Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture - 01
Lecture 01A - Motivation and Overview 1

Very good morning, welcome to the course on Applied Time-Series Analysis. I hope you are here because you have registered or and not because you are playing Pokemon Go and you happen to be here.

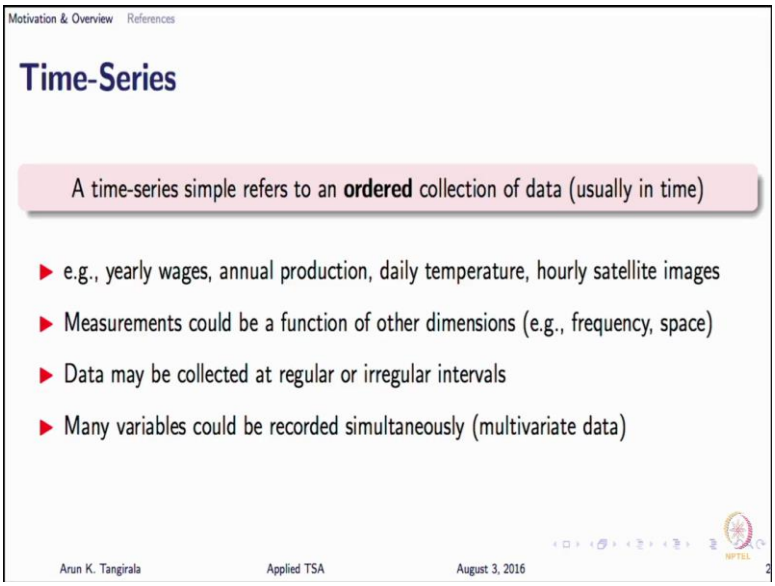
Today is kind of an introductory class, it is a kind of I would say a warm up class and I will of course, in this class give you an idea of what to expect to see in this course. This is one subject which I firmly believe that every engineer regardless of the discipline should be aware of because at some point in time each of you will be handling uncertainties either in your experimental data or in your simulations whatever may be the case, as engineers we are supposed to be well versed with handling uncertainties.

Unfortunately in lot of other courses, we are only taught how to handle exact problems, we have given exact numbers and you know very precise situations, which is a far cry from reality I should say. So, what we will do in this course is we will learn how to deal with the uncertainties in a particular context.

As the course reads, it is called applied time-series analysis and I believe some of you already have a feeling of what time-series analysis is about and the reason we have applied is because we are going to learn how to apply concepts of time series analysis is. The subject can be fairly theoretical, but what we will do in this course is we learn the theory and we will also learn how to practice it and no theory without learning how to practice is kind of meaningless as far as engineering discipline is concerned. So, we will place equal emphasis on practice and practical aspects of this subject and what I will do over the next maybe 15 to 20 minutes is kind of tell you what time series analysis is about give you a picture and then of course, you can make a decision whether you are here by accident or by intention.

Let us begin and ask ourselves as to what is a time series.

(Refer Slide Time: 02:30)



As many of you must have understood what a time series is, it is actually bunch of data arranged order in an orderly fashion in time we call this as time series, but gradually one has to learn to take a broader outlook and become independent of the domain in which we are looking at. In other words the generic word that we use is time series assuming that you are going to collect data or you are going to deal with data that has been collected over a period of time, but one does not have to do that you can actually collect data in space as well or as a function of frequency and so on.

So, what I am trying to tell you here is that this name or this phrase time series is a fairly generic one, you should not really get obsessed with the notion of time here, you can think of it as an index in which the data has been arranged in a certain order. So, order does matter in time series as against a lot of situations where you deal with steady state data. So, called steady state data where it does not matter when you collected the data, I mean if I have performed experiments and I have collected steady state data then I just need to have an idea of the time stamp, but the way in which I arrange the data for my analysis is probably in material, whereas in time series that is the key, the time stamp is the key because in time series analysis as you will learn in this course, we exploit patterns that occur in time that evolve in time what we call as correlations and so on and the moment you change the order of the data, these correlations can either break down or new correlations can appear and so on. One has to pay attention to the orderliness of data.

By the way you are free to ask any question at any time I do not like keeping things formal. So, please raise your hand so that the people in the recording studio can actually focus on you when you are asking the question.

So that is as simple as a definition that you can give for time series, of course, you will get a better feel of it as you start seeing it and where do you see time series data everywhere, there is no discipline that can actually claim ownership to time series data, in every discipline in engineering, social sciences, medicine or you can say biomedical engineering, econometrics, financial engineering, everywhere you would run into time series data which means that the principles that we learn in this course are fairly universal as well we will not really restrict ourselves to a specific domain as you will see even in the assignments and so on.

To just to give you a few examples of the resilience that you can see, you have yearly wages for example, or annual production, daily temperature and you know hourly satellite images and so on. So, when you are looking at time series data the other thing that needs to be kept in mind is that these message as I said earlier, these measurements could be a function of space frequency and not necessarily as a single function, you can have time series as a function of time and space or time and frequency or time space and frequency. So, you could have multi dimensional data as well and you can also have multi variable data. So, the examples that are being given here are only for a single variable like wages and so on, but in an experiment you would be recording many variables and that is typical in many industries in many applications if you take the atmospheric process you would be measuring for example, rainfall humidity sunshine temperature pressure and so on.

To deal with such data, we would need the principles of multivariate data analysis, but that is an advanced concept as it stands with respect to this course this course is an introductory one. By and large we will deal with univariate time series which means the single variable observations of a single variable and occasionally we will talk about by variate analysis that is when we look at 2 variables at the same time and of course, you know you can take these principles and understand the multivariate data analysis as well. And the other thing that should be kept in mind is that when we collect data and when I say, when we collect data, it does not mean that you have to collect data all the time data could have been acquired by someone else also.

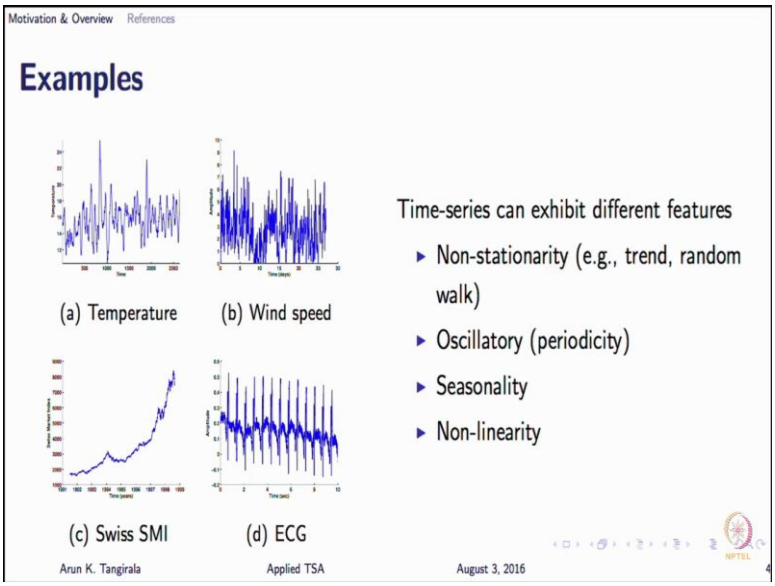
Let us be more passive and say the data that is with you could have been collected at regular time intervals or irregular time intervals and there maybe various reasons why you have irregular time series data what we mean by regular is the spacing in time is not uniform and this could happen by way of sampling itself that is it was perhaps not possible to observe data at regular time intervals. A classic example is if you walk into a process industry it is and you are the measurement that is being obtained is being obtained manually; that means, an operator actually takes a sample takes it to the lab uses analytical instrument and then obtains a reading all of this cannot be really regularized it depends on the operator shift when and so on.

There you can obtain irregular data that is a typical case or you can say for example, regular medical checkups of an individual an individual need not go at regular times to a doctor. So, if you are looking at blood pressure readings or sugar level readings and so on, this individual once may go in 3 months and an another occasion may go after 6 months and so on so that you will have irregular time series data where we encounter regular time series data is where you have an automated sensor which online sensor which is actually doing the measurements for you and of course, transmitting the data in some fashion to a computer.

In those cases, you can expect regular data even in such situations the sensor can fail at times and in such situations what happens is you will have occasionally irregular data and there is a whole lot of literature telling you how to deal with certain missing data. We will not really get into that such intricacies those are again considered advanced concepts we will assume by and large that data is obtained at regular time. So, we will all we will keep all the things at a very elementary level and yet there is enough to feel the heat.

And as I said earlier it is possible that many variables are being recorded simultaneously, but we will not really pursue that line by and large. So, to give you some examples of time series data I have just picked 4, but of course, you can one can actually run into zillions of examples.

(Refer Slide Time: 09:40)



So, what you see on the top left is the temperature measurement from an industrial dryer just to give you an example it is the x axis is a is a time and the y axis is a temperature and whatever you see is a time series, it is been collected over a single duration and then you have in the next top plot is a wind speed. Now this is the wind speed that has been recorded on campus there is an automated weather station on campus and still not sure if it is functional there are about apparently 600 and odd weather stations in our country which are actually collecting data you know at several atmospheric variables pressure temperature humidity and so on and wind speed is one of them and again you see the wind speed time series and the bottom left one is something that is of interest to many.

The stock market index which is perhaps also the reason why many students are enrolled in this course so of course, it is a Swiss stock market index. Over a particular set of years and you can see a different pattern there for the time series, it is completely different from what you have seen for the temperature and wind speed.

What is the difference that you see? There is an increasing trend. So, there is a trend that is the technical term that we use in time series analysis when you see such patterns and then you have at the 4th as a 4th example, ECG coming from a human, from a patient and that again as a different set of features you see oscillatory features and so on and this is going to be the case as you move from one time series to another time series one domain to another domain you will encounter data with different features and the idea is

to be equipped with tools that will allow you to handle this various or different kinds of time series and still analyze them, we have not yet talked about what we mean my analysis we are only talking about the first part of the course title time series - the analysis part is the most interesting and the challenging part.

What we want to know from this time series involves analysis, but we need to be equipped with some kind of universal methods that will allow us to deal with several different kinds of time series and that is the beauty of this course in the sense that whatever you learn is fairly universal, of course, the sum of the things can be domain specific. So, for example, the kinds of features that you see in the ECG data are missing from the temperature readings. So, you may not have to really invoke certain tools that are required to analyze this more complicated ECG data than the temperature data. In fact, what you generally see is different features as we classify technically as non stationarities, you can run into oscillatory kind of features like periodicities as you see in the ECG data what you see for the stock market index is an example of a non stationary series.

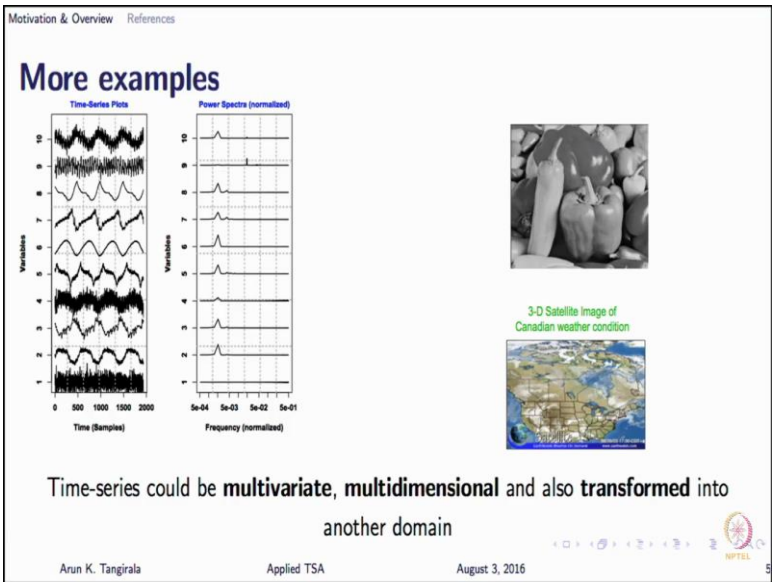
Now, we will go through technical definitions of all of this, at the moment I am just giving you just a flavor of what it looks like, we will not have a single equation today that is my promise to you. And then you can run into what are known as seasonalities, some of you are perhaps familiar with it, but we will not again go into the technical details of seasonality you can for now think of it as a generalization of periodicity and then you can have nonlinearities which today is one of them in non-linear time series analysis is one of the hottest topics that you can look around for in time series analysis and there are many processes which will not meet the assumptions that we will make in this course in this course we will by and large look at linear, so called linear processes where as many processes out there and actually non-linear in nature and you may require some advanced tools to handle that.

However, the principles that you will learn in this course are fairly generic and will serve as good foundations for you to understand non-linear time series analysis. Therefore, in all respects it is better to begin with an introductory course like this and of course, there are many universities that offer courses like this. But the unique feature is an engineering department is actually offering this because typically you will find this course being offered in the department of statistics.

And the nice thing about an engineering department offering this course is we look at the applied aspects more than worrying out ourselves about dilemmas and theorems and so on. We will assume, we will surrender to the statisticians and mathematician say yes you are right these theorems are correct no worries.

Now, as I said the series need not be univariate, you could have different kinds of time series.

(Refer Slide Time: 15:06)



On the left hand side you see what is known as a multi variable data where I show you time series from 12 different control loops in a pulp and paper manufacturing process, I will have to leave up to my parent department's name, I am from chemical engineering. So, one example has to come from chemical engineer having fulfilled that I am safe. So, this is actually simulated pulp and paper process from which we have drawn the readings. And on the left hand side you see what are known as a time domain readings from these control loops the objective of any control loop is to maintain those variables at their respective set points it is very simple, I mean you have air conditioner for example, here in this room or in the room where the other part of the class is seated the role of the controller in the air condition is to maintain the temperature of the room at the specified set point.

In industries typically one is concerned whether the controller is doing it is job and that is known as the control loop performance monitoring, we also have many control loops

operating within our bodies, they are trying to regulate temperature hormone level sugar levels blood pressure and so on. Here there also we are concerned with the health when we say I am concerned about my health or a go for a regular checkup what I am doing essentially is monitoring I am asking whether the control loops in my body are doing what they are supposed to be doing. So, likewise in industry we ask this question whether the controllers are doing what they are supposed to be doing as you can see from the time series data they are not actually doing what they are supposed to be doing why? So, the variables as you can see at least most of them even if one of them is actually oscillatory then there is a much it is a matter of concern.

Now, when you want to actually detect oscillatory features it is always nice to turn to the so called frequency domain which people dread to dread the frequency domain is considered a big no, no by many engineers and so on unfortunately. Whereas, the frequency domain analysis actually is one of the most power powerful domains for analyzing several features that you are searching for in time series. The moment you turn to frequency domain you can produce what is known as, you can calculate what is known as a power spectrum or a power spectral density and that is what you see on the right side of the plot as you can let me show you.

As you can see here, this is a time domain data that we have and then on the right, we have the spectral densities on the x axis, we have frequency that is common to all measurements, on the y axis we have what are known as normalized power spectral densities the values do not matter here for now because what we are searching for here are peaks in the power spectral density a peak in the power spectral density means that what does it mean? It means that frequency component is contributing significantly to the overall power in the signal which means in practical terms that frequency component is present significantly in the series.

As you can see here some of the signals of course, you can visually see whether these series have oscillations or not well, but for some others it is quite difficult because of presence of noise, measurement noise and so on or there may be multiple periodicities and so on. So, our ability to detect oscillations by visual inspection is fairly limited where as the Fourier analysis or the frequency analysis brings out these features in a very easy fashion in a very obvious way, as you can see here many of these series on the left have peaks in the power spectral their respective power spectral densities indicating that

they have oscillations in them and you can also figure out what is the frequency or the frequencies of those oscillations.

These examples is to give you a feel of what you will see in this course of course, we will deal with a single time series we learn how to compute power spectral densities what is a theoretical definition of power spectral density what is the theoretical definition of a periodic process and so on and then in that process go through a review of the Fourier transforms and so on.

Now, that is where usually students feel the heat and as I said earlier, there are certain ways of actually making sure that students can fit in one studio by beginning with Fourier analysis, but we will not do that. On the right hand side you have an image, it is a 2 dimensional image just to make things lighter and also lunch time is approaching. So, this image is indicative of that I am not saying that you will be fed with this, but maybe close to what you see.

This is also considered a time series an image is considered a time series, but it is not essentially a time series a spatial series in this case there is no notion of time the image is usually considered as the function of space and then at the bottom you have a 3 dimensional satellite image that you get again this is not a time series literally, but it is more of a spatial data that you see. What I am trying to say here is the time series that you will run into could be of different nature, it could be a function of time, could be a function of space, univariate multivariate, multidimensional and so on. So, we will of course, keep things simple to begin with, we will deal with univariate time series.