Course Name: Watershed Hydrology

Professor Name: Prof. Rajendra Singh

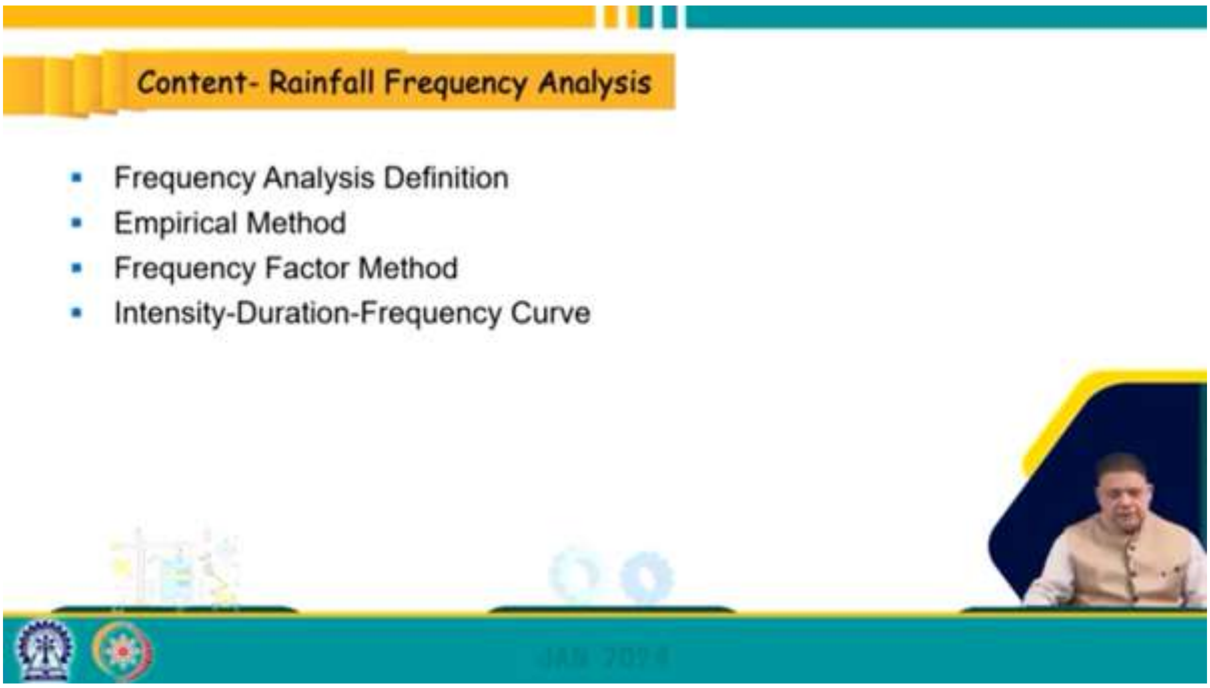
Department Name: Agricultural and Food Engineering

Institute Name: Indian Institute of Technology Kharagpur

Week: 01

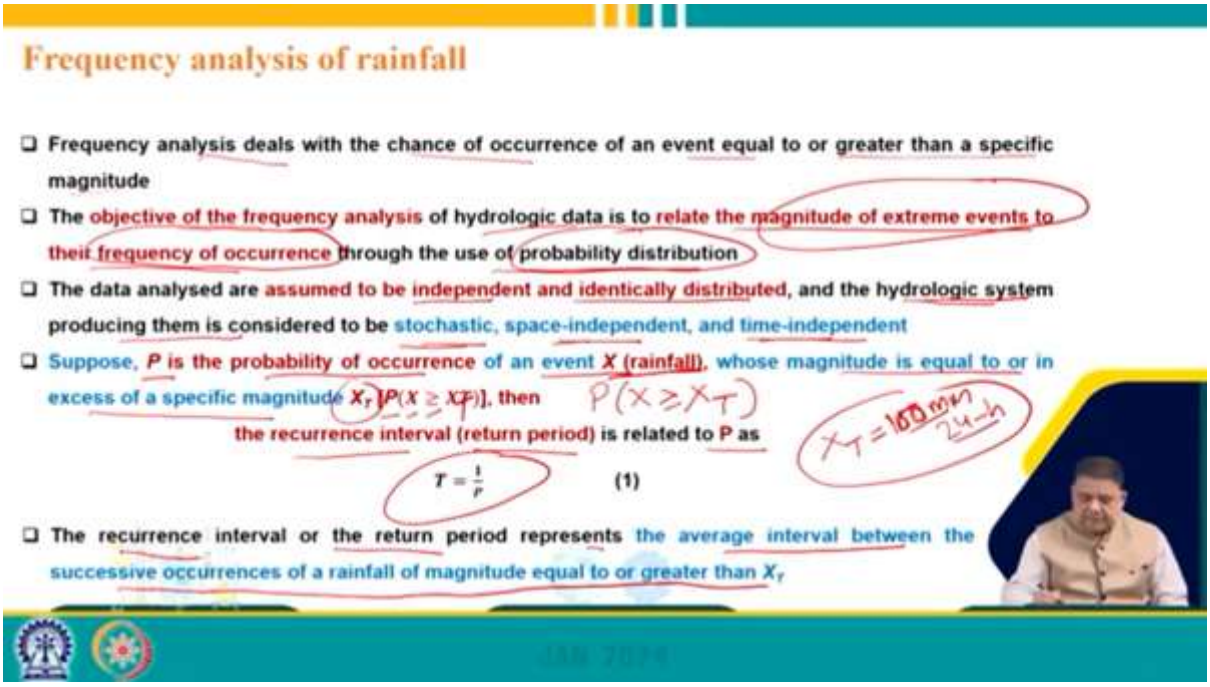
Lecture 05: Rainfall Frequency Analysis

Hello friends, welcome back to this online certification course on Watershed Hydrology. I am Rajendra Singh, a professor in the Department of Agriculture and Food Engineering at the



Indian Institute of Technology Kharagpur. We are currently in Module 1, and this is the final lecture, Lecture 5, focusing on Rainfall Frequency Analysis.

In this lecture, we will first define frequency analysis and then discuss empirical methods for carrying out frequency analysis. We will delve into the frequency factor method and conclude with a discussion on intensity-duration-frequency curves.



Now, regarding the frequency analysis of rainfall, it's important to note that while we are specifically discussing rainfall here, the principles remain applicable to any hydrological variable, whether it's floods, evapotranspiration, or any other parameter.

The concepts of frequency analysis remain consistent, with the discussion here focusing on rainfall. Regarding the definition of rainfall frequency analysis, it pertains to the probability of an event equal to or exceeding a specified magnitude occurring. If we have a certain magnitude of an event in mind, we aim to determine its frequency of occurrence, which is the purpose of frequency analysis. Therefore, the objective of frequency analysis of hydrological data is to establish a relationship between the magnitude of extreme events and their frequency of occurrence using probability distributions.

As mentioned earlier, the goal is to link the magnitude of extremes in any variable to the frequency of its occurrence, typically using probability distributions. This process is fundamentally a statistical procedure, assuming that the analysed data are independent and identically distributed. The system producing the data is considered stochastic, spatially and temporally independent. This means that the data are considered random; there is no correlation or relationship between a specific data point in a sequence and the following or preceding data points. Therefore, they are all considered independent by chance, following a certain distribution, and identically distributed, meaning that the entire data series follows a particular distribution.

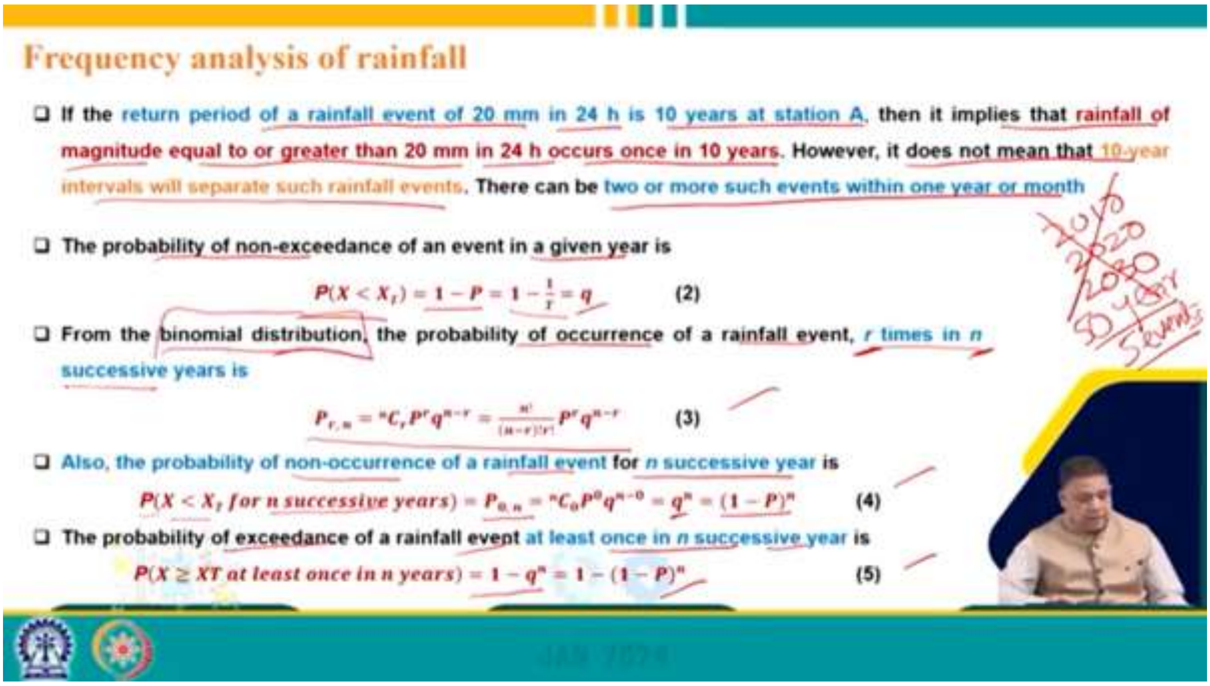
The system that produces a particular variable or data is considered stochastic, meaning there is no trend, and it is dependent solely on chance, being both space and time independent.

Let's consider P as the probability of occurrence of event X, such as rainfall, with a magnitude equal to or exceeding a specified value, denoted as x_t. This is generally expressed as P $(X \geq x_t)$, where x_t is the specified magnitude.

The recurrence interval or return period, denoted as t, is related to P as t = 1/P. This means that the return period or recurrence interval is simply the inverse of the probability of exceedance.

The return period or recurrence interval is significant as it represents the average interval between successive occurrences of rainfall with a magnitude equal to or greater than x_t

So, if we have a specific event in mind, let's say x_t equals 100 millimetres of rainfall in 24 hours, the first thing to determine is the probability of X being greater than 100 in 24 hours, denoted as P. The recurrence interval, denoted as t, is then calculated as 1/P. This recurrence interval represents the average interval between successive occurrences of rainfall of magnitude equal to or greater than 100 millimetres in 24 hours, which is also known as the return period.



For example, if the return period of a rainfall event of 20 mm in 24 hours is 10 years at station A, it means that rainfall of magnitude equal to or greater than 20 mm in 24 hours occurs once every 10 years. However, this does not imply that such rainfall events will occur exactly every 10 years; it simply indicates the average interval between successive occurrences. The return period of 10 years does not guarantee the periodicity of these events.

For instance, if a rainfall event with a 10-year return period occurred in 2010, it does not mean it will occur again in 2020 or 2030. The return period indicates the average interval between successive occurrences over a long period. For example, over a 50-year period, we can expect approximately 5 such events.

If we consider a 50-year duration, a 20-millimeter rainfall in 24 hours may occur around 5 times over that period, meaning there could be 2 or more such events within a year or month. Therefore, if an event occurred in 2010 with a recurrence interval, it may occur again in 2012 itself. The periodicity is not guaranteed, but over a long period, the average interval remains 10 years.

Regarding the probability of accidents, we have seen that the probability of non-accidents of an event, denoted as Q, where x_t is the specified magnitude, is given by (1-P) or 1 minus 1/t.

If we consider the binomial distribution, the probability of a rainfall event occurring r times in n successive years can be calculated using the formula:

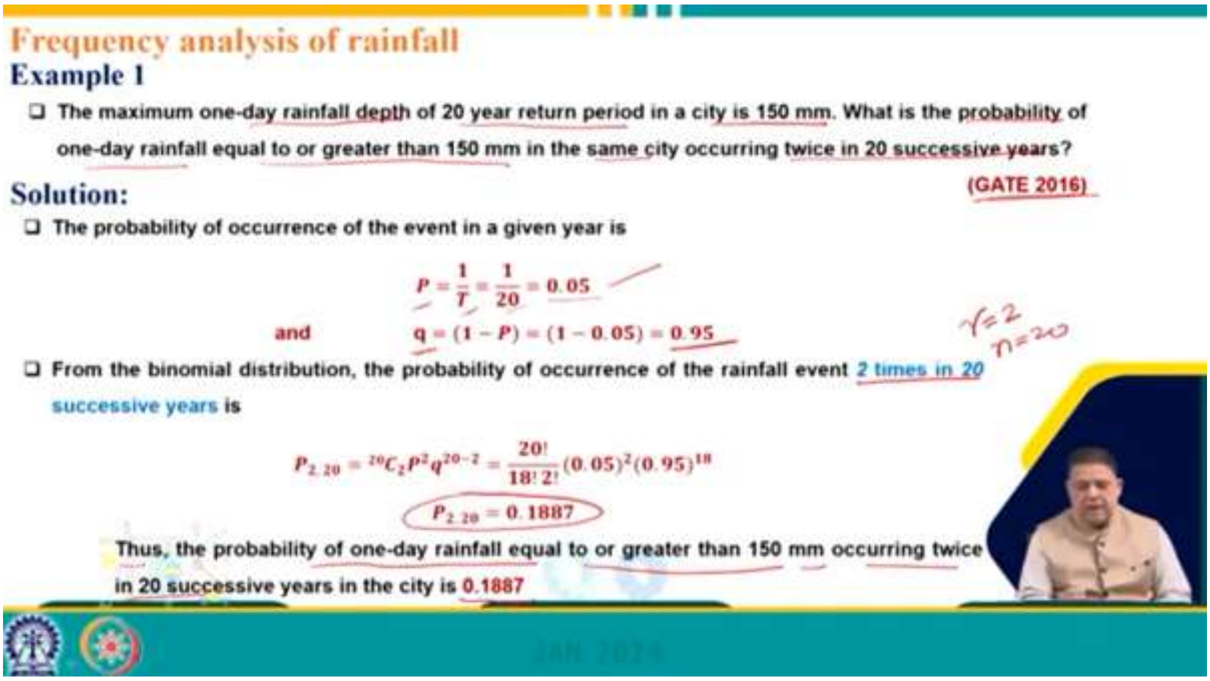
$$P\,(r,\,n) = \frac{n!}{r!(n-r)!} \times P^r \times Q^{n-r}$$

P is the probability of accidents, Q is the probability of non-accidents, n is the total number of years, and r is the number of times the event occurs.

Similarly, the probability of non-occurrence of a rainfall event for n successive years, denoted as P (0, n), can be found using the binomial distribution equation, which comes out to be Q^n or (1 - P) ^n.

Likewise, the probability of at least one occurrence of a rainfall event in n successive years can also be calculated, which is 1 - Q^n or 1 - (1 - P) ^n.

So, these three equations, 3, 4, and 5, can be used to calculate the probability of event occurrences using the binomial distribution, as we need to have a distribution in mind.

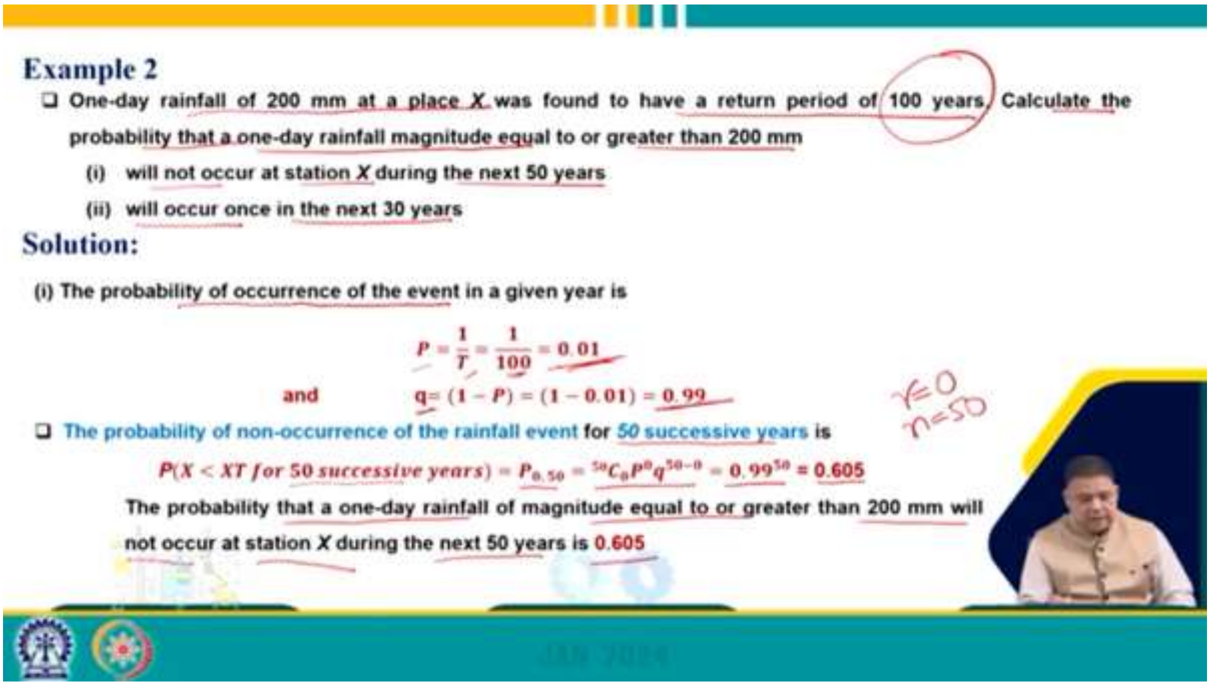


Let's consider an example: the maximum one-day rainfall depth with a 20-year return period in a city is 150 mm. What is the probability of one-day rainfall equal to or greater than 150 mm occurring twice in 20 successive years in the same city? This question is from the GATE 2016 exam. Since we are interested in the event occurring twice in 20 successive years, we can use the binomial distribution to solve this problem.

Firstly, we need to find out the probability of accidents, which we know is based on the 20-year return period.

The probability of the event occurring, P, is 1/20 or 0.05, and the probability of non-occurrence, Q, is 1 minus P, which is 0.95. Since we need to find the probability of the event happening twice in 20 years, we set R equal to 2 and n equal to 20.

Using the formula for binomial distribution, P (r, n) = nCr * P^r * Q^(n-r), we substitute the values and calculate P (2, 20) to be 0.1887. Therefore, the probability of one-day rainfall equal to or greater than 150 mm occurring twice in 20 successive years in the city is 0.1887.



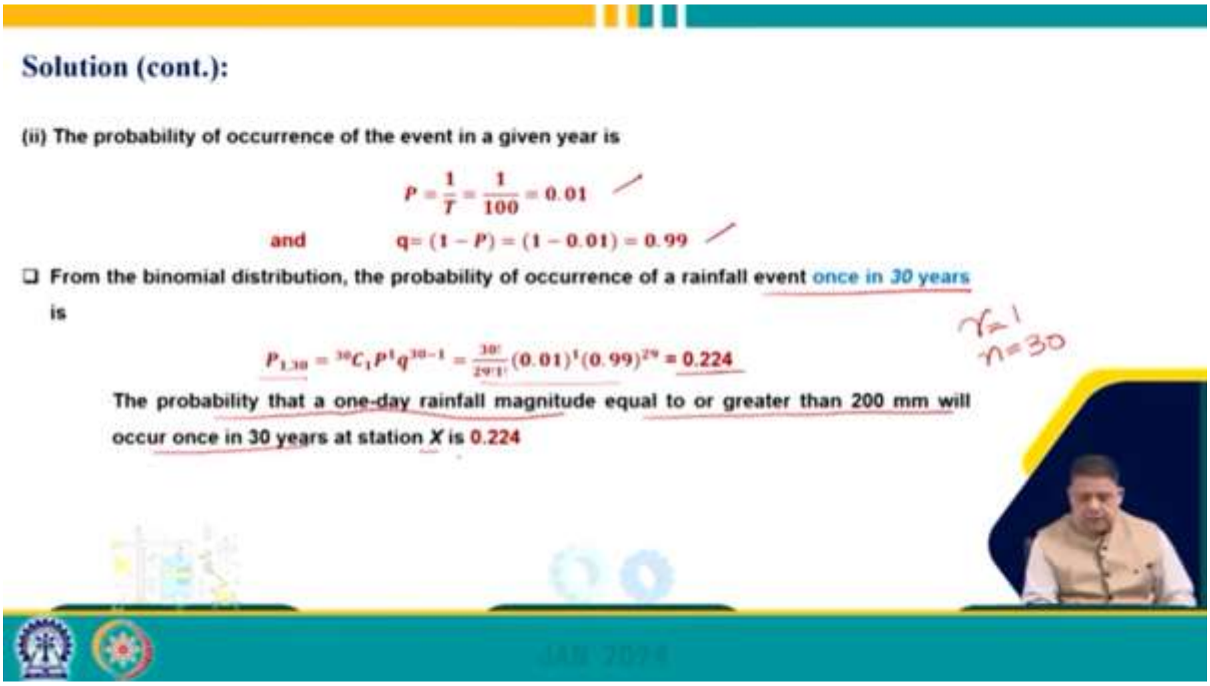
For another example, let's consider a one-day rainfall of 200 mm at a place X, with a return period of 100 years. We want to calculate the probability that a one-day rainfall magnitude equal to or greater than 200 mm will not occur at station X during the next 50 years but will occur once in the next 30 years.

Again, we can use the formulation of these two cases using a binomial distribution. Firstly, we need to calculate the probability of the event occurring, which is 1/T. In this case, the recurrence interval or return period is given as 100 years, so T equals 100. Therefore, the probability is 0.01.

The probability of non-occurrence, or non-accidents, is 1 minus P, which is 0.99. Now, the probability of non-occurrence of a rainfall event for 50 successive years, denoted as P (0, 50), can be calculated using the binomial distribution formula: nCr $\times P^r \times Q^{n-r}$.

Here, we know that R equals 0 and n equals 50. Plugging these values into the binomial distribution formula, we get a result of 0.

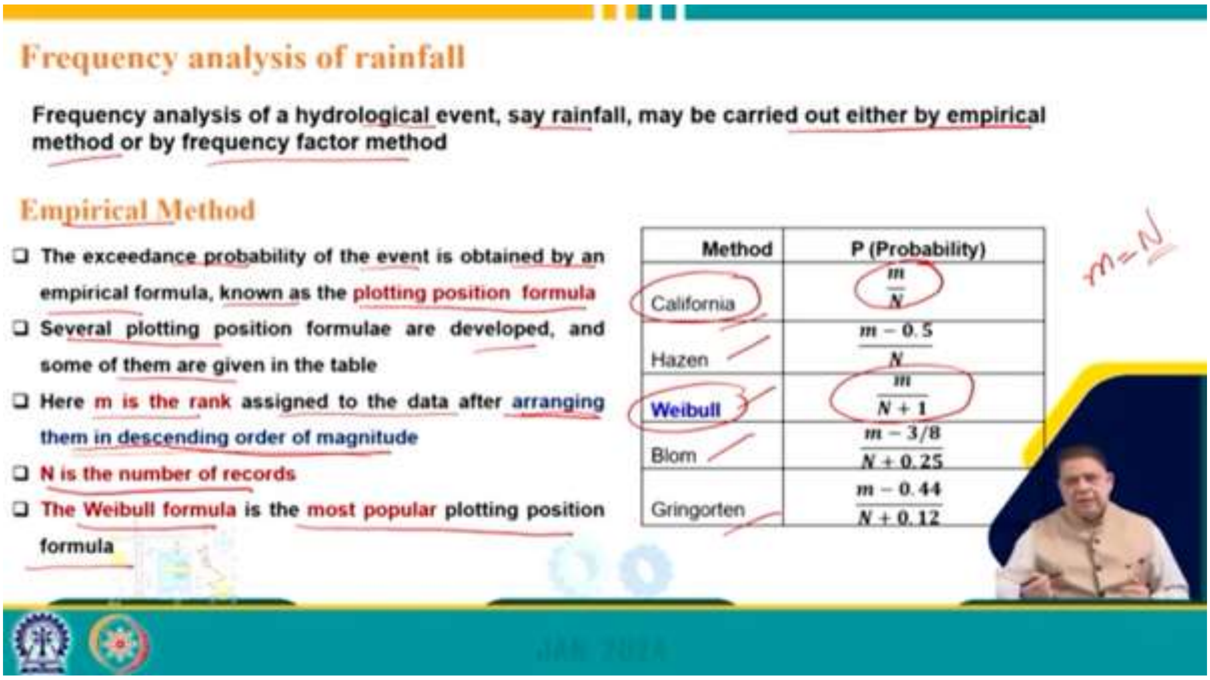
The calculation for the probability that a one-day rainfall of magnitude equal to or greater than 200 mm will not occur at station X during the next 50 years is 0.605, which is 0.99^ {50}. This means that the probability of this event not occurring is 0.605, and that answers the first part.



For the second part, the probability remains the same: P = 1/T = 0.01 and Q = 1 - P = 0.99, as we are still considering the same event. We need to find the probability of occurrence of rainfall once in 30 years, which means R is 1 and n remains 30. Using the binomial distribution formula,

P (1, 30) = $\binom{30}{1} \times 0.01^{1} \times 0.99^{29}$, we calculate the value to be approximately 0.224.

The probability that a one-day rainfall magnitude equal to or greater than 200 mm will occur once in 30 years at station X is 0.224. This is the answer to the second part.



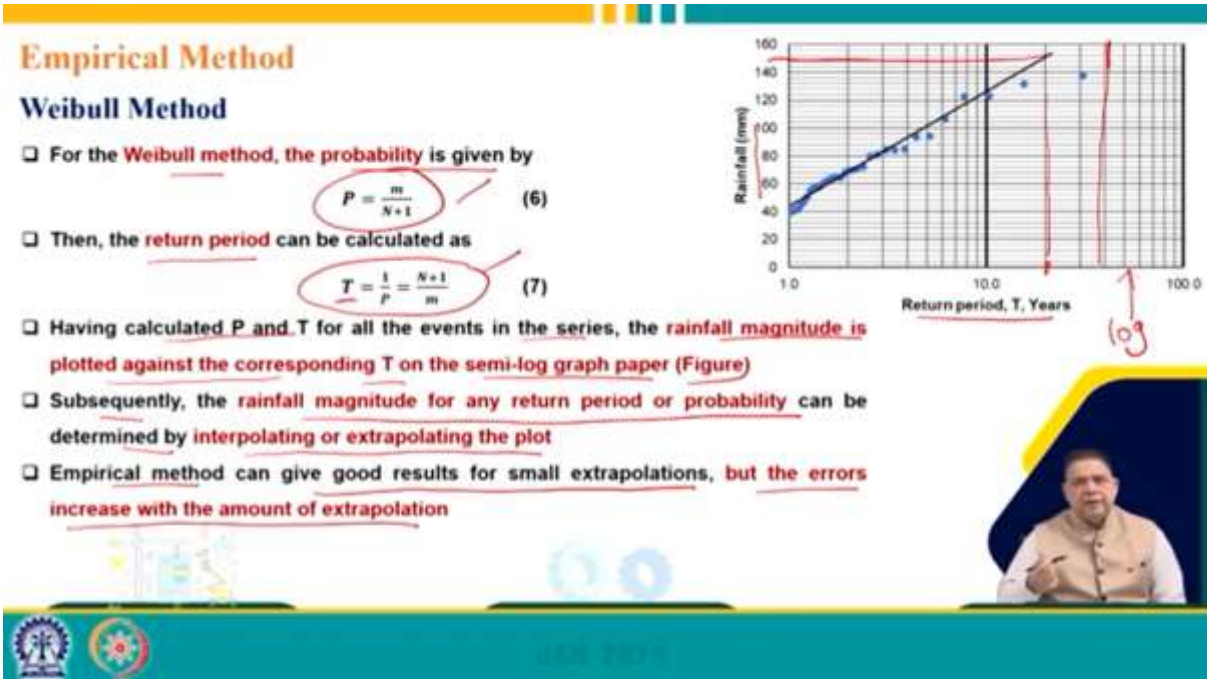
Now, let's talk about the frequency analysis of rainfall and hydrologic events. Frequency analysis can be conducted either through empirical methods or by frequency factor methods. In the frequency factor method, there are two possible ways of analysing frequency.

First, let's start with the empirical method.

In the empirical method of carrying out frequency analysis, the exceedance probability of an event is determined using a formula known as the plotting position formula. There are several plotting position formulas developed, some of which are listed in a table. For example, the California method defines the probability P as M/N, where M is the rank assigned to the data after arranging them in descending order of magnitude.

To analyse a data series using the California method, the first step is to arrange the data series in descending order of magnitude and then assign ranks to each data point. The highest value is assigned a rank of 1 (M = 1), the second-highest a rank of 2 (M = 2), and so on, with the last value assigned a rank of N (the total number of data points in the series).

Other plotting position formulas include Hazen, Weibull, Blom, Green Gorton, and several others, each involving formulations with M and N. Among these, the Weibull formula, P = M/N + 1, is the most popular for rainfall analysis.

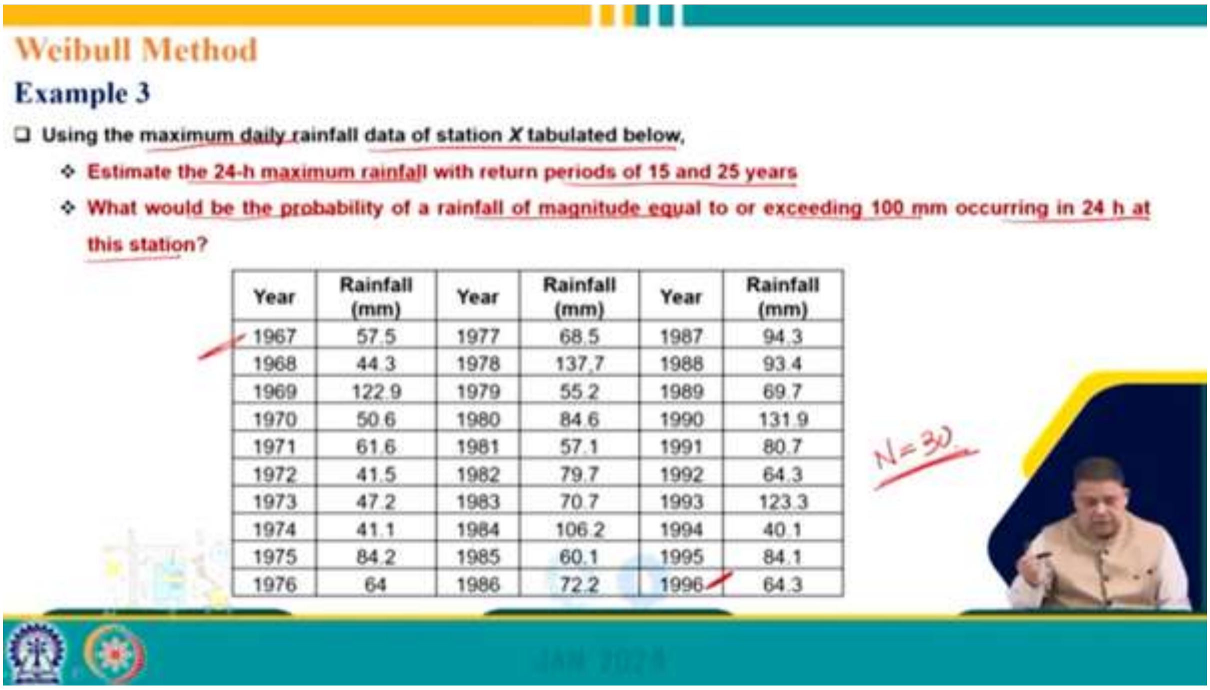


For the Weibull method, the probability is given by P = M/N + 1. The return period, T, is the inverse of P, or equivalently, P is the inverse of T. Therefore, the return period, T, can be calculated as N + 1 / M.

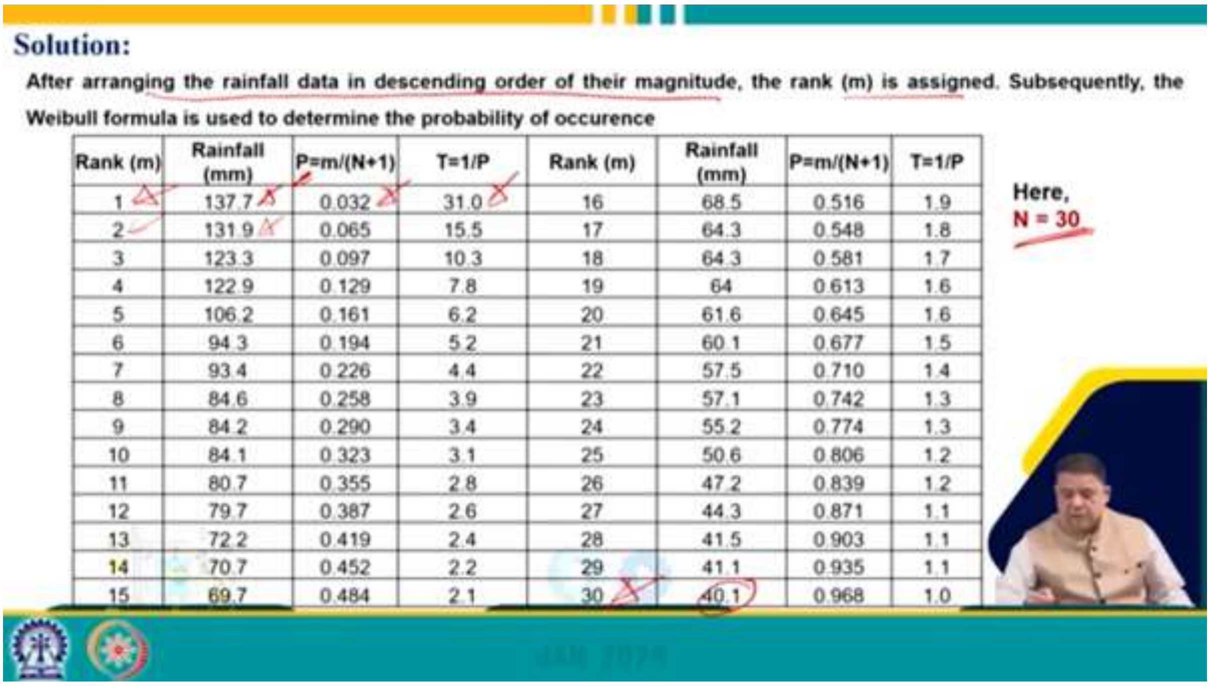
After calculating P and T for all events in the series, the rainfall magnitudes are plotted against their corresponding T values, the return periods, on semi-log graph paper as shown in the figure. On the y-axis, we plot the rainfall magnitude in millimetres, while the x-axis is logarithmic. This logarithmic scale is clearly indicated on the graph. By plotting the values, we aim to establish a straight-line relationship, as typically observed.

Subsequently, the rainfall magnitude for any desired return period or probability can be determined by interpolating or extrapolating from this plot. For example, if you need rainfall data for a 20-year return period, you can directly read it from the plot. However, if you require data for, say, a 50-year return period, extrapolation would be necessary.

Here, if you need data for 50 years, you will have to extrapolate. One drawback of the empirical method is that while it can provide accurate results for small extrapolations, the errors increase significantly with larger extrapolations. Therefore, it is advisable to avoid extrapolating whenever possible, and if extrapolation is necessary, it should be done to the smallest extent possible.

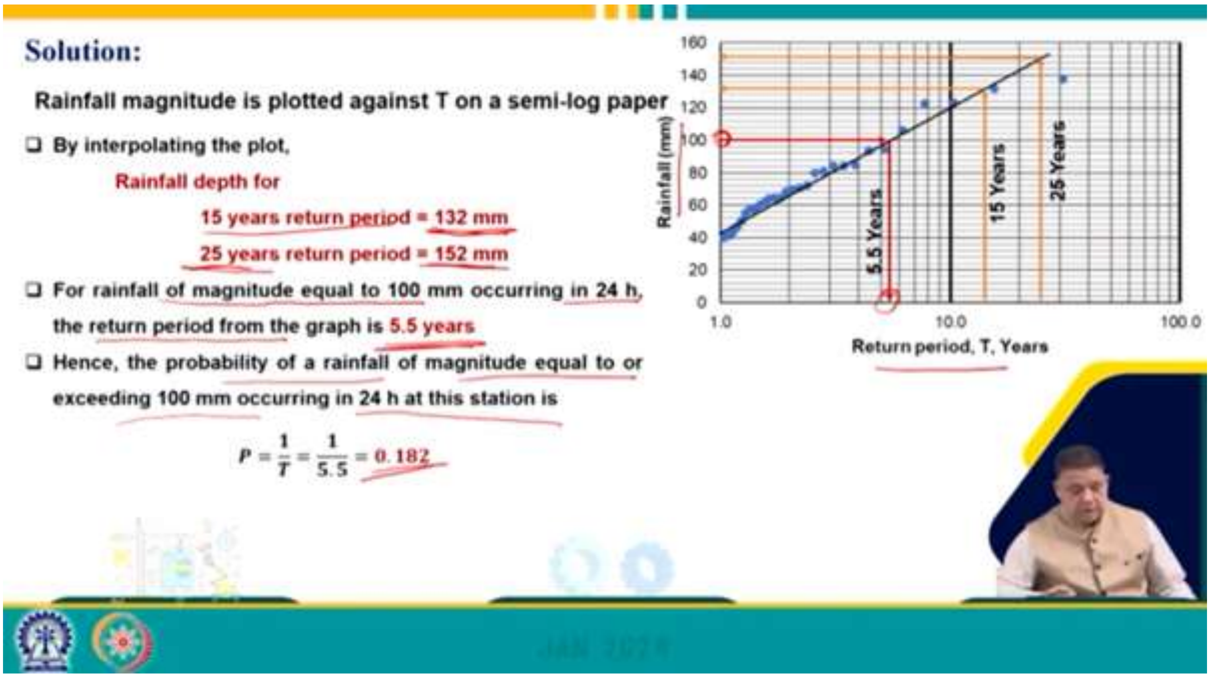


Let's consider example 3: using the maximum daily rainfall data tabulated for station X below, estimate the 24-hour maximum rainfall return periods of 15 and 25 years. Additionally, determine the probability of rainfall of magnitude equal to or exceeding 100 millimetres occurring in 24 hours at this station. The data spans from 1967 to 1996, providing 30 years of data, denoted as n = 30.



To begin, we need to arrange the data in descending order of magnitude and rank the data.

So, the first step is to scan through the data and identify the highest rainfall value, which is 137.7 in this case. This value is placed at the top, and we assign it a rank of 1. The next highest value, 131.9, is assigned rank 2, and so on, until the lowest value, 40.1, which is assigned rank 30, equal to the total number of data points (n). Once we have assigned ranks (m) to all values, we can use the Weibull's formula, P $=\frac{m}{n+1}$, to calculate the probability, and the return period, T $=\frac{1}{P}$.



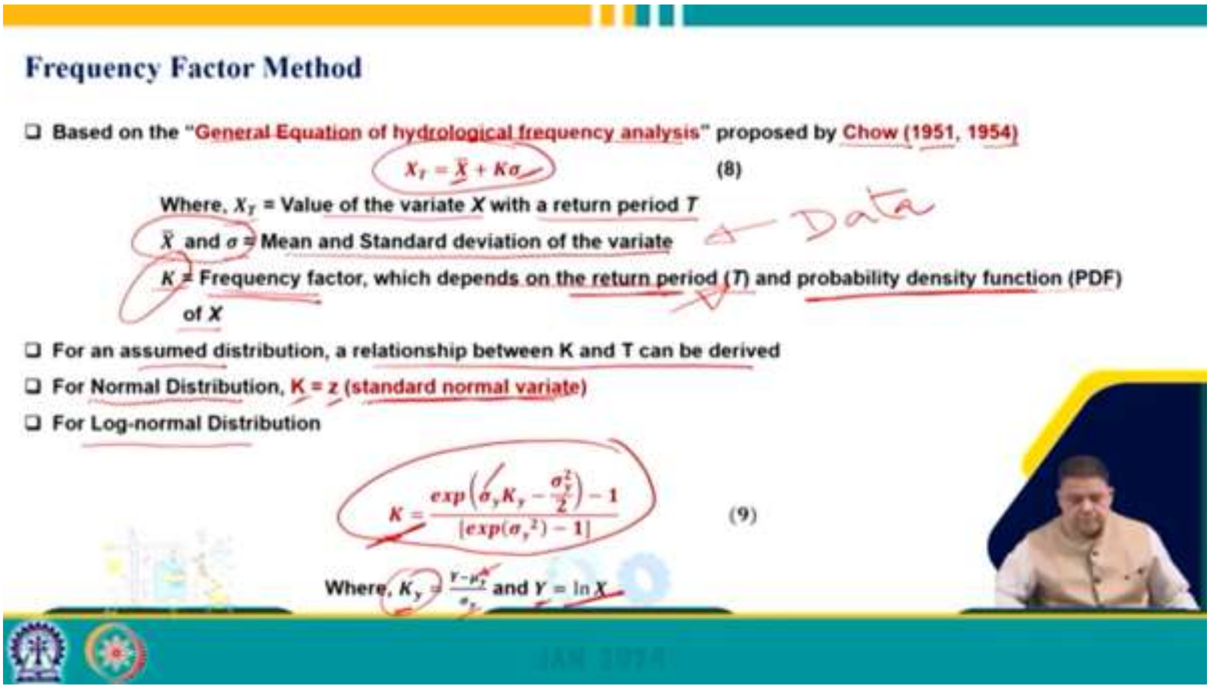
So, we can calculate the return period for all values. Once we have these values, we can plot rainfall magnitude versus return period on a graph paper and then interpolate. To find the rainfall depth for a 15-year return period, we draw a line representing 15 years and determine where it intersects the plot, giving us a rainfall magnitude of 132 mm through interpolation.

Similarly, to find the magnitude for a 25-year return period, we draw a vertical line representing 25 years and determine the corresponding rainfall magnitude, which is 152 mm.

These return periods are obtained through interpolation, as we are extrapolating beyond the available 30 years of data.

Next, to determine the return period for rainfall of magnitude equal to 100 mm occurring in 24 hours, we locate 100 mm on the plot and find the corresponding return period, which is 5 years.

Hence, the probability of rainfall of magnitude equal to or exceeding 100 years occurring in 24 hours is 1/T, which equals 0.182. This demonstrates how we can utilize the plotting position formula to obtain answers once we have plotted the data.

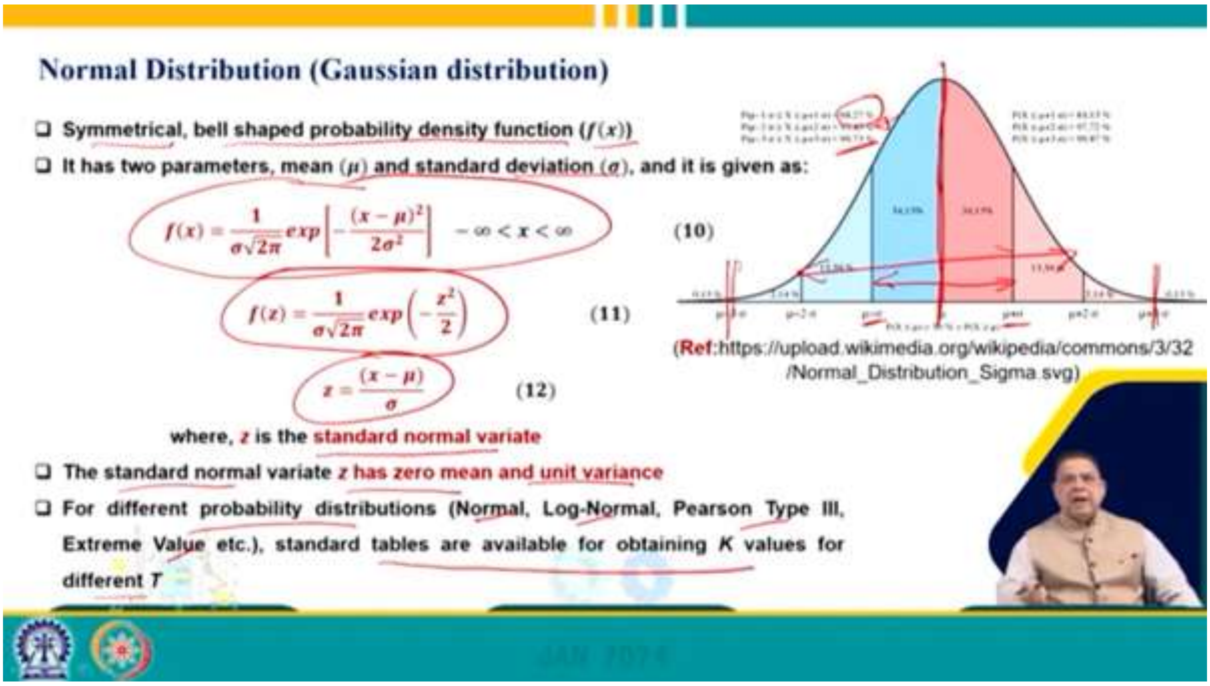


Next, let's discuss the frequency factor method. This method is based on the general equation of hydrological frequency analysis formulated by Chow, Vente Chow in 1951 and 1954. This equation, $x_t = \bar{x} + k\sigma$,relates the value of the variable x with the return period t. Here, x_t represents the value of the variable x with the return period t,$\bar{x}$ and σ are the mean and standard deviation of the variable obtained from the data series, and k is a frequency factor dependent on the return period t and the probability density function (PDF) of x. In essence, this equation allows us to derive the mean and standard deviation from the data series.

Also, we need to fit the probability distribution function to the data to determine which distribution our data fits into. There are various distribution functions available, and once we know the probability density function and the return period we are interested in, we can calculate the frequency factor k. With the knowledge of k, $\bar{x}$, and sigma, we can find the value of x for the desired return period using the general equation of hydrological frequency analysis.

For an assumed distribution, a relationship between k and t can be derived. For the normal distribution, which is commonly used, the frequency factor k is equal to z, which is a standard normal variate. This z value is familiar to those who have used the normal distribution before. Similarly, for the lognormal distribution, k can be determined using a relationship involving y, where y is ln(x), meaning the natural logarithm of x.

To apply the lognormal distribution, we first transform our data using the logarithmic function. Then, we fit the distribution and calculate the standard deviation sigma y, which is the square root of the variance, and the mean mu y. The standardized variable k y is calculated as y minus μ y divided by sigma y for each value of y. This standardized variable is used to find k for use in the general equation of hydrological frequency analysis.



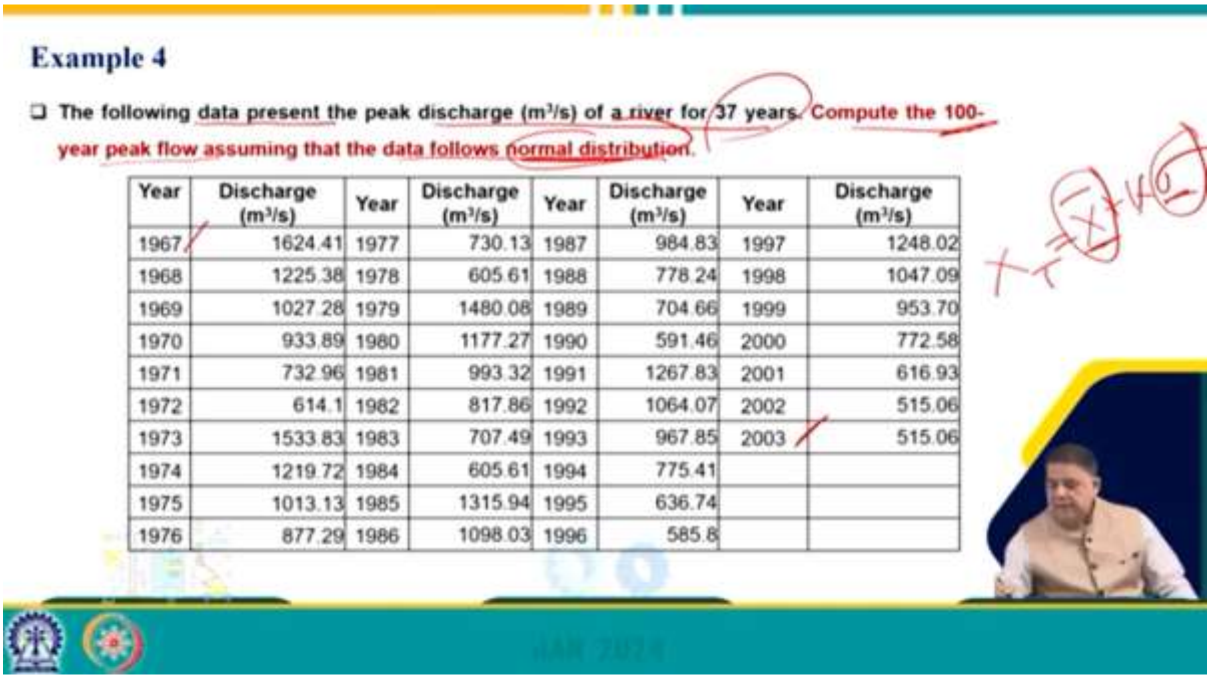
Now, to give you an idea, the normal distribution, also known as the Gaussian distribution, is a symmetrical bell-shaped probability density function, denoted by $f(x)$. It is a two-parameter model characterized by its mean and standard deviation. Mathematically, it is expressed as follows

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

You may have encountered the graph of a normal distribution in statistics or mathematics, where the peak represents the mean and the spread is determined by the standard deviation. The area between $\mu + \sigma$ and $\mu - \sigma$ covers approximately 68.27% of the total area under the curve. Similarly, the area between $\mu - 2\sigma$ and $\mu + 2\sigma$ covers about 95.47%, and the area between $\mu - 3\sigma$ and $\mu + 3\sigma$ covers nearly 99.73%.

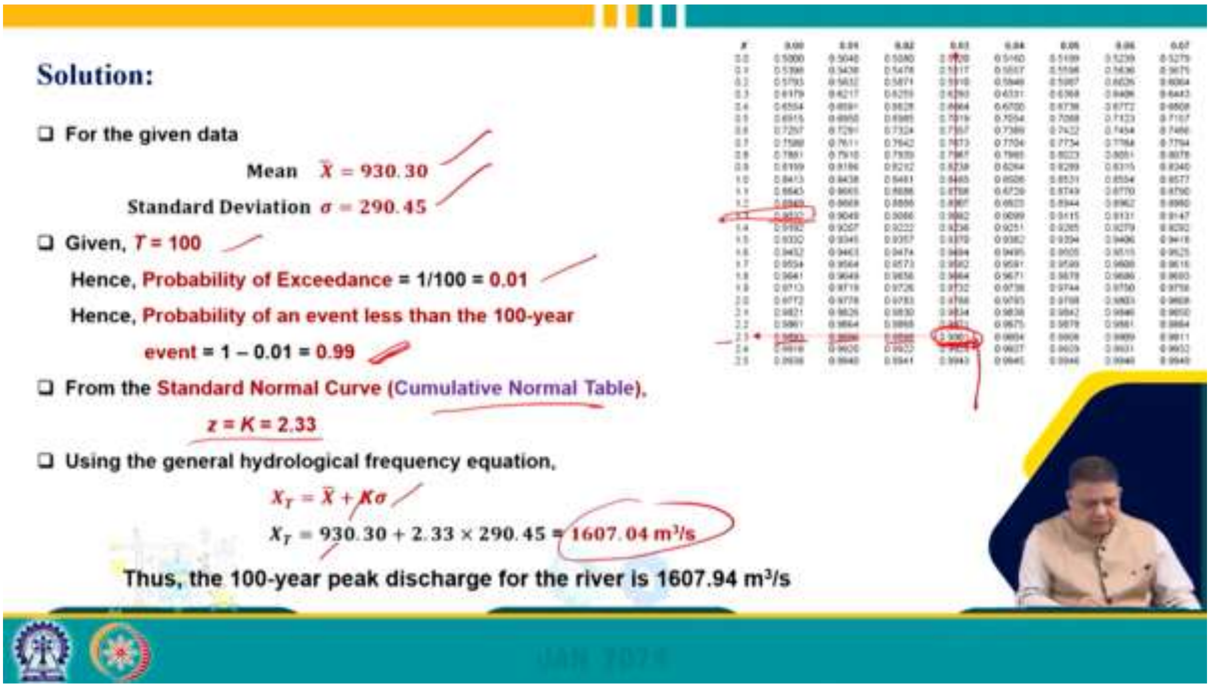
The standard normal variate, denoted by z , is defined as $z = \frac{x-\mu}{\sigma}$. It has a mean of 0 and a variance of 1. For different probability distribution functions such as normal, lognormal, Pearson type, and extreme value distributions, tables are available to obtain the frequency factor k . These tables provide the necessary information for calculations involving these distributions.

So, if you know the return period t, you can find the value of k to use in the general equation of hydrological frequency analysis. Let's consider an example: the following data presents the peak discharge of a river for 37 years. We are asked to compute the 100-year peak flow, assuming that the data follows a normal distribution.



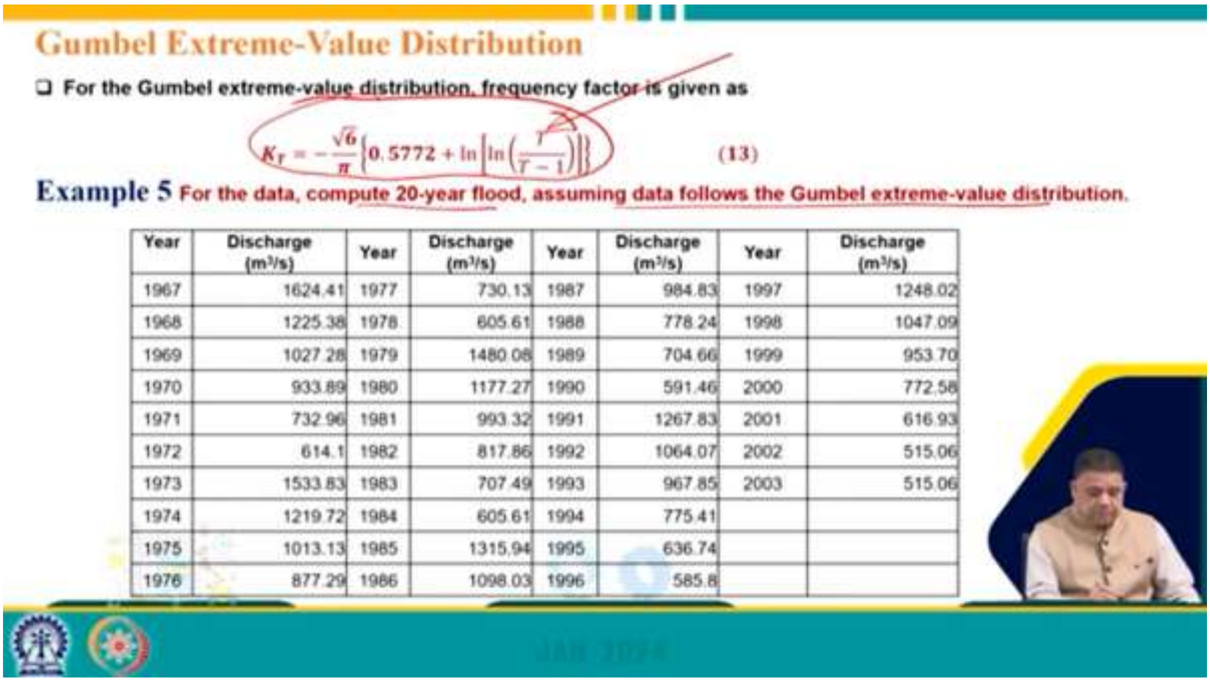
Year	Discharge (m^3/s)	Year	Discharge (m^3/s)	Year	Discharge (m^3/s)	Year	Discharge (m^3/s)
1967	1624.41	1977	730.13	1987	984.83	1997	1248.02
1968	1225.38	1978	605.61	1988	778.24	1998	1047.09
1969	1027.28	1979	1480.08	1989	704.66	1999	953.70
1970	933.89	1980	1177.27	1990	591.46	2000	772.58
1971	732.96	1981	993.32	1991	1267.83	2001	616.93
1972	614.1	1982	817.86	1992	1064.07	2002	515.06
1973	1533.83	1983	707.49	1993	967.85	2003	515.06
1974	1219.72	1984	605.61	1994	775.41		
1975	1013.13	1985	1315.94	1995	636.74		
1976	877.29	1986	1098.03	1996	585.8		

Given that the distribution is normal and 37 years of data are provided (from 1967 to 2003), we need to calculate the 100-year peak flow. Using the equation $x_t = \bar{x} + k\sigma$, we first need to determine the mean and standard deviation from the data series. The distribution function is typically fitted, but in this case, it is specified to be a normal distribution, meaning k should come from the standard normal variate. From the data, we find that, $\bar{x}$ = 930.30



3 standard deviations are 290.45. Given t = 100, the probability of accidents is 1/t, which is 0.01, and the probability of the event being less than 100 years is 0.99. To find k, we refer to the standard normal cumulative normal table, which is available in any textbook, and locate the z value corresponding to 0.99. In the table, the value for z = 2.33 is 0.9901, so k = 2.33 for 0.99. Similarly, for any other value, such as 90 percent, the nearest value is 0.9032, for example. So, we can take the value of k = 1.3 ·h. Once we know k, the mean, $\bar{x}$, standard deviation, and k, then k· x_t can be calculated as 1607.04 cubic meters per second, which is the 100-year peak

discharge for the river, or 1600.04 cubic meters per second. Similarly, we can fit an extreme value distribution, such as the Gumbel distribution. For the Gumbel distribution k· $x_t$is calculated by a formula that is a function of t only, which is the recurrence interval.



Gumbel Extreme-Value Distribution

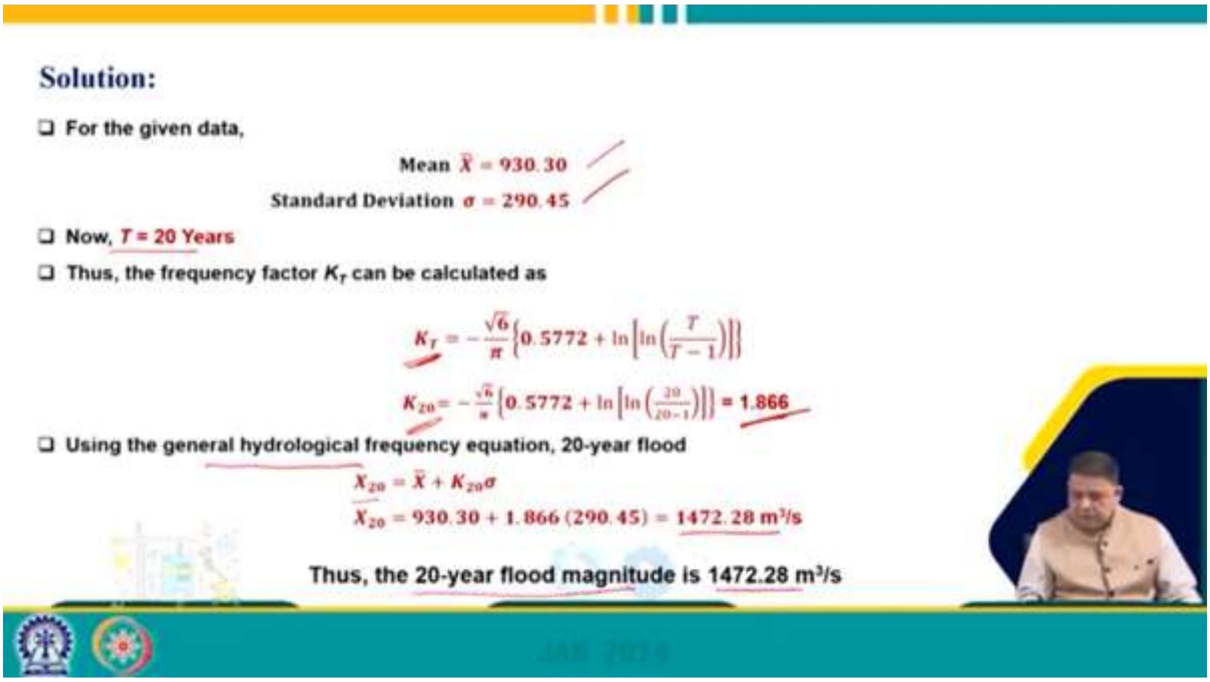
- For the Gumbel extreme-value distribution, frequency factor is given as

$$K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\} \qquad (13)$$

Example 5 For the data, compute 20-year flood, assuming data follows the Gumbel extreme-value distribution.

Year	Discharge (m³/s)	Year	Discharge (m³/s)	Year	Discharge (m³/s)	Year	Discharge (m³/s)
1967	1624.41	1977	730.13	1987	984.83	1997	1248.02
1968	1225.38	1978	605.61	1988	778.24	1998	1047.09
1969	1027.28	1979	1480.08	1989	704.66	1999	953.70
1970	933.89	1980	1177.27	1990	591.46	2000	772.58
1971	732.96	1981	993.32	1991	1267.83	2001	616.93
1972	614.1	1982	817.86	1992	1064.07	2002	515.06
1973	1533.83	1983	707.49	1993	967.85	2003	515.06
1974	1219.72	1984	605.61	1994	775.41		
1975	1013.13	1985	1315.94	1995	636.74		
1976	877.29	1986	1098.03	1996	585.8		

Another example is to compute the 20-year flood assuming the data follows the Gumbel distribution. Given the data and the calculated mean and standard deviation, we calculate the value k· $x_t$for t = 20 using the Gumbel extreme value distribution equation. This yields a k value of 1.866, and using the general hydrologic frequency equation, we find x_{20} to be 1472.28 cubic meter per second.



Solution:

- For the given data,

$$\text{Mean } \bar{X} = 930.30$$

$$\text{Standard Deviation } \sigma = 290.45$$

- Now, $T = 20$ Years
- Thus, the frequency factor K_T can be calculated as

$$K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\}$$

$$K_{20} = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{20}{20-1}\right)\right]\right\} = 1.866$$

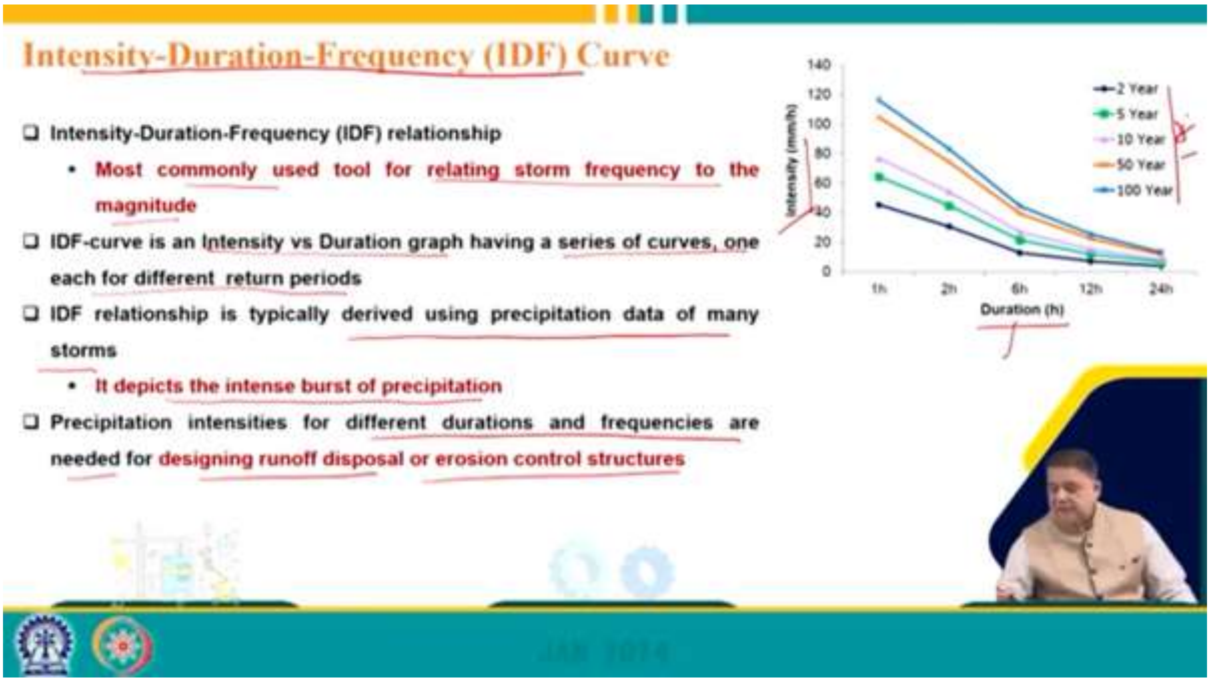
- Using the general hydrological frequency equation, 20-year flood

$$X_{20} = \bar{X} + K_{20}\sigma$$

$$X_{20} = 930.30 + 1.866\,(290.45) = 1472.28 \text{ m}^3/\text{s}$$

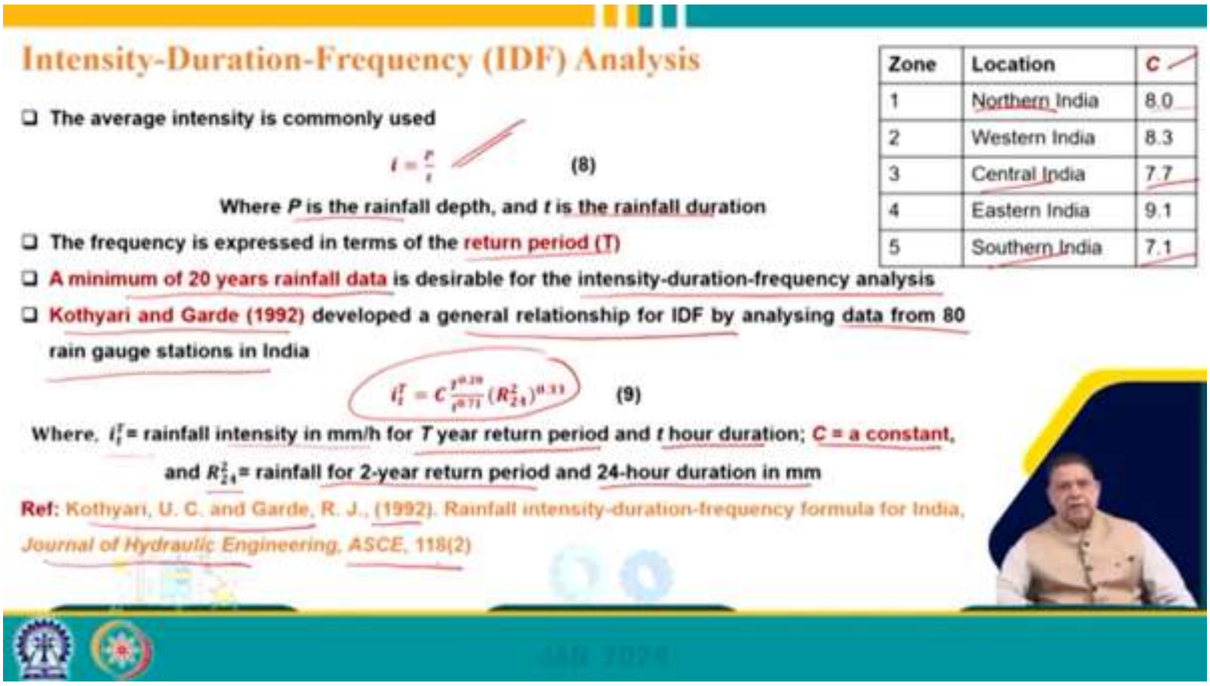
Thus, the 20-year flood magnitude is 1472.28 m³/s

The 20-year flood magnitude is 1472.28 cubic meters per second. So, that is how we can use the frequency factor method for finding out or fitting the frequency analysis.



Now, let us discuss the intensity-duration-frequency curve once we have talked about frequency. The intensity-duration-frequency curve is the most commonly used tool for relating the storm frequency to the magnitude, and it is an intensity versus duration graph with a series of curves, one for each different return period. Here, you can see that we have intensity on the y-axis and duration on the x-axis, and these are curves for different recurrence intervals or frequencies.

Using this curve, we can find out the intensity of rainfall for any recurrence interval and any duration. It is typically derived using precipitation data from many storms and depicts the intense burst of precipitation. Precipitation intensities for different duration frequencies are needed for designing runoff disposal or erosion control structures. So, anytime we design a runoff disposal or erosion control structure, we have to find out what is the 10-year maximum rainfall intensity or magnitude. Using this intensity-duration-frequency curve for different durations, we can find out what the intensity of the curve is. Typically, we use an overall intensity formula where $I = \frac{P}{T}$, where P is the rainfall depth, I is the rainfall duration, and frequency is in terms of the return period. We express that, and a minimum of 20 years of rainfall data is desired for intensity-duration-frequency analysis.



So, besides frequency analysis, people have also tried to derive general relationships of IDP. One such equation was derived by Kothari and Garde in 1992, who analysed data from 80 rain gauge stations across India and developed this relationship. Here, ITT, which is rainfall intensity in millimetres per hour for a T-year return period and T-hour duration, can be found out by T and T. Of course, we have defined R as 224 for rainfall of a 2-hour return period and 24-hour duration in millimetres. So, for that location, you have to have R 224 data, and C is a constant. Kothari and Garde also recommended values of C for different zones in the country. For the northern zone, the value of C is 8, and for the central zone, it is 7.

For southern India, it is 7.1. Using this C value, T, and T, and also knowing R 224, which is the rainfall of a 2-year return period and 24-hour duration, we can find out the rainfall intensity for the desired duration and recurrence interval. This equation is taken from Kothyari and Garde's paper published in the Journal of Hydraulic Engineering ASCE,118(2) in 1992. So, with this, we come to the end of this rainfall frequency analysis. Thank you very much for your patience in listening, and please feel free to give your feedback and also raise your doubts or questions in your forum, which can be answered. Thank you very much.