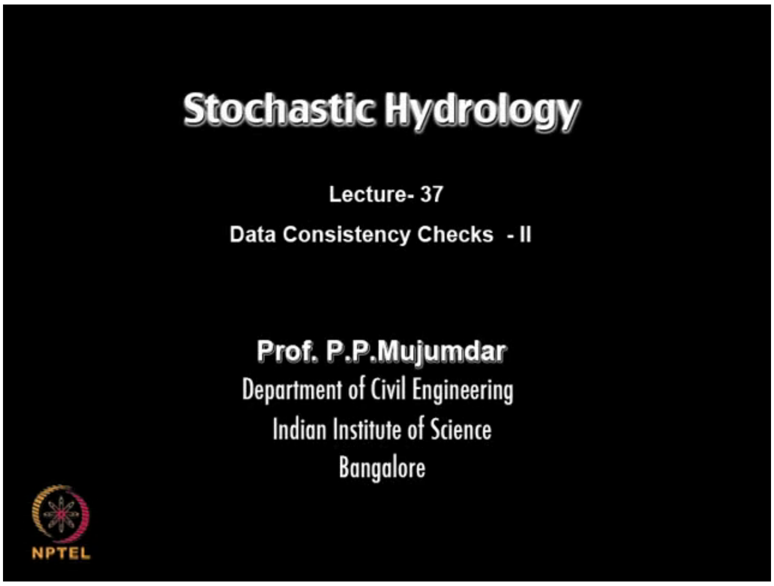
Stochastic Hydrology

Prof. P.P.Majumder
Department of Civil Engineering
Indian Institute of Science, Bangalore

Module No # 08
Lecture No # 37
Data Consistency Checks – II

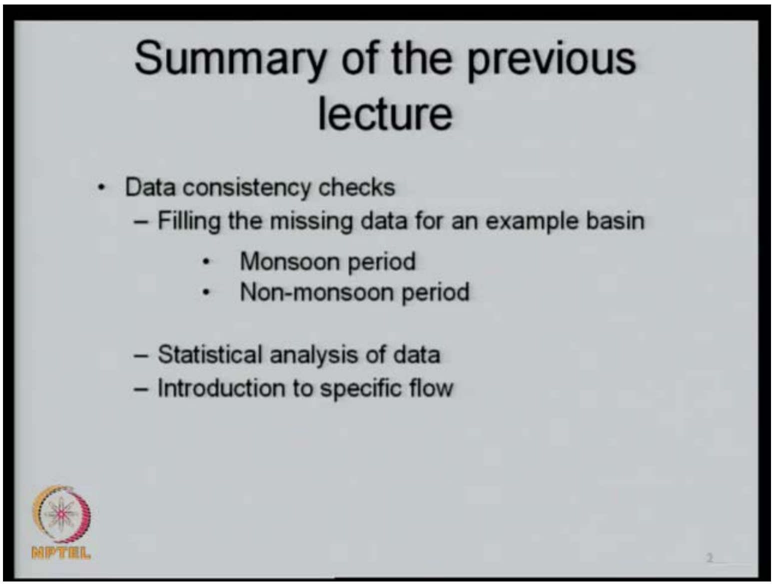
(Refer Slide Time: 00:11)



Good morning and welcome to this lecture number 37 of the course stochastic hydrology. Now, we are discussing some practical aspects of the data analysis. All the techniques that we have studied earlier deal with the observed data. When we have observed data, let say of the stream gauges, we need to do certain checks to make sure that the data we are using for the analysis is, in fact, good quality and it is consistent, so that, we can use this data in our analysis to make some decisions. In fact, why we do all this analysis is, finally, to make decisions on the hydrologic designs. Let say of the reservoirs or reservoir operating policies or even hydrologic designs, such as the capacity of the bears that we want to provide, the culverts that we would like to provide at certain locations and so on.

So, the analysis is essentially done to make sure that, finally, the designs that we make, the engineering designs that we make are representative of the data that we have observed. Therefore, in the last class, I introduced what are called as the consistency checks for the data. That means, once we have observed data at several locations, stream gauges and so on, we need to check whether the data series that we have for analysis is, in fact, consistent with data series that is available elsewhere at other gauges or in the same catchment or in the surrounding catchments and so on. So, essentially we look at the homogeneity as well as the consistency of the data.

(Refer Slide Time: 02:18)



In the last lecture, we examined, typically the methods for data filling. Let say that, you have observed data at a particular location for fifty years and something. In these fifty years, sequence of data in the fifty years series of data, there may be many values missing for several reasons. May be, it has been not observed or it has been faded in the sense that, they would have recorded on paper and then, for some reasons it has faded and you are unable to make out exactly what is the data or the gauge may not be working during those periods and therefore, no data is available and so on.

Typically, whenever we have hydrologic time series, especially in the developing countries like ours, we may not have continuous records. There may be certain values that are missing and it is important for our analysis to have a continuous time series of data. Therefore, we adopt the procedures that I discussed for missing of the data, the filling of the missing data.

If you recall, we look on the dates on which the data is missing and then simply fill it by the mean of the flow. If we are talking about the flow data, mean of the flow for that particular date, using all the remaining years of the data in which data is available for that particular date, calculate the mean of that and simply fill it. This is the simplest way and I have also introduced some other more rigorous ways of filling the data in the last lecture.

We typically used two different periods. For example, monsoon period, we adopt certain period and non monsoon period, where the flows are very small. In the peninsular region, where flows are very small in the non monsoon period, essentially due to the flows, are essentially due to the base flow during the non monsoon period. So, we adopt different procedures for the monsoon periods as well as non monsoon periods for filling of the data.

Once we have filled the data and now, we have a continuous time series available at a particular gauge. In fact, at all the gauges that we are interested in that particular catchment and then, we start the consistency checks. In the consistency checks, we ask the question, yes, we have the time series and how representative is this time series with respect to the time series that are observed in the other locations and also the same catchment. So, that is where we do the consistence test.

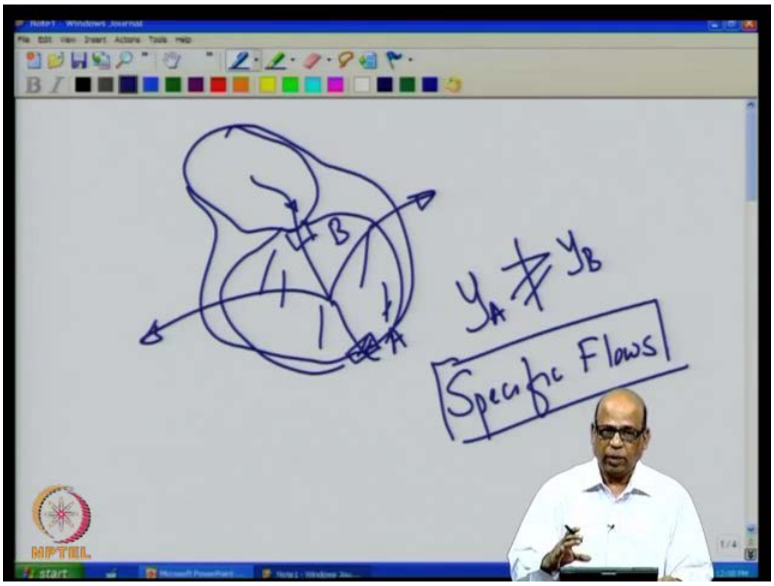
Like I have been mentioning all through the codes, the first thing we do is, whenever we have data, plot the time series first of the data and then, do the statistical analysis. Simply look at the mean, standard deviation, coefficient of variation and then, look at the skewness coefficient and so on. Simply get the statistics of the data. The statistics of the data is telling you a story of what has happened at that particular location. That is, how variable the flows are, what is the maximum flow that you can expect, what is the minimum flow that you can expect, what is the mean flow for long period of time and so on.

So this, we have to keep this in mind when we are doing the analysis using the stochastic hydrology methods. We have to keep in mind that, we are relying in all our decisions on the data that has been observed and the data that has been observed at a particular location. In fact, tells us what has happened at that particular location. It tells us the story of the processes that have happened there.

Therefore, we must respect the data, observed data and at the same time, we must make allowance for the errors in the data collection, errors in the data itself, the non reliability of the data, and the bad quality of the data. We have to make allowance for all of these. We must have some rigorous methods of making allowance for all of this and then use the data.

This is an important aspect and therefore, I keep on stressing this. Because, in practical applications, simply we pick up some data that is provided by the agencies or custodians of the data and then, start analyzing. The first thing is, you suspect the data whether the data is of good quality, and the data is reliable and so on. Therefore, it is important for us to make rigorous checks on the data that we have obtained with respect to the data is available elsewhere and then may proceed with the analysis.

(Refer Slide Time: 07:23)



The consistency checks that I introduced in the last lecture or typically looking at a particular gauge, let say that, we have two gauges. This is a stream flow and then, there are two gauges. Let say, gauge A and gauge B and the flow direction is like this.

So, if you want to check the consistency of this, remember the flows that are recorded at the site B are the flows that are contributed by the upstream catchment up to this point.

The flows that are recorded at gauge A, includes the flows that are recorded at gauge B and therefore, it will have this particular catchment. So, the flows at A must be greater than or equal to the flows at B. Typically, there will be this flow plus all the intermediate catchment flow that has not been recorded by the flows at B. So, typically this is intermediate catchment flow that comes in. Therefore, if you plot A, let say a double masker, in the sense that, flows at this location with respect to the total aggregated flows that have been recorded at B, you must expect a straight line. Typically, the flows at A will be the cumulative flows at A, as time progresses, and will be greater than the cumulative flows at B. So, this is of one analysis that we do. I will demonstrate this with a particular case study.

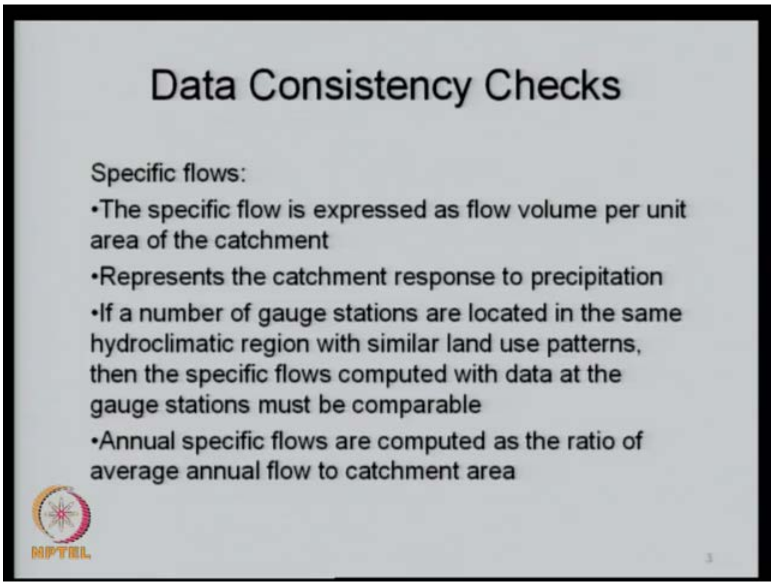
Now, if that does not happen, what does that mean? Let say that, the flows at A, I represent it as y a, should be greater than or equal to let say y b, if I represent y as the flow. If it does not happen, that means, if y a is not greater than equal to y b, what does that mean? It either means that, there is something wrong with the data that we have observed at A or it may mean that, you are using a lot of water between B and A. This is going either as loses or for utilization. Let say that you have a major lift irrigation scheme between B and A and then, you are continuously using the water at that point and then the intermediate catchment is not contributing enough to account for the utilization. Then, it is possible that the flows at A are not, in fact, equal to, in fact, greater than or equal to flows at B.

If this is not the case or if you are not able to account for why the flows, the fact that over the flows at A are not greater than equal to the flow at B, it means that there is something wrong with either our recording or our way of measurement itself. Therefore, you must suspect the inconsistency in the data.

The other part that I mentioned in the last class is, what are called as specific flows. Now, this is a concept that we often use in India for checking the consistency when we talking about large water resource systems. Let say, you are talking about the catchment of the size of Narmada river basin or Mahanadi river basin. It has lakhs of square kilometers in the area. In which, you will have sub catchments which may all be considered as hydrologically homogeneous regions in as much as the way they respond to rainfall in terms of runoff generation. These characteristics are the same across the sub basins and therefore, these sub basins are treated as hydrologically homogeneous. If a catchment is hydrologically homogeneous, it means that the specific flow or the flow per unit area that you get for a particular rainfall, must be the same across different case of catchment.

If let say some catchment of A and catchment of B are hydrologically homogeneous, it means that, the flow generation characteristics of the catchment are the same for both these catchments. Therefore, the specific flow which is the flow per unit area, must be the same for both these catchments. That is what we check in the consistency.

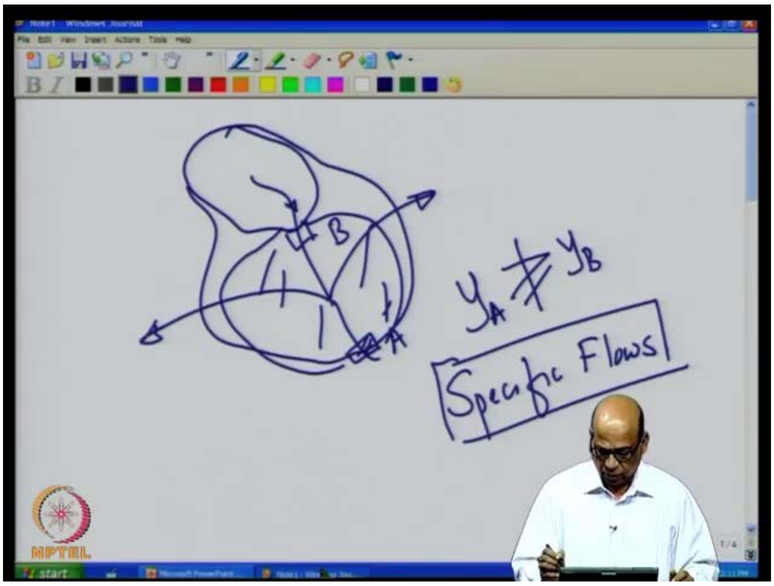
(Refer Slide Time: 12:17)



So with this, now we will progress and then look at some examples of what we mean by consistency of the data with respect to the mask curve as well as with respect to the specific flows. The specific flow is expressed as flow volume per unit area of the catchment. So, simply look at, let say a particular gauge, it has recorded a certain flow and you look at the annual average flow let say and divided by the catchment area. That will give you the flow volume per unit area of the catchment. Therefore, it will be typically in the depth unit or cubic meters, million cubic meters per square meter of the area. In that area, in those units, the specific flow is expressed.

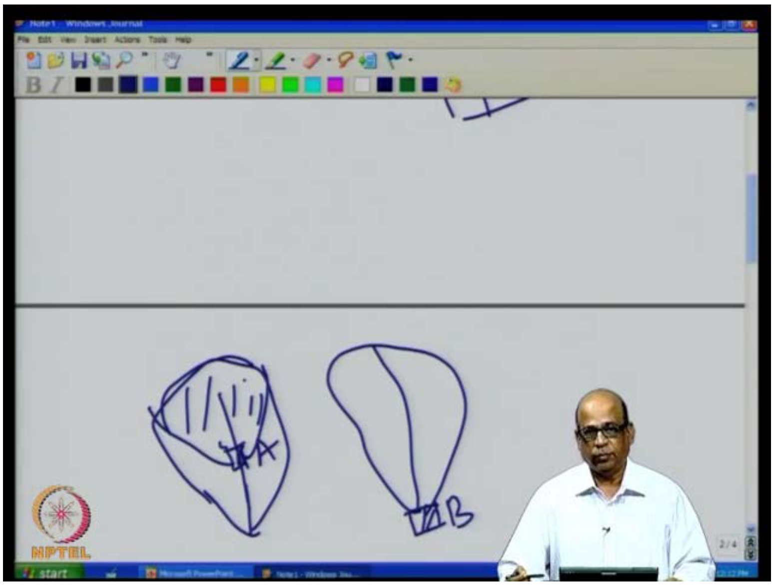
Now, as I said, the specific flow is, in fact, representative of catchments response to precipitation. There is a precipitation occurring. How much of this precipitation gets converted into runoff is what is given by specific flow. So, it is, in fact, the response of the catchment to an input that you are providing, the hydrologic response. Now, if a number of gauge stations are located in the same hydro climatic region, as I said hydrologically homogeneous region, then you must expect that the specific flows are comparable.

(Refer Slide Time: 14:14)



This again, I have told in the last class, but, because it is an important concept, let us again recall this, what I mean by that.

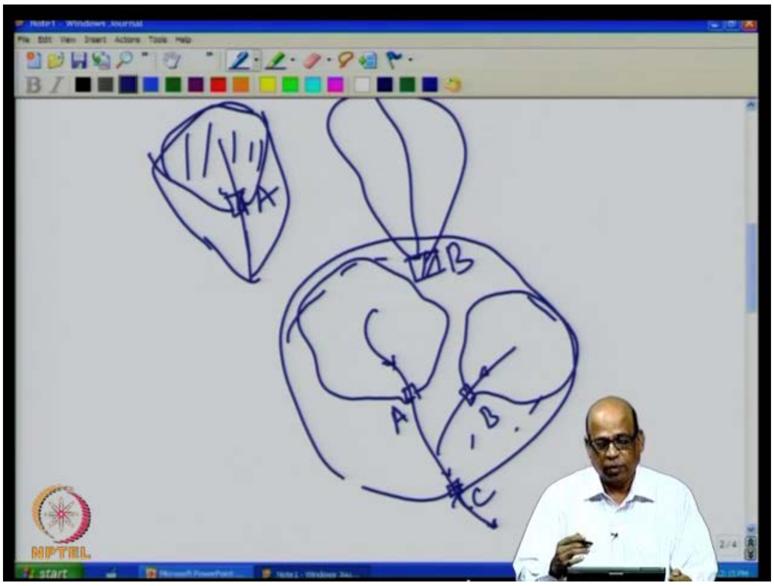
(Refer Slide Time: 12:18)



Let say that, you have two regions like this. There is one catchment here and the flows are taking place here. Then, you have recorded the flows here at this point A and this is the catchment that is contributing to the flows that bay A. Then, there is an another hydrologically homogeneous region to this catchment and you have recorded the flows at B. These two are different catchments.

Now, if these are hydrologically homogeneous in the sense that, their responses are the same to a given precipitation. Then, the specific flow at B, which is the average flow divided by the total area of this catchment must be comparable to the specific flow at A, which is the average flow that is recorded at this location with respect to the catchment area, divided by the catchment area.

(Refer Slide Time: 15:26)



Now, typically we do this when we are analyzing large water resource system, where here is one gauge here and there is another gauge here. May be, perhaps, you have another gauge here like this and the flow directions are like this. Let say, this is the flow direction. Now, we compute this specific flow at this location with respect to catchment area of this particular gauge. Similarly, we compute the specific flow of this with respect to the catchment area of this gauge. Let say, we call it as A, B and C. Now, the C will comprise of this entire area. These borders are coinciding here.

So, the gauge C will measure everything that is measured at B, everything that is measured at A plus the intermediate catchment flow between these locations. Now, when we compute the specific flow at B, we take the flows at this and take this area and compute the specific flow at B and similarly, specific flow at A and specific flow at C.

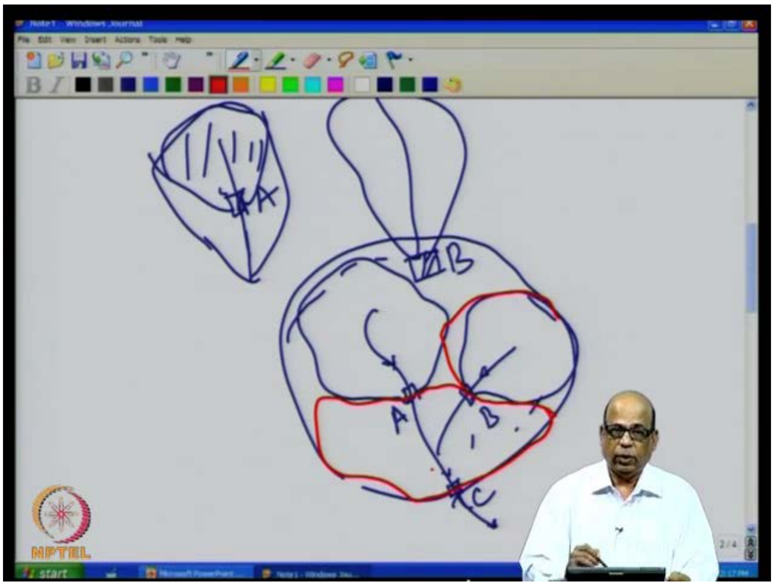
The areas are different and so are the flows at all of these locations, but, the ratios of the flows to the area will be approximately the same, if they are hydrologically homogeneous. Which means that, the precipitation pattern is the same in all the three catchments, as well as the lands used patterns are the same, the vegetation is the same, the amount of urbanization that has taken place in these catchments are the same, and soil is the same. In other words, the complete gamete of hydrologic variables that contribute to runoff generation are the same for all this catchments. In which case, the response of these catchment will be same for a given precipitation pattern and then, we say that these are hydrologically homogeneous catchments.

If they are, in fact, hydrologically homogeneous catchments, then the specific flow that we compute in that location must be the same or must be comparable. Ideally, it must be same as this, but, at least it must be comparable to specific flow competent at this location.

Let us say that a specific flow computed at C is much higher compared to the specific flows at A and B. Let us assume for the time being that specific flow at A and specific flow at B are comparable. Therefore, there is a good reason to believe that the precipitation patterns and hydrologic responses of these two catchments are similar.

However, the specific flow at C which represents its entire area and typically it accounts for this intermediate catchment area, which is not covered by both this A and B. If that specific flow is much higher compared to the specific flow at A and B, what that does mean?

(Refer Slide Time: 18:41)



It means that the intermediate catchment here let say that, I show this by some, let say all of this has been the intermediate catchment flow. In fact, these two are adjoining with each other. The entire flows has computed captured by A and B. Now, if the specific flows at C are much higher compared to the specific flows at A and B, it means one of the two things. It may mean that the precipitation pattern here, the precipitation values are much higher in this intermediate catchment when compared to the precipitation pattern here. Therefore, for the same hydrologic characteristics of the region you are getting more higher, more flows at this location, more observed flows at this location or it may also mean that the precipitation patterns may be the same, but, your catchment characteristics are such that for the same given precipitation it is producing a higher flow in that location.

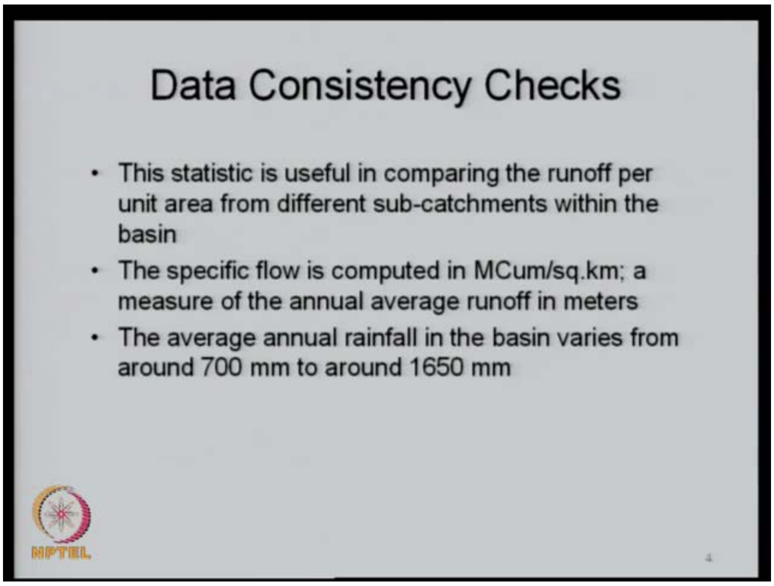
If you are gauged, if your recorded data is correct, then it means one of these two things. One is that, the precipitation values are much higher here or that the hydrologic characteristics are much different compared to these other two catchments. Therefore, it points to a non homogeneity of this catchment compared to the other two catchments.

So, these are the tests that we need to do for checking the homogeneity and the consistency of the data at that particular location. So, we will go back to the example and then see how these values look. Now, we compute the annual specific flows. Typically, when we are doing large scale system analysis, water resource system analysis, we look at the annual flows and then compute the specific flows.

So, let say fifty years of data is available and annual data is available. You take the annual average flow and then divide it by the area of the catchment and you get the annual specific flow. This can also be done for seasonal flows. Let say one season, for Monsoon season and non monsoon season, you can compute the specific flows.

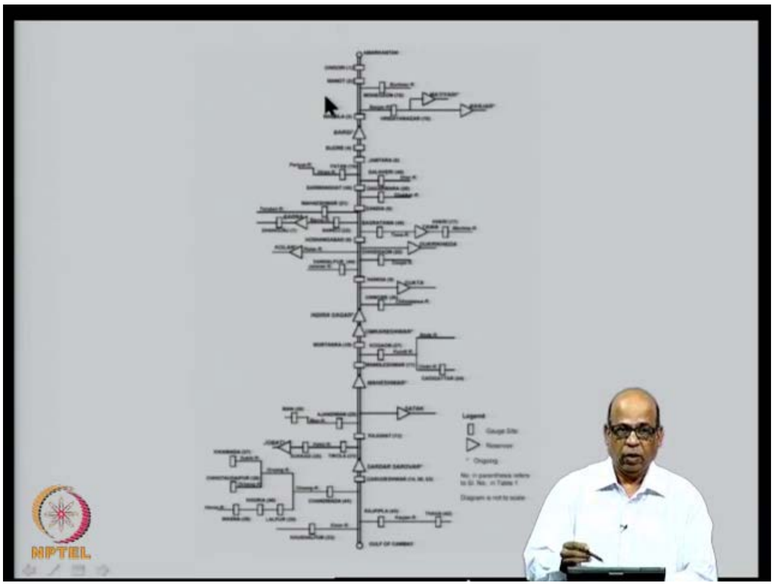
Now, as I have been mentioning, the specific flow is useful in comparing the runoff per unit area from different of catchment within the larger catchment and the units are typically million cubic meter per square kilometer. This is the measure of annual average runoff in depth units, typically in meters or centimeters or something.

(Refer Slide Time: 21:59)



The example that I have been showing in the last class was this. I was talking about Namarda river basin. This example has an average annual rainfall varying between 700 millimeter to about 1650 millimeters. As you go from east to west the rain fall is varying significantly. Therefore, you know, for large river basins as the rainfall patterns are changing, if your hydrologic responses are the same, then, you may expect different specific flows for different catchments. Let us see how it looks like.

(Refer Slide Time: 22:37)



So, this is the system view. It may not be very clear, but, do not worry about it. It only shows that there are large numbers of gauges, stream gauges. These rectangles here, small rectangles, they are stream gauges and this is the Narmada river. This is Amarkantak here at this point and this is the Bargi reservoir here and then, there is Indira Sagar reservoir here. There is Maheshwar reservoir here and Sardar Sarovar reservoir here and so on.

There are large numbers of gauges. In fact, fifty number of gauges at which the flows are recorded. On all of this, the available record at these gauges varies from as few as five years and the length of the record varies from as few as five years to as high as about forty years, fifty years and so on at several locations. We have to make use of this data to arrive at decisions. Let say that, you wanted to put another reservoir here and let say that there is an ongoing reservoir for which you want to develop operating policies. In which case, you have to analyze how the flows, flow pattern has been at that location and therefore, it is important for us to check.

Let say for Maheshswar reservoir, there is a gauge Mandleshwar. Now, the Mandleshwar gauge at this location, if I slightly zoom that or draw a zoom picture here, this is Mandleshwar here, at this location. Then, there is another gauge here and this is Mortakka and this is Maheshwar. That is Mandleshwar which contribute to Mandleshwar reservoir and there is also another small stream that is coming here and there is a gauge called Kogaon. So, just focus on this Mandleshwar gauge, Now, this has an intermediate catchment between Mandleshwar and Mortakka. Between this, there is a very small intermediate catchment and all also between Kogaon and Mandleshwar.

So, we look at the catchment for Mandleshwar, catchment at Mortakka, catchment at Kogaon and compute the specific flows because, they are quite close to each other and they are belong to the same sub catchment. If we can reasonably assume that there is hydrologic homogeneity in terms of the response of the catchment and sub catchments, then the specific flows at this location must be comparable. Similarly, you look at, let say Sardar Aarovar reservoir. Near Sardar Sarover reservoir, just upstream of that, there is a Rajghat which is a stream gauge. Then, there is also Tikola, which is coming from the site here and then, immediately downstream of that, there is a Garudeshwar stream gauge.

So, you will have records. Before this reservoir came up, you will have records in Garudeshwar and then, you also have records after the flow has been obstructed by the Sardar Sarovor reservoir. We will see what is to be done with the reservoirs and how to compute the normalized flows subsequently. But, assume that you have normalized flows. By normalized flow, we mean that the effect of the reservoir has been taken out and the flows have been converted into naturalized flow series.

So, if you have flows at this location, flows at this location and flows at this location, you can examine the consistency of the data of Garudeshwar with respect to Tikola as well as with respect to Rajghat. So, you check the consistency with respect to each other and then, you will be able to point out whether the data at a particular location is, in fact, inconsistent with respect to data at other locations. We would not be able to say for sure whether this data is correct or that data is correct. But, let say you have large number of gauges like this and all of them, in fact, many of them are consistent with each other, but, one of them is inconsistent. Definitely, you will be able to point out that there is something wrong with that data that is observed at this particular location.

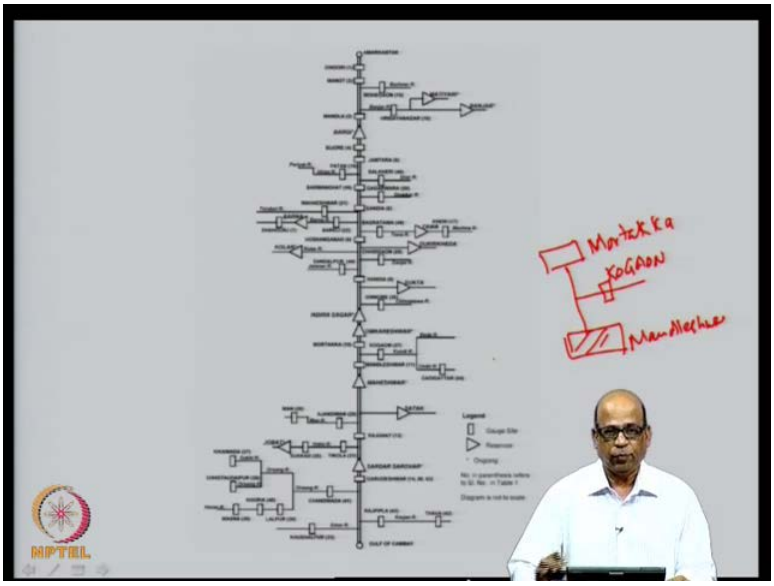
(Refer Slide Time: 27:10)

Annual and Seasonal Specific Flows for Gauge Sites

S.No.	Gauge site	Catch-ment area (sq. km)	Data used (period)	Duration (years)	Annual specific flows (MCum per sq. km)	Seasonal specific flows (MCum per sq. km)
1	Dindori	2,292.00	1988-1999	12	0.5460	0.4881
2	Manot*	4,667.00	1976-1999	24	0.6519	0.6113
3	Mandla Town	Not available	1977-1980 1993-1995	7	NA	NA
4	Bijore	14,561.00	1988-1999	12	0.5984	0.4361
5	Jamtara	17,157.00	1971-1999	29	0.5350	0.4569
6	Sandia	33,953.50	1978-1999	22	0.4247	0.3569
7	Hoshangabad	44,543.00	1972-1999	28	0.5148	0.4577
8	Handia	54,027.00	1977-1999	23	0.4684	0.4116
9	Mortakka	67,184.00	1970-1978 1988-1999	21	0.4659	0.4145
10	Mandleshwar	72,809.30	1971-1999	29	0.4565	0.4132
11	Rajghat	77,674.10	1971-1999	29	0.4349	[illegible]3943
12	Garudeshwar	87,892.00	1971-1975 1980-1999	25	0.3670	[illegible]278
13	Mohegaon	4,622.00	1977-1999	23	0.47[illegible]	[illegible]
14	Hridayanagar	3,370.00	1976-1999	24	0[illegible]	[illegible]
15	Patan	3,950.00	1979-1999	21	[illegible]	[illegible]

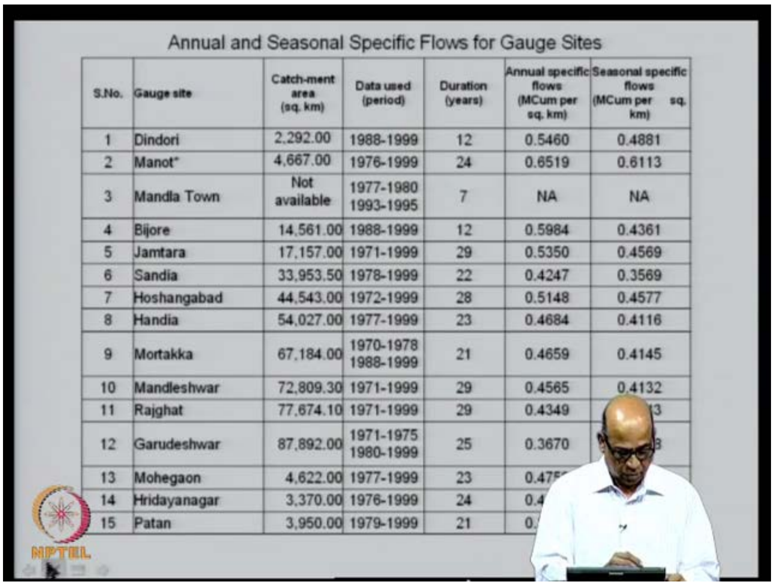
All right. So, when we do that, as I mentioned, for every gauge that we have, let say Dindori, Manot, Mandla town etcetera, you will have the catchment area.

(Refer Slide Time: 27:23)



Now, all these different gauges are pointed out are shown here. This is Handia, for example, and this is Hoshangabad and then this is Jamtara. This is Manot gauge and Hridaynagar and so on. So all this, at all these gauges, we have the catchments which are contributing to the flows that are recorded at those particular gauges. So, these are gauge sites and these are the catchments.

(Refer Slide Time: 27:55)



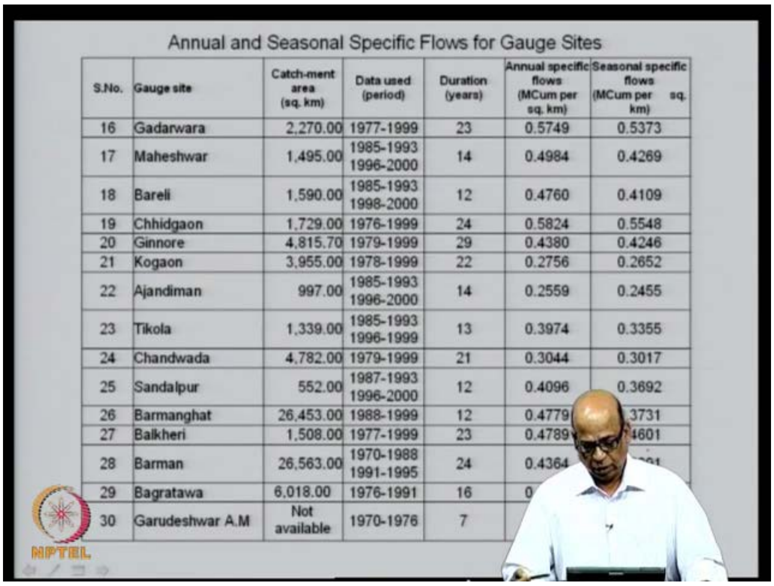
Annual and Seasonal Specific Flows for Gauge Sites

S.No.	Gauge site	Catch-ment area (sq. km)	Data used (period)	Duration (years)	Annual specific flows (MCum per sq. km)	Seasonal specific flows (MCum per sq. km)
1	Dindori	2,292.00	1988-1999	12	0.5460	0.4881
2	Manot*	4,667.00	1976-1999	24	0.6519	0.6113
3	Mandla Town	Not available	1977-1980 1993-1995	7	NA	NA
4	Bijore	14,561.00	1988-1999	12	0.5984	0.4361
5	Jamtara	17,157.00	1971-1999	29	0.5350	0.4569
6	Sandia	33,953.50	1978-1999	22	0.4247	0.3569
7	Hoshangabad	44,543.00	1972-1999	28	0.5148	0.4577
8	Handia	54,027.00	1977-1999	23	0.4684	0.4116
9	Mortakka	67,184.00	1970-1978 1988-1999	21	0.4659	0.4145
10	Mandleshwar	72,809.30	1971-1999	29	0.4565	0.4132
11	Rajghat	77,674.10	1971-1999	29	0.4349	[illegible]
12	Garudeshwar	87,892.00	1971-1975 1980-1999	25	0.3670	[illegible]
13	Mohegaon	4,622.00	1977-1999	23	[illegible]	[illegible]
14	Hridayanagar	3,370.00	1976-1999	24	[illegible]	[illegible]
15	Patan	3,950.00	1979-1999	21	[illegible]	[illegible]

So, for some places, you may not have the catchment available and also the duration is very small. Therefore, you can ignore those stations. So, the catchment area is given here in square kilometers and then, you have to calculate the annual specific flow million cubic meters per square kilometer.

The annual specific flow is computed based on the annual average flow recorded at that particular gauge site divided by the catchment area, which you are contributing to flow that particular region. So, these are the values that you get 0.546, 0.659 and so on. These are the values that you get.

(Refer Slide Time: 28:52)



Annual and Seasonal Specific Flows for Gauge Sites

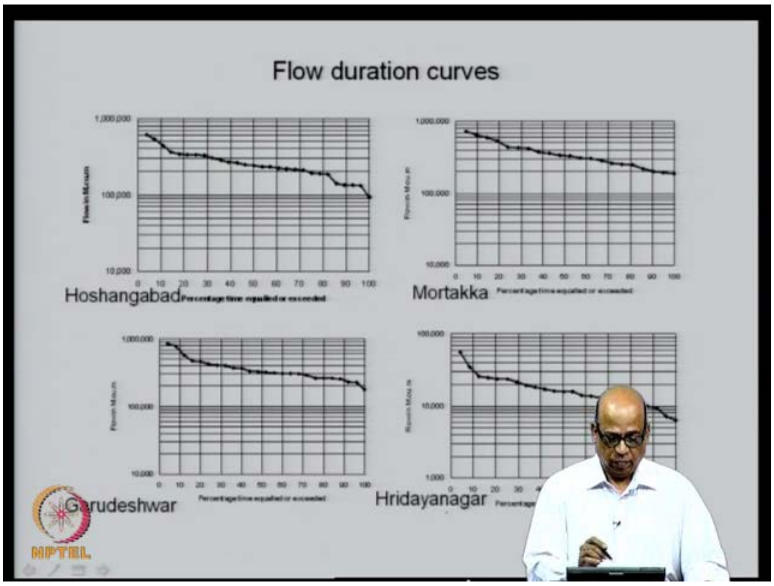
S.No.	Gauge site	Catch-ment area (sq. km)	Data used (period)	Duration (years)	Annual specific flows (MCum per sq. km)	Seasonal specific flows (MCum per sq. km)
16	Gadarwara	2,270.00	1977-1999	23	0.5749	0.5373
17	Maheshwar	1,495.00	1985-1993 1996-2000	14	0.4984	0.4269
18	Bareli	1,590.00	1985-1993 1998-2000	12	0.4760	0.4109
19	Chhidgaon	1,729.00	1976-1999	24	0.5824	0.5548
20	Ginnore	4,815.70	1979-1999	29	0.4380	0.4246
21	Kogaon	3,955.00	1978-1999	22	0.2756	0.2652
22	Ajandiman	997.00	1985-1993 1996-2000	14	0.2559	0.2455
23	Tikola	1,339.00	1985-1993 1996-1999	13	0.3974	0.3355
24	Chandwada	4,782.00	1979-1999	21	0.3044	0.3017
25	Sandalpur	552.00	1987-1993 1996-2000	12	0.4096	0.3692
26	Barmanghat	26,453.00	1988-1999	12	0.4779	[illegible]
27	Balkheri	1,508.00	1977-1999	23	0.4789	[illegible]
28	Barman	26,563.00	1970-1988 1991-1995	24	0.4364	[illegible]
29	Bagratawa	6,018.00	1976-1991	16	[illegible]	[illegible]
30	Garudeshwar A.M	Not available	1970-1976	7	[illegible]	[illegible]

Similarly, for several sites, we have chosen thirty sites here, in this particular case. All the thirty sites, we compute the annual specific flows. So, the first step is simply compute the annual specific flows and then, also look at the monsoon specific flows or the seasonal specific flows for monsoon period.

So, I have also showed here the seasonal specific flows. Look at these values. The seasonal specific flows must be smaller than the annual specific flow because, you are taking the total volume of flow that has occurred during the entire year in the annual specific flow and the total volume of flow that has happened or has occurred during the monsoon season, which is smaller than the flows in the annual scale. Therefore, the seasonal specific flows will be smaller than the annual specific flows.

Like this, we compute for all the thirty gauge sites. The specific flows, wherever it is possible, where the data is either very small or the catchment area or some other data is not available, we do not include that.

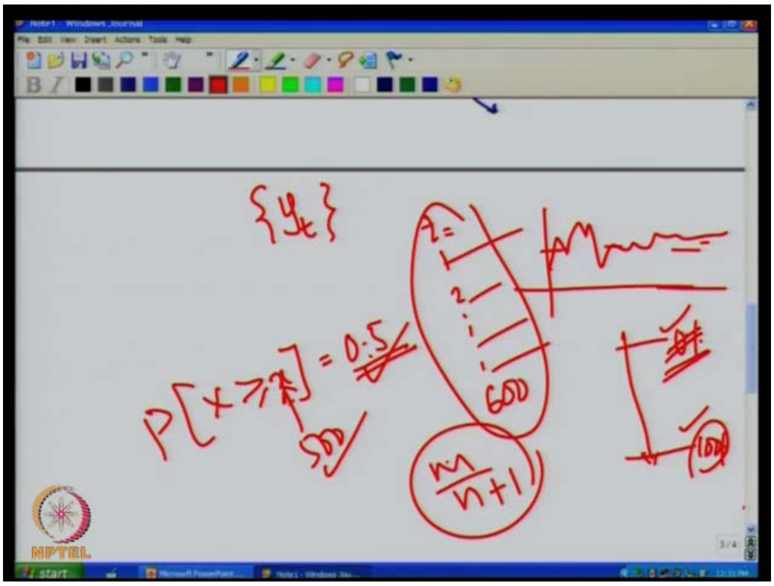
(Refer Slide Time: 30:02)



Then, the next step is, we plot what is called as flow duration curve. Remember, all of these we are doing to examine whether the data that we have for our analysis purposes is good to use or is there any difficulty. Is there any problem with respect to some gauges? In which case, those gauges we have to discard in our analysis.

The flow duration curve is also an important step in hydrology construction of flow. Duration curve is a important step in hydrology. This indicates, let say that at a particular series we have, let say we have the observed time series y t and these are flows. We are talking about monthly flows, let say of monthly flows and t is equal to 1, 2 etcetera.

(Refer Slide Time: 30:55)

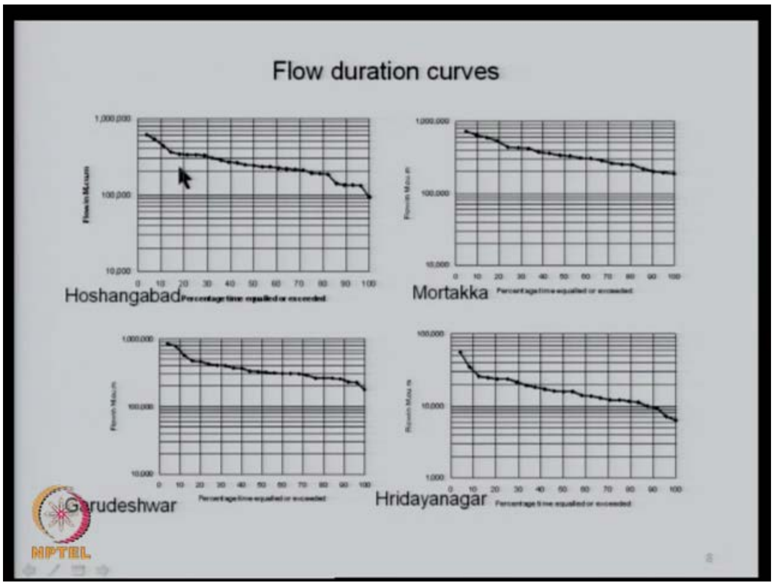


If you have fifty years of data, we may have 50 into 12, and that is 600 values. So, this is the time series of monthly flows that we have like this. This has a long term average of this and so on. Now, the flow duration curve indicate the percent of values of flows, let say we are saying here that probability of x being greater than equal to x is equal to, let say 0.5. What does it mean? We are saying with 50 percent probability, this flow value x will be exceeded. That is, the flow, let say this flow is 500. So, the flow stream, flow x will exceed a value of 500 with a probability of 0.5.

Similarly, if I increase this to 0.7, what happens? This flow magnitude will decrease. So, the probability that the flow will exceed, let say 700 units will be equal to 0.7 and so on. So, the flow duration curves give this tradeoff between the probability of exceedance and the magnitude of the flow. How do we compute? We simply use, let say either the Weibull's formula or something and arrange them in rank decreasing order, and ascend rank and then give the percentage of exceedance and plot the percentage of exceedance with the flow magnitude. That is how you obtain the flow duration curve.

If you are not using Weibull's formula, let say which is m divided by n plus 1. Forgetting the probabilities, you can simply get the fraction of flows that will be exceeded. Simply take it as and arrange them in decreasing order. Let say, you have 600 values and arrange them in decreasing order and then, assign the lowest value as 100 percent exceedance because, the minimum value will be exceeded 100 percent of time in the record. This comes to about, in fact, it has to be actually 0 percent, but, as I mentioned in dealing with or probability plotting positions, that you should not assign a 0 percent or 100 percent or you may go as close to 0 percent as possible and close to 100 percent as possible in this. That is why we use typically m divided n plus 1 as plotting positions for the estimation of probabilities. So, we use this kind of assignment of probabilities to each of these flows here and then, plot the flow duration curves. Let see how the flow duration curve look like.

(Refer Slide Time: 34:11)

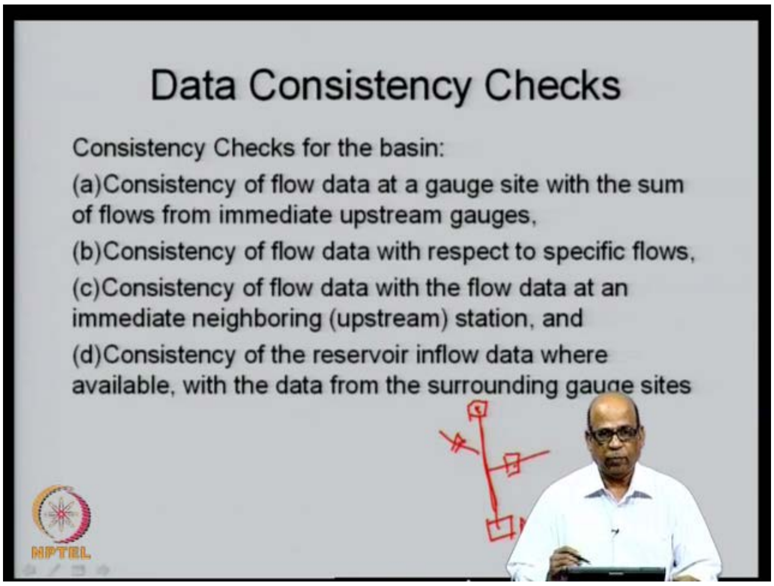


So, at different locations, we have the flow duration curve like this. At any given value here, the x axis is the percentage time equal or exceeded. Let say, you have 70 percent and you get a flow of the order of this is 10 and this will be 11. That is 100000 and this is 110000. This is flow in million cubic meter.

Like this, it indicates the flows that will be equaled or exceeded 50 percent of time, if you are looking at this location, 60 percent of time and so on. So, the flow duration curve gives information for design purposes. For example, at Mortakka, you also have 70 percent. You have slightly higher flow. This is 100000 here and this is also 100000 here. So, 60 percent you have, 100000 and 120000 and so on. So, compare these with respect to each other for the same given percentage, what is the flow that is exceeded at various locations.

If Mortaka is downstream of Hoshagabad, then for same given percentage, you must have a larger flow. Similarly, Garudeshear is downstream of Mortakka for a same given percentage, you must have larger flow and so on.

(Refer Slide Time:35:42)



Then, we do the data consistency check. Now, this is a large river basin that we are considering and therefore, we do the data consistency for sets of gauges. We look at a particular gauge with respect to its immediate up stream gauge and its immediate surrounding gauges and if there is a reservoir, we also look at the reservoir flow with respect to immediate surrounding gauges.

So, the type of consistency we do for the data are the first consistency of flow data at the gauge site with the sum of flows from immediate upstream gauges. Like I said that, you have a gauge here and you may have an immediate upstream gauge on the same catchment or you may also have certain gauge flows that are coming from tributaries. Let say, there is a tributary here and you have a gauge here. There is a main stream and you have a gauge here and so on. So, when you are checking the consistency of this location, you must look at what has happened at this particular gauge, at this particular gauge, as well as this particular gauge. So, you are looking at immediate upstream gauges.

However, if you have another gauge that has contributed to this flow, that gauge, you do not have to worry because, that has already been recorded. The flows at that gauge have already been recorded by this particular gauge. Therefore, you look at immediate stream gauges for these particular gauges. So, we check the consistency of flow data at a particular gauge with the sum of flows from immediate upstream gauges. That means, something is being contributed from here and something is being contributed from here. You add all of them up or you look at sum of these and then look at the consistency of data.

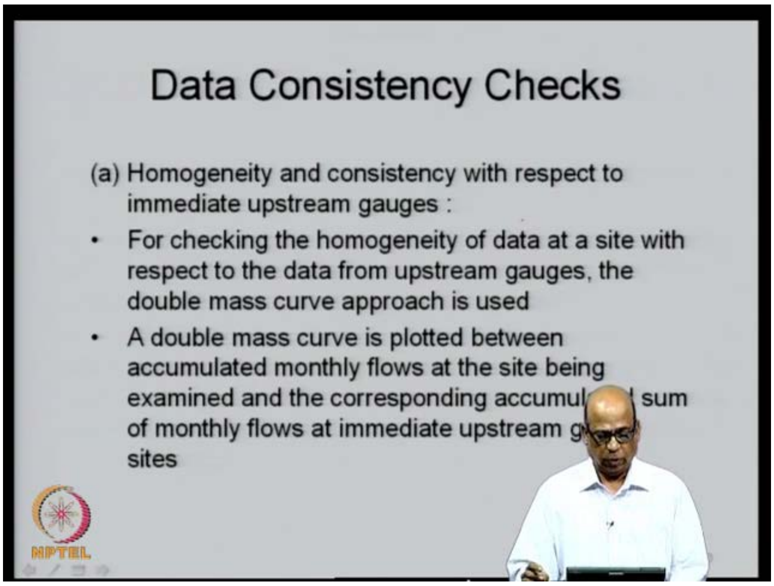
Then, we also look at the consistency of the flow data with respect to the specific flows. As I mentioned, you look at the catchment areas of each of these gauges and then, see the consistency with respect to the specific flows. Then, we look at the consistency of flow data with the flow data at an immediate neighboring upstream station. That means, the data at this point must be consistent with this individually. It must be consistent with this individually. So, one at a time also you look at.

In the first step, we consider all of this together and look at the sum of the data. Then, it must also be consistent with this individual gauges, stream gauges. Remember here, the underlying assumption is that all these gauges belong to a hydrologically homogeneous region. Precipitation pattern is the same and also the catchment response is the same at all of this location. If those are not there, then it will not be consistent.

Then, we also check the consistency of the reservoir inflow data were available. The data with the surrounding gauges, let say that around this point, immediately upstream of this, you have, let say a reservoir here and you have the reservoir levels. Typically, whenever you have reservoir sites, reservoir levels will be measured. Therefore, you have the flow or the inflow data to the reservoir, that is data on the inflow to the reservoir you will have.

Now, those should also be consistent with the data that you have observed at these locations. If you are downstream of the reservoir, then you have to account for the effect of this reservoir in terms of the control flows and then, look at the consistency of this.

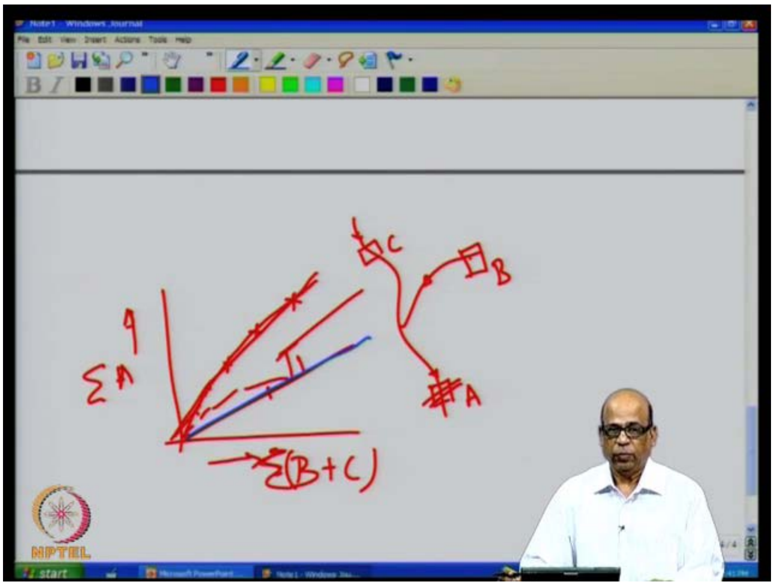
(Refer Slide Time: 39:40)



Then, we first start with homogeneity and consistency with respect to immediate upstream gauge. So, first we look at only two gauges. That is, a gauge A with respect to immediate upstream and gauge B and plot the double mass curve. That means, accumulated flows at this location with respect to the accumulated flows at this location. If there are more gauges that are coming from tributaries and so on, you also consider those gauges.

Then, we plot a double mass curve between accumulated monthly flows at the site being examined which is A in this case and the accumulated sum of monthly flows at all these upstream gauges. So, that will give you a double mass curve as I will show you now. Then, let say that the double mass curve that you plot, let me make that slight more clear now. What do I mean by the double mass curve?

(Refer Slide Time: 41:17)



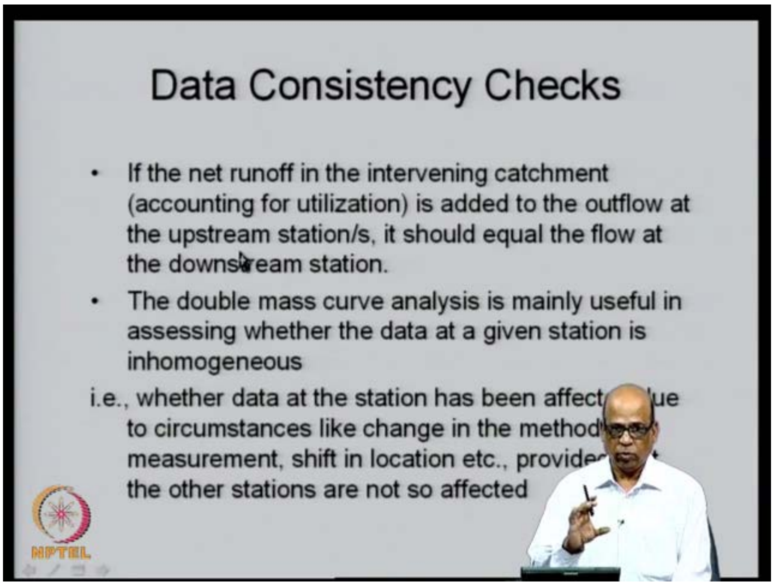
That, you have two gauges, let say this is gauge A and you have another gauge B and another gauge C here. This is the flow direction and this is the flow direction here. You are examining now the consistency at gauge A. B and C, that is the flow at B and flow of C, have contributed to flow at A along with the flow at intermediate catchment.

So, when you plot the double mass curve, what you will do is, you will take in some sense, in fact, you will take the sum of the flows at B and the flows at C and then plot the cumulative flows or sum of flows at B and C, with respect to the sum of flows at A. This is a cumulative flow feature and therefore, it will be continuously increasing.

Let say that, you get some such curve here. Actually, it should be a straight line as I mentioned. If it is perfectly consistent, it will be a straight line. Now, if you draw a 45 degree line here, this curve that you get for cumulative flows at A, which is a downstream gauge, must be above the curve that the values that you get for B plus C, which means this curve should be above the 45 degree line. That means that, the flows at this point are more than the flows at these locations, because of your contribution from the intermediate catchment, if there is no significant water use in the intermediate catchment or if the rainfall is not significantly lower.

So, this is what we examine. If it does not happen, let say that, the flow that you are getting here is something of this order. Let say that, you get a curve of this type, which is lower than 45 degrees, what does that mean? That means that, overall the flows at A are smaller than the flows at B plus flows at C, which should indicate that there is some problem with the data or we are missing the physical picture on what is happening between the gauges B and C and the gauge A. What is happening in the intermediate catchment? So, that is what should lead you to see or examine the physical features of this particular catchment. So, that is what we examine in this step.

(Refer Slide Time: 44:17)



So, if the net runoff in intermediate catchment, that is accounting for utilization, I keep on saying this because, normally we miss this kind of point when we are doing large scale data analysis. We have to also account for utilization. You may have used the waterform principal and industrial supply and then, you have a series that is recorded at a particular gauge. Unless you add this value at this flow that has been actually extracted from the river and then, you are reckoning the flows at this particular gauge by adding those, then the analysis will not be correct.

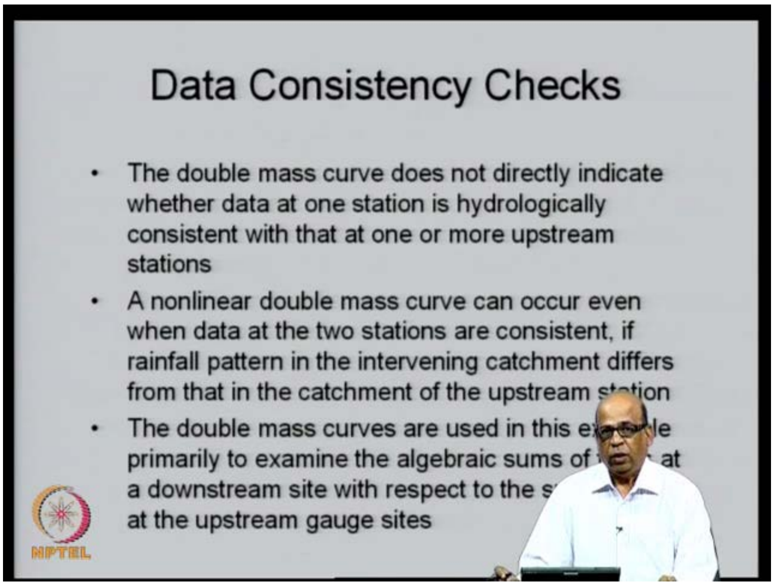
So, whenever you are doing analysis on hydrologic data series, you must account for all the upstream utilizations, all the loses and any disturbance that have taken place in terms of your catchment characteristics and so on. So, if the net runoff in the intervening catchment is added to the outflow at the upstream station, it should equal the flow at the downstream station, obviously.

That is, here the flow is happening like this. There is a upstream gauge A and a downstream gauge B. The flow at the downstream gauge B must be equal to the flow at upstream gauge A plus whatever has been contributed from the intermediate catchment between B and A, after accounting for any utilization that has been taken place between these two gauges.

Then, the double mass curve is mainly useful in assessing whether the data at a given gauge station is in homogeneous. That is, whether the data at this station has been affected due to circumstances like the change in the method of measurement or shift in the location itself, etcetera.

So, we are essentially examining, let say 50 years of data is available at a particular gauge. We are examining whether this 50 years of data is consistent in a statistical sense, with respect to the data that is observed in the immediate upstream gauges. If is not so, then you have to suspect that the method has been different or the gauge has shifted or the data is in homogeneous with respect to the other gauges.

(Refer Slide Time: 46:46)

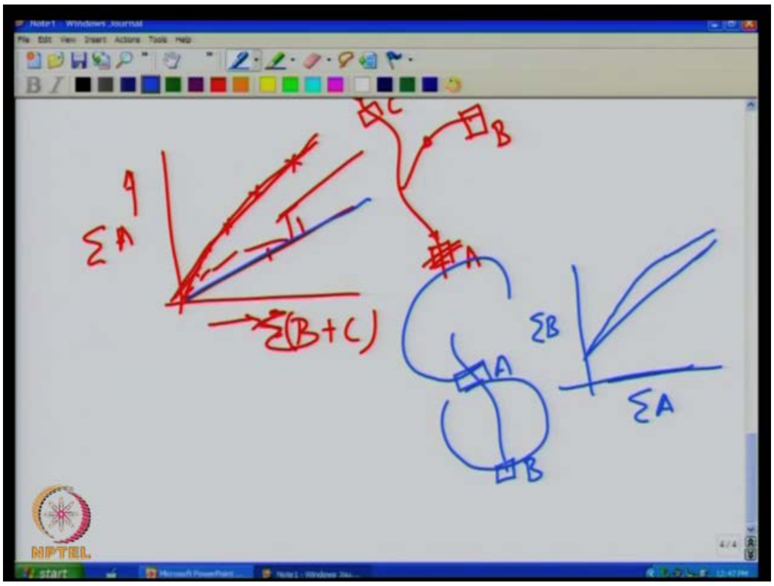


The double mass curve; however, does not directly indicate whether the data at one station is hydrologically consistent with that at one or more upstream station. It only shows that there is some problem here. But, it may not exactly say that it is, in fact, in homogeneous or it is in fact, inconsistent. We have to go into exactly what has happened at that particular site. That is what I keep on saying that, physically what has happened or what has been altered in the catchment is what we have to look at.

In many situations that I presently show, the double mass curve will be straight up to certain point and there is a sudden break and then, again it becomes straight a straight line. This means that, wherever there is a break, some change has happened in the catchment. Either the gauge has been shifted towards a different location and still we are labeling it as a same gauge name or the method has suddenly changed. Therefore, you have to measure the flows with a better accuracy or in a different manner all together. Therefore, there is a there is a change in the flow values with respect to the flow values that are recorded elsewhere.

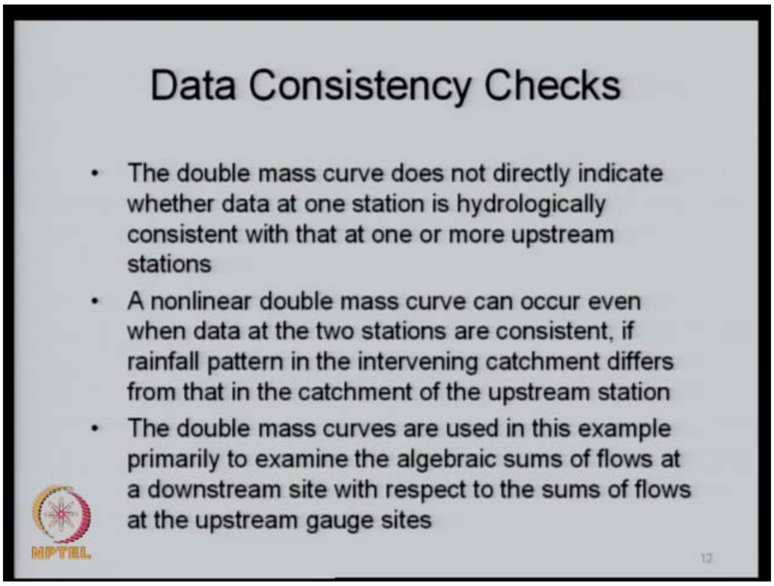
Now, while I keep on saying that this has to be a straight line, there will be situations where a non-linear double mass curve can occur even when the data at two stations are consistent. If the rainfall pattern in the intervening catchment differs from that in the catchment upstream location, what I mean by that is, let say that you have two stream gauges.

(Refer Slide Time: 49:01)



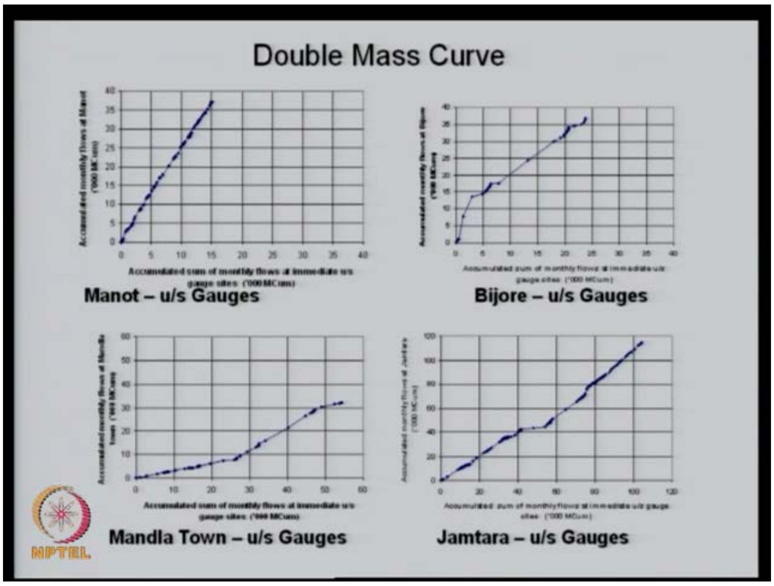
You take two stream gauges here. This is B and this is A. When you plot the double mass curve, you expect a straight line. Something like this, all right. But, you may get a non-linear curve like this. What does this point two? If you have a non-linear curve, let say this as summation of B and this is summation of A, flows at A. If you get the non-linear curve, it means that the rain fall pattern in intermediate catchment, which has contributed to the flows at B is much different from the rain fall pattern which is contributing to the catchment at A. Therefore, you can expect a non-linear relationship. If it is a linear, it means that the rainfall is the same and therefore, you are getting the flows at B as directly proportional to the flows A. This need not be the case, if your rainfall patterns are different. That is what we mean by this statement.

(Refer Slide Time: 50:12)



Now, the double mass curve used in the example that I am showing primarily to examine, if the algebraic sums of flows at downstream site with respect to the sums of flow at the upstream gauge site are consistent with each other. So, this the primary purpose for which we are using, especially for this example.

(Refer Slide Time: 50:41)



Now, we I will show you some typical mass curves. On the x axis, I am showing accumulated sum of monthly flows at immediate upstream gauge sites. So, you may have two gauge sites. You add the two gauge sites flows together and then, get the accumulated sums of flows. So, these are accumulated flow values in thousands of million cubic meters.

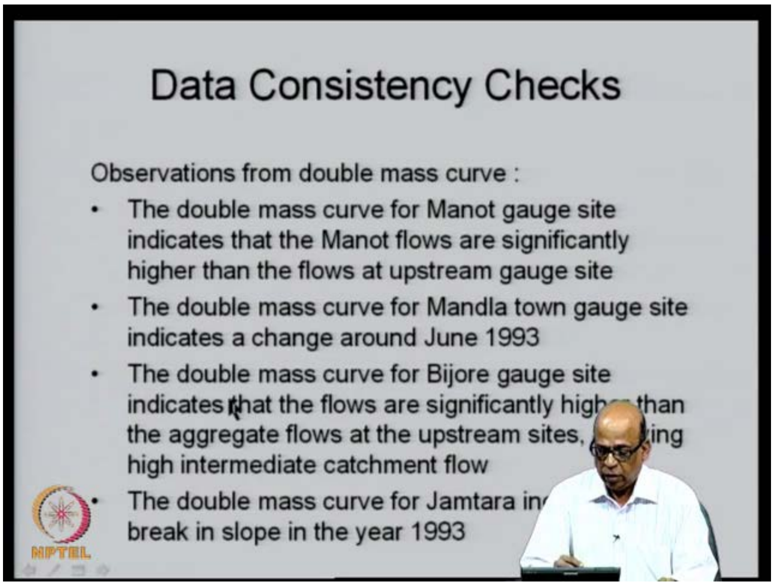
We are examining at Manot, at one of the gauge sites Manot. So, we plot on the y axis the accumulated monthly flows at Manot in thousands of million cubic meters and you typically get a straight line. This straight line is also above the 45 degree line that you may draw, indicating that there is no problem here. Similarly, we draw for various locations. Let say, this is Bijore. You may get certain kings here or certain breaks in the curve. This may be because of the way you have filled in the data and or because of the way the data has being observed or being recorded or because of certain changes that have happened in the catchment.

So, these breaks or these discontinuities that you see in the double mass curves, point to certain changes that have occurred, either in the way of recording or certain physical changes that have happened in the catchment.

Now, you see Mandla town with respect to upstream gauges. If you draw a line here, a 45 degree line, double mass curve here is well below the 45 degree line. Now, this will indicate that, there are certain things that are happening in the intermediate catchment in terms of the water utilization, which you have not recorded or also there is a possibility that the data that is available is so small compared to the upstream gauge data, that it has not been not able to capture or it is not well representative of the particular catchment itself.

In addition, there is also a break here, which indicates that some changes have happened in the catchment around this particular time. Like this, the streams, the double mass curves will indicate for a particular gauge, what has happened with respect to upstream gauge rain sites.

(Refer Slide Time: 54:22)



So, the double mass curve of Manot gauge, indicates that the Manot flows are significantly higher than the flows at upstream gauge site here. As I said, if you draw a 45 degree line, these flows are much higher compared to all the upstream gauge site, which indicates that the intermediate catchment flow, is much higher, compared to the actual observed flows in the upstream catchment. This is possible because, the intermediate catchment area itself may be large or the rainfall in the intermediate catchment may be significantly larger than the rainfall in the upstream catchment and so on.

Then, the double mass curve for Mandla town gauge site, indicates a change around 1993. So, this will show that there has been a change here and we can go back to the data and see where these changes are happening and around 1993 the change has happened.

We can go back to the records and see what kind of changes that have happened here and so on and then, put those changes in some statistical sense into our data.

The double mass curve for Bijore gauge site, indicates that the flows are significantly higher than the aggregated flows at the upstream sites, implying high intermediate catchment flow. Then, the double mass curve for Jamtare, indicates that the break in the slope and so on. So, the double mass curve essentially indicates whether the data at a particular point particular gauge is, in fact, consistent with the data at the immediate upstream gauge sites. We get the accumulated aggregate flows at all these upstream gauge sites and compare them with respect to the flows that are recorded at this particular gauge site.

The double mass curve is the first level consistency check that you do whenever you are analyzing large hydrologic systems, such as that one that I have shown in this particular example. Then, we move on to specific flows and then make sure that with respect to specific flow also, the data is consistent and so on.

So, this lecture as well as in the previous lecture, I have been emphasizing on the fact that the data that you obtained, may be unreliable or may be of poor quality or may have certain several breaks in the time series and may, in fact, be inconsistent with the data that that is recorded elsewhere in the same catchment.

So, these have to be examined before we start any hydrologic analysis. So, while the stochastic hydrology course, we have learned quite rigorous methods of analysis of the data and drawing inferences from the data. All these methods are as good as the data that you have for the analysis itself. Therefore, unless you are completely satisfied that the data that you are using is, in fact, of good quality and is in fact, consistent with respect to other gauges or with respect to whatever has happened in the catchment, you should be cautious about making decisions using the rigorous methods that we have studied in this particular course.

So, we will continue this discussion in the next class and then, complete the data and consistency analysis and also, perhaps look at how we account for the control structures. This means that, there is a reservoir that which has been operating for several years and then, you are recording the flows downstream of the reservoir. But, for analysis purposes you need the naturalized flows. How do we convert these control flow sequence into a naturalized flow sequence, which only should be used in the analysis. So, we will continue this discussion in the next class. Thank you for your attention.