Stochastic Hydrology

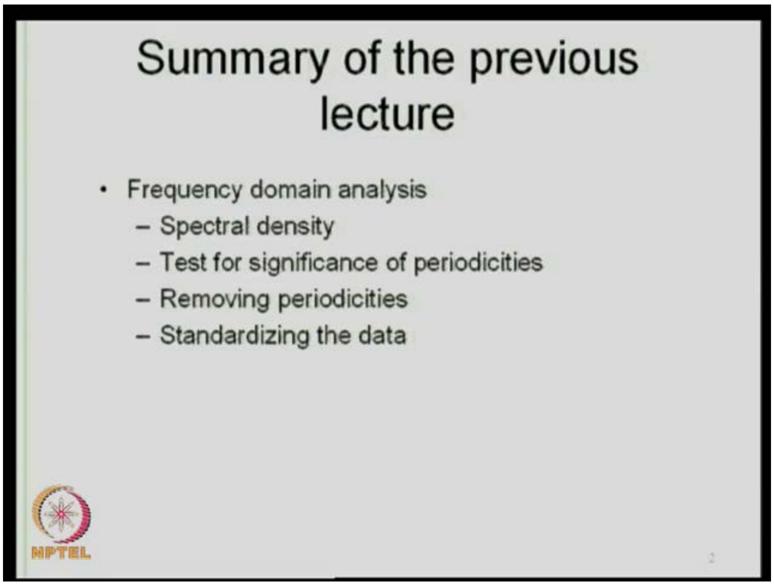
Prof. P. P. Mujumdar
Department of Civil Engineering
Indian Institute of Science, Bangalore

Lecture No. # 14

Frequency Domain Analysis - II and ARIMA Models - I

Good morning, and welcome to this the lecture number fourteen of the course stochastic hydrology.

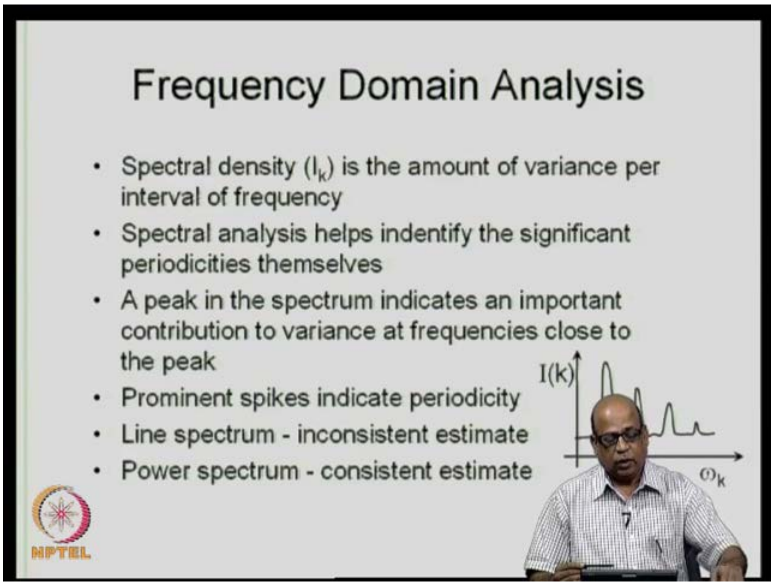
(Refer Slide Time: 00:26)



If you recall in the last lecture we covered essentially the details of frequency domain analysis. We introduced in that course how we convert the data in the time domain to frequency domain, and we introduced the concept of spectral density. And from the spectral density diagram, which is typically called as a line spectrum or smoothen spectrum is also called as the power spectrum. From this we identify periodicities inherent in the data, and then we also introduced the test the statistical test for testing, which of these periodicities that we identify through spectrum analysis or in fact statistically significant. And then once we identify the statistically significantly periodicities.

We also examined how we remove the periodicities from the data, and then in that process we are also introduced a method of standardizing the data. We examined what happens by standardizing the data, how the correlogram looks and how the spectral density or the line spectrum or the power spectrum looks for the standardized data visa v those form the original data. Towards the end of the last lecture we were discussing a numerical example of estimating the power spectrum and identifying the periodicities.

(Refer Slide Time: 02:18)

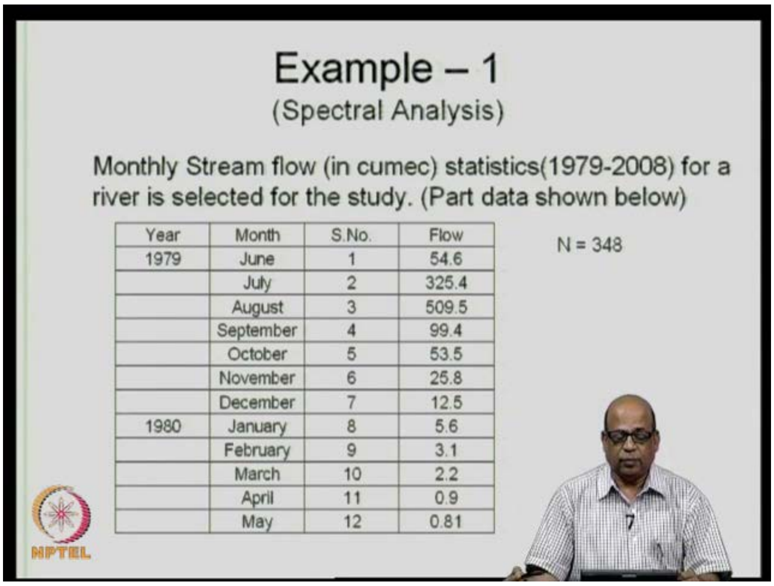


So, we will continue with that example. So, we considered the data for monthly stream flows at a location in that example. So, we will first recapitulate significant features of the frequency domain analysis, as I just mentioned spectral analysis helps us identify significant periodicities in the data. The correlogram gives us an idea that s periodicities are present in this data and then the correlogram the spectral analysis helps us identify which of these periodicities are in fact significant. The spectral density as we introduced it is also called as variance spectral density; it gives the amount of variance per interval of frequency.

A peak in the spectrum are the spikes like this for example, you may get spikes like this. These spikes correspond to periodicities in the data, the w value the omega value that we get corresponding to these peaks here they correspond to the periodicities in data. The lines spectrum as we introduced in the last lecture is an inconsistence estimate statically.

Whereas, the power spectrum or the smoother spectrum which is shown here are is a statistically consistent estimate.

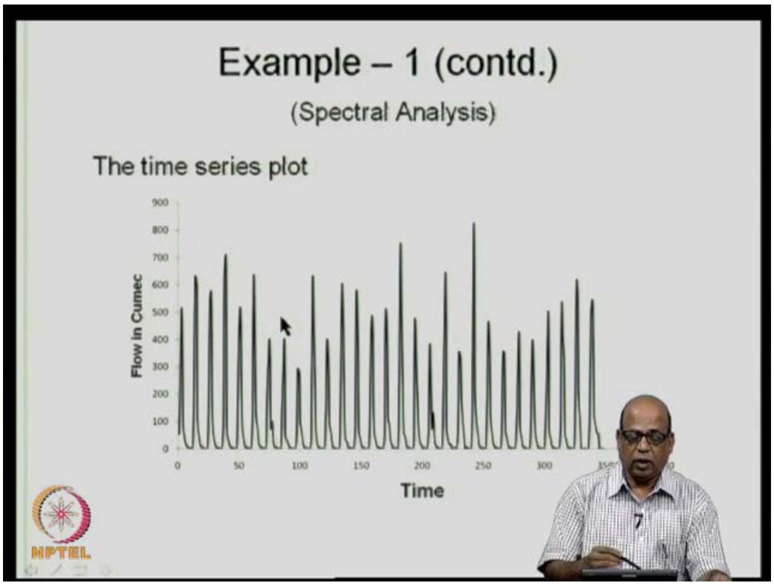
(Refer Slide Time: 03:56)



So, we will continue with the example that we were discussing in the last class, we are considering the monthly stream flow which is given in comics these are available for 1979-2008. So, there are total of 348 values only just a few values are shown here just to give you an idea of how the time series is arranged. For example June, July, August etc. It keeps on going until May of the next year and we have the corresponding flows, like this 1 to 12, it keeps going until N is equal to 348.13. For example, correspond to the month June.

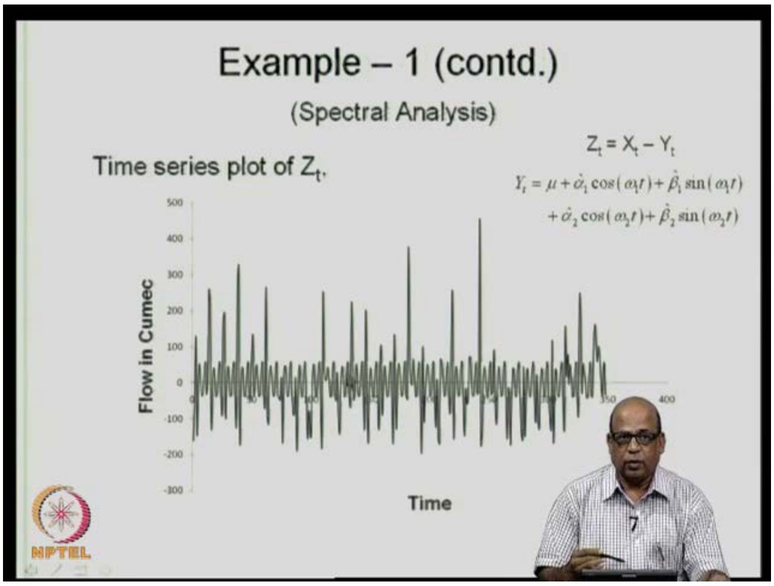
Similarly, 25 correspond to month June etc. So, the time P here goes from 1 to 348, but there is also an associated month index June, July, August etc, which is important when we go to reduction.

(Refer Slide Time: 05:04)



We will revisit this again until that point. So, we had a time series plot. So, you have 348 right values for the time series the moment you plot, you will realize that there must be a periodicity, because your significant peaks here and series itself at regular interval.

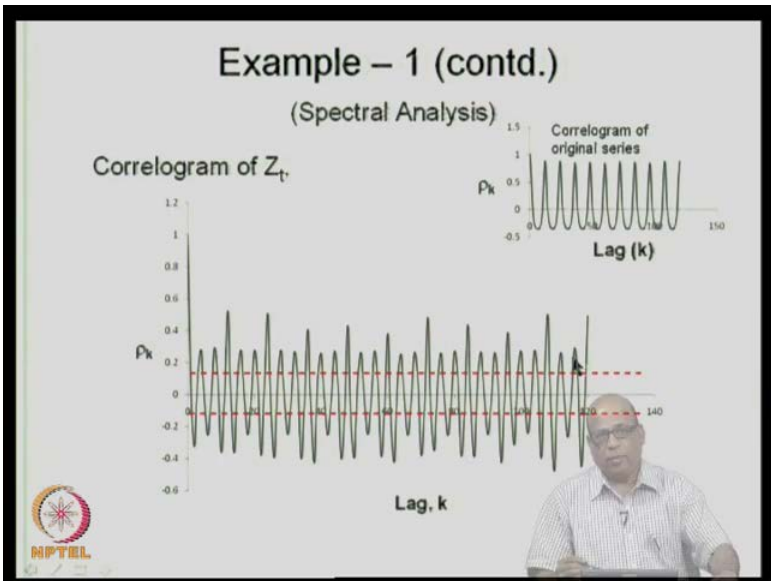
(Refer Slide Time: 05:31)



So, this information we look at more closely through the correlogram and spectral analysis, I have shown in the last class. Let us see what happens to the time series, when we standardize at t. If you recall what we did in the last class is last lecture is that we identify periodicities in the data through spectral analysis. We saw that there was there were several peak corresponding to 12 months corresponding to 6 months, 4 months, 3 months and so on. So, we wanted to examine what happens to the series when we remove the first two periodicities from the data. So, we reconstruct the time series as Z t is equal to X t minus Y t where X t is your original series and Y t is the series corresponding to the first two periodicities.

So, this is in the frequency domain the first two periodicities are here alpha 1 and beta 1, alpha 2 and beta 2, omega 1 and omega 2. So, this term corresponds to the first two periodicities in the frequency domain. So, we reconstitute the time series in the frequency domain as Z t is equal to X t minus Y t remove the first two periodicities and plot the time series. Look at how the time series looks we saw we the original time series. The original time series is here this is your X t and this is your Z t. So, the regularity of the peaks that we were seeing has not diminished.

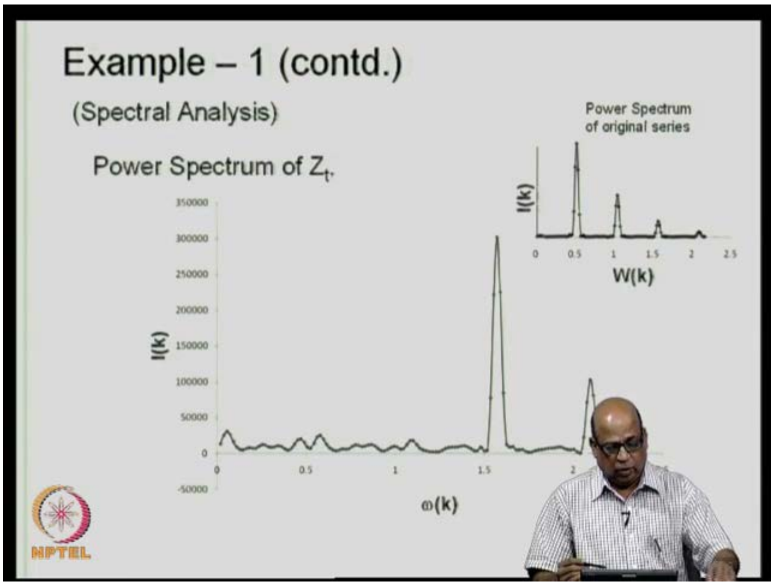
(Refer Slide Time: 07:24)



So, this is when we removed the first two periodicities. Let us see how the correlogram and the spectral density of this new series looks the original data had a correlogram. Like this indicating presence of a significant periodicity as you can see the correlogram is oscillating regularly and then if you take it further. Let us say you have 348 values may be if you go up to 200 or something, you will see that it is slowly decaying as time progresses with respect to lag the correlogram slowly decays.

For the Z t if you plot the correlogram here the significance levels were somewhere around point 1 here. So, most of the correlations were significant statically significant, if you plot the correlogram of Z t from which the first two periodicities have been removed this is how it looks most of the correlations are now insignificant. However there are some significant correlations at regular intervals as you can see. So, you again suspect that there may be some periodicities still inherent in this Z t; that means see one after you removed the first two periodicities there may be some periodicities still inherent in the new series Z t.

(Refer Slide Time: 08:53)

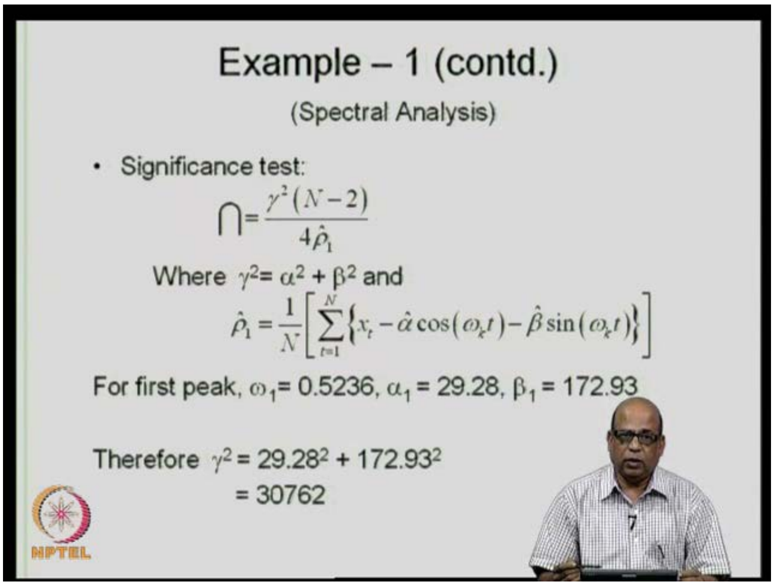


To examine this we plot the spectral density of the new series Z t. So, let us see how it looks as you can see here the periodicities which were inherent. This is a power spectrum for the original data; this is the power spectrum for the Z t, which is a transform data in from which the first two periodicity from the original data have been removed. So, you had a periodicity of somewhere near 0.5 and somewhere near 1.1 or something those are absent here. These periodicities have been removed and the periodicities corresponding to the 4 months and 3 months, this was for 12 months, this was for 6 months and 4 months and 3 months these became prominent now.

So, somewhere around 1.6 or something you had a spike here in the original data, that becomes prominent and similarly somewhere around 2.2 or something you had peak here and that becomes prominent. So, like this when you remove the periodicities, now you can remove this periodicity from Z t using the same transformation that we use on X t to determine Z t similarly on Z t you remove this periodicity then you will see that this periodicity becomes more prominent, and then if there are any other periodicities downstream of that that will become prominent and so on.

So, this is how you will examine visually whether you still have any periodicities present in the data. Now whether the periodicities that you have just determined is significant or not is a different story that we will take it with statistical test, but that this periodicity is in fact present in the data itself becomes apparent. Once you remove the periodicities that you have already identified earlier.

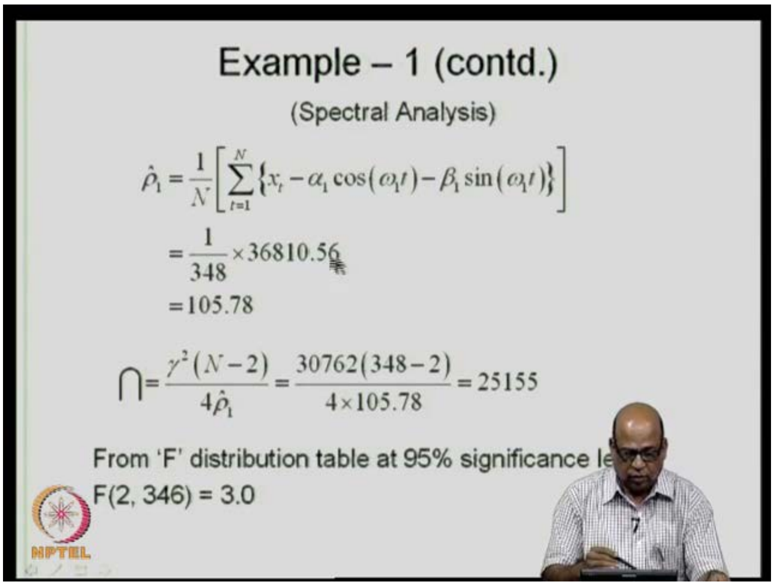
(Refer Slide Time: 10:46)



So, now we will do the statistical test to examine whether a particular periodicity is significant or not in the original series. When we plotted the power spectrum you had a peak corresponding to a w value of 0.5236. Now this is 0.5236 and we have seen that for the monthly data this periodicity corresponds to a periodicity of 12 months, simply 2 pi by omega p. Now, we will examine whether this is statistically significant or not to do that what do we form the statistic gamma square N minus 2 by 4 rho 1 cap. Where rho 1 is estimated or rho 1 cap is 1 by N, X t which is your original data alpha cap which is the alpha corresponding to the particular periodicity that you are examining.

So, for the first peak we had omega 1 is equal 0.5236 and alpha 1 is equal to 29.28 and beta 1 is equal to 172.93, which we had estimated in the last lecture. So, these are the values that we had towards the end of the last lecture. So, this omega is for that particular k that you are examining. So, you are examining for the first periodicities. So, you will take omega 1 and beta and alpha correspond to the same periodicities. So, alpha 1 and beta 1 you take. So, you can estimate rho 1 cap, N is the total number of values that you have in this case it will be 348.

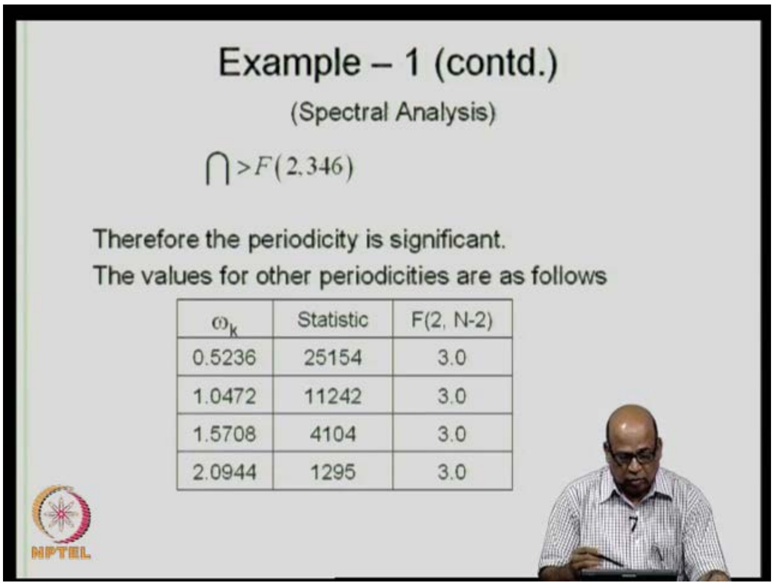
(Refer Slide Time: 12:51)



So, gamma square here is simply alpha square plus beta square you will get alpha square plus beta square is equal to 30762 and rho 1 cap you estimate from this. So, rho 1 cap is equal to 1 by 348 and all of this term together will be 36810. So, it is 105.78 and from that you estimate this is statistic gamma square N minus 2 by 4 rho 1 cap. So, N is 348, you will get 25155 with this value you go to the F distribution tables, which are available in any standard text books corresponding to 95 percent value 95 percent significance level. You get F with 2 degrees of freedom and for N minus 2 as 346; you will get a value of 3.0. If the value of the statistic that you have calculated is more than 3.0, it indicates that the corresponding periodicity in this case the first periodicity corresponding to omega 1 is in fact statistically significant.

So, in this particular case 25000 is far ahead far more than the corresponding F value here for 95 percent significance level and therefore, the periodicity corresponding to your value of omega 1 of 0.5236. Which corresponds to a periodicity of 12 months is in fact statistically significant.

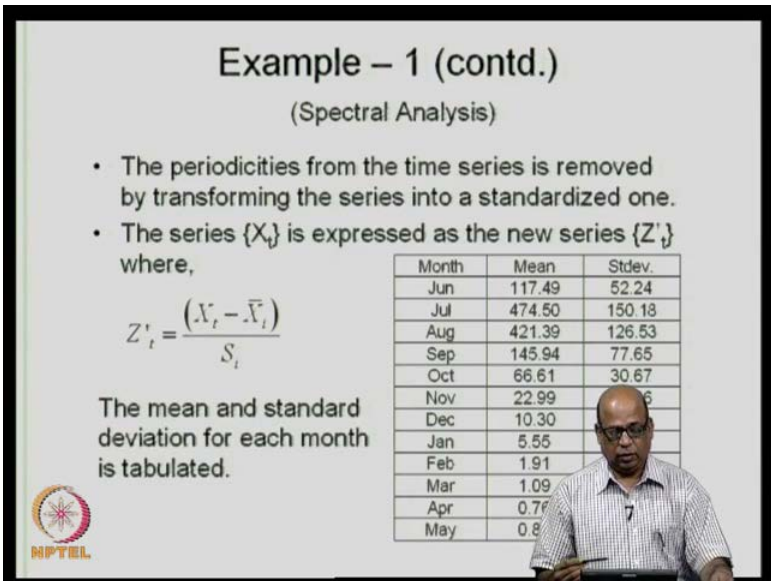
(Refer Slide Time: 14:24)



So, the periodicity that we have just examined is statistically significant. Similarly we do it for other periodicities as you can see here after the first peak you have a second peak corresponding to this value of omega 1, omega 2 and this value corresponding to omega 3 etc. So, these things also we examine one at a time remember here this statistic is written only for a particular value of omega. So, every time you test one periodicity at a time. When you do that you will get the values of statistic like this 11000, 4,000, 1295 and so on.

So, all of these are much above the critical value of F and therefore, they are all statistically significant. Now having identified that this periodicities this correspond to 12 months, this to 6 months, this to 4 months and this to 3 months having identified that these periodicities in the data are statistically significant. We now want to see how to remove these periodicities from the data. As I mentioned in the last lecture one simple way of doing this is simply remove the periodicity by standardizations simply standardize the data. That is Z t is equal to X t minus x bar over sigma or X t minus mu over sigma. Let us see what happens to the original time series plot correlogram and the spectral density when we remove the periodicity by standardizing.

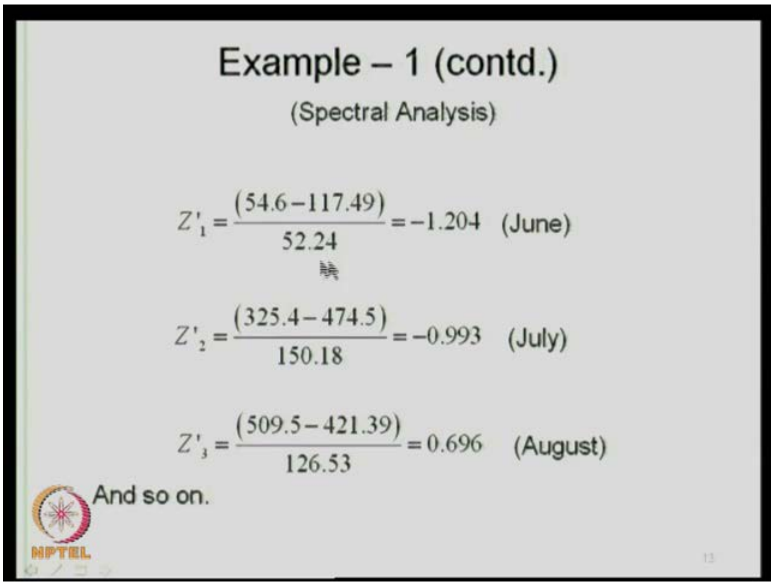
(Refer Slide Time: 16:13)



So, let us look at the standardized time series. So, what we are now doing is we are transforming the original series into a standardized one. We write because we have used the notations Z t earlier in the same problem. So, we will write this as Z dash t which is a standardized time series X t minus X i bar over S i then i here corresponds to the particular month to which t belongs. As I mentioned earlier t goes from 1, 2, 3, etc up to 348. Which are the total number of values, but correspond to every t we have an i which identifies it with the particular month for example, t is equal to 1 corresponds to i is equal to 1, t is equal to 3 corresponds to i is equal to 3, t is equal to 13 corresponds to i is equal to 1, again t is equal to 14 to i is equal to 2 and so on.

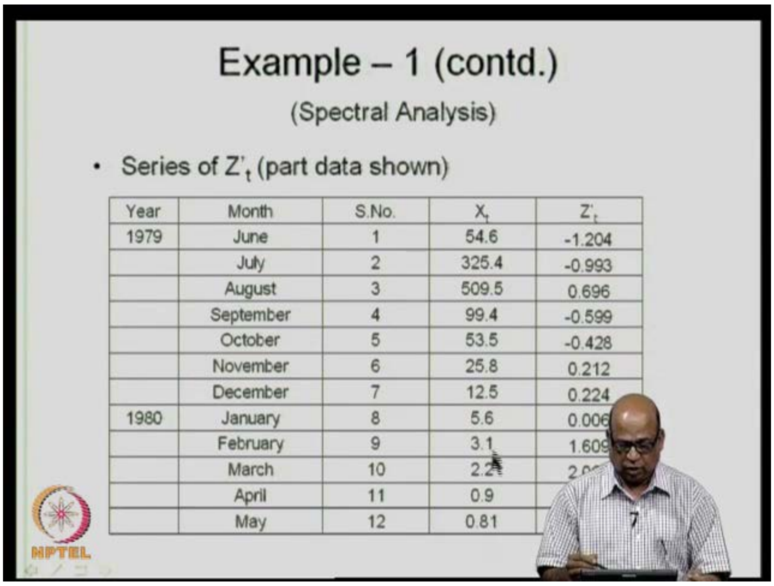
So, we identify the month to which the time period t belongs and deduct the particular mean of that particular month June has its own mean. So, you deduct the mean of June when you are taking i is equal to 1 and so on. Similarly, the standard deviation corresponding to that particular month.

(Refer Slide Time: 17:47)



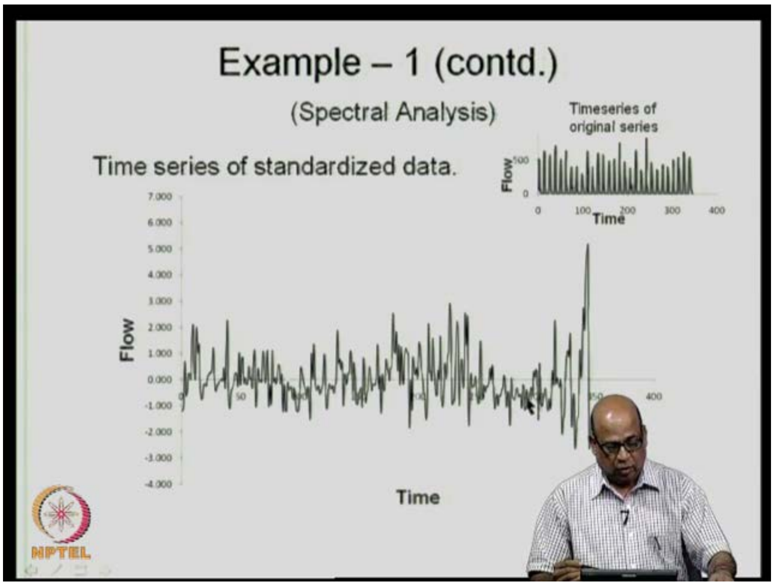
So, we formulate Z dash t by this particular transformation. So, the original time series X t is now transformed into Z dash t and for example, you are looking at the first value X 1. For the first value you deduct the mean of June month and standard divide by standard deviation 52.24 and so on, for this month June.

(Refer Slide Time: 18:30)



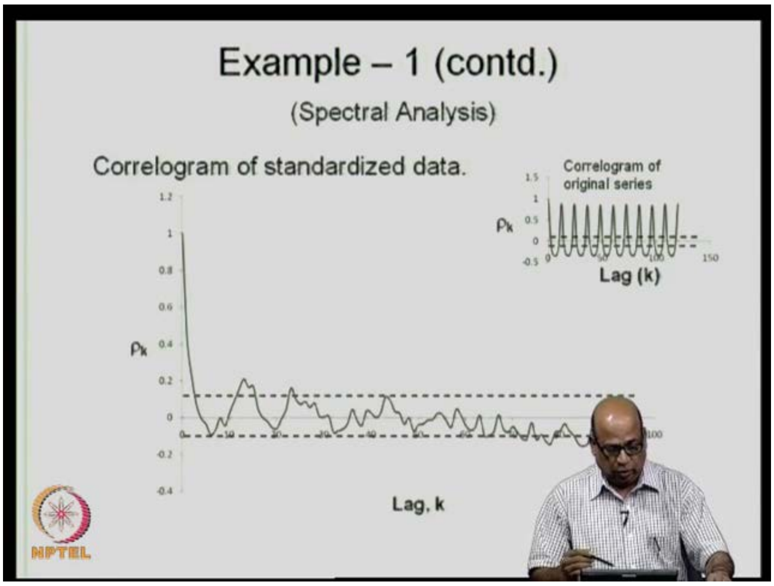
Similarly, do you take the July month 3 you take the August month and so on? 13 again you take June month, 14 July month and so on. So, like this you formulate the Z dash series this is called as the standardized series this is one way of standardizing, where you deduct the mean and divide by the standard deviation. So, you have now ready with you the Z dash series. So, Z dash series for example, looks like this as just the first year values I have shown first 12 months. So, X t and Z dash t. So, Z dash t series is like this. So, you will have 348 values of Z dash t for t is equal to 1 to 346 till 348, we will do the spectral analysis on this new series. Now Z dash t and see how it looks we saw we the spectral density figure or spectral diagrams of the original time series X t.

(Refer Slide Time: 19:12)



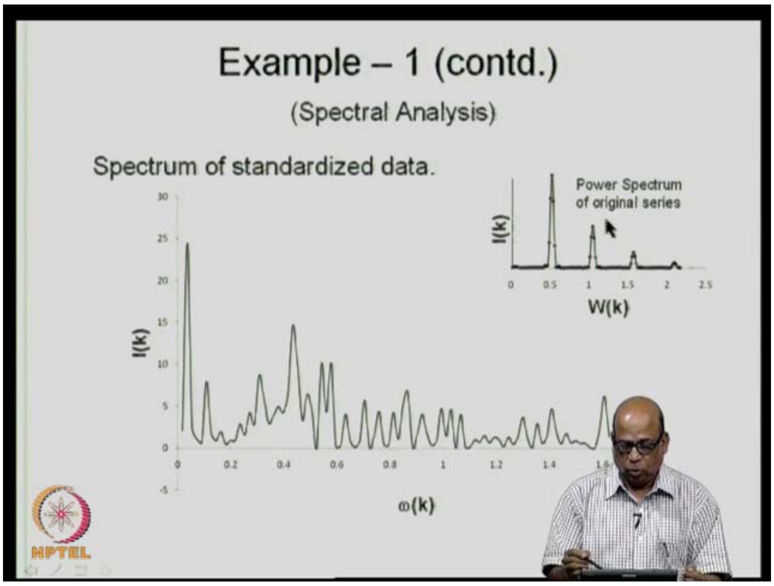
So, the time series when you plot for the standardized data it looks like this whereas, the original time series was this. So, this is a standardized data time series for the standardized data this flow is standardized flow. As you can see here this is a much more random series compared to the original time series the original time series had some periodicities and so on. So, you see some regularity in the original time series whereas, by standardizing you are seeing a much more random series here, but whether it is truly random or whether there are still inherent periodicities in this etc can be seen only if you plot a correlogram plot the correlogram of this series and do the spectral analysis on this series let us do that now.

(Refer Slide Time: 20:03)



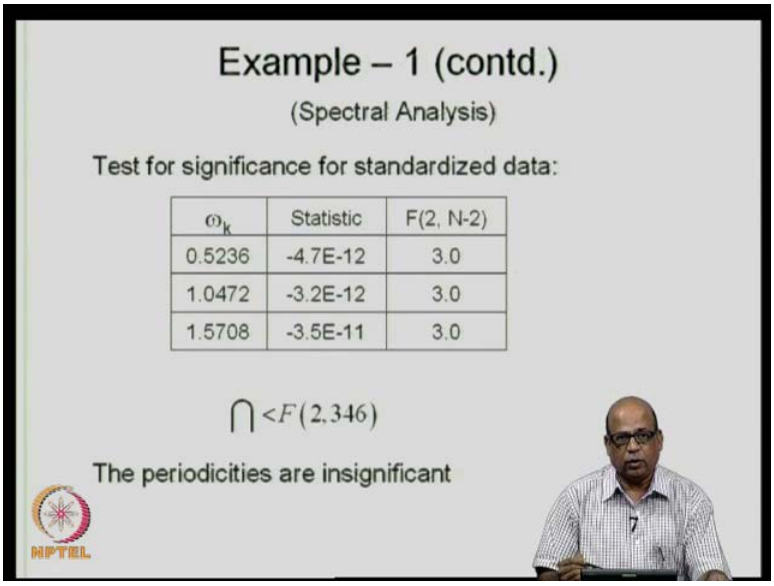
So, if you plot the correlogram your original correlogram look like this. That is a correlogram of the original time series looks like this and the standardized time series when you do the correlogram this looks like this.

(Refer Slide Time: 20:54)



So, most of the correlations are statistically insignificant here. So, the periodicity seems to have been removed in the standardized series, but this can be verified through the spectral density. So, the correlogram shows that is the statistically significant periodicities have all been removed and most of the correlations here except for the really. Early 1s they are statistically insignificant this information can be seen better in the spectral density.

(Refer Slide Time: 20:56)



If you look at the spectral density this is how it looks in your original series the power spectrum looks like this. The power spectrum of the original series showed significant spikes corresponding to the periodicities of 12 months, 6 months, 4 months and 3 months here. Whereas, the spectrum of the standardized data looks much more random here no single periodicity are no single spectrum spectral density is anymore different from any other periodicity here and therefore, the variance is spread more or less uniformly as you can see from the spectral density of the standardized data.

So, standardization removes the periodicities present in the data, but this may not be always true. So, you need to really examine after standardizing the data you need to examine whether the periodicities had in fact been removed by doing the transformation corresponding to standardization. For example you may have daily rainfall data what we have done, just now is for the monthly stream flow data which is much more smoothen process corresponding to compare to the daily rainfall data. For example, if you have daily rainfall data with several 0s and then once in a while it rains, but it rains very heavily during that time period and then this may have long term periodicities present in the data.

If you are then looking at standardization with respect to the daily mean and the daily standard deviation the periodicity the long term periodicities that are present in the data may not be removed we will see some applications of the spectral density towards the end of this course where we I will be dealing with only the applications of all the topics that we had covered in the course. During that time we will see how the rain fall time series behaves how the monthly stream flow time series behaves how aggregating the rainfall over season and then looking at seasonal time series of the rainfall will be much different from the spectral densities and the correlations of the daily time series and so on.

So, the information content in the time series comes to the surface by doing all these analysis. So, the inherent information that we have in the time series comes to the surface through the correlogram through the power spectrum and so on. So, what we have just examined is that the periodicities that were significantly shown up in the original time series have been removed by doing the standardization. So, standardization is one way of removing the periodicities. So, we see that corresponding to this value we had a significant periodicity. So, we will check again corresponding to that value of omega k whether the periodicity is present or not.

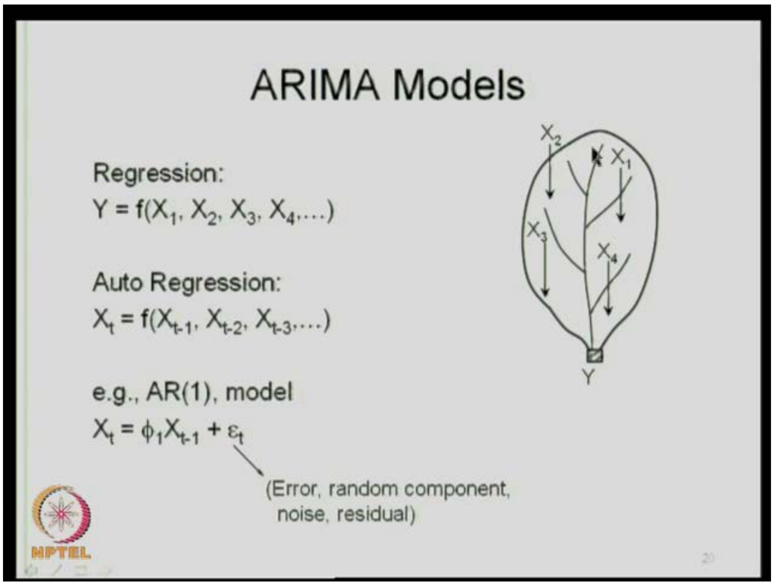
So, we again recalculate the statistic as we defined here earlier we calculate this is statistic gamma square N minus 2 therefore, rho 1 cap and then corresponding to this omega 1. We test the statistic the statistic value comes out to be minus 4.7 E to the power minus 12 very low values very insignificant values compared to the critical value of F which is 3.0 which shows that any of these periodicities are statistically insignificant or they can be taken to be absent in the data.

So, the periodicities are insignificant. So, essentially what we have tested this through the exercise that we just did is that in the original time series, we saw some periodicities which were all statistically significant we standardize the data and reexamined to make sure that the periodicities that were present corresponding to 12 months, 6 months and 3 months have all been removed from the data by ensuring that the periodicities corresponding to this are all statistically insignificant.

Now, we will go to an interesting topic in the time series analysis, these are the models that we will be dealing with now onwards are called as aroma models auto regressive integrated moving average models. Now these are box Jenkins types models and they are very popularly used in hydrology, especially for modeling monthly flows seasonal flows and such the process which are of interesting hydrology most hydrology applications. Where we want to use these models for forecasting as well as generation of the models data generations.

Recall, that we introduce a model earlier on may be about three lectures ago, for data generation using the first order Markov process and that we also called it as non stationery first order Markov process. We will see that was in fact an auto regressive model when we look only at the stationery model. So, before we went on to the non stationery model, we introduced the primary stationery model that in fact turns out to be the first order auto regressive model. So, the Arima models we use extensively in hydrology for data generation as well as for real type forecasting models. So, we will just over the next half an hour and also over the next lecture we will deal mostly with the Arima type of models.

(Refer Slide Time: 27:33)



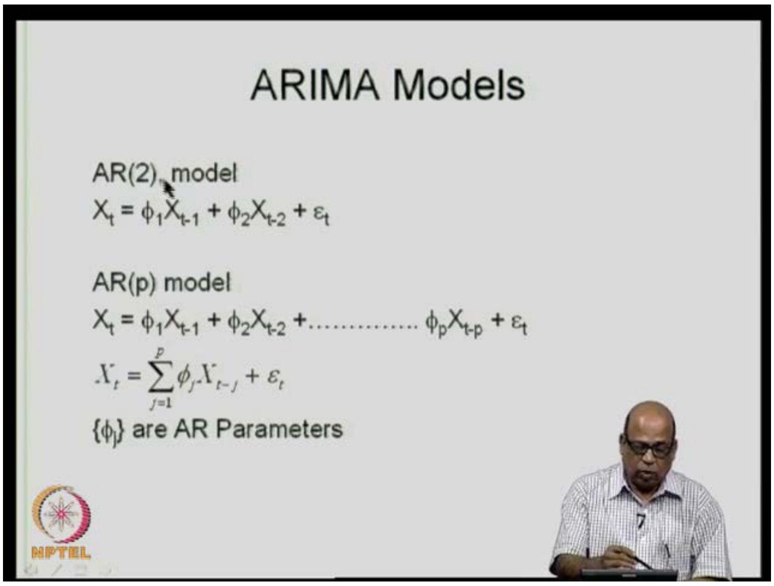
Let us look at what we did in our regression let say that you are looking at the flow at a particular location. Which is governed by this catchment and we call this variable as y and there were several variables. For example, rainfall was one of the variable soil moisture may be another variable the catchment slope or vegetation may be another variable and so on. Antecedent moisture which I just mentioned as soil moisture that may be another variable there may be several other variables.

For example evapotranspiration may be another variable and so on. So, we were essentially looking at the relationship between the runoff at this location with several different physical variables. That is what we did in regression actually the regression that I introduced was only simple regression in which only one variable was considered, but subsequently we will also deal with multiple regression in which the dependent variable will be governed by several independent variables like X 1, X 2.

For example as I said rainfall soil moisture evapotranspiration and vegetation and so on. So, you may have several variables all of which determine the variation in Y, this is called as whenever we talk about regression. We generally understand that your regressing one variable with other variables. In the auto regression what we do is we are regressing the variable upon itself, but the values that I have that the variable have taken at different time periods. So, for example, we are talking about X t being dependent upon X t minus 1, X t minus 2, X t minus 3 etc. So, we are not talking about different variables we are talking about the same variable.

For example, the flow at a particular location at different time periods t minus 1, t minus 2 etc. For example, this may be June months flow this may be may this may be April, March, February and so on. So, you are talking about a single variable, but the values taken by that variable at different time periods. So, that is called as auto regression. Let say you are talking about auto regression of first order, we may simply write X t is equal to some constant Phi 1 into X t minus 1 which is the flow, if X t denotes a flow at time period t your regressing X t with X t minus 1 the flow during the previous time period plus there is a error term or the random noise term here. So, this is the random component or noise or residual.

(Refer Slide Time: 31:14)

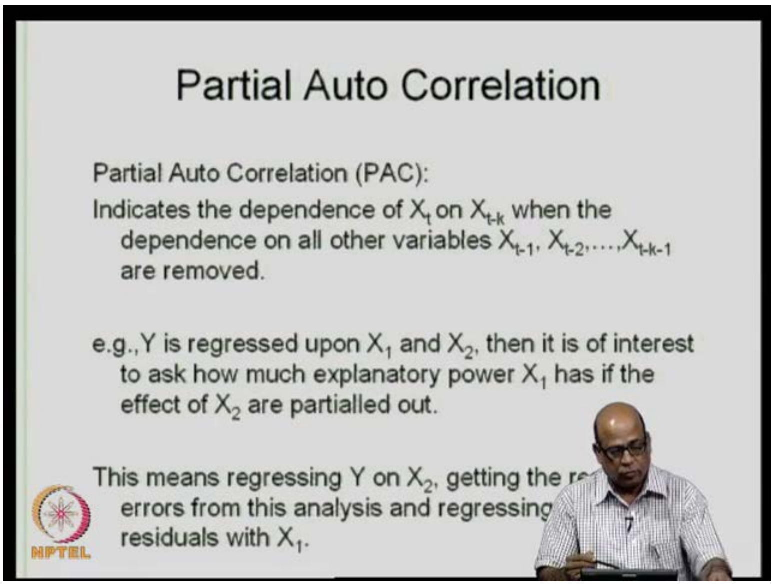


So, write you write X t is equal to Phi 1 into X t minus 1 plus epsilon t which is similar to what we did in regression simple linear regression what did we write y is equal to a x plus b. So, similar way we are writing X t is equal to Phi 1 X t minus 1 into plus epsilon t this is a simple AR 1 model let say you are writing auto regressive model of order 2. So, you will write X t; that means, when you are talking about order 2 you are saying X t depends on X t minus 1 as well as X t minus 2. So, two terms behind you are taking so, you will write AR 2 model X t is equal to Phi 1 X t minus 1 plus Phi 2 X t minus 2 plus epsilon t this is similar to writing a multiple regression model multiple linear regression model of the type y is equal to a 1 X 1 plus a 2 X 2 plus b plus some other constant b.

So, in the auto regressive two models you will write the model in terms of the two previous values X t minus 1 and X t minus 2 and like this. You write a general AR p model auto regressive model of order p by taking into account the previous p previous terms. So, X t is equal to Phi 1 X t minus 1 plus Phi 2 X t minus 2 etcetera, Phi p X t minus p plus epsilon t. So, if X t denotes the flow at a particular month. Let us say you are talking about the stream flow in the month of June and this is a monthly time series you may write that as Phi 1 X t minus 1 let say you are talking about Phi 12, X 12 is equal to Phi 1 x 11 plus Phi 2 X 10 and so on.

So, you are taking p previous terms. So, you write this in a more compact form as X t is equal to j is equal to 1 to p Phi j X t minus j plus epsilon t, these Phi j are called as AR parameters. So, the Phi j is called as the auto regressive parameters. So, given any model of this type we should be able to estimate Phi 1, Phi 2 etc from the data. So, you have the data on X t you write a model like this. So, there must be a way of estimating these parameters which we will see subsequently how we do this?

(Refer Slide Time: 33:37)



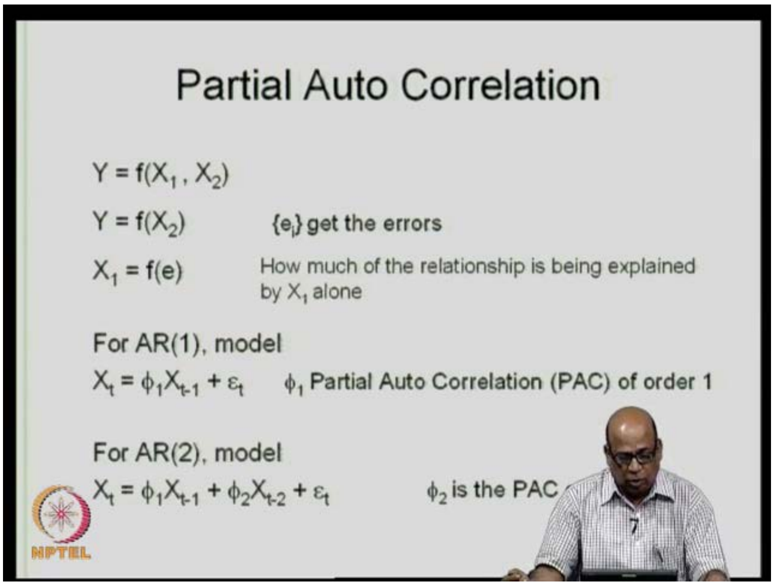
Similar to the correlation that we talked about we have a very important concept called as partial autocorrelation it is denoted as PAC. Before we go into a more generalized Arima model explanation let us see what we mean by PAC. What did the correlation coefficient indicate the correlation coefficient between two variables. The correlation between two variables indicates the linear dependence of one variable on the other. For example, if we say that rho 1 is equal to 0.6 between X and Y or the correlation between x and y is 0.6, it indicates the degree of dependence between of Y on X. Now the partial autocorrelation indicates the dependence of one variable X t on X t minus k. We are talking about auto correlation and therefore, the same variable. We are talking at various time periods X t on X t minus k, when the dependence on all other variables X t minus 1, X t minus 2 etc, X t minus k minus 1 are all removed.

What we mean by this is that when you are talking about dependence of X t on X t minus k all these other dependence the correlation. When we are talking about X t on X t minus k these dependence are also included in that, but if you have some mechanism by which partial out or you remove the dependence of X t on X t minus 1, X t minus 2 etc on all other variables except X t minus k. Then the remaining correlation is in fact the partial autocorrelation. So, it indicates the dependence on X t on X t minus k alone when the dependence of X t on all other variables has been removed.

To understand this better let us say that y is regressed upon X 1 and X 2 and then we are interested in how much explanatory power X 1 has if the effect of X 2 is partial out or removed; that means, Y is dependent on X 1 as well as X 2. So, you have regressed upon regressed y upon X 1 and X 2 both together, but now you ask the question out of X 1 and X 2 how much power X 1 alone has how do we answer these question. Let us say that why we regress only on X 2 first and get the residual out of that; that means, X 2 has been able to explain part of variables of Y the residuals. We take out and then the residuals, we regress with respect to X 1 and then see how much of these errors or these residuals can be explain by X 1 and that is in fact the explanatory power of X 1 on Y. When the dependence on X 2 has been taken out when the correlation on X 2 has been taken out on this.

So, this generally gives the idea of partial autocorrelation, now the partial autocorrelations are important indicators of what type of Arima models that we make use for the data that we have. And therefore, we must able to estimate the partial autocorrelations and infer from the partial autocorrelations what level of auto regressive terms that we may want to use for our model.

(Refer Slide Time: 37:34)

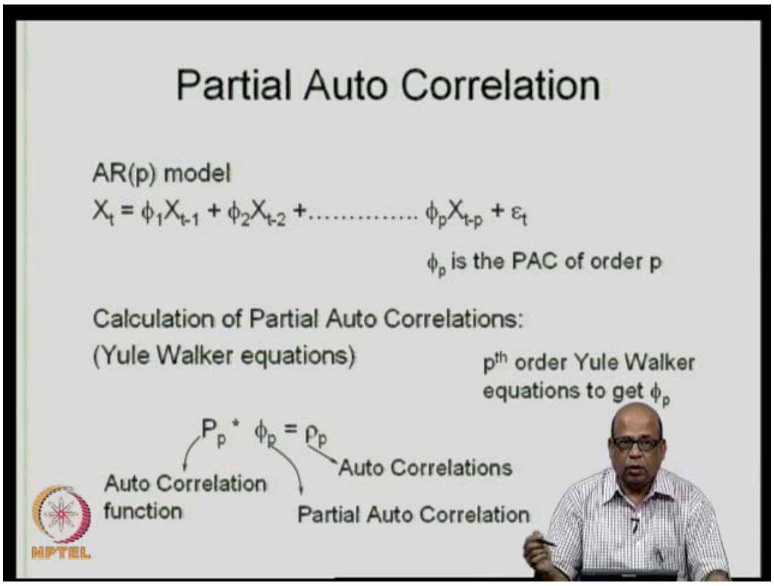


The partial autocorrelations this is what I just explained. For example, if you have Y is equal to F of X 1 comma X 2 first you regress only X 2 then you get the error terms and then regress X 1 on the error terms that you get. So, you will be able to tell how much of the relationship is being explained by X 1 alone how much of this relationship is being explain by X 1 alone and that indicates the partial autocorrelation.

The AR 1 model that we wrote as X t is equal to Phi 1 X t minus 1 plus epsilon t, because you are dealing with only one variable X t and X t minus 1 that is X t on X t minus 1. You are regressing X t on X t minus 1 the term Phi 1 here explains completely the explanatory power provides completely the explanatory power of X t minus 1 on X t and therefore, that itself becomes the partial autocorrelation for order 1 of order 1. So, Phi 1 of the AR 1 model is in fact the PAC or the partial autocorrelation of order 1.

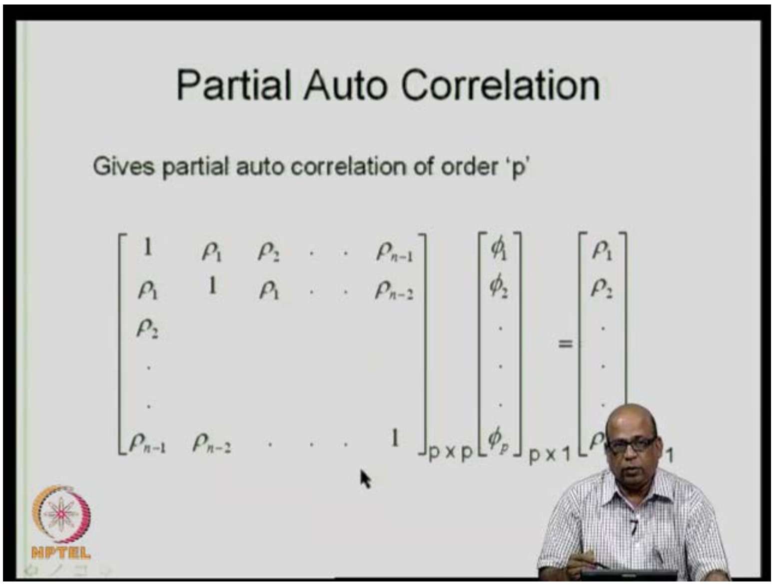
Let us write AR 2. Now AR 2 we write it as X t is equal to Phi 1 X t minus 1 plus Phi 2 X t minus 2 plus epsilon t for the AR 2 model Phi 2 is the partial autocorrelation of order 2. Remember the Phi 1 of AR 2 model is different from the Phi 1 of AR 1 model and the Phi 1 of AR 2 model does not I repeat does not indicate the partial autocorrelation of order 1 if you want the partial autocorrelation of order 1 look at the AR 1 model the Phi 1 of AR 1 model is a partial autocorrelation of order 1. If you want Phi 2 or the partial autocorrelation of order 2 you write the AR 2 model and the Phi 2 that you write here for the AR 2 model is in fact the partial correlation of order 2. So, in general the Phi p of the AR p model indicates the partial autocorrelation of order p.

(Refer Slide Time: 39:58)



So, any AR p model you write like this X t is equal to Phi 1 X t minus 1 etc. Phi p X t minus p plus epsilon t the Phi p of the p AR p model is a PAC of order p. There is an elegant way of estimating the partial autocorrelations, we have introduced earlier in the lectures we use the Yule Walker equations to obtain the Phi p or the partial autocorrelations. If you recall the partial the Yule Walker equations we wrote as p it is an auto correlation function of order p into Phi p, which is a partial autocorrelation is equal to rho p, which is the lag 1 lag they are just the autocorrelations.

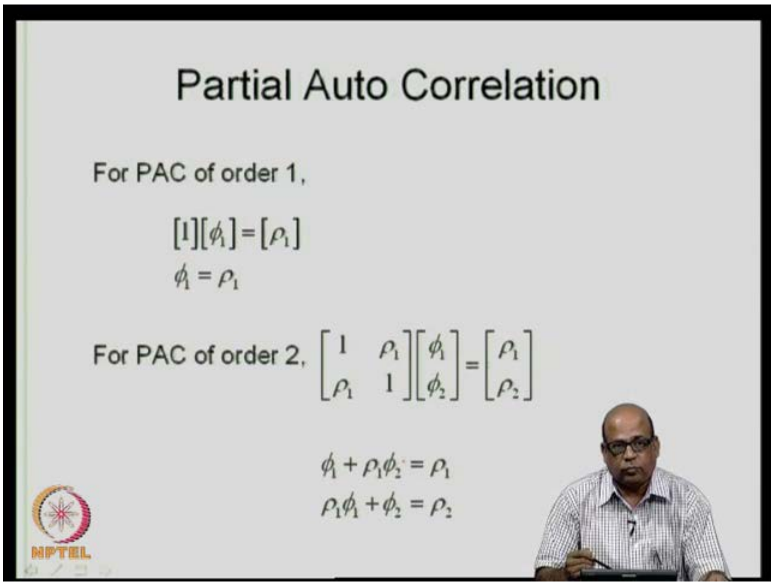
(Refer Slide Time: 40:56)



So, if you write this in the long form this is your P p or that we called them as the auto correlation function. So, the auto correlation function is 1 rho, 1 rho 2 etc rho n minus 1 like this it goes and the similar this is this has to be p this has to be p here this is p by p. So, similarly Phi 1, Phi 2 etc up to Phi p just half a minute I will just change the pen. So, this is rho p minus 1 etc rho p minus 2 and rho 1. Similarly Phi 1, Phi 2 etc, Phi p and this will be equal to rho 1 rho etc rho p. So, this is a p th order Yule walker equation. So, when you solve this p th order Yule Walker equation you will get the solutions for Phi 1, Phi 2 etc. Phi p and the Phi p which is a last term that you get is in fact the partial autocorrelation of the order p.

So, this is how you determine the partial autocorrelations of order p if you want Phi 1 you write the Yule Walker equation for a of order 1, if you want Phi 2 solve Yule Walker equations of order 2 and so on.

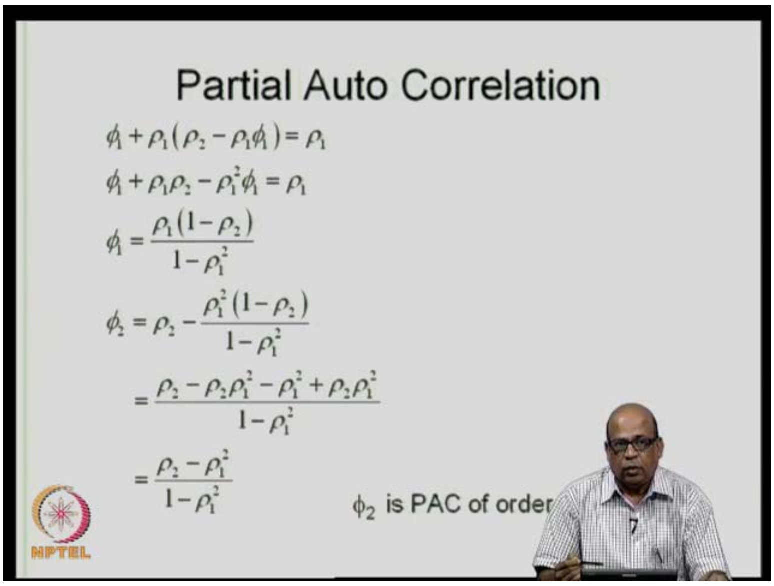
(Refer Slide Time: 42:40)



So, this is a p th order Yule Walker equations. So, for PAC of order 1 you write the Yule Walker equation of order 1 which is only 1 term here you will write the first term which is 1 here and then Phi 1 is equal to rho 1 therefore, the first partial autocorrelation is equal to the lag 1 auto correlation itself. So, Phi 1 is equal to rho 1. Similarly, for the second order partial autocorrelation you will write the Yule Walker equations of order 2 which is 1, rho 1, rho 1, 1 what I am doing now is that I am writing it for the second order. So, 1, rho 1 and rho 1, 1 is equal to into Phi 1, Phi 2 is equal to rho 1, rho 2 that is that is a second order equation.

So, I will write this as 1, rho 1, rho 1, 1. Phi 1, Phi 2 is equal to rho 1, rho 2 when we simplify this and multiplying this. So, Phi 1 plus rho 1, Phi 2 is equal to rho 1 that is the first equation similarly rho 1 Phi 1 plus Phi 2 is equal to rho 2 that is a second equation. So, you have two equations rho 1 and rho 2 is known. You can determine Phi 1 and Phi 2 rho 1 and rho 2 are the lag 1 and lag 2 autocorrelations these are determine from the data. So, from data you know rho 1 and rho 2 and therefore, you should be able to solve these simultaneously and get Phi 1 and Phi 2.

(Refer Slide Time: 44:21)



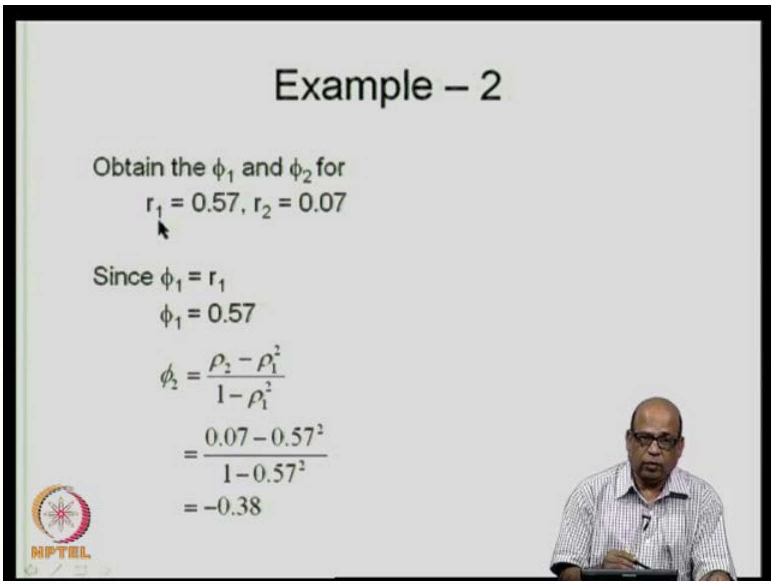
So, we will simplify that you write Phi 1 plus rho 1, rho 2 minus Phi 1 rho 1 Phi 1 is equal to rho 1, this is from this equation you get here Phi 1 plus rho 1 Phi 2 is equal to rho 1. So, you are putting it for Phi 2 here substituting for Phi 2 and then getting it to 1. So, Phi 1 you will get as let me explains this correctly. So, for Phi 2 you have rho 1 Phi 1 plus Phi 2 and this you put it in rho 1 Phi 2. So, Phi 2 you will substitute as rho 2 minus rho 1 Phi 1 and this you write it as equal to rho 1. So, by simplification you will get Phi 1 is equal to rho into 1 minus rho 2 by 1 minus rho 1 square rho 1 and rho 2 are known from the data. So, straight away you get Phi 1.

So, similarly you will get Phi 2 as rho 2 minus rho 1 square by 1 minus rho 1 square this. So, simple algebraic simplification, I repeat again when you solve the second order Yule

Walker equation which we are doing now. So, you are solving the second order Yule Walker equation to get Phi 1 and Phi 2, the Phi 2 that you get is the partial autocorrelation of order 2 remember that the Phi 1 that you get here is not the partial autocorrelation of order 1. If you want the partial autocorrelation of order 1 you have to solve the Yule Walker equation of order 1 which we did earlier and then said that Phi 1 is equal to simply rho 1. So, the Phi 1 that you get from the Yule Walker equation of order 2 will be different from the Phi 1 that you get from the Yule Walker equation of order 1.

So, the Phi 2 here is the PAC of order 2, in general Phi p is the PAC of order p, when you formulate and solve the Yule Walker equation of order p. So, we now know how to determine the partial autocorrelations at different order. So, p is equal to 1, p is equal to 2 etc. You know how to determine the partial autocorrelations of different orders from the data you will have all the autocorrelations rho 1, rho 2 etc rho p and look at the Yule Walker equations. If you have rho 1, rho 2 etc rho p you will be able to form the auto correlation function you have the partial autocorrelation function here and these are the autocorrelations up to lag p.

(Refer Slide Time: 47:42)

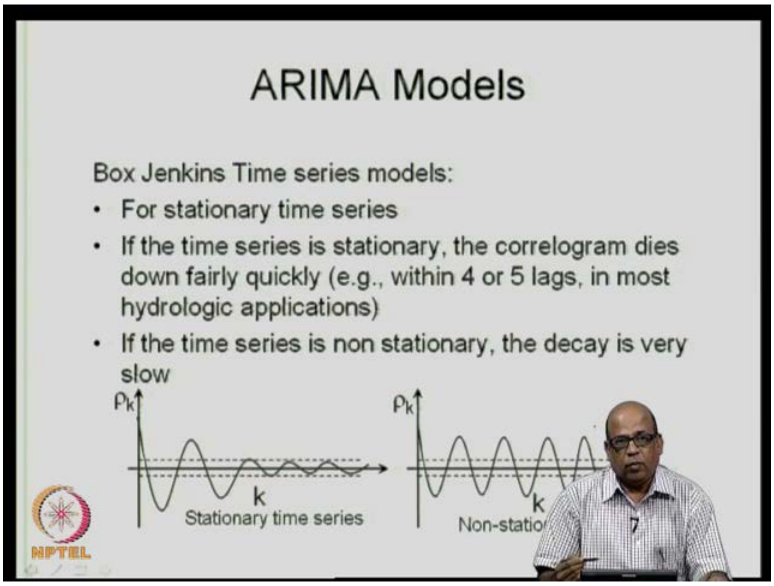


So, all you need is the autocorrelations up to lag p to get the partial autocorrelations and these we get from data, you know how to get the correlogram. Let us see from the data let us say that you have the r 1 which is an estimate for the lag 1 correlation as 0.57 and r

2 which is a estimate for rho 2 which is a lag 2 autocorrelations as 0.07 then Phi 1, which is the PAC of order 1 partial autocorrelation of order 1 is simply equal to r 1. So, this is you can put a cap here this is Phi 1 cap is equal to r 1 this is estimate. Similarly, Phi 2 cap I can write it as these are estimates sample estimates because you are estimating rho 1 from the sample and therefore, the sample estimates for the PAC are also denoted as Phi 1 cap and Phi 2 cap and so on.

So, you get Phi 2 cap is equal to rho 2 minus rho 1 square by 1 minus rho 1 square and rho 2 is r 2 here 0.07 minus 0.57 square by 1 minus point 57 square it is minus 0.38. The partial autocorrelation being correlations by themselves will have a range of minus 1 to plus 1 they are also like correlations they vary between minus 1 to plus 1 for all p.

(Refer Slide Time: 49:28)



Now, we will see how we use the information on the partial autocorrelation. So, we have the information on the correlogram. We also have the information of the spectral density, now we have added one more information which is that of partial auto that provided by the partial autocorrelation. So, just given the data; that means, that you have observed data at a particular location, let us say you are talking about observed stream flow at a particular location first we simply plotted the time series we saw that there is a some kind of a indication of some regularity in the data, it may be just let say the periodicity or it may be an increasing trend or it may be just a long term mean around which the value are fluctuating and so on.

So, you see just by plotting the data as the time series plot you get some indication of presence of some kind of a regularity. Now this information we further smoothened or we extracted much more information by plotting the correlogram. The correlogram may indicate presence of periodicities; that means, there is some periodicity that is indicated by the correlogram it may be 12 month periodicity 14, 24 months periodicity or long term decadal periodicity and so on. So, the correlogram gives an indication that yes there is a periodicity present in the data. We further find this information that is provided by the correlogram by plotting the spectral density. The spectral density brought to the 4 the presence of periodicities much more strongly than did the correlogram.

So, in the spectral density we could identify that there is a periodicity corresponding to 12 months, there is a periodicity corresponding to 6 months, 4 months and so on. In the example that we did we solved before coming to Arima models. Then we also examined how to test these periodicities and now we have introduced the partial autocorrelations, which means that we are going deeper and deeper into the information contained in the observed and time series. The basis for all of this is just the observed time series you may have one time series you may have several realizations of the time series, but we are simply going deeper and deeper into what information can be extract out of the observed values that we have.

So, the partial autocorrelations are another source of information from the data observed data. So, much the same way we plot the correlogram we should be able to plot the partial autocorrelation also. Now with all these information we should be able to build models for the particular process let say you are talking about stream flow, monthly stream flow at a particular location. So, we have extracted all the information that is contained in the data we must use these data to build models for that particular process.

The specific models that we will be now discussing as I just mentioned is are called as the Arima models autoregressive moving average models. These are the specific type of models that I will be discussing are called as the box Jenkins type of time series model they are written for stationery time series. So, the models that I will be discussing are all only for stationery time series there are also non stationery time series models, they are beyond the scope of this course, but keep at the back of your mind that much the same way we develop the time stationery time series models. You can also develop non stationery time series models adopting different other method.

If the time series that we are dealing with is non stationery first you must convert the time series into a stationery time series and only then develop or apply these models how do we identify that the time series is non stationery? first you look at the correlogram for a stationery time series the correlogram dies down very rapidly in fact in most hydrologic applications you may see that the stationery the correlogram dies down after 3 or 4 lags or 4 or 5 lags in most of the cases by correlogram dying down what I mean is that this correlations become insignificant as lag progresses with progress in lag. They become quickly insignificant whereas, if you have a non stationery time series the decay in the correlogram is not very fast, it may decay very slowly over a long period of time you may see that the correlations become insignificant.

So, this indicates as you can see here there is significant periodicity that refuses to die down even with significant lags and therefore, it indicates that the time series is in fact non stationery. So, first you need to identify whether the series is stationery or not and then only apply the time series procedure the Arima models that we will be discussing and so on. So, let us summarized now what we covered in this lecture, we started with the spectral density the problem that I was discussing in the last lecture we saw that in the monthly time series the offspring flow that we consider. For the example there were significant periodicities which were shown primarily by the correlogram.

The correlogram indicated a sinusoidal oscillation and therefore there by indicating there are periodicities present in the data. The spectral densities the spectral analysis brought much more strongly the presence of periodicities, but it also identify where exactly the periodicities are present. In the numerical example it identified that the there was a periodicities corresponding to 12 months, there is a periodicity corresponding to 6 months, 4 months and 3 months. Then we examined which of these periodicities are in fact statistically significant by doing the statistical test.

We saw that all the periodicities were statistically significant in that particular example. Then we saw the effect of standardization, we standardize the series and saw that all the periodicities that we had identified are statistically significant. Where absent in the standardized series indicating that standardization is one way of removing the periodicities. Then we went on to write the autoregressive moving average models autoregressive moving average models. So, the AR p model we write we wrote and then explained the auto regression process itself verses the regression model that we are used to.

Then we introduced the concept of partial autocorrelations the partial autocorrelations indicate the explanatory power of a particular variable X t minus k on X t. When the dependence on X t on all other terms have been removed and we defined a general autoregressive integrated moving average model. So, we will continue with this discussion in the next lecture, where we will write a more general autoregressive integrated model, integrated moving average model and then see the various steps involved in fitting an Arima model, general Arima model to a particular hydrologic process. So, we will continue the discussion in the next lecture thank you for your attention.