**Course Name: Watershed Hydrology**

**Professor Name: Prof. Rajendra Singh**

**Department Name: Agricultural and Food Engineering**

**Institute Name: Indian Institute of Technology Kharagpur**

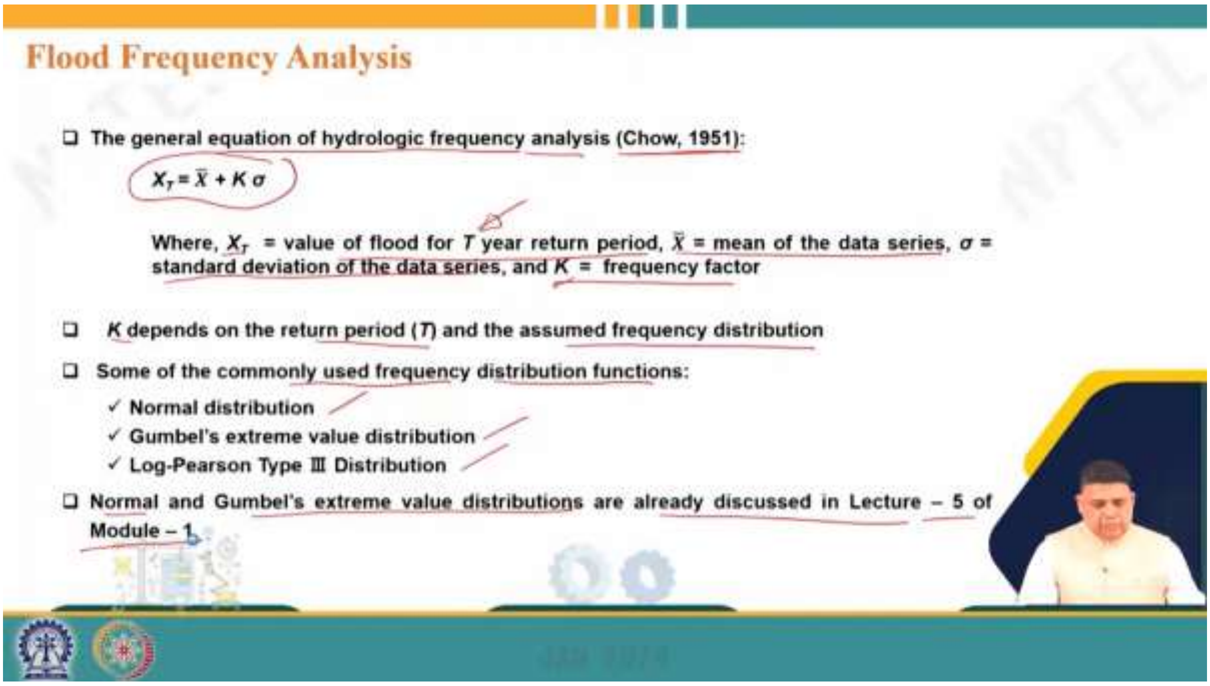
**Week: 09**

**Lecture 43: Flood Frequency Analysis**

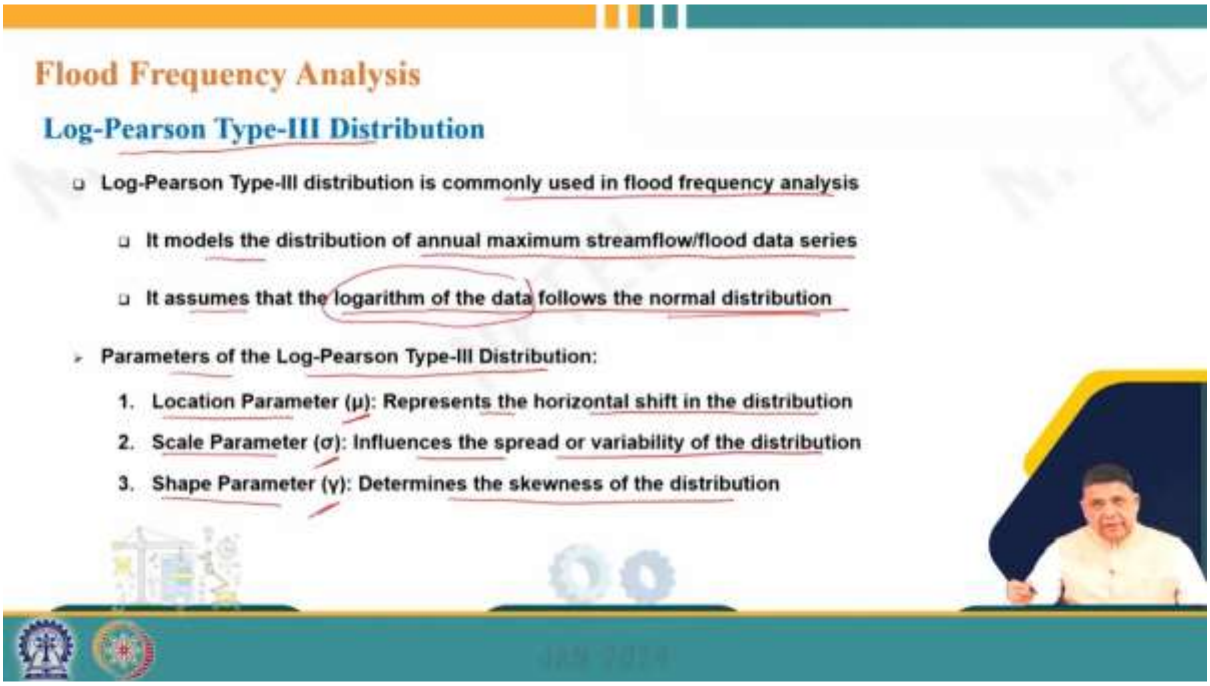
Hello friends, welcome back to this online certification course on Watershed Hydrology. I am Rajendra Singh, a professor in the Department of Agriculture and Food Engineering at the Indian Institute of Technology Kharagpur. We are in Module 9, this is Lecture Number 3, and the topic is Flood Frequency Analysis.


In this lecture, we will discuss flood frequency analysis, introduce confidence limits, and talk about risk, reliability, and safety factors.


Now, coming to flood frequency analysis, if you remember, we discussed frequency analysis during rainfall, and at that time, we mentioned that the techniques or methodology for frequency analysis remain the same, be it rain, flood, or any other hydrologic variable. So, if you recall, we discussed a general equation of hydrological frequency analysis given by Chow in 1951. According to this equation, $xt = x + k\sigma$, where $xt$ is the value of flood for a return period T, which is the period we are interested in to find the peak flood. X bar is the mean of the data series, $\sigma$ is the standard deviation of the data series, and $k$ is the frequency factor. So, obviously, you need to have the flood data with you, and then, by fitting the distribution, we can find out the mean and standard deviation. And of course, this $k$ depends on the return period T and the assumed frequency distribution. So, for the frequency distribution and T value, we can either calculate K or obtain it from the standard tables, which are available as we discussed earlier in this lecture. Some of the commonly used frequency distribution functions are the normal distribution, Gumbel's extreme value distribution, and the log Pearson type 3 distribution. Out of these, the normal and Gumbel's extreme value distributions were discussed in Lecture 5 of Module 1, where we covered rainfall frequency analysis. However, we will still go through an example of Gumbel's extreme value distribution and discuss the log Pearson type 3 distribution.

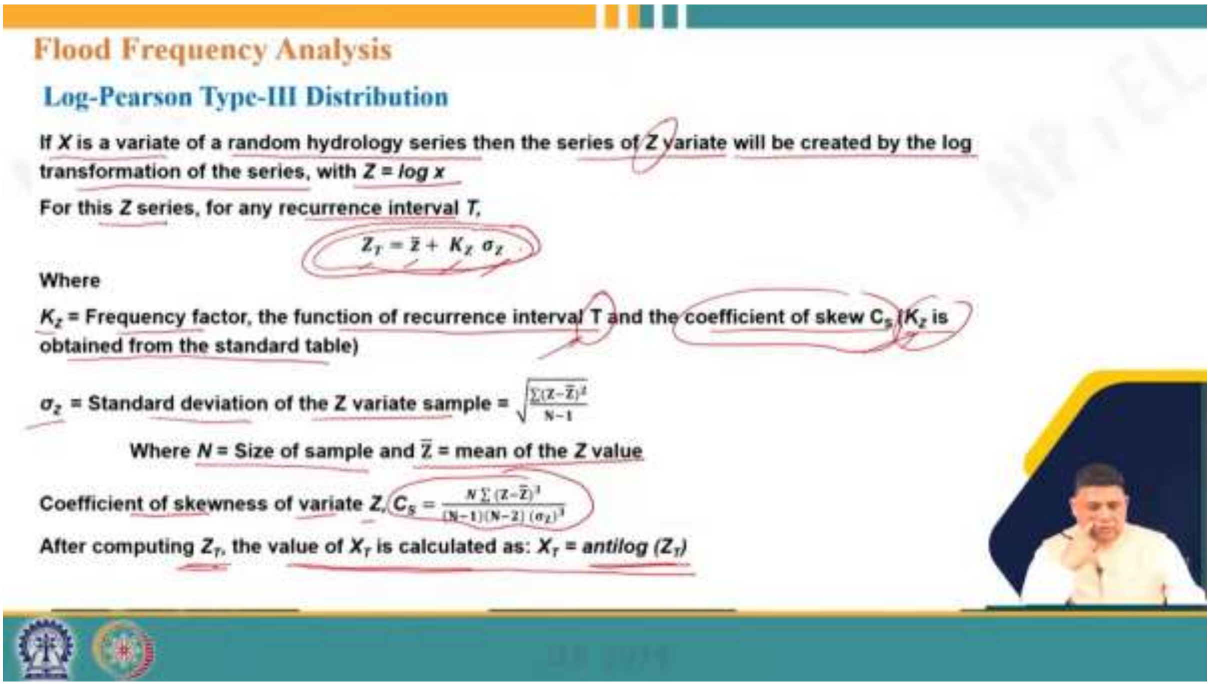



The log Pearson type 3 distribution is commonly used in flood frequency analysis. It models the distribution of inward maximum stream flow or flood data series and assumes that the logarithm of the data follows a normal distribution. As we discussed earlier, we first fit a distribution and try to convert it into a normal form, which is also the case with the log Pearson type 3 distribution. Essentially, it assumes that the logarithm of the data follows a normal distribution and is suitable for analysing inward maximum flood data series, although it can be used for other time steps as well. Now, regarding the parameters of the log Pearson type 3 distribution, it is a three-parameter model consisting of a location parameter $\mu$, a scale parameter $\sigma$, and a shape parameter $\gamma$. In contrast, the normal distribution is a two-parameter distribution, with mean and standard deviation being the two parameters. Here, we have scale, location, and shape parameters. The location parameter represents the horizontal shift in the distribution, meaning it affects the horizontal scale. The scale parameter influences the spread or variability of the distribution, while the shape parameter determines the skewness of the distribution which relates to its peakedness, as we discussed earlier. So, these are the three different parameters of the log Pearson type 3 distribution model.

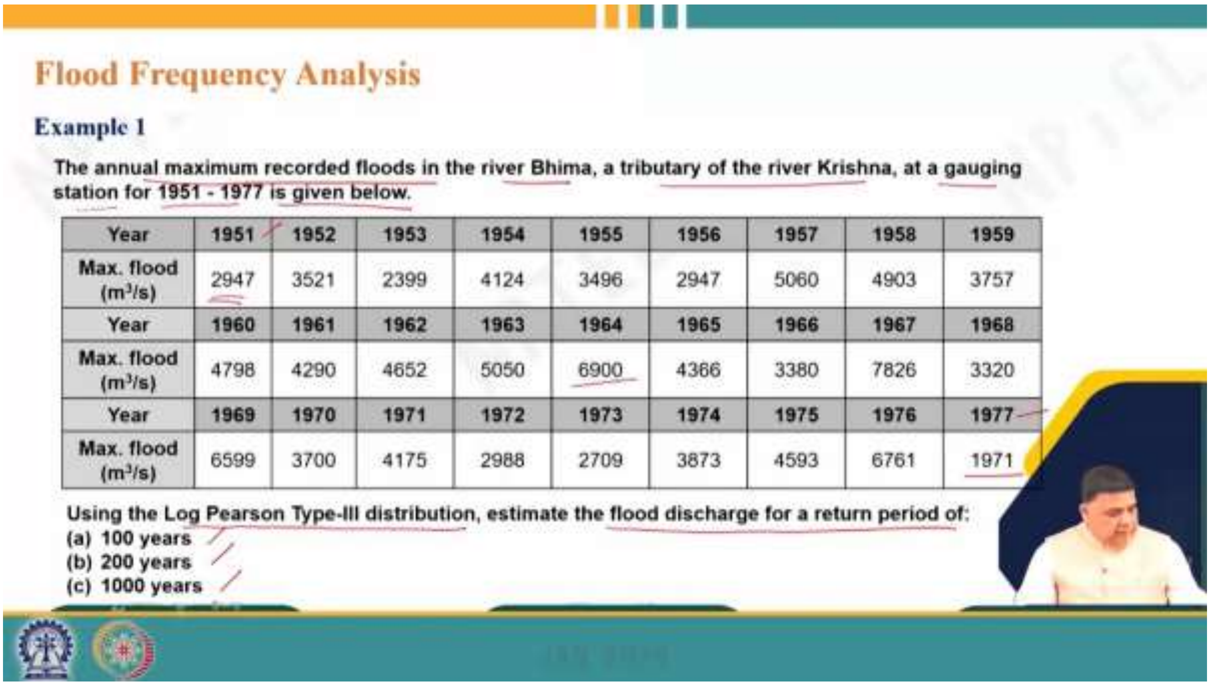



If $x$ is a variate of a random hydrological series, then a series of Z variate will be created by taking the logarithmic transformation of the series, denoted as $Z=\log(x)$. Essentially, if $x$ is the given series, we take the logarithm of that to obtain the transformed variate Z, which we refer to as the $Z$ series or the log-transformed series. For this $z$ series, for any recurrence interval T, the formula becomes $Zt=\bar{z}+Kz\sigma t$, where we are expressing this in terms of $Z$ to represent that $z$ is a log-transformed data, and $kz$ is the frequency factor, a function of the recurrence interval $t$, and the coefficient of skewness $Cs$. The value of $kz$ is obtained from standard tables available for the log Pearson type 3 distribution. The skewness coefficient in the case of the log Pearson type 3 distribution is calculated using the formula, and knowing the skewness and $t$, we can obtain the value of $kz$ from the standard tables available for the log Pearson type 3 distribution. $\sigma z$ represents the standard deviation of the $z$ variate sample, which is calculated using the standard formula $(z-\bar{z})2n-1n-1(z-\bar{z})2$, where $nn$ is the size of the sample, $\bar{z}$ is the mean of the $z$ values. The coefficient of skewness of the variate $z$ is given by the relationship, which again depends on $z\bar{j}$ mean, $\sigma$, the variate itself, and the number of data points. So, $nn$ represents the number of data points or the size of the sample, $\bar{z}$ and $\sigma$ are the mean and standard deviation of the transformed $z$ data, and $z$ is, of course, the variate.

After computing $zt$, the value of $xt$ is calculated as $xt=ezt$. So, once we know $Cs$ and the recurrence interval, we can obtain the value of $kz$ from the standard table. Using this, we can then find the value of $zt$. The last line indicates that after computing $zt$ using this generalized form of hydrologic frequency analysis equation, the value of $xt$ is calculated by transforming it back using $ezt$. We will demonstrate this process through an example.

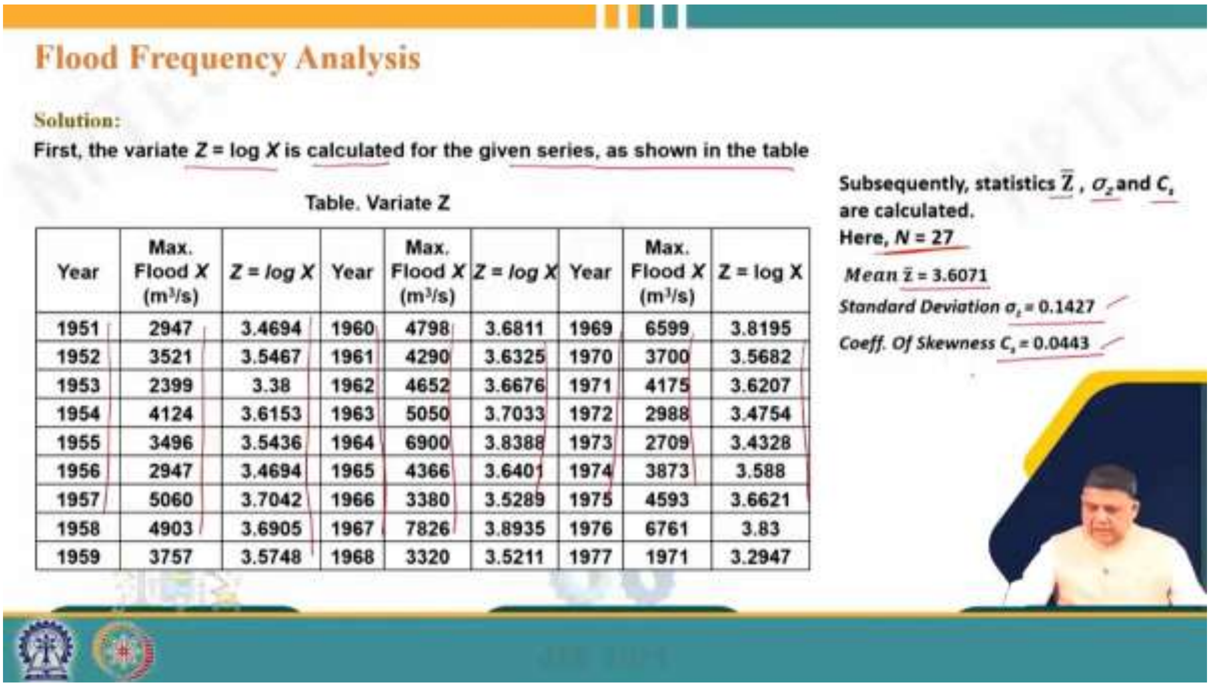



Now, let's consider Example 1: The annual maximum recorded floods in the river Vima, a tributary of the River Krishna, at a gaging site from 1951 to 1977 are given below. The maximum flood values recorded in cubic meters per second for each year are provided. Using the log Pearson type 3 distribution, we need to estimate the flood discharge for return periods of 100, 200, and 1000 years. We will find the flood values for these three different return intervals. As discussed, we first calculate the variate $z$ which equals $\log(x)$, for the given series, as shown in the table.



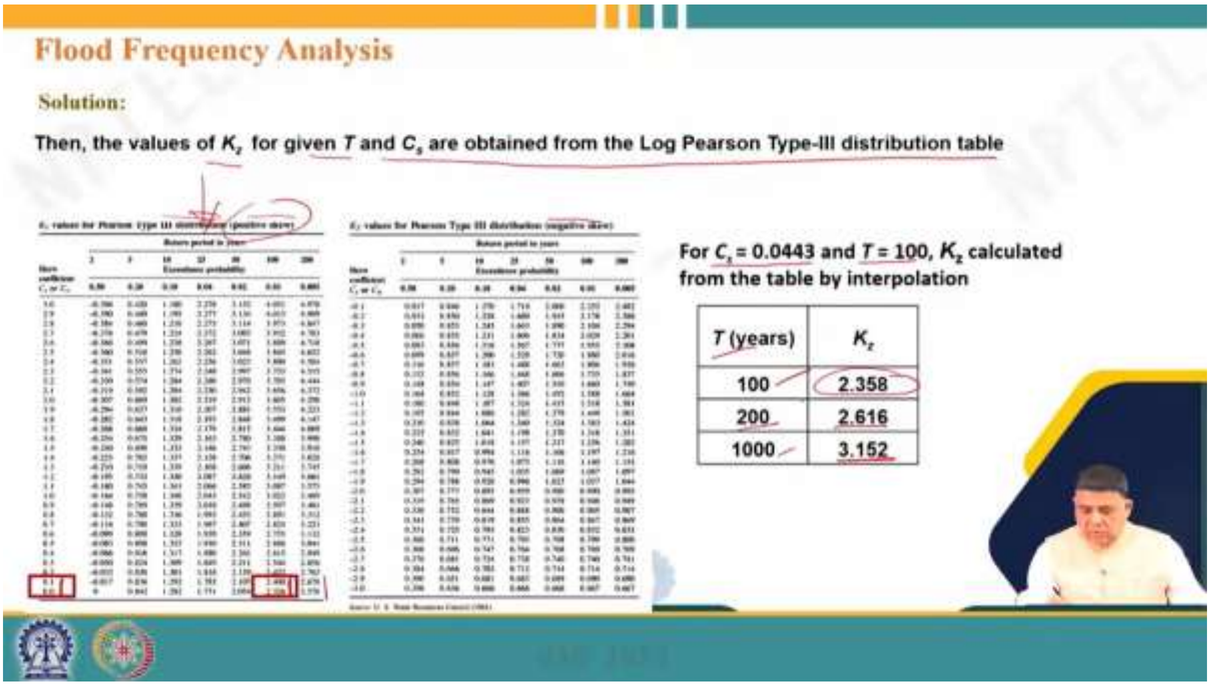

So, here we have the $x$ values listed. For different years, $x$ values are listed, and we will take the logarithm of each value to obtain the $z$ value. So, for each year $x$ and $z$ values are listed.

The first step we take is to log-transform the data because the log Pearson type 3 distribution assumes that the logarithm of the data series follows a normal distribution. Therefore, the first thing we have to do is use the log transformation. Then, once we have the $z$ series, we calculate statistics like $\bar{z}$, $\sigma$, $n$, and $C_s$. Here, $n$ is 27 because the data range from 1951 to 1977. By using this data, we can calculate $z$ bar, which is 3.6071, the standard deviation which is 0.1427, and the coefficient of skewness, which is 0.0443. To calculate these values, we need $n$, $z$, $\bar{z}$, $\sigma$, and $C_s$. Once we obtain the mean, we calculate the standard deviation, where we need the mean. To calculate the coefficient of deviation, we need both the mean and $\sigma$. We can calculate these using the standard formulas.

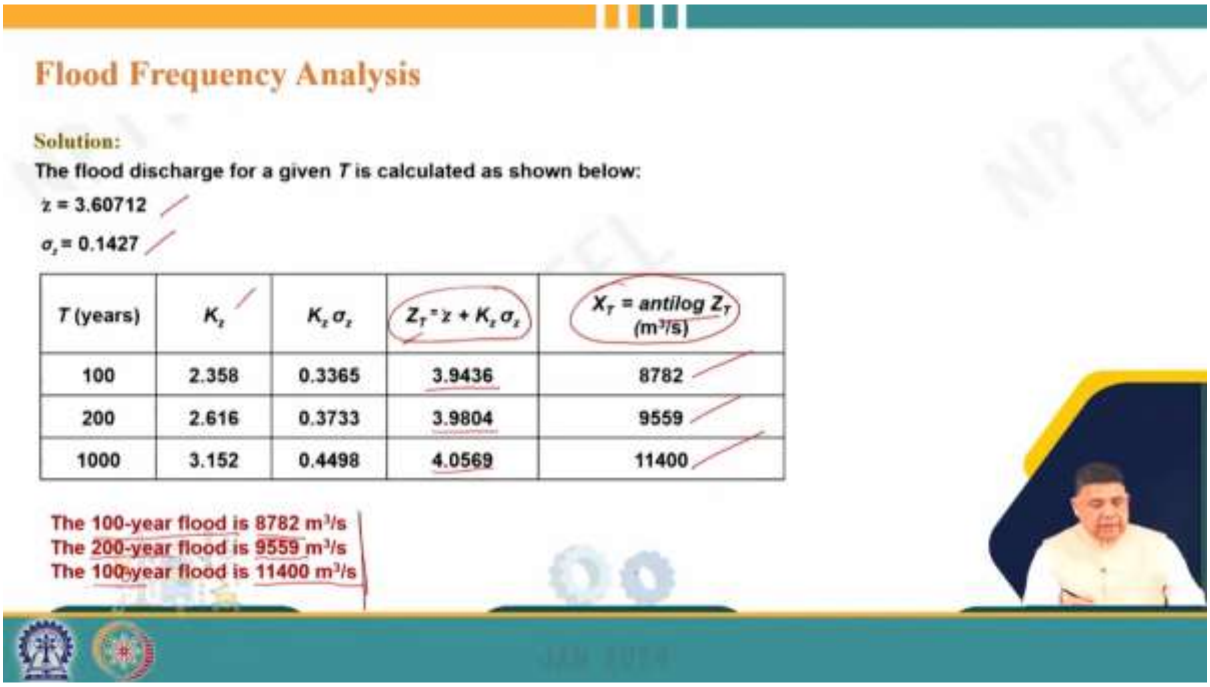



| Year | Max. Flood X (m³/s) | Z = log X | Year | Max. Flood X (m³/s) | Z = log X | Year | Max. Flood X (m³/s) | Z = log X |
|---|---|---|---|---|---|---|---|---|
| 1951 | 2947 | 3.4694 | 1960 | 4798 | 3.6811 | 1969 | 6599 | 3.8195 |
| 1952 | 3521 | 3.5467 | 1961 | 4290 | 3.6325 | 1970 | 3700 | 3.5682 |
| 1953 | 2399 | 3.38 | 1962 | 4652 | 3.6676 | 1971 | 4175 | 3.6207 |
| 1954 | 4124 | 3.6153 | 1963 | 5050 | 3.7033 | 1972 | 2988 | 3.4754 |
| 1955 | 3496 | 3.5436 | 1964 | 6900 | 3.8388 | 1973 | 2709 | 3.4328 |
| 1956 | 2947 | 3.4694 | 1965 | 4366 | 3.6401 | 1974 | 3873 | 3.588 |
| 1957 | 5060 | 3.7042 | 1966 | 3380 | 3.5289 | 1975 | 4593 | 3.6621 |
| 1958 | 4903 | 3.6905 | 1967 | 7826 | 3.8935 | 1976 | 6761 | 3.83 |
| 1959 | 3757 | 3.5748 | 1968 | 3320 | 3.5211 | 1977 | 1971 | 3.2947 |

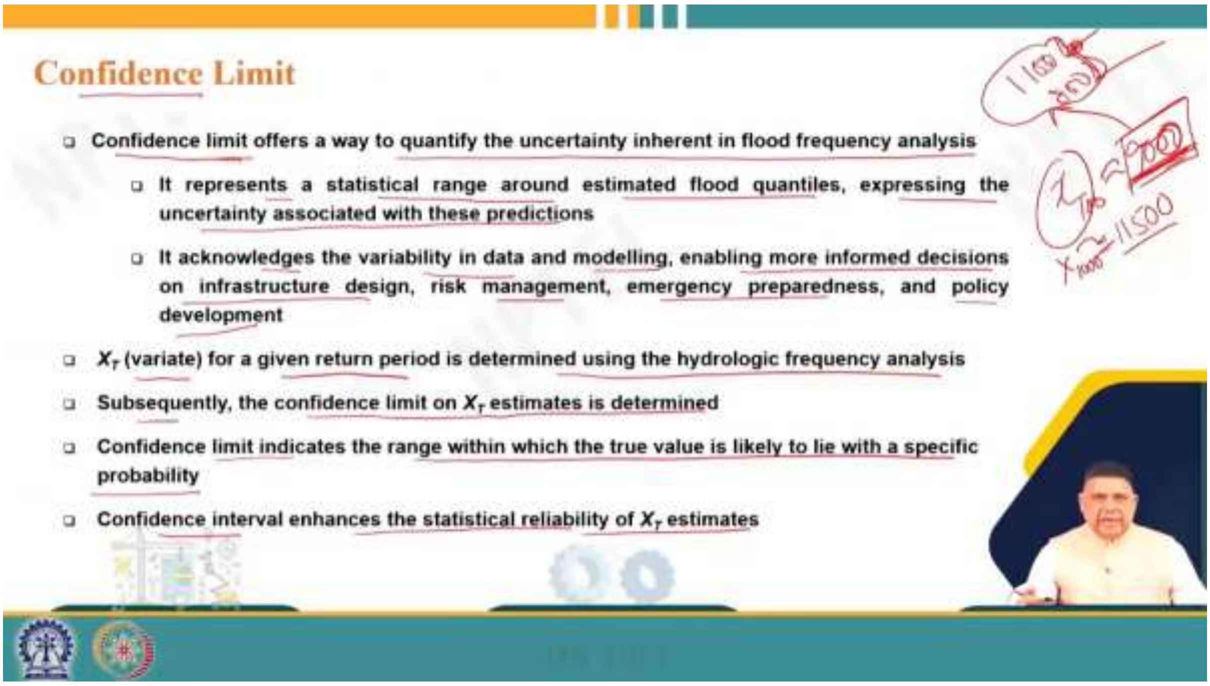
Now, the value of $K_z$ for a given $tt$ and $C_s$ is obtained from the log Pearson type 3 table. There are two tables provided; this is only a sample table. The key $t$ value for Pearson type 3 distribution is for positive skewness where $C$s is positive and for negative skewness where $C_s$ is negative. In our case, $C_s$ is positive, so we will follow this table to obtain the value of $K_z$. For $C_s$ = 0.0443 and $t$ = 100, as we have been asked to calculate for 100 years, 200 years and 1000 years. For each year and each of these return periods, we have to obtain the $K_z$ value from the table. So, for a return period of 100 years, our $C_s$ value lies between 0 and 0.1 with our value being 0.0443. For 100 years, the $K_z$ values provided in the table are 2.4 and 2.326. Hence, we need to interpolate the value, and upon interpolation, we obtain $K_z$ as 2.358 for a return period of 100 years. Similarly, for 200 years, the interpolated value lies between 2.576 and 2.670, resulting in $K_z$ being 2.616. Though the value for 1000 years is not shown in this table, there is a separate table available and from there, we find that $K_z$ is 3.152 for a return period of 1000 years. Thus, we now have the $K_z$ values.



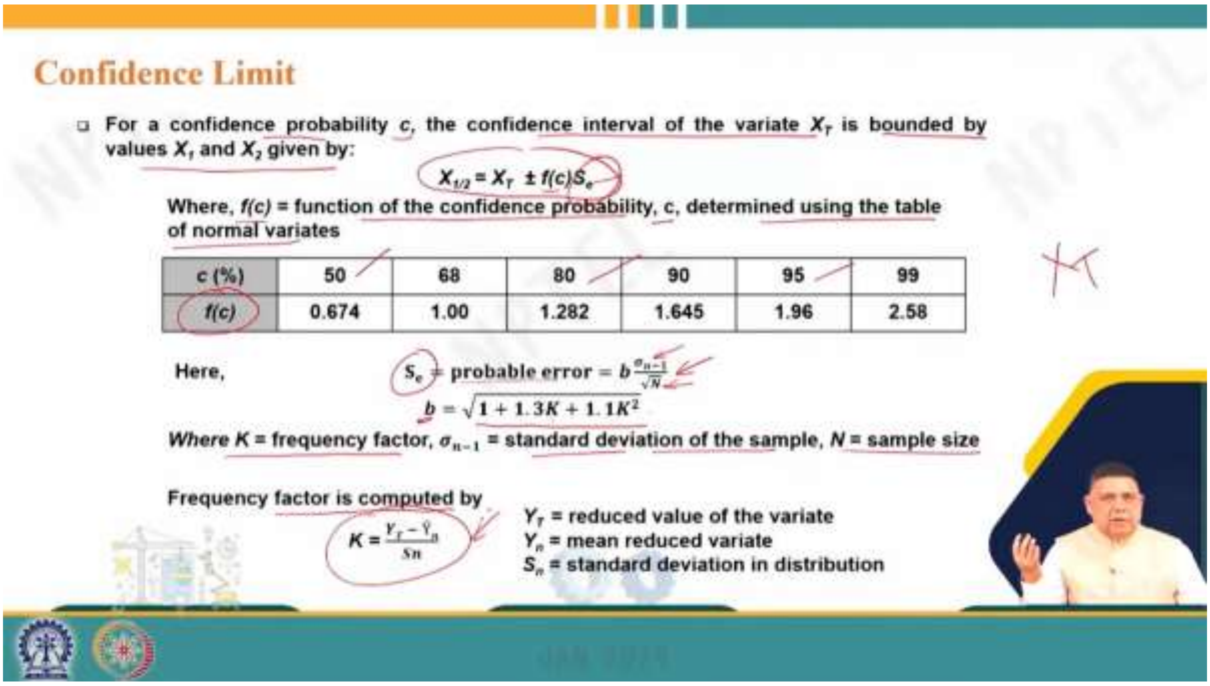

Having obtained $\bar{z}$, $\sigma$ and $K_z$ for different return periods, we can use the formula to calculate $z_t$ values for 100, 200, and 1000 years. The values obtained are 3.9436, 3.9804, and 4.0569, respectively. To obtain the peak values, we need to reverse the logarithmic transformation by taking the antilog. Upon doing so, the values are found to be 878, 9559 and 11400 respectively. Therefore, the 100-year flood for the given data series is 878 cubic meters per second, the 200-year flood is 9559 cubic meters per second, and the 1000-year flood is 11400 cubic meters per second. So, these are the values we calculated for the given data series using the log Pearson type 3 distribution. This is how we apply the log Pearson type 3 distribution function in our calculations.



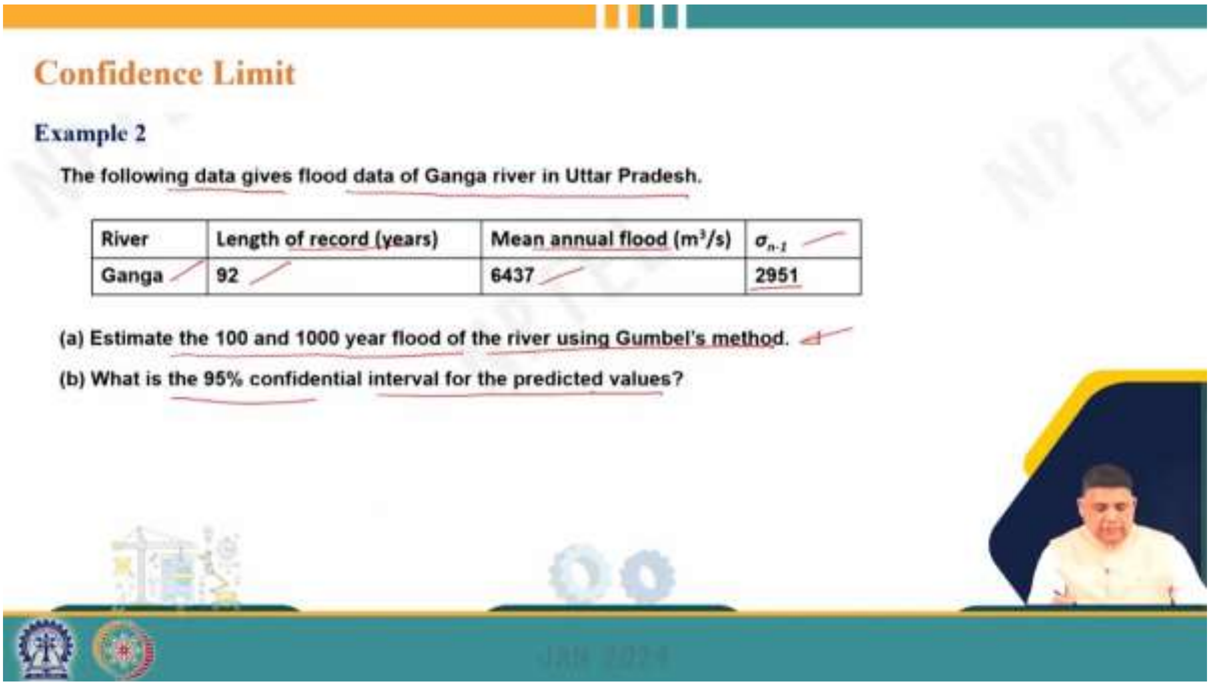
Now, let's move on to confidence limits. Confidence limits offer a way to quantify the uncertainty inherent in flood frequency analysis. They represent a statistical range around estimated flood quantiles, expressing the uncertainty associated with these predictions. Confidence limits acknowledge the variability in data and modelling, enabling more informed decisions on structure design, risk management, emergency preparedness, and policy development. Essentially, what we are saying is that we have already calculated the value of $x_t$. For example, in the previous problem, we calculated a certain value for 100 years, let's say around 9000, and similarly for 200 years. For 1000 years, it was around 11500, let's assume. These were absolute values calculated by this distribution. However, we are not very sure of the distribution function or the data or the parameters we have calculated. All these things are uncertain. To address this uncertainty, we obtain values around these estimated values. These values indicate that within this range, this particular estimated value may range over, for instance, for this 900, we may get 1100 and 800 as two limits. So, that means, if the flow value is between 800 and 1100, as we have calculated, this is the confidence limit within which this value may range. We have to be careful while making decisions. The $x_t$ variate for a given return period is determined using hydrologic frequency analysis. Subsequently, the confidence limit on $x_t$ estimates is determined. As I mentioned, for 900, it could be 1100 and 800. The confidence limit indicates the range within which the true value is likely to lie with a specific probability. Of course, when we estimate this confidence limit, we have to assign a particular probability, and the confidence interval enhances the statistical reliability of the $x_t$ estimate.



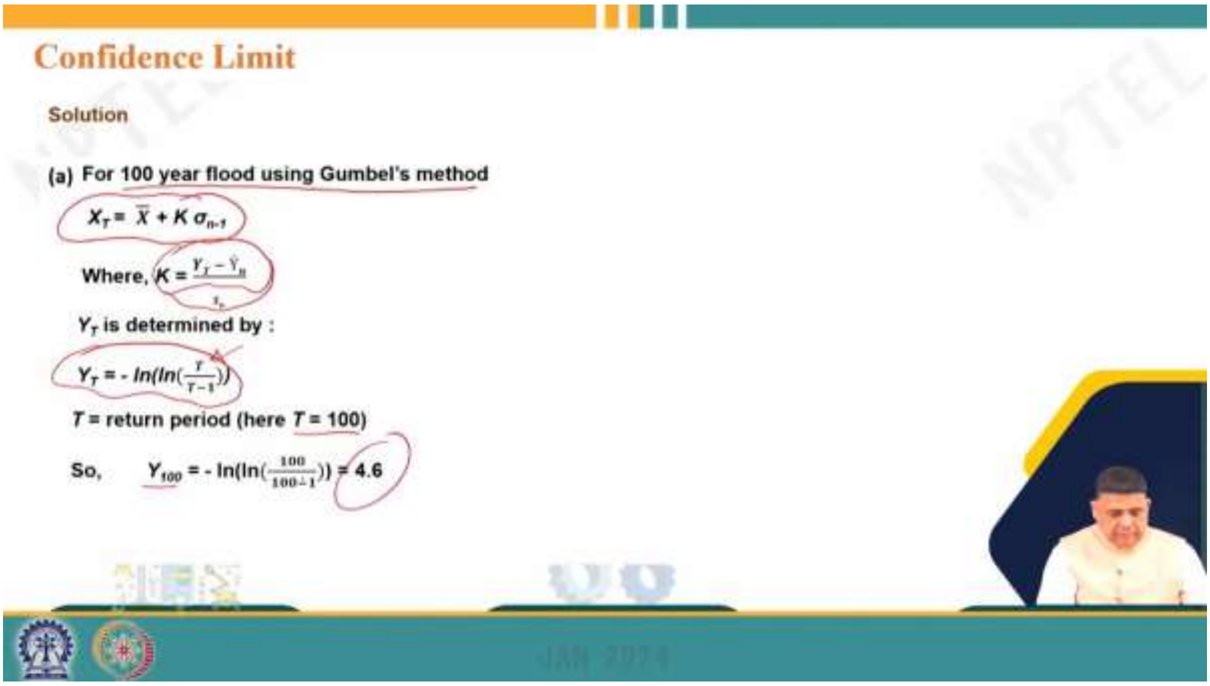

So, for a confidence probability $CC$, the confidence interval of a variate $xtxt$ is bounded by values $x_1$ and $x_2$, given by this relationship: $x_1$ or $x_2$ is $x_t \pm f_c S_{ext} \pm fc\ Se$, where $f_c$ is a function of confidence probability $C$ determined using the table of normal variates, which is given here. Depending on whether you take a confidence probability of 50, 80 or 95, your $f_c$ value will vary, and this $f_c$ value can be taken from this table. Another term used here in this equation is $SeSe$, which is referred to as the probable error and is given by this relationship: $B\sigma n(n-1)$ $B_n(n-1)\sigma$. Here, $\sigma\sigma$ is the standard deviation, $n$ is the number of data points. We already know this value, and $BB$ is a function of $k$. So, it is the square root of 1 plus 1.3k plus 1.1k squared, and $k$ is the frequency factor. We have already discussed how to determine this frequency factor. $\sigma\sigma$ is the standard deviation, and $nn$ is the sample size. So, basically, while deciding or computing the value of $x_t$, we will know all these things. We will know $k$, we will know $x_t$, and the only thing we have to decide is the confidence probability and get the value of $cc$ from here. Otherwise, with whatever we know, we can calculate the value of $S_e$. The frequency factor can be computed by this formula, depending on the probability function. This is the one used for Gumbel's distribution. Otherwise, we know how to determine that for the log Pearson type 3 distribution. Just now, we discussed earlier also we discussed for normal distribution that the normal table could be used. We also showed that there are direct equations which can be used for finding out the value of $k$.



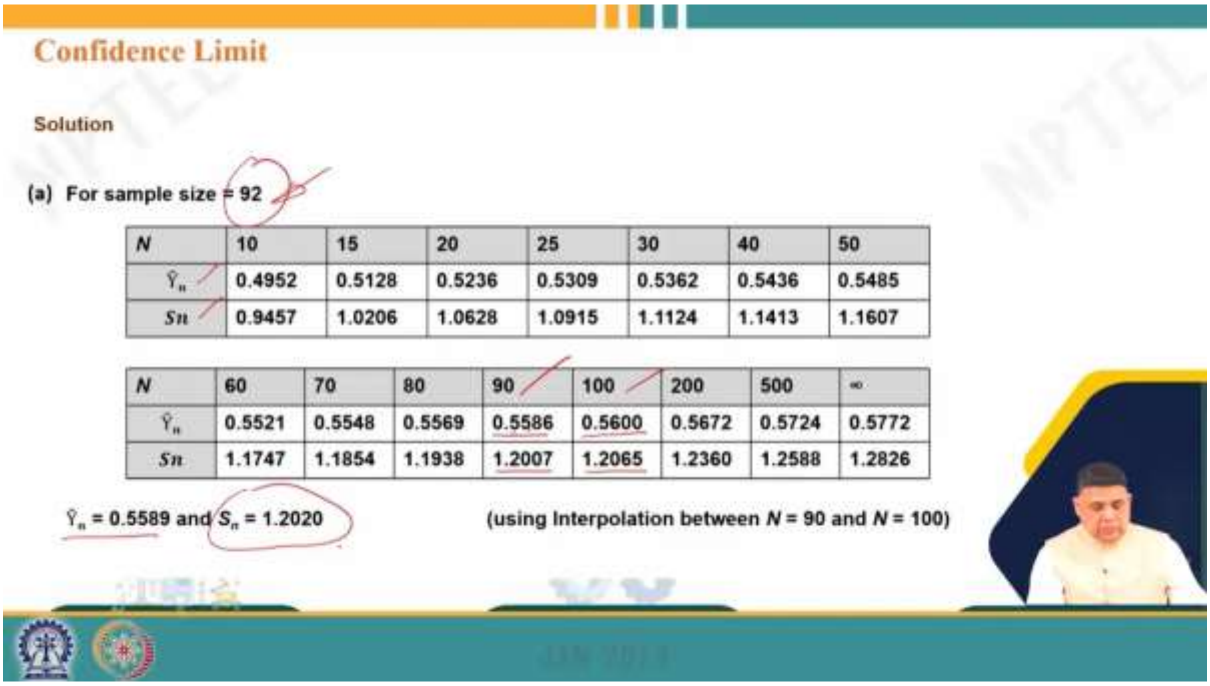

So, let's take an example here. The following data gives flood data from the Ganga River in Uttar Pradesh. The length of the record is 92 years, the mean annual flood is 6437 cubic meters per second, and the standard deviation is 2951. Estimate the 100 and 1000-year flood of the river using Gumbel's method, and what is the 95 percent confidence interval for the predicted value? So, we have to first get the $x_t$ value and then the 95 percent confidence interval on the estimated $x_t$ value. These are the two components we have to do.



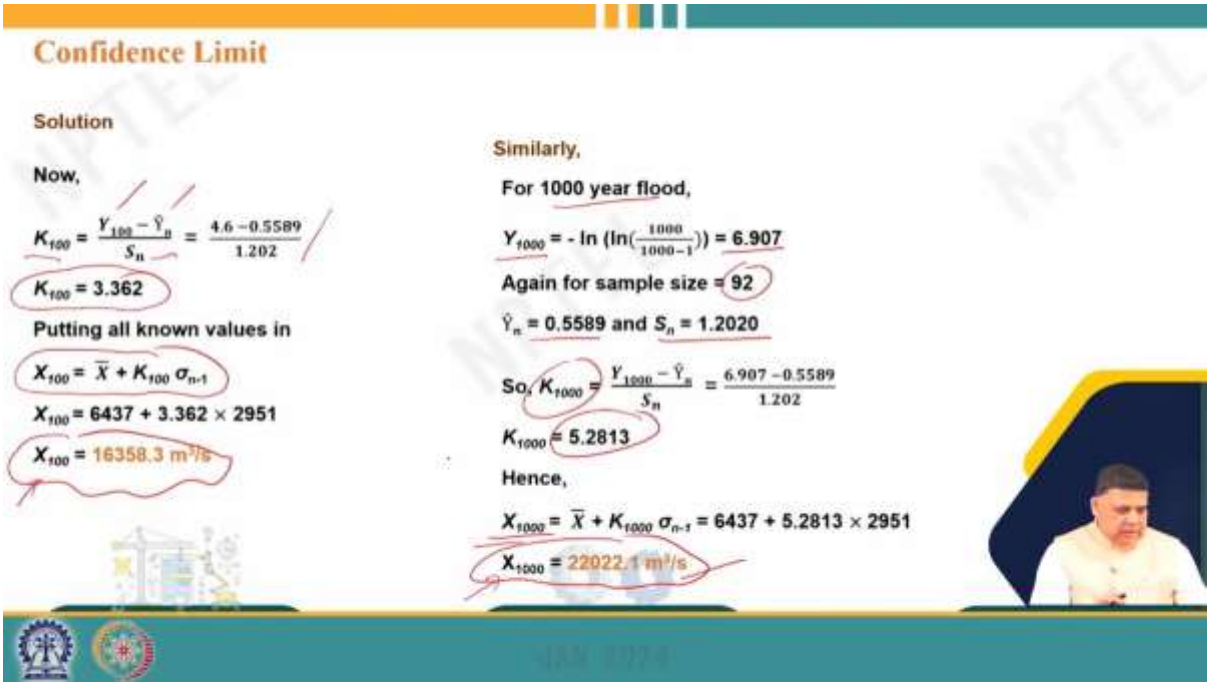

So, let's start with the 100-year flood using Gumbel's method. This is the generalized equation we know, where the $kk$ value is given by this relationship. Just now we discussed where $y_t$ is determined by this. It's a log-transform variate given by this relationship, $-\ln(t/(t-1))-\ln(t/(t-1))$, where it is just a function of $tt$. So, for $tt$ equals to 100 years, $y100y100$ can be calculated, and it comes out to be 4.6.



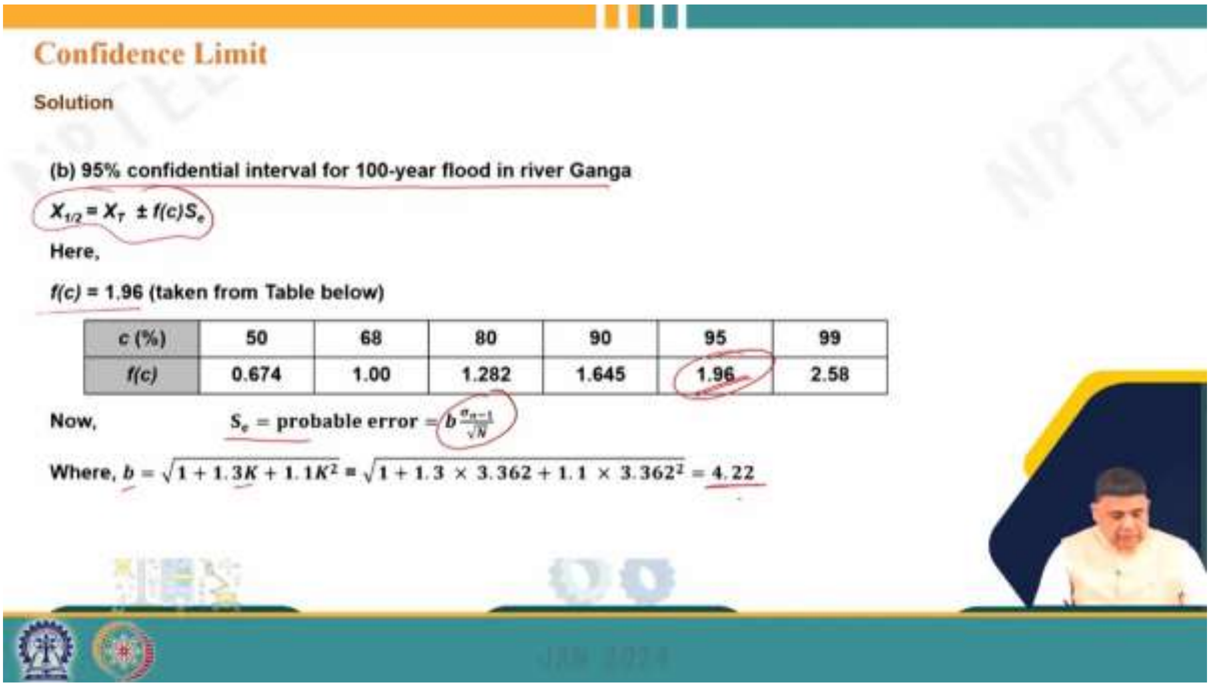

Then, for sample size 92, that is, $\bar{y}$ and $S_n$ values can be read from this table. Here, we have been given the values for 90 and 100. So, $\bar{y}$ is 0.5586 and 0.560, which will lie between these because our sample size is 92. So, we have to interpolate and through interpolation, we get a value of 0.5589. Similarly, $S_n$ has to be between 1.207 and 2.065. We have to interpolate and here we get for 92 $S_n$ equals to 1.2020.



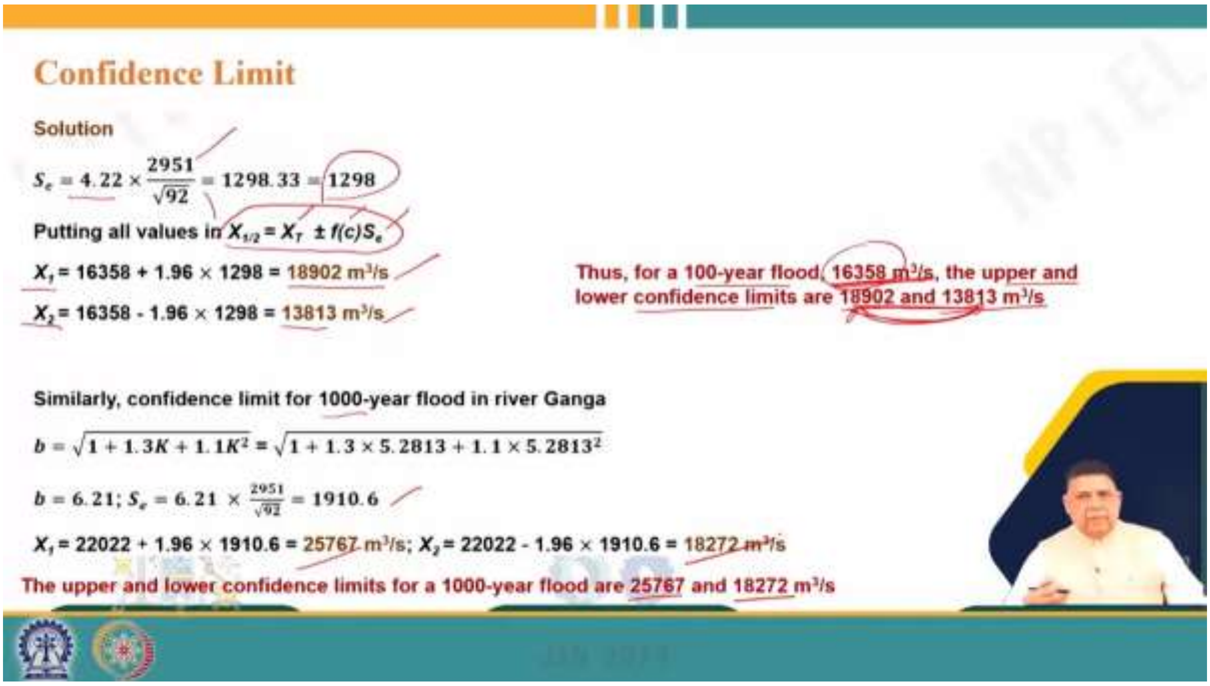

So, that means, now we know we can calculate $k_{100}$ because we know $y_{100}$, we know $\bar{y}$, we know $S_n$. So, putting the required values, we get $k100$ equals to 3.362. Putting in this generalized equation, all the knowns because we know $\bar{y}$, we know $\sigma\sigma$ because these values can already be given. So, $x_{100}$ comes out to be 16,358.3 cubic meters per second. So, that is the magnitude of the flood for a 100-year return period. The same procedure can be adopted for a 1000-year flood, and for a 1000-year flood, $y_{1000}$ comes out to be 6.907. Samples are 92, so $\bar{y}$ and $S_n$ values will remain the same. So, $k_{1000}$ comes out to be 5.2813, and then using the generalized equation we get $k\times1000$ as 22,022.1 cubic meters per second. So, the first part is that using the Gumbel's distribution, we have obtained the flood magnitude for 100 years and 1000 years return period. So, these two things we have done first.



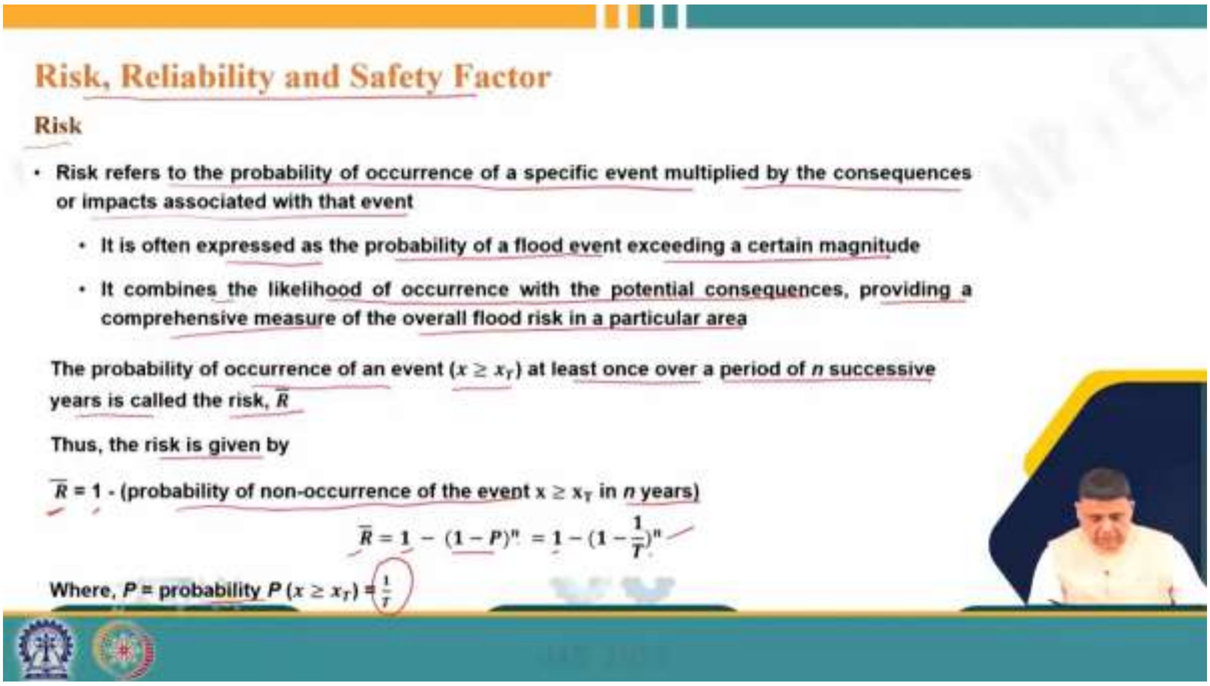

Now, we will go into the confidence interval, basically, and so, the 95 percent confidence interval for the 100-year flood in the Ganga River. We have to use this formula where $f_c$ has to be taken from this table. So, it comes out to be 1.96. So, $f_c$ is 1.96, probable error $S_c$ is $B\sigma n-1$, and already we have calculated the value of $kk$. So, putting the value of $kk$ in here, we can calculate the value of $B$ equals to 4.22.



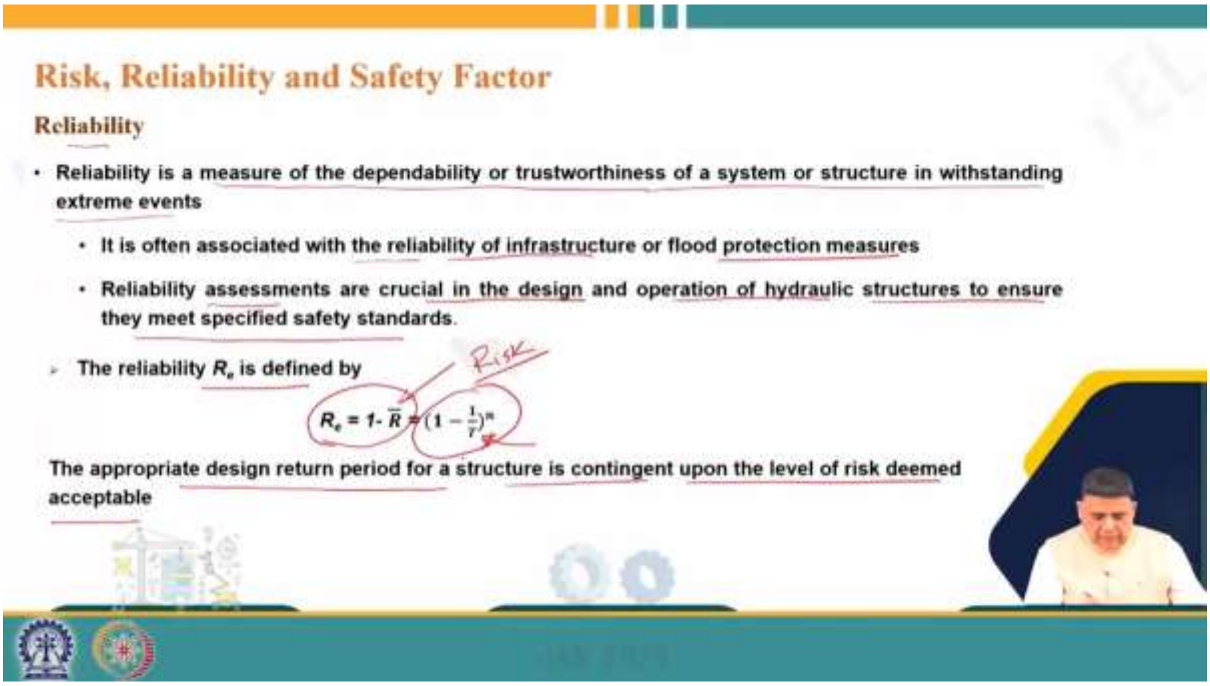

And once $BB$ is known, we can calculate the value of $ScSc$ as 1298. These values are already given to us. So, it comes out with 18,902 and 13,813 cubic meters per second. So, for a 100-year flood which the magnitude was 16,358, the upper and lower confidence limits are 18,902 and 13,813. So, that means, this confidence limit says that though we have estimated a value of 16,358, but keeping the uncertainty, this value could lie anywhere between these two limits. Similarly, for a 1000-year flood, we will calculate the value of $B$, and then we will calculate the value of $x_1$ and $x_2$, and so, for $x_1$, it is 25,767, and $x_2$ is 18,272. So, the upper limit for a 1000-year flood is 25,767 and 18,272. So, this is how we can once we know the $x_t$ value, we can also determine the confidence limits on the values.



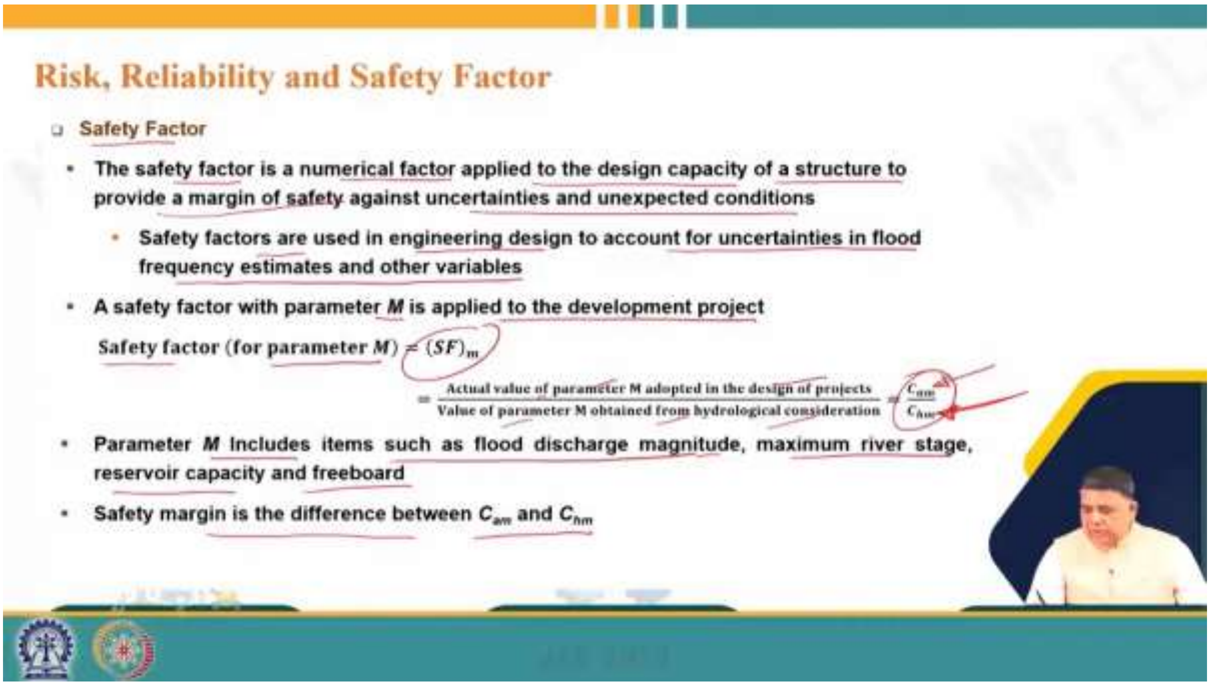

Now, we go into risk reliability and safety factor. Risk refers to the probability of occurrence of a specific event multiplied by the consequences or impact associated with that event. So, obviously, when we talk in terms of risk, that means, there is some chance of a mishap and we are ready to face the consequences in case of failure. That is the risk we are talking about. It is often expressed as the probability of a flood event exceeding a certain magnitude, combining the likelihood of occurrence with the potential consequences, providing a comprehensive measure of the overall flood risk in a particular area. So, obviously, you are fully aware of the consequences when you take a risk actually. The probability of occurrence of an event $x \geq x_t$ at least once over a period of unsuccessful years is called the risk. So, risk is given by $\bar{R}\bar{R}$, which is 1−1− the probability of non-occurrence of the event $x \geq x_t$ in $nn$ years or $\bar{R}\bar{R}$ is $1-(1-P)n$, and it is in terms of the recurrence interval or return period $1-(1-t)n$, where $P$ is the probability and $tt$ is equal to $1/t$ or the recurrence interval. So, this is basically what we get.



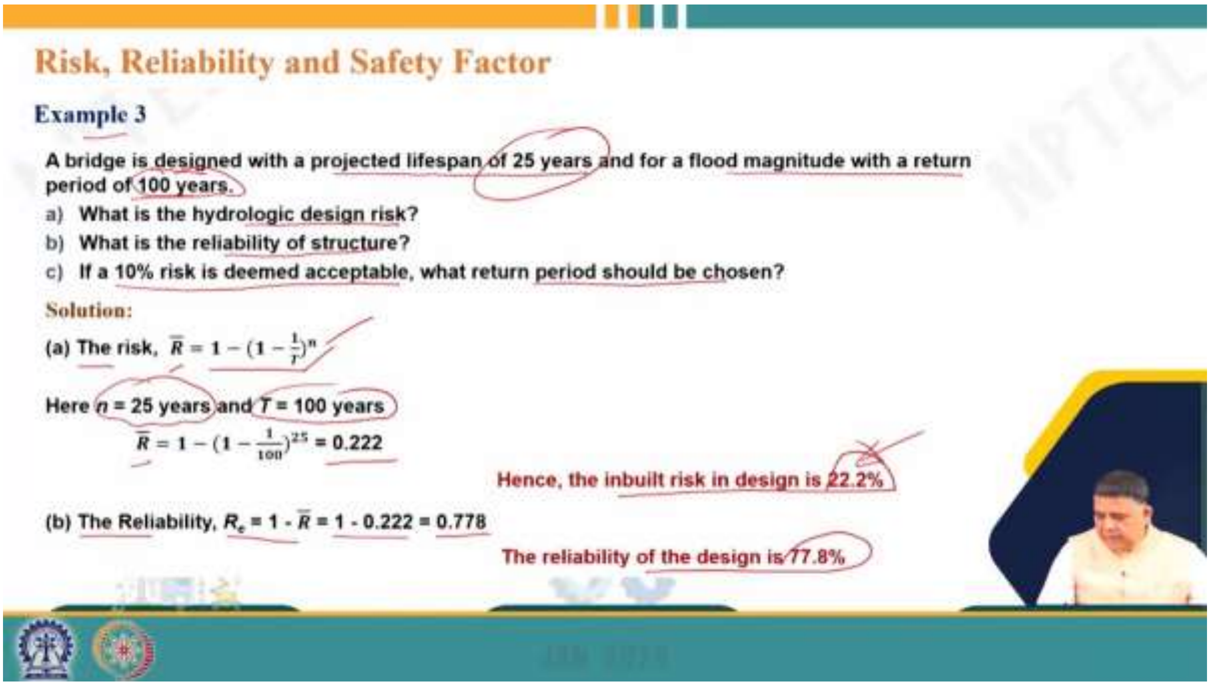

And if you talk about reliability, it's the measure of the dependability or trustworthiness of a system or structure in withstanding extreme events. So, how reliable the design is, that is what we are talking about. In the previous case, we thought about what risk we are going to take; in this case, how reliable our system is. So, just two different things we are talking about, often associated with the reliability of infrastructure or flood protection measures. Reliability estimates are crucial in the design and operation of hydraulic structures to ensure they meet specified safety standards, and reliability is basically defined as $1-\bar{R}$, where $\bar{R}$, as you know, is nothing but risk. So, risk is calculated, and then 1−risk is reliability. So, it is $1-1t_n$. So, the appropriate design return period for a structure is contingent upon the level of risk deemed acceptable. So, how much risk you are going to take, of course, depends on the recurrence interval we are going to adopt for a particular design.



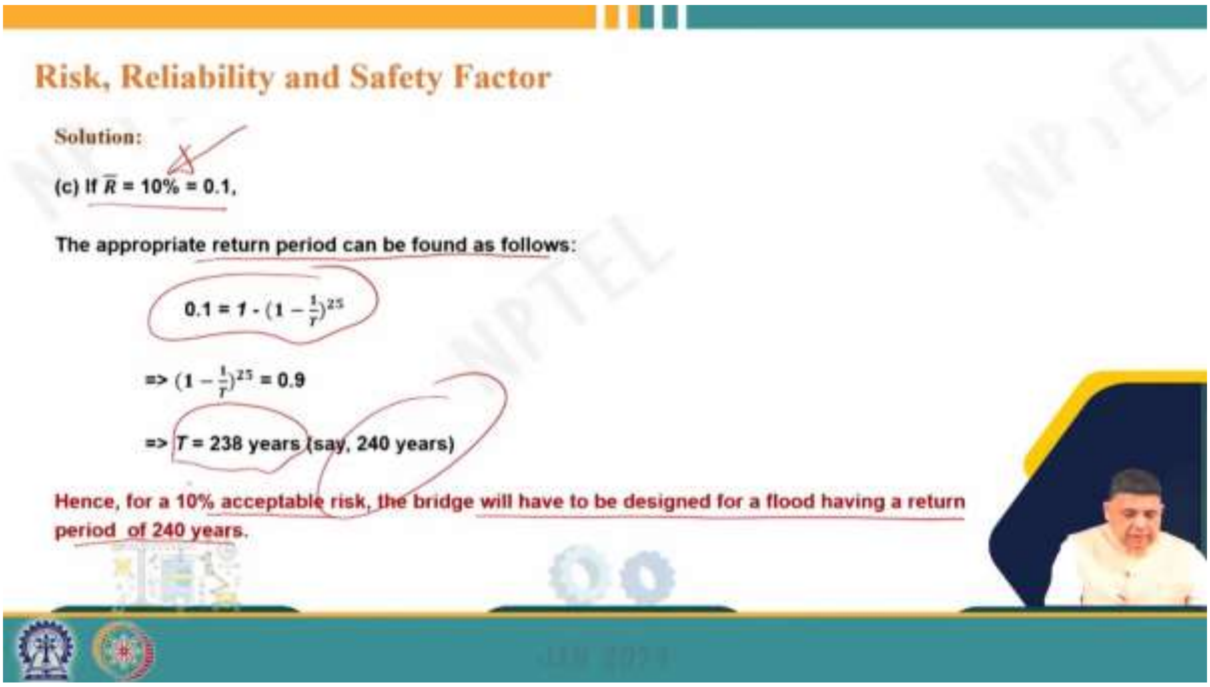

Then we come to the safety factor, which is here. The safety factor is a numerical factor applied to the design safety capacity of a structure to provide a margin of safety against uncertain, unexpected conditions. Safety factors are used in engineering design to account for uncertainties in flood frequency estimates and other variables. So, what safety you are going to provide in your design, that is the safety factor. The parameter $mm$ is applied to the development project, and the safety factor for parameter $mm$, $Sf_m$, is the ratio of the actual value of the parameter adopted in the design to the value of parameter $mm$ obtained from hydrologic considerations, so $Cam$ by $Chm$. So, basically, this parameter value we are obtaining after analysing the data, but over and above that, we want to provide some kind of safety. So, that is why we adopt a higher value of $Cam$, and that is how this safety factor is there. Parameter $mm$ includes items such as flood discharge magnitude, maximum river stage, reservoir capacity, and freeboard. So, anything can be used as a parameter, and the safety margin is the difference between $Cam$ and $Chm$. So, this is the actual value, and this is the adopted value. So, the difference between the two is referred to as a safety margin.



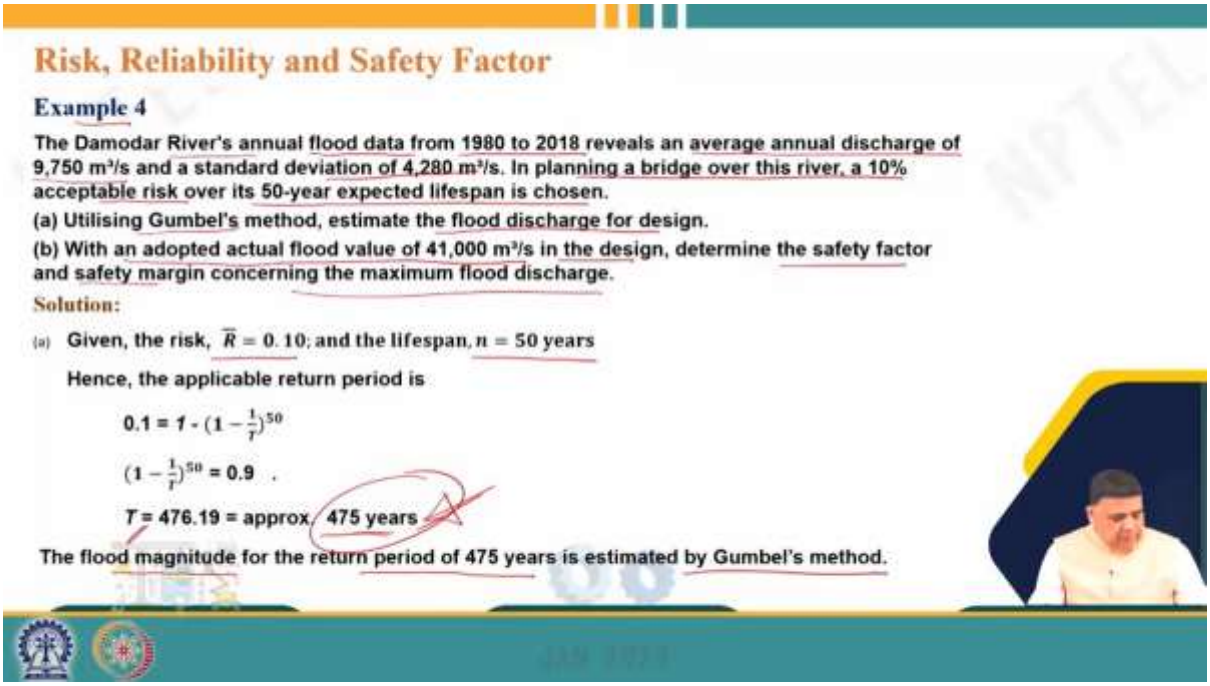

Let's take an example here: a bridge is designed with a projected lifespan of 25 years for a flood magnitude with a return period of 100 years. What is the hydrological design risk? What is the reliability of the structure? If a 10 percent risk is deemed acceptable, what return period should be chosen? That is what we are talking about. So, we have three different bits. The risk, as we already know, $\bar{r}\bar{r}$ is $1-t_n$. Here, $nn$ is because the projected lifespan is 25 years, so we are talking about $n$=25 years, and the return period considered is 100 years, so $t$=100 years. Putting values of $tt$ and $nn$ here, we get $\bar{r}$=0.222, which simply means that the inbuilt risk in this design is 22.2 percent. So, when we are designing this structure for 25 years with a 100-year return period, then we have an inbuilt risk of 22.2 percent that we are willing to take; that is what this means. Now, coming to bit B, reliability, and we already know the reliability is nothing but $1-\bar{r}$. So, that means 1−0.222 or 0.7780.778, which simply means the reliability of the design is 77.8 percent. So, our design, if we use this design, is 78 percent reliable.



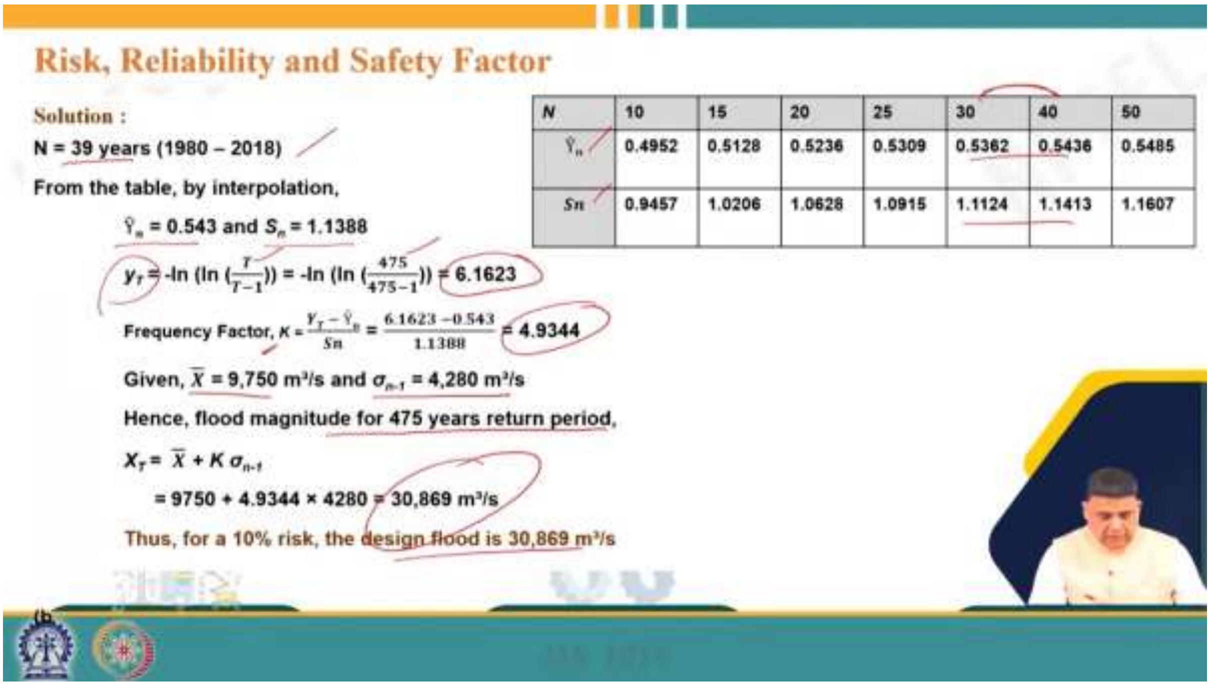

And then, if $r^-$ is 10 percent or 0.1, the appropriate return period can be found using this relationship, and from here, we get $t$=238. So, that simply means, earlier we saw that when we were using a return period of 100 years, our risk was 22 percent. So, if you want to take a risk of only 10 percent, then you have to choose a return period of 240 years for designing a structure. And that is why, as we discussed earlier, when we talked about large structures, we take $t$=1000 years, I mean, the lifespan may be 100 years, but we take a return period of 1000 years because we do not want to take that higher risk, and we want to minimize the risk in that particular design. So, for a 10 percent capital risk, the bridge will have to be designed for a flood having a return period of 240 years. So, with this, if you take $r^-$=1 percent, then obviously, you can guess what will be the return period, which will be much, much higher.



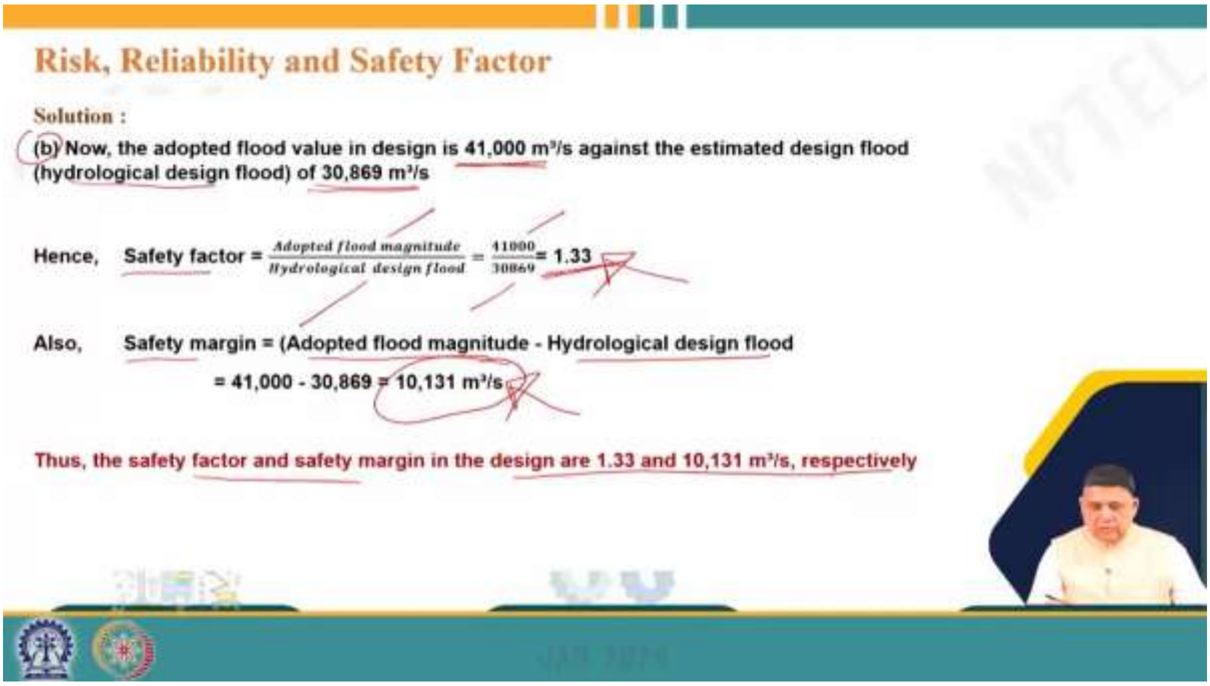
Let's take another example: the Damodar River's annual flood data from 1980 to 2018 reveals an average annual discharge of 9750 cubic meters and a standard deviation of 4280 cubic meters per second. In planning a bridge over this river, a 10 percent acceptable risk over its 50-year expected lifespan is chosen. Utilizing Gumbel's method, estimate the flood discharge for the design with an adopted actual flood value of 41000 cubic meters per second. In the design, determine the safety factor and safety margin concerning the maximum flood discharge. So, obviously, in this case, the risk value is 0.1; all other data is given, and we have to first find out the flood discharge. Then, of course, because we have already designed with a certain value, we need to calculate the safety margin. So, obviously, the first thing we have to do is calculate the value of $T_e$, the return period, which comes out to be 475. So, for this 10 percent risk, the lifespan, or rather the return period, has to be 475 years. So, obviously, we need to estimate the flood magnitude for a return period of 475 years using Gumbel's method.




And for Gumbel's method, we know that we need $\bar{y}$ and $S_n$, and our $nn$ value is 39 years. So, from here, with 39 years, we have to interpolate the values. Thus, we obtain $\bar{y}$ bar as 0.543 and $S_n$ as 1.1388. The $y_t$ value, which we need to calculate the variate in terms of $T_e$, which is 475, comes out to be 6.163. From this equation, we can calculate the frequency factor as 4.9344. Given $\bar{x}$ as 9750 and standard deviation as 4280, the flood magnitude for a 475-year return period using Gumbel's method comes out to be 30,869 cubic meters per second. Therefore, for a 10 percent risk with the given data, our flood magnitude is 30,869 cubic meters per second. Thus, the design flood is this value.




Now, regarding bit B, the safety factor and safety margin: the adopted flood magnitude is 41,000 cubic meters per second, whereas the calculated hydraulic design is 30,869. Therefore, the safety factor is the adopted flood magnitude divided by the hydrological design flood, which results in 41,000 divided by 30,869. Thus, the safety factor comes out to be 1.33, and the safety margin is the difference between the two, that is, the adopted flood magnitude and the hydrologic design flood. Consequently, that comes to be 10,131 cubic meters per second. Hence, we have adopted a safety factor of 1.33 and a safety margin of around 10,000 cubic meters per second in the design. So, the safety factor and safety margin in the design are 1.33 and 10,131 cubic meters per second, respectively.



With this, we come to the end of this lecture. We have carried out flood frequency analysis using the log Pearson type 3 method. We have also seen how to calculate the risk, reliability, and safety factor or safety margin in a particular design. Please give your feedback and feel free to raise any questions or doubts. We shall be happy to answer them on the forum.

Thank you very much.