Agent Based Modelling for Social Research

Introduction to modelling Part 1

Hello, my name is Martin Hinsch. In this talk, I'm going to give a brief general introduction to modelling.

There are two types of models that I want to talk about. First predictive models and then explanatory models. So let's start with a few examples. In the top left, you see the famous Miller–Urey experiment from 1952 where Stanley Miller attempted to test how life might have originated on Earth. Below that, we see a scaled down model of a ship in flowing water. On the top right, we see a population dynamics model of two interactive populations, a predator and a prey population. And finally, we have the output of a complicated weather model.

Let's focus on the lower two examples first, and we'll take a small detour via something that is not a model. So look at these two examples. Let's assume we want to throw a ball and hit a net. Usually, what we'll do is we would not start calculating gravitation and friction, and laws of motion and things like that. But we would use our intuition and our experience and would be usually quite fine with it. On the other hand, if we want to send somebody on to the moon, as you can see, on the right hand side, we would be quite mad to use intuition or experience alone. Instead, we would take recourse to a huge set of tools in the form of physical laws and calculations, approximate methods, computers and so on. Since otherwise, we would not be able to do that without probably the entire thing ending in a disaster.

So what you see is that some problems for some problems, our brains are not equipped to handle them, we're just not able to, to think ahead precisely enough. So we have to use some form of formula/tools/calculations. What does that have to do with modelling? Let's look at another example. This time proper model.

This is the cases of COVID-19 in parts of China and the rest of the world at the beginning of this year. As we all know, governments have since then struggled to contain the pandemic and to find ways to minimise fatalities, while at the same time keeping the economy on track. And the way they have doing that is usually by following some kind of modelling. So this, for example, is a very, very simple, classic epidemiology model. It's called an SIR model, with a susceptible and infected and recovered root population.

Models like that, in a much more sophisticated form, have actually been used in the real world by governments to find out which scenarios would be most advantageous. So what they do is they use that model, they plug in different initial conditions, then test how the pandemic evolves and see which scenario works best and the given boundary conditions. And from that, they can then deduce which measures to take more or less successfully.

So on the surface, this looks very much like the previous problem. So we have a process that is very, very difficult or very complicated, has very high stakes, so it's too difficult for the unaided brain. So we have to use formulas, a formal description of the process and some kind of mathematical computational method to find the best course of action. So same motivation. But the difference is that in one case case of the moon landing, most of the processes involved can be described with really high precision and most most cases, we actually know how precisely we can calculate the process. In the case of this pandemic, this is not the case we don't fully understand the process. There are many things that happened that we don't know about. Or if we know about them, we can't properly formalise them. So for example, we don't know exactly how the wearing of masks reduces infectivity. We don't know how many people are going to wear masks if we tell them to and so on. So this is a an approximation, a very rough one. And that's essentially the the basis or the the essence of modelling modelling means that we want to approximate a process but can't do it precisely.

So predictive modelling means we want to predict how a process behaves under given conditions, our intuition is not sufficient to let us predict the process well enough. We do not know enough to describe the process precisely. So we have to approximate. And that approximation is then used as a prediction of the behaviour of the real process. So, this covers are two examples, perhaps, oh, wait a moment, we have to do something about the ship. So that's actually quite easy, we just remove the formal part and then we're fine. So any kind of approximation actually does, it doesn't necessarily have to be a mathematical approximation, it might also be a physical approximation. And then we have a predictive model of our system.

Now, for the upper to upper two examples, what is that? Let's go through them in terms. First, the Miller–Urey experiment, performed by Stanley Miller in 1952, was an attempt to explain how life originated on early Earth. And what he did was essentially he took two glass flasks, one of them he filled with water, the other one with a few basic chemicals and assumed to reflect the conditions on early Earth and then connected them, heated the water and regularly sent sparks to the other compartment. And that setup he had run for a couple of days a week and after that, you see this brown sludge developing. And what he found was that this sludge actually contains five of the essential amino acids that make up every single living being on Earth. Recent reanalysis of the original setup by sea, by the way found that all 20 amino acids occur. So that was the basic setup.

The other example is a very famous model in population biology and ecology. The last Lotka-Volterra Equations. Vito Volterra was inspired by his future son in law, who observed that during World War One, the proportion of predator and prey species and individuals in fish catches fluctuated. So, what he tried to do is to understand why that was the case. So, he did that by writing down the population dynamics of two populations when predator or prey species and then adding interaction terms to them. So the effect of the predator on the prey species and vice versa. And extra extrapolating from that, what he found was that he actually found fluctuations under certain parameter conditions.

So, if we compare the two examples, what you can seeis that the basic structure is pretty similar. So, we have a very general question, we implement some kind of system, some kind of model. And from the behaviour of the model, we make inferences concerning the original question. It is important to note that in neither of those two cases, we are after any kind of prediction, so, we are not interested in predicting the number of predator individuals next month or the number of prey species left over after a certain time, or trying to predict whether life is going to emerge on a given plant. Instead, we are really after understanding the system.

So, if we look at the questions more thoroughly, we can see that our models do not actually answer those questions, so we can't answer why predator and prey species fluctuate given the model. We can't answer how life originated given given Miller's experiment. The questions we can answer are much more specific. So what you can answer is whether the fluctuations in these populations might be a consequence of the interactions. The model clearly shows that they might and are the conditions of early Earth sufficient to spontaneously generate ingredients for life. The experiment clearly shows that yes, they might be.

So, let's look at the, at the structure of what we've been doing in a bit more detail. What we have in general is we have some kind of reality that we find difficult to understand. We observe some aspects of the behaviour of that reality. We try to copy the process that generated that behaviour by implementing a model and hopefully, the behaviour that that model generates is similar or identical to the behaviour that we have observed. And the first class of models that we looked at so the predictive models, what we're interested in is actually trying to replicate that behaviour as accurately as possible. So we want to know exactly if we change the conditions of it, how is the behaviour going to change, or if we just look at the future, how is the behaviour going to be in the future, which means that in a predictive model, we're actually not really interested in the upper part of the process, we can, in principle, ignore that entirely. And there are actually cases where people have have done that. So instead of implementing a model of process, they have used, for example, artificial neural nets to directly predict the future behaviour or alternative behaviours from the observed behaviour.

Now, if we look at the other type of model, the explanatory model, the situation is slightly different. What we're interested in here is, of course, the behaviour, but only insofar as it tells us something about the process that's happening. So we want to know how this part up here, the complicated reality generates the behaviour that you observe. And, to do that, what we have to do since we can't observe that part directly, we can't, we don't really have a handle on it in most cases. That's the reason why we start modelling in the first place, what we can, what we have to do is we have to, to pick parts of that process that we think is relevant for our observed behaviour, since we can't really look at it directly implemented in a model and then see whether the behaviour we get out of it is similar to the behaviour of observed. And the thing we can prove with that structure is that if we find one particular implementation of that process, that generates behaviour that is similar to the one observed that this is a sufficient logical structure to generate that behaviour.

So in the end, what this type of model really does, it serves as a proof of concept that demonstrates that one particular mechanism is sufficient to cause a specific type of behaviour. To some this up extplanatory modelling really means that we observe a phenomenon, we have an idea of the conditions that may cause this phenomenon. So the the conditions that have to be prevalent, and the causality that causes the behaviour that we observe, but it is insufficient as a proof. So we can't just conclude from that idea that it has to be true. Therefore, we need to replicate the conditions in some kind of approximate representation, either formal or physical. And then run it to see if the behaviour that you observe that you want to prove actually occurs in that presentation.

So some parting words. First model is always just a tool. It's really just as good as the context we use it. For a given part of reality, how we model it depends entirely on what we want to know. So for example, if we want to model a river, in order to test whether a certain area is prone to flooding, we can do that we can build a model for that. But for example, if we want to know how long that river is going to retain chemicals that were spilled into it, we can model that as well. But the model would look completely different than the one that we used for the flooding. Or if we want to know how much fishing we can allow in that river. So if you want to know the carrying capacity of fish, we can move that as well. But again, that one will look completely different than the first, first two.

A model without question is essentially just a toy. So we can do that. It's probably fun to play with. But it does not really fulfil the purpose . Ading things to the model,making it more precise or comprehensive does not necessarily make it better. It really always depends on the question you're asking.

And finally, the only thing a model can really show is the consequences of a set of assumptions. That is all it can do. So it really depends on the right question how useful the model is. And that's it for this part. Thank you for listening.