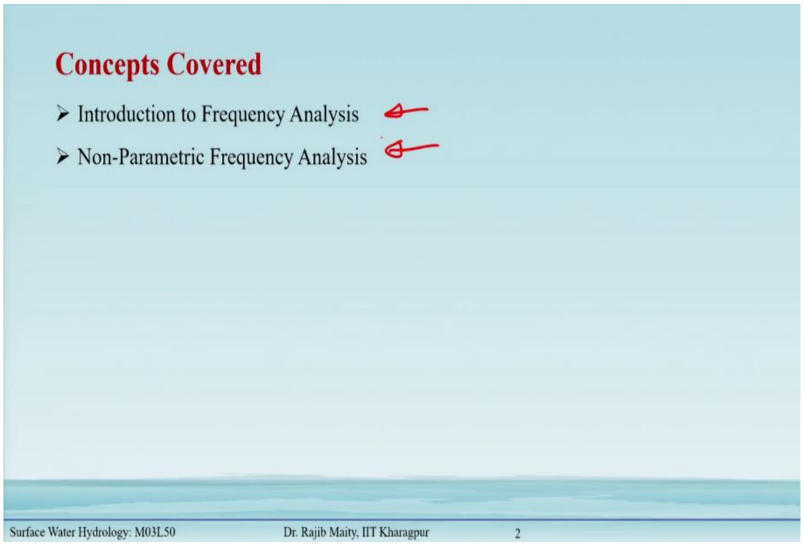
Surface Water Hydrology
Professor Rajib Maity
Department of Civil Engineering
Indian Institute of Technology, Kharagpur
Lecture 50
Introduction to Frequency Analysis

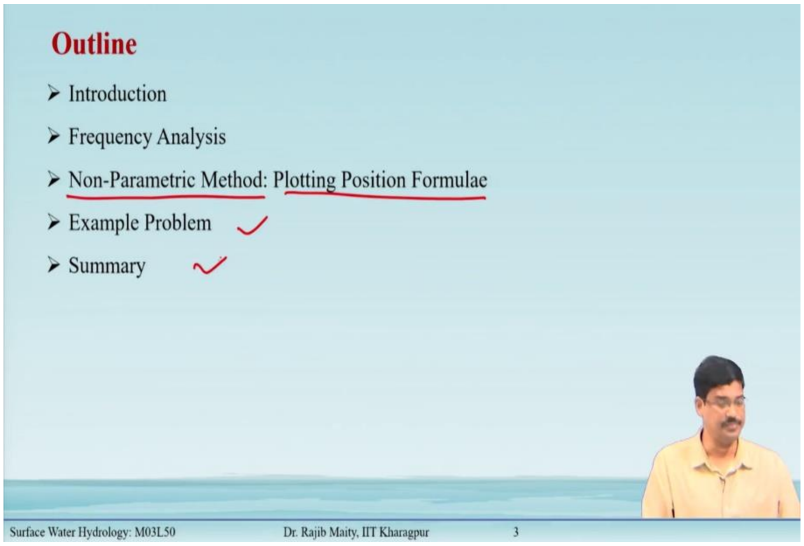
In today's lecture, we are starting with the introduction of frequency analysis that will include the nonparametric way of carrying out this exercise that we will see.

(Refer Slide Time: 00:38)



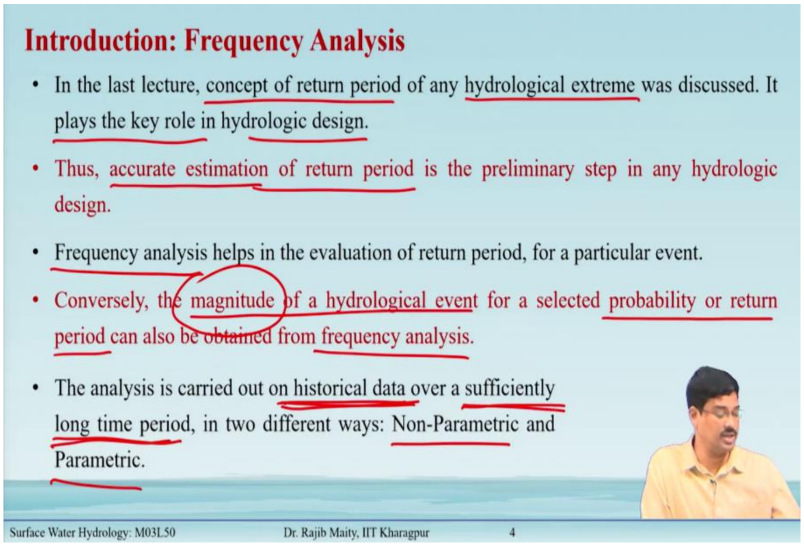
Under this concept covered, two things will be covered the first one is this frequency analysis, and out of the two major categories that are parametric and nonparametric in this lecture Non-parametric will be considered.

(Refer Slide Time: 00:54)



The outline for this lecture goes like this first introduction to this frequency analysis, then the under this Non-parametric method, that plotting position formula will be discussed and then there will be an example problem for application purpose and then coming to the summary.

(Refer Slide Time: 01:14)

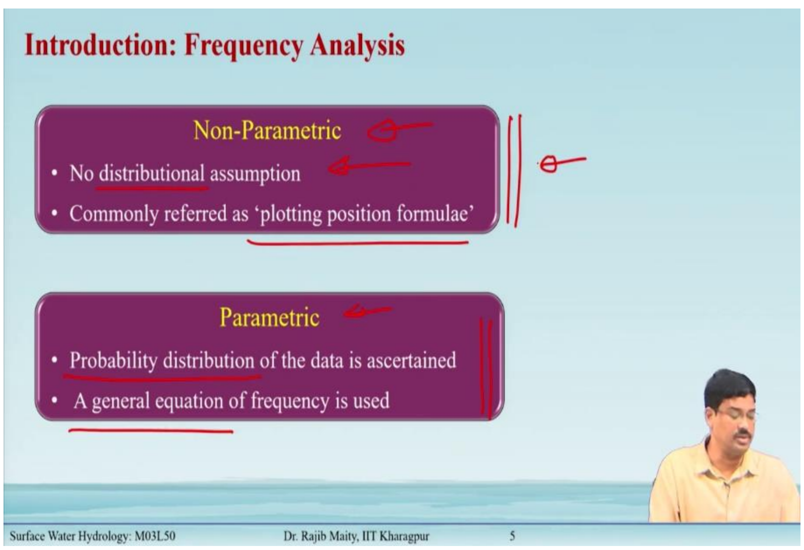


Introduction: Frequency Analysis

In the last lecture, the two important things that were discussed one them was the concept of return period and this concept of return period for any hydrological extreme event that was discussed and this is the first very important information that plays a key role in any hydrologic design. So, that, estimation of this return period, we get from this frequency analysis. So, this frequency analysis helps in the evaluation of the return period for a particular event. Conversely, it can be also stated that the magnitude of the hydrologic event for a selected probability or the return period can be obtained from the frequency analysis, the analysis is carried out on the historical data what the sufficiently long time period.

Now, there are two things here. So, we say that it should be the recorded data historical data for that particular variable and it should have a sufficiently long duration period. Now, these two things themselves sometimes create a different situation that we have to go a different way for this solution. But, if we have sufficiently long historical data, then the general method that is available to us that can be categorized into two parts, one is called the non-parametric way and the other is the parametric way.

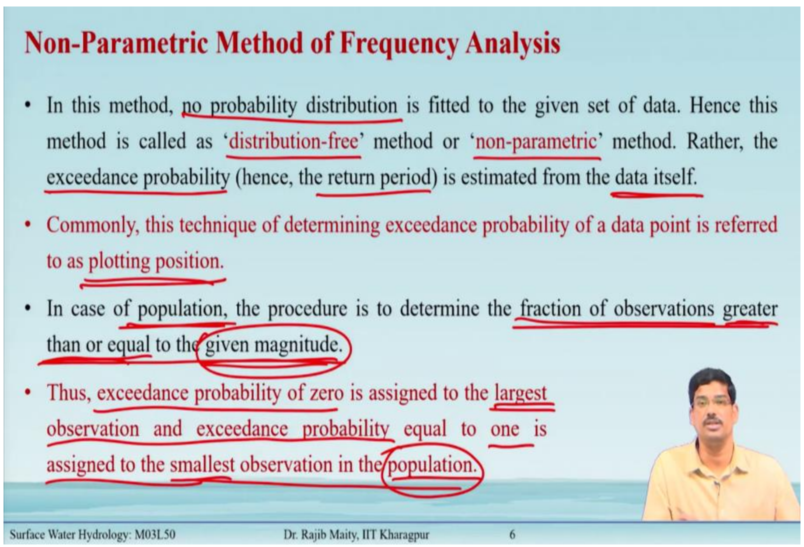
(Refer Slide Time: 03:48)



Under this non-parametric category, there is no distributional assumption here the distribution. So, no probability distributional assumption is necessary so, far as the non-parametric method is concerned, we purely depend on the data, and whatever we develop the probabilities whatever we compute different probabilities exceedance non-exceedance Probabilities. So, the estimation of the probabilities is purely based on the data itself. And from that data when you calculate those probabilities, those are there is a set of equations are available that are referred to as plotting position formulae.

On the other hand, when we go for this parametric category, the first thing that we do is that we ascertain one probability distribution that best fits the data. And then we use the expression to find out either its frequency or the magnitude for a certain return period. And for that one, a general equation of the frequency is used for that purpose.

(Refer Slide Time: 05:24)



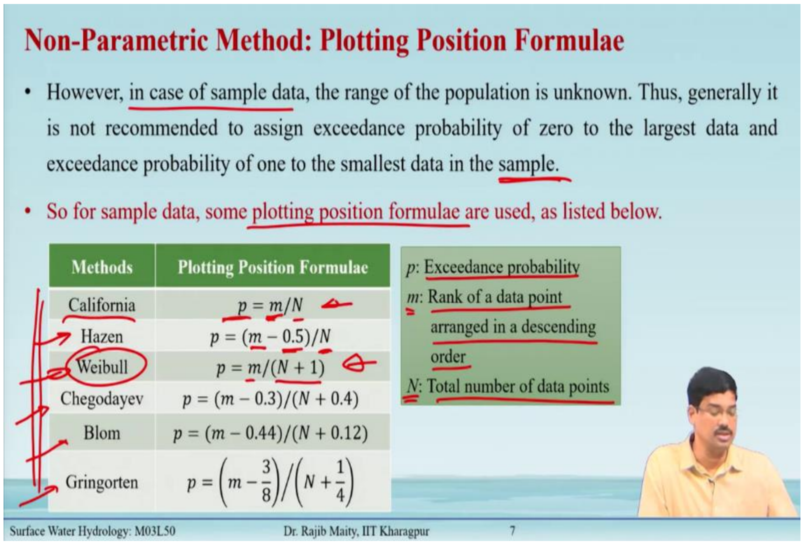
Non-Parametric Method of Frequency Analysis

In this method, no probability distribution is fitted to the given set of data. Hence this method is called a 'distribution-free' method or 'non-parametric' method. Rather, the exceedance probability (hence, the return period) is estimated from the data itself.

Commonly, this technique of determining the exceedance probability of a data point is referred to as plotting position. In the case of population, the procedure is to determine the fraction of observations greater than or equal to the given magnitude.

Thus, the exceedance probability of zero is assigned to the largest observation and the exceedance probability equal to one is assigned to the smallest observation in the population.

(Refer Slide Time: 09:00)



Non-Parametric Method: Plotting Position Formulae

However, in the case of sample data, the range of the population is unknown. Thus, generally, it is not recommended to assign an exceedance probability of zero to the largest data and an exceedance probability of one to the smallest data in the sample.

So for sample data, some plotting position formulae are used, as listed below.

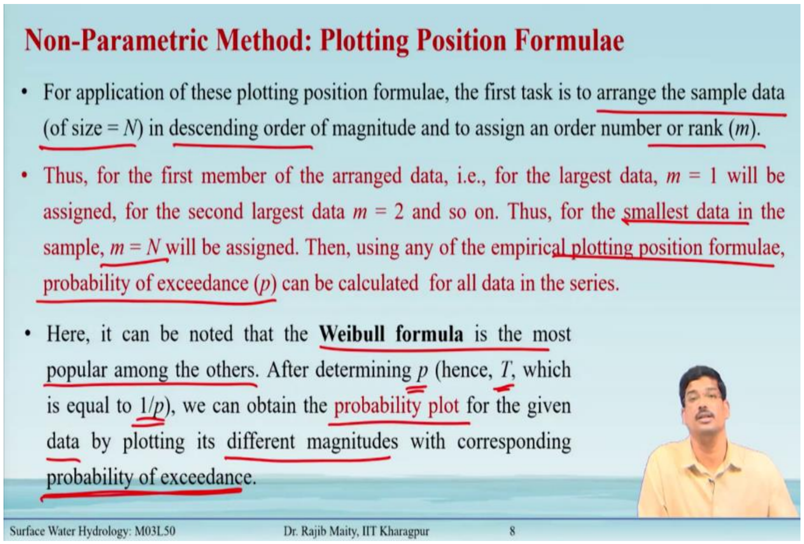
Methods	Plotting Position Formulae
California	$p = m/N$
Hazen	$p = (m - 0.5)/N$
Weibull	$p = m/(N + 1)$
Chegodayev	$p = (m - 0.3)/(N + 0.4)$
Blom	$p = (m - 0.44)/(N + 0.12)$
Gringorten	$p = \left(m - \frac{3}{8}\right)/\left(N + \frac{1}{4}\right)$

Where p: Exceedance probability

m: Rank of a data point arranged in a descending order

N: Total number of data points

(Refer Slide Time: 11:50)



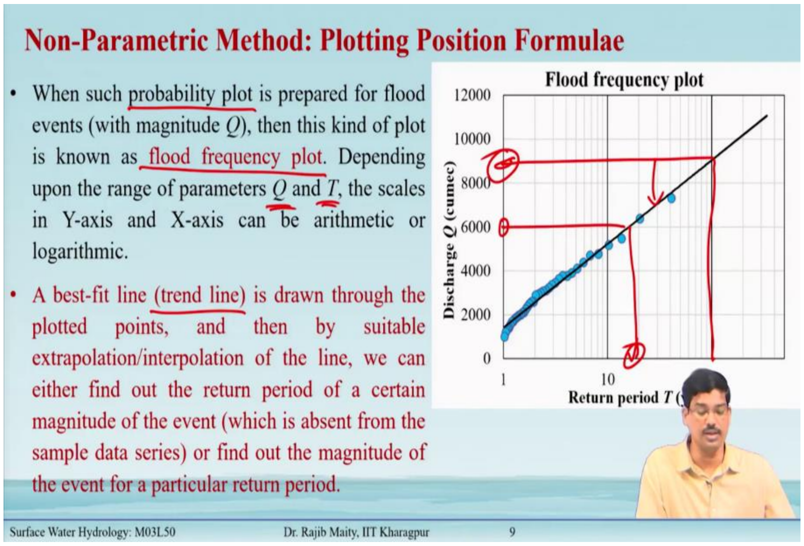
Non-Parametric Method: Plotting Position Formulae

- For the application of these plotting position formulae, the first task is to arrange the sample data (of size = N) in descending order of magnitude and to assign an order number or rank (m).

- Thus, for the first member of the arranged data, i.e., for the largest data, m = 1 will be assigned, for the second-largest data m = 2, and so on. Thus, for the smallest data in the sample, $m = N$ will be assigned. Then, using any of the empirical plotting position formulae, the probability of exceedance (p) can be calculated for all data in the series.

- Here, it can be noted that the Weibull formula is the most popular among the others. After determining p (hence, T, which is equal to $1/p$), we can obtain the probability plot for the given data by plotting its different magnitudes with a corresponding probability of exceedance.

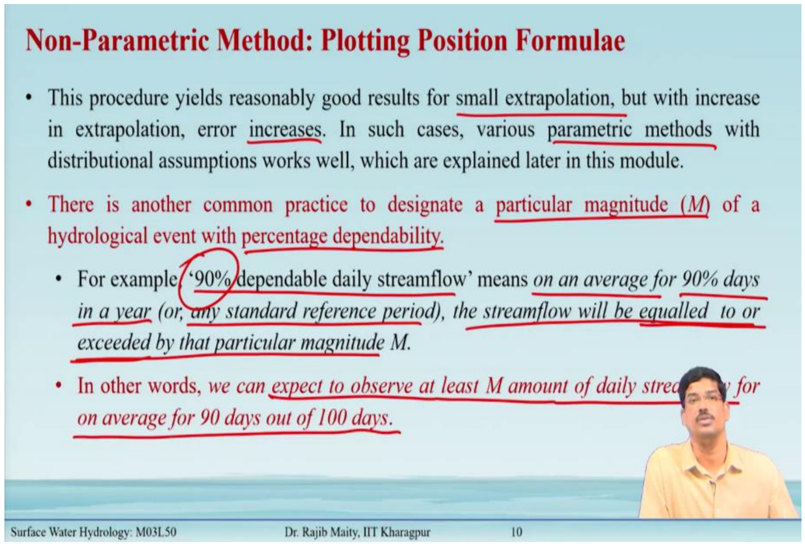
(Refer Slide Time: 13:08)



- When such a probability plot is prepared for flood events (with magnitude *Q*), then this kind of plot is known as a flood frequency plot. Depending upon the range of parameters *Q* and *T*, the scales on the Y-axis and X-axis can be arithmetic or logarithmic.

- A best-fit line (trend line) is drawn through the plotted points, and then by suitable extrapolation/interpolation of the line, we can either find out the return period of a certain magnitude of the event (which is absent from the sample data series) or find out the magnitude of the event for a particular return period.

(Refer Slide Time: 14:33)



- This procedure yields reasonably good results for small extrapolation, but with an increase in extrapolation, error increases. In such cases, various parametric methods with distributional assumptions work well, which are explained later in this module.

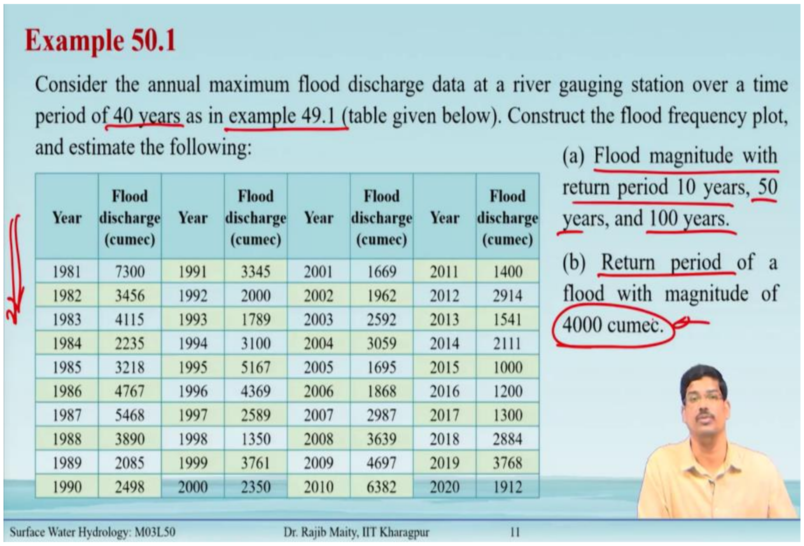
- There is another common practice to designate a particular magnitude (M) of a hydrological event with percentage dependability.

    - For example, '90% dependable daily streamflow' means on an average for 90% of days in a year (or, any standard reference period), the streamflow will be equalled to or exceeded by that particular magnitude M.

    - In other words, we can expect to observe at least M amount of daily streamflow on average for 90 days out of 100 days.

.

(Refer Slide Time: 16:06)



Example 50.1

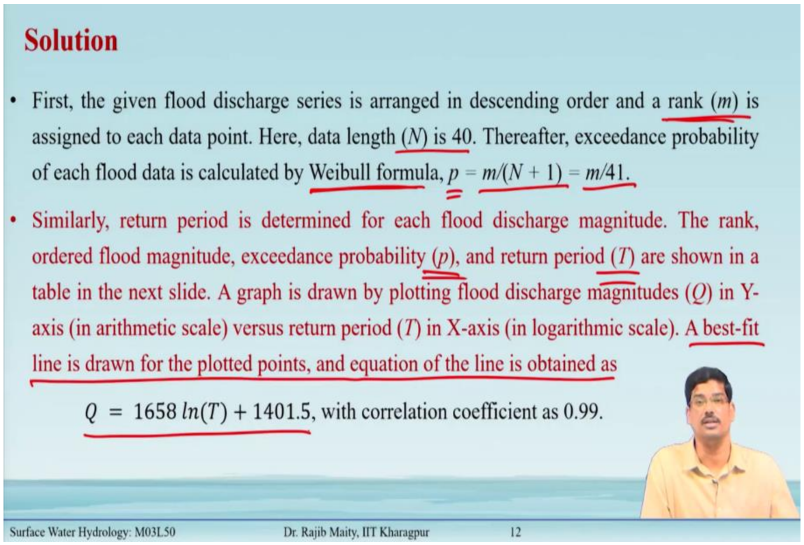
Consider the annual maximum flood discharge data at a river gauging station over a time period of 40 years as in example 49.1 (table given below). Construct the flood frequency plot, and estimate the following:

Year	Flood discharge (cumec)	Year	Flood discharge (cumec)	Year	Flood discharge (cumec)	Year	Flood discharge (cumec)
1981	7300	1991	3345	2001	1669	2011	1400
1982	3456	1992	2000	2002	1962	2012	2914
1983	4115	1993	1789	2003	2592	2013	1541
1984	2235	1994	3100	2004	3059	2014	2111
1985	3218	1995	5167	2005	1695	2015	1000
1986	4767	1996	4369	2006	1868	2016	1200
1987	5468	1997	2589	2007	2987	2017	1300
1988	3890	1998	1350	2008	3639	2018	2884
1989	2085	1999	3761	2009	4697	2019	3768
1990	2498	2000	2350	2010	6382	2020	1912

(a) Flood magnitude with return periods of 10 years, 50 years, and 100 years.

(b) Return period of a flood with a magnitude of 4000 cumec.

(Refer Slide Time: 17:42)



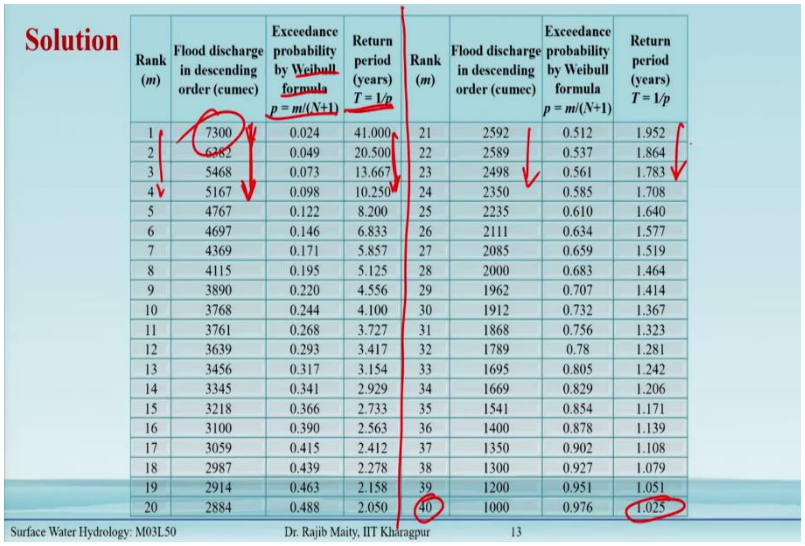
Solution

First, the given flood discharge series is arranged in descending order and a rank (m) is assigned to each data point. Here, the data length (N) is 40. Thereafter, the exceedance probability of each flood data is calculated by the Weibull formula, $p = m/(N + 1) = m/41$.

Similarly, the return period is determined for each flood discharge magnitude. The rank, ordered flood magnitude, exceedance probability (p), and return period (T) are shown in a table in the next slide. A graph is drawn by plotting flood discharge magnitudes (Q) on Y-axis (in arithmetic scale) versus return period (T) on X-axis (in logarithmic scale). A best-fit line is drawn for the plotted points, and the equation of the line is obtained as

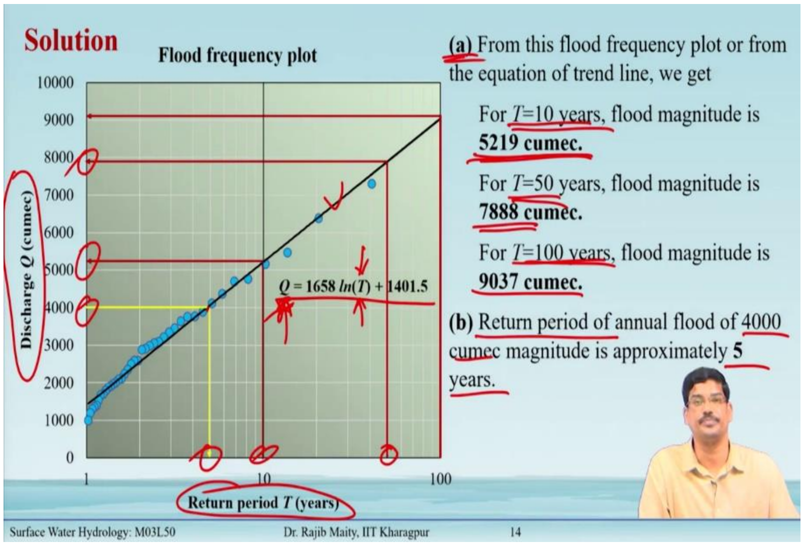
$$Q = 1658\, ln(T) + 1401.5,$$ with a correlation coefficient of 0.99.

(Refer Slide Time: 18:56)



Rank (m)	Flood discharge in descending order (cumec)	Exceedance probability by Weibull formula $p = m/(N+1)$	Return period (years) $T = 1/p$	Rank (m)	Flood discharge in descending order (cumec)	Exceedance probability by Weibull formula $p = m/(N+1)$	Return period (years) $T = 1/p$
1	7300	0.024	41.000	21	2592	0.512	1.952
2	6382	0.049	20.500	22	2589	0.537	1.864
3	5468	0.073	13.667	23	2498	0.561	1.783
4	5167	0.098	10.250	24	2350	0.585	1.708
5	4767	0.122	8.200	25	2235	0.610	1.640
6	4697	0.146	6.833	26	2111	0.634	1.577
7	4369	0.171	5.857	27	2085	0.659	1.519
8	4115	0.195	5.125	28	2000	0.683	1.464
9	3890	0.220	4.556	29	1962	0.707	1.414
10	3768	0.244	4.100	30	1912	0.732	1.367
11	3761	0.268	3.727	31	1868	0.756	1.323
12	3639	0.293	3.417	32	1789	0.78	1.281
13	3456	0.317	3.154	33	1695	0.805	1.242
14	3345	0.341	2.929	34	1669	0.829	1.206
15	3218	0.366	2.733	35	1541	0.854	1.171
16	3100	0.390	2.563	36	1400	0.878	1.139
17	3059	0.415	2.412	37	1350	0.902	1.108
18	2987	0.439	2.278	38	1300	0.927	1.079
19	2914	0.463	2.158	39	1200	0.951	1.051
20	2884	0.488	2.050	40	1000	0.976	1.025

(Refer Slide Time: 20:15)



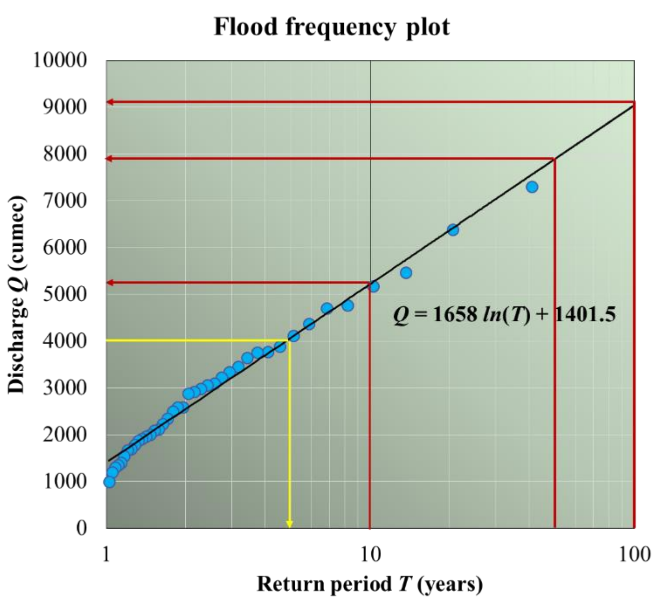


Figure 1 shows the flood frequency plot of example 50.1

(a) From this flood frequency plot or the equation of the trend line, we get

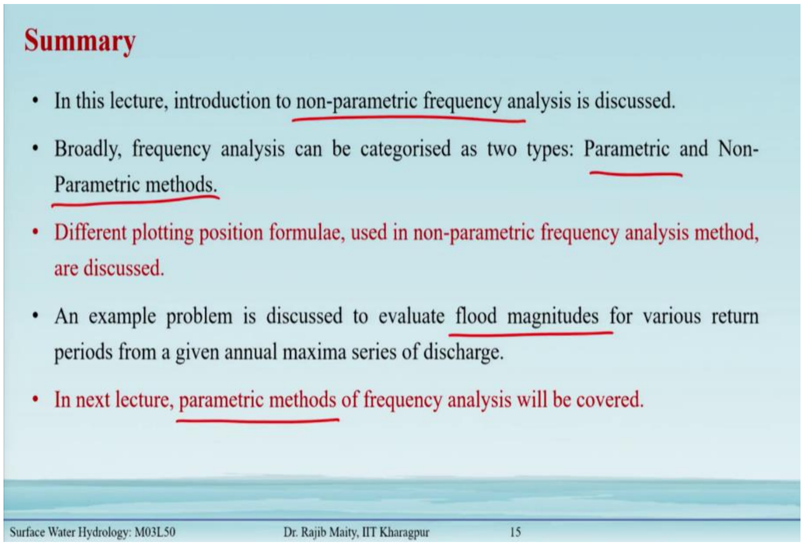
For T=10 years, flood magnitude is **5219 cumec.**

For T=50 years, flood magnitude is **7888 cumec.**

For T=100 years, flood magnitude is **9037 cumec.**

(b) Return period of an annual flood of 4000 cumec magnitude is approximately **5** years.

(Refer Slide Time: 22:08)



Summary

In summary, we learned the following points from this lecture:

- In this lecture, an introduction to non-parametric frequency analysis is discussed.
- Broadly, frequency analysis can be categorized into two types: Parametric and Non-Parametric methods.
- Different plotting position formulae, used in the non-parametric frequency analysis method, are discussed.
- An example problem is discussed to evaluate flood magnitudes for various return periods from a given annual maxima series of discharge.
- In the next lecture, parametric methods of frequency analysis will be covered.