ESRC Microsimulation Seminar Series

Seminar 1: 'Scaling up; scaling down'
London, 2nd April 2009

Dynamic micro-simulation with spatial interactions

Belinda Wu

[EDITED TRANSCRIPT]

First of all thank you for giving me this opportunity to introduce our dynamic spatial microsimulation model to you.

This presentation has been constructed in the following layout, so firstly I'll give a short introduction of our project and then get into some more details of the model and the seminar and sort of the initial results then we try to improve our model, then finally come to the conclusion and the future work.

So what is MOSES? It actually stands for Modelling and Simulation of the eSocial Science and it's a phase one project for the National e Social Science Centre and the specific object of it is to develop an individual based representation of the UK population at a fine spatial scale of wards and we want to project this synthetic population into the future for thirty years from year 2001 to year 2031 and explore the different potentials for public policy planning through various scenarios.

Since we are talking about some larger scale microsimulations I think maybe it's worthwhile to talk through quickly about some larger scale microsimulations. Since Orcutt invented DYNASIM, microsimulation has come a long way, for example CORSIM has inspired several versions of national level microsimulation such as the Canadian version and the Swedish version of national based microsimulation model. Such spatial microsimulation model are very useful because they will model in local context so these models have been widely used and tested but their experience tell us it's often very demanding in terms of resource and skills. So for example APPSIM which is now being developing by NATSEM? (03.07) in Australia with another 13 department agencies and it needs to mention the EUROMOD is, has researchers involved in 19 countries now I believe. So for us it's a tiny team of three researchers over 3 years duration, so we are not expecting to attempt such sophistication as in many areas of these models, but we do expect that new approaches like the Grid enabled eSocial Science approach will enable and encourage more attempts like us and maybe make it more accessible for users to experiment as well.

So what we actually do? MOSES actually once you provide a dynamic representation of key demographic events or transactions in a geographically identified population. We believe microsimulation and macrosimulation are just alternative ways of realising the process. So we use a spatial microsimulation of the population and the dynamics but the structure actually parallels the multi state cohort component projection model. As always microsimulation model depends on good data and important transitions experienced by individuals, where we have some problems ,as our other speakers have already point out as well.

So we are experimenting with an alternative approach called agent based model for a sub-population, students, where we find it's extremely difficult to get empirical data. So we intended to provide individual based representations, forecasts and scenarios, but what does that mean? For starters, the baseline population for Leeds areas- throughout this presentation I use Leeds as an example- that's over 720,000 individuals and there are over 60 millions for the UK, and each individuals have lots of variables, including 6 individual variables, about 20 household variables plus area variables and we need various probabilities for the dynamic simulation. For example we tried to produce localised sex and single year of age based mortality rates for Leeds and we ended up with a matrix with 6,666 cells, so that will give you an idea of the amount of work and resource this model requires.

Needless to mention there are distinctive behaviours from various population groups in different demographic processes and also to add on the problem sometimes you often find the inter- dependency of householder and individual variables. For example if during the fertility module someone is born, that will have an effect on the household as well.

The demographic processes modelled in this microsimulation actually were 6 modularised process, namely ageing, mortality, fertility, marriage, migration and health. Some are very straightforward, there are simple processes like ageing

and mortality which is easy to do, but some are more complicated and needs to be dealt with multi stage processes, such as migration, just to put it simple you have to decide who to move and where to locate them and during this process, simulation of household formation and dissolution can be observed as a result.

After the first version of MOSES, we produced some results and did some analyses on the initial results. Here is an example of the standard age-sex pyramid representation of the Leeds population and we compare that against the ONS (?08.50) projection, so immediately you can tell the difference. That is basically because there are 3 differences in our models than the ONS model assumptions and in mortality we assume a constant mortality rate throughout the whole cycle of the system for 30 years (09.19), so does fertility rate, but ONS actually assumed that there would be a steady improvement in life expectancy as well as increase in the fertility rates. So therefore that explains the larger number of older people and young people. Also about migration, ONS prediction actually allows acceleration of young population such as students after year 2001 census, so that's a major difference caused the differences in the results. But now we are looking at this and trying to look at the result alignment by matching the assumption used in ONS and look at, I suppose it's a step towards the validation of their model. That's at a higher level. When we brought it down to small areas at the level of wards, we find that in the suburban areas such Aireborough and Cookridge, the model results are more robust, reasonable ageing pattern is emerging. However when we look at wards within the city centre where student migration has a great impact, that is not the case, we find it has failed to reproduce younger population in the area, we've lost these peaks for aged 22 to 24, so that meant the microsimulation still didn't capture the subtlety of the local migration pattern. So we try to improve the migration model through a spatial interaction model.

We combine two approaches, first we use a person specific general model using probabilities of migration derived from British Household Panel Survey and then apply that to cloned records from 2001 census SAR data. Then we use 2001 census Spatial Migration Statistics for the location specific information about migration intensities in small areas. Then we use that to modify the results rather than person specific model. And this model has a two stage procedure, the Migrant generation procedure and the Migrant distribution procedure. And I'm going to talk about them in a bit more detail, how am I doing with the time?... OK.

So first we assess migration probabilities using the British Household Panel Survey data for households, groups of movers and individual movers. Then we find those listed variables are important for the different types of migrations. Then we find the groups of movers are not that huge, they're just small numbers, so we combine that with individuals. So for household movers we use age of head of household, household size and household type as the major drivers, for individuals we use age, household size and their marital status. Then they are locally adjusted by using the Special Migration Statistics and this process is explored through a number of simplified assumptions and basically it is a vacancy chain model of household migration.

So the problem can be described as follows, first estimate migration rates by location, age, household size and household type and this process will create a stock of vacant housing at the same time. And for each migrant by location and household type, which is age and size, find a destination location by locating the house type, then we calibrate this process using data on known moves and the known assignment of household type to the house type, which are from Spatial Migration Statistics and BHPS data specifically.

And just a quick recap, this procedure starts with a database of individual based records, then we generate the migrants, aggregate to migration population and generate an array of vacant dwellings(at the same time) and we put them into this spatial interaction model, in which the flow of the migrations react positively to the population characteristics of the destination and the housing characteristics of the destination, and inversely related to the distance between the origin and the destination.

And through the spatial interaction model we then generation the probabilities applied to these migrations, then according to that, located them a new destination and then update these characteristics.

So after we combined microsimulation and spatial interaction model we found to a certain degree it has succeeded, this is a fitness (assessment) of the model. So in suburban areas the results are really good but still inner city centre around the student populated area it's still less satisfactory. So we had to look at students migrants as a sub population and deal with them with another alternative model approach.

So first we look at is characteristics of the student migrants. Students unlike our other population they are highly mobile during their stay at the universities but they mostly they only move around the university area, not in the suburb areas and more importantly most of them will leave the city once they finish their study instead of settling and growing old in the suburb areas. And due to this renewal of university student population each year, the population in the wards within university areas tends to remain younger than other wards. So we considered to use agent based modelling approach to try to improve the results of the model. This is alternative approach that models individuals as agents but agents can be any other things, it can be other things as well not just individual person, it can be a firm, or an organisation, anything really, but in our case that's just an individual person.

So it can model them through the interactions with each other and also the environment they live in. And also it's very flexible to introduce heterogeneous agents with distinctive behaviour through the built-in rules, so you can just define the rules from them and they will act according to their built-in rules. So it's quite useful in modelling features in the model where knowledge and theory is lacking.

So to do this we first here recognise the following rules. There four kinds of students, first year under graduates, other under graduates, the reason of differentiating the two kinds of under graduates is because first year students tend to live on campus more then other under graduates as they grow familiar with an area they tend to move out, and Master students and Doctorate students. We then apply some simple rules for them, so each group is allowed set years to stay in the area and the students prefer to stay with their fellow students. This is a hypothesis developed through other (residential) segregation models, like Schelling's models. The same concept used here. Students stay close to the university,, subject to housing availability, and they don't do marriage and fertility. So it's a comparison between the results of a pure microsimulation and after combined with an agent based model . (We can see) the improvements in the results.

So from pure microsimulation as you can see, students are everywhere in Leeds and also when you look into details of the bands of the number of students, it's quite different than the observed results. Students are almost evenly distributed throughout the city, which is not the case in reality. After combining with agent based modelling we find that the results are quite encouraging, you can see clearly the concentration of the student population at the city centre and also the pattern of the scales to population in the suburban area, and also the natural break of the number of students are quite similar to the observed results, observed numbers of students. This is a 3D representation so you easily interpret the difference.

Mark has talked about a lot of usage of the model, here I am just providing another one for the health planning perhaps. The options are really endless, you can develop or experiment your own policy areas on this population model.

So after simulate the current population into the future for 30 years we find that the people with limited long term illness has increased in suburban areas, so this has an implication on the public service provision in suburban areas, but this can, like I said it can be applied to any policies really.

And conclusions. We think we have built the foundations of an ambitious hybrid model which combines microsimulation, spatial interaction model and agent based features. The next steps are, the first one, we are going to continue our efforts in a second phase project called Genesis, Generative e-Social Science, with our partners in UCL. And specific tasks, the first one results alignment towards the validation by matching assumptions using the ONS projections. And secondly, learn from the model and improve various sub models according to recent population trends until satisfactory "reality" is being reproduced. And thirdly, the one to explore the potential usage of agent based model in conjunction with microsimulation to explore the potentials about interactions between individuals and the environment. We have experimented a little bit in the student migration processes, so students have to interact with their fellow students to find out where they are and to find out whether they, so they can stay for long, whether it is suitable for this type of students, that's interaction with the environment and individuals. But we want to explore more sophisticated interaction. And also this can lead to more sophisticated behaviour modelling as well. And we'd like also to play around with the agents can look into his or her history, personal history and carry this information on and this will have an impact in their future decisions or changes. For example it's a well known fact that some people like miners, they do need to carry their personal history around in terms of the mortality, their mortality probabilities don't suddenly change because they just move to a very leafy area. It's just a matter of ideas to explore with this.

Thank you very much. That's the end of my presentation. CLAPPING

QUESTIONS

Paul Williamson – So any specific questions on this?

Belinda Wu - That's really good, everybody can understand my models, so they can use my model now!

Female question – Sorry I'm losing my voice, the start population say in 2001 and then how you get to 2006, is it going to be the same population which age the ? or whatever or do you start with a different sort of preparation ????

So really it's a ?? process ???

Belinda Wu - Yeah, each year they go through these processes and then produce a new baseline population for next year

Female question – So it's not like start from scratch each year, so it still continues?

Belinda Wu - Ah ah, it continues yeah. But this modularised process, so you can, we actually model the changes, components of change separately so, but obviously they interact with each other and affect each other, but you can actually take out a certain process out if you don't want them, for example I can just produce natural change without migration or whatever, it's very flexible to experiment. Yeah.

Paul Williamson – You were talking about agent based approaches, to what extent is that just a blurring of continuum of microsimulations that are agent based modelling or are they really genuinely different things?

Belinda Wu - Yeah, there has been a lot of debate on this but I think conceptually maybe there is some difference there, I'm thinking maybe microsimulation is more like a statistical procedure which is probability driven, but agent based model is more like rule based, each agent can have its own intelligence to, there is actually a type of agents called BDI agents, belief, desire, intention, so they can actually react to the environment or other individuals and make their own action, make their own mind. And computationally ? speaking, I don't know, should I mention it here or not! But anyway I think microsimulation is more like a central controlled but agent based model can be distributed computing.

END OF RECORDING