Stochastic Hydrology

Prof. P. P. Mujumdar

Department of Civil Engineering

Indian Institute of Science, Bangalore

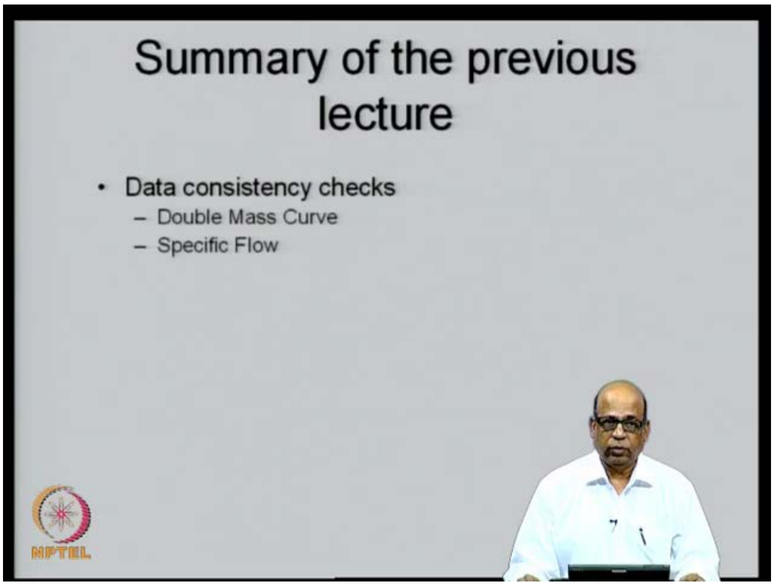
Lecture No. # 38

Data Consistency Checks - III

Good morning and with this the lecture number 38 of the course, Stochastic Hydrology. As we draw closer to the course, I am just discussing now the data issues, how you test the consistency of the data. And then today we will also discuss something about naturalization of the data or normalization of the flows specifically when you have control structures, the time series that you need to consider for the analysis.

As I keep mentioning, normally these are the issues that we tend to forget that, whenever we have the data without looking at where the data has come from or what what is the source of the data; in the sense that, whether it is in fact representing the natural flows at a particular location if we are dealing with the flow data. We tend to start analyzing that, but then the analysis may lead to certain inferences or conclusions which may in fact not be correct, because the flow data that we are considering will not in fact be naturalized data. And therefore, even before starting the analysis, we need to be alert to the situation of the type of data that we are using, and what the data actually represents. We will discuss these issues today along with the consistency of the data itself that I have been discussing in the previous lecture.

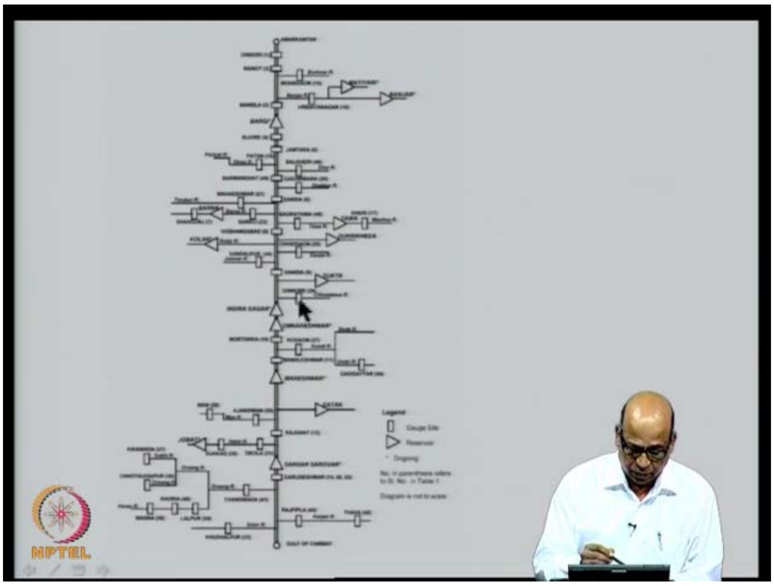
(Refer Slide Time: 01:49)



So, in the previous lecture if you recall, we were talking about the data consistency checks; and specifically we dealt with the double mass curve and the specific flow concepts for checking the data consistency. Typically, the specific flow we we use this method of specific flows, when we are dealing with the flow data. And double mass curve we can use both for rainfall as well as a flow data, which are the two major variables that we consider in most hydrologic applications. So, the double mass curve if you recall, is the curve between the accumulated flows at all the gauges typically upstream of the particular gauge, specific gauge at which you are checking the consistency of the data with respect to the accumulated flows at the candidate gauge itself.

And the specific flow is the flow per unit area. So, typically we reckon annual specific flows or the seasonal specific flows; so, you take the annual average flows indicated at a gauge divided by the catchment area of that particular gauge, catchment area contributing to the flows at that particular gauge that is what we reckon as the specific flow.

(Refer Slide Time: 03:21)



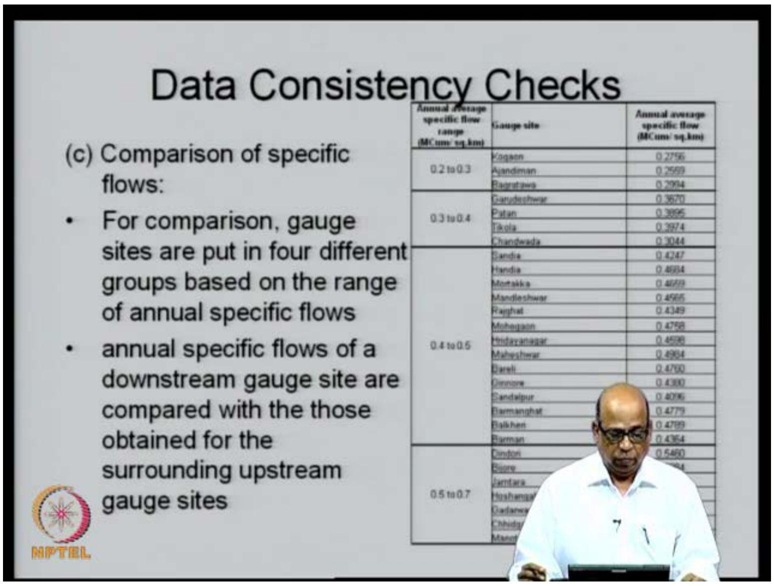
We were discussing the particular case study as I mentioned, this is a Narmada case study and this is Amarkantak here and this is a gulf of (( )); you have Sardar Sarovar reservoir here, Maheshwar, Omkareshwar, Indira Sagar and Bargi is here. And there are a large number of other reservoir forms on the tributaries. And these rectangular boxes that you see here are in fact the gauge locations. So, there are a large number of gauges at which the data is available.

The first exercise that we do in such large systems is to check the consistency of the data with respect to each other. That is consistency of the data at a particular gauge with respect to the gauges in the surrounding region. And in the in the previous lecture, I have discussed the double mass curve for this particular system and then we arrived at certain conclusions.

As I mentioned the double mass curve will indicate whether the flow at a particular gauging station; let us say, you take Handia for example, now this flow has been contributed by these tributaries on the left and right side here, as well as what is coming on the main stream itself. So, Hoshangabad is here, so the gauge at Hoshangabad plus that is the flow at Hoshangabad plus the flow at let us say this is Chirgaon and then this is Sandalpur. These flows will get added to come to Handia, in addition there is also an intermediate catchment that also contributes to the flow at Handia.

Therefore, when you plot the double mass curve at Handia, the Handia flows must be greater than the flows accumulated flows at all these three stations together. That is what we examine in the double mass curve.

(Refer Slide Time: 05:23)

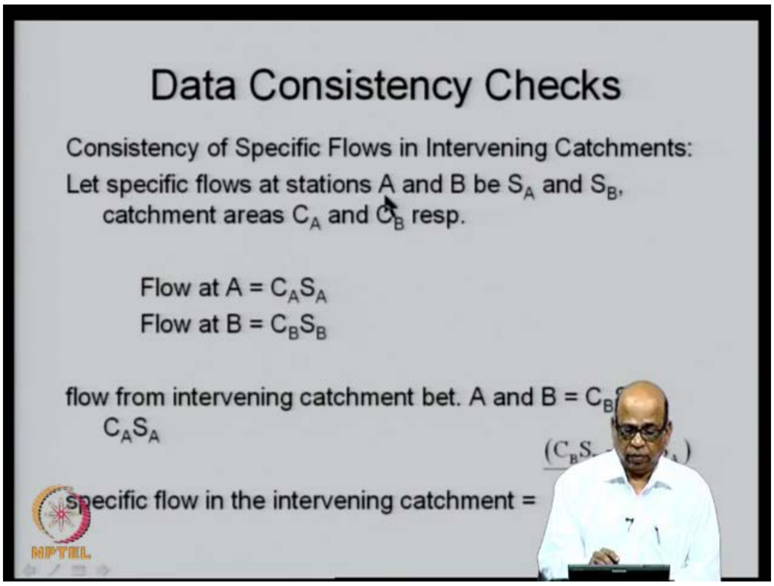


Then we come to the specific flows, we compute the average annual specific flow. So, at each of these locations I have shown in the last class last lecture, the catchments for each of these gauges are known, and the average annual flows are computed; you simply compute based on these two values the specific flows. When you compute those specific flows, there we typically classify the specific flow into three or four classes.

There are certain stations, which have a specific flow of 0.2 to 0.3 then certain other stations between 0.3 and 0.4, 0.4 and 0.5 and so on. This will give an idea of which regions in the large basin are contributing to flows in a certain range, range of flows. Now, you have a look at it, the specific flows typically ranged from around 0.25 to around 0.7. This means that 0.7 million cubic meter per square kilometer is the yield of that particular basin.

So, the annual specific flows of a downstream gauge site then, we compare it with the surrounding upstream gauge sites. Now, this is just a statistic this is, this will give you an overview of what kind of specific flows can be expected from the system. Now, we go and then start comparing with the specific flows of the surrounding gauges.

(Refer Slide Time: 07:19)



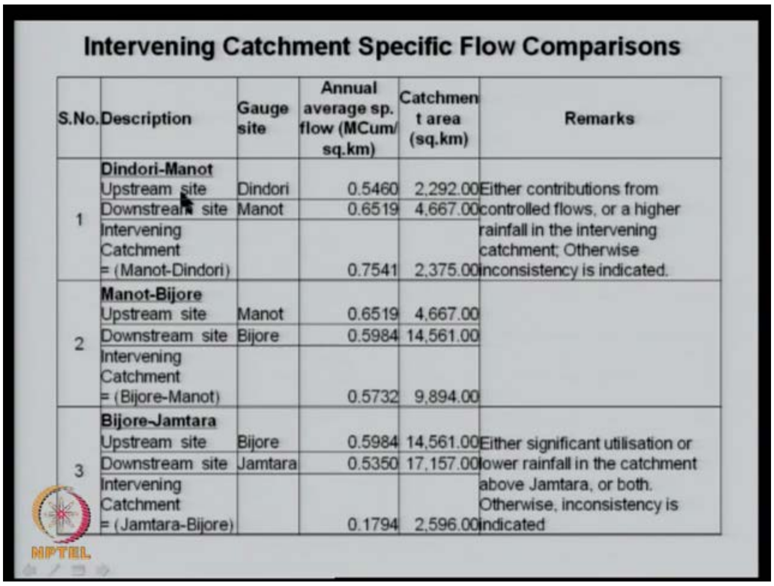
When we do this we look at let us say that, there are two stations we are considering station A and station B, and we get a specific flow of S A and S B respectively; and the associated catchments catchment areas are C A and C B. So, the flow at A is simply catchment area into the specific flow, this is C A into S A, this is the average flow that you can expect at that particular gauge.

Then similarly flow at B is C B into S B. Now, if A and B are upstream and downstream gauges here let us say A and B (Refer Slide Time: 07:59). So, you get the flow at B as C B into S B; flow at A as C A into S B S A. So, the difference that you get must be the intermediate catchment flow.

So, from the intervening catchment you get C B S B minus C A S A, this is the flow that is contributed by this particular intermediate catchment. If there are no anthropogenic disturbances that are taken place in in the sense that, you are not using the water in between or there is no addition of water apart from the precipitation. Now, the specific flow of the intermediate catchment we compute that means this intermediate catchment has contributed C B minus C B S B minus C A S A.

And then the total catchment area is C B minus C A, because C B is the catchment area here, C A is the catchment area here. So, we compute the specific flow in the intervening catchment at this location. Now, this is an important indicator of how the region behaves; let us say that, you get the intermediate catchment specific flow here and then another sub catchment you consider take the intermediate catchment specific flow there, and then you can compare the intermediate catchment specific flows. That is what we will do now.

(Refer Slide Time: 09:30)



Intervening Catchment Specific Flow Comparisons

S.No.	Description	Gauge site	Annual average sp. flow (MCum/ sq.km)	Catchment area (sq.km)	Remarks
1	**Dindori-Manot**				
	Upstream site	Dindori	0.5460	2,292.00	Either contributions from controlled flows, or a higher rainfall in the intervening catchment; Otherwise inconsistency is indicated.
	Downstream site	Manot	0.6519	4,667.00	
	Intervening Catchment = (Manot-Dindori)		0.7541	2,375.00	
2	**Manot-Bijore**				
	Upstream site	Manot	0.6519	4,667.00	
	Downstream site	Bijore	0.5984	14,561.00	
	Intervening Catchment = (Bijore-Manot)		0.5732	9,894.00	
3	**Bijore-Jamtara**				
	Upstream site	Bijore	0.5984	14,561.00	Either significant utilisation or lower rainfall in the catchment above Jamtara, or both. Otherwise, inconsistency is indicated
	Downstream site	Jamtara	0.5350	17,157.00	
	Intervening Catchment = (Jamtara-Bijore)		0.1794	2,596.00	

Just for an example I will give you between two stations Dindori and Manot we take. Manot is here, Dindori is here they are adjacent gauges here quite close to each other. So, we take between Dindori and Manot, the upstream site is Dindori and downstream site is Manot. Now, if you look at the average annual specific flow it is 0.546 at Dindori and 0.6519 at Manot, and the associated catchment areas are given.

However, when you look at the intervening catchment specific flow by adopting this method now; you look at the intervening catchment it is 0.7541, which is higher than either of these two specific flows (Refer Slide Time: 10:26). Now, what does this mean? It means that, either the contributions from controlled flows, or a higher rainfall in the intervening catchment. That means is there any controlled flow that is coming in between is what you have to check.

If there is no controlled flow that is coming in between which is true in this particular case, there is no reservoir here and therefore, there is no addition that is coming here. Then it may indicate that, in this small catchment there is a significantly higher rainfall that is happening here, or a higher rainfall in the intervening catchment; if neither of them is true then, there is a inconsistency of the data. This is the conclusion that you arrive at.

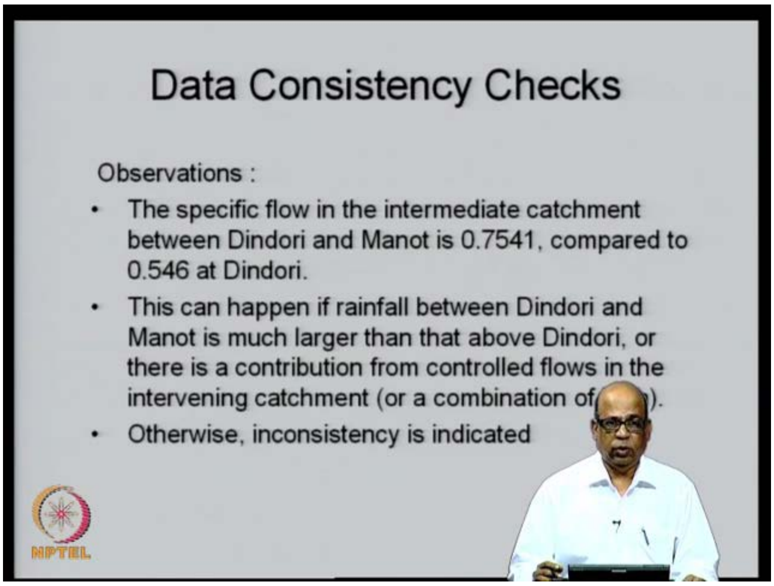
Similarly, you look at Bijore and Jamtara, now Manot and Bijore there is straight forward thing, there is Manot here at this location, and Bijore is at this location, there is a Bargi reservoir in between. So, Manot and Bijore, there is one reservoir here. So, if you look at Manot and Bijore, the upstream site is Manot, it has a specific flow of 0.6519. The downstream site is Bijore, which has a flow of 0.5984.

And intervening catchment when we apply that method, you get (( )) which is quite close to 0.59; and therefore, there is no problem and it is quite consistent with what we are seeing here. However, when you are going on to Bijore and Jamtara, now Bijore is here, Jamtara is here just almost close to each other.

Between Bijore and Jamtara you see again that, the intervening catchment specific flow is much smaller compared to the specific flows at either of these stations, Bijore or Jamtara. Now, what does this mean? It means that, either there is a significant utilization or significantly lower rainfall in the catchment above Jamtara, or both of these are true. That means you are using the water and therefore, the intervening catchment is yielding less or there is significant significantly lower rainfall in that particular catchment.

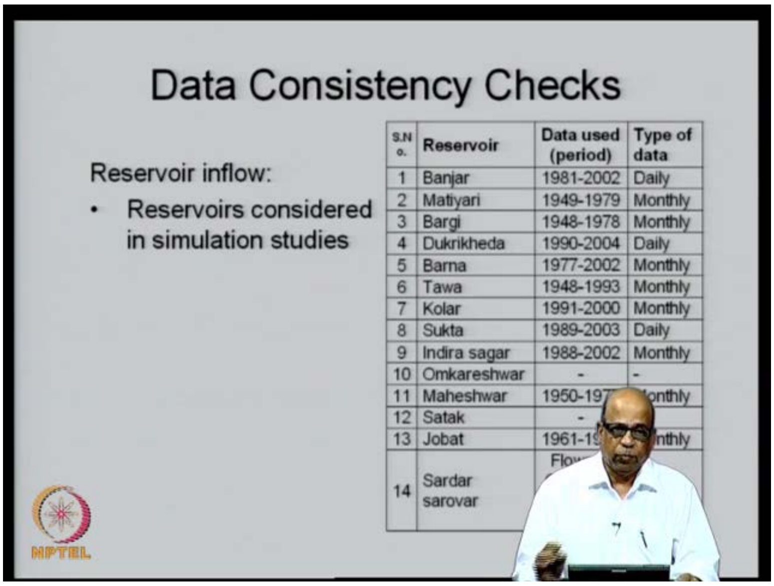
And therefore, you are getting a smaller specific flow in the intervening catchment, or both of them may be true. If both of these are not true which means from the data you examine this and ensure that, there is no significant utilization, if both of them are not true then you have to say that, the data is inconsistent. So, this is what we do using the specific flows.

(Refer Slide Time: 13:29)



So, from the data consistency checks we observe in this particular case that, the specific flow in the intervening catchment between Dindori and Manot is 0.7541, compared to 0.546 at Dindori, which is quite high. This can happen as I just mentioned, if the rainfall between Dindori and Manot that is the catchment area between Dindori and Manot is much larger than that above Dindori, or there is a control flow that is coming may be there is a small reservoir or some such thing which is letting in water. And that is why the intermediate catchment specific flow is larger, if neither of them is correct from the data then inconsistency is indicated.

(Refer Slide Time: 14:20)



Now, similar to what we did for the gauged flow, we also do for the reservoir flows. Remember, these data consistency we are doing because, we are dealing with large systems where there are number of reservoirs, there are also a number of stream flow gauges; and we want to be sure that, the time series that we consider for this analysis using the stochastic hydrology techniques are in fact, consistent with each other. And that is why we do the data consistency test first.

So, much the same way as we did for the rain gauges, as the stream flow gauges, we also do the same exercise with the reservoir flows. So, these are the type of data that you have; and we consider typically the monthly data and for the specific flows, we consider the annual or the seasonal data.

(Refer Slide Time: 15:18)

Statistics of Annual Inflows

S.No	Gauge site	Data used (Period)	Duration (years)	Annual Average (MCum)	Maximum (MCum)	Minimum (MCum)	Standard deviation (MCum)	Coeff. of variation (%)
1	Matiyari*	1949-1979	31	80.43	168.17	23.43	32.25	40.10
2	Bargi*	1948-1978	31	7,392.65	15,430.00	2,152.00	2,957.96	40.01
3	Barna#	1977-2002	26	500.12	1,208.03	67.14	269.11	53.81
4	Tawa	1948-1993	46	3,768.41	9,444.75	1,787.68	1,721.83	45.69
5	Kolar	1991-2000	10	219.09	470.17	78.34	119.71	54.64
6	Sukta	1989-2003	15	71.03	98.81	32.95	22.82	32.13
7	Indira Sagar+	1988-2002	15	10,594.85	23,737.80	4,036.20	5,854.42	55.26
8	Maheshwar	1950-1977	28	27,822.55	56,125.10	11,298.90	9,454.72	33.98
9	Jobat#	1961-1980	20	299.49	807.10	39.20	203.16	67.84
10	Sardar sarovar	Flows at Garudeshwar will be used	-	-	-	-	-	-

And as I said any data set first you look at the statistics, annual average, maximum flow, minimum flow, standard deviation, coefficient of variation etcetera you just compute that. This will tell you the story of what is what can be expected from this particular system. So, each of them has its own statistics here.

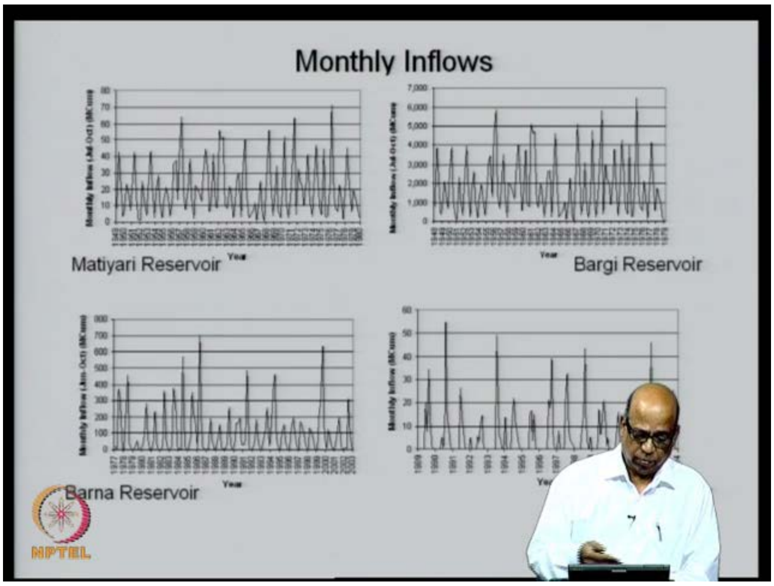
See for example, Matiyari which is a small reservoir on the tributary it has an average annual flow of 80.43 compared to that at Bargi, which is 7392 million cubic meters. So, this will give you an idea of the magnitude of the difference that you can expect.

If you as you come down these are major river Narmada, as you come down look at Maheshwar here, it is 27822 million cubic meters, the average annual flow which is far far higher compared to the 80.43 million cubic meters that we are looking at at any of these tributaries. Similarly, Jobat is a small tributary which has a 299.49 and so on.

Now, you should as as students of stochastic hydrology, you should develop a feel for the physical aspects of these flows. Why, this is smaller compared to this? Immediately we should be able to say that, this is of much downstream reservoir which accumulates flow as it comes from upstream to downstream; and therefore, the flows are much much higher here. Sardar Sarovar I have not shown here, but in a Sardar Sarovar will be much larger compared to even Maheshwar; and Indira Sagar is also quite large compared to Matiyari and so on.

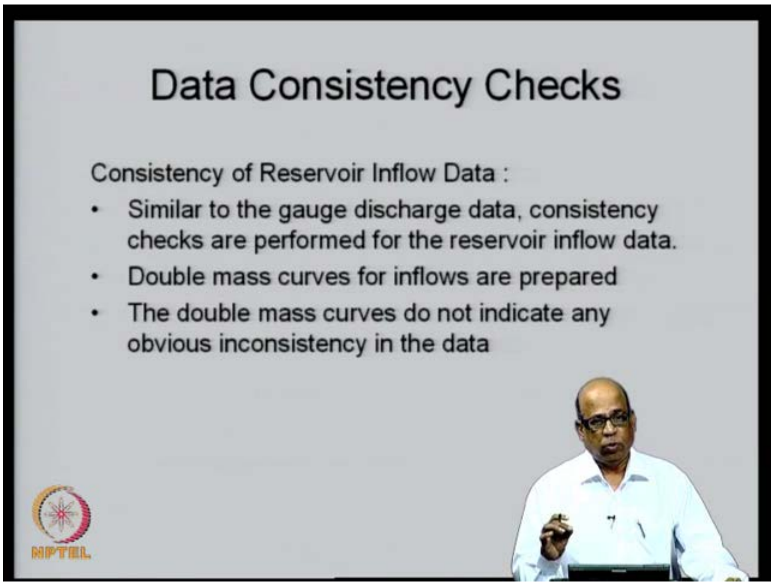
So, your maximum flows follow a similar trend, minimum flows follow a similar trend and so on. So, the first level exercise as I have been repeatedly telling is just to look at the statistics, compute all the overall statistics of the flows, typically when we are dealing with hydrologic designs we will be dealing mostly with the flows; and then get a feel for what you can expect from that particular system.

(Refer Slide Time: 17:44)



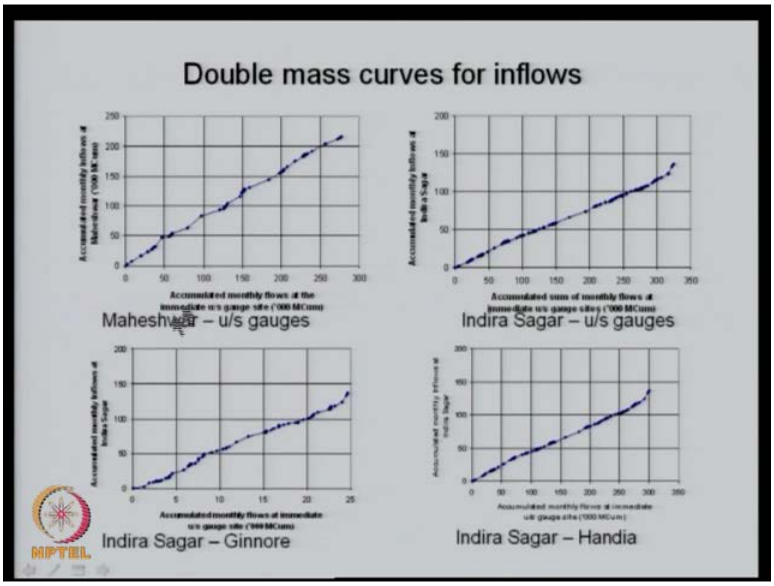
And then we do as as I mentioned and perhaps, we have discussed this also through other case studies in the time series analysis topics, just plot the time series. The moment you get the data in in the form of flow recorded at various time durations, simply plot the time series; that itself will tell you how the variations are happening in the series. Then we do the data consistency checks.

(Refer Slide Time: 18:11)



I am now talking about the reservoir inflow. Similar to the gauge discharge data, we examine the data consistency of the reservoir flows. So, what do we do? First you look at the particular reservoir and then, look at all the upstream gauges including the upstream reservoirs; and then see whether the data that is recorded at this particular location is in fact, consistent with the data that is recorded at the upstream points. For doing which, we will also use the double mass curves and this is how the double mass curves look at.

(Refer Slide Time: 18:50)



For example, you are looking at the Maheshwar, all the upstream gauges of Maheshwar here and then the accumulated flows at Maheshwar itself; similarly, Indira Sagar all the upstream gauges and the accumulated flows; and between Indira Sagar and Ginnore and the upstream gauges here; and Indira Sagar and Handia.

Here, we are looking at all the upstream gauges, here only one upstream gauge with respect to Indira Sagar, another upstream gauge with respect to Indira Sagar and so on. So, like this we do several types of consistency checks with respect to several different combinations of the upstream gauges.

As you can see, most of these will indicate consistency there is no inconsistency indicated at all from the double mass curves of inflows at particular reservoir; this I had just shown as an example. The same exercise we carry out with respect to other reservoirs also. So, this is what we do in data consistency checks.

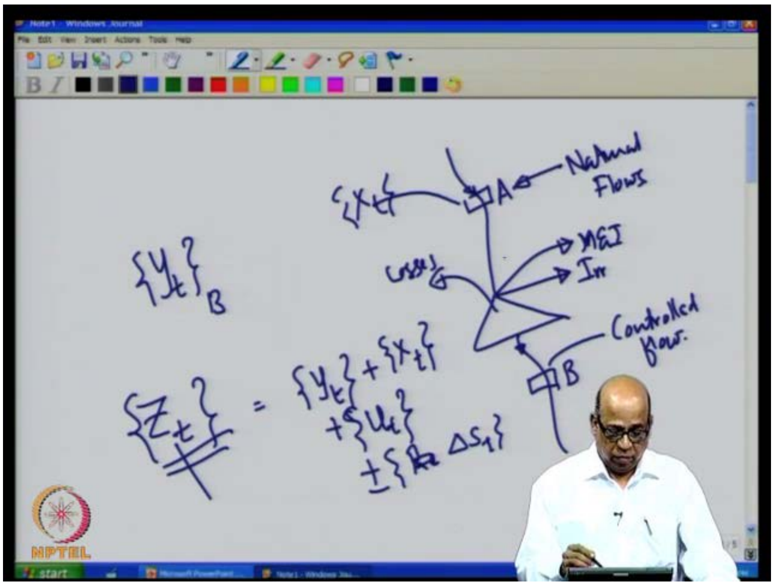
Now, still on the data, there are other issues that you must be alert too. As I have been mentioning, the data as typically the flow data that we use for analysis must be normalized or naturalized flow data.

Typically, you know in in countries like ours you have data for last about 30 years, 40 years and so on, the actual measured data. But, in the last 30, 40 years and even beyond, there have been major anthropogenic interventions in the rivers. For example, there are being dams built, there have been lift irrigation schemes that have been built, there have been barrages, major anthropogenic changes have taken place on the river systems.

And therefore, the data that we have will contain the signals of these anthropogenic changes what I mean by that is. Let us say that, you have a gauge and the gauge data has been made available to you and this gauge is just downstream of a reservoir that has been constructed let us say 20 years ago, and you have the data at this gauge for the last about 60 years let us say; out of these 60 years of data, the last 20 odd years may be 10 years for construction plus 20 years, since it has been in operation.

So, the 30 years of data the most recent 30 years of the data is in fact control flows, it reflects the control flows. And therefore, we must have a means of naturalizing the flow sequence at this location. So, this we will discuss briefly; so, that we will know how to normalize the the flows.

(Refer Slide Time: 22:11)



What we are discussing at this point is let us say that, you have a gauge I will call it as gauge B and then there is a gauge A upstream of the reservoir; from the reservoir, you are letting out water in a controlled manner you have let us say gated spillway and then, you are letting out water in a controlled manner. So, what you are seeing here is a controlled flow.

And if there are no major structures or if there are no structures upstream of A, what you are seeing here is natural flows. Now, we want to do the analysis on the data that we have obtained at gauge B, but what we have obtained the series at gauge B; let us say I call it as y t at B is controlled flows plus a little bit of uncontrolled flows that is contributed from this intervening catchment.

And therefore, we must convert this time series y t at B to a naturalized series naturalized flow series. What do we do for that? Let us say, I write it as z t which is a naturalized flow is equal to whatever we have observed y t, these are the observed flows. Now, upstream of this, there may be utilizations that have taken place; let us say that, you have put the water for irrigation then, there is also some water supply meant for municipal and industrial usage and there is some losses that are taken place. Now, if this reservoir was not there then, all of this water would have come here.

And therefore, you have to add whatever upstream utilization that has taken place upstream of this to this sequence in time period t, let us say you are talking about annual series then every year let us say t is equal to 1 to 50, 50 years sequence you want to reconstruct. So, 1 to 50 every year, how much of water that has been actually utilized, these are called as the upstream utilizations.

So, you add all the utilizations I will call that as u t which is upstream utilization. So, what we have done now is whatever that is observed here plus upstream utilization, this also includes losses. Remember, when I write u t it is utilization plus losses then, there is a change in the storage that has taken place here. So, this we will write it as plus or minus reservoir regulation or delta s t I will write.

So, if we do it for a long sequence of time, the effect of storage will be taken care of by this delta s t. So, in certain cases you may have plus delta s t, because from whatever that is coming here, you may have accumulated by releasing less amount of water; and therefore, that accumulates, in other periods you would have released more than what has come and therefore, the delta s t will be negative.

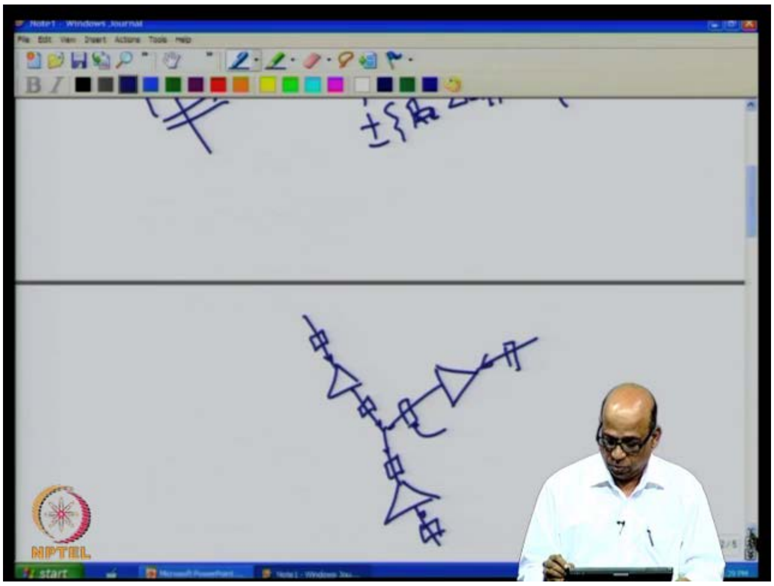
So, you will have plus or minus delta s t. So, that is how you account for the storage effect here and then reconstruct your series z t here. And it is on this reconstructed series that, you make the analysis and not on the actual observed at this particular location. Now, to this actually you have to use what has been observed at an upstream location immediate upstream location, where completely naturalized flows completely natural flows have been observed.

So, y t it should consist of what has come here at this particular point plus what has been utilized from that plus whatever plus or minus whatever storage effect that is taken place. So, I will also add to this, the flow that is naturally occurring at this particular location. So, this you have to do may be you can use the work sheet like excel or some such work sheet and then do it for a long period of time.

Typically, if you are talking about monthly flows you want to reconstruct the monthly flow sequence here; then do it for 30 years 40 years etcetera. Concurrent data should be available on all the utilizations and (( )) operating policies that you have adopted; and the gauge flow immediately downstream of this, and the storage fluctuations that you have etcetera. Reconstruct the sequence and then do all your analysis on z t here and z t will be the reconstructed time series at this particular location. And these are called as the naturalized time series naturalized flows.

So, the point that I am emphasizing over and over again is that, whenever you are adopting any stochastic hydrology techniques that we have dealt in this particular course, you must deal with the naturalized flows, if you are doing the analysis on the flows. So, if you have let us say several such structures; typically, this is a case of most existing system.

(Refer Slide Time: 28:38)

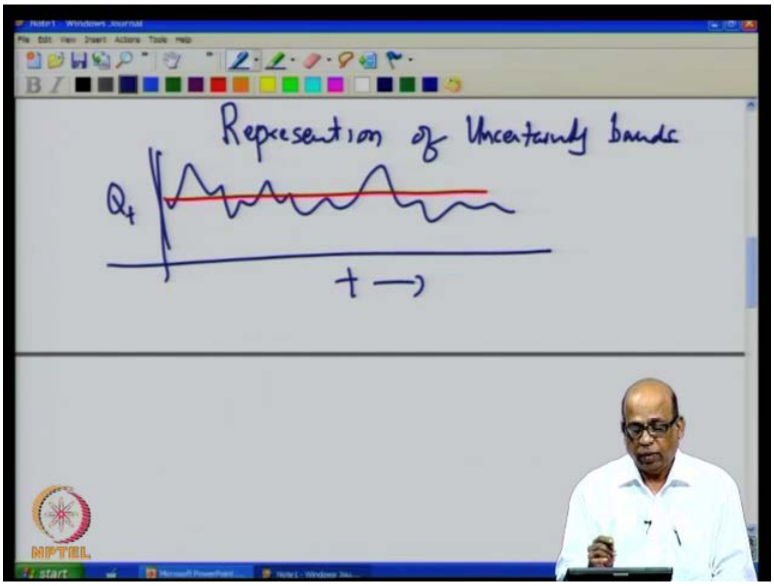


Let us say, you have one reservoir that is coming from here, another reservoir coming from here let us say you have another reservoir here and then a third reservoir coming from here etcetera. And you want to do the analysis at several occasions, let us say you have a gauge at this point, gauge at this point and so on, you have also gauges at upstream of that flow directions are like this.

Now, at all these points you must reconstruct the series to reflect a naturalized flow here, naturalized flow here, naturalized flow here and so on. Now, this is an important exercise to begin with any time series analysis that you would like to do any analysis on the data that you would like to do especially on the flows, you must make sure that the anthropogenic effects are taken out of that time series.

It is particularly pertinent, relevant in the context of climate change analysis that we do, where we use many of these stochastic hydrology techniques. It is important for us to always deal with naturalized flow sequences, which I will be dealing with in the next class and perhaps towards the end of this class on how to address the climate change effects on hydrology; so, this one one aspect of the data. There is another smaller aspect a minor aspect of representation of the data itself we will cover that, before I go to a specific applications specific recent application.

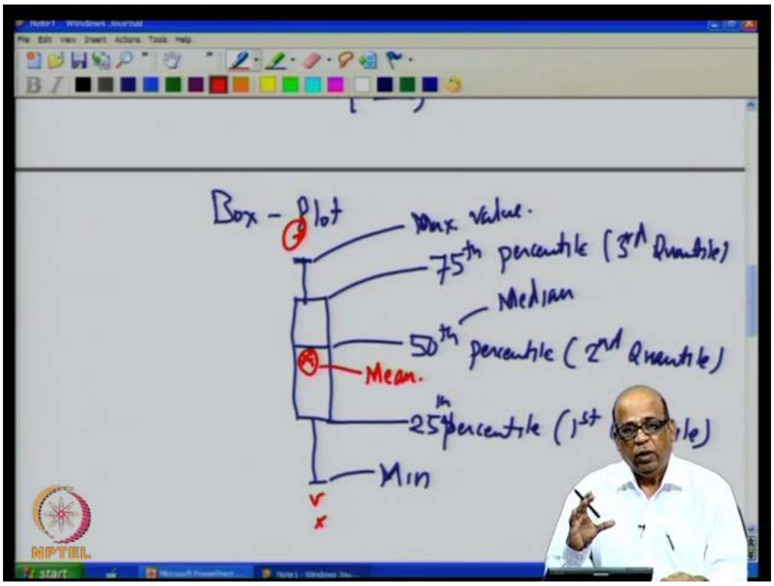
(Refer Slide Time: 30:42)



Let us say that, this is on the representation of the data that is representation of uncertainty bands. Let us say, you have a time series of flow for the last about 50 years, this is time and this flow let us say it is Q t and this is typically 50 years of some such thing.

Then you have a long term average flow like this around which the actual flows Q t are fluctuating. Now, this particular time series or the data set has a certain minimum flow, it has a certain maximum flow and it has a certain mean flow certain median, certain 25 percentile, certain 75 percentile and so on. So, to compress the information provided by this, visually we prepare what are called as the box plots.

(Refer Slide Time: 32:06)



So, the best way of representing the entire time series is through a box plot entire that is the information contained in the entire time series, we can prepare using a box plot. I will just draw a box plot and then we will see, what it contains what information it contains. Typically, we show the box plot like this (Refer Slide Time: 32:34), it is also called as a whisker plot. So, we may have lines like this, so this is the typical box plot; assumed that this is in fact a straight line, so this is a rectangular box that I have drawn here.

Now, the lowest point corresponds to the minimum of the data or the sample minimum value in the sample; the lower edge of this box corresponds to the 25 percent as 25th percentile I will call it as 25th percentile or it is also called as the first quartile. There is a line that is drawn inside the box this corresponds to the 50th percentile or the second quartile.

Now, the 50th percentile is in fact the median, so this is also the median not also it is the median. Then the upper edge of this box is a 75th percentile or the third quartile; and then you have a line, this is the maximum value. So, the time series the information contained in the time series has now been visually expressed through a box plot, you have 1 2 3 4 5, Five typical values. There are different variations of this, but typically it contains these five values.

So, this is a minimum value let us say you have 50 years time series, simply pick up the minimum value out of that, that is the minimum value; 25th percentile indicates the particular value in the time series below which 25 percent of the values lie.

And then 50th percentile, the particular value in the time series below which 50 percent of the values lie, this is in fact the median, this is called as the second quartile. Then 75th percentile, which is a third quartile 75 percent of values in the time series lie below that particular value and then you have the maximum value.

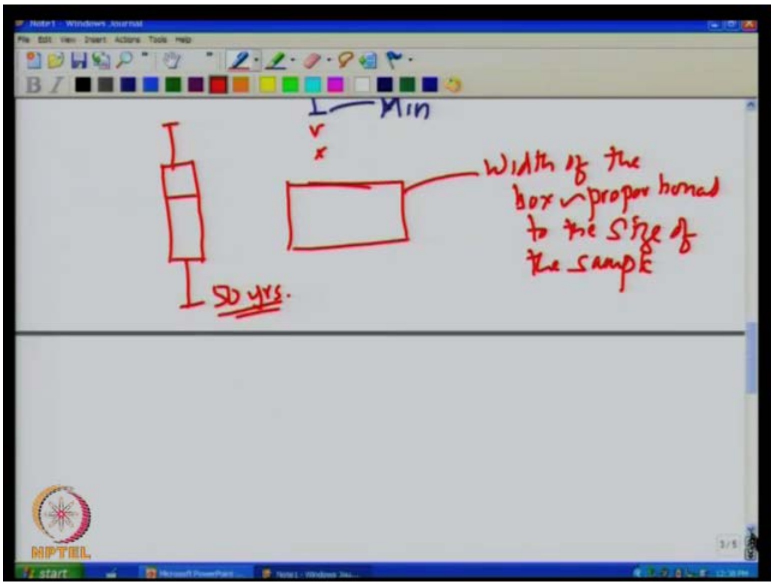
There are also certain out layers, we may also indicate certain out layers. What do we mean by out layers? These are value that, somehow do not belong to the other data contained in this time series; may be there is a 1000 years flood that has occurred in that particular sequence that you have, which means it is much much higher than the maximum value that you can expect; if you delete that particular value and then reckon only the remaining values.

So, there are techniques of identifying out layers which is slightly, these techniques are slightly beyond the scope of this particular course we have not discussed this. But, remember that out layers do not belong to the data set consisting of all other values. And in fact, in in a series of let us say 100 years of data, you may not have you may have not 1, but many out layers. So, there are ways of there are statistical ways of identifying what are these out layers, and out layers may also come on the other side.

So, this is the way we represent in a compact form the entire time series, typically using the five values. In addition, many times we also indicate the mean, this is the median and we may also have a mean here. And as I said, there are variations of representing the box plot, but box plots are generally are essentially used to represent what kind of uncertainties exist in that particular time series; because here as you can see this is a range that you can expect minimum and maximum, and this is the median that you can expect, this is a mean that you can expect and so on.

This is especially useful when you are comparing two three different time series. Let us say, you are comparing flows at a particular location with respect to flow at another location. Then once you plot this then, you will know what kind of mean, how the mean is changing, how the median is changing between the two places, what is the kind of maximum value that you can get from the two places and so on. So, the box plots are very useful in doing any analysis using the data. There is another small thing that we must know.

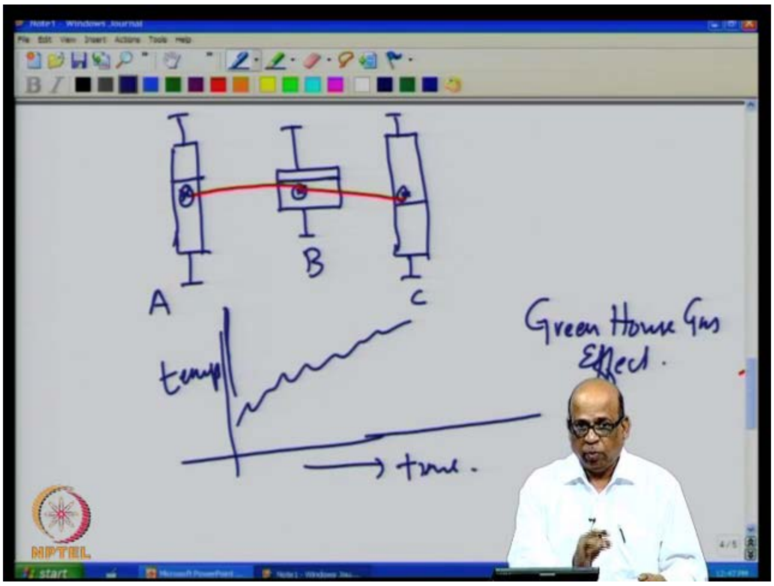
(Refer Slide Time: 38:00)



Let us say, I have a box plot like this, this box plot has been prepared with a data of 50 years and I have another series time series, the data available there is 100 years and we may have similar quartiles; we may have a maximum value, minimum value and then the 25th percentile, 50th percentile and so on; for that other data series, which is based on 100 years of data.

So, to reflect that the sample size is different in the two series, we use the width of the box itself. So, the width of the box may be made proportional proportional to to the size size of the sample. So, we may have let us say three or four sequences, why we use this analysis we will a see presently.

(Refer Slide Time: 39:31)



Let us say, I have a time series which whose box plot is something like this; we will assume that the out layers are not there now. And then you have a 50 year sequence, the 50th percentile then you will have a mean here, this is at location A or time series A let us say.

Then you have a time series B, which is of a different sample and this has a mean of this, a maximum of this and a median somewhere here and a mean here. Similarly, you may have certain other boxes box plots like this (Refer Slide Time: 40:24). We would be interested in seeing, how the mean has been changing from one point to another point. So, we simply join these let us say this is at point B and this is at point C.

So, we will be interested in checking how the mean is changing, how the median is changing, how the minimum values are changing and so on. So, that is why we plot we draw these box plots side by side to get a visual picture of how these various quartiles are changing from one location to another or from one time to another.

You may have time windows let us say 2010 to 2030 in the future you have a certain box plot then, 2030 to 2050 another box plot, 2050 to 2070 another box plot like this. When you are doing projections using a climate change or otherwise you will have time windows specifically like this 20 year time window. And 20 year time windows, these may be varying of course the width of the box in that case will be same, because your sample size will be the same.

So, the box plots are essentially visual depiction of the entire time series, how the time series is likely to how the time series is in fact with respect to various quartiles. If the construction of this is pretty simple, you get the minimum value extracted from that 25th percentile value, which is that particular value below which 25 percent of the values lie. So, you arrange them in a decreasing order and then pick up the 25th percentile value and so on. Mean you know how to compute, median you know how to compute and so on.

So, given a data series given a time series you know how to compute, how to get these box plots. And box plots are in fact good tools in most time series analysis. So, on the data we have done the data consistency checks and then we have also just seen how to represent the uncertainties in the data using the box plots.

There are variations of these box plots. For example, you may not necessarily have the first quartile or you may not necessarily have only these six values, you may also have certain other quartiles as as deemed important in the particular analysis. So, the remaining part of this lecture small part of this lecture and the next bulk of the next lecture, I will devote on the most recent applications of stochastic hydrology.

We have seen several applications earlier for example, data generation; we have seen how to fit the probability distributions and then make inferences from the probability distributions; how to use the information contained in the data for hydrologic designs and so on. The most recent and a burning issue in hydrology has been the issue of climate change.

I will just give a brief introduction in today’s class on what are the issues related with climate change. And then we will see how to address the climate change issues for hydrologic designs using the techniques that we have learnt in this particular course. And specifically I will deal with what is called as the statistical down scaling technique. What is this climate change issue? Most of you would be aware that, the global averaged temperatures have been rising.

So, if you emphasize if you look at the global average temperatures or the last about 1 century or something on certain scale this looks like this. So, these are the temperatures, this is with respect to time. Now, this is known to be or it has been said that, these rise in temperature has been because of the green house effect green house gas effect; because the temperatures are changing in in a rapidly increasing manner, this is global average temperature you know locally it may be not necessarily following this particular trend; but, global average temperatures based on a large number of measurements all across the globe or a last century it has been shown that, the temperatures are continuously increasing; because, the temperatures are increasing they have effects on the atmospheric circulation and because of which the precipitation patterns are changing all across the globe and also the sea levels are rising.

So, there are three major visible signals of climate change namely that, the global average temperatures are changing, the sea levels are rising and the precipitation patterns are changing. All the three of these have a direct implication on most of the hydrologic processes. As the temperatures change, a direct implication of that will be on the evapotranspiration, directly evapotranspiration may start increasing, if all other the influencing variables are held constant.

Of course, the global average temperatures will also have implications on several other processes, but direct implication will be on the evapotranspiration. And evapotranspiration has implications on the irrigation demands, crop water demands. The sea levels increase will have implications on the salt water intrusion into the lands apart from having implications on the coastal inundation. That is the sea levels will start ingressing into the main lands. And therefore, there is a coastal inundation that is possible.

The precipitation patterns changing will have implications on flooding, there may be high intensity, short duration in rainfalls increasing; and therefore, it will have implications on the flooding. Also it may have implications on the droughts on the water level amenity, because precipitation will contribute to water level amenity.

So, the question that is burning in hydrologic sciences today is, how do we assess what is likely to happen to let us say the water availability in a particular location with respect to climate change; as the climate is known to be changing rapidly in a very short time, what is likely to happen to water level amenity, what is likely to happen to the floods flood situation, what is likely to happen to the drought situation. These are the hydrologic extremes that we have seen in this particular course.

So, there may be deficit water, there may be excess water, because of which the floods flood disasters are caused and so on. Similarly, what is the implication on evapotranspiration, which directly determines the consumptive use of water. The ways of assessing the impacts of climate change on hydrology and water resources in a particular region are burdened with a large amount of uncertainties, as I will discuss in the coming lecture next lecture and perhaps part of this lecture.

Because, we are standing at a particular point in time now and then, we are looking ahead into the time what is likely to happen to let us say Narmada river in the 2020's 2040's 2060's etcetera. So, we have to look beyond into the future. Our conventional way of analysis would have been look at the historical data, fit probability distributions and then hold these probability distributions sacrosanct intact and then project it into the future.

All the time series analysis methods that I have discussed in this particular course are in fact based on this premise that, the probability distributions of the various processes that we have studied let us say stream flow etcetera. You fit the probability distributions based on the historical data hold the probability distributions intact sacrosanct and then, use the same probability distributions for extending into the future.

The data generation techniques that I discussed for example, one step Markov chain model the which is also Thomas (( )) model or the arma type of models etcetera, which we use to generate the data into the future; let us say 50 years into the future, 100 years into the future, which are all important for hydrologic designs.

They are all based on the historical data. We have collected the data at a particular location for the last 50 years, we use all our regress analysis techniques that we have gone through this particular course, fit the probability distributions and then use the data generation techniques; to arrive at how the system is likely to behave into the future. This based on the premise of stationarity hydrologic stationarity, why stationarity? Because we are saying that, the future is likely to be the same as the past in some statistical sense in some probabilistic sense. And therefore, we use the historical data we use all our techniques and then, make an assessment of how the future is likely to be.

So, the premise in all the statistical methods that we have discussed in this course is that, the history provides a valuable clue to the future; how the future will likely to be will be assessed based on how the history has been. This would have been perfectly fine, if there was no climate change issue.

The single most challenge that the climate change has thrown to the hydrologist is, one of non stationarity. Suddenly, we are not sure how the future will be, just based on we cannot be sure how the future will be, just based on the historical data. The future, they say in climate change is not likely to be the same as the history or the past. So, the past no longer provides a clue to the future.

And therefore, we must have ways and means of looking at how the future might be going beyond the historical data. So, we must have a methodologist developed of looking at how the future will be, in spite of the story that the historical data tells us. And this is what we do using the climate models, we rely on the climate models global climate models or the general circulation models, which provide us the simulations of how the climate will be in the future.

And then we use those simulations and relate it with hydrology of flows at a particular location, rainfall at a particular location and so on. So, we relate them using the stochastic methodology stochastic hydrology methodologies that we have dealt with, and then project provide projections of the future hydrologic scenarios.

These are extremely important exercises in hydrology. And they have large implications on how large water resources systems perform. The example of Narmada I have been talking about in this class and previous class, these are large water resource systems, the way they operate in terms of hydro power generation, in terms of irrigation, in terms of the municipal and water supply demands etcetera, will depend on what kind of flows that can be expected, and what kind of demands that can be expected from this; which are all related to the climate through precipitation through temperature and through other means. And therefore, we must have methodologies to assess how these systems are likely to perform into the future. And this is a extremely important and timely topic that, we deal with in stochastic hydrology.

So, we will start with in today’s lecture, I have just about 1 or 2 minutes now. And in the bulk of the next lecture, I will deal with this issue of relating the Global Climate Model simulations or the G C M simulations. G C M’s are also General Circulation Models, which are developed by atmospheric scientists.

We take the simulations provided by the G C M’s and then relate them with the hydrologic simulations to provide future projections for various hydrologic processes; like rainfall at a particular location or stream flow at a particular location and so on. So, this is what we call as hydrologic impacts of climate change (Refer Slide Time: 53:58).

And then when we are dealing with this particular topic, the uncertainty is a major issue there and quantification of uncertainty becomes important. So, we will see various techniques that are available to assess the impacts of climate change. We continue that discussion in the next class, which is lecture number 39. So, I am just coming to the close of this course and therefore, this advance topic I will deal with in in a compact form in 1 lecture I will just deal with the scale issues, as well as quantification of uncertainties in hydrology in the context of climate change.

So, essentially in today’s lecture, we continued our discussion on data consistency checks using the double mass curve, as well as using the specific flow concept. And then I dealt with representation of uncertainties in the data through the box plot; and also briefly touched upon naturalization of flows, when you have flow measured at a particular location, if it has contributions from controlled flows arriving out of reservoirs or barrages etcetera. How do we construct a series, which represents naturalized flows; which means, we are removing the effect of the anthropogenic interventions through reservoirs barrages and other constructions.

And then finally, I have just given a brief introduction to the climate change issues, which are extremely important issues in stochastic hydrology today. So, we will continue this discussion in the next class, and there I will deal with specific techniques. We typically use the techniques that we have discussed in the course all through; we may use multiple regression, we may use principal component analysis, we may use some stochastic models to relate the climate change simulations with respect to the the hydrology, we may use stochastic down scaling technique down disaggregation techniques from larger time to smaller time scales and so on. So, all of this I will give a just a broad flavor in the next lecture, thank you for your attention we will continue the discussion.