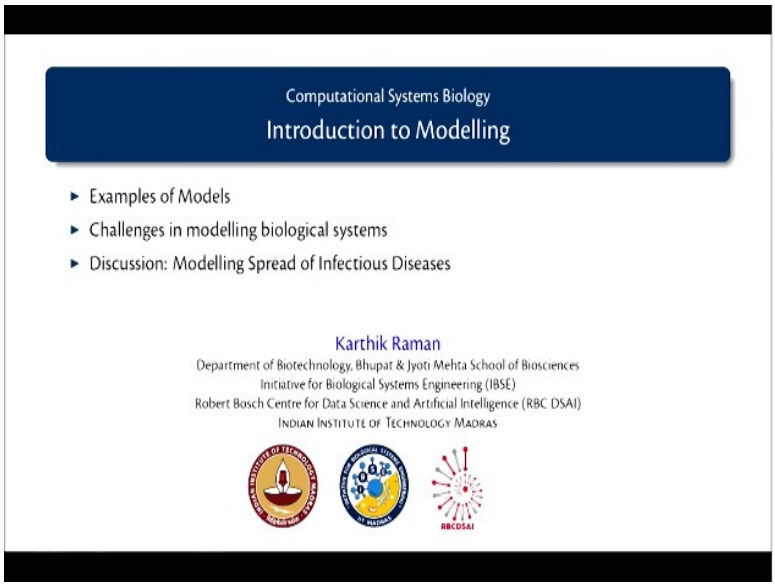
Computational Systems Biology
Karthik Raman
Department of Biotechnology
Indian Institute of Technology – Madras

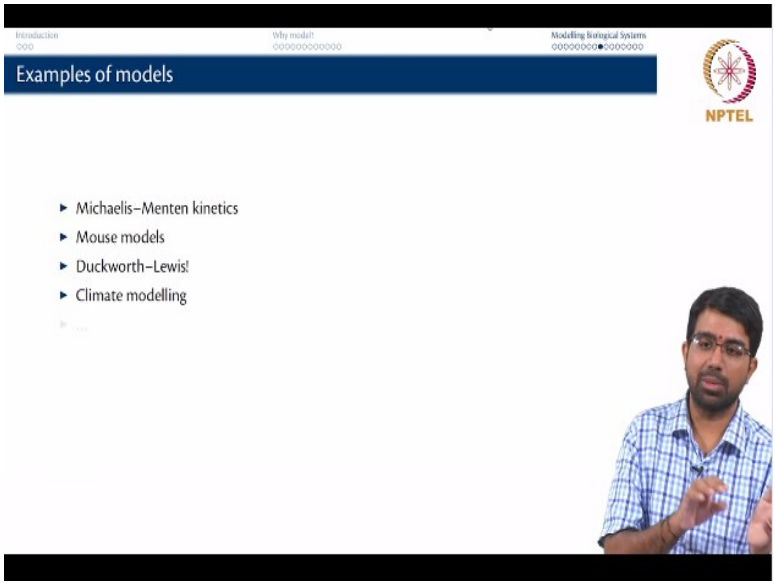
Lecture - 03
Introduction to Modelling

(Refer Slide Time: 00:11)



In this video, we will continue with the introduction to modelling and we will look at examples of some models, the challenges in modelling biological systems in terms of their heterogeneity, how they evolve and so on as well as a brief discussion on a very motivating example which is the modelling the spread of infectious diseases.

(Refer Slide Time: 00:29)



So, going back to the examples of models, there is another model that is very popular in cricket. What is the model? Duckworth-Lewis, right? There's also a 'S' there, I forget what it is. It is now called DLS system, so the DLS system actually is a model, right? What kind of a model is it? It is used to again predict something. It assumes a particular understanding of the game and it has certain assumptions.

Can you think of what are the assumptions? Undeclared assumptions? For example, it doesn't take into account who the players are. So, it only takes into account a match situation. There are like 6 wickets down or 8 wickets down or 3 wickets down or there are 10 overs left or 15 overs left or 20 overs left and the first thing that you might want to think about is, is there one DLS model for ODIs versus T20s, right?

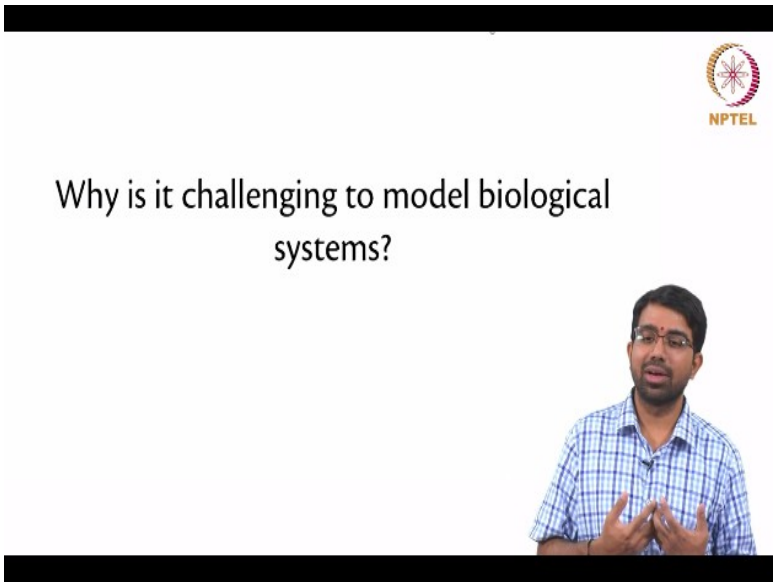
So, that is a classic place where you will see that assumptions break down. You cannot take your ODI assumptions and fit it on to T20 and have a model that predicts well in a T20 scenario right but you can potentially build different models for different situations and this is exactly what one needs to do in biology as well. If you look at Michaelis-Menten, Michaelis-Menten will only tell you what is the initial rate of a reaction.

If you want a rate after a steady state or something, you will have to resort to a different set of modelling tools. So, any other assumptions in Duckworth-Lewis that you can think of? Yeah, it doesn't worry about the teams, it doesn't consider the ground conditions, right? And it is a model that has been built on average over several matches that have been played in the past.

It knows that if somebody has 3 wickets left with 7 overs to go, this is how many runs they can get with some sort of an average or a good statistical average of what has been in the past and that is essentially what's being applied right. It doesn't understand that if it is a very swinging morning the conditions are completely different whereas if it is a flat pitch. Those kind of things.

So, you have to always be aware of what assumptions went into your model because you always keep complaining about the model, right? When you want to complain about the model you need to understand that there are several assumptions that were made in the model, many of these assumptions may not hold and may be they have to be changed if you want to apply the model to a different scenario.

(Refer Slide Time: 03:17)



What are the challenges in modelling biological systems? We talked about how it is important to model biological systems but what are the challenges in modelling biological systems? Why are the biological systems unique in some way? Complex nonlinear but you have many complex nonlinear systems in engineering as well. What is the big deal about a complex nonlinear system?

I am sure every chemical plant is a complex nonlinear system I would say. Yeah, so no big deal. Yes, that is a major issue, right? Biological systems are non-homogeneous. There is so much of heterogeneity when you look at biological systems. So, this heterogeneity in biological systems, non-homogeneity becomes a big problem. A classic example is if you took a bunch of cells-identical, clonal, xerox copy yeast cells and you grew them together.

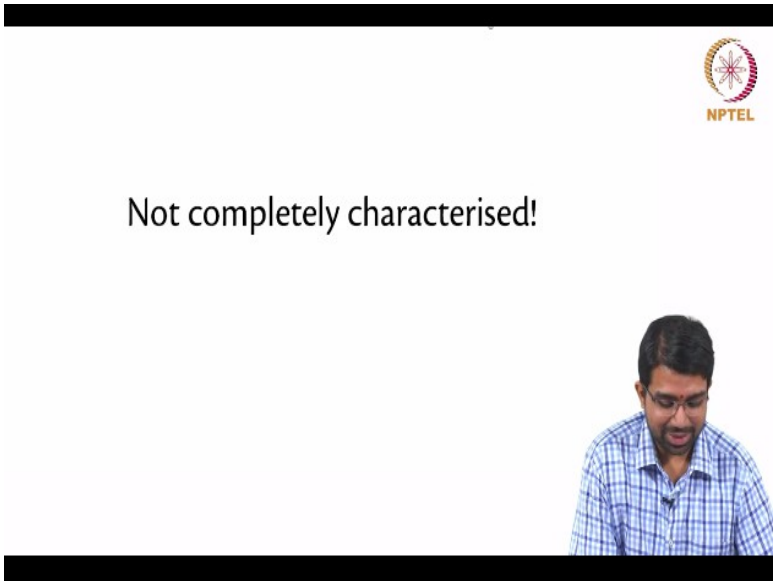
You will find I can plot a histogram of growth rates, you will see that not all cells have the same growth profile Not all cells will have the same expression profile. You will have an expression profile, you won't have an expression point. They will express different proteins differently. There is so much inherent noise and inherent differences between identical cells. So this becomes a major challenge, right?

Because no two cells are identical. So, we obviously have ways to get around this We look at population averages for growth, we look at population averages for expression of genes and so on. But typically if you looked at a single cell or a bunch of single cells, there are so many

single cell experiments you can do today, you can sequence single cells, you can look at single cell transcriptomics and so on.

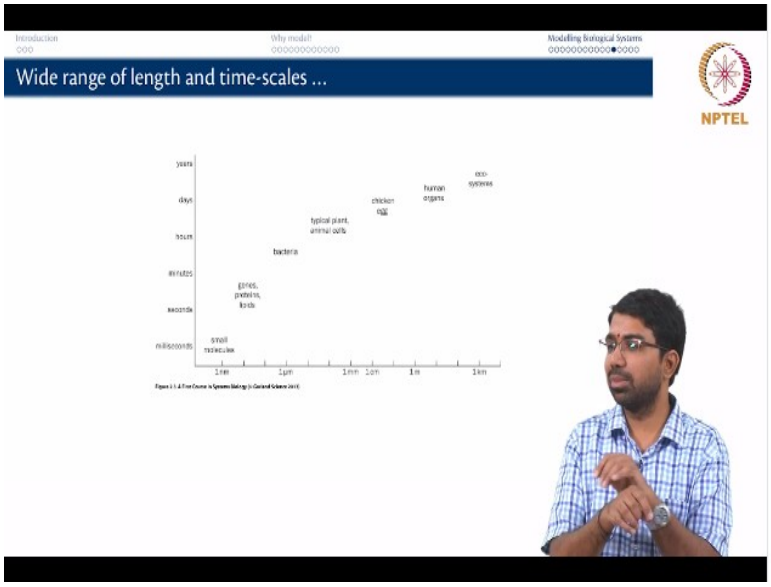
You will see that there is always a profile, right? So, this becomes a very interesting challenge. And the other major problem that you have in biology is it's an unknown box, it is a black box or a grey box. You know part of it, you don't know the whole system.

(Refer Slide Time: 05:27)



The system is not completely characterized.

(Refer Slide Time: 05:32)



And they have a wide range of length and time scales. So, you have on the length axis you have small molecules which are like in the nanometer time scale but then you also have organs or you know like neurons in your body can be a meter long and so on. So, you have

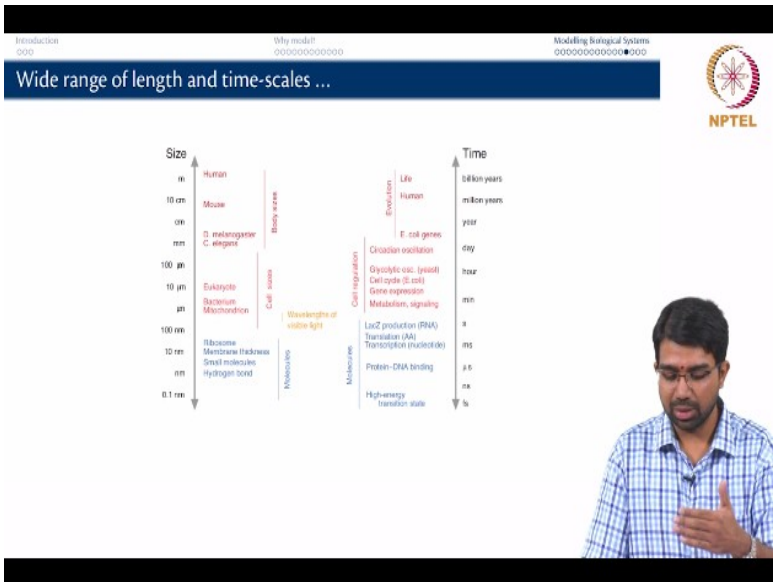
different time scales within the same organism, I mean with different length scales. You also have different time scales.

A very wide range of time scales that are in operation. Metabolism will happen over a few seconds or few minutes. Signaling can happen over a few microseconds and so on. And there are growth development happens over years, right? There are so many things for example, gestation, it happens over several months or there are so many such processes, they may be triggered at a particular point but there will be a very slow time scale over which the process evolves.

Evolution happens across millions, billions of years. So such large time scales variation in time scales when you look at biology and the challenge is sometimes you may need to model, obviously nobody is going to model metabolism and evolution on the same time scale, right? So, that is never going to be a problem but you may still have to marry a couple of time scales.

You may have to look at a millisecond time scale and a nanosecond time scale or a millisecond time scale and hours time scale. So, these two systems may be interacting with one another and that is going to give you some challenges when you start modelling them.

(Refer Slide Time: 07:09)



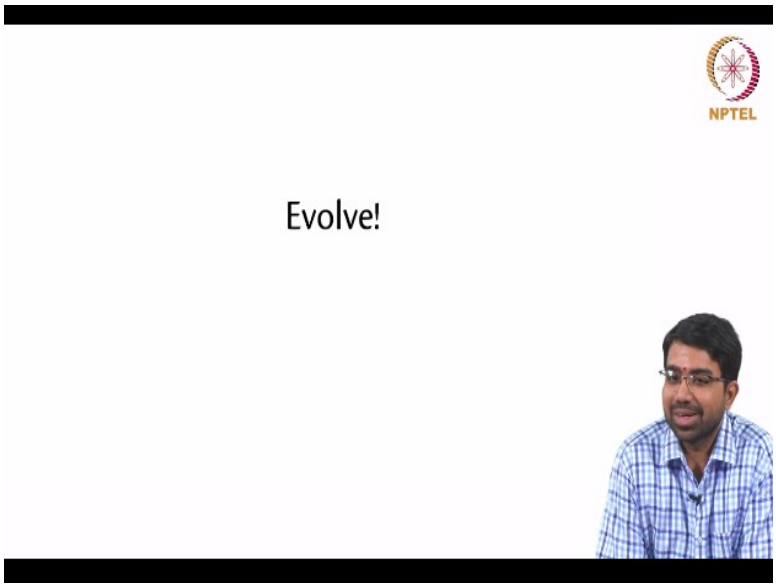
And of course, there are you know this again lists some of the various time scales So, you have transitions happening in femtosecond time scales, then you have transcription happening

across milliseconds, minutes for metabolism, gene expression, circuit and oscillations have a day's time period and so on. And evolution is even longer.

(Refer Slide Time: 07:31)

And we already discussed about the non-homogeneity of biological systems.

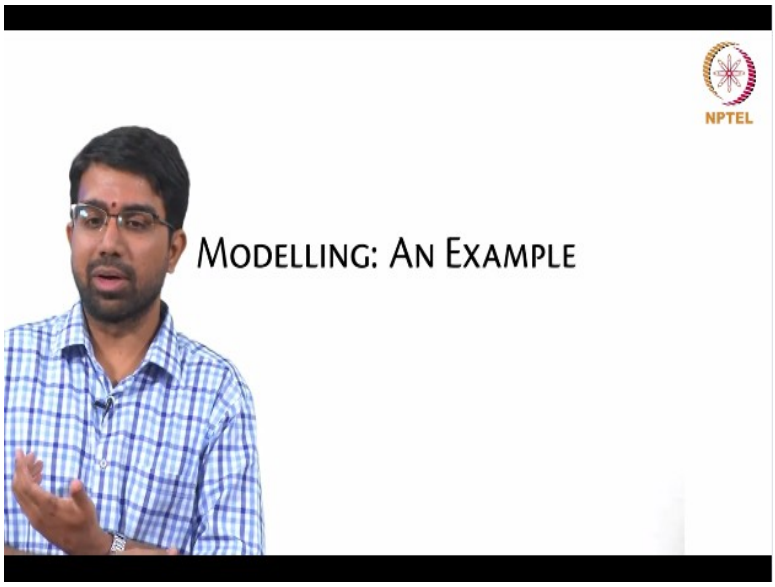
(Refer Slide Time: 07:35)



They evolve, right? So, you took a cell today, you studied it, you characterized it completely, you come back tomorrow you have a slightly different cell. So, how do you handle it, right? So, you make a genetic modification, it is something we will do in this course. We will try and predict what is the genetic modification that you need to do to make *E. coli* produce a little more of a particular compound.

But then how stable is this modification going to be? That is not something that you can easily predict using a mathematical model. So, there are ways to see you know make sure that you know some there are good experimental techniques that have been designed to make sure that there are better ways to do this but it always is a challenge that your system just evolves. So, it is always a moving target in some sense that you need to hit.

(Refer Slide Time: 08:24)



So let's start discussing about an interesting model and a classic model as one of you mentioned was how do diseases spread in a population. So, for understanding this you will have to first tell me what is that you want to predict? Based on that we have to make a decision on whether you need to use a static model or a dynamic model. I presume it's going to be a dynamic model because we want to see how something evolves, right?

You already have committed to a time axis. You want to see how the prevalence of a disease evolves in a population. And you need to also tell me whether we need to look at a communicable disease or a non-communicable disease. And maybe you are interested in a communicable disease because you want to see otherwise you have to model other things like if you want to say how diabetes changes in a population or cancer changes in a population.

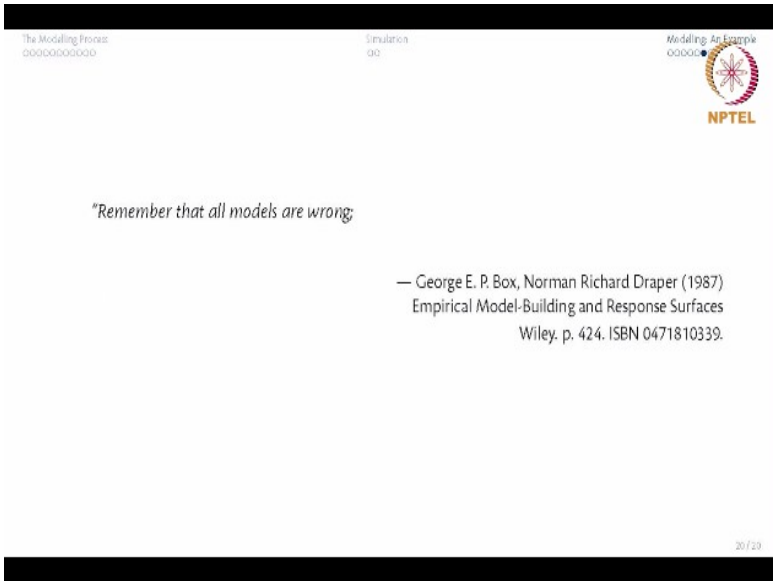
You may have to look at completely different things. But let's say we want to say how viral infection spreads through a population. So we have to start establishing certain equations. Before we even get to equations what is it that you want to predict? You may want to predict number of people who get affected. What is the population of people that is infected at time t?

And fair enough. What else? **"Student conversation"** May be you want to say what is the rate at which infected people die. That may become a parameter or something like that. We will get to those things. What else? **"Professor - student conversation"** At what stage of the disease do they die? So one other cardinal rule of modelling is we will try to start with the simple model and add all the bells and whistles as we go on.

So, you may want to start with a very simple model in the beginning and then we will start adding many features to the model. We will throw away every atrocious assumption that you started off with and incrementally build towards a very perfect, very versatile model. So to start with we will say there are people who are susceptible to infection, there are people who are infected and there is another population which has recovered from the infection.

Because we like to make this distinction because the recovered population may or may not be susceptible to infection. Some people may have become immune to infection through the process of infection itself.

(Refer Slide Time: 11:06)



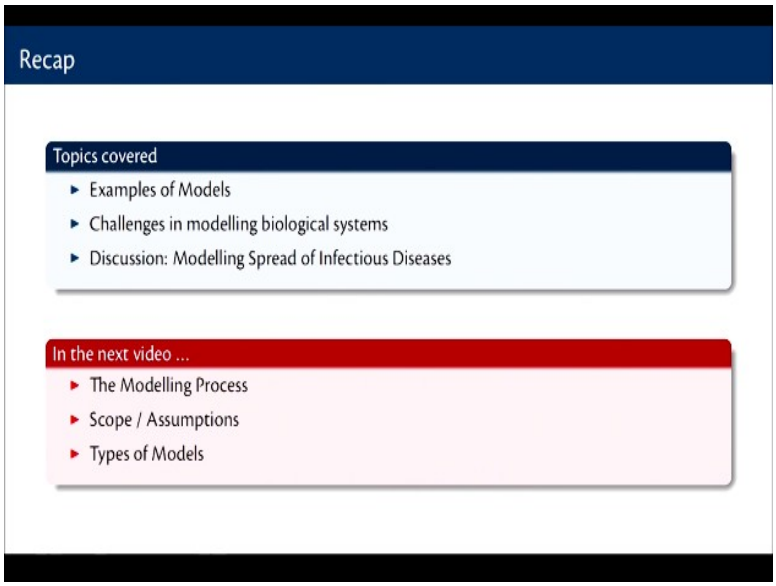
First rule for a modeler, remember that all models are wrong but although the sentence is a little complex so the question is they are all useful right, how useful is a model. And my favorite example is one a colleague gave is that you can assume earth and moon as point objects and have Mangalyaan that actually reaches Mars on time, right? So I am sure they did a little more modelling than that.

But I don't think the escape velocity value is very different from the 11.2 kilometers per second you assume, you compute by using point masses. So, it's a criminal assumption to assume that earth is a point mass and so is the sun and so on but it's good enough for calculating escape velocity. This basically tells you that yeah it's a model, it's wrong but it's very useful.

So, this is the same philosophy you need to apply and you have to bring this into your psyche very quickly. Because I have seen even Ph. D students who feel like not really convinced Is this good? We are leaving out so many things. Yes, we already agreed in the morning that we can't model the universe. So, we are trying to model some part of the universe which is very relevant to us. We have chosen to leave out certain aspects, we have chosen to retain certain aspects.

And we are going to make our best prediction given these restrictions. We have certain computational resources, certain data resources based on which you model and so on. So, what is the best that you can do given these?

(Refer Slide Time: 12:42)



In this video, we have covered some examples of models and what are the challenges involved in modelling biological systems. And finally we had a brief discussion about modelling the spread of infectious diseases which is a nice canonical model to look at to ask several questions may be in the next video and so on.

In the next video, we will start focusing on the modelling process. What is the scope of a model? What are main assumptions that go into modeling? How important assumptions are And the different types of models that exist.