Surface Water Hydrology

Professor Rajib Maity

Department of Civil Engineering

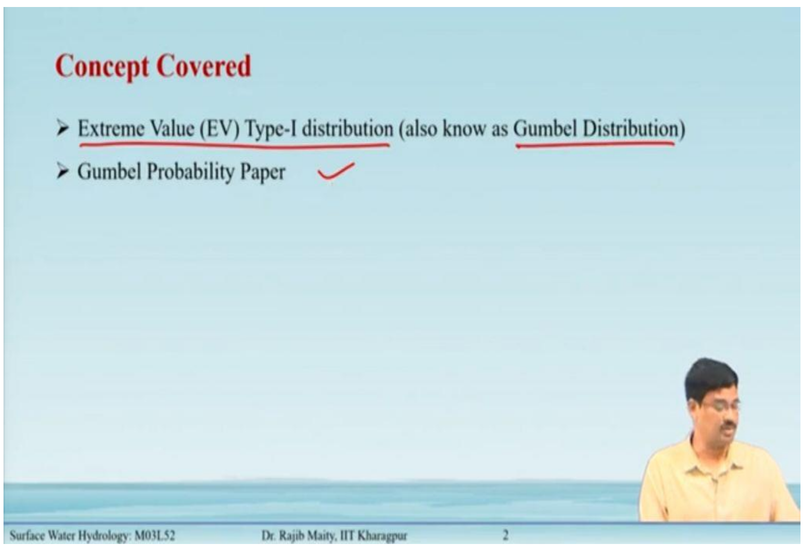
Indian Institute of Technology, Kharagpur

Lecture – 52

Frequency Analysis with Extreme Value Type-I Distribution

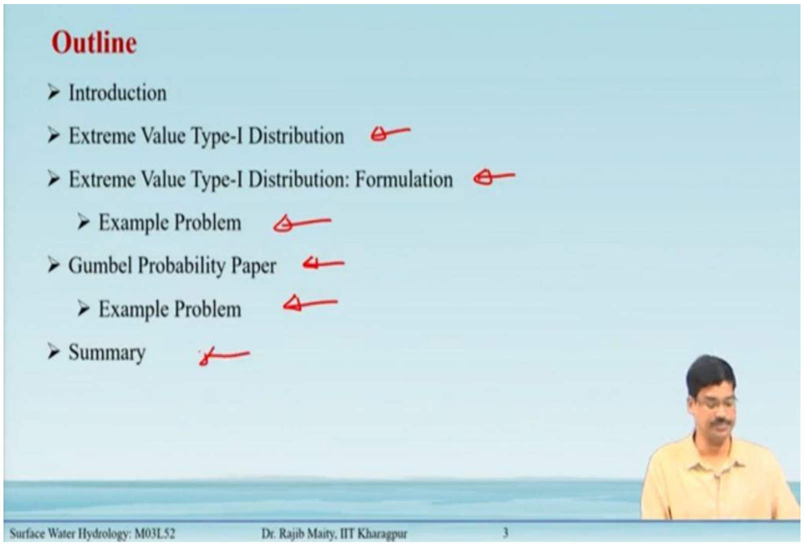
In this specific lecture, we are talking about the Frequency Analysis with Extreme Value Type-I distribution.

(Refer Slide Time: 00:56)



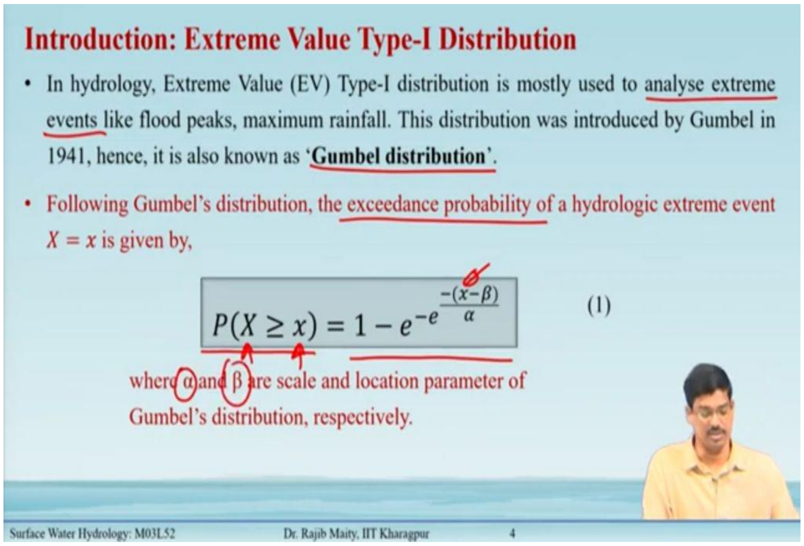
In today's lecture, we are using the Extreme Value Type-I distribution, along with that there will be a discussion on this Gumbel probability paper also.

(Refer Slide Time: 01:12)



The outline goes like this. First, we will give some introduction, then we will discuss this Extreme Value Type-I distribution. And then we will see how the different formulations could be there for the frequency analysis with some example problems. Then we proceed to the Gumbel probability paper, with that one also we will take one example problem and finally go to the summary.

(Refer Slide Time: 01:45)



Introduction: Extreme Value Type-I Distribution

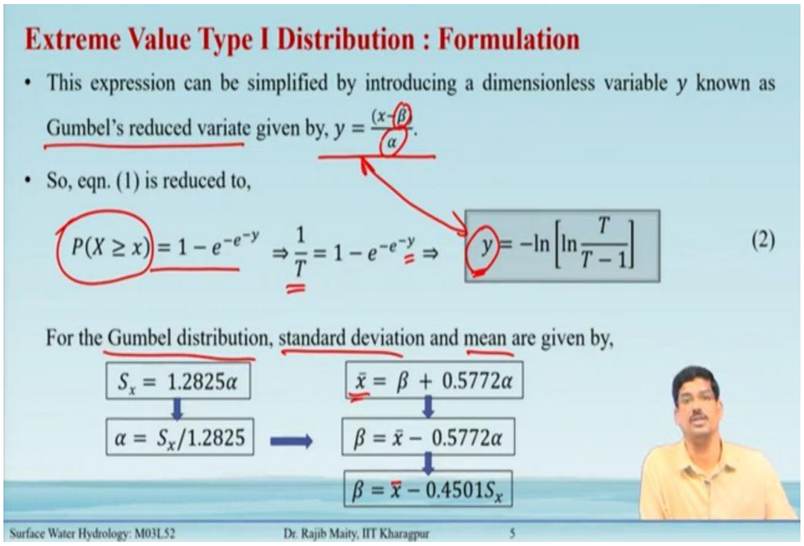
In hydrology, Extreme Value (EV) Type-I distribution is mostly used to analyze extreme events like flood peaks, and maximum rainfall. This distribution was introduced by Gumbel in 1941, hence, it is also known as 'Gumbel distribution'.

Following Gumbel's distribution, the exceedance probability of a hydrologic extreme event X=x is given by,

$$P(X \geq x) = 1 - e^{-e^{\frac{-(x-\beta)}{\alpha}}} \qquad (1)$$

Where α and β are the scale and location parameters of Gumbel's distribution, respectively.

(Refer Slide Time: 03:17)



Extreme Value Type I Distribution: Formulation

- This expression can be simplified by introducing a dimensionless variable y known as Gumbel's reduced variate given by, $y = \frac{(x-\beta)}{\alpha}$.
- So, eqn. (1) is reduced to,

$$P(X \geq x) = 1 - e^{-e^{-y}} \Rightarrow \frac{1}{T} = 1 - e^{-e^{-y}} \Rightarrow y = -\ln\left[\ln\frac{T}{T-1}\right] \qquad (2)$$

For the Gumbel distribution, standard deviation and mean are given by,

$$S_x = 1.2825\alpha$$

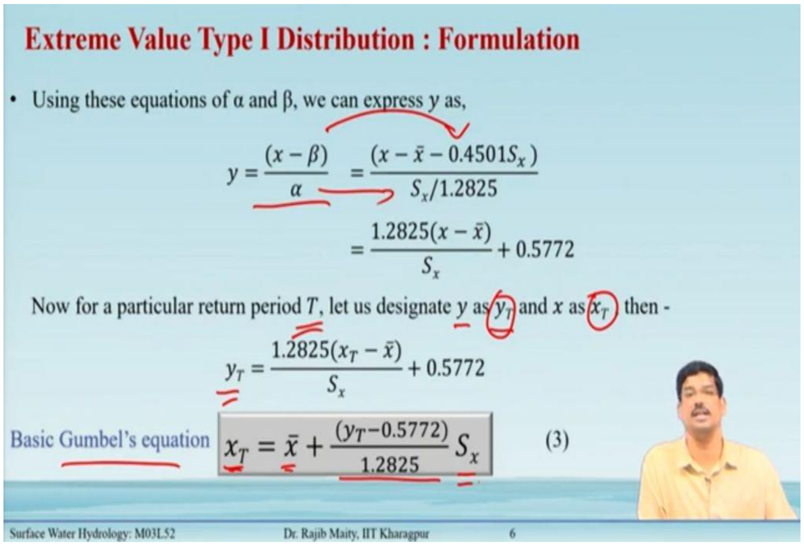
$$\alpha = S_x/1.2825$$

$$\bar{x} = \beta + 0.5772\alpha$$

$$\beta = \bar{x} - 0.5772\alpha$$

$$\beta = \bar{x} - 0.4501S_x$$

(Refer Slide Time: 05:45)



Using these equations of α and β, we can express y as,

$$y = \frac{(x-\beta)}{\alpha} = \frac{(x - \bar{x} - 0.4501S_x)}{S_x/1.2825}$$

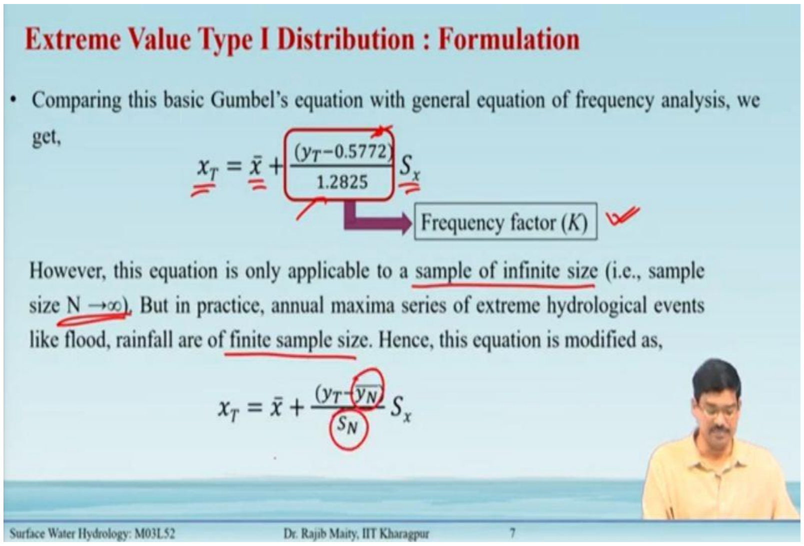
$$= \frac{1.2825(x-\bar{x})}{S_x} + 0.5772$$

Now for a particular return period T, let us designate y as y_T and x as x_T , then -

$$y_T = \frac{1.2825(x_T - \bar{x})}{S_x} + 0.5772$$

Basic Gumbel's equation $x_T = \bar{x} + \frac{(y_T - 0.5772)}{1.2825} S_x$ (3)

(Refer Slide Time: 07:06)



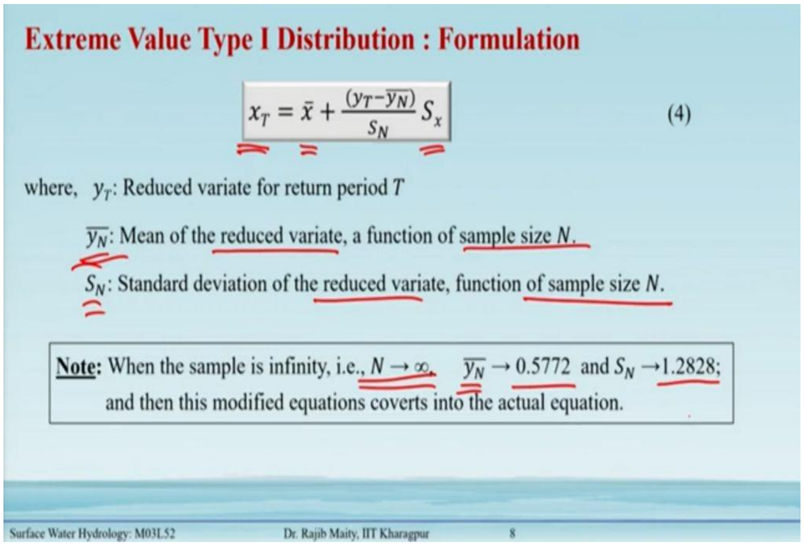
Comparing this basic Gumbel's equation with the general equation of frequency analysis, we get,

$$x_T = \bar{x} + \frac{(y_T - 0.5772)}{1.2825} S_x$$

However, this equation is only applicable to a sample of infinite size (i.e., sample size N →∞). But in practice, annual maxima series of extreme hydrological events like flood, and rainfall are of finite sample size. Hence, this equation is modified as,

$$x_T = \bar{x} + \frac{(y_T - \overline{y_N})}{s_N} S_x$$

(Refer Slide Time: 08:04)



$$x_T = \bar{x} + \frac{(y_T - \overline{y_N})}{S_N} S_x \quad (4)$$

where y_T: Reduced variate for return period T

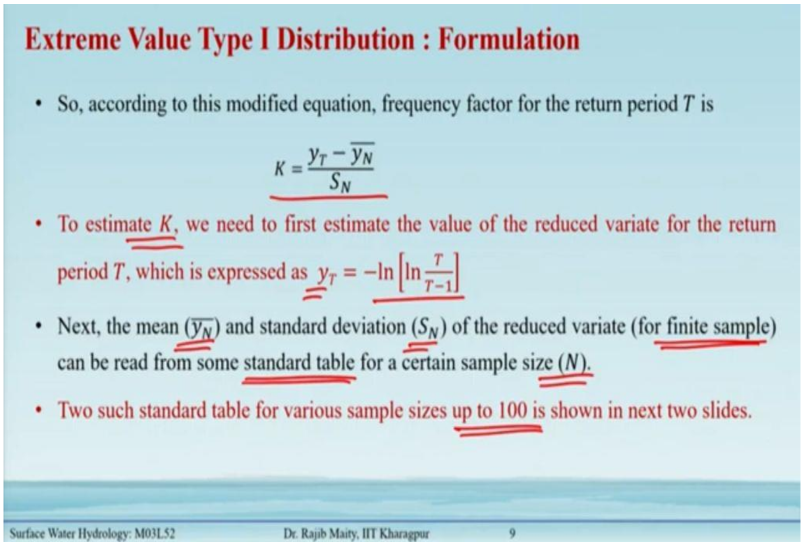
$\overline{y_N}$: Mean of the reduced variate, a function of sample size N.

S_N: Standard deviation of the reduced variate, the function of sample size N.

It may be noted that when the sample is infinity, i.e., $N \to \infty$, $\overline{y_N} \to 0.5772$ and $S_N \to 1.2828$;

And then these modified equations coverts the actual equation.

(Refer Slide Time: 08:55)



- So, according to this modified equation, the frequency factor for the return period T is

$$K = \frac{y_T - \overline{y_N}}{S_N}$$

- To estimate K, we need to first estimate the value of the reduced variate for the return period T, which is expressed as $y_T = -\ln\left[\ln\frac{T}{T-1}\right]$
- Next, the mean ($\overline{y_N}$) and standard deviation (S_N) of the reduced variate (for finite sample) can be read from some standard table for a certain sample size (N).
- Two such standard tables for various sample sizes up to 100 are shown in the next two slides.

(Refer Slide Time: 10:09)

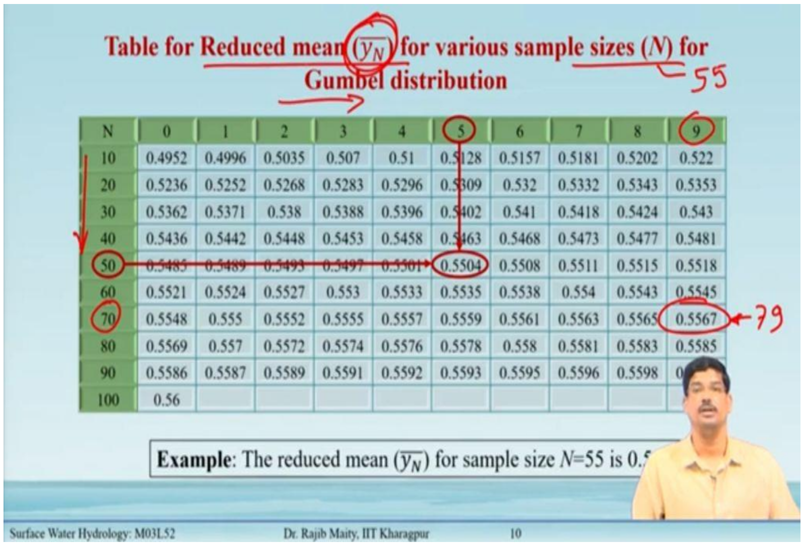


Table for Reduced mean ($\overline{y_N}$) for various sample sizes (N) for Gumbel distribution

N	0	1	2	3	4	5	6	7	8	9
10	0.4952	0.4996	0.5035	0.507	0.51	0.5128	0.5157	0.5181	0.5202	0.522
20	0.5236	0.5252	0.5268	0.5283	0.5296	0.5309	0.532	0.5332	0.5343	0.5353
30	0.5362	0.5371	0.538	0.5388	0.5396	0.5402	0.541	0.5418	0.5424	0.543
40	0.5436	0.5442	0.5448	0.5453	0.5458	0.5463	0.5468	0.5473	0.5477	0.5481
50	0.5485	0.5489	0.5493	0.5497	0.5501	0.5504	0.5508	0.5511	0.5515	0.5518
60	0.5521	0.5524	0.5527	0.553	0.5533	0.5535	0.5538	0.554	0.5543	0.5545
70	0.5548	0.555	0.5552	0.5555	0.5557	0.5559	0.5561	0.5563	0.5565	0.5567
80	0.5569	0.557	0.5572	0.5574	0.5576	0.5578	0.558	0.5581	0.5583	0.5585
90	0.5586	0.5587	0.5589	0.5591	0.5592	0.5593	0.5595	0.5596	0.5598	0.5599
100	0.56									

Example: The reduced mean ($\overline{y_N}$) for sample size N=55 is 0.5504

(Refer Slide Time: 11:08)

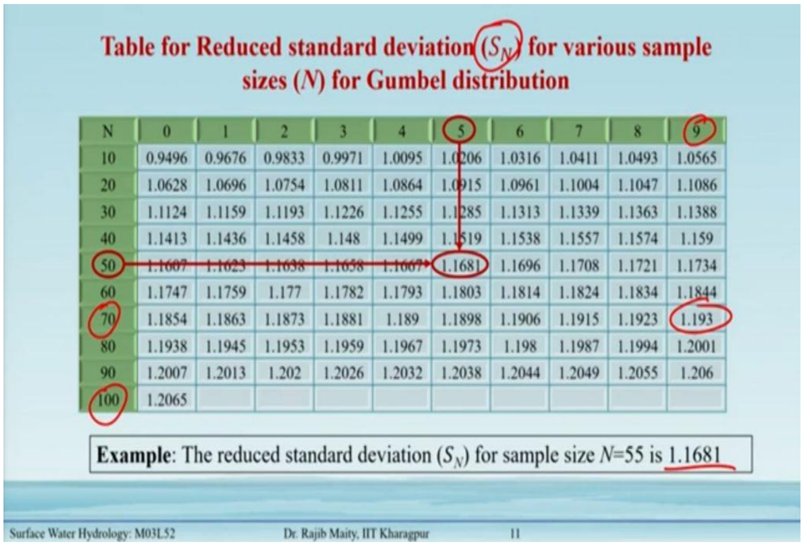
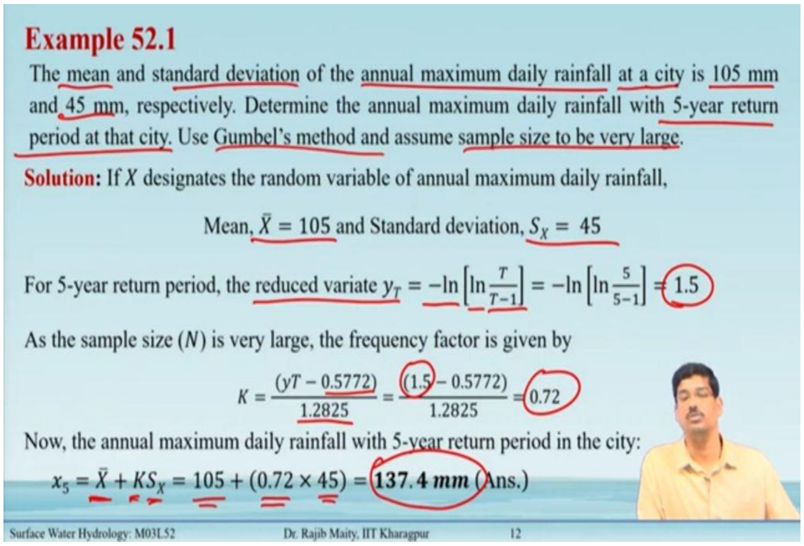


Table for Reduced standard deviation (S_N) for various sample sizes (N) for Gumbel distribution

N	0	1	2	3	4	5	6	7	8	9
10	0.9496	0.9676	0.9833	0.9971	1.0095	1.0206	1.0316	1.0411	1.0493	1.0565
20	1.0628	1.0696	1.0754	1.0811	1.0864	1.0915	1.0961	1.1004	1.1047	1.1086
30	1.1124	1.1159	1.1193	1.1226	1.1255	1.1285	1.1313	1.1339	1.1363	1.1388
40	1.1413	1.1436	1.1458	1.148	1.1499	1.1519	1.1538	1.1557	1.1574	1.159
50	1.1607	1.1623	1.1638	1.1658	1.1667	1.1681	1.1696	1.1708	1.1721	1.1734
60	1.1747	1.1759	1.177	1.1782	1.1793	1.1803	1.1814	1.1824	1.1834	1.1844
70	1.1854	1.1863	1.1873	1.1881	1.189	1.1898	1.1906	1.1915	1.1923	1.193
80	1.1938	1.1945	1.1953	1.1959	1.1967	1.1973	1.198	1.1987	1.1994	1.2001
90	1.2007	1.2013	1.202	1.2026	1.2032	1.2038	1.2044	1.2049	1.2055	1.206
100	1.2065									

Example: The reduced standard deviation (S_N) for sample size N=55 is 1.1681

(Refer Slide Time: 11:49)



Example 52.1

The mean and standard deviation of the annual maximum daily rainfall in a city is 105 mm and 45 mm, respectively. Determine the annual maximum daily rainfall with a 5-year return period in that city. Use Gumbel's method and assume the sample size to be very large.

Solution: If X designates the random variable of annual maximum daily rainfall,

$$\text{Mean, } \bar{X} = 105 \text{ and Standard deviation, } S_X = 45$$

For a 5-year return period, the reduced variate $y_T = -\ln\left[\ln\frac{T}{T-1}\right] = -\ln\left[\ln\frac{5}{5-1}\right] = 1.5$

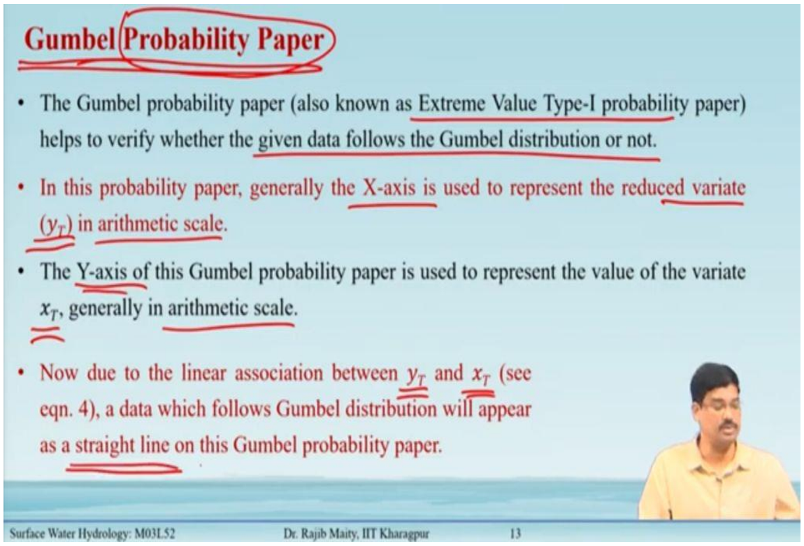
As the sample size (N) is very large, the frequency factor is given by

$$K = \frac{(yT - 0.5772)}{1.2825} = \frac{(1.5 - 0.5772)}{1.2825} = 0.72$$

Now, the annual maximum daily rainfall with a 5-year return period in the city:

$x_5 = \bar{X} + KS_X = 105 + (0.72 \times 45) = \mathbf{137.4\ mm}$ (Ans.)

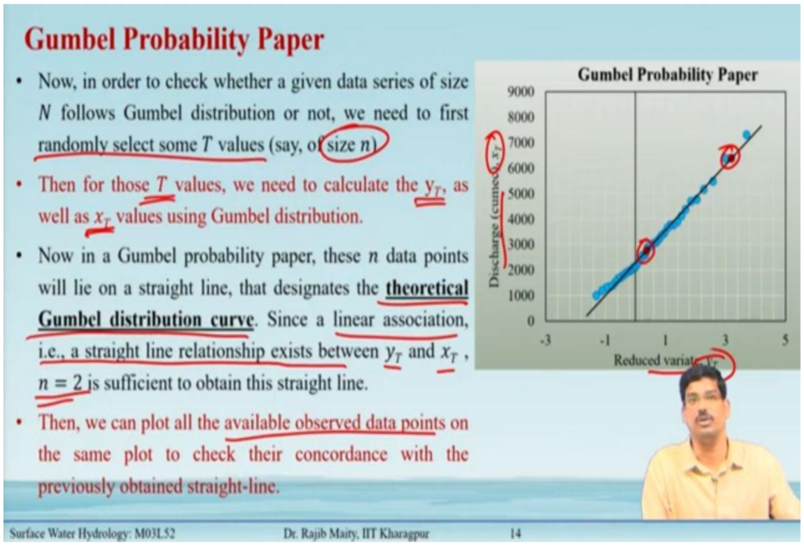
(Refer Slide Time: 13:56)



Gumbel Probability Paper

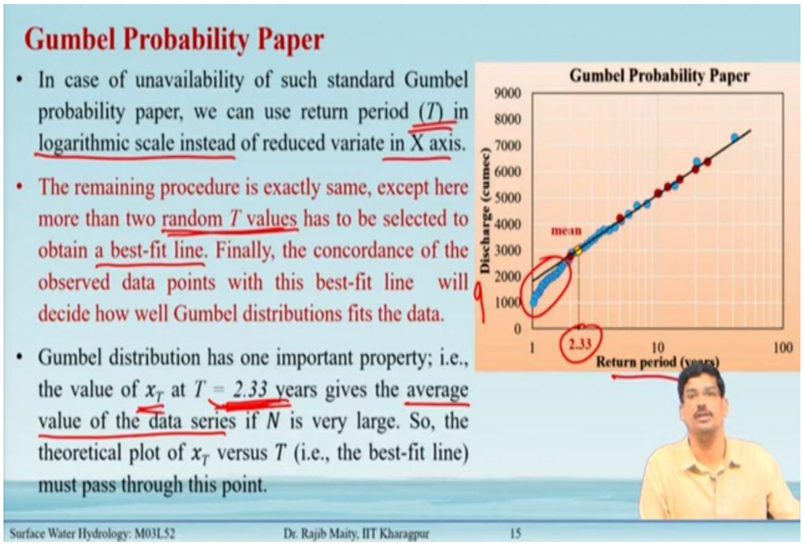
- The Gumbel probability paper (also known as Extreme Value Type-I probability paper) helps to verify whether the given data follows the Gumbel distribution or not.
- In this probability paper, generally, the X-axis is used to represent the reduced variate (y_T) in the arithmetic scale.
- The Y-axis of this Gumbel probability paper is used to represent the value of the variate x_T, generally on an arithmetic scale.
- Now due to the linear association between y_T and x_T (see Eqn. 4), data that follows the Gumbel distribution will appear as a straight line on this Gumbel probability paper.

(Refer Slide Time: 16:06)



- Now, to check whether a given data series of size N follows Gumbel distribution or not, we need to first randomly select some T values (say, of size n).
- Then for those T values, we need to calculate the y_T, as well as x_T values using Gumbel distribution.
- Now in a Gumbel probability paper, these n data points will lie on a straight line, that designates the theoretical Gumbel distribution curve. Since a linear association, i.e., a straight line relationship exists between y_T and x_T , $n = 2$ is sufficient to obtain this straight line.
- Then, we can plot all the available observed data points on the same plot to check their concordance with the previously obtained straight-line.

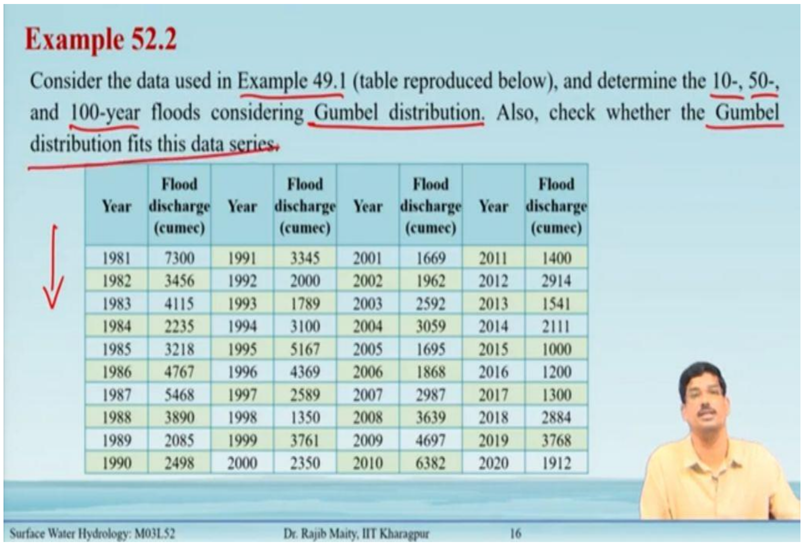
(Refer Slide Time: 17:48)



- In case of unavailability of such standard Gumbel probability paper, we can use return period (*T*) in logarithmic scale instead of reduced variate in X-axis.
- The remaining procedure is the same, except here more than two random *T* values have to be selected to obtain a best-fit line. Finally, the concordance of the observed data points with this best-fit line will decide how well Gumbel distributions fit the data.
- Gumbel distribution has one important property; i.e., the value of x_T at $T = 2.33$ years gives the average value of the data series if N is very large. So, the theoretical plot of x_T versus T (i.e., the best-fit line) must pass through this point.

.

(Refer Slide Time: 20:04)

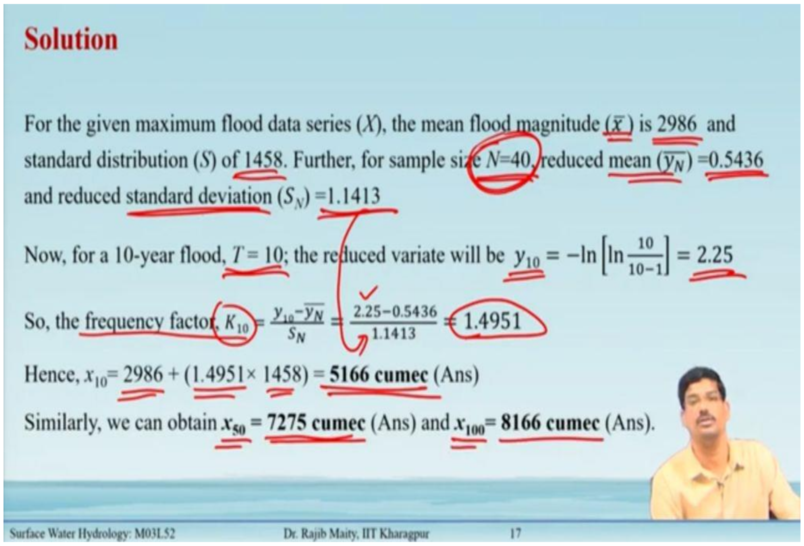


Example 52.2

Consider the data used in Example 49.1 (table reproduced below), and determine the 10-, 50-, and 100-year floods considering Gumbel distribution. Also, check whether the Gumbel distribution fits this data series.

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: 20:40)



Solution

For the given maximum flood data series (X), the mean flood magnitude ($\bar{x}$) is 2986 and standard distribution (S) of 1458. Further, for sample size N=40, the reduced mean ($\overline{y_N}$) =0.5436 and reduced standard deviation (S_N) =1.1413

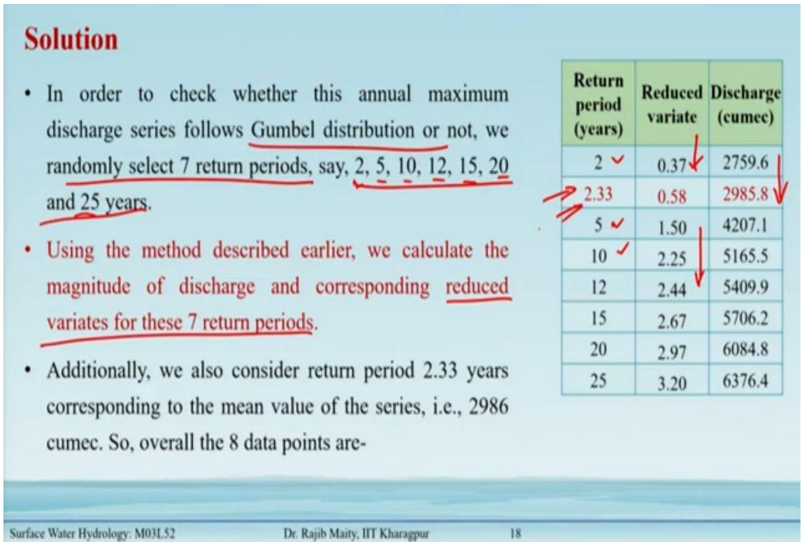
Now, for a 10-year flood, T = 10; the reduced variate will be $y_{10} = -\ln\left[\ln\frac{10}{10-1}\right] = 2.25$

So, the frequency factor, $K_{10} = \frac{y_{10}-\overline{y_N}}{S_N} = \frac{2.25-0.5436}{1.1413} = 1.4951$

Hence, x_{10}= 2986 + (1.4951× 1458) = **5166 cumec** (Ans)

Similarly, we can obtain $\boldsymbol{x_{50}}$ **= 7275 cumec** (Ans) and $\boldsymbol{x_{100}}$**= 8166 cumec** (Ans).

(Refer Slide Time: 22:48)



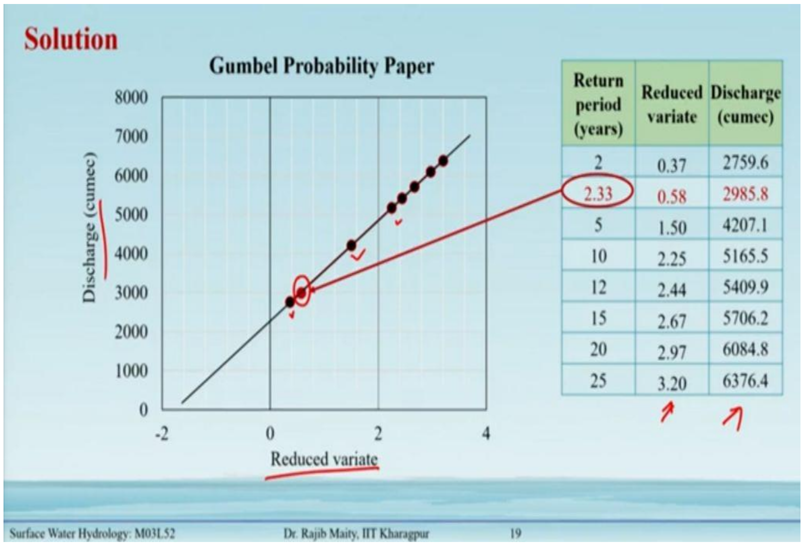
To check whether this annual maximum discharge series follows Gumbel distribution or not, we randomly select 7 return periods, say, 2, 5, 10, 12, 15, 20, and 25 years.

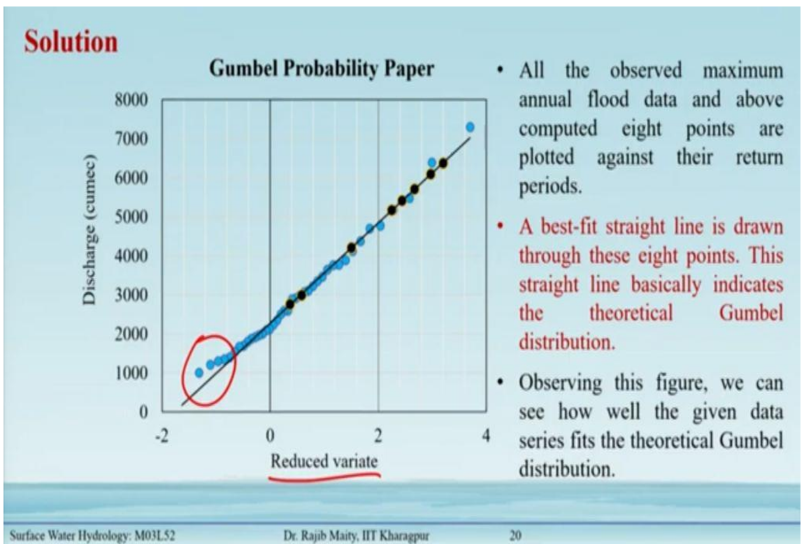
Using the method described earlier, we calculate the magnitude of discharge and corresponding reduced variates for these 7 return periods.

Additionally, we also consider the return period of 2.33 years corresponding to the mean value of the series, i.e., 2986 cumec. So, overall the 8 data points are-

Return period (years)	Reduced variate	Discharge (cumec)
2	0.37	2759.6
2.33	0.58	2985.8
5	1.50	4207.1
10	2.25	5165.5
12	2.44	5409.9
15	2.67	5706.2
20	2.97	6084.8
25	3.20	6376.4

(Refer Slide Time: 24:09)





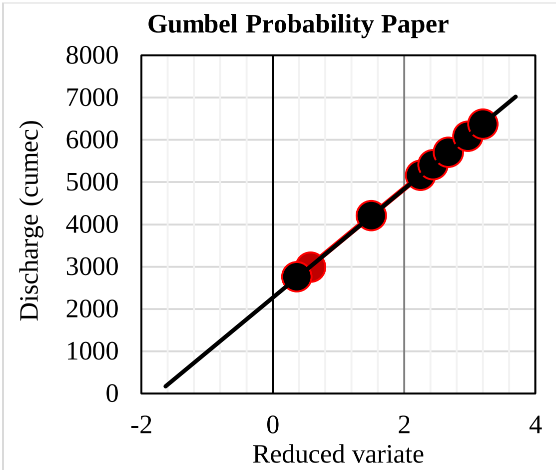


Figure 1 shows the flood frequency plot by Gumbel's Distribution of example 52.1

All the observed maximum annual flood data and above computed eight points are plotted against their return periods.

A best-fit straight line is drawn through these eight points. This straight line indicates the theoretical Gumbel distribution.

Observing this figure, we can see how well the given data series fits the theoretical Gumbel distribution.

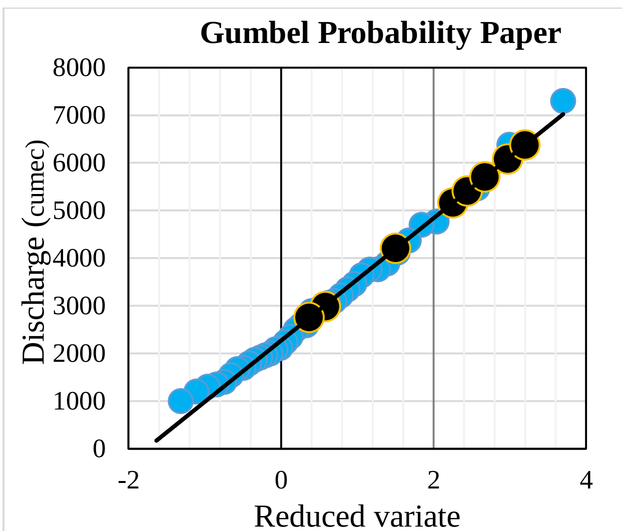
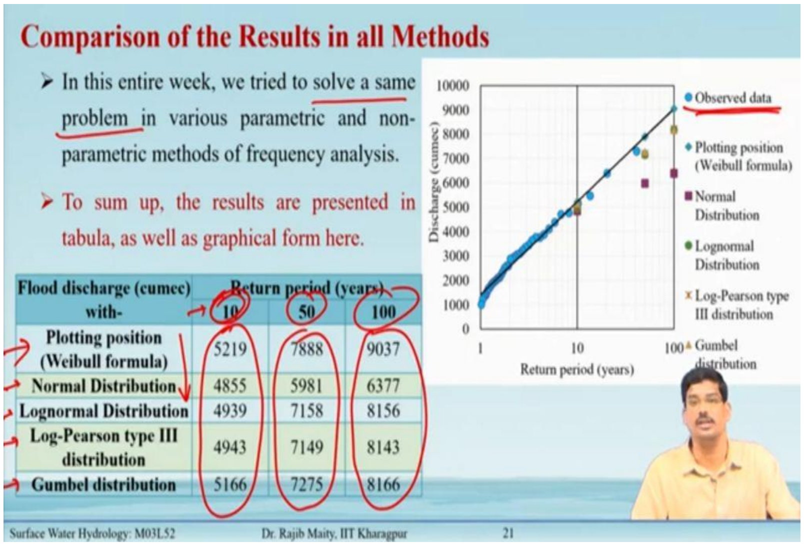


Figure 2 shows the flood frequency plot of example 52.1

(Refer Slide Time: 25:18)



Comparison of the Results in all Methods

During this entire week, we tried to solve the same problem in various parametric and non-parametric methods of frequency analysis.

To sum up, the results are presented in tabula, as well as graphical form here.

Flood discharge (cumec) with-	Return period (years)		
	10	50	100
Plotting position (Weibull formula)	5219	7888	9037
Normal Distribution	4855	5981	6377
Lognormal Distribution	4939	7158	8156
Log-Pearson type III distribution	4943	7149	8143
Gumbel distribution	5166	7275	8166

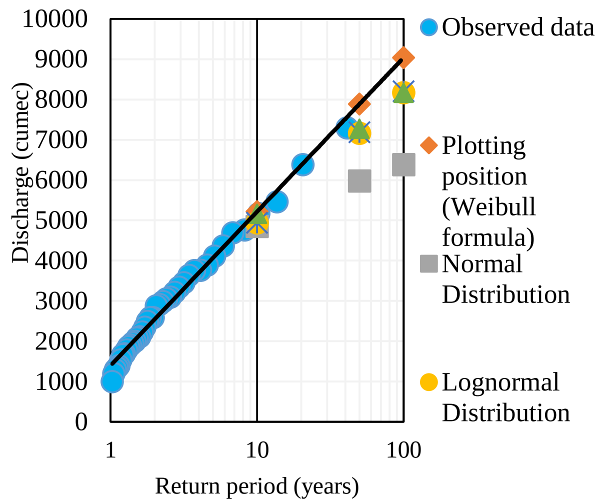
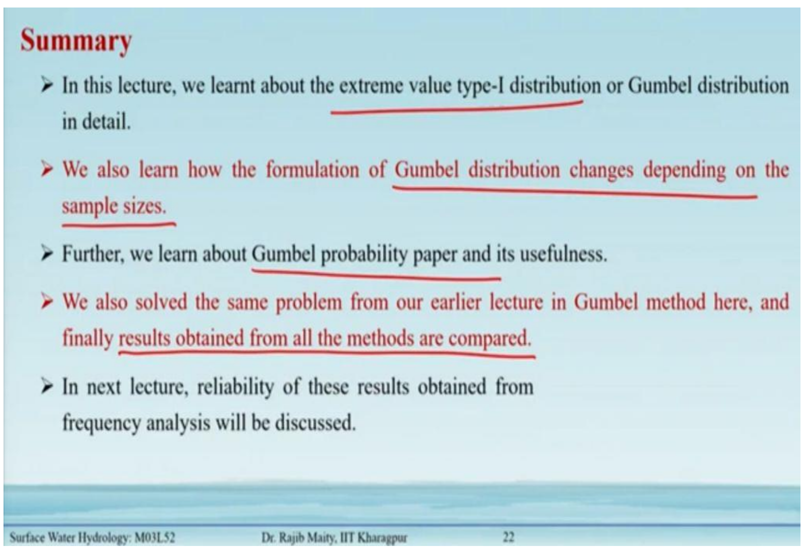


Figure 3 shows the comparison between different types of distribution

(Refer Slide Time: 27:14)



Summary

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

- In this lecture, we learned about the extreme value type-I distribution or Gumbel distribution in detail.

- Further, we also learn how the formulation of Gumbel distribution changes depending on the sample sizes.
- We learn about the Gumbel probability paper and its usefulness.
- We also solved the same problem from our earlier lecture in the Gumbel method here, and finally, the results obtained from all the methods are compared.
- In the next lecture, the reliability of these results obtained from frequency analysis will be discussed.