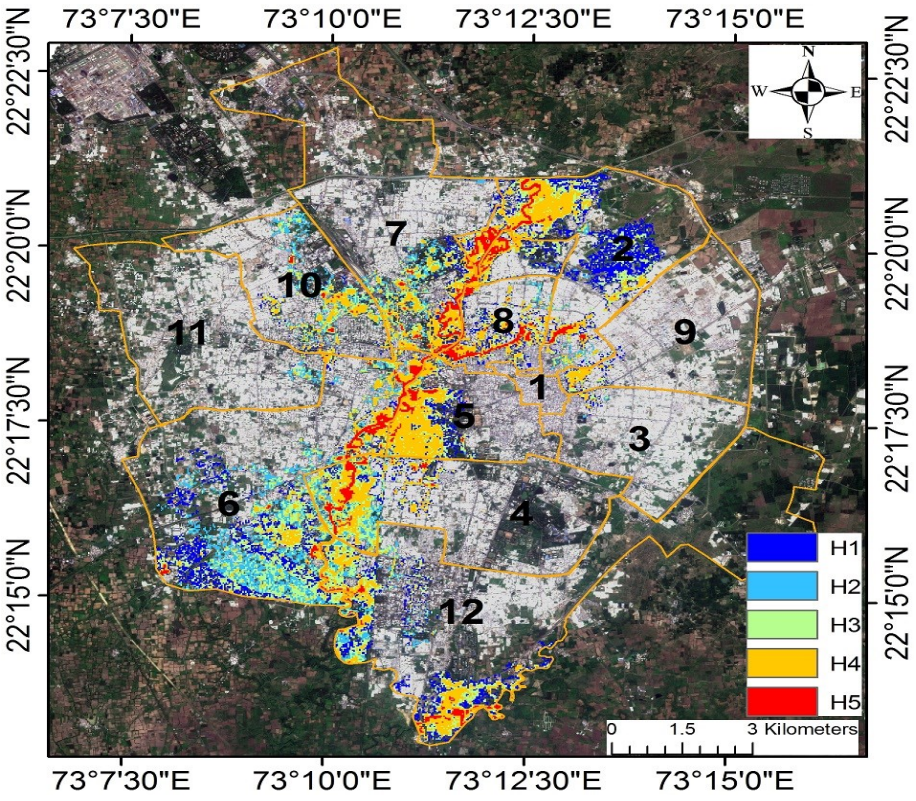


Figure 5.53: Classification of flood hazards based on MLIT water depth for Vadodara city.

Table 5.30: Inundation statistics for the Vadodara city.

Ward no.	Ward name	Ward area km^2	Min depth (m)	Max depth (m)	Range (m)	Mean Depth (m)	Inundated area (km^2)	Percentage of area Inundated
1	Nyay Mandir	1.12	0.048	2.05	2.00	0.62	0.01	0.56 %
2	Harni	13.73	0.002	12.26	12.26	1.97	5.46	39.79 %
3	Waghodia	9.39	0.001	4.05	4.04	1.30	0.13	1.38 %
4	Pratap Nagar	15.20	0.002	12.06	12.06	2.40	3.33	21.92 %
5	Raopura	7.08	0.008	11.23	11.22	2.54	2.53	35.69 %
6	Akota	22.83	0.001	13.00	13.00	1.66	8.59	37.64 %
7	Fatehgunj	22.47	0.005	13.19	13.19	2.75	3.67	16.34 %
8	Nagarwada	4.55	0.029	17.25	17.22	2.66	1.81	39.86 %
9	Ajwa	10.87	0.008	5.10	5.10	1.61	0.39	3.54 %
10	Subhanpura	9.61	0.001	8.87	8.87	1.85	1.86	19.31 %
11	Vasna	14.88	0.015	5.09	5.08	1.36	0.20	1.37 %
12	Makarpura	28.84	0.006	10.42	10.41	1.87	4.86	16.86 %

Table 5.31: Flood hazard classification based on water depth according to the MLIT.

Flood Hazard	Flood Depth (m)	Hazard Classes	Percentage of total flood extent	Hazard Description
H1	<0.5	Very low	55.65%	Flood does not pose hazard to people and on-foot evacuation is not difficult.
H2	0.5–1	Low	7.26%	Flood water poses hazard for infants and on-foot evacuation of adults becomes difficult; evacuation becomes more complicated.
H3	1–2	Medium	14.69%	Flood depth can drown people; people may be safe inside their homes.
H4	2–5	High	17.73%	People are exposed to flood hazard even inside their homes and evacuate towards the roof of their homes is suggested.
H5	>5	Extreme	4.67%	Built-up structures like homes may get covered by the flood; people may get drowned even if they evacuate towards the roof of their homes.

6. **Objective:** To quantify the effects of urban land forms on land surface temperature and modeling the spatial variation using machine learning. The models can help to predict land surface temperature under temporary cloud cover spots, which are present in the data at the time of the acquisition, using neighboring biophysical (cloud-free) independent variables relationship with land surface temperature.

LST distribution over the study area

The estimated land surface temperatures for the summer and winter seasons are shown in Figure 5.54. The urban heat in Vadodara city in the summer and winter is significant because

the high temperature zones are generally consistent with the urban built-up areas near the ward numbers 3, 9 and northeast of ward number 12, and the densely built-up areas in the ward numbers 3, 5, 6 and 12. However, the intensity of land surface temperature in winter is relatively low. In summer, the high temperature zone and the extremely high temperature zone occupy approximately 13.41% and 16.16% (Table 5.32) of the total study area, respectively. Furthermore, in winter, areas with high and extremely high temperatures are approximately 13.39% and 15.68% of the total study area, respectively. These two temperature classes show a 0.5% increase in area from winter to summer.

Table 5.32: Calculated area and percentage of each temperature class for summer and winter.

Temperature classes	Summer Area km² (%)	Breakpoint value for summer season image	Winter Area km² (%)	Breakpoint value for winter season image
Low temperature zone	20.03 (12.49)	LST<39.39	21.42 (13.35)	LST<29.95
Secondary low temperature zone	29.10 (18.14)	39.39≤LST<40.20	26.77 (16.69)	29.95≤LST<30.45
Medium temperature zone	34.29 (21.38)	40.20≤LST<41.00	34.82 (21.71)	30.45≤LST<30.95
Secondary high temperature zone	29.55 (18.42)	41.00≤LST<41.81	30.74 (19.17)	30.95≤LST<31.45
High temperature zone	21.51 (13.41)	41.81≤LST<42.61	21.48 (13.39)	31.45≤LST<31.95
Extremely high temperature zone	25.92 (16.16)	LST≥42.61	25.15 (15.68)	LST≥31.95

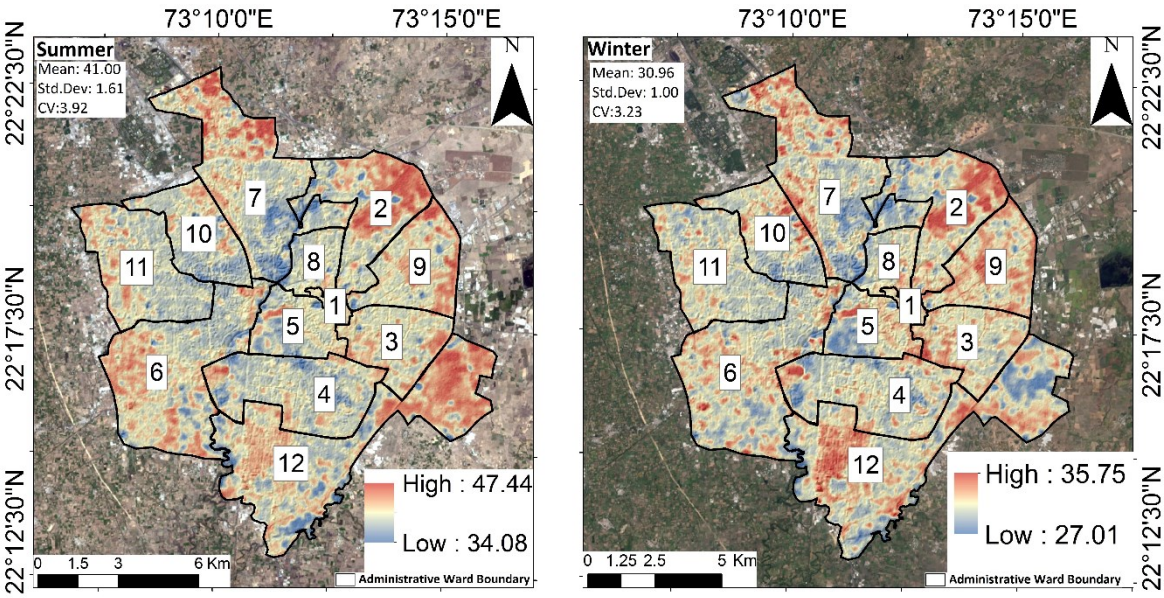


Figure 5.54: Estimated land surface temperatures for the summer and winter seasons using Landsat 8 data.

Temperature Variations for Different Land Cover Types

To understand the relationship between land surface temperature and land use/land cover, the investigation of the thermal signature of each land use/land cover form is important. A comparison was therefore carried out between land use/land cover and land surface temperature. The land surface temperature descriptive statistics of each land use/land cover category for both summer and winter in the study area are shown in Figure 5.56. The mean temperature of each landform category was calculated by averaging all consistent pixels of a given landform category. The average LST at 11:02:51.19 AM in summer reached up to 41.0 °C, whereas in winter it was 30.9 °C. Figure 5.54 shows the spatial distribution of land surface temperature in both summer and winter. In addition, LST in summer and winter had an almost similar and low coefficient of variation (CV), which indicate that temperature fluctuation in winter and summer are not substantial.

The average land surface temperature values of four land use/land cover types from high to low are baresoil > builtup > vegetation > water. The results indicated the highest land surface temperature was recorded for baresoil while the lowest was recorded for water bodies for both the seasons. In the study period, the city of Vadodara showed lower surface temperatures in residential urban areas as compared to the outskirts of the city. It is caused by heat from the sun in the surrounding areas that is directly absorbed into the ground, causing it to heat up faster than other ground cover ranges. This could be because of the different values of the surface albedo and land surface temperature on residential urban areas and baresoil. The residential areas in the city are generally painted with light colors which increases the albedo value than the baresoil land. In contrast, asphalt roads, pavements, buildings, concrete and other features that make up the urban surfaces tend to slowly release the heat absorbed. A high albedo means the surface reflects the majority of the radiation that hits it and absorbs the rest. However, black asphalt or roads in the urban areas tend to have high land surface temperature and low albedo value due to its thermal characteristic. Black asphalt or roads have a high tendency to absorb solar radiation. In other words, builtup areas tend to retain heat longer than other classes, such as barren land in the city areas, which does not retain heat for as long. These results are consistent with the findings of Kant et al., (2009) which found that areas with baresoil and builtup areas had higher land surface temperature levels, while other categories, such as water and vegetation, had lower land surface temperature values during the day. The results of this study suggest that the wastelands / barren lands have higher temperatures than residential urban areas.

Considering the relationship between landforms and thermal signatures is the most efficient approach in understanding the effect of different landforms on land surface temperature. Spectral indices are the most widely used and applicable method in large-scale research to measure the urban surface characteristic. Evidence from the past studies shows the precise and significant results in urban surface characteristic computation using spectral indices. To investigate the connection of LST with biophysical variables, indices such as NDVI, NDWI and DBSI were derived from Landsat 8. The NDVI has been used extensively to define the overall vegetation and green area conditions. A higher NDVI shows a higher vegetation likelihood. The DBSI can reveal the builtup and barren land of urban areas. High DBSI values generally signify areas with baresoil while mid-range values signify intensive urban development. Based on reflected near-infrared radiation and visible green light, NDWI enhances the open water features. A higher NDWI shows a higher water body likelihood. The relationship between land surface temperature and urban surface characteristics was examined using the Pearson correlation coefficient. The overall results indicated a statistically significant correlation (significant at the 0.01 level (2-tailed)). The NDVI map of Vadodara city with scatter plots are illustrated in Figure 5.55-A and Figure 5.57. The value of the NDVI ranged between −0.169 and 0.519, the areas with high NDVI values can be identified with dark green color. The strong negative correlation (r=-0.650) with LST and NDVI (Figure 5.57−A), shows that high areas of vegetation are the most likely to regulate the surface heating effect. Results describe the high temperature in less dense vegetative areas and low temperature in highly vegetated areas. Lower temperatures in vegetation areas are due to processes like transpiration and evapotranspiration. DBSI values ranged between −0.052 and 0.255, the DBSI value over baresoil and builtup classes showed a positive correlation with LST (Figure 5.57 B−C). It was found that the baresoil and builtup areas have a noticeable effect on the surface urban heat. Water bodies have a little thermal response and are known to be an efficient absorbent of radiation. The result presented in Figure 5.55−B indicates the NDWI spatial distribution, the NDWI values ranged from -0.458 to −0.166. A negative correlation was observed between LST and NDWI (Figure 5.57−D) over water bodies. Surface water characteristics reflect the pattern of heat flow and can be used to minimize the impact of urban heat.

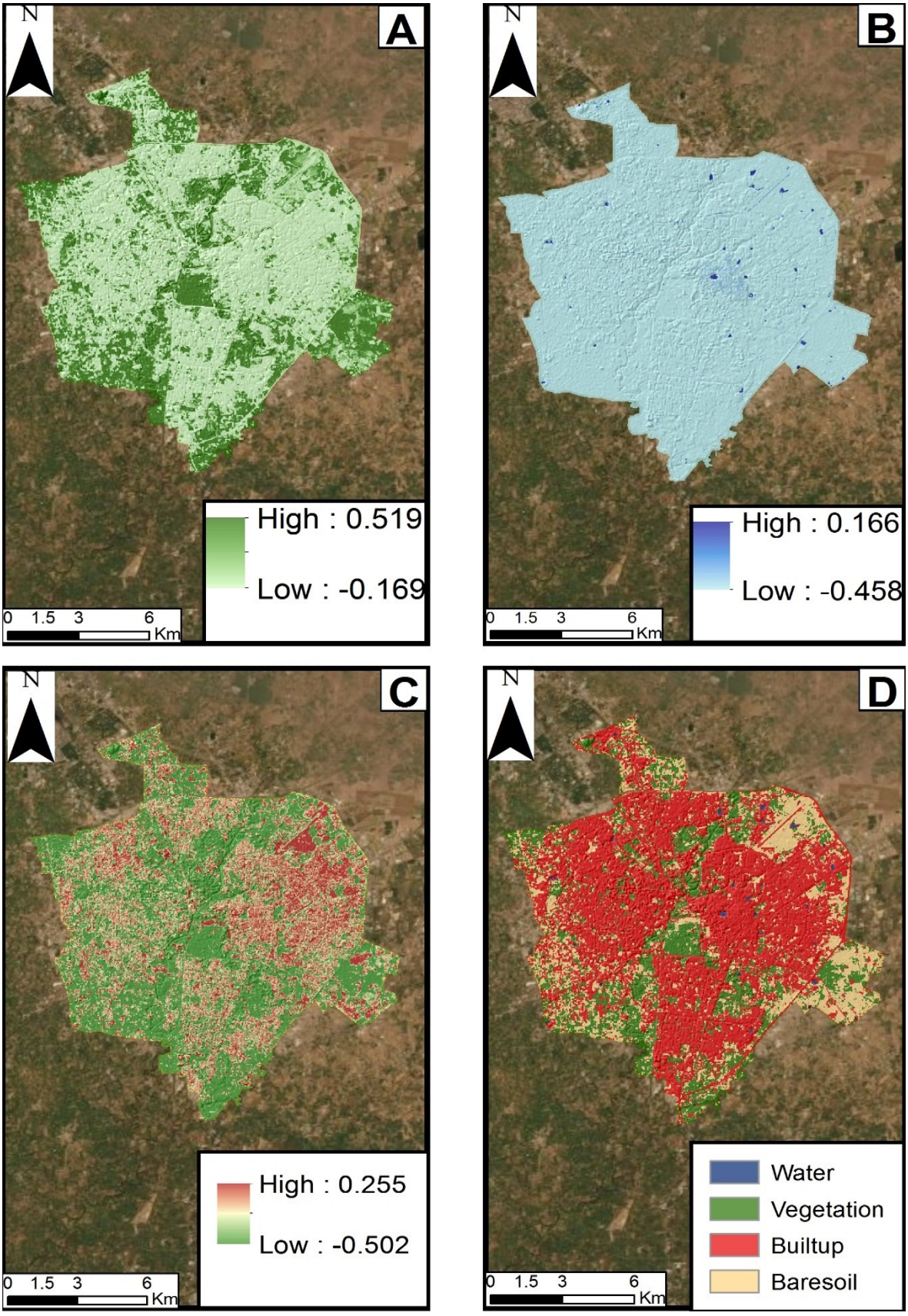


Figure 5.55: Spatial distribution of (A) NDVI, (B) NDWI, (C) DBSI and (D) Land use/land cover over Vadodara city.

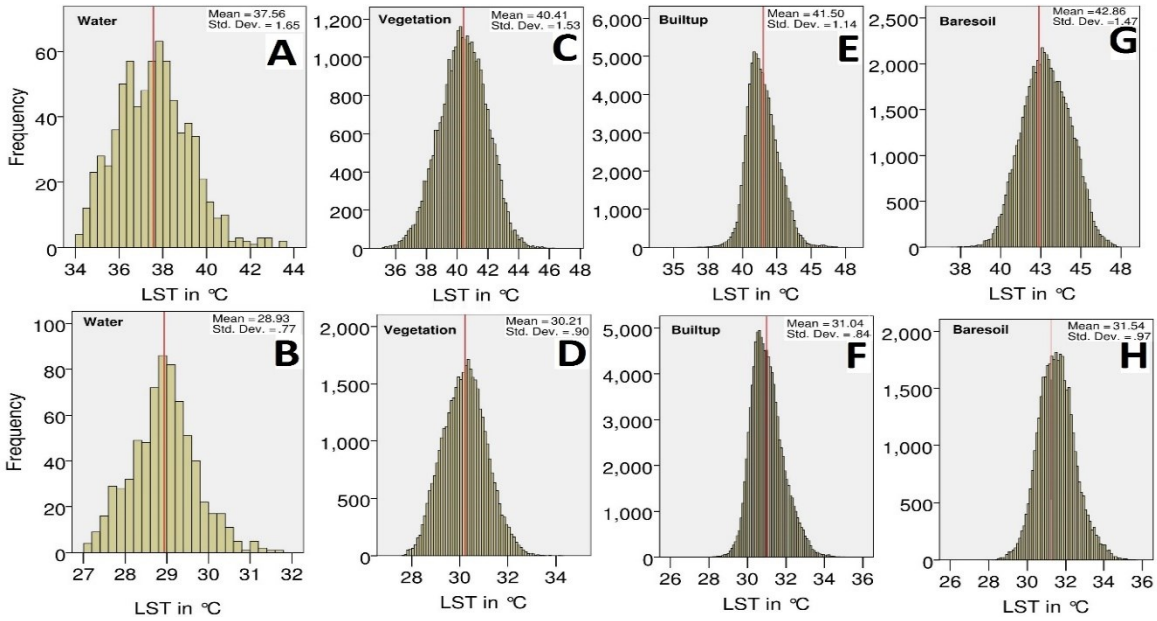


Figure 5.56: Land surface temperature descriptive statistics of each land use/land cover category for both summer (A-G) and winter (B-H) in the study area.

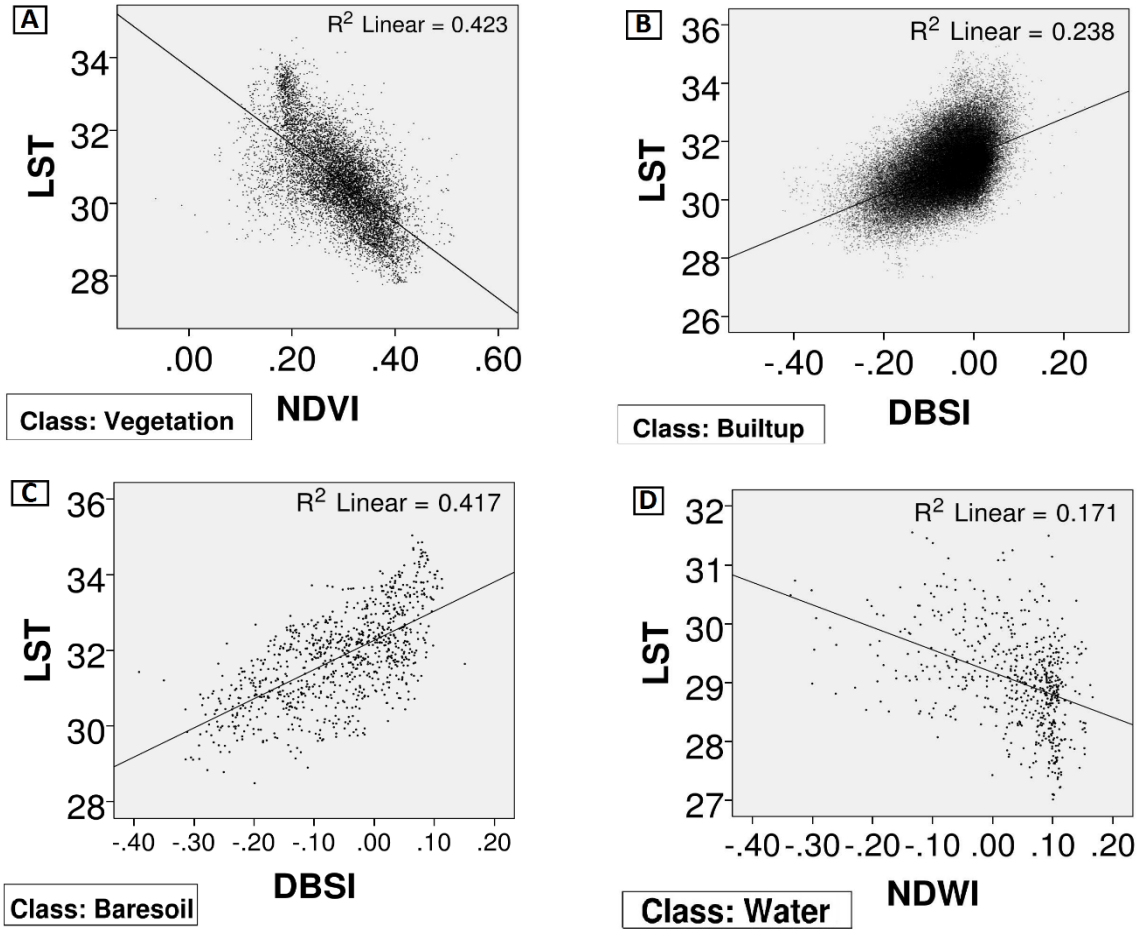


Figure 5.57: Scatterplots of (A) LST-NDVI over vegetation, (B) LST-DBSI over builtup class, (C) LST-DBSI over baresoil class and (D) LST-NDVI over water class.

Land Contribution Index (CI) and Landscape index (LI)

To accurately identify the connection between the surface parameters and the trajectory of the land surface temperature in the area under study, a contribution index for each type of land cover was calculated for the summer and winter seasons. The influence of each of the specified land use/land cover types on the research area's land surface temperature was investigated and illustrated in Table 5.33. The results of the calculated contribution indices of the land use/land cover types show the dominance of the baresoil surface in relation to the impact on the overall land surface temperature regime of the study area. This trend might be explained by the fact that the baresoil land cover type is one of the two groups with a positive net contribution index. Second, for both seasons, this contribution index is greater than all others combined, suggesting that the baresoil region contributes to more surface heating than any other land use/land cover in the study area or has the largest heat generating capability on the surface. Among the four types, vegetation provided maximum heat mitigating impact in the study area. The contribution index value of water remained the same for both the seasons, indicating that temperature variation over water tends to be less variable due to its high thermal capacity. As expected, baresoil and builtup land had high contribution index in summer, which was significantly lower in winter due to lower solar radiation. Apart from baresoil and builtup land, the other types also provided less heat contribution in winter. These observations can be explained by the rainfall season prior to winter season which leads to more vibrant urban green space and therefore more heat sinking. From Table 5.34, it can be seen that all the values of LI of source landscape increased from summer to winter season, thus indicating that the intensity of the land surface temperature decreasing. This depicted that the contribution of the source and sink landscapes weakens the intensity of land surface temperature in the winter season. LI value of less than 1 in summer shows that contribution of source and sink landscape promoted the intensity of land surface temperature. In a study conducted by Zölch et al., (2016), it was found that tree planting had the strongest impact on average physiological equivalent temperature (PET) reduction at 13 percent compared to existing planting. Open spaces with shade of trees and green facades provide evapotranspirational cooling. To identify administrative wards with green spaces requirements (such as woodlands, parks, street trees, green roofs and facades), values were assigned on the scale of 1 (low temperature zone) to 6 (extremely high temperature zone) on land surface temperature rasters of both the seasons. A combined score value was later used to divide the administrative wards into low, medium and high green spaces requirement wards. Figure 5.58 and Table 5.35 show that ward numbers 2, 9 and 12 belong to high green spaces requirement wards. Strategically planting vegetation in such heat-exposed areas will be more

effective than merely selecting a large percentage of the green cover. This strategy will moderate the city climate because shading and evapotranspiration reduce the thermal load outdoors during hot weather conditions.

Model fitting and evaluation

To evaluate the machine learning models' performances for K-Nearest Neighbor (K-NN) regression, Neural Networks (NN), Random Trees (RT) regression and Support Vector Machine (SVM) regression with the mean moving kernel (observation grid) of 2×2 and 5×5 for each explanatory variable (NDVI, NDWI and DBSI). Three performance measures, namely, coefficient of determination, bias and RMSE were used. Figure 5.61 shows the model performances for KNN, NN, RT and SVM with mean moving kernel in the winter season. When considering both the scenarios (2×2 and 5×5), more than 60% of the land surface temperature variation was explained by explanatory variables in each model except RT (2×2 and 5×5) (Figure 5.59), which was nearly about 20% only. Maximum land surface temperature variation was explained in NN 5×5 (about 64.1%), followed by, KNN 5×5 (62.6%), SVM 5×5 (62.1%), KNN 2×2 (61.7%). While, variation was explained equally in NN 5×5 and SVM 2×2 (61.3%).

For KNN 2×2, KNN 5×5, RT 2×2, SVM 2×2 and SVM 5×5 the land surface temperature absolute error map ($|\text{predicted LST} - \text{actual LST}|$) are shown in Figure 5.60–A1,E1,C1,D1 and H1, the mean value of error map ($\text{predicted LST} - \text{actual LST}$) or Bias and standard deviation were found to be -0.0076 ˚C, -0.0097 ˚C, -0.0006 ˚C, -0.0622 ˚C and -0.0642 ˚C, and 0.561 ˚C, 0.549 ˚C, 0.889 ˚C, 0.621 ˚C and 0.612 ˚C, respectively (Table 5.36). The corresponding frequency histograms of error (Figure 5.60–A2, E2, C2, D2 and H2) indicate that the above mentioned models underestimated the predicted LST. For NN 2×2, NN 5×5 and RT 5×5 the land surface temperature absolute error maps are shown in Figure 5.60–B1, F1, and G1, the mean value of error map or Bias (◦C) and standard deviation were found to be 0.0047 ˚C, 0.0011 ˚C and 0.0034 ˚C, and 0.617 ˚C, 0.595 ˚C and 0.890 ˚C, respectively (Table 5.36). The corresponding frequency histograms of error (Figure 5.60–B2, F2, and G2) indicate that the models slightly overestimated the predicted LST. The worst performances were observed by RT models (2×2 and 5×5) (Figure 5.61–E and F). However, NN 2×2, NN 5×5, SVM 2×2 and SVM 5×5 models performed moderately good with overall RMSE of 0.617 ˚C, 0.594 ˚C,0.623 ˚C and 0.615 ˚C, respectively. Although the average estimation by all the models is very close to the average LST calculated using the Landsat 8 data, boxplots show that none of the models was capable to predict the extreme values Figure 5.62. The comparative results revealed that the K-NN algorithm outperformed the

other models. The lowest overall RMSE was calculated at a value of 0.549 ◦C for KNN 5×5, followed by, KNN 2×2 (0.561◦C).

Table 5.33: The calculated land contribution indexes of each land cover types.

	Summer					Winter				
Class	**Area Km²**	**Ti**	**M**	**Pi**	**Ci**	**Area Km²**	**Ti**	**M**	**Pi**	**Ci**
Water	0.65	37.56	41.00	0.004	-0.01	0.62	28.94	30.97	0.004	-0.01
Vegetation	27.01	40.41	41.00	0.168	-0.10	34.28	30.21	30.97	0.214	-0.16
Builtup	90.27	41.50	41.00	0.563	0.28	90.27	31.04	30.97	0.563	0.04
Baresoil	42.45	42.86	41.00	0.265	0.49	35.21	31.55	30.97	0.220	0.13

Table 5.34: The calculated landscape index of sink and source landscapes.

Class		***Sink***	***Source***
Summer	**Area Km²**	27.66	132.72
	Ti	38.98	42.18
	M	41	41
	Pi	0.17	0.83
	Ci	-0.3467	0.9821
	LI	0.353	
Winter	**Area Km²**	34.9	125.48
	Ti	29.58	31.29
	M	30.97	30.97
	Pi	0.218	0.782
	Ci	-0.303	0.2561
	LI	1.183	

Table 5.35: Administrative wards with green spaces requirement.

	Summer		Winter			
Ward no.	**Majority class**	**Individual score**	**Majority class**	**Individual score**	**Combined score**	**Green spaces requirement**
1	Secondary high temperature zone	4	Secondary high temperature zone	4	8	Medium
2	Extremely high temperature zone	6	Extremely high temperature zone	6	12	High
3	Medium temperature zone	3	Secondary high temperature zone	4	7	Medium
4	Medium temperature zone	3	Medium temperature zone	3	6	Medium
5	Medium temperature zone	3	Low temperature zone	1	4	low
6	Secondary high temperature zone	4	Medium temperature zone	3	7	Medium
7	Medium temperature zone	3	Medium temperature zone	3	6	Medium
8	Medium temperature zone	3	Medium temperature zone	3	6	Medium
9	Medium temperature zone	3	Extremely high temperature zone	6	9	High
10	Secondary low temperature zone	2	Medium temperature zone	3	5	Medium
11	Secondary low temperature zone	2	Medium temperature zone	3	5	Medium
12	Extremely high temperature zone	6	Extremely high temperature zone	6	12	High

Table 5.36 Calculated bias and RMSE of the predictive models.

	Bias(◦C)		**RMSE(◦C)**	
	2X2	5X5	2X2	5X5
KNN	-0.0076	-0.0097	0.561	0.549
NN	0.0047	0.0011	0.617	0.594
RT	-0.0006	0.0034	0.888	0.890
SVM	-0.0622	-0.0642	0.623	0.615

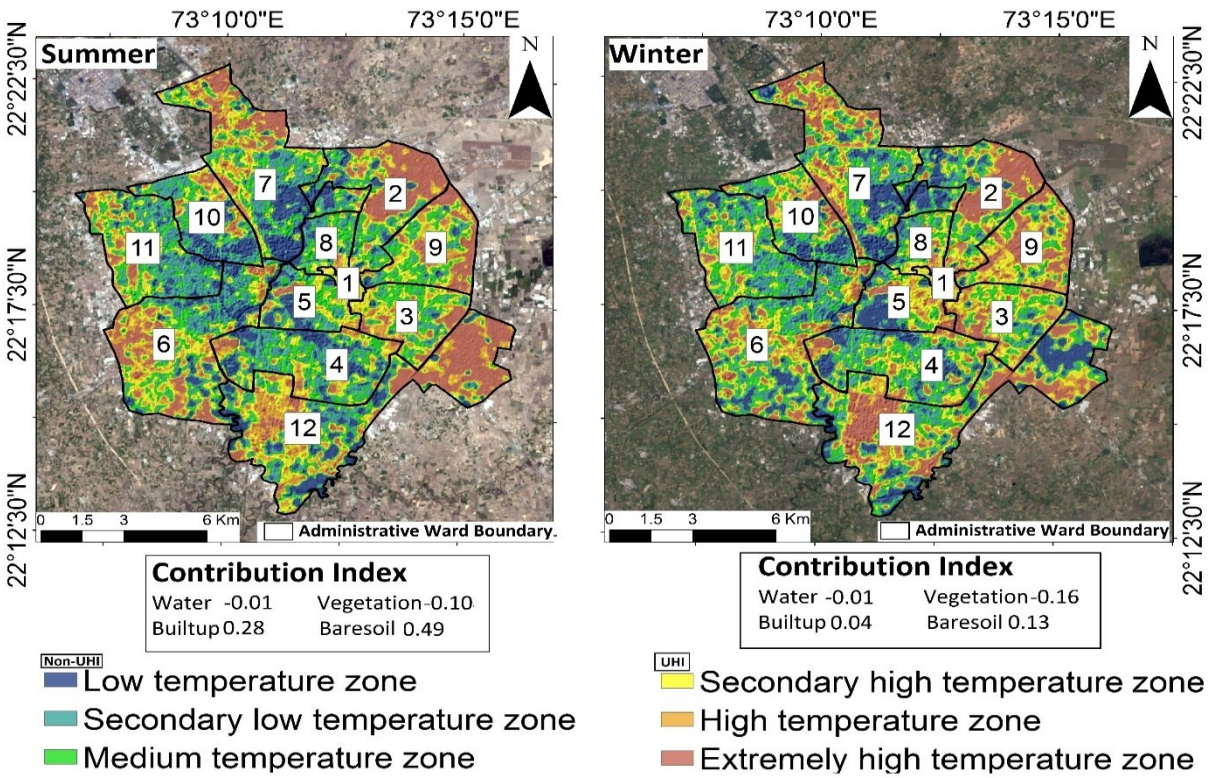


Figure 5.58: Classification of heat zones into UHI and non-UHI zones, and contribution index of land use/land cover classes in summer and winter seasons.

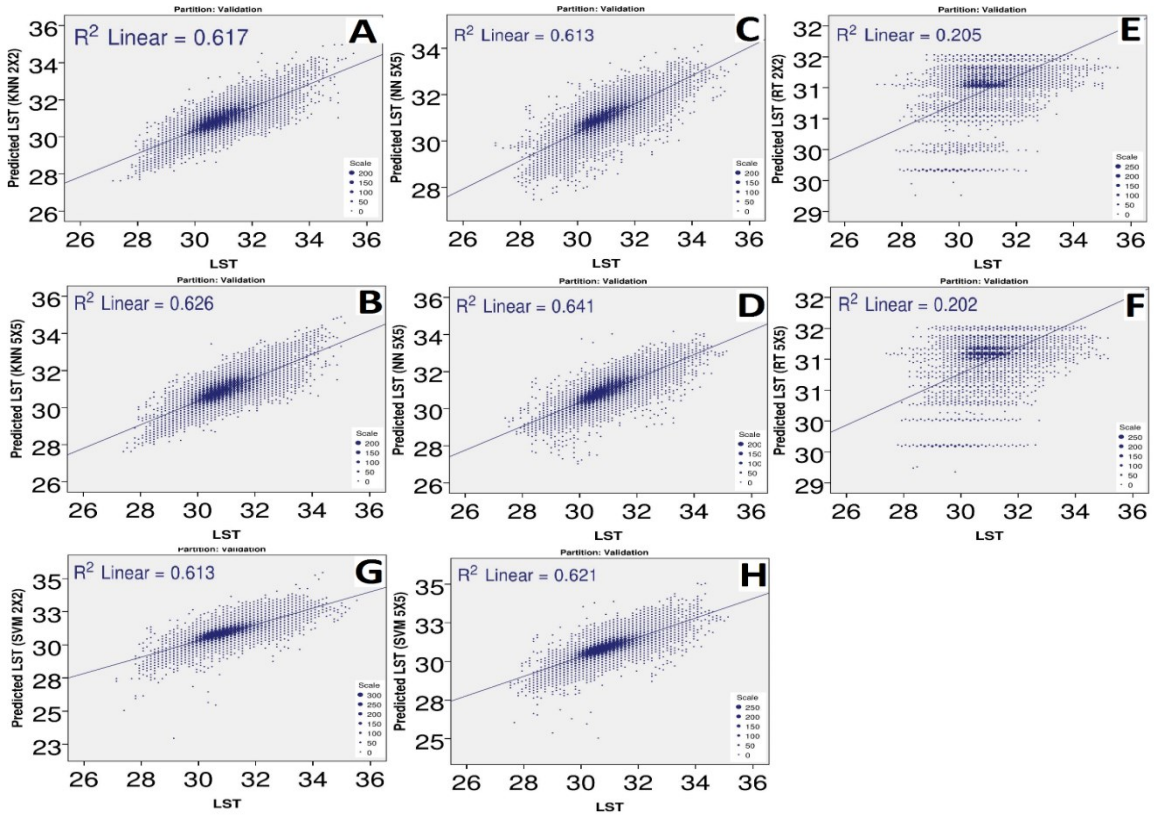


Figure 5.59: Coefficient of determination between the LST calculated from Landsat 8 and the predicted LST using the K-NN (2×2), K-NN (5×5), NN (2×2), NN(5×5) ,RT (2×2), RT (5×5), SVM (2×2) and SVM (5×5) models, respectively.

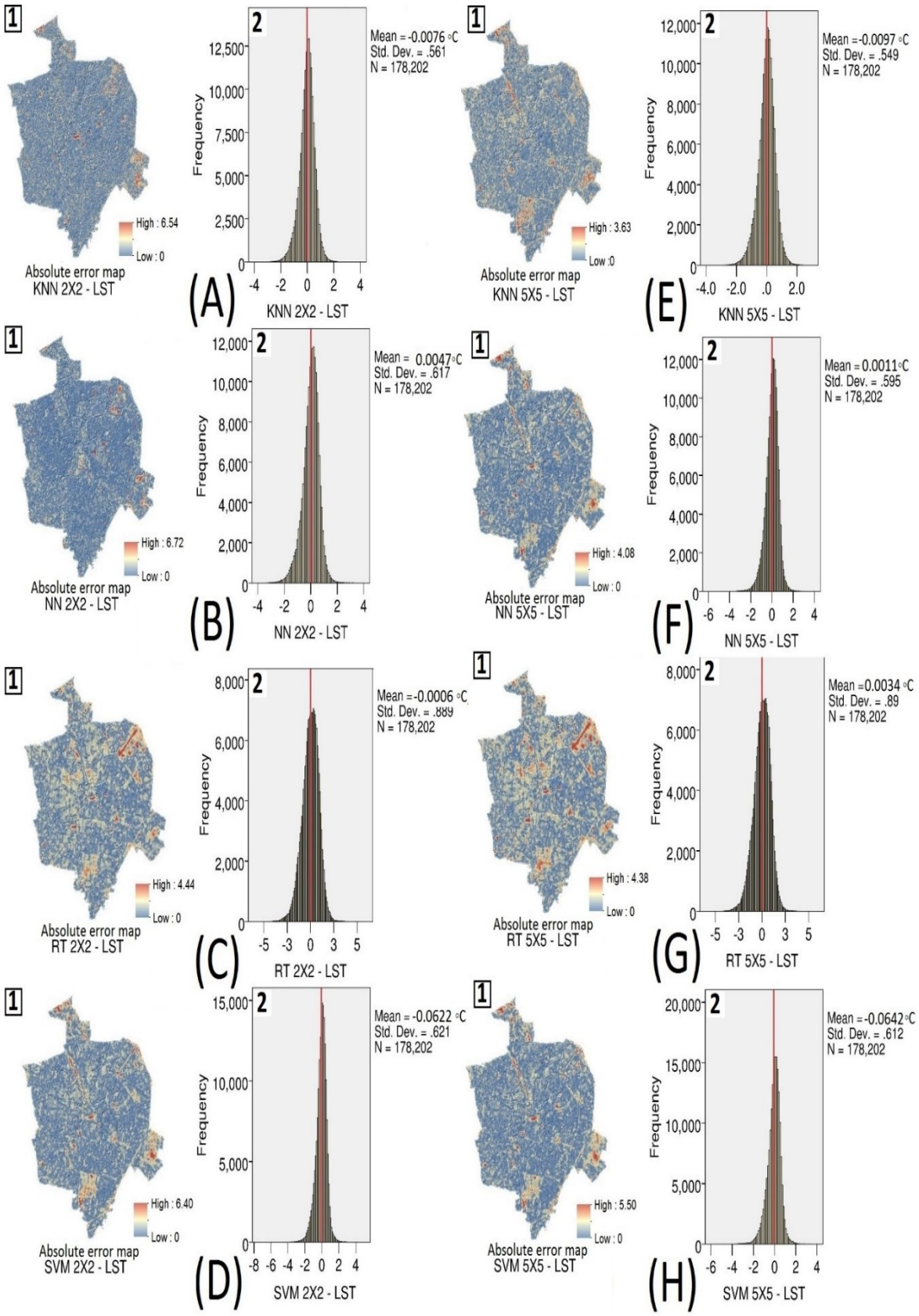


Figure 5.60: Absolute error maps A1-H1 and corresponding frequency histogram of error A2-H2 of predictive models.

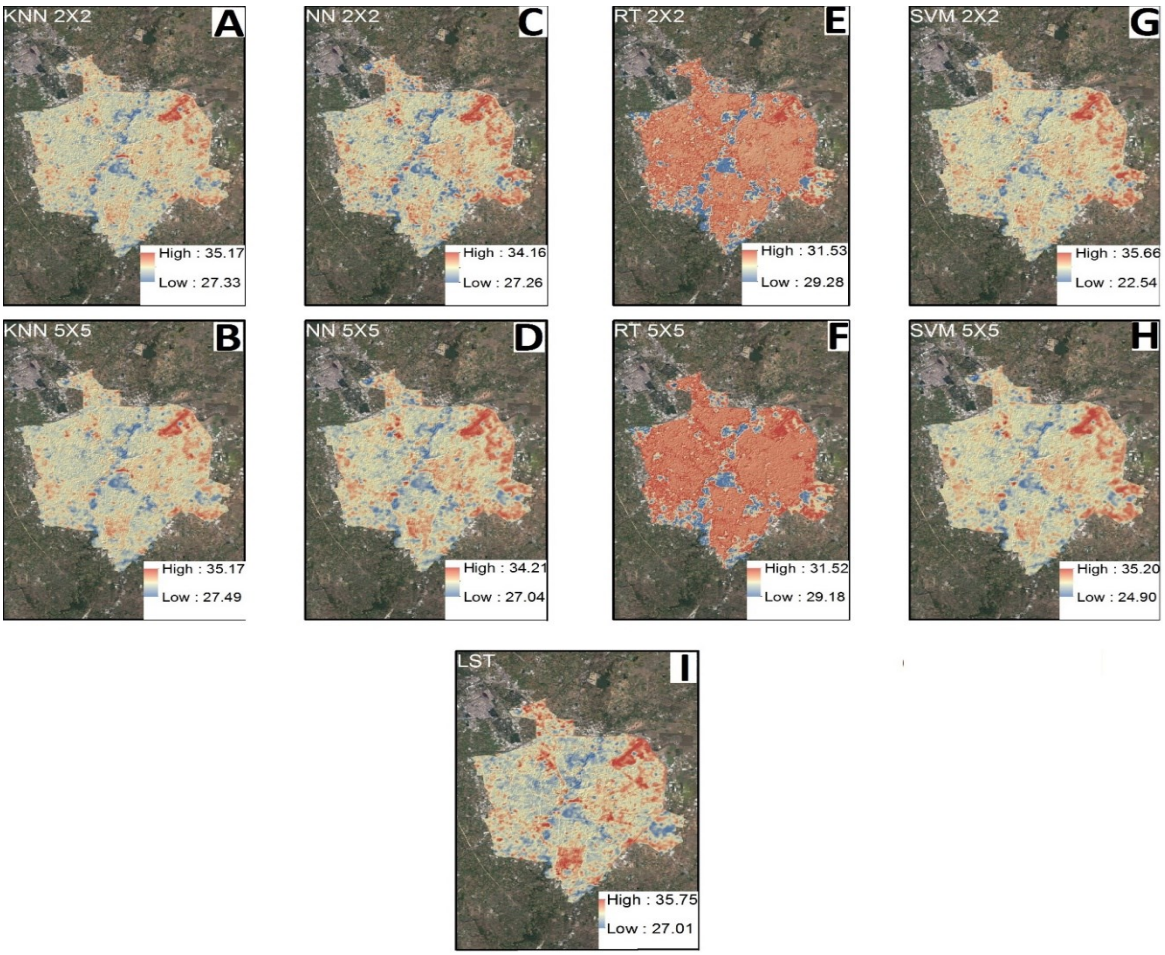


Figure 5.61: Maps A,B, C, D,E,F,G and H show the predicted LST using the K-NN (2×2), K-NN (5×5), NN (2×2), NN(5×5) ,RT (2×2), RT (5×5), SVM (2×2) and SVM (5×5) models, respectively. Map I shows the observed LST estimated using Landsat 8.

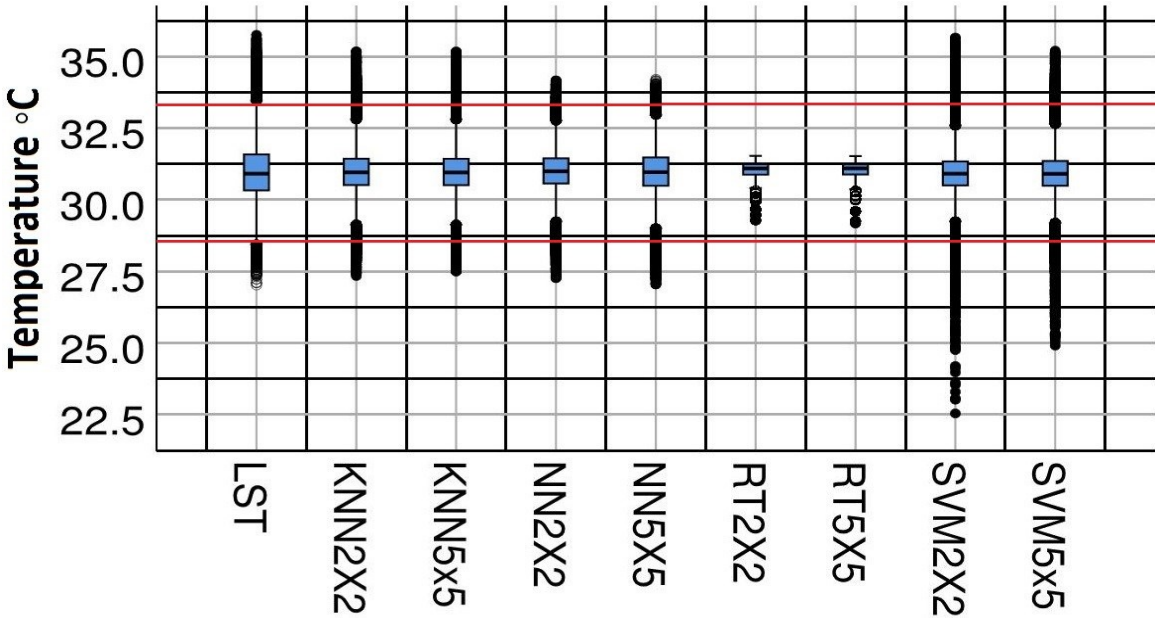


Figure 5.62: Box plot of observed LST estimated using Landsat 8, K-NN (2×2), K-NN (5×5), NN (2×2), NN(5×5) ,RT (2×2), RT (5×5), SVM (2×2) and SVM (5×5) models, respectively.