

Agent Based Modelling for Social Research

Documentation of a agent-based models

Now this talk, I will talk about something that's not in the forefront of most people's minds when they think about agent-based modelling, but it's very important. And it's about documentation of agent-based models and I want in particular talk about the ODD+D. standard for documenting and reporting agent-based models.

Now, while I will be talking about the ODD+D as an example, I will not so much focus on my largest part of my talk will be about why it's important to somehow properly and accurately document your model more than about the how in the context of the ODD+D. Because once you have these kind of standards, it's somewhat straightforward to apply them, I think, especially with the ODD+D but coming up with these kind of standards and realising that they are important things is the most important step to make.

So let me start with highlighting why such standards are useful and necessary in an agent based modelling. And I called this section of problems and implementing simulation models. And just okay, so this is now if you're interested in more, learning more about this, I highly recommend this paper it's already bit older from 2006 by Richiardi, Leombruni, Saam, and Sonnessa published in the Journal of artificial societies and social simulation, where they propose their own protocol for agent based social simulation. So I'm personally not not I've never used this protocol as I find other protocols a bit more and more easy to apply and more useful, but their discussion of why such protocols are useful, the thing is very enlightening and very, very interesting.

So, they highlight that that agent base models have solid methodological foundations, but they yield lots of freedom and degrees of freedom in terms of for example, how you design your model, which elements it has and how you analyse your model. And these degrees of freedom and this lack of coherence standard has led to what they call some kind of anarchy in the field when it comes to design and analysis and presentation of models. So, for example, there is no standard classification for the different ways in which a student can exchange information and communicate, there is no standard way to treat the artificial data that the model generates meaning that everybody every researcher kind of analyses their model at hoc, at least that used to be the case, without following a standard standard way of going about that, then, often there are no clear or no straightforward standards for or at least they are not applied in terms of how the experimental data so the outcome that the model generates should be compared with the empirical data that you have for validating your model outcomes, then the parameter values are sometimes chosen without proper discussion. So it's not clear why particular value has been chosen for a certain parameter and not other values. And, most importantly, often it is not important, it's not possible to understand the details of the implementation of an agent-based model just from the verbal description. And I agree with this statement that this makes replication difficult and sometimes impossible. And thereby, it violates

basic principles of scientific practice, namely that what we do should be at least in principle, be replicable by other researchers to validate what we have been doing and to call to confirm the veracity of our models and our results.

Then, a bit more in detail. There are some some pitfalls in describing papers name describing agent-based models. And the main difficulty here is that agent-based modelling is comparatively new, or at least used to be comparatively new. By now it's a bit more established. But mainstream journals often have difficulties accommodating agent-based models or descriptions, because they deviate from what has been done before, right? If you have a regression model that you apply to a population level data set, like the European Social Survey, there are certain standard templates that you can follow in describing your model, your research and your analysis that you did and your the results that you obtain that nobody really questions, right, because it's so common.

But agent-based models deviate from that or the description of agent based models deviates from that because for example, the models that we have typically read, deviate radically from the existing literature. So we're doing often, so, often you're in a situation where the first one to apply an agent base model to a specific research question, therefore, you cannot really draw on standard templates for describing what you're doing.

And then same goes for the structure of the model. So there are many elements that are essential for a model, but for which they are typically no guidelines for how to report this, whether i were to report this in a in a journal article, right. So for example, as we have seen also in this course treatment of time can be essential for how the model behaves. So but in a journal, you will rarely see a section that asks you for indicating whether you use discrete or continuous time, then often, it's not required to discuss or specify whether your model is exclusively stochastic or deterministic, how you treat heterogeneity in your model, which are all points that are important for for modelling for agent based modelling which which are not commonly reported or not, at least commonly featured in in standard templates for journal articles. And often this leads to a situation where you can't have.

So I personally, that's now my ultimate personal experience that I also colleagues that I talked to shared is when you try to submit an agent based model to a mainstream journal. You have to, attend to the needs of audiences at the same time, right on one and, you might want to make it understandable and some for creative designers in a way your description in a way that makes it understandable for standard readers of the journal who might not have a background in simulation modelling. But at the same time, you want to satisfy people who know a bit more about agent-based modelling to provide an accurate description of your model, and to enable them to replicate your model. And these are two goals that are some times really not compatible with each other, right? So often to make it through the review process, especially when you have non expert reviewers when it comes to simulation modelling, you need to boil it down to to have a simple story to use as few as quick equations as possible. And those equations that you have to explain them intuitively so that people understand what you're doing if they're not used to this kind of methods, which then goes at the cost of accuracy, right and detail. And so I skipped this now, because I said that already.

So there are some solutions to this problem. The first solution is I think this should be standard. It's not yet but many people are arguing for that that should be standard that you publish your code and input data alongside your paper. Because the code of the model is the only way in which the model can be described 100% accurately. Any verbal description of what your model does is only an approximation and therefore might be erroneous and ambiguous. So you might put a lot of effort into describing your model, but still, somebody might misunderstand what you're doing and with code you can circumvent this misunderstanding, directly showing what you have actually been doing.

And there are certain ways for making your code available. For example, a popular platform in the area of agent-based modelling is called the compass net or open ABM, where you'll find a huge collection of agent based models. Using different languages for implementing models, the language doesn't really matter. But it's the focus is your more on agent-based models within this library.

Then coming more from computer science or software development, there is GitHub. And I think it's also increasingly used for sharing code for analysis in research is GitHub, which is a bit less specialist more general, which I see increasingly being used also at our institute. And also people who do not necessarily use it for simulation but for example, writing a piece of code that accesses the Facebook API or Twitter API to extract data.

And then there are general databases like to have a database where you can upload the data for any kind of study that you have conducted. Finding agent based models is a bit less common and then the least recommendable thing to do is publishing code on the publishers websites because there might be access restrictions that people for example, can only download the book if they add the code to to a paper if they bought access to the book or the journal on respective journals webpage.

Now, the second step, that's important is, next to publishing your code up to prominence with publishing your code is that not everybody might be able to read your, your programming language, right? So I'm not fluent in Python, I get the basic understanding but checking model based on Python would be difficult for me and I might not understand certain things. So there are standards that can be used to give a verbal description of an agent based model that's more detailed than you would have more standardised than you would provide.

If you would just write an article for mainstream journal and starting from ecology, the ODD standard has been evolved developed, which has later been developed into the ODD+D standard for agent based models that involve human individuals who make decisions. And it's the the ODD comes from the three main elements, overview, design concepts and details which has been enriched in latest iteration with decisions that directly address individual decision-making behaviour. And this ODD standard has certain standard elements that are sorted into these three main areas overview design concepts and details. Within these sections, you have smaller sub sections, for example, your

way to describe the purpose of the model to state variables and scales, and so on and so on. And within each of these sections, you have certain sub sections.

And the idea is so that you start from such a list like you have here, which gives you all the elements of the of the standard starting from overview. So in if you write the description based on the standard, your work, your your first section would be the overview section where the first sub section would be purpose of your model. And then you want to try to answer the sub question. So what's the purpose of the study? What was the goal? was it scientific experiment? was the illustration? was meant for educational purposes?, and so on, and so on, so on, and you define for whom the model was designed.

And now without going into all the details, and all the different questions that you have here of which there are quite a number of those, I think that having such a list makes it ensures that you don't forget important elements of a model that you might ordinarily not think of when describing a model in laymen terms. But which would be important for implementing the model. So for example, for example, is there a collective learning in the model? Can your agents actually learn something? or instead, So heterogeneity it, that which is a central feature of agent based models and the last point here, heterogeneity is an exquisite element of this description.

And the idea is that you always list your answers to each of these elements or you say something to each of these elements, so that people can say, is your model or your agency to heterogenous, or not, even if they are not, it's important to explicitly state that because then people know and don't have to figure it out themselves from what you have written. And as I said, Before, I don't want to go into the details of all the aspects of the standards, you can download the paper here, from environmental modelling and software, we get a detailed example applied example of how to apply the standard, I just wanted to give you a flavour of what the standard looks like, and how you can, why it's important or beneficial to apply such a standard.

And now one central argument, as the last point that's often used against these kind of standards is, yeah, okay, this, this is my standard and if I write my paper this way, other models will understand, but there is no way that I get the paper published. It's written in this way, right? If I would submit this report to journal I would be rejected because reviewers don't know what to do with it because it doesn't follow standard, standard protocol. I think a useful approach is to while you published the paper with a mainstream description that's makes sense in the context of the journey also, next to this publish the model code together with the standardised subscription, like using the ODD+D protocol, on let's say, open ABM so that people who are interested in it can actually access it. The downside of this is that this creates extra work because you basically have to write the paper twice, but it's not exactly the case because you can largely sample from the main paper in your standard description and expand it a little bit but it's a bit more work but in my experience, it's worthwhile because people want to learn more about your model and want to understand really need to benefit greatly from that.