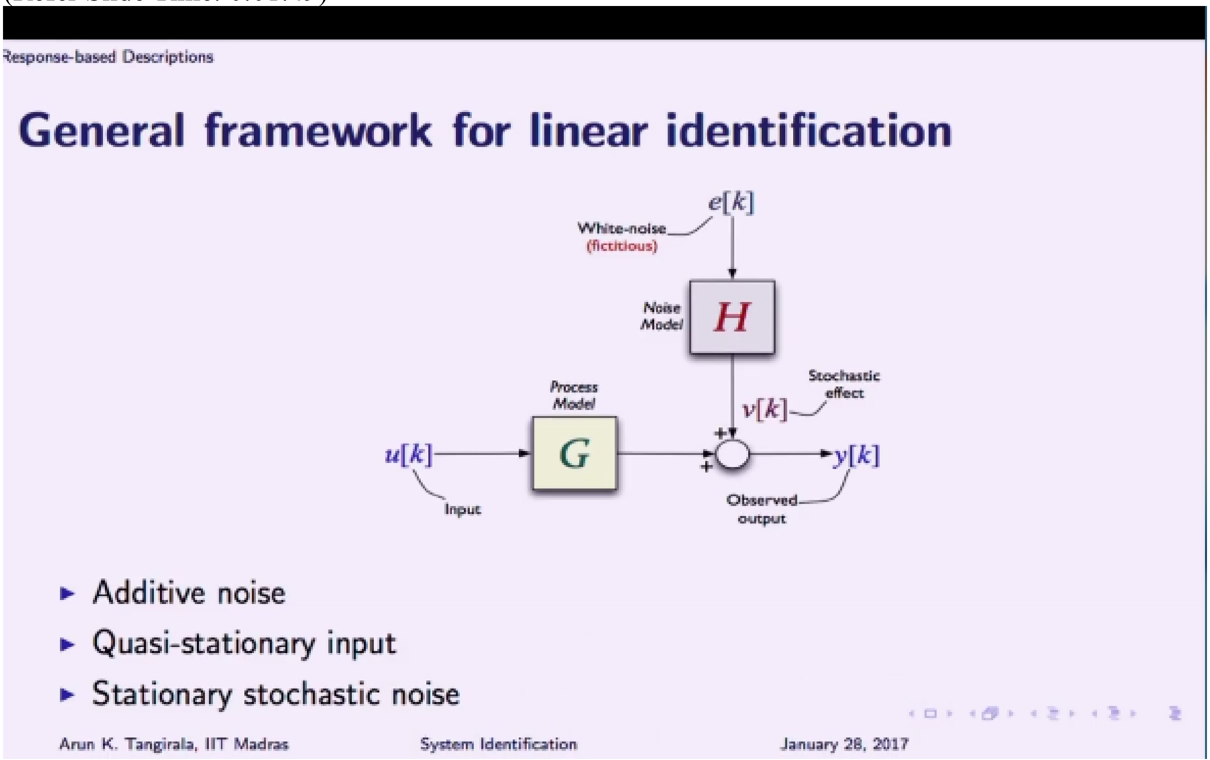
Lecture – 12

Response based Description-1

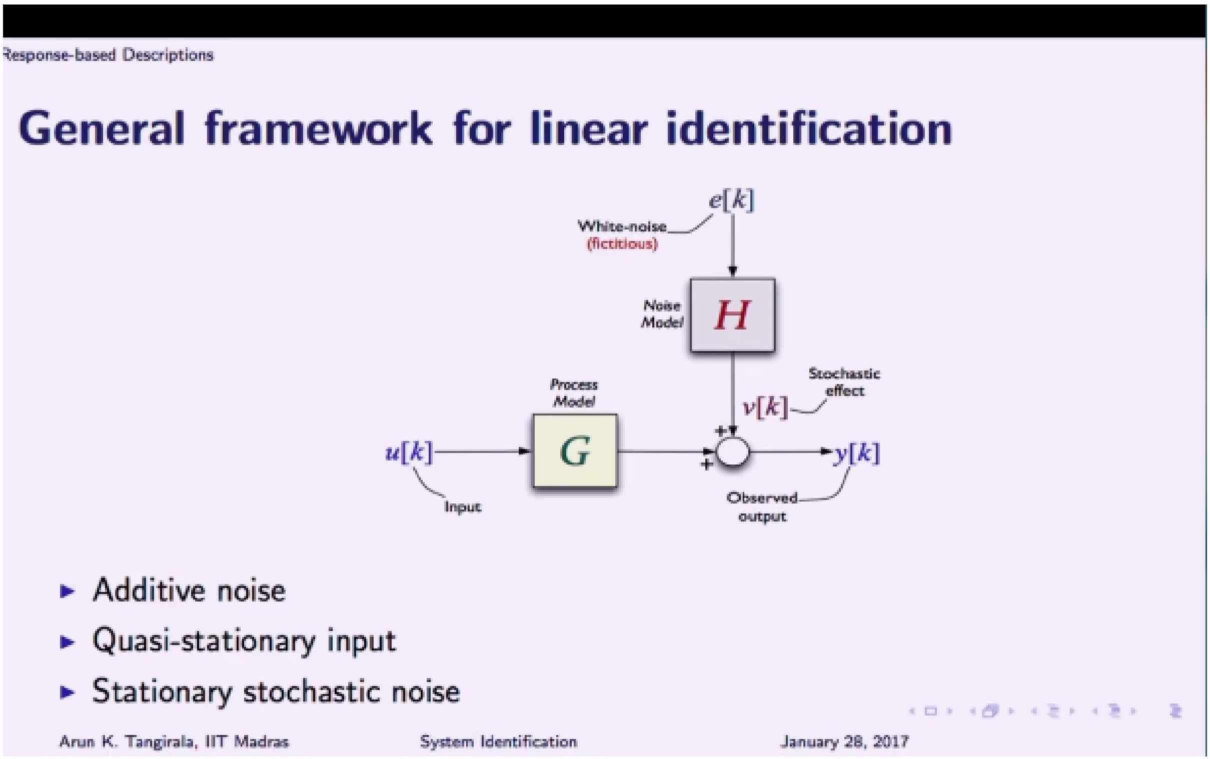
Here onwards our journey begins-- the theoretical journey begins. And remember, in fact, you should recall that system identification is a confluence of four different fields, right? You need to know the linear systems theory; you need to have some background in estimation theory and random processes and some amount of signal processing as well. So, let us understand, at least let me give you some overview of how the rest of the course is laid out with respect to this schematic that we keep seeing all the time in this course and maybe in-- typically in any system identification exercise. You have seen this before, but now we have more symbols in place. Remember, this is a framework that we had set out for identification. We said that, there is an input here, which excites a system G, the deterministic system G, the response of which is corrupted by v, which is a lump defect, right? It's a lump defect that contains contributions from sensor noise, from effects of unmeasured disturbances, and in practice, it may also contain modelling errors. And then comes out your measurement y. So, we have u, there which is the input and output y. In fact, very often in a system identification you keep asking why, why all of this-- This because of this y here, okay?

(Refer Slide Time: 0:01:49)



Now, the important thing that you should observe is, we have mentioned earlier is Additive noise. We have assumed additive noise, then we assume that the inputs are so called Quasi-stationary, which will define later on. Essentially it means that the inputs are well behaved in the sense that they don't explode in time, that or they just don't die out in time, persistently exciting and so on. Okay? So that is one of the assumptions. And the third is that this noise v, which is a lump defect, is so-called stationary. It's a stochastic signal. It's a stationary signal, now stationary should not give you the impression that it is at steady state. [02:37 inaudible] in the random signal world means that it statistical properties are at steady state.

(Refer Slide Time: 0:02:41)

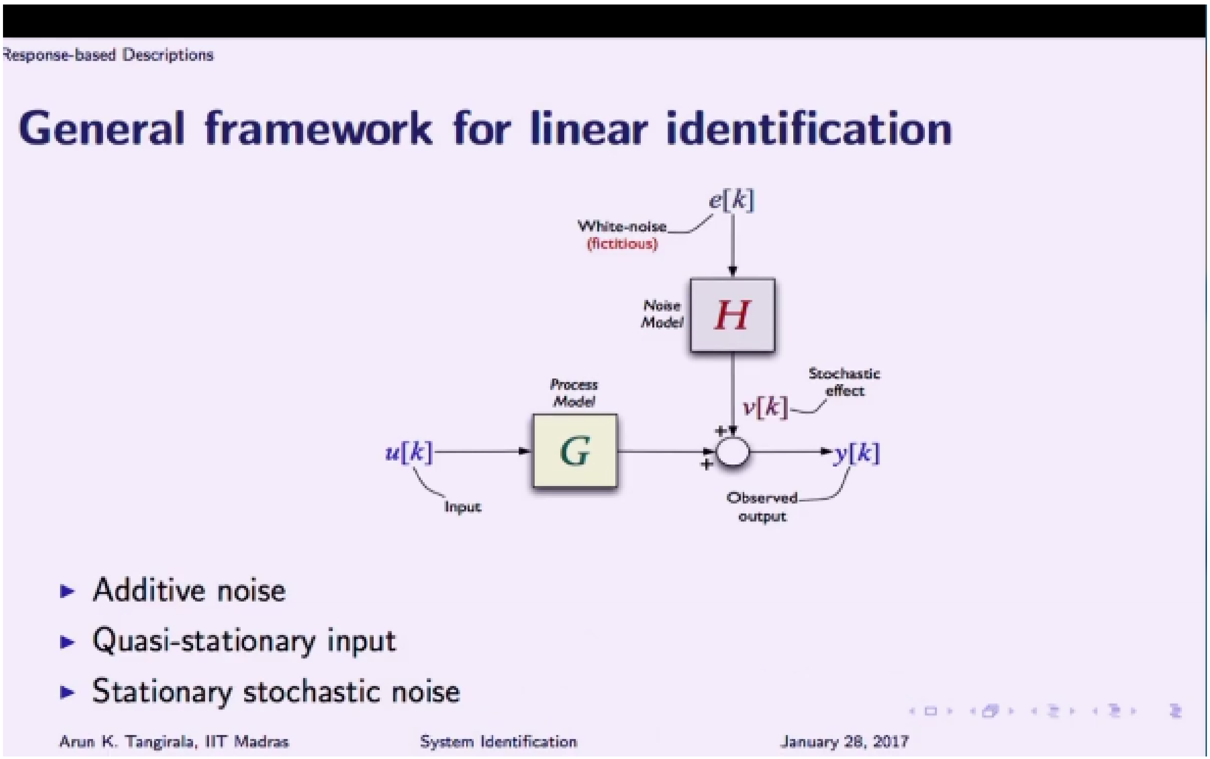


Okay, what do we mean by statistical properties: mean, variance, the nature of correlation, internal correlation between the two samples or sorry, two observations and so on. So we'll talk about that a bit later, but if you are already some of you, if you're enrolled in A Time Series Analysis course, you should know or some of you've already taken Time Series Analysis course and gone through it successfully. You should know what a stationary signal is and so on, okay? The reason for putting up this slide, in fact, there are two reasons.

One is that upfront, it is clear to you what is the framework under which we are developing the model. We know the walls, we know the peripheries of our modelling. So that at some point in time having tried everything and still the model doesn't work, you know, perhaps that the model is not within the walls of this. Maybe you have to step outside, maybe heard assume nonlinear models, right. Of course, have not mentioned the G and H are linear, but I already said it is linear identification. So which means it's assumed that G and H are linear and time invariant. So, if all your efforts fail to fetch a good working model, you have to knock at the doors of these assumptions and maybe open them, enter a different world where you may have to consider a nonlinear model or a linear time varying model, or maybe multiplicative noise or you know or maybe non stationary disturbances and so on.

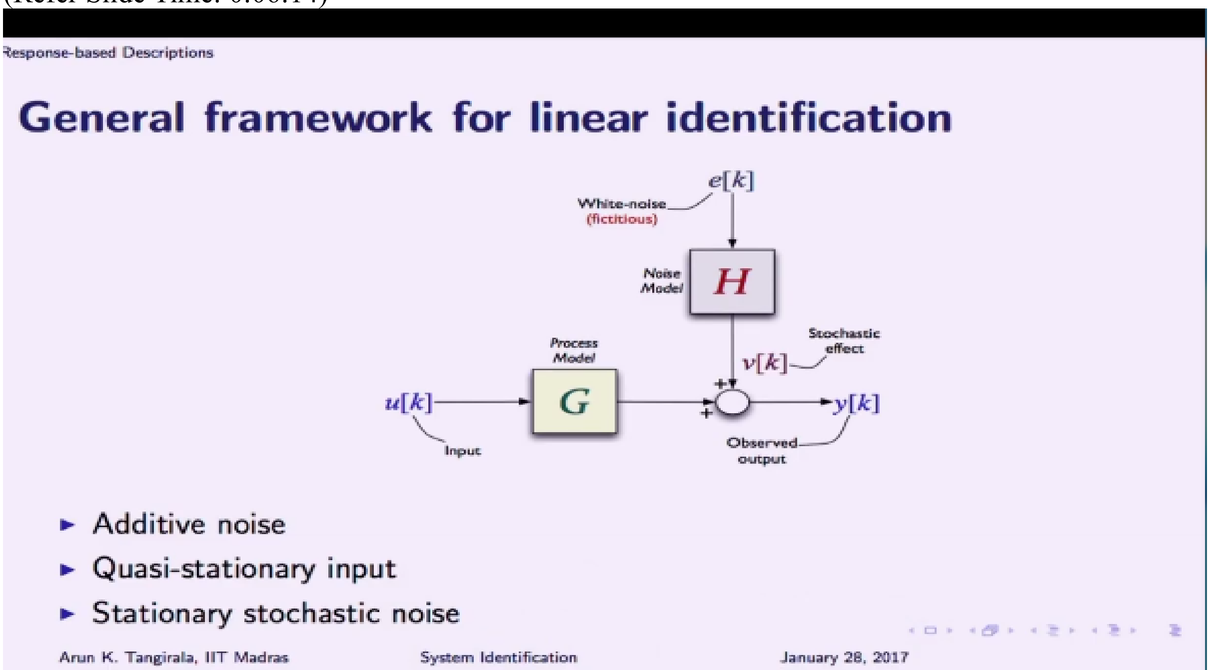
One or more of these assumptions have to be question but fortunately for us in this course, we that need may not arise, most of the times. Okay. Alright, so, now where are we proceeding for the immediate few lectures? Let's understand this. Given input and output I need to build G and H. What is the kind of theoretical knowledge or the knowledge that I should be equipped with? The first thing that I should be equipped with is, what are the mathematical descriptions available for G? Because after all, what am I going to do? I'm going to fit a mathematical model, right? That's the first thing. Likewise, what are the mathematical descriptions for H?

(Refer Slide Time: 0:05:13)



Assuming that v(k) is stationary. Of course, we have already assumed that this v(k) can be represented as white noise passing through this filter. Time series analysis tells you when you can do this, you know, under what conditions this assumption works, this setup works for you and how to estimate H, if you were to be given v and all of that. But the fact is that I still need to know for identification, what mathematical descriptions exist for edge? And of course, I need to know what is the statistical description of this white noise, e? And what kind of inputs are allowed we have already said, Quasi-stationary, okay? Let's say that, we will worry about that a bit later. Then I need to know, given a mathematical description for G and H and given the data, how to estimate that?

(Refer Slide Time: 0:06:14)



Each mathematical description will have some unknowns to be estimated like we have seen either impulse response coefficients or parameters and so on. So, estimation theory has to be learned. And very often, we may build models either in time domain or in some other transformed domain like frequency domain. How do I go from one domain to another domain? So I need some familiarity with Fourier transform because that is the transform that we are going to work with, whenever we talk of frequency domain models. So there's a lot of things to be learnt, but we learn them in a stepwise fashion. What I normally prefer to do is discuss the theory first. Of course embellish it with some practice that is, keep showing from time to time how to practice that theory, without going into the theory of the practice. There is a theory associated with practice also.

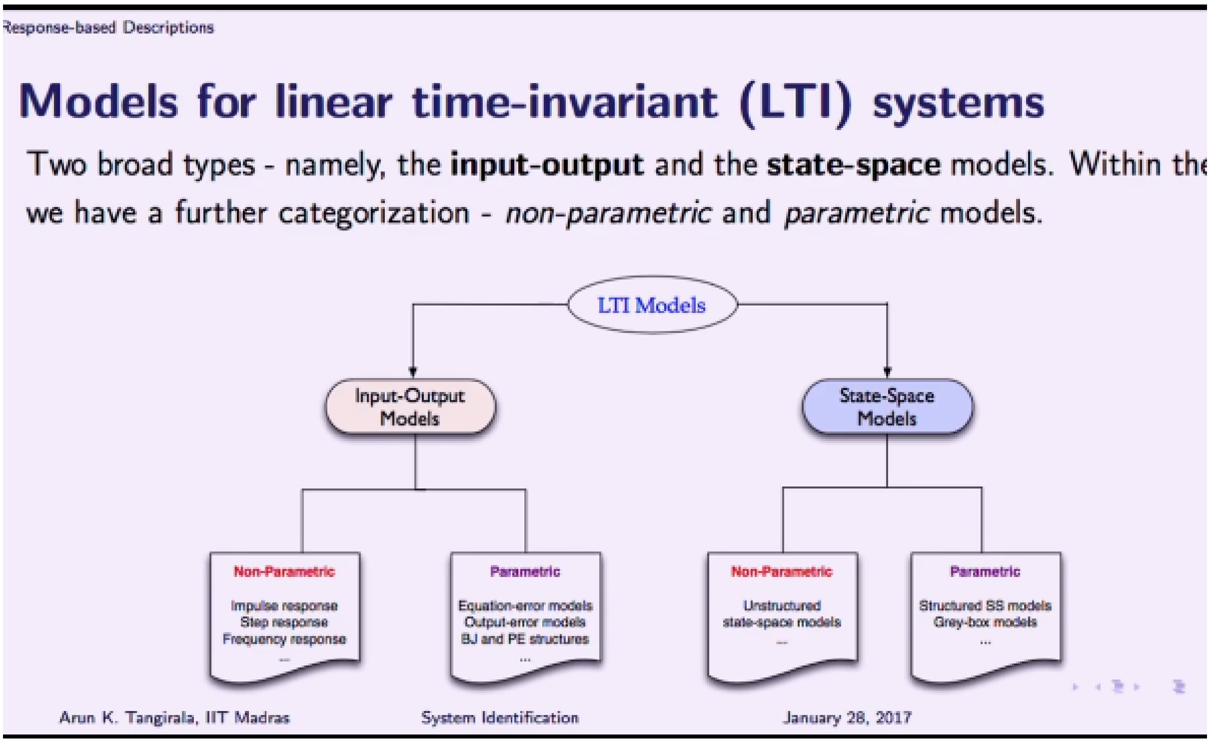
Once the theoretical concepts are through, then we'll get into how to estimate. In other words, we will spend time first on understanding, what mathematical descriptions exist for G and H? Because unless I know what models are available, I can't even proceed further, because that is the starting point in identification. It's pretty much like this, I mean; I want to connect to people, myself and someone else or the rest of the world and the one device that connects me to the rest of the world is a mobile phone, let us say. There was a phone. There are so many models that are available. I need to know what models are available, right.

What are the features; if I pick Samsung, then it may have some features, within Samsung itself there are different models. Within Apple there are other models and then within Xiaomi and so many things, I think the number of cell phone models has outperformed bet all the kinds of soaps, toothpaste, everything put together. So, there are so many models, but the point is, whatever model I pick, I should know its features. Then I will make a decision, right? If any model that I pick there are going to be some merits and demerits. And there are going to be practical aspects of it. So, if I buy a phone as big as-- there are some phones which are as big as a face itself, right, somebody speaking and you don't even you can't see 50% of the face, right. Okay, that's great for in many ways, but not so great in terms of mobility.

I can't really put it in my pocket. I have to put it under my shirt, right? So then, you know, it can actually give quite a misleading impression. So, there are disadvantages to that, and there are advantages to it. Likewise, when I choose a model, there are going to be some criteria. But before even apply that criteria, I need to know how this models work. So, we will start our journey with understanding what models exist for G, which means we'll focus on the deterministic [09:29 inaudible]. And remember, we'll restrict ourselves to linear time invariant systems.

When it comes to the world of linear time invariant systems, there are primarily two classes of representation in the time domain to begin with. Then will slowly talk about, of transformed domain or frequency domain. You can even say that this classification applies to transform domain as well, so, in that sense it is quite broad. And what are these two classifications? Or classes of models? One is your input-output model, where this mathematical description directly relates input to the output, there is no broke up. There is no middle agent, nothing just input and output talk to each other through a mathematical equation, that is an input output model.

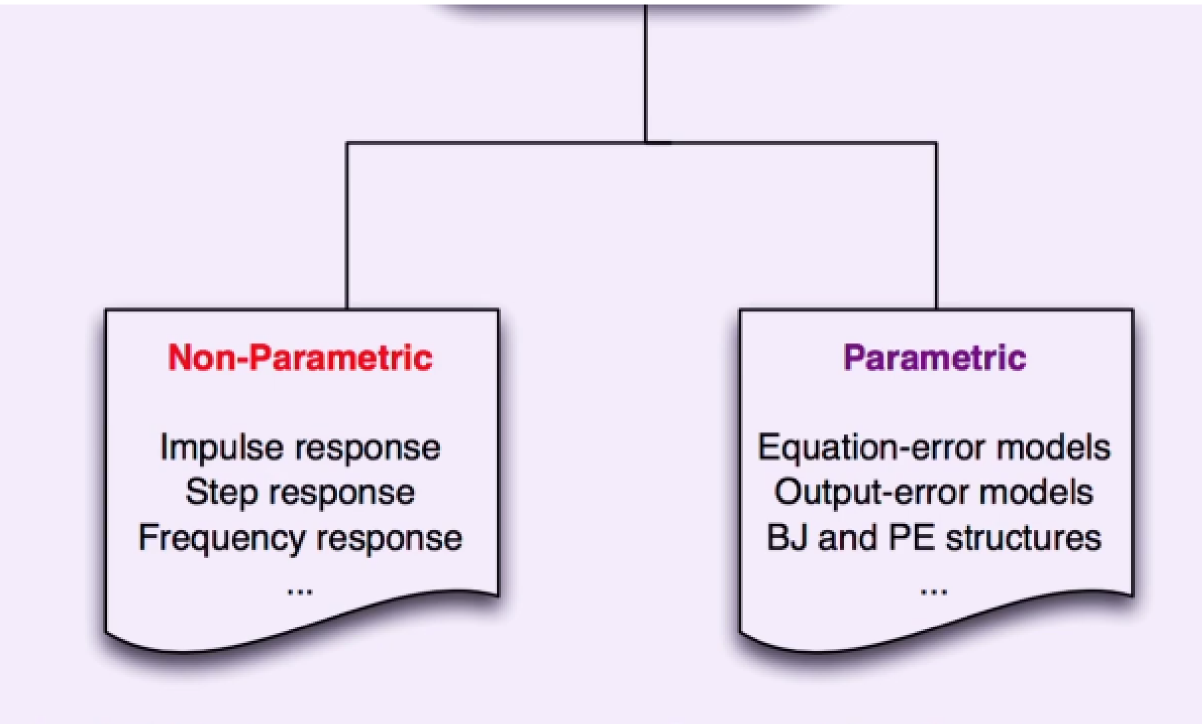
(Refer Slide Time: 0:10:21)



On the other hand, you have State-Space models, where between the input and output there is a quasi that means there is a state. They don't talk to each other directly, unless there is a feed through term, like you've seen in state-space. There is always a messenger that takes the message from the input and then delivers it to the output. And that messenger essentially is called state, so you are going to work in the space of states and therefore the name, and as we have briefly mentioned earlier State-Space models are not necessarily unique. But there are merits and demerits again to each of these classes of models. Input-output models have their own charm and their own limitations, and State-Space models also have their own charm and not limitations I would say, complications. Okay? State-Space models are generally do not suffer from the limitations that input-output models have. But they can unnecessarily, you can unnecessarily complicate the problem by choosing a State-Space model over an input-output description. Know it-- what happens is regardless of whichever class you choose, there are two again, branches.

Fortunately, they have the same names called the non-parametric and parametric. Let me zoom that for you because the font size may be too small. So within the Input-Output models-- is it visible to everyone now, at the back? Okay. So we have spoken briefly about these two classes: Non-Parametric and Parametric. We'll go more into detail, shortly. What is essentially non-parametric mean that it means that I do not assume any so-called "Structure" of the model, how the model looks like.

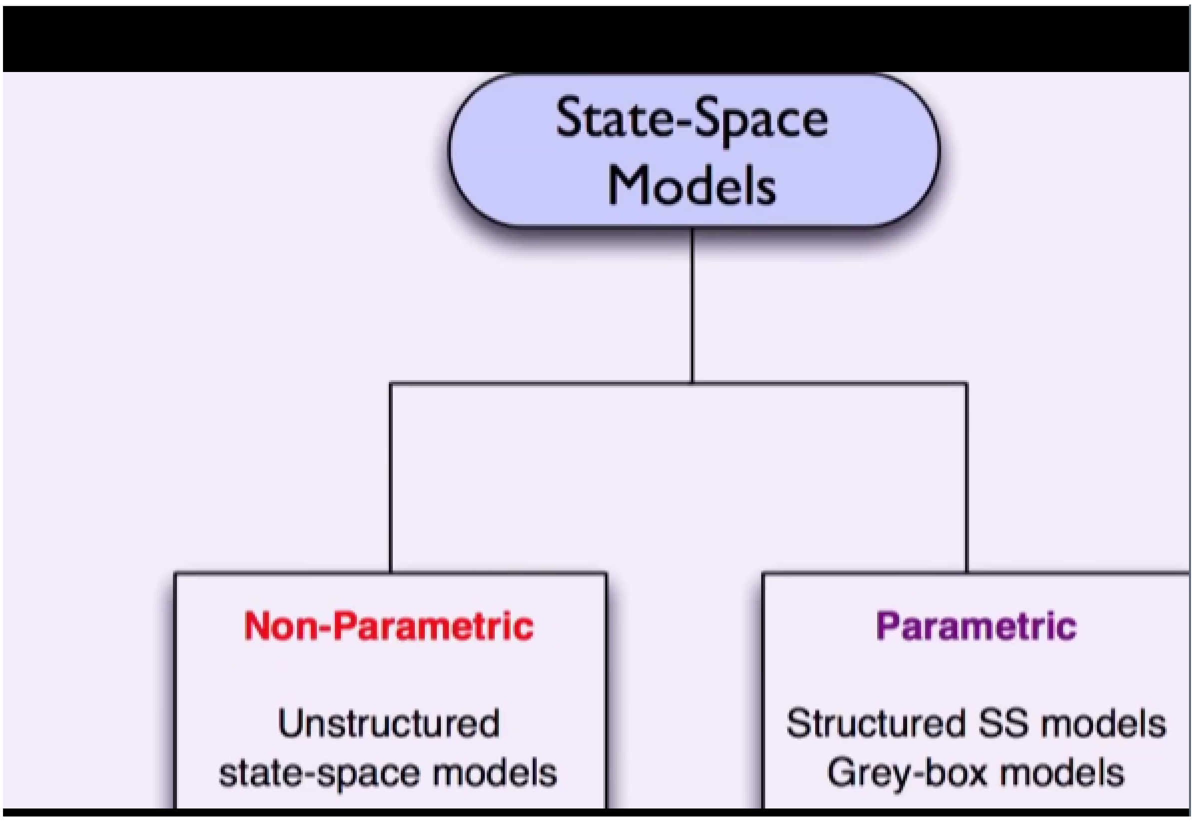
(Refer Slide Time: 0:12:22)



I do not dictate. They just fall out of some basic assumptions like linearity and time invariance. Okay? And under this class of non-parametric input-output models, normally you would encounter what are known as response based descriptions, such as the ones that you've seen in the liquid level case study, impulse response, step response and frequency response. And the counter one is a parametric model, a family, where there is a specific structure to the family-- to the models, there is a delay associated with it, there is an order associated with it and then there is an order associated with the numerator of the input dynamics, so it's quite structured. The user has to specify how the model looks like? At least some basic pieces of information the user has to supply. Although I say here equation error models and so on, you should understand that I've already clubbed the deterministic and stochastic parts there. In the deterministic world, we don't use the term equation error model. We use the term transfer function models, differential equation models and so on. So, please do make a note of that that this terminology that you see here is for the composite system-- deterministic plus stochastic system.

Otherwise, this terminology does not apply to the deterministic work. Within the state-space model again, you have this classification: Non-parametric and Parametric. What is the difference between the non-parametric and parametric for state-space description? Again the same story. Remember from the liquid level case study they were this matrices A, B, C, D and so on.

(Refer Slide Time: 0:14:22)



If you do not assume any particular structure on those matrices, that is, you do not care really what are the states? As long as the state-space model maps the input to the output you are okay. Then you're dealing with what is known as a non-parametric or an unstructured state-space model. Sometimes they are also known as freely parameterized state-space model. From a parameter viewpoint, you say the parameters can enter the matrix in any which fashion and go away. That means they can, they can be seated anywhere. Pretty much like an Indian wedding. You go to Indian weddings or, you know, for ceremonies in India, you can pick your place and sit.

On the other hand, I don't know, if you have attended weddings in the west, everything is pre-arranged for you, right? You go there, you're given the table and sometimes even the chair, you have to sit only there. So it's like structured, you can sit anywhere you want. You will be actually frowned upon. But that is how it is there, right? That's their custom. So then you're looking at a structured arrangement. So when you are looking at that in a state-space framework, where you know that the parameters that are-- parameters mean unknown that you're estimating, they are only situated at certain entries in A, B, C, D, not all entries of A, B, C, D and K will be necessarily estimated.

You know that some of them are zeros or you know their prior values. Only a few have to be estimated and you know their location also. That means you have an idea of the structure of this matrices [16:20 inaudible] and you're dealing with what is known as a parametric state-space model.

So, under this family you encountered what are known as structured state-space models. A sub family of features are called Gray box state-space models. Why Gray box because maybe the knowledge of the structure of this state-space models is being driven from your preliminary idea from your first principles approach. You know that certain entries have to be zero because I know from physics this state cannot depend on another state and so on. So the sequence of assuming a structure on a state-space model means implies that you do not want as a state-space model, you want a particular state-space model. That means a choice of states is fixed. Obviously. Because you're not allowing A, B, C, D, K to be anything you want. You're saying A has to be this, B has to look like this, C, D, they all have to look like this, the states cannot be changed. You're forcing the algorithm to seek a model in a particular space of states or a basis. Now, why should I worry about this distinction? Can you actually answer this question based on your experience with the liquid level case study? Why do I care? Let's see if you're able to answer this. How do I care? I mean, I can, why can't I just say lump them all together, just tell me what all models are available? From an identification viewpoint, how do I care? Why so much silence? In other words, how does it matter whether I worked with a parametric model or a non-parametric model in identification?

STUDENT 1: Cost of estimation parameter.

ARUN: Okay, good. So you're worried about estimation aspects, essentially. And one aspect of estimation that is the effort that goes into estimating. Good, that's one of the reasons why we want to study these two. Which one typically has higher cost of estimation associated with?

STUDENT 2: Non- parametric.

ARUN: Non-parametric. Let us see if anybody agrees. Anyone else agrees with that? So the claim is that the non-parametric model has a higher estimation cost associated-- estimation effort associated with it. Does anyone agree? Why so much hesitation? It's okay to be wrong. It's of course fine to be correct. Why? I mean, it can be a yes or no, it's not like your feedback from strongly disagree, agree and so on. Okay? What do you think? Yes, right. Why, because there are simply so many unknowns, even you can argue it either with the state-space example from a state-space model family or input-output model family. But the fact is typically in any data analysis, or curve fitting, or model estimation exercise, non-parametric models will always have large number of unknowns. Fine? So the estimation effort is larger.

So you may say, why even consider when you know very well that non-parametric models have relatively-- relative to parameter models, they have a larger estimation cost. Then why don't they just discard them? Why do I even discuss them? Sorry.

STUDENT 3: [20:25 inaudible]

ARUN: Okay. Anyone else?

STUDENT 4: And you need a preliminary knowledge from the non-parametric for a curve fitting a parametric model?

ARUN: Okay, so you can put it the other way around, I mean, you can say something that complements saying parameters models require some prior knowledge, which you may not have and therefore it may not be a wise thing to jump straight away to parametric model. It's okay to do it, nobody is going to issue an arrest warrant for it. Okay, it's not a crime, even from an estimation viewpoint. Only thing is that it just means that you're going about things in a blind manner, not in a systematic manner. And it may involve unnecessarily, a lot of effort which you can minimize by first spending time with non-parametric models.

This is a trade-off, at least among the many trade-offs that that non-parametric models have large number of unknowns associated with them, but the plus point is the user does not have to really be fully awake. You can do it even with your eyes closed, something like that. That is you just need to make minimal assumptions and you'll be able to estimate it. Whereas parametric models require quite a bit of information depending on kind of model that you're fitting but definitely some valuable information or some important information has to be pumped in, which you may not be able to find anywhere, nobody is selling it, or you don't know it, you have to derive it from data. Or you have to do something else. So it's good to go through a non-parametric step. But the advantage of a parametric model is that fewer unknowns, mainly, it, fewer unknowns.

Typically, parametric model, if you're chosen it correctly, will have much less unknowns, then what a non-parametric model would have. But nature is not going to be so kind to you; it's not going to allow you to reach that quickly. You can actually sit in a car when you're five-years- old and start driving it. You could do it depending on the country you are in, but typically, so you have to reach a certain stage and an age, likewise here you can't jump straight away to parametric model, you go through the non-parametric model [23:07 inaudible] routine and it's good to do that. So that is one of the reasons why we do distinguish between these families. And the other reason is the kind of algorithms that are involved in estimating a parametric model versus a non-parametric model.

Although I mentioned this in a system identification course, it is true across the board, across any signal processing arena, beat spectral density estimation or be it probability density function estimation, even in probability density function estimation, you will come across this a non-parametric and a parametric. What would be the case there in probability density or density fittings? What will be mean by parametric probability density function?

You do assume the, form of the PDF. Right? You may assume it to be going to be Gaussian to be uniform, chi-square. So, you are actually already-- you know the shape of the PDF. You know the mathematical equation that parameterizes that shape. And the only thing that remains to be estimated are the parameters of that PDF. That's a parametric PDF estimation, whereas a non-parametric PDF estimation would be simply estimating the density, which will give you the probability in a certain interval. That's all. That means you have to estimate the probabilities in each interval of outcomes. A lot of unknowns to be estimated whereas, if a parameterize the PDF, all I have to do is if I fit a Gaussian or any other, you know, typically two or three parameters at the max, those are the only two or three unknowns that I have to estimate from data.

So by choosing to work with the parametric model, I am pumping in a lot of information and coming back to the point, the algorithms that are used in estimating, these two families of models are in general and not necessarily the same. And this is particularly true of state-space models. For input-output models there may not be a stark difference between the algorithms but for state-space models, this there is clearly a difference between the class of algorithms that you use for estimating unstructured or non-parametric state-space models. And those algorithms are called subspace identification algorithms versus those that estimate structured state-space models, where you would use again, what are known as prediction error, minimization algorithms and so on. Okay?

So you should keep this in mind. So let's quickly spend maybe 5 or 10 minutes and then we'll continue in the next class. Any questions now? Very important to understand is distinction. Many people have misconceptions that non-parametric models even I had when I began learning this subject, I thought non-parametric would mean, I wouldn't see a single parameter are known to be estimated in the model. Which is absolutely wrong, and it took me a long time to figure out because probably I was just learning it by myself. The distinction between non-parametric and parametric models. A good way to remember is structure and no structure.