Spatial Interaction Modelling

Hello and welcome to this NCRM short video introducing you to Spatial Interaction Modelling. My name is Andy Newing from the University of Leeds, and I apply spatial interaction models in my own research and teaching.
In this short video, we're going to introduce you to these models and illustrate some of the ways that they can be used to support research, where we're trying to model or estimate flow data of some form. In the accompanying online resources, you can download a PDF document that gives you a chance to consider some of the concepts that we're working with in this short video. I suggest you download that PDF now and have a look at the example contained within. You'll see that that example introduces you to a hypothetical study area that contains 9 small neighbourhoods, or demand zones, and 3 different retail stores. We could think of these as grocery stores or supermarkets that most of you probably use on a daily or weekly basis. The task gets you to think about where some of the consumers living in those zones are most likely to shop. I'd recommend that you pause this video now and take 10-15 minutes to work through that activity.

Welcome back. Hopefully you've had a chance to work through that accompanying activity. It got you to think about some of the decisions that we need to make when we're trying to model or predict flows. In the example, you were working with consumer flows, thinking about where people are most likely to shop for their groceries. Actually, underpinning that is essentially a spatial interaction model. The decisions that you've made are exactly the same decisions that we ask a spatial interaction model to make. As you've thought through that activity, you were probably thinking about a number of different concepts related to that application. You were thinking about the underlying demand on each one of those origin zones, and in this case it was the money that people had to spend. You were trying to allocate that money to the different stores that people would shop in, and in doing so you probably thought about how attractive each of those stores might be - recognising that some of those were larger than others. You probably thought about geographical proximity, or accessibility, recognising that people might want to shop closer to where they live or perhaps where they work. And crucially, you recognised that this was complex because there was competition, or intervening opportunities. Consumers were having to make a decision about where to shop based on a range of alternatives that may be suitable for them. By thinking about those 4 different things - demand, accessibility, attractiveness and competition - you've run your first spatial interaction model. You've decided how to apportion those flows, or that spending, across the different places that those consumers could visit. So although spatial interaction models may seem complex when we start to look at the equations behind them, fundamentally it's doing no more than the decisions that you've just made. And if you want to check the answers that you've put down to some of those questions, then you can also download the accompanying answer sheet, which is part of the online resources to support this video. So spatial interaction is an important concept in Geography, and it's used to measure or predict flows between any particular set of origins and destinations. Those flows might be people, as in the case of our consumers who we thought about there. But they don't have to be. It could be flows of money, or it could be flows of goods, in an example to do with trade. Crucially, this is capturing some sort of interaction between an origin and a destination. And typically those origins, as in the example you've used, will be zonal. They will be some sort of small neighbourhood, or small zone, that contains our population of interest or that's producing some sort of good that possibly is going to be transported to a whole range of destinations. Our destinations might be another set of zones. In the case of the commuting example, we might be modelling the flow of people from the area in which they live, to the area in which they work. But commonly, our destinations are individual points. In the example that you've just looked at, they were grocery stores. But those points could be any form of service or facility - hospitals or schools, for example. And in our model, we denote these origins with the letter 'i' and the destinations generally with the letter 'j'. And so, when we start to think about this in the form of an equation, we're trying to model the flows between any particular origin, 'i', and any given destination, 'j'. So we're using our spatial interaction model to model those flows between any given origin and any given destination. And in mathematical terms when we write that down, we're modelling a flow, which we denote with the letter 's', from a given origin - 'i' - to a given destination, 'j'. And because we have multiple origins and destinations in our model, we're modelling lots of different Sij's - the flow from i1 to j1, or origin 1 to destination 1. From i1 to j2, that's our first origin to our second destination - and so on, right the way though all of our different origins and destinations. And so if we think about this in the form of the example that you worked with earlier, we've got 27 possible flows that we can model - that's the flow from every one of our 9 origin or demand zones to every one of our 3 destination stores, in this case. So we have 27 separate flows, from an origin 'i', to a destination 'j'. And typically we think of this in the form of a matrix, as you can see on the slide behind me.

We think about these models as being organised generally with our origins down the left-hand side and our destinations along the top. That matrix in the middle is where we estimate our flows. So in our model, if we think back to the example that we looked at earlier, the flow that we're predicting between that origin zone, 'i', and a given destination 'j', is going to be driven by a number of factors. It's going to be driven by the demand that's available in our origin zone - the amount of money, in this case, that people have to spend. And we know that we need to spread that money across the different destinations. It will be driven by the attractiveness of a given destination. It will also be driven by how accessible that destination is. And we know that we've got competition involved here. There are a number of competing destinations, so it will depend on the relative attractiveness and the relative accessibility of that one destination versus all of the other destinations that consumer could have visited. Now in verbal terms, that's quite straightforward to understand, particularly in the context of a retail example, as we've been working with, because they're the decisions that you make on a daily or weekly basis when you shop. So what we've just talked through there makes a lot of sense when we think about it in words, because it's the sort of behaviours and decisions that we're making on a daily basis when we choose where to shop. But people often get daunted when we try and break down a spatial interaction model into its constituent form as an equation. What you can see behind me is the typical equation for a spatial interaction model of the form that we have just been talking about. It looks daunting because of some of the terms that are in here, but it really doesn't need to. On the left-hand side of that equal sign, we've got our S i j. This is the thing that we are trying to predict. It's the flow, in this case of money, from our origin zone 'i' to our destination 'j'. That flow is proportional to everything that's on the right-hand side of this equation. Now firstly, it's driven by the amount of demand available in that origin zone. In this case, we refer to that as 'Oi' - the demand in demand zone 'i'. As we said, it's also driven by some notion of attractiveness at the destination in this case 'Wj', the attractiveness of destination 'j'. It's also driven by the accessibility of that destination and as we can see in the equation here, it's inversely proportional to the distance, the 'dij', between our origin zone 'i' and our destination 'j'. We know that the flows will diminish with distance, because in the case of the shopping example, people want to shop close to home. The hardest one for many people to understand is our 'Ai' term. And in the case of a retail model like this, we would call that our competition factor. And that reflects the fact that consumers are making this decision for a whole range of competing destinations. Now this 'Ai' term is really important in the model. It makes the model balance, and it takes account of the fact that consumers have a number of opportunities to shop, and a number of different destinations in which they can do so. And so hopefully what we've just talked through there, gives you the idea that this simple example that you can see behind you can be broken down into an equation that can be used to build these models using other software.

So although that example is intuitively straightforward, it's worth acknowledging that spatial interaction models are actually really powerful models, that can be used to model flows in a whole range of different contexts. And underpinning this model is a mathematical relationship, which accounts for those observed flows. And we can use that mathematical relationship in an applied context, to estimate missing data in a matrix of this form, or to predict things going forward if we were to make some sort of change, either at the origin or at the destination. And in the retail context, that often involves changing either the amount of demand available, or more commonly, changing something on the supply side - closing stores, or adding in new stores, and thinking about how the consumer behaviours may change as a result of that. Spatial interaction models have had quite a long history, particularly in geography and the social sciences. They were initially based on Newtonian analogies, and were termed 'gravity models' using the terminology that was borrowed from physics. But in a human context, those models really lacked some of the justification in their handling of human behaviour. And it was really in the late 1920's, with Reilly's 'law of retail gravitation', that we started to think about how these models could be applied to human-related problems. And in his case, it was understanding the probability of a consumer shopping at a particular retail centre based on its trade area, or its gravity - its ability to pull in consumers. But it was really work by Sir Alan Wilson, now of the Turing Institute, who operationalised these models in a human context, particularly in the fields of geography and urban planning. Sir Alan recognised that these models did lack some of the justifications needed for handling human behaviour. And he recognised that using entropy maximisation, a technique from statistical mechanics, that these models could have a better theoretical justification in their human applications. And it was really his work that introduced the term 'spatial interaction model', moving away from that cruder 'gravity model' approach. But in industry, we still find these models referred to quite frequently as ‘gravity models’. So using this entropy maximisation approach, Sir Alan started to apply these models in other fields, particularly in transport planning. And as a result of that work, they were applied more widely across urban planning, and that's really where this model started to take on its role as a location model, and got picked up by the commercial sector, where it's now applied very widely to make location-based decisions, particularly in the retail sector, in the context of the example that we looked at earlier. Sir Alan proposed a family of 4 variants of the spatial interaction model, which reflected 4 different contexts that you may wish to capture. Those 4 variants of the model have different constraints applied to them, reflecting either the demand side, which he called the production side of the model, or the destinations, which he called the attraction side of the model. And those constraints are used to constrain the values that the modelled flows can take, based on known information that we put into the model, and that we use at the model calibration stage. The first variant of the model is the production-attraction constrained, and this is used in a context where we know a lot of information about both the demand and the supply side. In this case, the production side, or the origin totals are known, as are the destination, or attraction side totals - the bits that are shaded in green in the example that's behind me. In this case, the modelled flows that go into the centre of that matrix have to conform to all of the known information that's highlighted in green. So this would be used in a situation where we know our marginal totals, but we need to predict the flows that contribute to those. The opposite of that scenario is our unconstrained model, where we have no constraints in our model. We don't know any of our origin or destination totals, and so that values that go into that matrix, as predicted by our model, are unconstrained. We have no known information that we can use to influence and constrain those flows. More commonly, we might use an attraction or a production constrained model. The attraction constrain model that you can see behind me now knows something about the destination totals, but it doesn't know the information for the origins. And so the flows in that model are constrained by our totals at the attraction side of the model - the bit highlighted in green behind me. The retail example that you've been looking at as part of the accompanying online resources is a production-constrained version of the model. In this case, we know a lot about the origin side of the model. We know how much money is available to spend in each of those demand zones, and so in our model, our totals on the production side are constrained - the modelled flows have to correspond with what we know about those origin totals. And so the cells highlighted in green behind me are those that we would use as part of our model calibration process. And we would expect the flows that are modelled in that matrix to conform with that information that we know about our origin totals. So hopefully that whistle-stop tour of the different variants of the SIM has given you an idea of how it could be applied as quite a versatile model, in a number of different contexts that you may be working in. These models can appear complex at first, particularly when you see them in equation form. But hopefully, by thinking about that retail example, you've seen that once the basics make sense, they can actually be quite straightforward. In this short overview, we had to gloss over a number of important points, but hopefully you can use the accompanying reading list and suggested papers to enhance your knowledge, particularly in your own application areas. I particularly highlight the papers you can see behind me at the moment as a very good introduction to these models, that's written in an accessible style - originally written in fact, for A-Level students when these models were part of the curriculum. I'd just like to finish by pointing out that the spatial interaction model has become one of geography and social sciences' most successful applications, particularly in the commercial sector. And so please don't be put off by the terminology or the equations behind these models. They are robust and reliable tools that have been used to make a number of commercial sector decisions, and I hope that you enjoy starting working with them.