Stochastic Hydrology

Prof. P.P. Mujumdar

Department of Civil Engineering

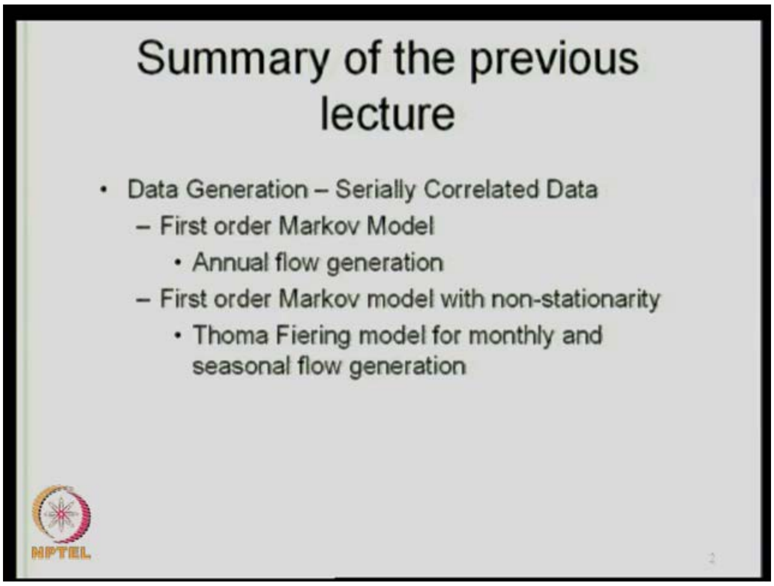
Indian Institute of Science, Bangalore

Lecture No. # 13

Frequency Domain Analysis – I

Good morning and welcome to this lecture number 13 of the course stochastic hydrology. If you recall in the last lecture, we dealt with data generation techniques especially, the stream flow generation using the first order Markov processes; and we introduced the and for annual flow generation the first order Markov model. Recall that this will express X t using the lag 1 correlation and its which indicates its dependence on X t minus 1. And the model that we introduce will preserve the mean, the standard deviation and the correlations of the historical data. So, if you have annual series of stream flows, you can compute these moments, and then use the model Markov model first order Markov model and generate series of annual flows.

(Refer slide Time: 01:19)



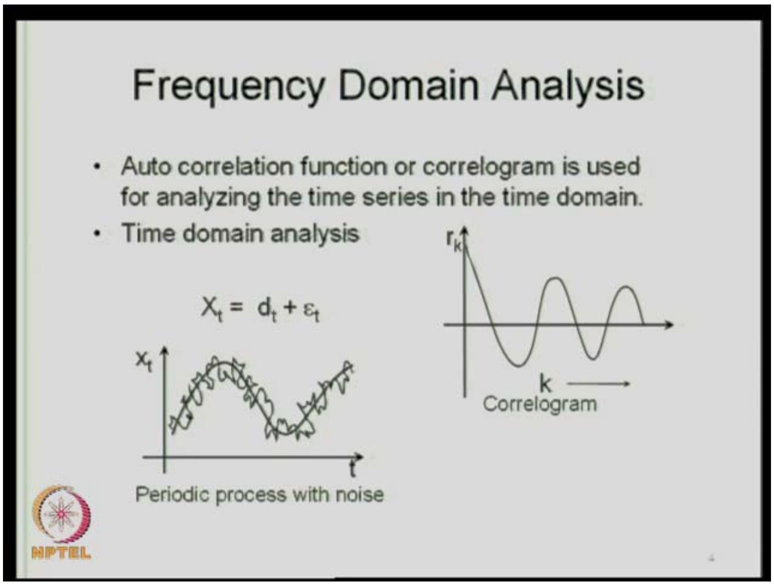
Then we relax the requirement that the series be stationary and we consider the non stationary first order Markov model, where for each of the month or season, you have different means, the standard deviations and the lag on correlations. We incorporated that into the Markov model, and then introduced the non-stationarity first order Markov model. This in the hydrologic literature is famously known as the Thoma Fiering model, and typically it is used for flow generation for monthly and seasonal time periods; and in some cases it has also been used for flow generation during smaller time intervals like 10 day time intervals and fortnightly time intervals etcetera. Recall that the requirements for the Markov models as we introduced in the last lecture are that the flow time series can be approximated to be a Markov chain, and also that the flow series can be approximated to follow normal distribution.

The Thoma Fiering model that we introduced in the last lecture requires the flows to follow normal distribution. And it also generates negative values because we are flowing the normal distribution and in the applications the negative values are set to 0 if you are using the sequence that you. So, generated using the Thoma Fiering model into let say simulation of reservoir operation at the time of applications you set the negative values to 0, but while using the generating model itself you keep the negative values to generate the next values.

Because it generates negative values often using the log transformation is advantageous. So, we also have the log transform Thoma Fiering model where simply you transform the X t into log X t and use the Thoma Fiering model on the log X t series that is also quite popularly applied in hydrologic literature. Now, we move to the next topic which i introduce just towards the end of the last lecture this is called as the frequency domain analysis. So, far whatever we have been doing on the time series analysis is on the time domain for example, we calculate the correlations and then plot the coreelaogram all of these are done with respect to time. We talk about correlation at a particular lag and these lag that we considered are with respect to time.

We can also transform the observed series X t into a frequency domain and then do the analysis in the frequency domain and often this is very advantageous especially in determining the periodicities inherent in the data. So, the frequency domain analysis we carry out mainly to identify the periodicities in the data, if you recall we expressed X t the time series as consisting of d t, a deterministic component plus epsilon t which is a stochastic component.

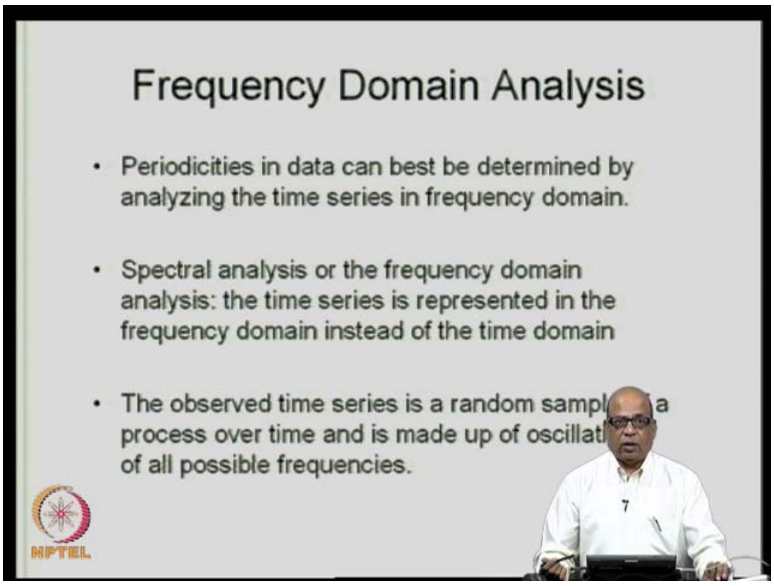
(Refer Slide Time: 05:07)



In the beginning we mention that the d t that is a deterministic component can consist of let say a long term mean around which the values are fluctuating or you may have a trend increasing or a decreasing trend around which the values are fluctuating or you may have a periodicity like this around which the values are fluctuating randomly apart from having a jump jump in the series and so on. Identification of these periodicities is an important problem in hydrology, the correlogram as as you can see here if you have a periodic process the correlogram will typically look like this in the in the time domain if you plot the correlogram r k versus k or rho k versus k the periodic process will look like this.

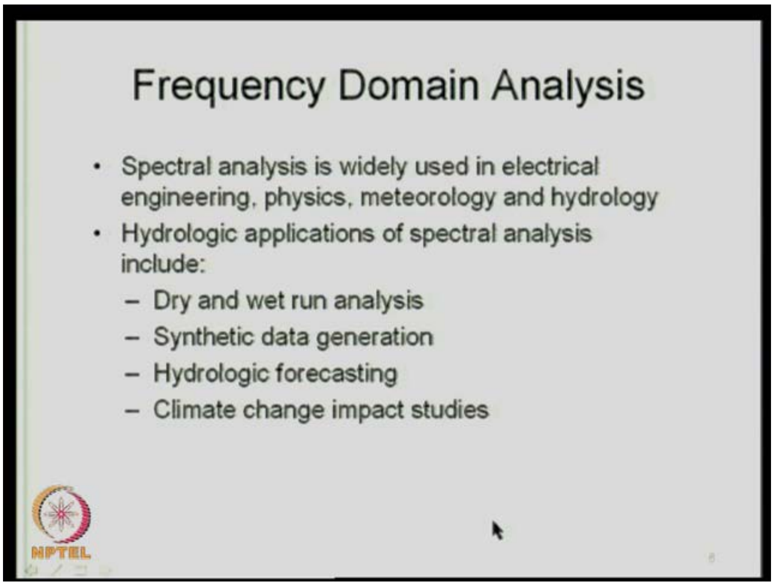
So, where the correlogram is oscillating periodically and typically it dies down slowly in most of the hydrologic process. The correlogram indicates that yes there are periodicities inherent in the data, but identification of the exact periodicities and the significant periodicities inherent in the data is better done with the spectral analysis where we transform the time series X t into the frequency domain and then we analyze the time series in the frequency domain.

(Refer Slide Time: 06:49)



This analysis is called as the spectral analysis. So, the periodicities in the data which were indicated by the correlogram can best be determined by analyzing the time series in the frequency domain and such analysis is called as the spectral analysis or analysis in the frequency domain. The basic premise or the basic hypothesis with which we start is that the observed time series is a random sample of a process over time and is made up of oscillations of all possible frequencies.

(Refer Slide Time: 07:43)

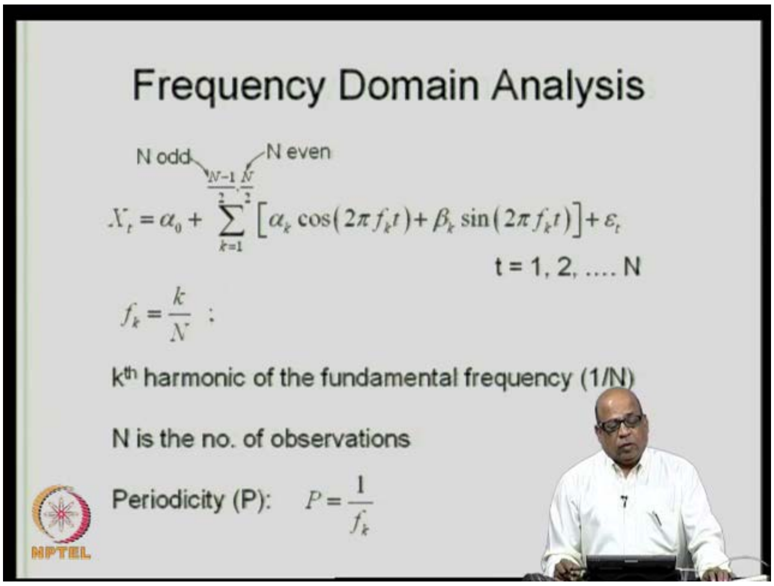


So, this this is the basic premise on which the frequency domain analysis is based. Then we look at the time series as having consisted of several frequencies and then we look at which of these frequencies are dominant in the observed time series. The spectral analysis as we introduce now is widely used in several other fields also like: electrical engineering, physics, metrology, apart from hydrology. In hydrology specifically the applications of spectral analysis include dry and wet run analysis let say you have a large sequence of rainfall data and you would be interested in frequencies of dry periods or wet periods, periodicities of dry and wet periods that is where we use the frequency domain analysis.

We also use it for developing models for synthetic generation in which we would be incorporating into the models the periodicities inherent in the data and for that identification of periodicities becomes important. Even in the hydrologic forecasting models when we are talking about medium term forecast let say or the next season 6 months, 1 year or 3 years medium term forecast. When we are talking about we need to incorporate the periodicities in the process into the models and that is where we use the spectral analysis and the recent research area in hydrology which deals with impacts studies of that is the climate change impact studies on hydrology, such studies also require identification of periodicities oscillations decadal oscillations and so on. In the particular process that we are talking about.

So, frequency domain analysis is an important topic in stochastic hydrology. We will not go into the theoretical development of several aspects of the frequency domain analysis that we will be dealing with rather what we will do is we will pick up directly the expressions that are available and then see how we apply to hydrologic time series, how we interpret the results and how we identify the periodicities, how we check the significance of the periodicities. So, identified and what do we do with these periodicities in subsequently building up the time series models.

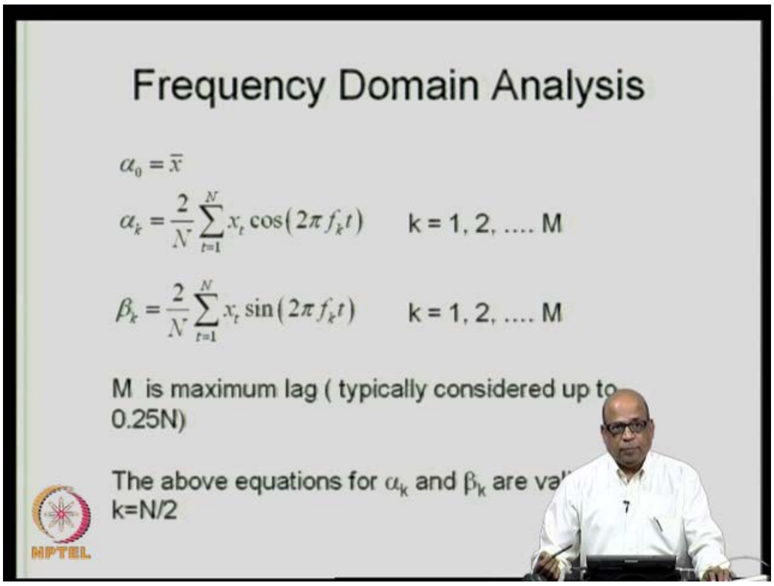
(Refer Slide Time: 10:45)



So, the first level as I said we express the time series X t as consisting of different frequencies for example, in this case we take a sinusoidal representation of the time series. So, we write X t as a constant alpha naught and then summation over k to N by 2 or N minus 1 by 2, where N is the number of data points that we have if N is even we write the summation up to N by 2, if N is odd you write up to N minus 1 by 2. Alpha k cos 2 pi f k t plus beta k sin 2 pi f k t summation ends here plus epsilon t.

So, this is how we are expressing the time series as consisting of a sinusoidal component, a random component and a constant. For a given k this f k here is the k th harmonic of the fundamental frequency. The fundamental frequency is simply 1 by N. So, f k is written as k by N. So, for a given k we will have expressions for determining alpha k you know f k you are writing for the time period t therefore, this t is known beta k we have an expression I will just present it, again f k is known t is known and epsilon t is the random component.

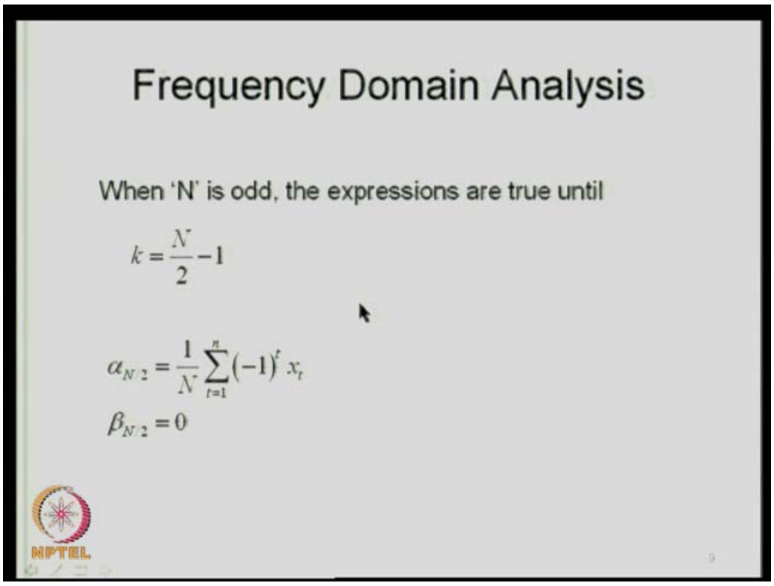
(Refer Slide Time: 12:44)



The periodicity when we write X t in this form is simply 1 by f k. So, f k is the frequency; obviously, periodicity will be 1 by frequency. So, in this expression we determine alpha naught which is a constant here by simply the estimate of the mean which is x bar and alpha k we determine by this expression 2 by N t is equal to 1 to N x t cos 2 pi f k t. You are summing it over all the time periods t is equal to 1 to N and this is the original time series x t f k is the particular frequency that we have determined for that particular k that is k by N and t is the particular time period.

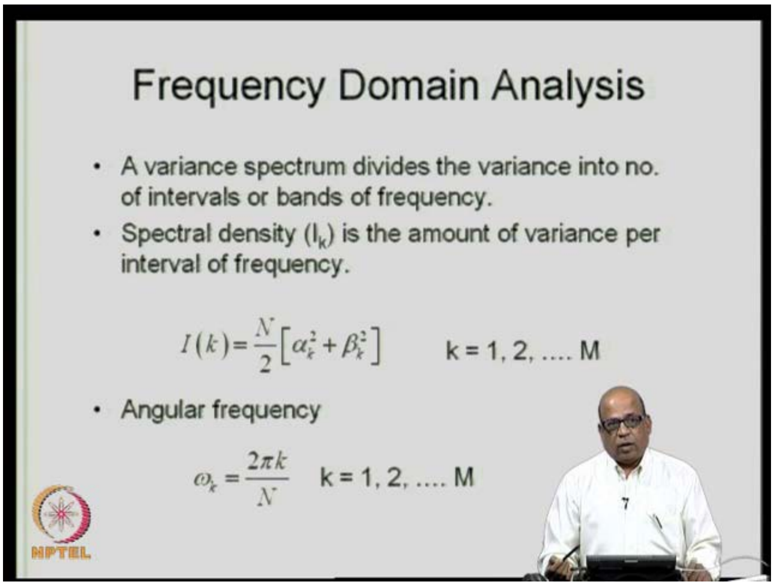
Similarly, beta k we determine by x t sin 2 pi f k t. Now these expressions are in general taken up to about a maximum lag of M and this lag typically. We consider up to about 25 percent of the total amount of data total number of data points that we have. So, about 25 percent of the data lag these expressions for alpha k and beta k are valid up to k is equal to N by 2 as as mentioned here.

(Refer Slide Time: 14:04)



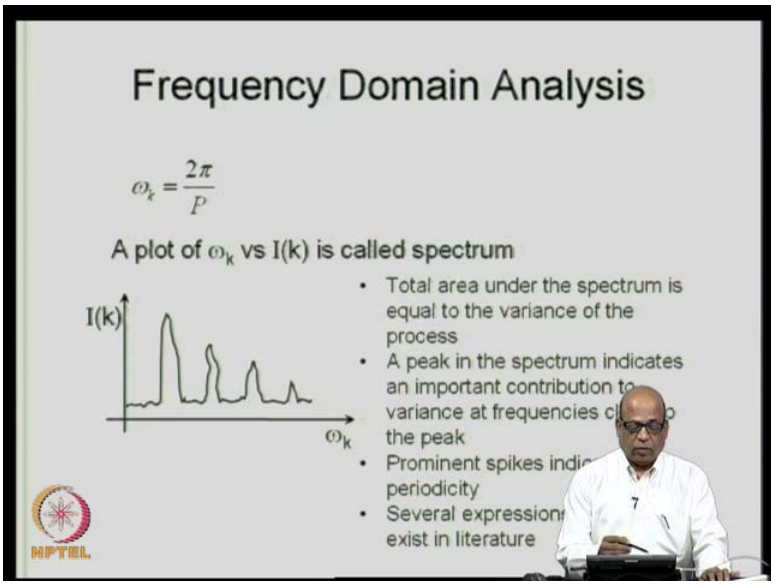
When N is odd the expressions are true until k is equal to N minus 1 by 2 and the last values of alpha N by 2 and beta N by 2 will be given by this beta N by 2 will be equal to 0.

(Refer Slide Time: 14:28)



Now let us see what are the interpretations of this the way we have introduced the spectrum is the called the variance spectrum. So, it divides the variance that is observed variance into number of intervals of frequency or bands of frequency. The spectral density as we write here it indicates the amount of variance per interval of frequency. So, we write the spectral density I K as N by 2 alpha k square plus beta k square written for k is equal to 1 to M, M is a maximum lag. Similarly we write the angular frequency the frequency that we introduce namely f k is equal to k by N is now converted into angular frequency by writing it as 2 pi k by N. So, omega k here is 2 pi k by N again written for k is equal to 1 to maximum lag.

(Refer Slide Time: 15:53)



A plot of I K versus omega k. Omega k on the x axis I K on the y axis is called as the line spectrum or simply spectrum in many cases and it looks typically like this. So, with P as given by N by k we have omega k is equal to 2 pi by P, P is the periodicity. Now we plot omega k versus I K, omega k is determined by 2 pi k by N and I K is determined by this expression, alpha k is determined as earlier alpha k is determined using this beta k is determined using this. Therefore, we know all the terms given a time series we can determine for various values of k the w k values and you plot w k versus I K this is called as spectrum in general. These are the various salient features that we must remember the total area under spectrum when we are talking about various spectrum the total area under the spectrum is equal to the various of the process.

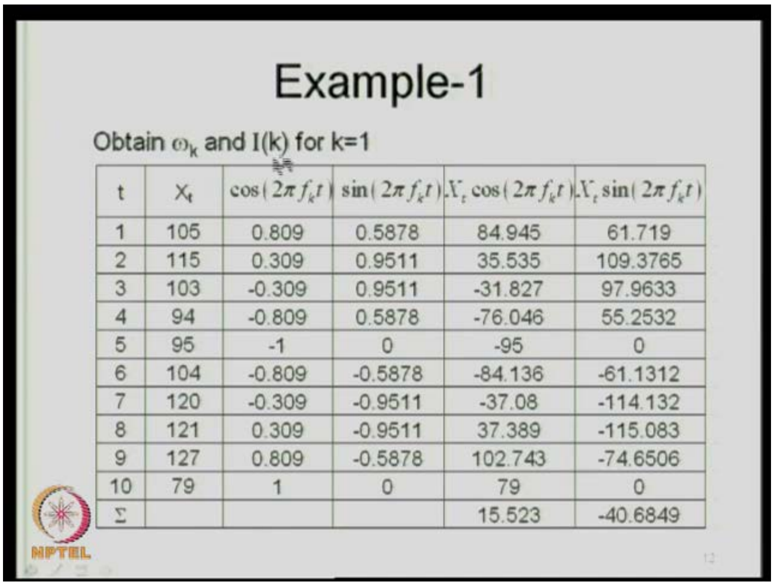
A peak like this in the spectrum indicates an important contribution to spectrum that is the whole range here this whole range, we are talking about the distribution of the spectrum in several bands of frequencies. So, whenever you get a peak here this indicates that there is an important contribution to the variance of the process around these frequencies. The prominent spikes here indicate that there is a periodicity corresponding to the angular frequency that we are talking about. So, when you plot the spectrum you capture these peaks look at the omega k values convert that into the periodicity and that indicates that a periodicity corresponding to that particular time exist in the data.

Now, the particular expression that we just introduced are one among several possible expressions for spectrum estimates. So, we are using one of them in the text and in the literature you may find several different ways of expressing the spectrum, but we will add here to that particular type of expression that we have just introduced. So, essentially then what we are doing in spectral analysis is to convert the X t which is the observed time series into a sort of a 4ier transform where we are using a sinusoidal representation of the time series and then estimating the spectral density I K we are plotting I K versus omega k versus I K, where omega k is the angular frequency.

When we plot omega k versus I K or I K on the y axis and omega k on the x axis. We see spikes or peaks in the spectral density typically. Now these peaks will indicate the periodicities inherent in the data, the omega k corresponding to a particular peak you convert that into the periodicity 2 pi by omega k and that periodicity in the time domain will indicate the periodicity that is present in the data for example, if you are looking at monthly stream flow data and then you plot a spectrum plot the line spectrum. Typically you may get an omega k of omega k which corresponds to a periodicity of 12 months or may be 6 months, 3months etc. Depending on the type of stream flow data that you have.

So, when we look at certain examples these points will become clear. We will see one example, now where first we introduce how exactly we estimate alpha k, beta k the coefficients. For a specific value of k remember for a given k f k is equal to simply k by N. N is the number of the observations that you have.

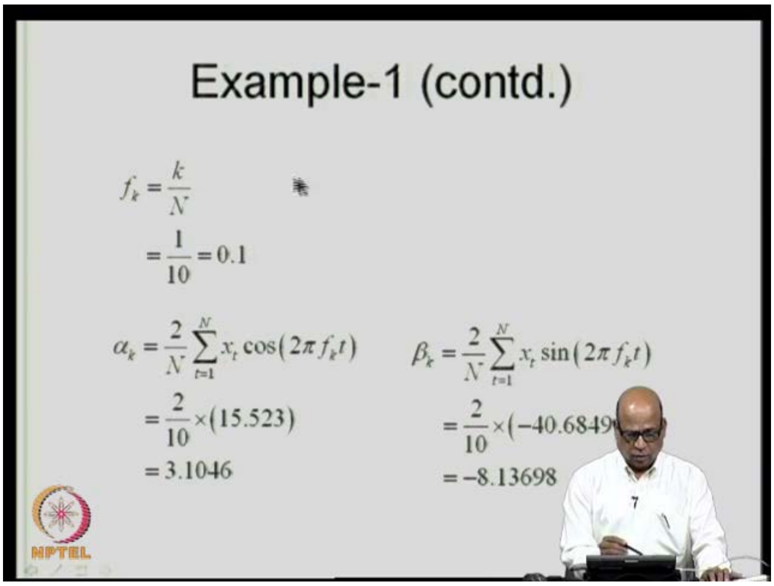
(Refer Slide Time: 20:30)



t	X_t	$\cos(2\pi f_k t)$	$\sin(2\pi f_k t)$	$X_t \cos(2\pi f_k t)$	$X_t \sin(2\pi f_k t)$
1	105	0.809	0.5878	84.945	61.719
2	115	0.309	0.9511	35.535	109.3765
3	103	-0.309	0.9511	-31.827	97.9633
4	94	-0.809	0.5878	-76.046	55.2532
5	95	-1	0	-95	0
6	104	-0.809	-0.5878	-84.136	-61.1312
7	120	-0.309	-0.9511	-37.08	-114.132
8	121	0.309	-0.9511	37.389	-115.083
9	127	0.809	-0.5878	102.743	-74.6506
10	79	1	0	79	0
Σ				15.523	-40.6849

So, let us get an example here this is the time series we have 10 values here this is just for demonstration you have 10 values of X t. We are determining omega k and I K for k is equal to 1. Because k is equal to 1 you know f k is equal to k by N. N is 10 in this case therefore, f k is equal to 0.1.

(Refer Slide Time: 20:47)



So, here we write t is equal to 1 X t is known 1, 0, 5 these are the observed values of the time series 2 pi into 1 by 10 into t in this case it is 1 in radiance. So, cos of 2 pi f k t is 0.809. Similarly sin of 2 pi f k t is 0.5878 and so on. Then we write X t cos 2 pi f k t and

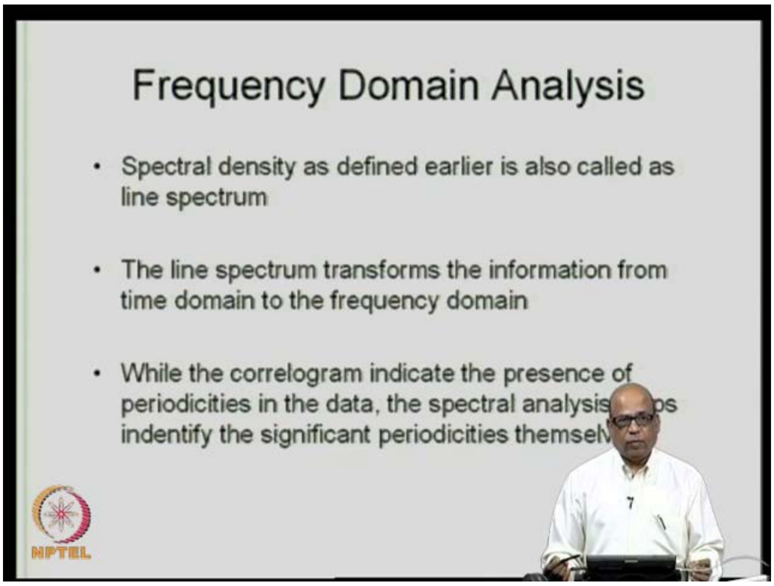
X t sin 2 pi f k t what are we doing here? We are looking at this expression I K is equal to N by 2 alpha k square plus beta k square. So, for a given value of k we determine alpha k and beta k and for that given value of k we are determining I K.

So, the example is for k is equal to 1. So, for a given k first to determine f k by k by N and then apply this these expressions. So, X t cos 2 pi f k t, X t sin 2 pi f k t and you get summations here alpha k. We will look at alpha the expression for alpha k here is 2 by N X t cos 2 pi f k t t is equal to 1 to N. So, this is the expression we are using to estimate alpha k similarly for beta k we are using this expression. So, we sum over sum this over all the N time periods and get this value similarly sum this term you get this value. You get alpha k as 2 by N, which is 2 by N into the summation of X t cos 2 pi f k t and you get alpha k as 3.1046.

Similarly, beta k as 3.1046 similarly beta k you sum all the sin terms here X t sin 2 pi f k t terms over t is equal to 1 to n and you get beta k as minus 8.1398. So, for a specified k how to get alpha k and beta k. Once alpha k and beta k for all k you determine I K, once you have determine alpha k and beta k for let say k is equal to 1 to M. M is the maximum lag, you can determine I K for all these k because alpha k and beta k are know.

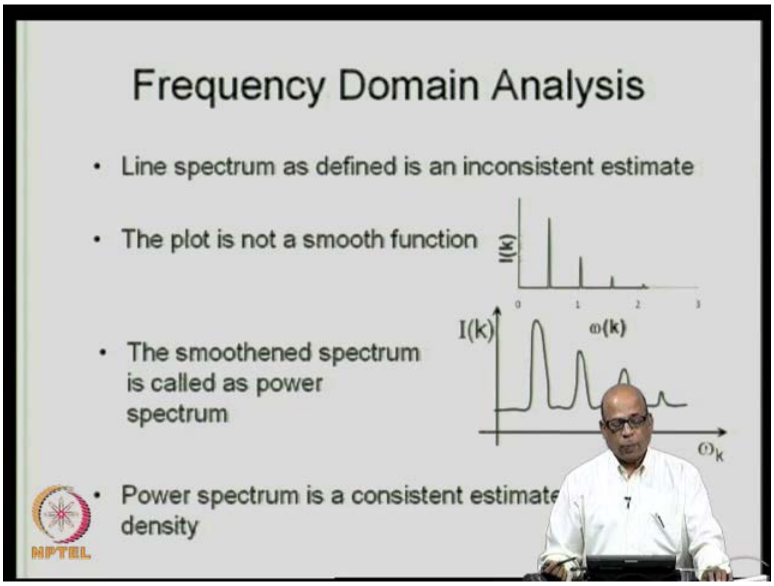
Similarly, you know omega k 2 pi k by N. So, for a given k you know omega k and I K plot omega k on the x axis and I K on the y axis you will get the spectrum. So, this is how we estimate the spectrum. Now what we are doing while calculating alpha k beta k and then omega k etc is essentially to capture the inherent periodicities in the data. So, if you have plotted a correlogram in the time domain. Correlogram would have already given you some idea about existence of periodicities, now this will be verified in the spectral analysis and also the information contained the information on the frequencies or the periodicities inherent in the data comes out more prominently in the frequency domain and you can also examine the significance of the periodicity.

(Refer Slide Time: 25:05)



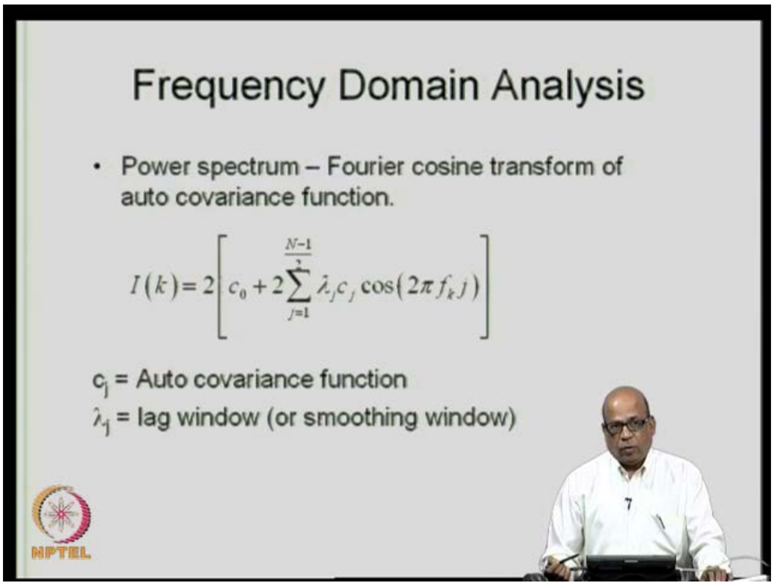
So, you identify the periodicities and then examine the significance of the periodicities. The way we estimated the spectrum and plot of I K versus omega that we just did is called as the line spectrum. Now the line spectrum transforms the information from the time domain to the frequency domain. As I said while the correlogram indicates the frequency presence of periodicities in the data. The spectral analysis helps identify the significant periodicities themselves. We will do some examples by which we can demonstrate this fact.

(Refer Slide Time: 25:30)



The line spectrum as we had defined is a statistically inconsistent estimate and the plot that we get typically is a is not a really a smooth function you get spikes like this and then these need not be actually zero, but the there would be some lines like this and then there another peak another peak and so on. So, this is not a really a smooth function that we have defined. We smoothened the function and redefine the line spectrum this is generally called in hydrology literature at least as the power spectrum. So, the smoothened spectrum is called as the power spectrum for the same graph it may appear something like this. So, you may have smoothen values this is the power spectrum as you define present just now we will be is called as the it is a consistent estimate of the spectral density.

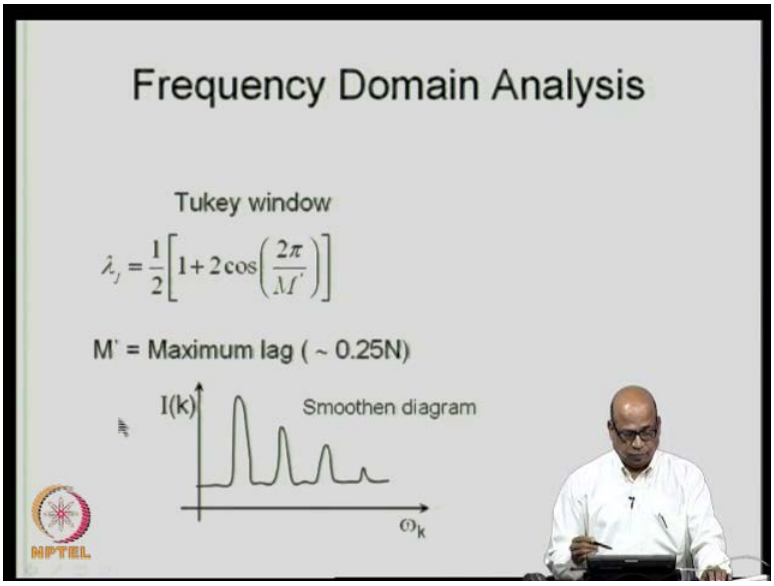
(Refer Slide Time: 26:40)



So, we define the power spectrum is a this is actually a Fourier cosine transform of the auto covariance function. So, we know how to determine the auto covariance functions c k if you recall and we transform the auto covariance function as a Fourier cosine function and then define the power spectrum function. So, the power spectrum I (k) is given by 2 c naught plus 2 into summation j is equal to 1 to n minus 1 by 2 lamda j, c j, c j are the auto covariance functions cos 2 pi f k into j. So, in this summation you are talking about j here these are called as lamda j's are called as a lag windows and there are ways of estimating the lag windows for example, we talk about tukey window window and so on.

We will introduce one of the methods of estimating lamda j's and all other terms are defines c naught is your auto covariance function at zero which is just the covariance. In fact, it becomes the variance.

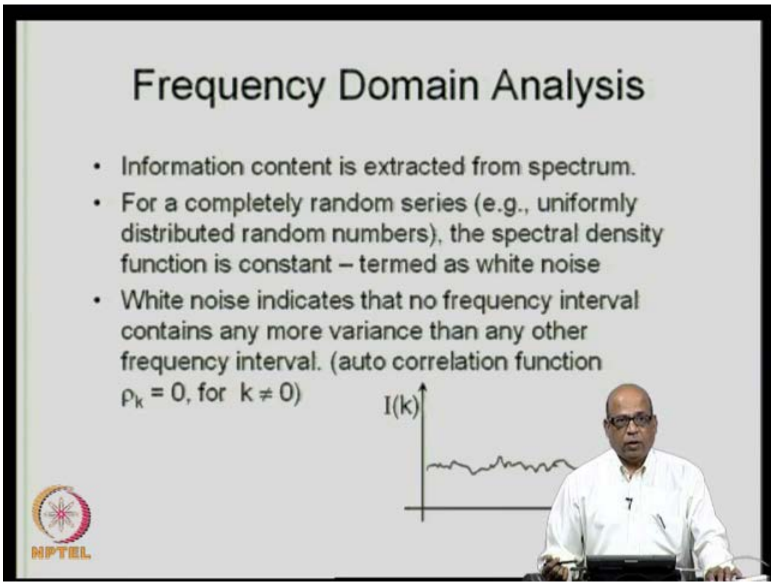
(Refer Slide Time: 28:01)



The lamda j's are estimated by what are called as as a there are various expressions for expressing lamda j for determining estimating the lamda j's. So, lamda is equal to 1 by 2 this is given by tukey window and this is most commonly used tukey window and window these are typically used for estimating lamda j's. We introduce the tukey window here. So, lamda j is equal to 1 by 2, 1 plus cos 2 pi by M dash. Now, this M dash is slightly different from the M that we talked about the maximum lag. Now M dash can also be about point 25 N, a general guide line for M dash or the maximum lag is that it should not be too small nor should it be too large.

So, that you are not missing the information contained in the frequency distributions. So, typically we use about 25 percent of the data and some author also recommend 2 root N you go up to 2 root N. If you have 100 data 2 into root of 100. So, about 20 whereas, if you have 100 data this may indicate about 25 values. So, typically it should not be too large nor should it be too small the smoothened spectrum which is omega k versus I K once you estimate lamda j's you can estimate I K here the covariance functions are known. So, you estimate the covariance functions and you can estimate I K and omega k is the same as what we did earlier 2 pi k by N and you plot w k versus I K this is called as the power spectrum. So, this will be a much smoother diagram and it is also a consistent estimate.

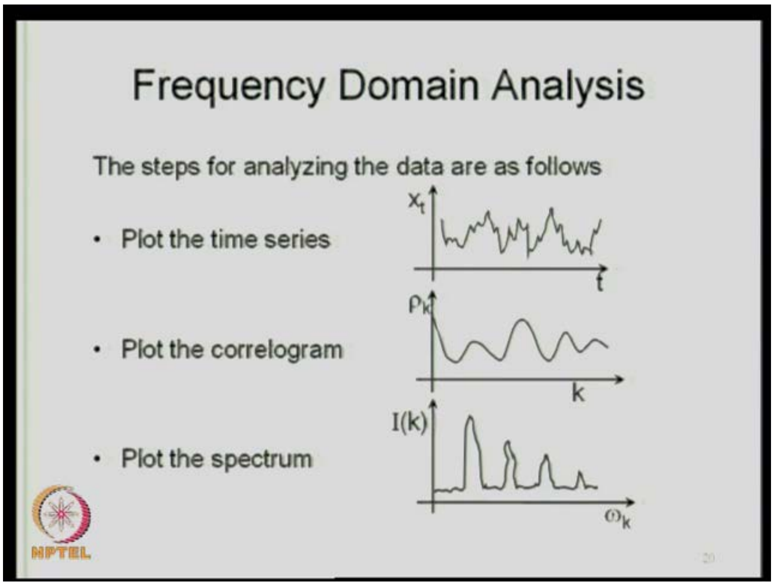
(Refer Slide Time: 30:04)



Now, in the frequency domain analysis the information is extracted from the frequency domain that is from the spectrum. Let say you have a completely random sequence for example, you generate using calculator a series of uniformly distributed random number in the interval 0 and 1 and you plot the spectral density for a completely random sequence you will see that the spectrum oscillates like this; that means, you are unable to get any particular peaks in the spectrum. What does this mean? This means that the variance is rather uniformly spread there is no particularly frequency band, which has more variance contained than any other band such processes are called as white noise in the spectral.

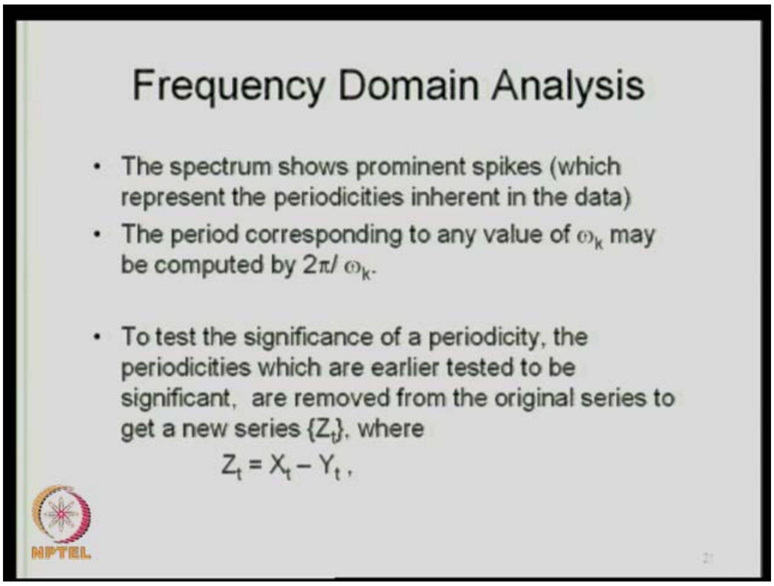
So, the white noise indicates that no frequency interval here contains any more variance than any other frequency interval and you recall that if you have a purely random sequence your rho k is equal to 0 for k not equal to 0 and this is the theoretical rho k and if you have sample estimate r K the r K will be all significant for all k not equal to 0. So, r K will be insignificant for all k not equal to 0 what do I mean by significance if you recall we draw a band of 2 by root N. Actually 1.96 by 2 root N on either side plus 1.96 by root N and minus 1.96 by root N this is the 95 percent significance level. So, if all your r K or the lag k correlations for within this band they are all insignificant. So, for a purely random process you will have all r K's being insignificant and for a purely random sequence if you draw a line spectrum or a power spectrum the spectrum looks like this. Which indicates that no frequency band contains anymore variance than any other frequency interval that you consider.

(Refer Slide Time: 32:36)



Typically what we do is in the in the run up to building time series model first we plot the time series, then we plot the correlogram. We may suspect that the correlogram indicates some frequencies or some periodicities inherent in the data then we would like to be sure that we remove the periodicities form the time series. Because when we are building time series models X t is equal to d t plus e t you want to first identify the d t once you identify the d t you remove the deterministic component. So, that you can model only the stochastic component.

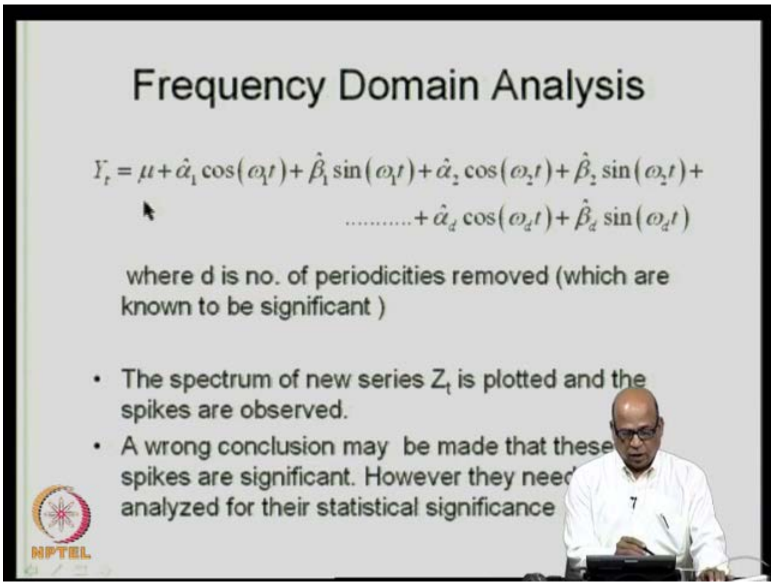
(Refer Slide Time: 33:44)



To identify the periodicities we then plot the spectrum either the line spectrum or the power spectrum, smoothen spectrum and indentify these periodicities. We also check for which of these periodicities that we have identified are. In fact, significant. So, that you can use them in your models the spectrum as we showed just now, it shows several spikes like this. Prominent spikes which indicate that there are periodicities inherent in the data. The period corresponding to any particular value of omega k may be computed by 2 pi by omega k. So, let say you have a spectrum like this you identified that this is the peak and this corresponds to a particular omega the period corresponding to that omega k will be simply 2 pi by omega k. Then let say that we identify that this is a period and this is significant, but we are not sure that this is significant whether this is significant and so on.

How do I test once we know that a particular peak is significant let say this periodicity is significant. We remove the periodicity corresponding to this reconstruct the series re-plot the power spectrum and then examine the periodicity for the next. So, the way we do is we test for the significance of periodicities one by one. Let say you remove the periodicity and test for the next highest highest peak remove that and test for the next highest peak and so on. So, this is the way we test for significance of periodicities.

(Refer Slide Time: 35:47)

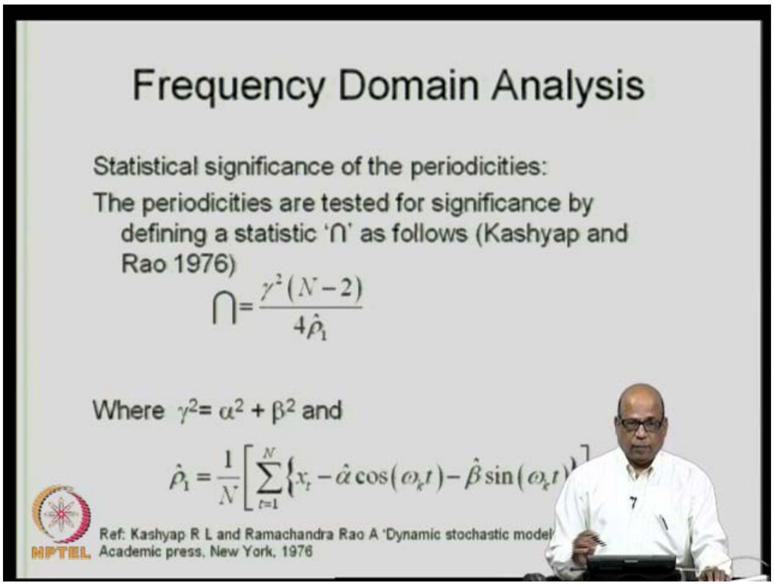


So, what do we do we reconstruct the time series Z t is equal to X t minus Y t in which Y t is the series containing the previous periodicities. So, we reconstruct it as X t minus Y t where Y t is defined as let say that you want to remove d number of periodicities: 1, 2, 3 etc up to d number of periodicities. So, you write Y t as your frequency domain notation as Y t is equal to mu plus alpha 1 cos 1 omega 1 t N and. So, on this is corresponding to the first periodicity omega 1 this term is corresponding to the second periodicity omega 2 and so on. So, you deduct terms corresponding to d number of periodicities omega d t.

So, we are actually removing d periodicities through this series Y t and your constructing Z t is equal to X t minus Y t with d number of periodicity is removed typically in hydrologic application. We remove one periodicity, two periodicities, three periodicities and so on not more than that. So, we construct the series Y t and reconstruct for X t and the spectrum of new series Z t is plotted and the spikes are again observed. Now, when you reconstruct this typically what we see again we see some prominent spikes, let say you removed one periodicity and then reconstruct the time series plotted the spectrum for the reconstructed time series you will see again significant peaks visually. You can see that there are significant peaks because you have removed the earlier significant periodicities the remaining periodicities now appear very prominently in the series in the spectral density spectral diagrams.

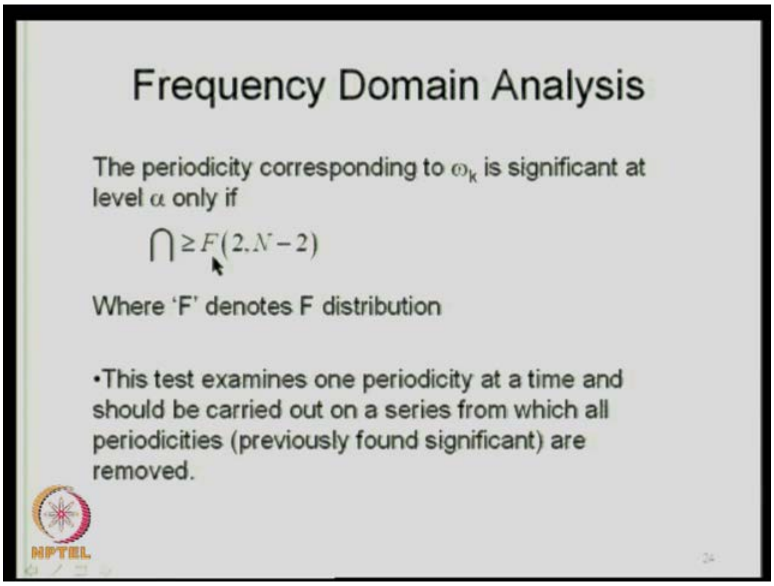
Just by visually examining we may tend to conclude wrongly that these periodicities that we are seeing are statistically significant because they appear very prominently in the time series in the transform time series, but it is necessary for us to have statistical methods by which we can test the significance of the the periodicities that we identify and that is what is done we introduce a statistics gamma square N minus 2 by 4 rho 1 cap this is given by Kashyap and Rao 1976 the reference is here.

(Refer Slide Time: 38:08)



So, gamma square is alpha square plus beta square and rho 1 cap is estimated by this t is equal to 1 to N this is the particular periodicity which we are examining. So, omega k as I have identified in your spectral analysis you have identified a particular k and this is the periodicity which you want to examine for significance. So, omega k is known and t you are summing from t is equal to 1 to N. So, X t minus this particular periodicity removed that gives you rho 1 k.

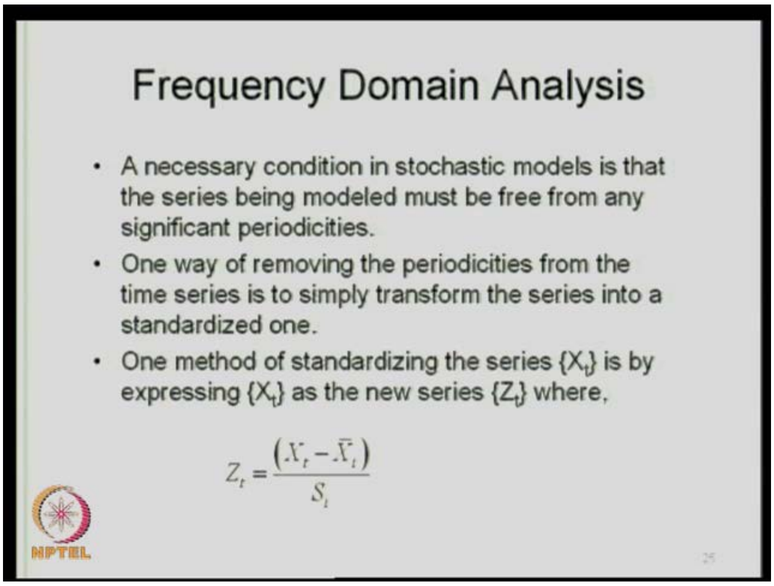
(Refer Slide Time: 39:21)



This alpha and beta are also for that particular k that you are considering. So, you know gamma square and n is known rho 1 cap is estimated and therefore, you know this statistic we check against the f distribution with 2 degrees of freedom N is the number of data points. So, if the statistic that we computed is greater than the corresponding f value then it is significant the periodicity at level alpha is significant. This test examines one periodicity at a time and it should be carried out on a series from which all the periodicities previously found to be significant are removed. So, this is done one by one remove the first periodicity test for the second one remove the second one test for the third one and. So, like this you can do.

Now, when we build stochastic models for example, you are building models arima type etcetera which will subsequently introduce you need the series to be devoid of any periodicities. So, you would have identified the you should identified the periodicities first remove the periodic component form the series reconstruct the series and then on the reconstructed model reconstructed series you build the time series model stochastic models.

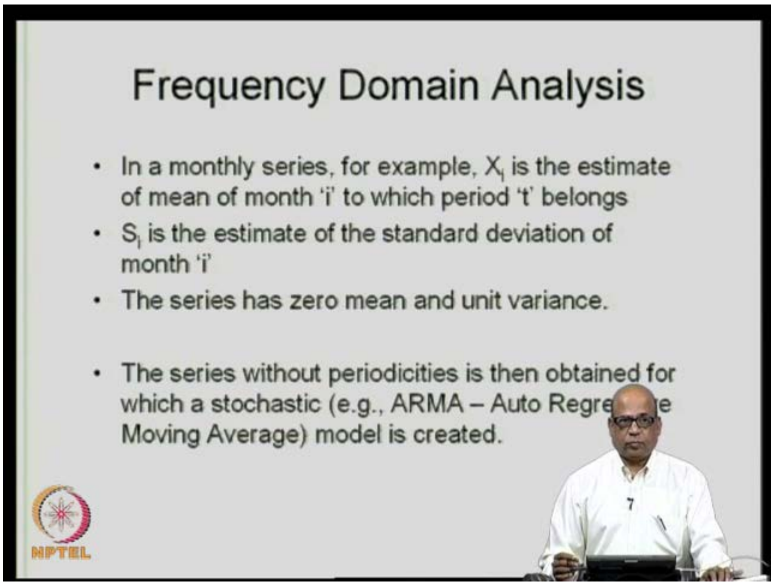
(Refer Slide Time: 40:49)



So, a necessary condition in stochastic models in most stochastic models is that the series being modeled must be free from any significant periodicities now let sat that we have identified the series consist of several significant periodicities how do we remove these periodicities now one way of doing this that is we want to remove all the periodicities inherent in the data one simple way of doing it is simply standardize the series that is equal to X t minus mu over sigma that is the long term mean by the standard deviation.

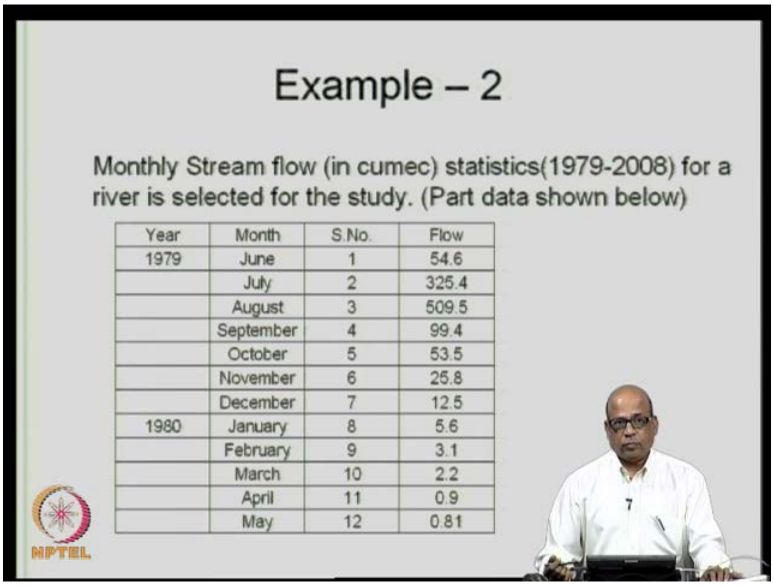
If we normalize the series and then examine the particular spectral density you may see that most of the periodicities are removed in many of the hydrologic applications this standardization works is a standardization or normalization works as a as a first step then there are also techniques for differentiating and. So, on of the series which will remove certain deterministic components.

(Refer Slide Time: 42:00)



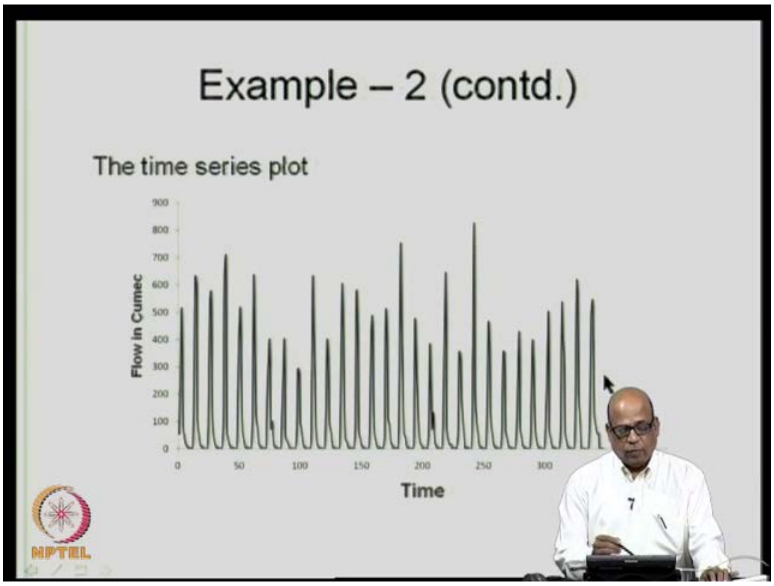
So, in a monthly time series for example, in this expression we may write X t minus x i bar over S i let say you are talking about monthly time series where t is equal to 1 to twelve and you are talking about the flows. Now X i bar here indicates the mean of the particular month of which t belongs for example, mean of the january mean of the february and so on. Similarly S i is the standard deviation of the month to which t belongs. So, like this we standardize the series the series that we standardized has a 0 mean and unit variance. So, this series is we can examine we will do some examples subsequently we will examine that the Z t that we. So, obtained by transforming X t will be in most cases divide of any periodicities we can use Z t in our stochastic models.

(Refer Slide Time: 43:14)



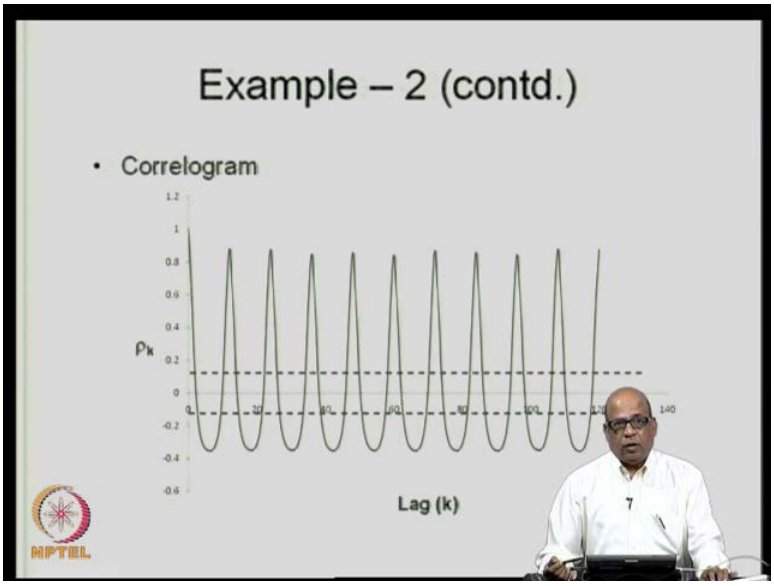
Let us look at a time series the statistics are available let say the flow are available from 1979 to 2008 they will be available in this form that is the monthly stream flow will be available in this form I have shown only one year data from june to may, but like this it is available from 1979 to 2008.

(Refer Slide Time: 43:41)



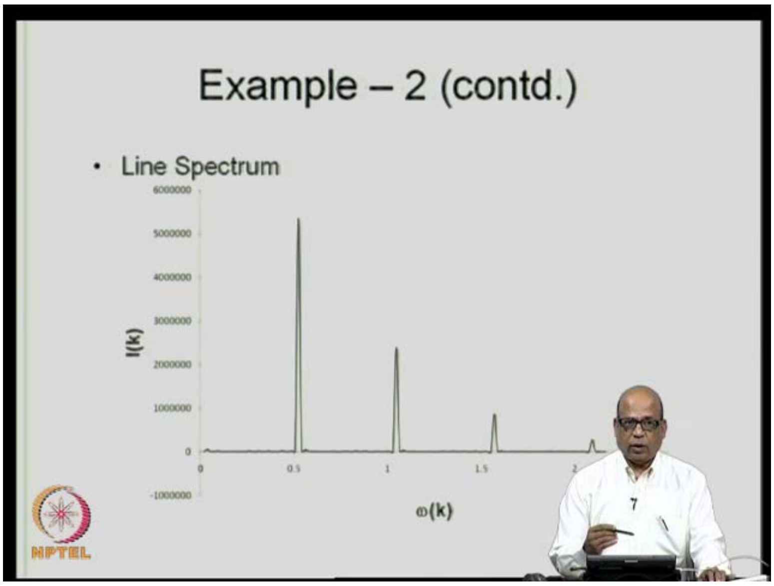
So, you have the series of flow observed we plot the time series first. So, you have around 3 and 40 some or values here. So, you plot the time series the time series looks like this, immediately just the visual examination of the time series indicates that there are there must be some periodicities inherent in the data. As you are seeing the time series itself shows some oscillations there and therefore, you suspect that there must be some periodicities inherent in the data.

(Refer Slide Time: 44:10)



Then we plot the correlogram how do we plot correlogram you know for different k you can estimate rho k and then you plot rho k versus k the correlogram clearly indicates that there are periodicities here. You can get the peaks of the correlation on either side and then you can suspect that the periodicities are let say corresponding to k is equal to 6, k is equal to 12, k is equal to 18 and so on. This indication that the time series and the correlogram together have given, namely that there are periodicities inherent in the data we now confirm or reconfirm with the spectral analysis.

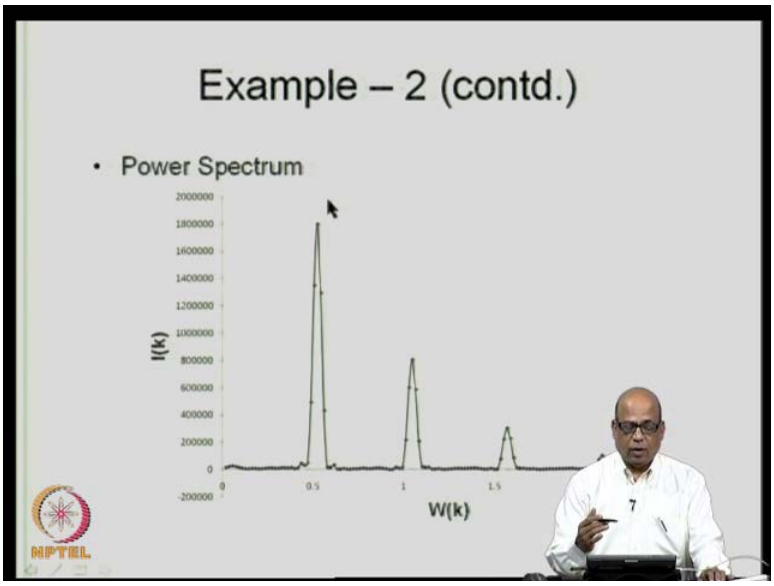
(Refer Slide Time: 45:26)



So, we use the time series and draw the line spectrum, while the correlogram showed a smooth oscillations like this. In the time domain we convert this the time series into frequency domain and then do frequency domain analysis plot the line spectrum immediately you will see that there is one significant spikes at a certain omega k somewhere around 5. Similarly, another omega k around 1.1 or such thing and so on. So, the line spectrum immediately extract the information of the frequencies and then shows that there may be a periodicity here etc. If you recall again that what this shows is that there is a significant contribution to variance of the process around these frequencies much more significant contribution corresponding compare to any other frequencies somewhere here.

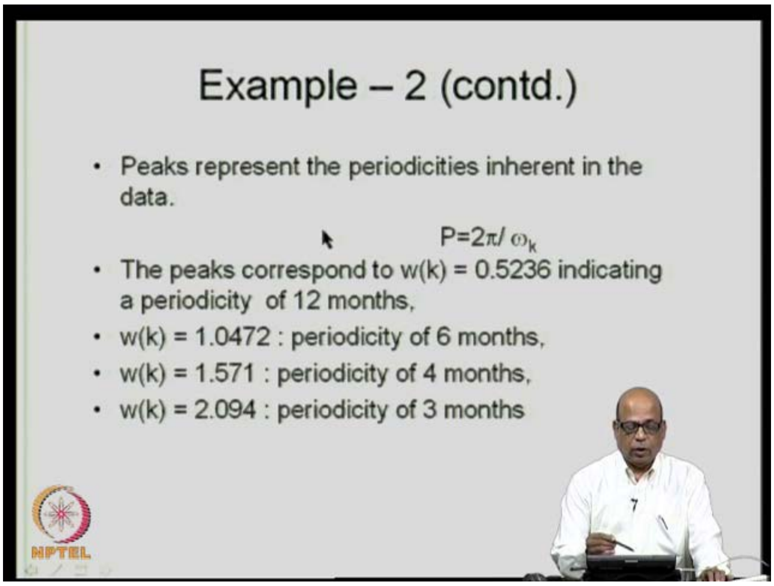
Similarly, suddenly you come across another spike where there is a significant contribution to variance around these frequencies and so on. So, these indicate that hese are periodicities that are present here. This is the line spectrum which is an inconsistent estimate and then we also provide the power spectrum.

(Refer Slide Time: 46:25)



So, power spectrum also as you can see between the line spectrum and the power spectrum this is a slightly more smoothen of the version. Where you have a slightly broader and smoother band around which the variance the contribution to the variance are prominent.

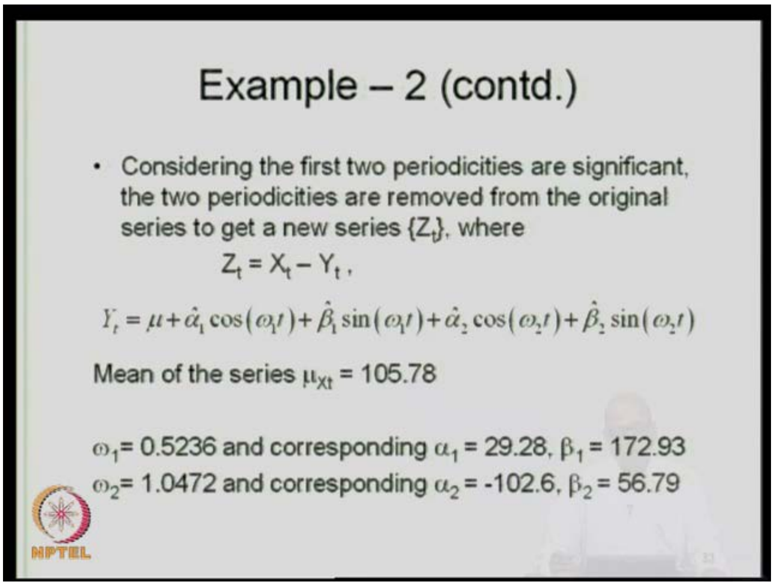
(Refer Slide Time: 46:57)



So, we identify this particular omega k value where we are seeing the peak in this particular case it is 0.5230. So, this peak occurs at 0.256 both for the line spectrum as well as for the power spectrum and the corresponding periodicity p is simply 2 pi by omega k where the omega k is the value of omega k corresponding to this particular peak. So, 0.5236 will be the periodicity corresponding to this will be 2 pi by omega k which corresponds to 12 months this is a monthly time series. So, we directly get the periodicity as 12 months.

The next peak here occurs at omega k of 1.0472 this corresponds to 6 months then you do 2 pi by 1.0472 it. You come to around 6 the actual value may be just 6 point something etc, but you can you can say that the periodicity is 6 months similarly the next one corresponds to 1.57 months and that indicates the periodicity of 4 months next corresponds to 2. 094 here and that corresponds to a periodicity of 3 months. So, from the time series data you now identify buy doing the spectral analysis. You now identify that there is a periodicity 12 months and 6 months, 4 months and 3months inherent in the data, but whether all of these periodicities need to be included in our time series models all of these are significant. Now this needs to be tested by the test that we just indicated.

(Refer Slide Time: 48:52)

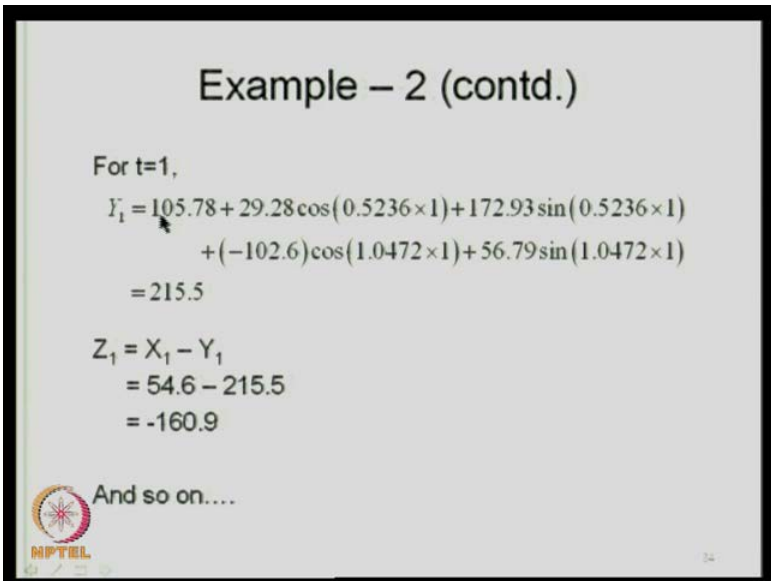


So, let say that we want to examine whether the third periodicity we will assume that 12 months and 6 months let say they are significant, but we are not. So, sure that whether the 4 month periodicity is significant whether the 3 month periodicity is significant and so on. So, to test this what we do is first you remove the first two periodicities that we have identified; that means, corresponding to omega k of 0.5 something and omega k of 1.0 something these two periodicities we remove. Reconstruct the data reconstruct the time series and then re-plot the power spectrum and then see whether these periodicities are significant.

So, first we reconstruct the time series as the t is equal to X t minus Y t we are removing the first two periodicities here. So, the periodicity corresponding to omega 1 is removed here, periodicity corresponding to omega 2 is removed here. where omega 1 and omega 2 are here this is omega 1 and this is omega 2. So, these two periodicities we are removing and then reconstructing the time series as Z t is equal to X t minus Y t. So, omega 1 is this and the corresponding alpha 1 is 29.28 and beta 1 is 72.93. Now these are obtained from your earlier expressions if you recall alpha k and beta k they are written here. So, for a given value of alpha, for a given value of k you know how to get alpha k you know how to get beta k. So, k is equal to 1 in the first case. So, you determine those and retransform the series.

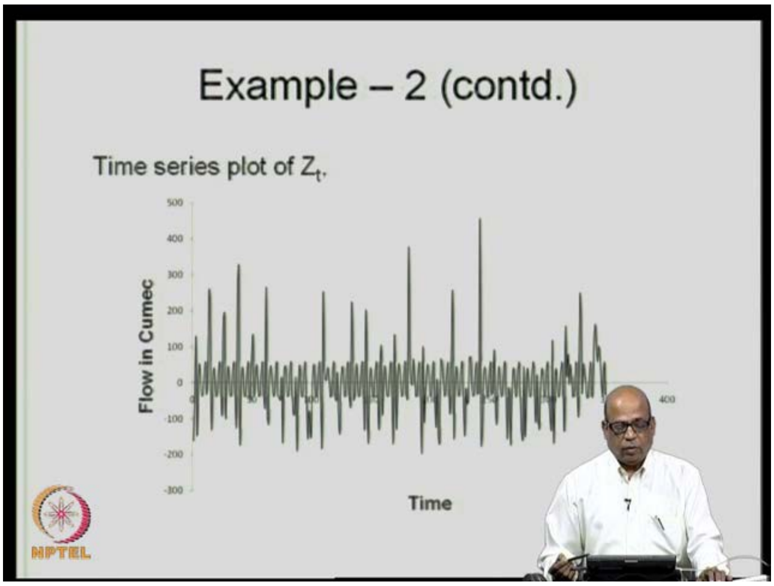
So, this is how you determine alpha 1 and beta 1 similarly for the second spike that we saw in the spectral density. You have omega 2 is equal to 1.0472 you get the corresponding alpha 2 and beta 2. So, here all the terms are known construct Y t series and then construct Z t series from your original X t from by deducting Y t from your original X t.

(Refer Slide Time: 51:18)



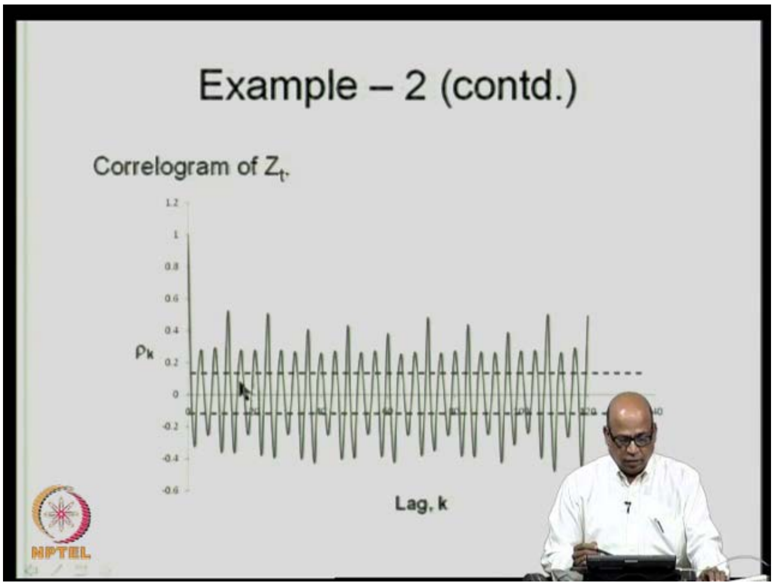
Say for example, for t is equal to 1 by putting all these values, you get Y 1 is equal to 215.5 and Z 1 is equal to X 1 minus Y 1. So, X 1 was in this data as you can see 54.6. So, 54.6 minus 215.5 you get minus 160.9 and so on.

(Refer Slide Time: 51:51)



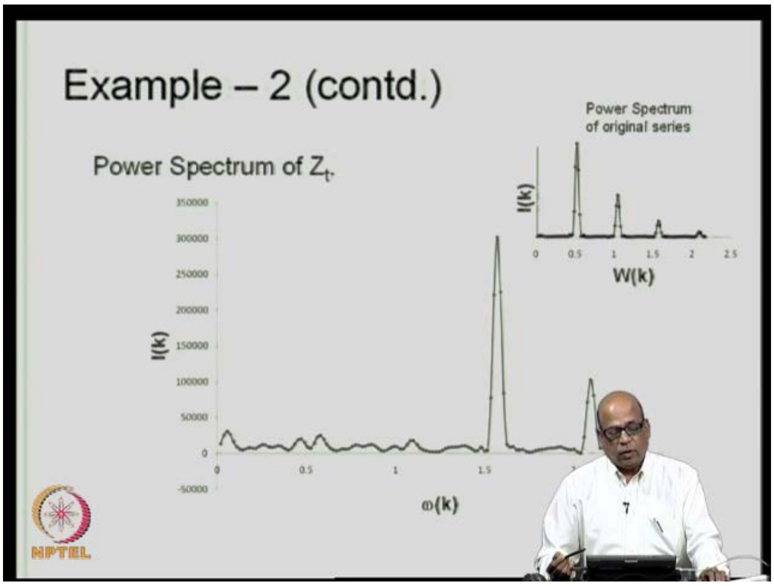
So, like this now you have constructed the Z t series. Now on the Z t series now this is how the time series plot of Z t looks. So, initially you started with X t time series the original observed data, you plotted the time series you suspect that there are periodicities present in that you plotted the correlogram which confirm that there are periodicities. Then you plotted the spectral density which brought out the periodicities corresponding to 12 months, 6 months, 4 months and 3 months. You removed the periodicities corresponding to 12 months and 6 months from the data by transforming the data Z t is equal to X t minus Y t. Where Y t is a series corresponding to the first two significant periodicities suspected significant periodicities and that is the series that have plotted here now.

(Refer Slide Time: 52:44)



So, this series looks something like this. Then we plot the correlogram, now correlogram these are the significant bands correlogram looks something like this. There are some peaks here it is still oscillating and there are some peaks which are way beyond the significance bands.

(Refer Slide Time: 53:06)



We plot the power spectrum, this is how it looks. Now the original power spectrum was somewhere here is shown here. In the original power spectrum, there was a peak at somewhere around 0.5 there is no peak here, because we have removed that. There was another peak at 1 point some value that is also removed. So, this is corresponding to 12 months, this is corresponding to six months, both of them are not absence not present here. The third peak which was occurring somewhere around 1.56 or some such thing that becomes prominent now. So, the effect of transformation that we did was to remove the linear periodicities, and bring out to the 4 the other periodicities. So, these two become prominent.

Now, we can remove this from this figure it appears as if this has to be significant, but we have the test that we can make for examining this significance. If we know that is significant remove that and then again re-plot the power spectrum and then check which other periodicities come out. The significance test of periodicities or the identified periodicities from the spectral analysis. How significant they are we conduct the significance test to answer that question that we will discuss in the next lecture.

So, in today's lecture what we did is we introduce the frequency domain analysis; in the frequency domain analysis we transform the time series X t into a series consisting of typically in the way we have introduced sinusoidal terms. And then we estimate the spectral density and we plot the spectral density against the angular frequency omega k and this gives the periodicities present an idea of periodicities present in the present in the data.

Then the way we have introduced the first expression that we have introduced is called as a line spectrum which is an inconsistent estimate and we converted that into a power spectrum is actually we did not convert the line spectrum into power spectrum. We introduced another expression for power spectrum which is a Fourier transform of the covariance function and we plot the power spectrum again I k versus omega k and then see that there are prominent spikes the prominent spikes that we see either in the line spectrum or in the power spectrum indicate the particular periodicities. The omega k corresponding to a spike can be transform it to the corresponding periodicity by P is equal to 2 pi by omega k.

So, in the monthly time series that we saw as an example, we got periodicities whether statistically significant or not needs to be tested, but we got periodicities corresponding to 12 months corresponding to 6 months, 4 months and 3 months. We remove the first two periodicities for example, reconstruct the time series by removing the first two periodicities, redraw the power spectrum of the transform series, and then we see that first two periodicities are not are not present in the revised power spectrum, but the third one becomes prominent, the fourth one becomes prominently visible. We need to test whether the periodicities that we have so identified from the spectral density, spectral analysis are in fact, statistically significant. This exercise we will do in the next lecture; thank you very much for your attention.