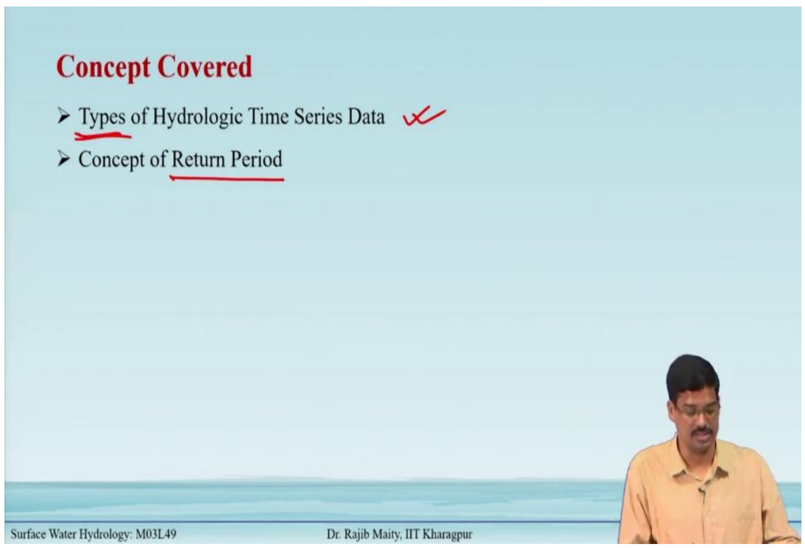
Surface Water Hydrology
Professor Rajib Maity
Department of Civil Engineering
Indian Institute of Technology, Kharagpur
Lecture 49
Types of Data Series and Concept of Return Period

In today's lecture two basic things we will try to learn. The first one is the types of data series and a very important concept is the concept of the return period.

(Refer Slide Time: 00:33)

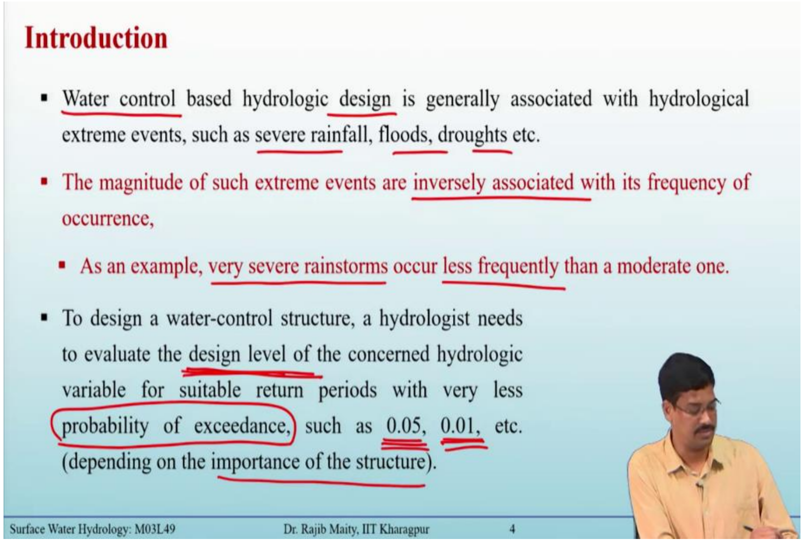


In the concept covered, there will be two things we will cover one is the types of hydrologic time series data. And another important concept that is also needed for frequency analysis is called the concept of the return period.

(Refer Slide Time: 01:09)

The outline goes like this the first step brief introduction then different types of hydrologic data series, how to pick out the extreme related values, and then we come to the concept of the return period. Under this, we will solve some example problems also, before we go to the summary of this lecture.

(Refer Slide Time: 01:32)

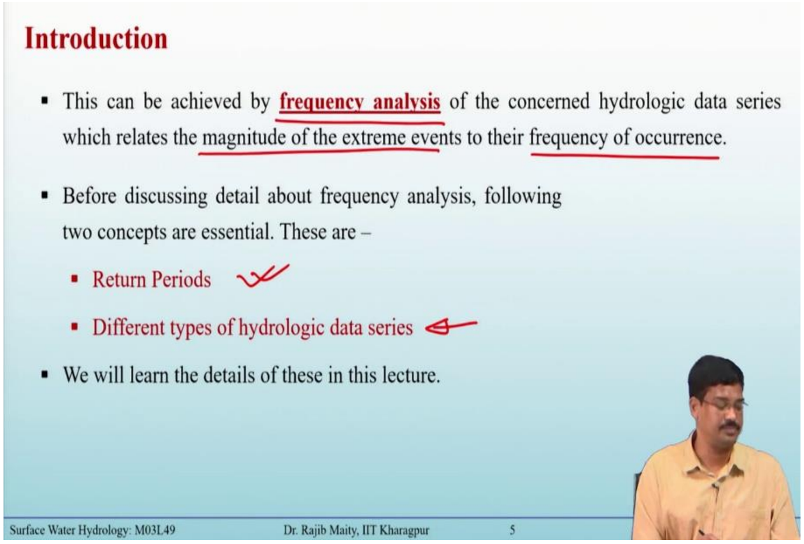


Introduction

In the case of water control-based hydrologic design, this is generally associated with the hydrological extreme events such as severe rainfall, floods, droughts, etc. Now, the magnitude of such extreme events is inversely associated with their frequency of occurrence. For example, if we take the example of the very severe rainstorms, the more and more severe the rainstorm is the less and less frequency of that particular event.

Now, to design a water control structure, a hydrologist needs to evaluate the design level of the concerned hydrologic variables. So, our target is to understand the design value that we should select before we go further with design aspects and these are generally suitable for the return periods with a very less probability of exceedance. The design value or the design level that we consider to be exceeded with a very less amount of probability in the natural events, we cannot always ensure that that can never be exceeded.

(Refer Slide Time: 03:53)



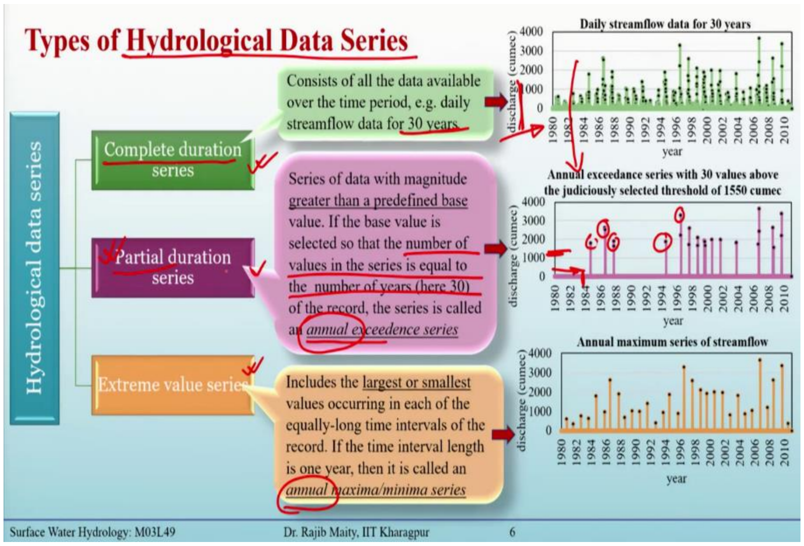
To get that design value, frequency analysis is essential. So, this one will give the magnitude of the extreme event that is required for this design purpose, it should have a relation to the frequency of the occurrence. So, this relationship we need to understand through this frequency analysis.

Before discussing detailed frequency analysis, the following two concepts are essential. These are

➢ Return Periods

➢ Different types of hydrologic data series

(Refer Slide Time: 04:55)



If we take the types of hydrological data series it can be broadly divided into three general groups.

1. **Complete duration series**

Consists of all the data available over the time period, e.g. daily streamflow data for 30 years

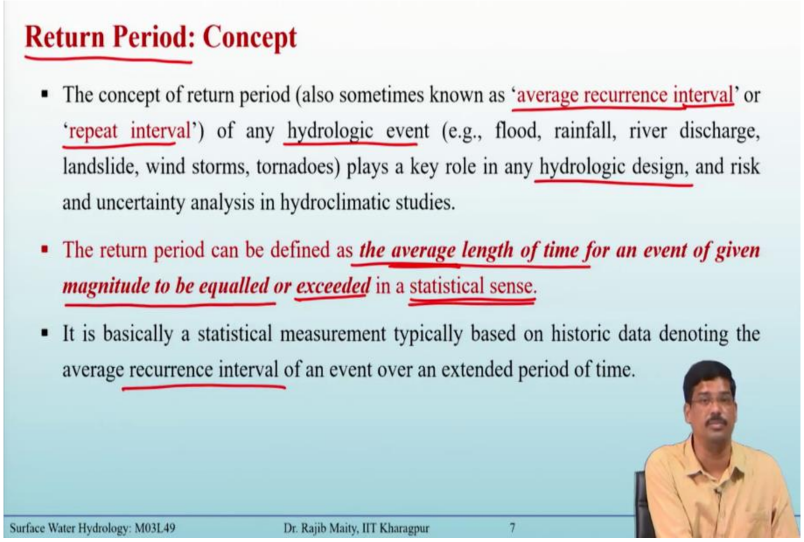
2. **Partial duration series**

Series of data with magnitude greater than a predefined base value. If the base value is selected so that the number of values in the series is equal to the number of years (here 30) of the record, the series is called an annual exceedance series

3. **Extreme value series**

Includes the largest or smallest values occurring in each of the equally-long time intervals of the record. If the time interval length is one year, then it is called an annual maxima/minima series

(Refer Slide Time: 09:32)



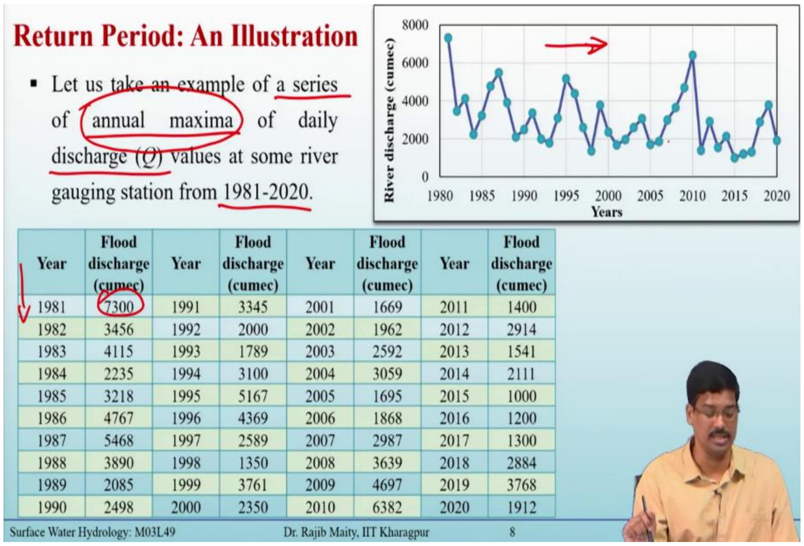
Return Period: Concept

The concept of the return period (also sometimes known as 'average recurrence interval' or 'repeat interval') of any hydrologic event (e.g., flood, rainfall, river discharge, landslide, wind storms, tornadoes) plays a key role in any hydrologic design, and risk and uncertainty analysis in hydroclimatic studies.

The return period can be defined as the average length of time for an event of a given magnitude to be equaled or exceeded in a statistical sense.

It is basically a statistical measurement typically based on historic data denoting the average recurrence interval of an event over an extended period of time. So, in the definition of the return period, three things are very important to mention the first one is the average length of time, the second one is that of a given magnitude to be equaled or exceeded, and thirdly in a statistical sense. So, it is basically a statistical measurement typically based on the historical data denoting the average recurrence interval also another term is the recurrence interval of an event over an extended period of time.

(Refer Slide Time: 13:07)

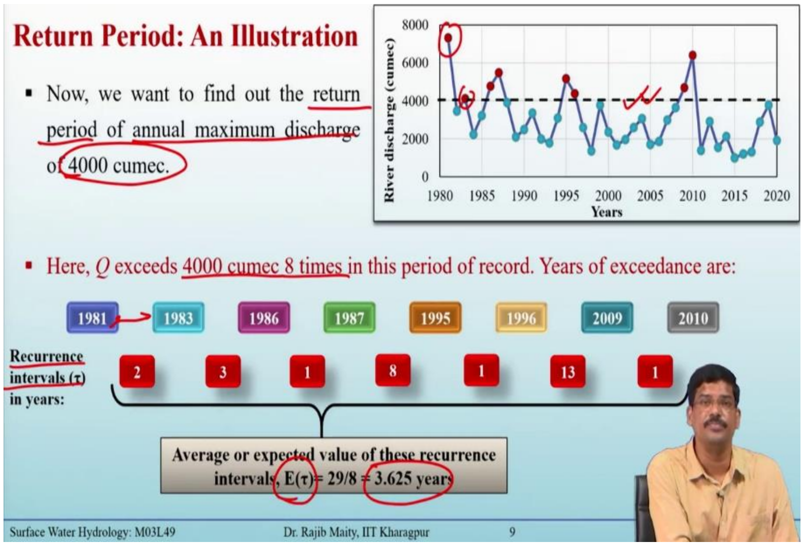


Return Period: An Illustration

Let us take an example of a series of annual maxima of daily discharge (Q) values at some river gauging stations from 1981-to 2020.

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

(Refer Slide Time: 14:15)



We want to find out the return period of the annual maximum discharge of 400 cumecs.

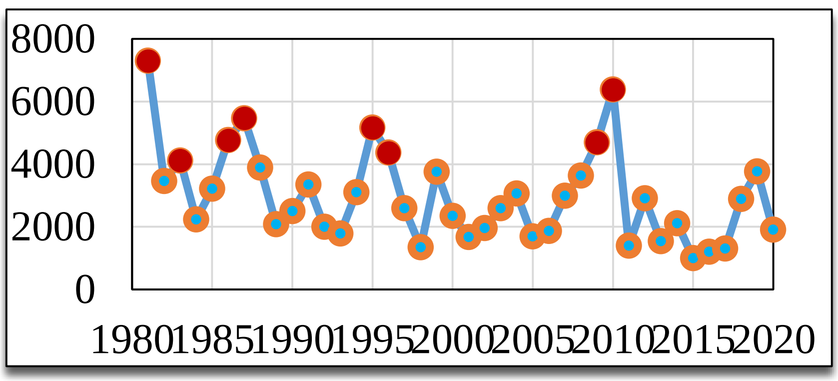


Figure 1 Time series graph of discharge data from 1980-to 2020

Here, Q exceeds 4000 cumec 8 times in this period of record. Years of exceedance are:

1981 -1983 =2

1983 -1986=3

1986 -1987=1

1987 -1995 =8

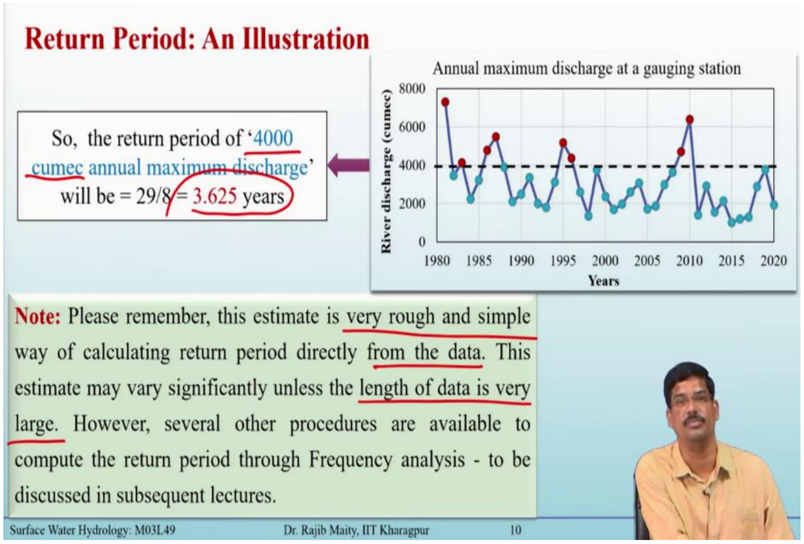
1995 -1996 =1

1996 -2009 =13

2009-2010 =1

The average or expected value of these recurrence intervals, E (τ)= 29/8 = 3.625 years

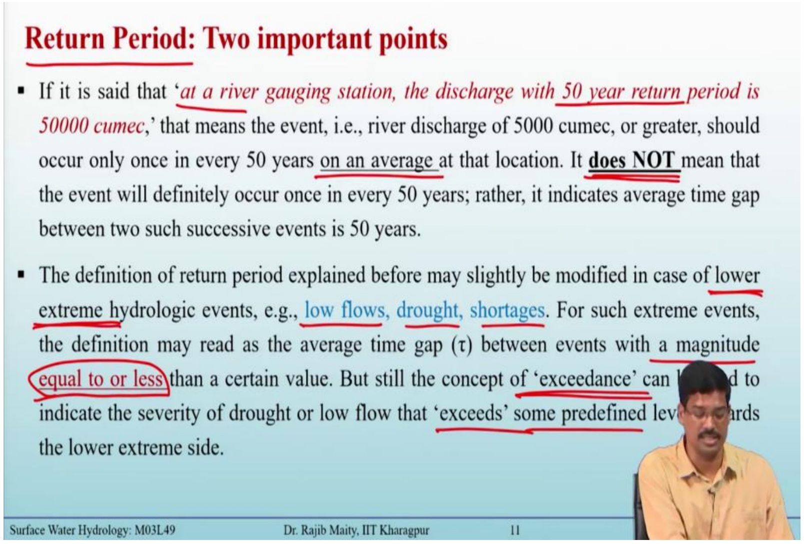
(Refer Slide Time: 16:14)



So, the return period of '4000 cumec annual maximum discharge' will be = 29/8 = 3.625 years

It may be noted that this estimate is a very rough and simple way of calculating the return period directly from the data. This estimate may vary significantly unless the length of data is very large. However, several other procedures are available to compute the return period through Frequency analysis - to be discussed in subsequent lectures.

(Refer Slide Time: 17:26)

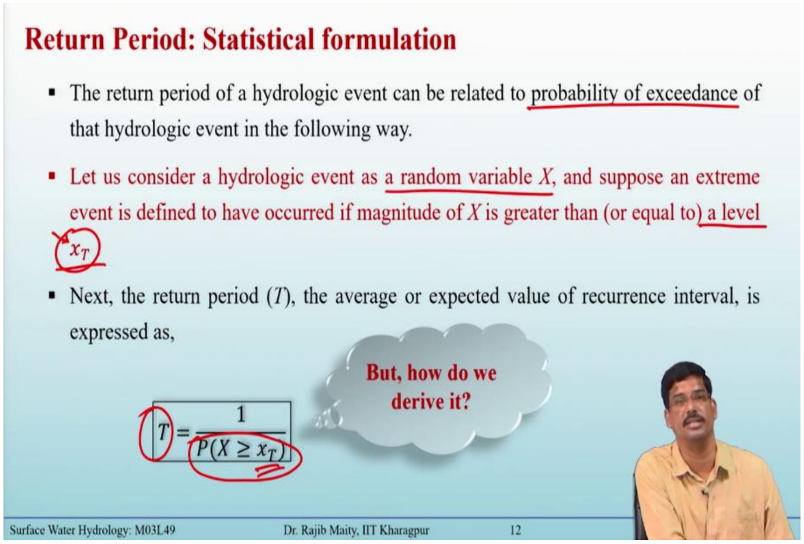


Return Period: Two important points

- If it is said that 'at a river gauging station, the discharge with 50 year return period is 50000 cumec,' that means the event, i.e., river discharge of 5000 cumec, or greater, should occur only once in every 50 years on an average at that location. It does NOT mean that the event will definitely occur once in every 50 years; rather, it indicates the average time gap between two such successive events is 50 years.

- The definition of the return period explained before may slightly be modified in case of lower extreme hydrologic events, e.g., low flows, drought, shortages. For such extreme events, the definition may read as the average time gap (τ) between events with a magnitude equal to or less than a certain value. But still, the concept of 'exceedance' can be used to indicate the severity of drought or low flow that 'exceeds' some predefined level towards the lower extreme side.

(Refer Slide Time: 19:45)



Return Period: Statistical formulation

- The return period of a hydrologic event can be related to the probability of exceedance of that hydrologic event in the following way.

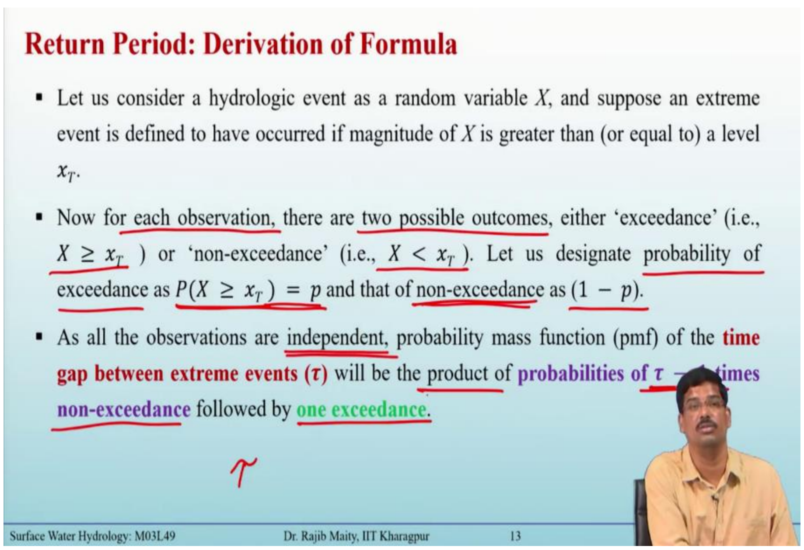
- Let us consider a hydrologic event as a random variable X, and suppose an extreme event is defined to have occurred if the magnitude of X is greater than (or equal to) a level x_T .

- Next, the return period (T), the average or expected value of the recurrence interval, is expressed as,

$$T = \frac{1}{P(X \geq x_T)}$$

So, x_T is the threshold here. So, this is the probability and the inverse of this one is showing the return period. So, the thing is that the next question immediate question may come how we establish this relationship that is the inverse of the probability of exceedance is the return period.

(Refer Slide Time: 21:23)



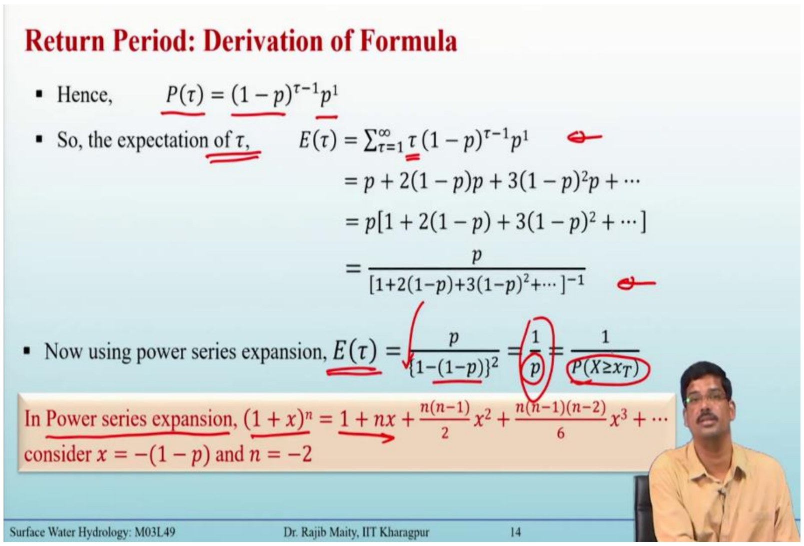
Return Period: Derivation of Formula

- Let us consider a hydrologic event as a random variable X, and suppose an extreme event is defined to have occurred if the magnitude of X is greater than (or equal to) a level x_T.

- Now for each observation, there are two possible outcomes, either 'exceedance' (i.e., $\geq x_T$) or 'non-exceedance' (i.e., $X < x_T$). Let us designate the probability of exceedance as $P(X \geq x_T) = p$ and that of non-exceedance as $(1 - p)$.

- As all the observations are independent, the probability mass function (pmf) of the time gap between extreme events ($\boldsymbol{\tau}$) will be the product of probabilities of $\boldsymbol{\tau - 1}$ times non-exceedance followed by one exceedance.

(Refer Slide Time: 23:30)



- Hence, $P(\tau) = (1-p)^{\tau-1}p^1$

- So, the expectation of τ, $E(\tau) = \sum_{\tau=1}^{\infty} \tau \,(1-p)^{\tau-1}p^2$

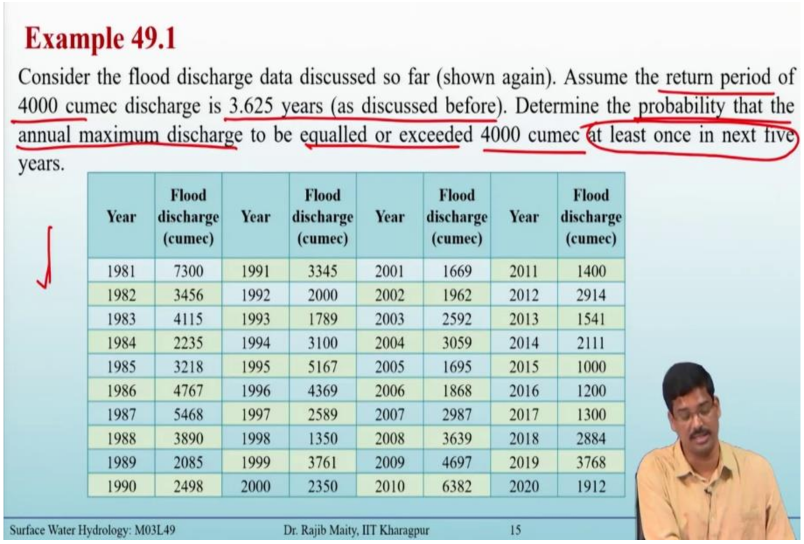
$$= p + 2(1-p)p + 3(1-p)^2p + \cdots$$
$$= p[1 + 2(1-p) + 3(1-p)^2 + \cdots]$$
$$= \frac{p}{[1 + 2(1-p) + 3(1-p)^2 + \cdots]^{-1}}$$

- Now using power series expansion, $E(\tau) = \frac{p}{\{1-(1-p)\}^2} = \frac{1}{p} = \frac{1}{P(X \geq x_T)}$

- In the Power series expansion, $(1+x)^n = 1 + nx + \frac{n(n-1)}{2}x^2 + \frac{n(n-1)(n-2)}{6}x^3 + \cdots$

- consider $x = -(1-p)$ and $n = -2$

(Refer Slide Time: 25:01)

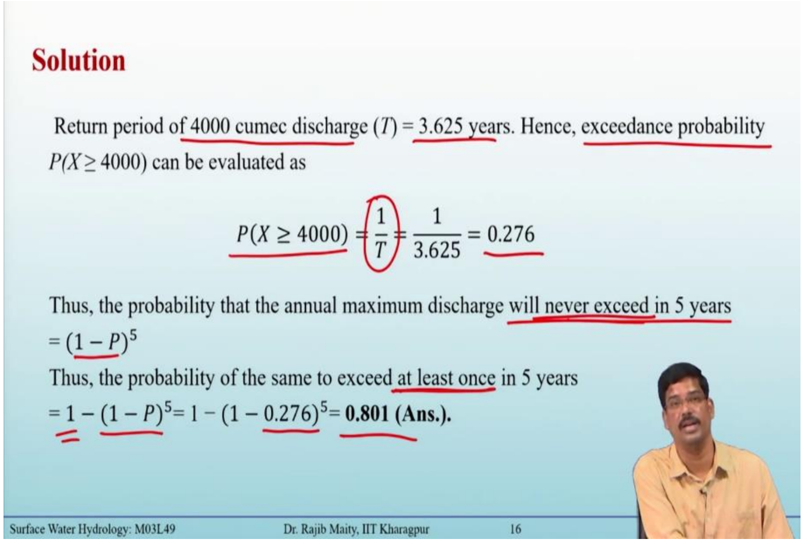


Example 49.1

Consider the flood discharge data discussed so far (shown again). Assume the return period of 4000 cumec discharge is 3.625 years (as discussed before). Determine the probability that the annual maximum discharge to be equalled or exceeded 4000 cumec at least once in the next five years.

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

(Refer Slide Time: 26:05)



Solution

Return period of 4000 cumec discharge (*T*) = 3.625 years. Hence, exceedance probability $P(X \geq 4000)$ can be evaluated as

$$P(X \geq 4000) = \frac{1}{T} = \frac{1}{3.625} = 0.276$$

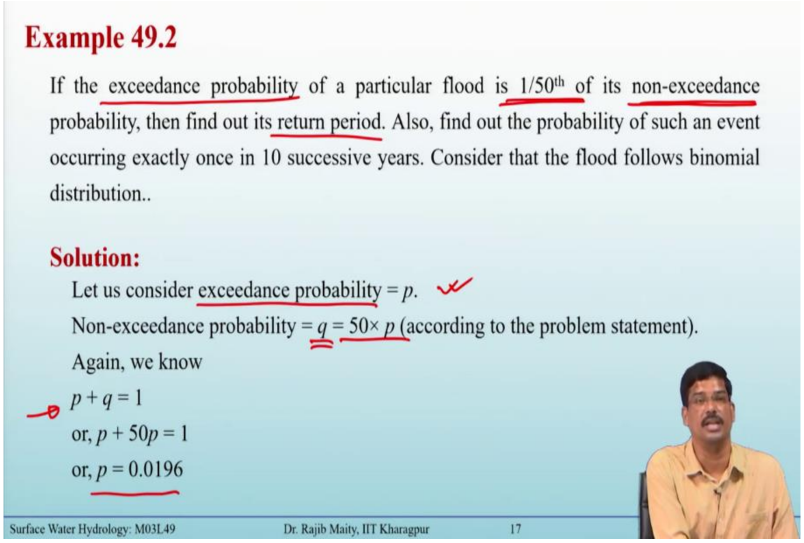
Thus, the probability that the annual maximum discharge will never exceed 5 years

$$= (1 - P)^5$$

Thus, the probability of the same exceeding at least once in 5 years

$= 1 - (1 - P)^5 = 1 - (1 - 0.276)^5 =$ **0.801 (Ans.).**

(Refer Slide Time: 27:22)



Example 49.2

If the exceedance probability of a particular flood is 1/50th of its non-exceedance probability, then find out its return period. Also, find out the probability of such an event occurring exactly once in 10 successive years. Consider that the flood follows a binomial distribution.

Solution:

Let us consider exceedance probability = p.

Non-exceedance probability = $q = 50\times p$ (according to the problem statement).

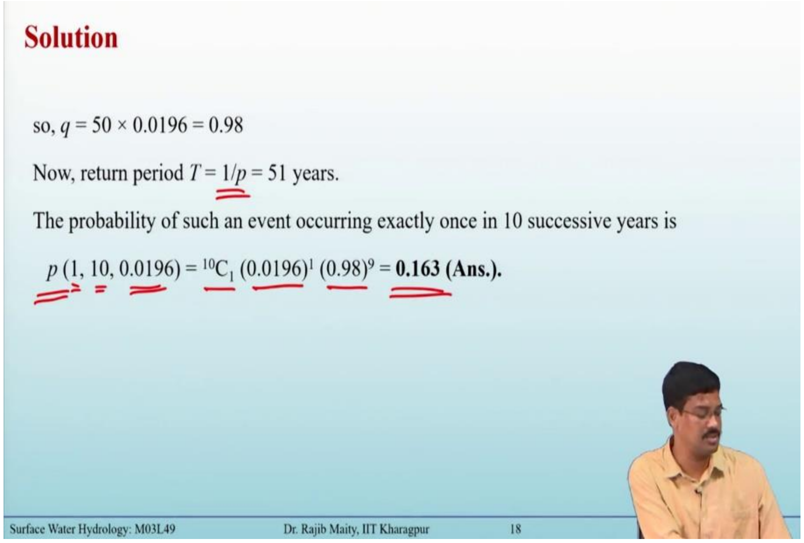
Again, we know

$p + q = 1$

Or, $p + 50p = 1$

Or, $p = 0.0196$

(Refer Slide Time: 28:34)



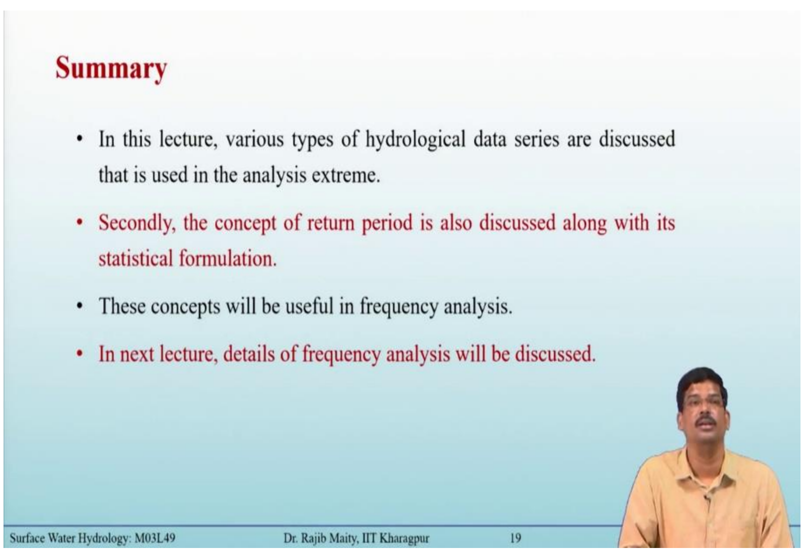
So, $q = 50 \times 0.0196 = 0.98$

Now, return period $T = 1/p = 51$ years.

The probability of such an event occurring exactly once in 10 successive years is

$p\,(1, 10, 0.0196) = {}^{10}C_1\,(0.0196)^1\,(0.98)^9 =$ **0.163 (Ans.).**

(Refer Slide Time: 29:28)



Summary

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

- In this lecture, various types of hydrological data series are discussed that are used in the analysis extreme.
- Secondly, the concept of the return period is also discussed along with its statistical formulation.
- These concepts will be useful in frequency analysis.
- In the next lecture, details of frequency analysis will be discussed.