Course Name: Watershed Hydrology

Professor Name: Prof. Rajendra Singh

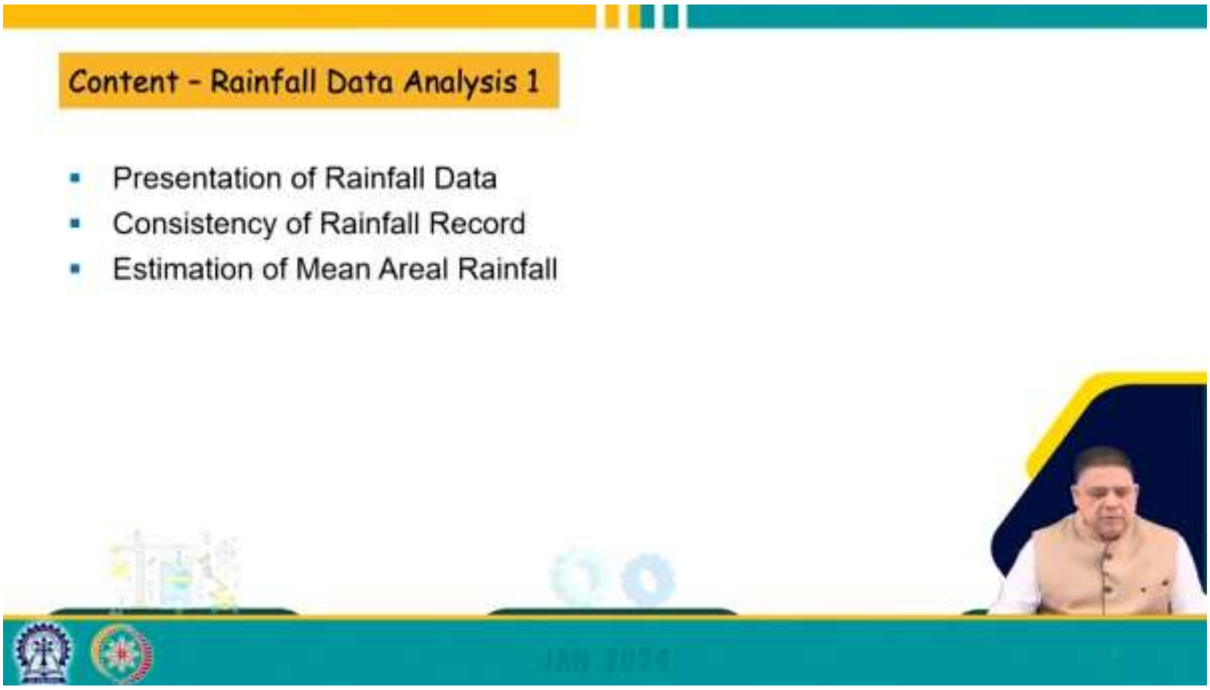
Department Name: Agricultural and Food Engineering

Institute Name: Indian Institute of Technology Kharagpur

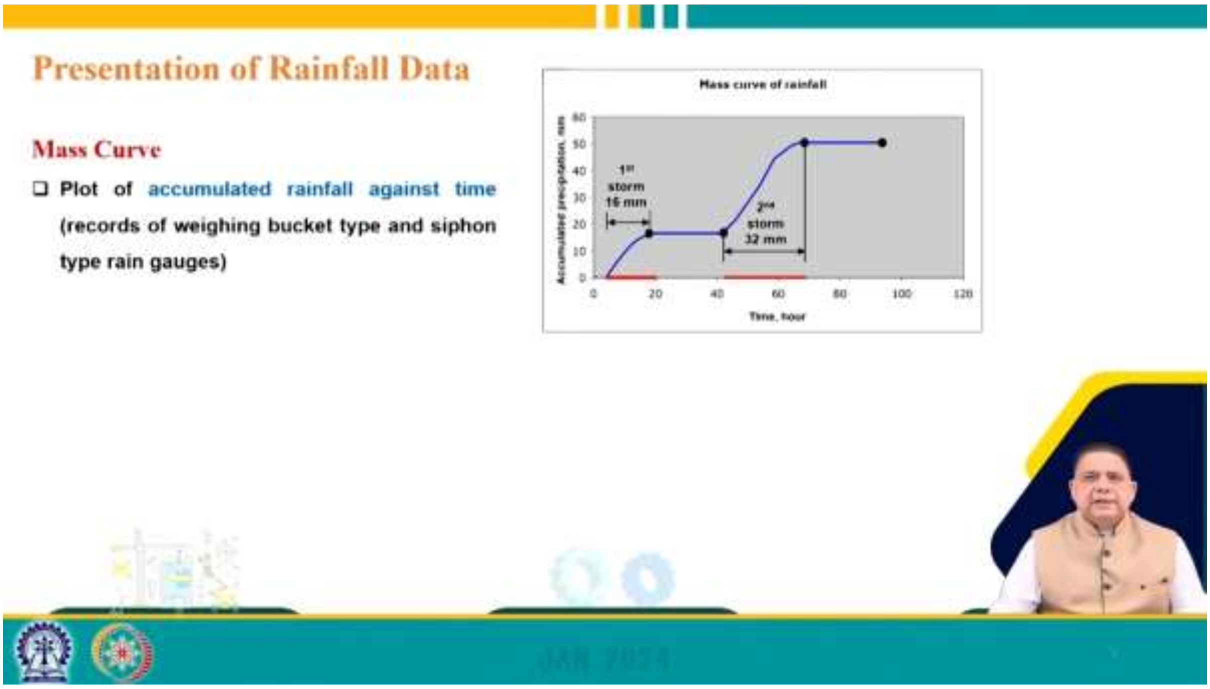
Week: 01

Lecture 03: Rainfall Data Analysis-I

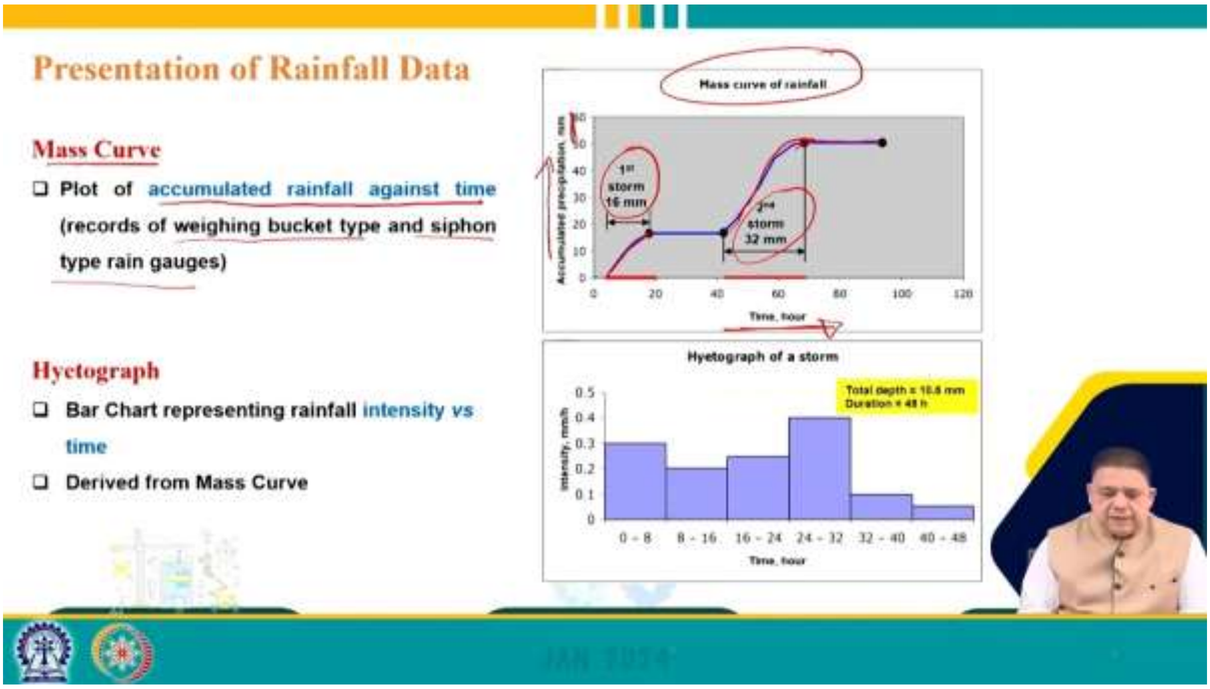
Hello friends, welcome back to this online certification course on Watershed Hydrology. I am Dr. Rajendra Singh, a professor in the Department of Agriculture and Food Engineering at the Indian Institute of Technology, Kharagpur and we are in Module 1. This is Lecture Number 3 and the topic is Rainfall Data Analysis Part 1. In this lecture, we will talk about the presentation of rainfall data, the consistency of rainfall records and the estimation of mean rainfall.



Now, coming to the presentation of rainfall data, if you remember from the previous lecture, we discussed the measurement of rainfall. We mentioned that the typical instrument used is called a rain gauge which comes in two types: recording and non-recording. The non-recording type includes Simon's rain gauge while the recording type includes float or siphon types as well as weighing bucket or tipping bucket rain gauges. These rain gauges along with radar and satellites are used for measuring rainfall. Once rainfall data is measured typically by institutions such as the India Meteorological Department the data must be preserved and presented in a certain form.

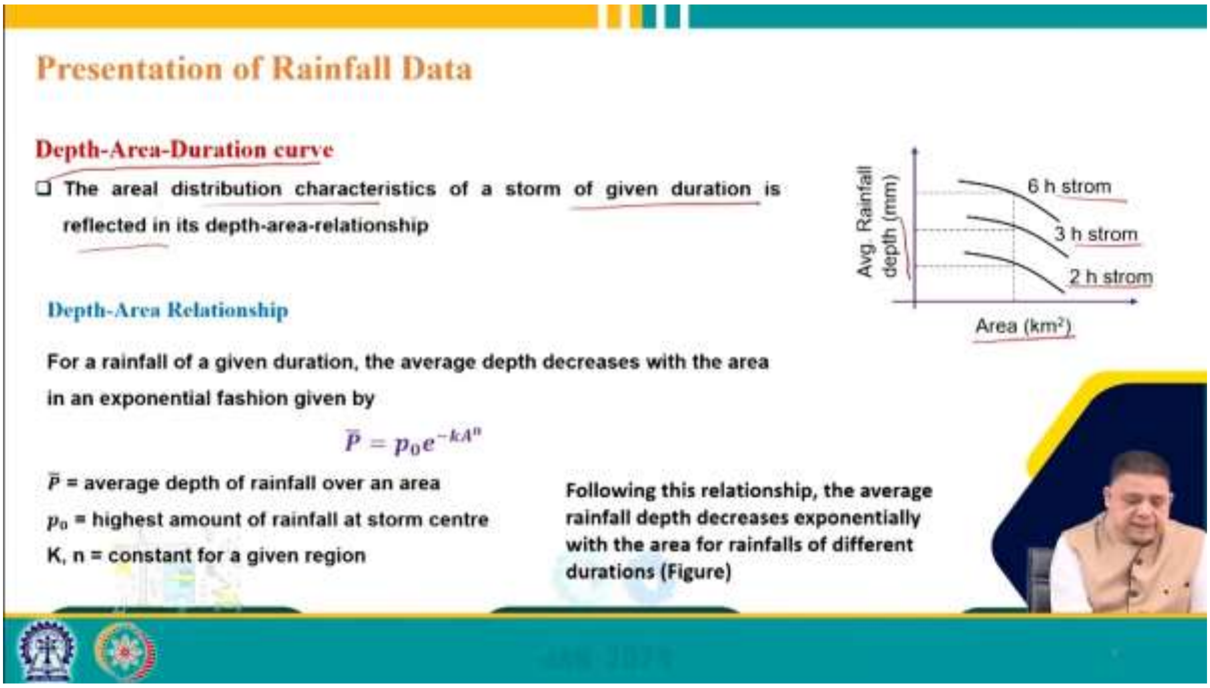


Now, one of the most common ways of preserving or presenting rainfall data is in the form of a mass curve which is basically a plot of accumulated rainfall against time. Here, you can see this mass curve representing rainfall. As you can observe accumulated precipitation in millimeters is plotted against time in hours. This curve essentially indicates cumulative rainfall signifying instances of rainfall events. For instance, the initial spike indicates the first storm which had a rainfall of 16 mm. Subsequently, the flat horizontal line denotes a period with no rainfall occurrence between 20 and 40 hours. Following this dry spell a second rainfall event is depicted resulting in approximately 32 mm of rainfall. Then, once again the curve remains flat indicating no further rainfall. This method allows us to record data either on a daily or weekly basis. Interestingly, it's noteworthy that both the weighing bucket and siphon type rain gauges automatically generate mass curves. This automated process makes them commonly used for preserving rainfall data.



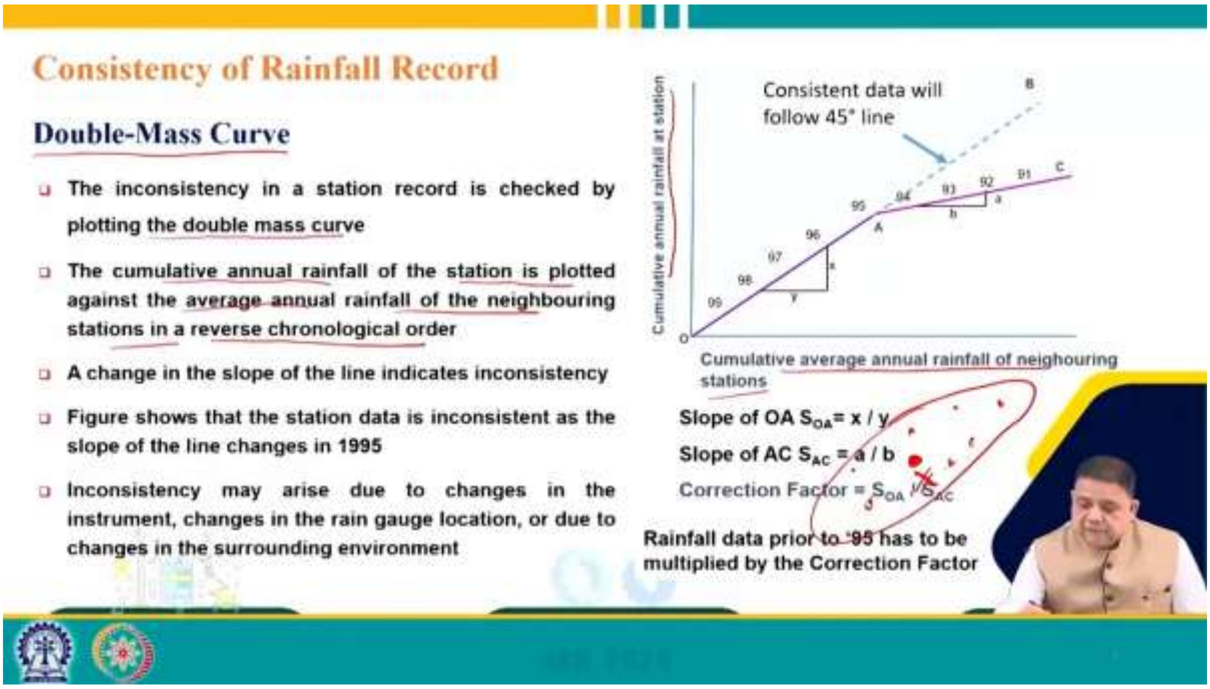
Another method of preserving rainfall data is through hyetographs which are essentially bar charts representing rainfall intensity over time. In this representation rainfall intensities are plotted against time in hours depicted as bars on the chart. Consequently, we need to consider time intervals or slices. In this case 8-hour slices are utilized: 0 to 8 hours, 8 to 16 hours and so forth. Calculating the area under each bar reveals the amount of rainfall within that specific time slice providing a comprehensive view of rainfall distribution. Summing these areas over the entire period yields the total rainfall depth. For instance, over a 48-hour period the total rainfall amounts to 10.6 millimeters.

Hyetographs are typically derived from mass curves which might seem complex at first glance. However, they play a crucial role in various applications including the analysis of infiltration rates. In future lectures, we will delve into the significance of hyetographs and their application alongside discussions on average infiltration rates. Another method of preserving rainfall data is through depth area duration curves which we will explore further in subsequent discussions.



So, basically as you can see here the depth of rainfall versus the average rainfall depth is plotted for different durations of rainfall. These 2-hour storms, 3-hour storms simply indicate the duration of the storm. The aerial distribution characteristics of a storm of a given duration are reflected in its depth-area relationship. As you can see the depth area and storm duration are all shown here. Typically, when we discuss the depth-area relationship for a given duration of rainfall, the average depth decreases with the area in an exponential fashion as given by this formula where $\bar{p}$ represents the average depth of rainfall over the area, p_0 is the highest amount of rainfall at the storm centre and k, h and r are constants with a representing the area. Now, as you can see the average depth decreases with the area exponentially which we can also observe here. The average depth decreases as the area increases showing an exponential decrease.

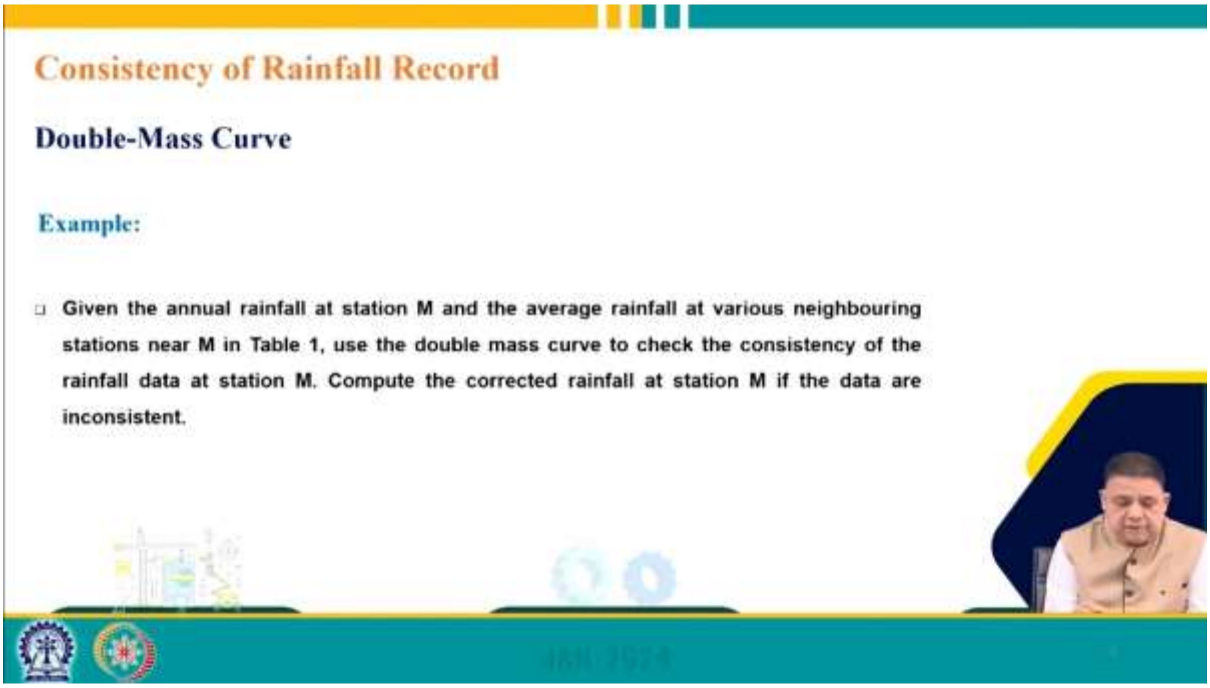
Following this relationship the average rainfall depth decreases exponentially with the area for rainfall of different durations. So, as you can see here for different durations of course when the duration of the rainfall is longer, the magnitude of total rainfall will be higher. However, in terms of depth versus area. It decreases exponentially. This is how the depth-area duration curve preserves the data.



Then, we come to the consistency of rainfall records. Basically, the consistency of rainfall records is analysed using a double mass curve. The consistency in a station's record is checked by plotting the double mass curve. So, whenever we take the data of a particular station and as part of quality analysis, we want to determine whether the data of the station are consistent or not over a period of time over the period of the record. We can do this analysis by plotting the double mass curve. Here, the cumulative annual rainfall of the station is plotted against the average annual rainfall of neighbouring stations in reverse chronological order.

So, as we've observed in a catchment, we can have several rain gauges installed. Typically, we design the gauge network accordingly. Suppose, for example we have a station here and we want to analyse the data from this particular station. Let's call it X. Obviously, to carry out a consistent analysis, we collect data from all neighbouring stations.

To plot the double mass curve, we plot the cumulative annual rainfall of station X against the average annual rainfall of neighbouring stations. Therefore, we collect data from neighbouring stations to determine their average annual rainfall and then plot the cumulative average annual rainfall of these neighbouring stations against the cumulative annual rainfall of station X. This can be observed in the plot displayed.



If the data are consistent, it is expected that they will follow a straight-line relationship akin to a 45-degree line. However, a change in slope of the line indicates inconsistency. If the slope of the line changes, it suggests that the data are not consistent.

Examining the figure it's apparent that the station data is inconsistent. There's a noticeable change in the slope of the line specifically in 1995. This indicates the year in which the inconsistency occurred. Therefore data prior to 1995 and post-1995 follow different trends or patterns.

Today, we're going to discuss why we plot data in reverse chronological order. Essentially, the reason behind this approach is quite significant. We opt for reverse chronological order wherein the latest data is plotted first and the oldest data comes last simply because it's anticipated that we will continue to utilize the same instruments, locations and observe consistent data trends in the future.

In a double-mask curve when we correct data, our objective is to align inconsistent data with the current trend. Therefore, we plot the data in reverse chronological order and then analyse whether the data points are consistent or not. Inconsistencies in rainfall records can stem from various factors such as changes in instruments alterations in the rain gauge's location or shifts in the surrounding environment.

For instance, let's consider a specific station where the rain gauge was initially functioning on one wall but later the instrument itself was replaced. Consequently, the recording pattern might have been altered due to the change in instruments. Similarly, in certain regions a slight relocation of the rain gauge can occur. This relocation can result in changes in the recorded rainfall patterns due to the shift in the gauge's location and its surrounding environment.

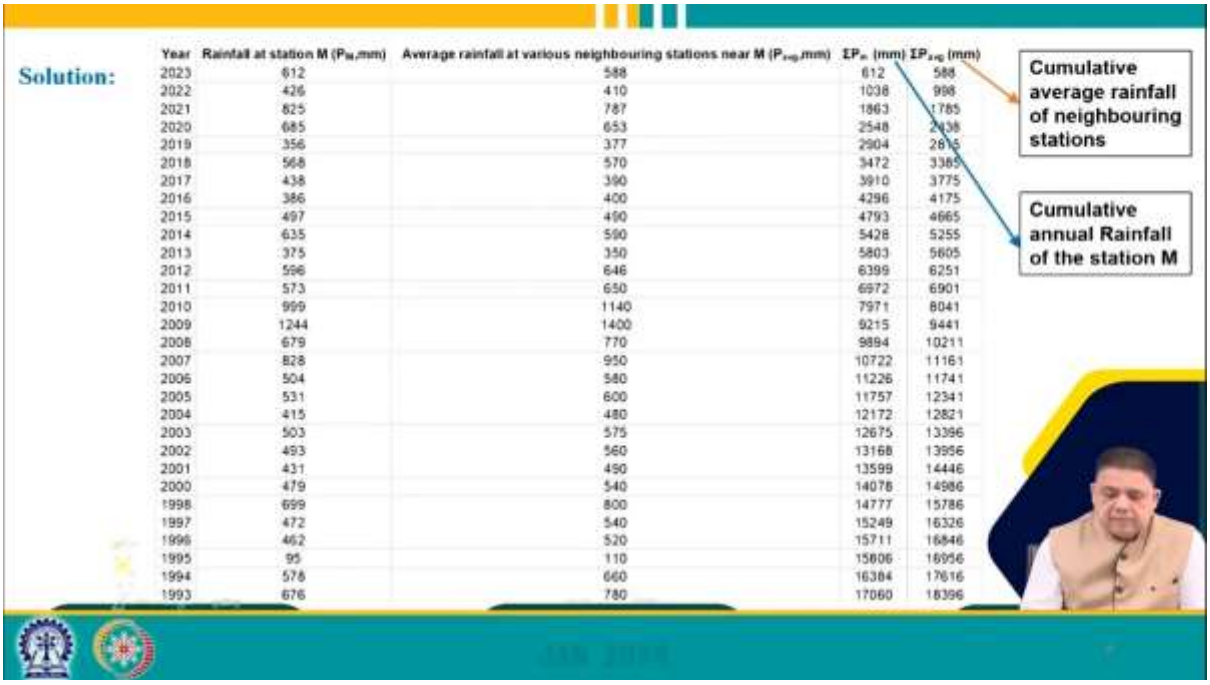
Year	Rainfall at station M (P_U, mm)	Average rainfall at various neighbouring stations near M (P_{av}, mm)
2023	612	588
2022	426	410
2021	825	787
2020	685	653
2019	356	377
2018	568	570
2017	438	390
2016	386	400
2015	497	490
2014	635	590
2013	375	350
2012	596	646
2011	573	650
2010	999	1140
2009	1244	1400
2008	679	770
2007	828	950
2006	504	580
2005	531	600
2004	415	480
2003	503	575
2002	493	560
2001	431	490
2000	479	540
1998	699	800
1997	472	540
1996	462	520
1995	95	110
1994	578	660
1993	676	780

Moreover, changes in the surrounding environment can also contribute to inconsistencies. For example, if a location with a rain gauge suddenly sees the construction of a tall building nearby or the growth of a new tree, it can affect the rainfall trend observed at that particular location. So, understanding these factors is crucial when analysing rainfall data as it helps us discern and account for any inconsistencies that may arise over time.

So, due to changes in surrounding conditions, the data may become inconsistent. Essentially to correct the data we analyse the slope of the line OA which represents x divided by y. Similarly, we examine the slope of line AC which is A divided by B. Then we determine the correction factor SOA divided by SOV using the slopes of the two lines. This correction factor is crucial for adjusting the data. Rainfall data prior to 1995 needs to be multiplied by this correction factor. I believe clarity will emerge once we delve into an example. Let's consider one now.

We'll analyse the annual rainfall at station M and the average rainfall at neighbouring stations near M as presented in Table 1. We'll employ the double mass curve to assess the consistency of rainfall data at station M and calculate the corrected rainfall if inconsistencies exist.

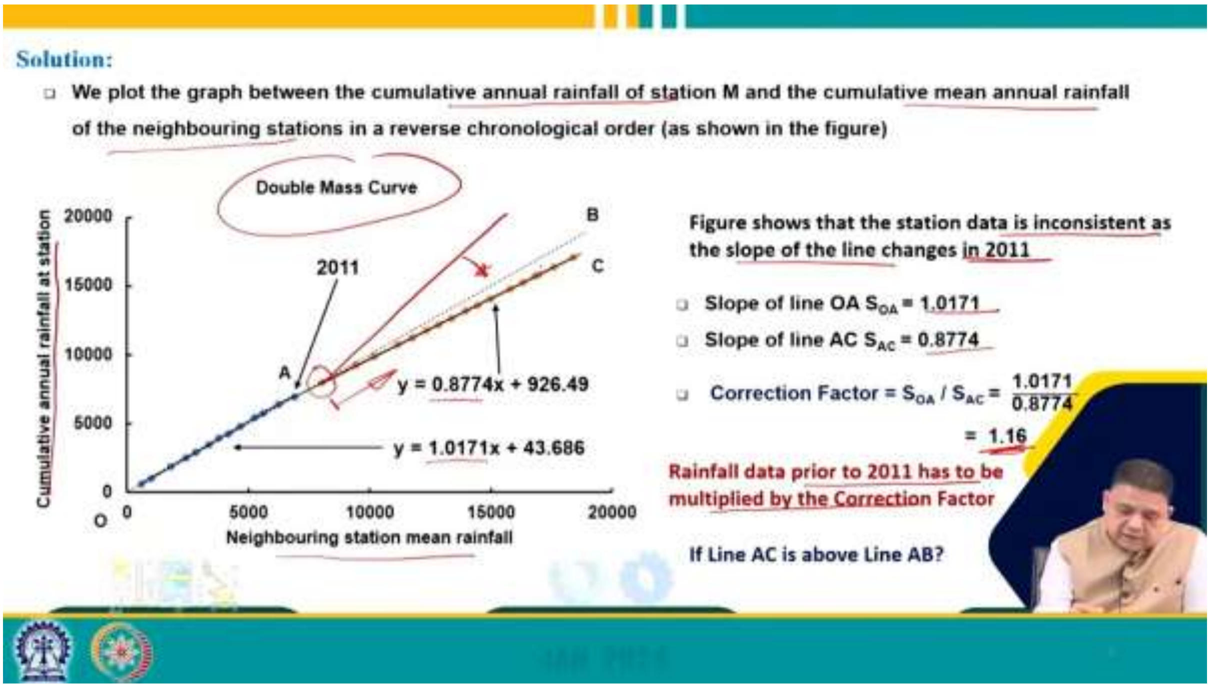
Here's the table displaying rainfall data at station M (in millimeters) and the average rainfall at various neighbouring stations near M (also in millimeters). The data spans from 1993 to 2023. Notably, the data is already arranged in reverse chronological order as required for the double mass curve analysis. If the data were not in this order, we would have to rearrange it accordingly.



Year	Rainfall at station M (P_M,mm)	Average rainfall at various neighbouring stations near M (P_{avg},mm)	ΣP_m (mm)	ΣP_{avg} (mm)
2023	612	588	612	588
2022	426	410	1038	998
2021	825	787	1863	1785
2020	685	653	2548	2438
2019	356	377	2904	2815
2018	568	570	3472	3385
2017	438	390	3910	3775
2016	386	400	4296	4175
2015	497	490	4793	4665
2014	635	590	5428	5255
2013	375	350	5803	5605
2012	596	646	6399	6251
2011	573	650	6972	6901
2010	999	1140	7971	8041
2009	1244	1400	9215	9441
2008	679	770	9894	10211
2007	828	950	10722	11161
2006	504	580	11226	11741
2005	531	600	11757	12341
2004	415	480	12172	12821
2003	503	575	12675	13396
2002	493	560	13168	13956
2001	431	490	13599	14446
2000	479	540	14078	14986
1998	699	800	14777	15786
1997	472	540	15249	16326
1996	462	520	15711	16846
1995	95	110	15806	16956
1994	578	660	16384	17616
1993	676	780	17060	18396

Now, as we observed in the double mass curve, we need to plot the cumulative annual rainfall of the station against the cumulative rainfall of neighbouring stations. So, the first thing we have to do is to obtain the cumulative rainfall at station M. It is essentially the cumulative data of this column. The first value of course remains 612. The second value is the sum of these two values, which is 612 plus 426 resulting in 1038 and so forth. Similarly, the P sum P average millimeters which is the cumulative average rainfall of neighbouring stations is derived from summing up the cumulative values of this average rainfall value.

The first value here is 588. The second value will be the sum of these two which is 998. The third value of course will again be a let 787 to 817, 85 and so on. Thus, cumulative values are calculated and then of course we plot the graph between the cumulative annual rainfall of station M and the cumulative mean rainfall value of neighbouring stations in reverse chronological order. This is how we generate the graph. Therefore, if we plot the graph between the cumulative annual rainfall of station M and the neighbouring station's mean rainfall then this is the graph we obtain referred to as the double mass curve.



Now, we have to analyse whether the data are consistent or not. If the data are consistent, then all the data will follow the 45-degree line indicating no change in the slope of the line anywhere. However, as you can see here, there is a point where the slope of the line changes. This implies that between O and A and A and C, the slope of the line is not the same; it changes.

So, from the year 2000 to 2011, specifically in 2011, there was a significant change observed in the slope of the line, indicating inconsistency in station data. Consequently, any data collected prior to 2011 needs correction. To achieve this correction, we must determine a correction factor which we've already identified as the ratio of the slopes of two lines.

To calculate this, we've employed a method using Excel to fit straight lines resulting in the equations displayed here. For the segment OA, the equation yields a slope of 1.0171. Similarly, for the segment AC, the slope is 0.8774. Therefore, the slope of line SAC is 0.8774. The correction factor is then the ratio of these slopes resulting in 1.16.

Hence, all data preceding 2011 must be multiplied by this correction factor to align with the current recording trend at this station. This instruction is indicated by 'rainfall data prior to 2000 has to be multiplied by the correction factor.

Solution:

Year	P_M (mm)	P_{avg} (mm)	ΣP_M (mm)	ΣP_{avg} (mm)	Final values of P_M (mm)
2023	612	588	612	588	612
2022	426	410	1038	998	426
2021	825	787	1863	1785	825
2020	685	653	2548	2438	685
2019	356	377	2904	2815	356
2018	568	570	3472	3385	568
2017	438	390	3910	3775	438
2016	386	400	4296	4175	386
2015	497	490	4793	4665	497
2014	635	590	5428	5255	635
2013	375	350	5803	5605	375
2012	596	646	6399	6251	596
2011	573	650	6972	6901	573
2010	999	1140	7971	8041	1158.84
2009	1244	1400	9215	9441	1443.04
2008	679	770	9894	10211	787.64
2007	828	950	10722	11161	960.48
2006	504	580	11226	11741	584.64
2005	531	600	11757	12341	615.96
2004	415	480	12172	12821	481.4
2003	503	575	12675	13396	583.48
2002	493	560	13168	13956	571.88
2001	431	490	13599	14446	499.96
2000	479	540	14078	14986	555.64
1998	699	800	14777	15786	810.84
1997	472	540	15249	16326	547.52
1996	462	520	15711	16846	535.92
1995	95	110	15806	16956	110.2
1994	578	660	16384	17616	670.48
1993	676	780	17060	18396	784.16

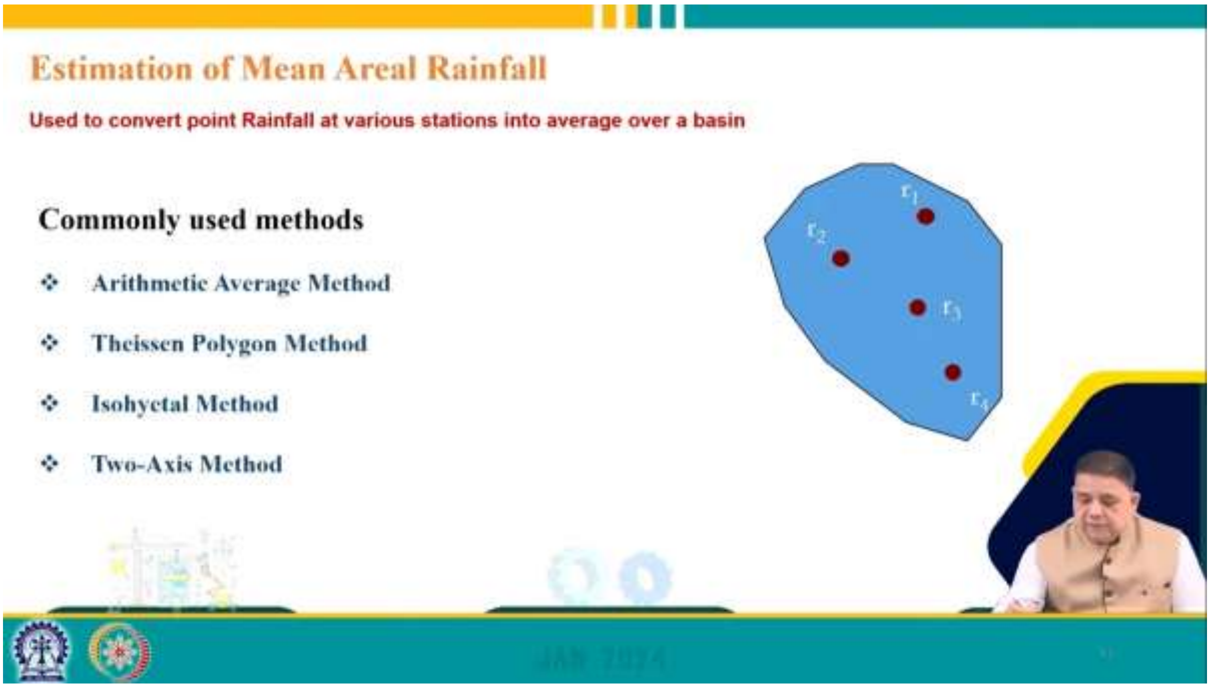
- Rainfall data prior to 2011 has to be multiplied by the Correction Factor of 1.16
- In the table, for 2010, P_M = 999 mm
- Thus, the corrected value of P_M for 2010 = 999 * 1.16
  = 1158.84 mm

It's worth noting that while the figures I've presented illustrate a typical case. There are instances where the line might deviate. In such cases, we still need to calculate the slope and correction factor to adjust the data to align with line AB.

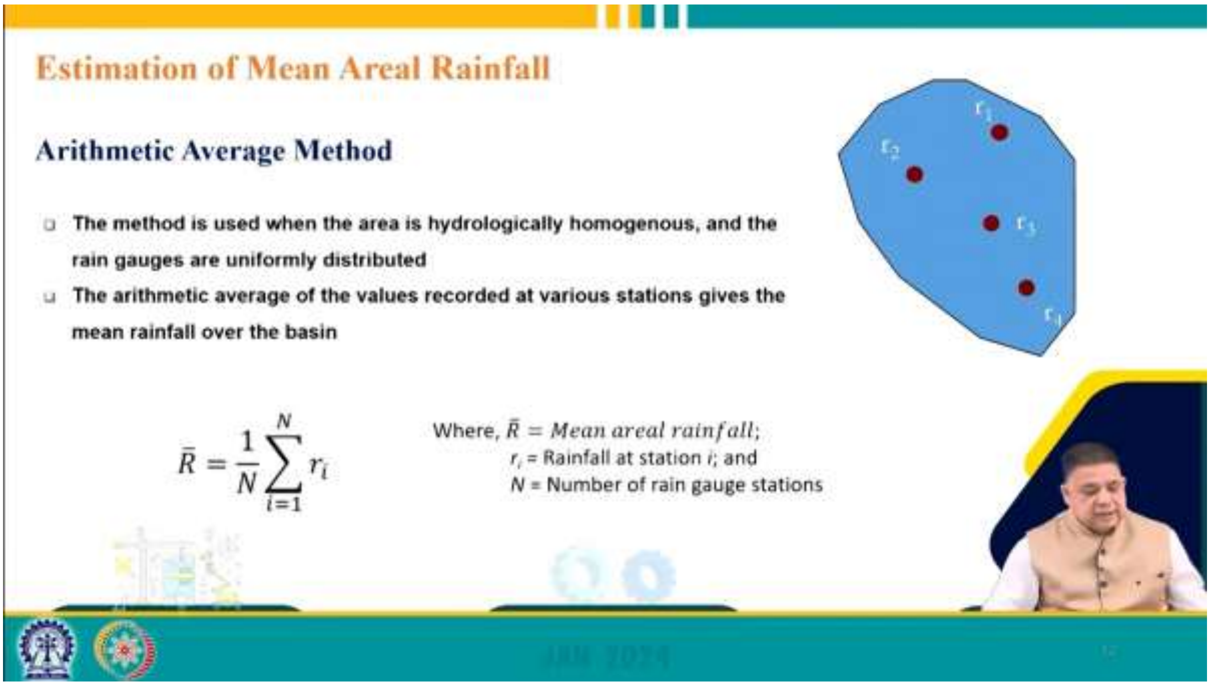
So, this may not always be the case; it won't always fall below the 45-degree line. Therefore, for the final values if you observe the PM, it means that for the station prior to 2011, we will need to obtain fresh values. Prior to 2011, the data is multiplied by a factor of 1.16. In this table the PM value in 2010 is 998.

This is the 9th, so it needs to be multiplied resulting in 1158.84. In fact, all data prior to 1993 needs to be multiplied by 1.16 in order to obtain corrected data. This is the consistent data we will be using for further analysis.

Now, let's discuss the estimation of mean rainfall which is essentially the rainfall measured by rain gauges in a catchment. We have multiple rain gauges and the recorded rainfall at these stations is referred to as point rainfall. So, the data we obtain from a specific station is known as point rainfall. However, to derive a representative value for the entire catchment, we need to convert the point rainfall from various stations into an average over the basin.



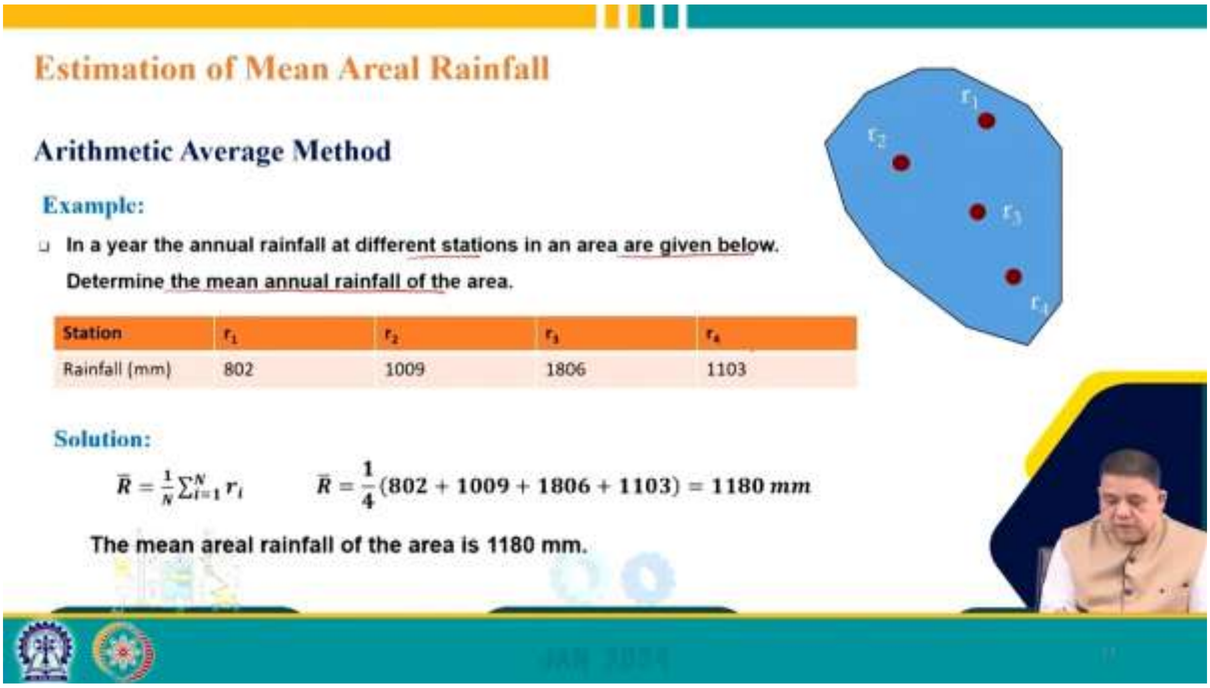
The commonly used methods for this purpose include the arithmetic average method, Thiessen polygon method, isohyetal method and two-axis method. Let's delve into each of these methods one by one.



Let's start with the arithmetic average method. As the name suggests, it's the simplest method where we calculate the arithmetic average of all the recorded points. However, there's a condition: this method is applicable when the area is hydrologically homogeneous and the rain gauges are uniformly distributed over the catchment.

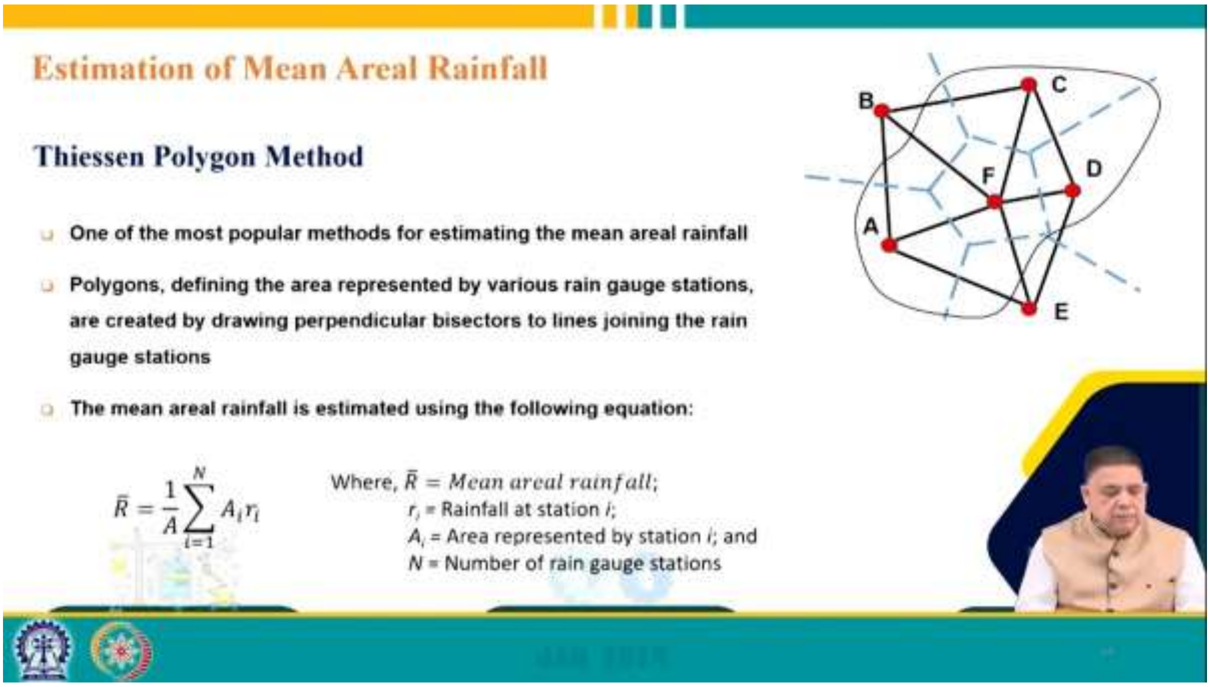
The first condition is that the hydrological behaviour of the catchment should be uniform throughout. So, there are ways to determine whether the area is hydrologically homogeneous. Another condition is that rainfall should be uniformly distributed over the catchment to represent the data from all parts of the catchment. If these conditions are met, then simply taking the arithmetic average of the values recorded at various stations gives the mean rainfall of the basin.

The mean rainfall ($\bar{r}$) is calculated as the sum of all the station values divided by the number of rain gauge stations. In essence, in the arithmetic average method each station is given equal weight which is n where n is the number of stations. So, in this method each station contributes equally to the final average.



Let's illustrate this with an example: In a year, the annual rainfall at different stations in an area is given below. We need to determine the mean annual rainfall of the area.

So, there are 4 stations: r_1, r_2, r_3 and r_4 and the recorded rainfalls in the year are 800, 2009, 1806 and 1103 respectively. Obviously, we presume or assume that the area is hydrologically homogeneous and the rain gauges are uniformly distributed. In that case, we can simply take an arithmetic average of the values. So, the sum of the values divided by the number of rain gauge stations will give us the average value which is 1180 millimeters. Thus, the mean annual rainfall using the arithmetic average method comes out to be 1180 mm.



Now, the next method is the Thiessen polygon method which is one of the most popular methods for estimating the mean aerial rainfall. In fact, most computer software also uses this method. Here, polygons defining the area represented by various rain gauge stations are created by drawing perpendicular bisectors to the lines joining the rain gauge stations. We will see the detailed procedure in the next slide.

Then, the mean annual rainfall is estimated using the following equation:

$\sum_1^n i = a_i r_i$

Where r_i is the rainfall at station a_i is the area represented by station i and n is the number of rain gauge stations. In this case, a_i divided by the sum of all a_i's gives the weighing factor for the particular station.

Today, we'll delve into how weights are assigned based on the area of the polygon represented within a particular catchment. First, let's outline the procedure. We start by drawing the catchment area to scale and marking the rain gauge stations on it. Here, we have six rain gauge stations labelled from 'a' to 'f'. It's worth noting that some stations may lie outside the catchment boundary.

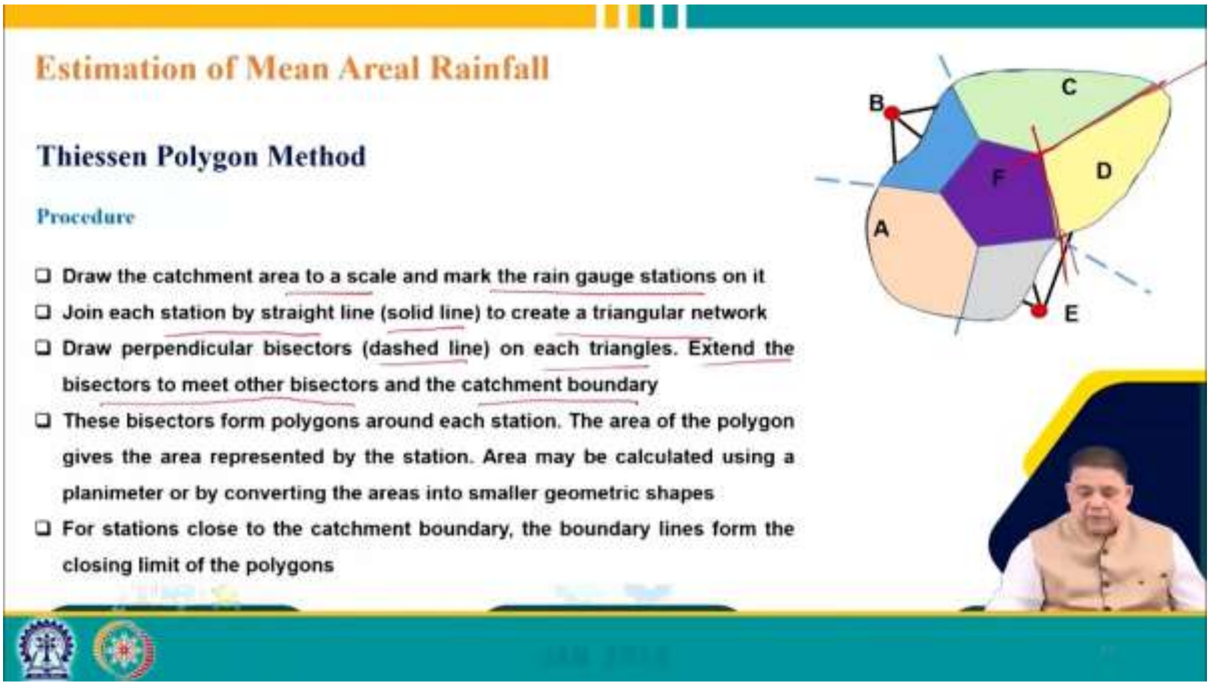
Next, we connect each station with straight lines forming a triangular network. This network is depicted by solid lines on the diagram.

Moving forward, we draw perpendicular bisectors within each triangle represented by blue dashed lines. These bisectors are extended to intersect with each other and with the catchment boundary. This process is repeated for each triangle within the network.

The resulting bisectors enclose polygons around each station, defining the area represented by that station. For instance, the polygon around station 'a' represents its area within the catchment.

To calculate the area of each polygon one can use a planimeter or convert the area into a smaller unit. Alternatively, the entire network can be plotted on a graph sheet to obtain representative values.

For stations located near the catchment boundary, the boundary lines serve as the closing limit of the polygon. This aspect is crucial, as demonstrated by station 'e' which lies outside the catchment boundary.



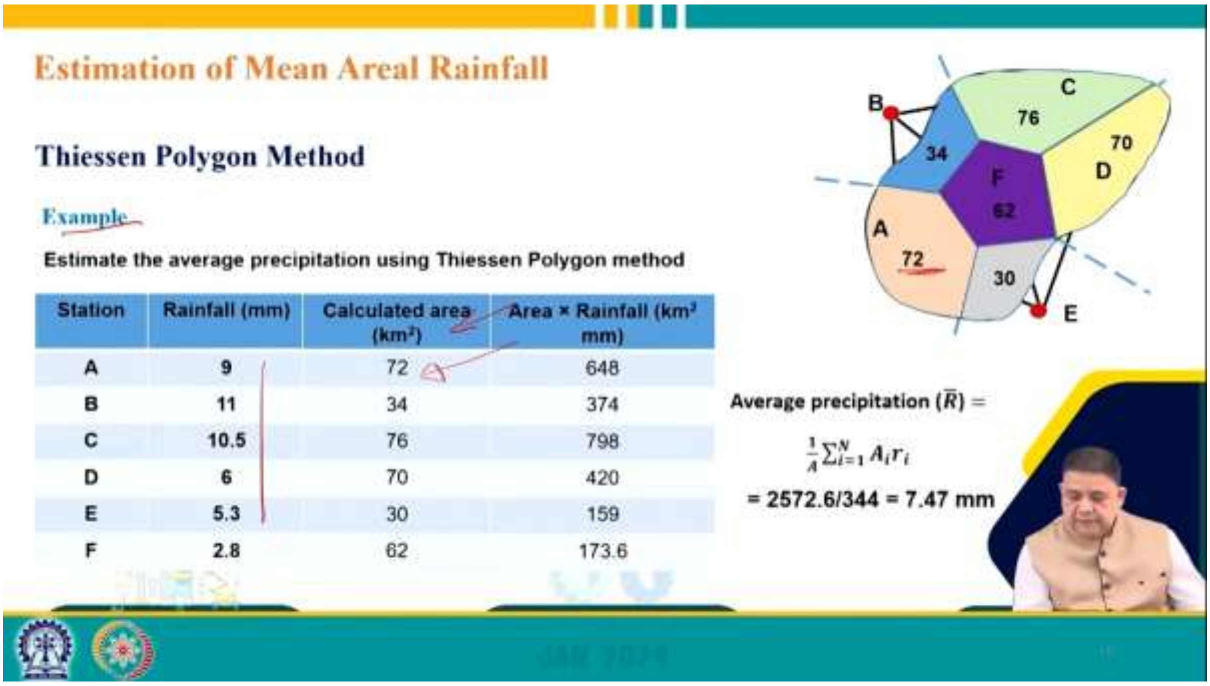
In summary, this method allows us to assign weights based on the area represented by each station within the catchment, facilitating accurate analysis and prediction of rainfall patterns.

That means the weights are assigned based on the area of the polygon represented within the particular catchment. Now, coming to the procedure, what we do is draw the catchment area to scale and mark the rain gauge stations on it. So, you can see here we have 6 rain gauge stations from 'a' to 'f' and some stations may be marked outside the catchment boundary. Now, what we do is join each station by a straight line (shown here as a solid line) to create a triangular network. So, as you can see, a triangular network is created.

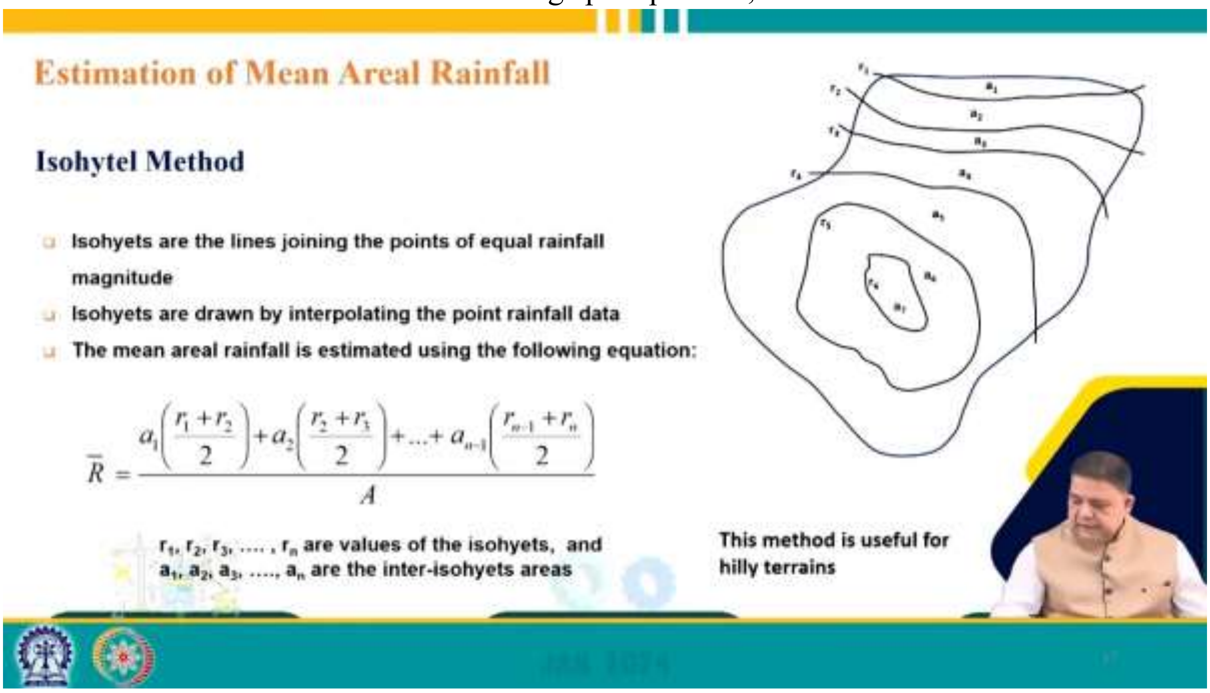
Next, what we do is draw perpendicular bisectors (shown here in blue dashed lines) on each triangle and extend the bisectors to meet other bisectors and the catchment boundary. So, we draw these perpendicular bisectors like this and also extend them so that they intersect each other and the catchment boundary, here and here and so on. That is the procedure followed in each case. Now, these bisectors form a polygon around each station. So, as you can see for station 'a' here this is the polygon.

So, that is the area being represented by Station A. These bisectors form polygons around each station. The area of each polygon gives the area represented by the station and it may be calculated using a planimeter or by converting the area into small units. Alternatively, you can use a graph sheet to plot this entire thing and obtain the representative value. For stations close to the catchment boundary, the boundary lines form the closing limit of the polygon. This is important because, for Station E in this case, although it is outside the catchment boundary, it still holds significance.

So, basically, this is the area. The catchment boundary forms the limit for this area. The area within the polygon, but bounded by the catchment boundary on one side represents Station E. Let's take an example to estimate the average precipitation using the Thiessen polygon method. These are the stations with recorded rainfall. As discussed, we will create the Thiessen polygon network, extend it, find the representative area of different stations, and then determine the catchment area.



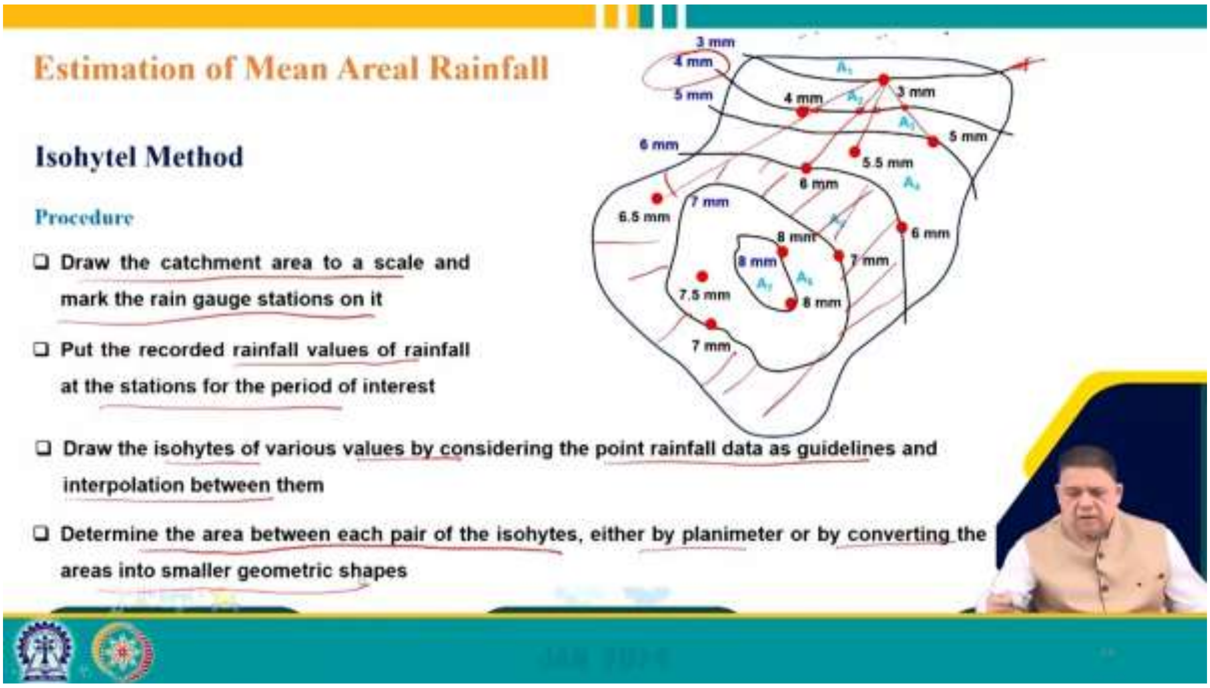
So, for example, let's consider station A, where the area represented by the station is 72 square kilometers. Then, of course we multiply the area by the catchment rainfall to get the total rainfall for that area. To calculate the average precipitation, we know that it's the summation of



the product of area and rainfall divided by the total area. So, when we sum these values, we get 2572.6 and when we sum the areas, we get a value of 344 square kilometers.

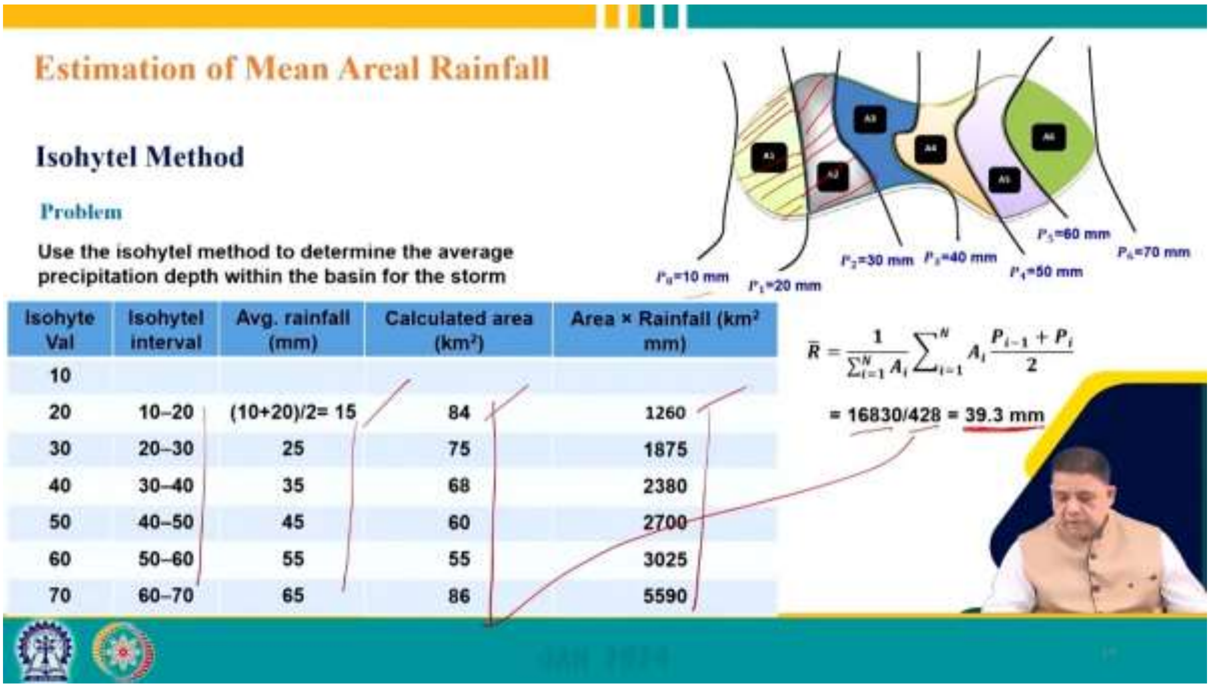
Thus, the average precipitation comes out to be 7.47 millimeters using this method. Now, let's move on to the last method which is the isohyetal method - the third method. in fact, isohyets are the lines joining points of equal rainfall magnitudes and they are drawn by interpolating point rainfall data. So, essentially isohyets represent areas of consistent rainfall magnitude.

The mean annual rainfall is estimated using the following equation: the area included between two isohyets is determined and the average of the two stations within that area is taken into account. The sum of these values divided by the total area provides us with the average rainfall. This method is particularly useful in hilly terrains.



Let's consider an example here. The first step is to draw the catchment area to scale and mark the rain gauge stations on it. This is precisely what we have done here; we've marked the catchment area and the rain gauge stations, indicating the recorded rainfall values. Record the rainfall values at each station for the period of interest. In this case, we are dealing with daily rainfall data. Next, draw the isohyets of various values by utilizing the point rainfall data and interpolation.

For instance, to draw the 3 mm isohyet we require surrounding data points. If, for example, we aim to depict 4 mm we already have the recorded value. However, to establish where 4 mm lies between 3 and 5 we employ interpolation. Similarly, we interpolate to determine the position of 4 mm between 3 and 5.5 and between 3 and 6. By utilizing data points such as 3 and 6.5, we can accurately interpolate the location of 4 mm. Through this interpolation process, we obtain sufficient points to delineate the isohyet of a known value.



So, that is what we have done here. Similarly, we will draw 5, 6, 7, 8 and so on. That is how, for the given area, we will draw the isohyetes of different values. Then, we determine the area between each pair of isohyetes. That means between 6 and 7 for example we will find out the area between 6 and 7 and so on. This can be done either by planimeter by converting the areas into smaller geometrical shapes or even by using a large-sized graph sheet.

Now, let's take an example or problem on the isohyetal method. We will use the isohyetal method to determine the average precipitation depth within the basin for the given storm for which the isohyetes are already provided. The isohyet interval is also given. The average rainfall values between the isohyetes such as between 10 and 20, 15, 20 and 30 will be 25 and so on. Then, we have to find out the catchment area. For example, for A_2 which lies between 20 and 30, we need to find the area within the basin.

A_1, A_2 and so on, will measure this area. Then, we multiply the area by the average rainfall to obtain the sum of the values. The sum of these values is 16,800 and the total area of the catchment is provided. Dividing the sum by the total area, we get the average aerial rainfall of the basin which comes out to be 15,800 mm.

With this, we close today's lecture. Thank you very much for listening, and please feel free to give your feedback and raise questions on the forum so that we can address them. Thank you very much.