CHAPTER 4
METHODOLOGY

4.1 GENERAL

This chapter briefly shows the methodology adopted for the individual objective. The chapter is divided into six sections, each section contains the specific methodology for the specific objective.

4.2 METHODOLOGY

Objective 1: To study the changes in Land Use/Land Cover (LULC) /Urban Sprawl over a period, that influences the runoff characteristics of the study area.

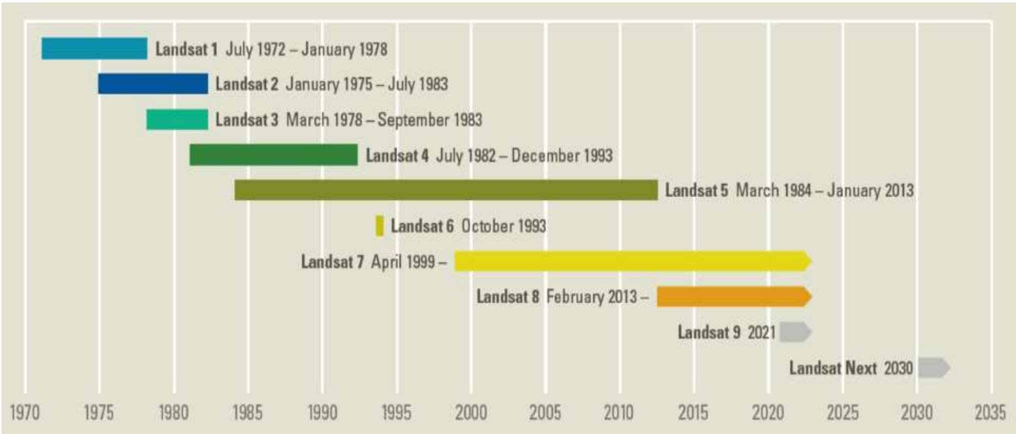

Figure 4.1: Collection of Landsat Satellite Images Provided by NASA/USDA

Since 1972, Landsat satellites have continuously acquired space-based images of the Earth's land surface, providing uninterrupted data to help land managers and policymakers make informed decisions about our natural resources and the environment. Fig. 4.1 shows the collection of various Landsat images launched in various different years.

Land use/ land cover (LULC) change detection based on remote sensing data is an important source of information for various decision support systems. Information derived from land use and land cover change detection is important for land conservation, sustainable development, and management of water resources. Landsat Imagery data was collected for the years 1977, 1988, 1993, 1998, 2003, 2008, 2013 and 2018 pertaining to the area under study.

Different land use/ land cover has been recognized and demarcated pertaining to the area under study, using visual interpretation techniques of satellite data. The focus is on defining the broader levels of land use zones inside the city limits; the specifics are categorized with a viewpoint towards mapping land use and land cover. Google Earth Pro is used for feature verification and inclusion when needed to identify typical land use zones, such as developing built-up areas, places with natural vegetation, and agricultural areas. The goal of this study is to map urban settlements, agricultural, natural vegetation, and water bodies. Where necessary, feature incorporations and verifications have been done. The software that delineates land use classes facilitates the identification of land use zones, allowing for the analysis and association of spatial patterns of growth profiles and its pattern of evolution.

4.3 LAND USE / LAND COVER ANALYSIS

In principle, land cover mapping from satellite data consists of four steps: data acquisition, pre-processing, analysis/classification, product generation and documentation.

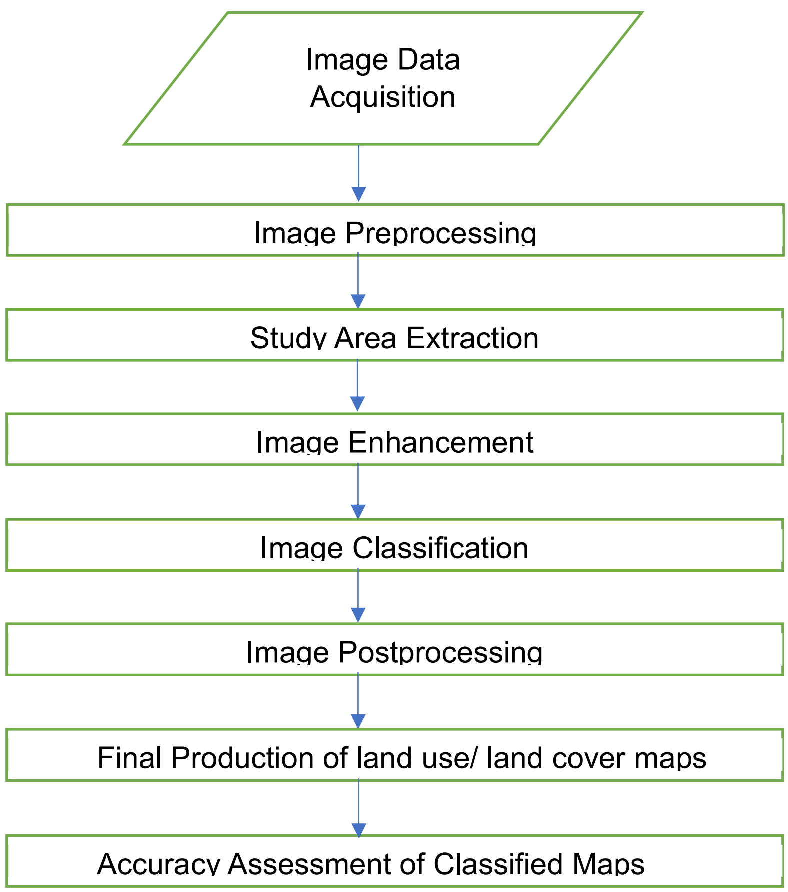

Figure 4.2: Flowchart for LULC detection

Fig. 4.2 shows the flowchart of the methodology adopted for carrying out the LULC classification. Data acquisition of satellite imagery for the Vadodara city is done from the USGS website as in Table 3.1. Further preprocessing of the acquired image is carried out.

Image Pre-processing

In principle, it entails geometric and radiometric corrections.

In remote sensing, data for georeferencing refers to the process of assigning to image data from some map coordinate. Geometrical distortion is an error on an image between actual image acquisition (sensor) coordinate data and ideal image coordinate data.

Here for images, georeferencing is done using ArcMap software. Vadodara lies in the following UTM Zone: WGS_1984_UTM_Zone_43N. Details on radiometric corrections is presented in the subsequent paragraphs.

Coarse resolution data:
Previously, certain classification tasks utilized low-resolution data and were conducted using individual, mostly cloud-free photos. Nevertheless, this approach is inherently constrained because to the diminishing likelihood of cloud-free images as the coverage area of a single picture expands. Obtaining acceptable photos for land cover mapping is quite challenging, particularly when the required time span is limited. Moreover, these photos exhibit systematic inaccuracies caused by atmospheric influences that vary with the distance travelled, as well as a consistently shifting level of detail for sensors with lower resolution. Classifying them is so challenging and necessitates interactive refinement for every input scene employed, along with subsequent procedures to harmonize disparities between neighbouring scenes, guaranteeing uniformity throughout the mapped region. In recent years, research has prioritized the utilization of picture composites due to these factors. During the compositing process, the visual result is carefully constructed to include as much information as possible regarding the characteristics of the land surface. The primary goal of the technique is to choose the measurement with the least number of clouds from the available options for a certain pixel in the composite image, considering that clouds usually occupy a significant portion of the pixels.

Fine resolution data:
In the past, most land cover studies employing high resolution data were carried out with single images (hereafter called ‘scenes’), parts of scenes or an assembly of such scenes from different areas. In these cases, radiometric consistency was not an issue because the classification could be optimized individually for each scene. When classifying a scene composite (i.e., a mosaic of scenes), the

situation is more complicated. In principle, two options are possible. First (case I), one can classify each scene separately and subsequently reconcile the classes across the mosaic. Another approach (case II) is to assemble a mosaic of scenes for the entire area, establish radiometric uniformity across the mosaic, and then classify it as one entity. In case I, each scene is treated as a separate data set to be classified, using ancillary data that are appropriate for the classification procedure employed.

Due to the limited frequency of satellite revisits, the process of combining high-resolution data over extensive regions (known as case II) involves using complete scenes rather than individual pixels from low-resolution data. Therefore, while radiometric noise remains, it manifests in various distinct forms. Initially, the impact of air pollution is minimized by exclusively selecting landscapes that are predominantly free of clouds and haze (ideally less than 10% coverage). Furthermore, bidirectional issues are significantly less problematic, especially when considering nadir-looking sensors with a limited field of view like the Landsat Thematic Mapper (TM).

Layer stack is one of the useful Raster image preprocessing method. It is a process of joining all the available band in a single layer for further processing.

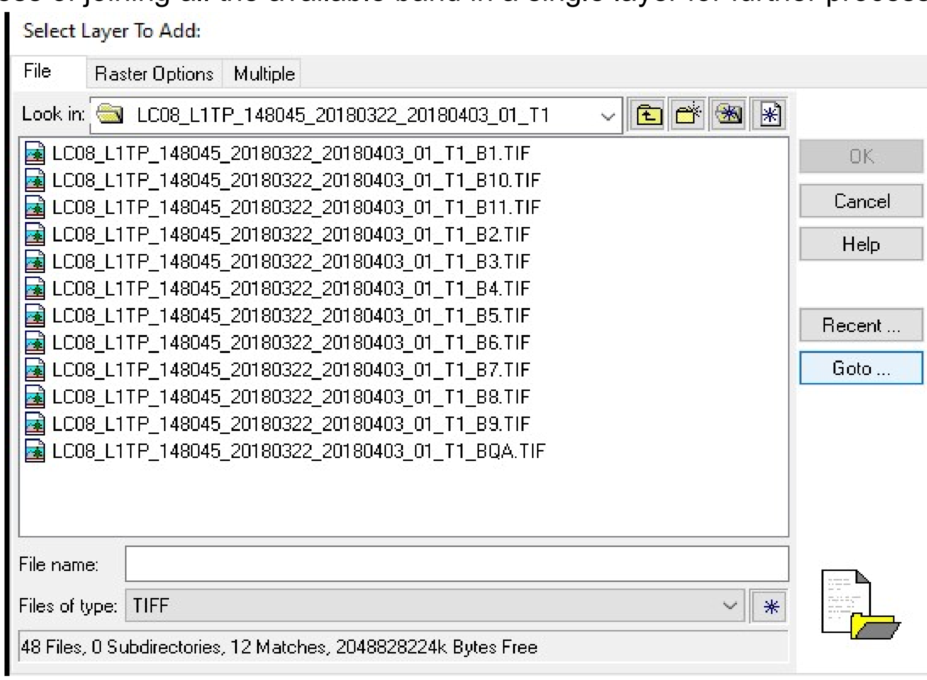

Figure 4.3: Number of bands in Satellite Image of a Particular Year

Fig. 4.3 shows the different bands available in a particular year image. In layer stacking technique these bands are stacked together to form a single layer. Once the layer is prepared, one can move to study area extraction.

Figure 4.4: Layer Stacking of the Satellite Image

Fig. 4.4 shows the image after layer stacking. This image is further extracted for the study area.

Study Area Extraction

Study area extraction was performed in Erdas Imagine software, with the help of mosaic tool. This step requires the shapefile of the study area either directly available or to be prepared. In the present study the shape file of Vadodara City was downloaded from the given link (https://dataspace.niua.org/dataset/ward-office/resource/79797beb-7012-463b-89c0-40264045ae5f) for the year 2001 and was also verified with the manual maps uploaded by Vadodara Municipal Corporation (https://vmc.gov.in/TDOTP.aspx). For study years excluding the year 2011, areas are extracted after carrying out the LULC classification, using the visual interpretation technique.

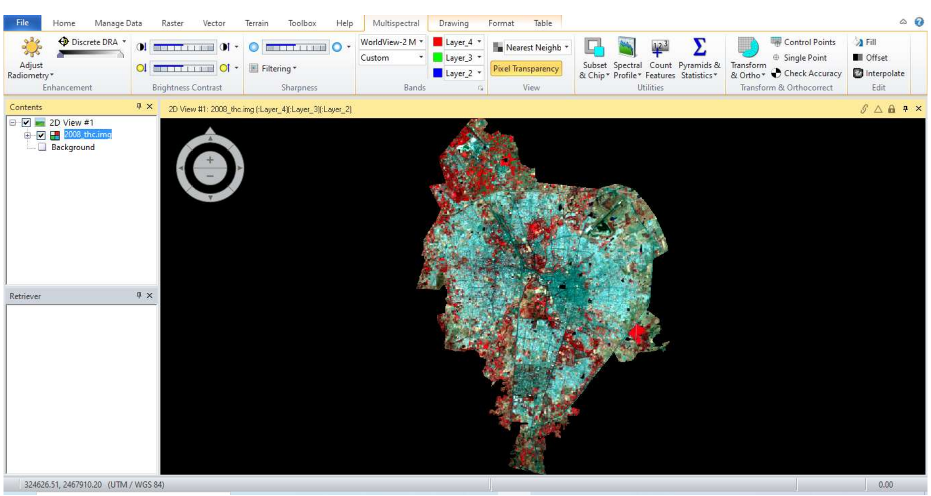

Figure 4.5: Study Area Extraction from the Satellite Image

Fig. 4.5 shows the extracted image of the study area. Further to classify this image, some enhancement is to be carried out.

Image Enhancement

Image enhancement is the process of making an image more interpretable for a particular application. Enhancement makes important features of raw, remotely sensed data more interpretable to the human eye. Changing the band combination helps to identify the features of land use. The combination of bands is referred to as color composites.

Colour Composites

A multispectral image data has six bands (nominally 30m per pixel) where the sensor on the satellite measures the intensity of reflected light in six distinct bandwidths filters ranging from blue light to near infrared and short-wave infrared. A colour composite is an image created using combinations of these image band data displayed in red, green and blue display channels. We must create true and false colour composites for general image enhancement. False colour composites (FCC) can be generated using different combinations of the Landsat bands displayed in red, green, blue (RGB) image display channels.

Image Classification

The classification of unsupervised data through ERDAS Imagine helped in identifying the terrestrial objects in the Study Image. The spectral pattern present within the data for each pixel was used as the numerical basis for categorization. The first analysis of the Image involved the use of generalized Unsupervised Classification with 5 categories namely Waterbodies, Urban Settlement, Natural Vegetation, Agricultural Land and Barren land. The result of the Unsupervised Image was used to create another image by using Supervised classification, which is named as Combined classification.

Image clipping was conducted to create a fully classified image, Computer-guided (Unsupervised) and User-guided (Supervised) procedures are used on image clipping to view each classification one at a time and stack them into a complete classified image. Land cover information that can be gleaned from satellite images is the spectral and spatial attributes of individual cover types. Table 4.1 shows the LULC classification scheme used in the study area.

Table 4.1: LULC Classification Scheme used in the Study Area

LULC Class	Description
Waterbodies	Rivers, streams, ponds and lakes
Urban Settlement	Residential areas, commercial areas, industrial areas, urban areas, roads and airport
Natural Vegetation	Trees which have grown naturally without human aid and has been left undisturbed by humans for a long time.
Agricultural Land	All the cultivated and uncultivated land used for agriculture, crops, and horticulture
Barren Land	Locations outside of and inside of cities that are completely devoid of vegetation, including playgrounds, waste grounds and exposed earth

By contrasting a map created from remotely sensed data with another map acquired from another source, image accuracy is evaluated. The landscape frequently changes quickly. As a result, gathering the ground reference as soon as feasible after the date of remote sensing data acquisition is ideal. The most popular method for evaluating a classified map's accuracy is to generate a collection of random points from the ground truth data and use a confusion matrix to compare them to the classified data.

Classification Accuracy

The classification made using classification algorithms discussed earlier does not give always perfect result. Therefore, the classified image contains lot of errors due to labelling clusters after unsupervised classification, preparing of training areas with wrong labelling, undistinguishable classes and correlation between bands, imperfect classification algorithm, etc. Therefore, a common question of digital satellite remote sensing is: "How accurate is the classification?". An accuracy assessment of a classified image gives the quality of information that can be obtained from remotely sensed data. Accuracy assessment is performed by comparing a map produced from remotely sensed data with another map obtained from some other source. Landscape often changes rapidly. Therefore, it is best to collect the ground reference as close to the date of remote sensing data acquisition as possible.

Classification Error Matrix

One of the most common means of expressing classification accuracy is the preparation of a classification error matrix (confusion matrix or contingency table). To prepare a error matrix first task is to locate ground reference test pixels or sample collection, based on which an error matrix is formed. There are many mathematical approaches in this regard. Generally, it is suggested that a minimum of 50 samples of each land use/ landcover classes should be included. If the study area is large or the numbers of land use /landcover classes are more than 12, the sample should be 75 to 100. Data sampling can be done using various procedures such as: random, systematic, stratified random, stratified systematic unaligned, and cluster. An error matrix compares the relationship between known reference data (ground data) and the corresponding results obtained from classification.

From the error matrix several measures of classification accuracy can be calculated using simple descriptive statistics as discussed below:

a) Overall accuracy
b) Producer's accuracy
c) User's accuracy
d) Kappa coefficient (k)

a) Overall accuracy: represents the total classification accuracy. It is obtained by dividing the total numbers of correctly classified pixels by the total numbers of reference pixels. The drawback of this measure is that it does not tell us about how well individual classes are classified. The producer and user accuracy are two widely used measures of class accuracy depends on the omission and commission accuracy respectively.

b) Producer's accuracy: it refers to the probability that a certain feature of an area on the ground is classified as such. It results from dividing the numbers of pixels correctly classified in each category by the numbers of sample pixels taken for this category (column total).

c) User's accuracy, it refers to the probability that a pixel labelled as a certain class in the map is really this class. It is obtained by dividing the accurately classified pixels by the total numbers of pixels classified in this category. The producer accuracy and user accuracy are not same typically. For example, the producer accuracy of water is 100% whereas the user accuracy is 93%.

d) Kappa coefficient: it is a discrete multivariate method of use in accuracy assessment. In classification process, where pixels are randomly assigned to classes will produce a percentage correct value. Obviously, pixels are not assigned randomly during image classification, but there are statistical measures that attempt to account for the contribution of random chance when evaluating the accuracy of a classification. The resulting Kappa measure compensates for chance agreement in the classification and provides a measure of how much better the classification performed in comparison to the probability of random assigning of pixels to their correct categories.

4.4 URBAN SPRAWL

"Urban sprawl" is known as a "form of spatial development, characterized by low density, scattered and discontinuous leapfrog expansion, and a segregation of land use". It consists of prominent characteristics such as single use development, fragmentation, shape irregularity, inequality/low concentration and linear development. Hence, it has contributed significantly to convert rural land use topologies into urban land use and depleted the number of rural lands in term of sprawl is the opposite to be compact urban development as well as it more concern about unplanned and uncontrolled growth of a city.

Various indices have been identified to measure and analyse the urban sprawl, which are described in detail below:

1. Land Consumption Rate (L.C.R.)

Land Consumption Rate (L.C.R.) is an estimate of the rate at which the land is consumed by the developing area. Land consumption rate is the rate at which urbanized land or land occupied by a city/urban area changes during a period of time (usually one year), expressed as a percentage of the land occupied by the city/urban area at the start of that time. (unstats.un.org)

It is calculated from the following formula:

L.C.R = Areal extent of the city / Population

2. Population Density

Population Density shows the ratio of total population to the built – up area of a particular year. The method to measure urban sprawl focus on population density. Low population density numbers are suggestive of high degrees of urban sprawl. (unstats.un.org)

3. Urbanness

Urbanness is defined as the percent of neighbourhood that is built – up. The neighbourhood is the total considered area. The ratio of built – up area to the total area. (proceedings.esri.com)

4. Urban Expansion Index

Urban expansion index (UEI) quantifies the urban sprawl typologies observed in the study areas (**Shukla & Jain, 2019**). This method quantitatively categorizes the development pattern of recent built-up patches surrounding the past developments. The distance between the existing developed areas and newly developed built-up patches is crucial to identify the urban sprawl typology. UEI is the ratio between the length of a common boundary between past and recent built-up patches and the patch perimeter of the recent built-up patches.

UEI quantifies the built-up growth pattern into three categories such as edge expansion ($0 < UEI < 0.5$), scatter development ($UEI = 0$), and ribbon development ($0.5 < UEI < 0.75$) (**Wu et al., 2016**). Edge expansion patches are the new built-up patches formed as an extension of existing built-up patches. Scatter development occurs independently and without any overlap to the existing built-up patches. Ribbon development occurs along the 100 m proximity of major roads within an urban area. (**Chettry (2022))**

5. Landscape Expansion Index (LEI)

An infilling type refers to a situation when the space or gap between existing patches, or inside a single patch, is filled by a newly grown patch. (**Weerakoon(2017**)). The edge-expansion type refers to the expansion of a patch that grows in a unidirectional manner, forming parallel strips from an edge. If the recently developed area is discovered to be separate from the original, it would be classified as an outlying category (**Liu et al., (2010**)).

They introduce the following formula to calculate the

$$LEI=100\times A0\ /A0+Av$$

LEI - landscape expansion index
Ao - intersection between the buffer zone and the occupied category
Av - intersection between the buffer zone and the vacant category

The LEI value ranges from 0 to 100. The LEI values ranging from 1 to 50 indicate infill growth, while values from 51 to 100 indicate edge expansion. A LEI value of zero is referred to as outlying growth. **Liu et al. (2010)** employed the Land Expansion Index (LEI) to discern the classified forms of urban land expansion in a Chinese metropolitan region.

Three complementing growth indices were utilized to measure the rate, intensity, and patterns of urban expansion. **Akubia and Bruns (2019)** The following indicators were included: Average Annual Urban Expansion Rate (AUER), Urban Growth Coefficient (UGC), Expansion Intensity Index (UEII), and Urban Expansion Differentiation Index (UEDI).

6. Average Annual Urban Expansion Rate (AUER)

The AUER quantifies the average yearly increase in developed land over the whole study area within a specified time period, which includes the base year and the final year. The index provides an estimation of the rate at which the developed land in a specific region is changing, as described by **Acheampong (2017).**

The AUER calculates the mean annual rate of expansion of built-up land for the entire study area between two periods—the base year and the final year. The index yields an estimate depicting the quantum rate at which built-up land of a given region is changing.

$AUERi$= [($ULAi^{t2}$ / $ULA\ i^{t1}$)^(1/t2−t1) −1] * 100

Where,
AUERi is Annual Urban Expansion Rate;
$ULAi^{t2}$ and $ULAi^{t1}$ are the area of urban built-up land at times t2 and t1, respectively.

Once the rate of urban expansion was quantified, Urban Growth Coefficient is calculated to determine whether urban growth is sprawling or densifying. A composite metric that utilizes the rate of urban expansion was used for this exercise. The coefficient calculation is shown in the formula.

7. Urban Growth Coefficient (U. G. C.)

The Urban expansion Coefficient was computed to ascertain whether urban expansion is characterized by sprawl or densification. (**Oyesiji (2023)).** This exercise employed a composite statistic that incorporates the pace of urban expansion, as described by Rode et al. (2017). The calculation of the coefficient is demonstrated using the following formula:

UGC = Rate of Urban Expansion/Rate of Urban Population Growth

Rode et al. (2017) found that when the UGC (Urban Growth Coefficient) is greater than 1, it signifies a rapid expansion of built-up land in relation to the population growth in a specific area. Conversely, a UGC value below 1 indicates densification.

8. Urban Expansion Intensity Index (U. E. I. I.)

The UEII is a measure of the yearly average proportion of newly expanded urban built-up area in relation to the total area that underwent changes (**Lu, Guan, He,**

& Zhang, 2014). The UEII formula calculates the average annual growth area, which is adjusted based on the total area of a specific spatial unit.

$UEII_i = (ULA_i^{t2} – ULA_i^{t1}) / (TLA_i * dt) * 100$

where,

UEIIi is Urban Expansion Intensity Index of unit i;
ULA^{t2i} and ULA^{t1i} i are the areas of urban built-up land at times t_2 and t_1, respectively; TLA_i is the total land area within the study area i and dt is the study time period (i.e., t_2 and t_1).

UEII, as described by **Abdullahi and Pradhan (2017)**, serves as an indicator of the future trajectory and capacity for urban growth. It also allows for the comparison of the rate and extent of urban land-use change across different time periods. The subsequent indices were formulated as a standard for evaluating UEII output data. The expansion rates can be categorized as follows: very slow expansion (<0.28), slow expansion (0.28–0.59), medium-speed growth (0.5–1.05), high-speed expansion (1.05–1.92), and very high-speed expansion.

9. Urban Expansion Differentiation Index (U. E. D. I.)

The UEDI, as defined by **Lu et al. (2014)**, quantifies the relationship between the rate of urban expansion in a specific spatial unit and the rate of urban growth in the whole research region. The UEDI formula measures the discrepancy in urban built-up land expansion across various geographical units. This index enables the comparison of urban built-up land expansion patterns across various spatial units and assists in identifying areas of intense urbanization. The study conducted by **Qiuying et al. in (2015)**

$UEDIi = [| ULA_i^{t2} − ULA_i^{t1} | * ULA_i^{t1}] / [| ULA_i^{t2} − ULA_i^{t2} | * ULA_i^{t1}]$

Where,

$UEDI_i$ indicates the Urban Expansion Differentiation Index of unit i;
ULA_i^{t2} and ULA_i^{t1} represents the areas of urban land of unit i at times t_2 and t_1, respectively; and
ULA^{t2} and ULA^{t1} indicate the total areas of urban built-up land in the study area at times t2 and t1, respectively.

This index basically compares urban expansion of a given unit.

There are three reference categories of UEDI that can be determined: (1) if the differentiation index of the constituent spatial unit is greater than 1, it is considered a "fast" growing area compared to the region; (2) if the differentiation index of the district is less than 1, the area is classified as a "slow" growing area relative to the region; and (3) if the differentiation index of the district is equal to 1, it is regarded as a "moderate" growing area in relation to the region.

Objective 2: To study and analyse the Rainfall over a number of decades in the study area.

4.5 RAINFALL ANALYSIS

Rainfall Variability

To analyse the rainfall variability four months of the monsoon namely June, July August and September are considered. The rainfall Variability is calculated as: Variability = standard deviation *100/ mean

The results of variability of rainfall, suggests that the precipitation patterns remain relatively stable or consistent when observed for minimum variability. Conversely, when observed for high variability in rainfall pattern, it implies significant fluctuations in the amount of rain that falls over time, which can lead to more unpredictable weather conditions.

Thus, monsoon rainfall is the main source of water for irrigation, drinking, industrial and domestic purposes. So, rainfall variability is analysed for the months of June, July, August and September. The coefficient of variation (CV) in rainfall variability is a statistical measure of the dispersion of data points in a data series around the mean and it has been computed in order to investigate spatial pattern of seasonal and annual rainfall variability over the study area.

Trend Analysis Test

Statistical methods for testing and estimating trends

Typically, statistical tests that are either parametric (depending on a certain distribution) or non-parametric (independent of any distribution) can be employed to determine the presence of a statistically significant trend. A test is considered parametric if the variation measured by the test can be defined in relation to one or more parameters. Linear regression is a parametric test. Parametric testing approaches are extensively employed in classical statistics. Parametric testing requires the assumption of an underlying distribution for the data, typically the normal distribution, and assumes that data observations are independent of each other.

A long-term change analysis was performed in this study using a linear approach. A straight line was adjusted to each pixel location in the time-series data by using the least squares linear regression method. The linear regression line can be computed as follows:

$$y = a + bx$$

The extracted regression line was then used to estimate the slope, where x (time) and y (drought indices) are the explanatory variable and dependent variable, respectively, while a & b are the intercept and slope, respectively. The slope indicates the temporal change of the studied variable at the pixel level. A positive slope indicates an increasing trend, while a negative slope indicates a decreasing trend.

Mann Kendall Test

The Mann-Kendall test is a commonly employed statistical test for analysing trends in climatological and hydrological time series data. There are two benefits of utilizing this examination.

Firstly, it is a non-parametric test and does not require the data to be normally distributed.

Furthermore, the test exhibits a limited ability to detect sudden disruptions caused by unevenly distributed time series. Non-detectable data is included in the dataset by assigning them a uniform value that is smaller than the minimum measured value. Based on this test, the null hypothesis H0 posits that there is no discernible pattern in the data (it is independent and randomly arranged). This is contrasted with the alternative hypothesis H1, which implies the presence of a pattern or trend.

The computational procedure for the Mann Kendall test considers the time series of n data points and Ti and Tj as two subsets of data where i = 1,2, 3…, n-1 and j = i+1, i+2, i+3, …, n. The data values are evaluated as an ordered time series.

Each data value is compared with all subsequent data values. If a data value from a later time period is higher than a data value from an earlier time period, the statistic S is incremented by 1.

On the other hand, if the data value from a later time period is lower than a data value sampled earlier, S is decremented by 1. The net result of all such increments and decrements yields the final value of S

The Mann-Kendall S Statistic is computed as follows:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sign}(Tj - Ti)$$

$$\text{sign}(Tj - Ti) = \begin{cases} +1 \text{ if } Tj - Ti > 0 \\ 0 \quad \text{if } Tj - Ti = 0 \\ -1 \text{ if } Tj - Ti < 0 \end{cases}$$

Where, Tj and Ti are the annual maximum daily values in years j and i, j > i, respectively.

If n < 10, The value of |S| is directly compared to the theoretical distribution of S, as determined by Mann and Kendall. The two-tailed test is employed. At a given significance level, H0 is rejected in favour of H1 when the absolute value of S is equal to or greater than a specified value, Sα/2. Sα/2 represents the smallest S value that has a probability less than α/2 of occurring in the absence of a trend. A positive number of S implies an upward trend, whereas a negative value of S suggests a downward trend. For n ≥ 10, the statistic S is approximately normally distributed with the mean and variance as follows:

E(S) =0

The variance (σ^2) for the S-statistic is defined by:

$$\sigma^2 = \frac{n(n-1)(2n+5) - \sum t_i(i)(i-1)(2i+5)}{18}$$

In which t_i denotes the number of ties to extent i. The summation term in the numerator is used only if the data series contains tied values. The standard test statistic Zs is calculated as follows:

$$Z_s = \begin{cases} \frac{s-1}{\sigma} & \text{for } S > 0 \\ 0 & \text{for } S = 0 \\ \frac{s+1}{\sigma} & \text{for } S < 0 \end{cases}$$

The test statistic Zs is used a measure of significance of trend. In fact, this test statistic is used to test the null hypothesis, H0. If | Zs| is greater than Zα/2, where α represents the chosen significance level (e.g.: 5% with $Z_{0.025}$ = 1.96) then the null hypothesis is invalid implying that the trend is significant.

Sen's Slope Estimator Test

Sen's Slope Estimator Test: The magnitude of trend is predicted by the Sen's estimator. Here, the slope (Ti) of all data pair is computed as (sen, 1968)

$$T_i = \frac{x_j - x_k}{j - k}$$

Where, x_j and x_k are considered as data values at time j and k (j>k) correspondingly. The median of these N values of Ti is represented as Sen's estimator of slope which is given as:

$$Q_i = \begin{cases} T_{\frac{N+1}{2}}, & \text{N is odd} \\ T_{\frac{N}{2}} + T_{\frac{N+2}{2}}, & \text{N is even} \end{cases}$$

Sen's estimator is computed as Q_{med} = T (N+1)/2 if N appears odd, and it is considered as Q_{med} = [$T_{N/2}$ + $T_{(N+2)/2}$]/2 if N appears even. At the end, Q_{med} is computed by a two-sided test at 100(1-α) % confidence interval and then a true slope can be obtained by the non-parametric test. Positive value of Qi indicates an upward or increasing trend and a negative value of Qi gives a downward or decreasing trend in the time series.

Objective 3: To determine and analyse the runoff in the study area using quantitative techniques like Soil Conservation Services-Curve Number (SCS-CN) method using Remote Sensing and Geographic Information System (GIS).

4.6 COMPUTATIONS OF RUNOFF

The National Engineering Handbook (NEH-4) released in 1956 by the U.S. Department of Agriculture's Soil Conservation Service (now known as the Natural Resources Conservation Service) contains documentation on the Soil Conservation Service Curve Number (SCS-CN) approach, which was created in 1954. It is one of the most popular method used techniques for estimating the amount of surface runoff from small urban, forest, and agricultural watersheds during a certain rainfall event. For ungauged watersheds, the method is stable and advisable to execute. Because it considers the majority of the features of runoff-producing watersheds—soil type, land use/treatment, surface condition, and antecedent moisture condition—it is widely applicable and acceptable.

The ratio of actual infiltration (F) to potential maximum retention (S) determines the amount of direct surface runoff (Q) relative to total rainfall (P). The second hypothesis establishes a connection between the initial abstraction (Ia) and the probable maximum retention. Consequently, the SCS-CN approach comprises the following components:

(a) Water Balance Equation:

$$P = I_a + F + Q$$

(b) Proportional Equality Hypothesis:

$$Q/P\text{-}I_a = F/S$$

(c) I_a -S hypothesis

$$I_a = \lambda S$$

Where P= total rainfall; I_a =initial abstraction; F= cumulative infiltration excluding I_a; Q= direct runoff; and S= potential maximum retention or infiltration.

$$Q = (P - I_a)^2 / (P - I_a + S)$$

Equation is valid for $P \geq I_a$. For $\lambda = 0.2$, the equation can be written as:

$$Q = (P - 0.2\ S)^2 / (P + 0.8\ S)$$

Above equation is the rainfall-runoff relationship used in the Curve Number Method. By knowing the probable maximum retention value S, one may estimate the runoff depth based on the rainfall depth. This greatest potential retention mostly refers to the infiltration that takes place once runoff has commenced. The infiltration process is governed by either the rate of infiltration at the soil surface, the rate of transmission in the soil profile, or the water-storage capacity of the profile, depending on which factor is notably limiting.

Thus, the existing SCS-CN method is a one parameter model for computing surface runoff from daily storm rainfall, for the method was originally developed using daily rainfall-runoff data of annual extreme flows. S is a constant and is the maximum difference of (P-Q) that can occur for the given storm and watershed condition. S is limited by either the rate of infiltration at the soil surface or the amount of water storage available in the soil profile, whichever gives the smaller S value. Since parameter S can vary in the range of $0 \leq S \leq \infty$, it is mapped into a dimensionless curve number (CN), varying in a more workable range $0 \leq CN \leq 100$, as follows: (Actually, the CN limit should be $0 < CN \leq 100$)

$$S = (25400/CN) - 254$$

The underlying difference between S and CN is that the former is a dimensional quantity (L) whereas the latter is a non-dimensionless quantity. The CN theoretically varies from 0 to 100.

The **Curve Number** is a dimensionless parameter indicating the runoff response characteristic of a drainage basin. In the Curve Number Method, the CN is related to land use, land treatment, hydrological soil group, and antecedent soil moisture condition in the drainage basin.

CN = 25400/(254+S)

Factors Affecting the Curve Number are given below:

(a) Hydrological Soil Group

Soil properties greatly influence the amount of runoff. In the SCS method, these properties are represented by a hydrological parameter: the minimum rate of infiltration obtained for a bare soil after prolonged wetting. The influence of both the soil's surface condition (infiltration rate) and its horizon (transmission rate) are thereby included. This parameter, which indicates a soil's runoff potential, is the qualitative basis of the classification of all soils into four groups. The Hydrological Soil Groups, as defined in the SCS-CN method, are:

Group A**:** Soils having high infiltration rates even when thoroughly wetted and a high rate of water transmission. Examples are deep, well to excessively drained sands or gravels.

Group B**:** Soils having moderate infiltration rates when thoroughly wetted and a moderate rate of water transmission. Examples are moderately deep to deep, moderately well to well drained soils with moderately fine to moderately coarse textures.

Group C: Soils having low infiltration rates when thoroughly wetted and a low rate of water transmission. Examples are soils with a layer that impedes the downward movement of water or soils of moderately fine to fine texture.

Group D: Soils having very low infiltration rates when thoroughly wetted and a very low rate of water transmission. Examples are clayey soils with a high swelling potential, soils with a permanently high-water table, soils with a clay pan or clay layer at or near the surface, or shallow soils over nearly impervious material.

(b) Practice in relation to Hydrological Condition

Land treatment mostly pertains to agricultural land utilization. It encompasses mechanical techniques like contouring or terracing, as well as management techniques such as crop rotation, grazing control, or burning.

Rotations are planned sequences of crops (row crops, small grain, and close-seeded legumes or rotational meadow). The hydrological rotations vary in quality from low to high. Inadequate rotations typically refer to the utilization of one type of crop (monoculture) or a combination of row crops, small grains, and fallow periods on a piece of land. Effective rotations typically involve the inclusion of legumes or grasses that are planted in close proximity.

The hydrological state is categorized as poor, fair, or good for the purpose of controlling grazing and conducting controlled burns in pasture ranges and forests. The classification of pasture range as bad occurs when it is severely grazed and less than 50% of the area is covered. It is categorized as fair when it is not heavily grazed and between 50% to 75% of the area is covered. Lastly, it is classed as good when it is lightly grazed and more than 75% of the area is covered. Woodlands are categorized as impoverished when subjected to intense grazing or frequent burning; as moderate when grazed but not burned; and as thriving when shielded from grazing.

(c) Antecedent Moisture Condition

Another major aspect determining the final CN value is the soil moisture condition in the drainage basin prior to runoff. The Curve Number Method categorizes the soil moisture status into three classes known as Antecedent Moisture status (AMC) Classes:

AMC I: The soils in the drainage basin are practically dry (i.e. the soil moisture content is at wilting point).

AMC II: Average condition.

AMC III: The soils in the drainage basins are practically saturated from antecedent rainfalls (i.e. the soil moisture content is at field capacity).

These classes are based on the 5 – day antecedent rainfall (i.e. the accumulated total rainfall preceding the runoff under consideration), as illustrated in Table 4.2. In the SCS method, a distinction was made between the dormant and the growing season to allow for differences in evapotranspiration.

1. Growing Season:

During the growing season, vegetation is present, and the land surface has a higher capacity to absorb and retain rainfall. This is typically associated with periods when vegetation is growing and actively transpiring, leading to lower runoff potential.

2. Dormant Season:

The dormant season refers to a period when vegetation is not actively growing, such as during the winter months or in regions with a dry season. During this time, the land surface may have a lower capacity to absorb rainfall, and runoff potential can be higher compared to the active season.

In the SCS CN method, the curve number is adjusted based on whether the area is in a growing or dormant season. Different curve numbers are assigned for different land use and land cover conditions during these two states. The purpose is to account for the seasonal variability in land surface conditions and its impact on runoff generation.

Table 4.2: Seasonal rainfall limits for AMC classes (**SCS, 1972**)

Antecedent Moisture Condition	**5 Day Antecedent rainfall (mm)**		
Class	Dormant Season	Growing Season	Average
I	< 13	< 36	< 23
II	13-28	36-53	23-40
III	> 28	> 53	> 40

(d) Land Use/ Land Cover

Table 4.3 shows land use categories and associated curves number according to soil group.

Table 4.4 shows the land cover categories and associated curve numbers. For the present study the, the values of curve numbers for different land use/ land cover categories are chosen from these tables.

(e) Agricultural Management Practices

Agricultural management system involves different types of tillage, vegetation, and surface cover. Mouldboard plough increases soil porosity from 10-20%, depending on the soil texture and, in turn, increases infiltration rates as compared to those for the non-tilled soils. Also, an increase in the organic matter content in the soil lowers the bulk density or increases porosity, and hence increases infiltration and in turn, decreases the runoff potential.

Table 4.3: Land Use Categories and Associated Curve Numbers

(TR – 55)

Land-use description	*Hydrologicc soil group* A	B	C	D
Cultivated land[1]: without conservation treatment	72	81	88	91
with conservation treatment	62	71	78	81
Pasture or range land: poor condition	68	79	86	89
good condition	39	61	74	80
Meadow: good condition	30	58	71	78
Wood or forest land: thin stand, poor cover, no mulch	45	66	77	83
good cover[2]	25	55	70	77
Open spaces, lawns, parks, golf courses, cemeteries, etc.				
good condition: grass cover on 75 percent or more of the area	39	61	74	80
fair condition: grass cover on SO to 75 percent of the area	49	69	79	84
Commercial and business areas (85 percent impervious)	89	92	94	95
Industrial districts (72 percent impervious)	81	88	91	93
Residential[3]:				
Average lot size — Average percent impervious[4]				
1/8 acre or less — 65	77	85	90	92
1/4 acre — 38	61	75	83	87
1/3 acre — 30	57	72	81	86
1/2 acre — 25	54	70	80	85
1 acre — 20	51	68	79	84
Paved parking lots, roofs, driveways, etc.[5]	98	98	98	98
Streets and roads:				
paved with curbs and storm sewers[5]	98	98	98	98
gravel	76	85	89	91
dirt	72	82	87	89

Table 4.4: Runoff Curve Numbers for Hydrologic Soil Cover Complexes

Land Use	**Cover**		**Hydrologic soil group**			
	Treatment or practice	**Hydrologic condition**	**A**	**B**	**C**	**D**
Cultivated	Straight row		76	86	90	93
Cultivated	Contoured	Poor	70	79	84	88
		Good	65	75	82	86
Cultivated	Contoured &	Poor	66	74	80	82
	Terraced	Good	62	71	77	81
Cultivated	Bunded	Poor	67	75	81	83
		Good	59	69	76	79
Cultivated	Paddy		95	95	95	95
Orchards	With understory cover		39	53	67	71
	Without understory cover		41	55	69	73
Forest	Dense		26	40	58	61
	Open		28	44	60	64
	Scrub		33	47	64	67
Pasture	Poor		68	79	86	89
	Fair		49	69	79	84
	Good		39	61	74	80
Wasteland			71	80	85	88
Roads (dirt)			73	83	88	90
Hard surface areas			77	86	91	93

[Source: K. Subramanya]

(f) Initial Abstraction and Climate

The initial abstraction consists of interception, surface detention, evaporation, and infiltration. The water held by interception, surface detention, and the infiltration at the beginning of a storm finally goes back to atmosphere through evaporation. The effect of the climatic condition of watershed is accounted for by the existing SCS-CN method in terms of the initial abstraction. It is the amount of initial abstraction for a given rainfall amount in watershed. Thus, the initial abstraction reduces the runoff potential of the watershed and the curve number.
(g) Rainfall Intensity and Duration and Turbidity

A high intensity rainfall or raindrop breaks down the soil structure to make soil fines move into the soil surface or near-surface pores, leading to the formation of crust that impedes infiltration. The crust formation actually decreases the effective soil depth responsible for infiltration and also the soil porosity, decreases S or increases CN. It is for this reason that a fallow land exposed to rain, produce a higher runoff for a given rainfall amount than does the unexposed or covered land. The term turbidity refers to impurities of water that affect infiltration by the process of clogging of soil pores and consequently, affecting the soil conductivity or ease with which water is transmitted into the soil. The contaminated water with dissolved minerals, such as salts, affects the soil structure and consequently, infiltration. (Source: http://edepot.wur.nl/183157)

Objective 4: To analyse the impact of Land Use/ Land Cover changes on Runoff in the study area.

4.7 IMPACT OF LULC ON RUNOFF

To analyse the impact of land use/ land cover changes on runoff three scenarios have been considered, where in for each of the years of 1977, 1988, 1998, 2008 and 2018, five maximum daily rainfall events of each year have been considered, to compute and analyse the runoff. The purpose of considering these five values of maximum rainfall is that those values are most likely to affect the runoff in the area.

Scenario 1 – Considering 1977 as a base year and superimposing the LULC maps of 1988, 1998, 2008 and 2018 on the base map and considering the extent of area to be same as that of 1977, the changes in runoff values are analysed. The changes in runoff will definitely occur due to the changes in LULC for different years.

Scenario 2 – Considering 1988 as a base year and superimposing the LULC maps of 1998, 2008 and 2018 on the base map, the changes and considering the extent of area to be same as that of 1988, the changes in runoff values are analysed. The change in runoff will definitely occur due to the changes in LULC for different years.

Scenario 3 – Considering 1998 as a base year and superimposing the LULC maps of 2008 and 2018 on the base map, the changes and considering the extent of area to be same as that of 1998, the changes in runoff values are analysed. The change in runoff will definitely occur due to the changes in for different years.

Alternative Approach

Scenario 4: Considering 1977 as a base year, daily maximum rainfall of base year, the runoff of base year is compared to the runoff of the years 1988, 1998, 2008 and 2018 respectively by swapping their LULC maps.

Scenario 5: Considering 1988 as a base year, taking top 5 daily rainfall values from this year, daily maximum rainfall of base year, the runoff of base year is compared to the runoff of the years 1977, 1998, 2008 and 2018 respectively by swapping their LULC maps.

Scenario 6: Considering 1998 as a base year, taking top 5 daily rainfall values from this year, runoff will be computed considering the LULC of study years 1977, 1988, 2008 and 2018 and then, comparing these values of runoff with values of base year runoff i.e., 1998.

Scenario 7: Considering 2008 as a base year, taking top 5 daily rainfall values from this year, runoff will be computed considering the LULC of study years 1977, 1988, 1998 and 2018 and then, comparing these values of runoff with values of base year runoff i.e., 2008.

Scenario 8: Considering 2018 as a base year, taking top 5 daily rainfall values from this year, runoff is computed considering the LULC of study years 1977, 1988, 1998 and 2008 and then, comparing these values of runoff with values of base year runoff i.e., 2018.

Objective 5: To determine rainfall and runoff for various recurrence intervals i.e., for 2, 5, 10, 15, 20, 25, 50, 75 and 100 years return period.

4.8 DETERMINATION OF RAINFALL AND RUNOFF FOR VARIOUS RECURRENCE INTERVALS

An IDF curve is a mathematical function that correlates the intensity, duration, and frequency of rainfall. These curves are frequently employed in hydrology to predict floods and in civil engineering for designing urban drainage systems. The establishment of a rainfall Intensity-Duration-Frequency (IDF) relationship is a fundamental need for designing storm water drainage systems in urban areas. Water resource managers utilize the rainfall depths obtained from the intensity duration frequency relationship to plan, design, and operate water resource projects. In order to determine the hydrologic risks, it is crucial to evaluate the occurrence of intense rainfall and develop Intensity-Duration-Frequency (IDF) curves. The IDF is a statistical correlation between the intensity of rainfall (I), the length of rainfall (D), and the frequency of occurrence (T).

Examination of the daily and hourly rainfall data using the Gumbel distribution for statistical analysis is carried out. The IDF development, in this study used the yearly maximum daily data.

The step-by-step procedure used is given below.

- Find out year wise annual maximum daily rainfall series.
- Rank (m) the rainfall totals.
- Find out return periods (T).
  T = (N +1) / m
  where N is the number of years data.
- Plot the graph of rainfall depth v/s return period on semi log paper.

Find out the equation of regression line to find out precipitation depth.

$$R_{24}^{2} = P_{(mm)} = a\ln(T) + (b)$$

where, P = maximum daily precipitation.

- Finding out rainfall intensity for different return periods and different durations by using formulas as described as below.

The following procedure is followed for preparation of IDF curves.

Step 1: Collection of annual daily rainfall data.
Daily rainfall data from the years 1961 to 2016 is used to obtain recent trends in the rainfall intensity.

Step 2: Preparation of short duration rainfall data using IMD formula as in equation 1.
Shorter duration rainfall series is generated from Daily Rainfall data and shown in Table 4.5

$$P_t = P_{24}\left(\frac{t}{24}\right)^{\frac{1}{3}}$$

Where,
P_t = Rainfall in mm for t hours duration,
P_{24} = Daily Rainfall data in mm and
t = Shorter duration in hours

Step 3: Probability Distribution

In this study, Gumbel's Extreme Value distribution method is used for probability distribution for each selected duration data series. Chow (1951) has shown that hydrological studies can be expressed by the following equation of the hydrologic frequency analysis: The rainfall (P_T) corresponding of a given return period (T) using the Gumbel's Distribution is given by:

$$P_T = \sigma + K.S$$

Where σ = Average Annual Daily Maximum Rainfall
S = Standard Deviation of Annual Daily Maximum Rainfall
K = Frequency Factor given by:

$$K = -\frac{\sqrt{6}}{\pi}\,[0.57720 + \ln[\ln(\frac{T}{T-1})]]$$

Table 4.5: Sample of data showing Shorter Duration Rainfalls (in mm) Derived from Maximum Daily Rainfall using IMD 1/3rd rule

Year	1 hr	2 hr	6 hr	12 hr	24 hr
1981	32.10	20.22	9.72	6.12	3.86
1982	60.67	38.22	18.37	11.57	7.29
1983	44.20	27.85	13.39	8.43	5.31
1984	39.35	24.79	11.92	7.51	4.73
1985	63.86	40.23	19.34	12.18	7.67
1986	31.89	20.09	9.66	6.09	3.83
1987	36.82	23.19	11.15	7.02	4.42
1988	31.20	19.66	9.45	5.95	3.75
1989	36.40	22.93	11.02	6.94	4.38
1990	54.29	34.20	16.44	10.36	6.53
1991	28.95	18.24	8.77	5.52	3.48

Using equation for frequency factors, Frequency Factors for the return periods of 2, 5, 10, 15, 20, 25,30, 50, 75 and 100 years are calculated, and those values are 0.16, 0.72, 1.30, 1.63, 1.87, 2.04, 2.19, 2.59, 2.91 and 3.14 respectively. The above values of frequency factor are used to obtain P_T corresponding to return periods of 5, 10,15, 20, 25,30, 50, 75 and 100 years for durations of 1 to 24 Hours. Then, the intensity of rainfall (I_T) is obtained for the return period T from the below equation: $I_T = \frac{P_t}{T_d}$

Where T_d is duration in hours.
The frequency of the rainfall is usually defined by reference to the annual maximum series, which consists of the largest values observed in each year.

Step 4: Preparation of IDF Curves by using these data, Intensity Duration Frequency curves are plotted for various return periods.

IDF Curves by Empirical Equations

The IDF formulas are the empirical equations representing a relationship among maximum rainfall intensity (as dependant variable) and other parameters of interest such as rainfall duration and frequency (as independent variables). There are several commonly used functions found in the literature of hydrology applications (Chow et al., 1988), four basic forms of equations used to describe the rainfall intensity duration relationship are summarized as follows:

Talbot equation:

$$i = \frac{a}{d + b}$$

Bernard equation:

$$i = \frac{a}{d^e}$$

Kimijima equation:

$$i = \frac{a}{d^e + b}$$

Sherman equation:

$$i = \frac{a}{(d+b)^e}$$

where i is the rainfall intensity (mm/hour); d is the duration (minutes); a, b and e are the constant parameters related to the metrological conditions.
Using equation frequency factor, frequency factors for the return periods of 2, 5, 10, 15, 20, 25,30, 50, 75 and 100 years are calculated, and those values are 0.16, 0.72, 1.30, 1.63, 1.87, 2.04, 2.19, 2.59, 2.91 and 3.14 respectively. The above values of frequency factor are used to obtain P_T corresponding to return periods of 5, 10,15, 20, 25,30, 50, 75 and 100 years for durations of 1 to 24 Hours. Then, the intensity of rainfall (I_T) is obtained for the return period T from the below equation:

$$\mathbf{I_T = \frac{P_t}{T_d}}$$

where,
T_d is duration in hours.

The frequency of the rainfall is usually defined by reference to the annual maximum series, which consists of the largest values observed in each year.

Preparation of IDF Curves by using these data, Intensity Duration Frequency curves are plotted for various return periods.

- Similarly, to compute Runoff Intensity in mm/hr. at various durations and return periods, a rainfall runoff model is prepared. Various models were developed, but the non-linear model with polynomial degree 5 was best suited for this purpose. Now using the rainfall and runoff data, a non-linear, polynomial model was developed for the dataset of 70% – 30% (Training - Validation). The non-linear model equation is $Y = a + bx + cx^2 + dx^3 + ex^4 + fx^5$, where Y = Runoff and x = Rainfall and a, b, c, d, e and f are constants:

Objective 6: To predict probable Land Use/ Land Cover changes in the study area

4.9 ESTIMATION OF LULC CHANGES

To predict the probable Land Use/ Land Cover various statistical models like linear model, logarithmic model, exponential model, power model and polynomial model up to 2nd order are prepared. Then after, with all these models the probable land use/ land cover changes are estimated.