Modelling Migration Routes

Martin Hinsch, Jakub Bijak

1st November 2018

1 Introduction

Forced and non-forced migration are ubiquitous as well as fundamentally important demographic phenomena. Not only can migration have substantial cultural, political, economic and demographic effects in all countries involved (i.e. origin, transit and destination), but it has also always been an important and contentious political topic (with new, increased relevance in recent years, see Leurs and Smets, 2018; Ekman, 2018). There are therefore strong scientific as well as practical incentives for understanding migration. However, while there is substantial empirical research on migration, theoretical studies are more sparse and largely focused on voluntary economically motivated migration (Massey et al., 1993).

The advertised - and admittedly lofty - aim of this project is to take the next step in modelling migration by integrating data on the micro-processes that make up migration, including human decision making, to produce comprehensive agent-based simulation models that will be analysed using advanced statistical techniques.

2 Sidenote: the role of models in studying complex systems

At this point it might be helpful to briefly discuss which role models can play in analysing complex social phenomena. In a more general sense there are various purposes for models (Epstein, 2008), here we are however specifically interested in the application of models to the study of complex systems.

Such systems, i.e. systems of many components with non-linear interactions, are notoriously difficult to analyse. Even under best experimental conditions emergent effects can make it nearly impossible to deduce causal relationships between the behaviour and interactions of the components and the global behaviour of the system (Johnson, 2010). This issue is greatly exacerbated in those systems that are not amenable to experimentation under controlled conditions because they can neither be easily replicated nor manipulated, such as for instance large-scale weather, a species' evolutionary history or most medium- to large scale social systems. In these cases modelling can be an extremely useful - and sometimes the only - way to understand the system.

At its heart a model - whether it is a computer simulation or a mathematical model - is a deduction engine, that is a tool to rigorously and automatically infer the consequences of a set of assumptions thereby augmenting limited human reasoning (Godfrey-Smith, 2009; Johnson, 2010). We can distinguish two fundamental ways in which such a tool can be used in the context of studying complex systems:

Proof of causality Understanding causality in complex systems can be challenging since the links between micro- and macro-behaviour or between assumptions and dynamics tend to be opaque. A model can be used in this situation to infer specific chains of causality. By modelling a set of micro-processes or assumptions we can demonstrate (rigorously, assuming no mistakes have been made) which behaviour they produce. This can be used to directly prove or disprove a pre-existing hypothesis about the system. Alternatively, by iterating, the (minimum) set of assumptions required to produce a specific behaviour can be discovered (see Grimm et al., 2005; Weisberg, 2007; Strevens, 2016).

Extrapolation For many complex systems we are interested in their behaviour under conditions that are not directly empirically accessible, such as future behaviour or the reaction to specific changes in circumstances. Assuming we already have a good understanding of a system we can use a model to replicate the mechanisms responsible for the aspects of the system we are interested in and use it to extrapolate the system's behaviour.

At this point it is important to note that everyday use of language tends to obscure what we really do when building a model. We tend to talk about real world systems in terms of discrete nouns, such as "the weather", "this population" or "international migration".

This has two effects. First, it implies that these are things or objects rather than observable properties of dynamic, complex processes. Second it suggests that these phenomena are easy to define with clear borders. This leads to a - surprisingly widespread - "naive theory of modelling" where we have a "thing" that we can build a canonical, best model of in the same way we can draw an image of an object.

In reality, however, for both types of inference described above, how we build our model is strictly defined by the problem we use it to solve: Either by the set of assumptions and behaviours we attempt to link, or by the specific set of observables we want to extrapolate. That means that for a given empirical "thing" (such as "the weather") we might build substantially different models depending on what aspect of that "thing" we are actually interested in. In short, which model we build is determined by the question we ask.

3 Planned modelling efforts

With data collection on migration and decision making still ongoing, current efforts in this part of the project are currently largely focused on exploratory and methodological questions.

3.1 Questions

For this part of the project we are broadly interested in three general questions:

- How does migration work?
- What happens to migration if...?
- How well do we and *can* we understand migration?

Under the umbrella of these very general (and partly overlapping) questions we have a number of specific projects planned.

- How consistent are earlier macroscopic models of migration with a model based on a bottom-up approach? Are the simplifying assumptions in earlier models justified?
- How predictable is migration *intrinsically*? That is, assuming perfect knowledge, how stochastic is the process per se (see Bijak and Wiśniowski, 2010; Azose and Raftery, 2015)?
- What determines the number of migrants arriving at a specific destination?
- Under a number of realistic scenarios in terms of policy, climate and geostrategy, what would happen with migration?
- How well are we able to model migration in terms of *mechanisms* (Courgeau et al., 2017)? Given the uncertainty concerning the processes involved, how authoritative can our model be (see Poile and Safayeni, 2016)?
- What is the intrinsic numerical uncertainty of a best effort model? That is, assuming perfect data, how certain are our predictions concerning migration?
- Given best efforts in terms of structure how does the uncertainty of the available data affect the model's numerical uncertainty?

3.2 Challenges

As with any complex system the first major challenge in this case consists in defining and delineating the system. First, horizontally - that is, which part of the world do we consider peripheral and which parts should be part of the model? Second, vertically - how much detail do we consider important?

The answers to both questions of course depend crucially on the purpose of the model in question. Beyond that, however, there are a number of principal and practical complications.

Practically we are constrained by various factors such as availability of data, complexity of implementation and computational and analytical tractability of the simulation (Silverman, 2018).

More severely and even ignoring practical considerations, however, there is fundamentally no straightforward way to determine which processes need to be part of the model and which do not (Barth et al., 2012; Poile and Safayeni, 2016). The issue is less pronounced as long as we are working in the context of a proof-of-causality modelling effort, since finding which assumptions produce a specific kind of behaviour is precisely the aim of this type of modelling. However, as soon as we intend to use our model to extrapolate system behaviour we get into trouble. Trying to include all processes that might affect the dynamics we are interested in, but leaving out those that only unnecessarily complicate the model becomes a difficult task.

In the current project the situation is further complicated by the fact that empirical data on many processes is quite sparse or non-existent. For example there is strong anecdotal evidence that smugglers play an important role not only in transporting migrants across the Mediterranean, but also in helping them for instance along the Balkan route. Empirically it is, however, extremely difficult to asses the prevalence of smuggling on these routes since all involved parties have a vested interest in understating these numbers. As another example it is obvious that borders and border patrols are an extremely important factor in determining how many migrants arrive in which EU country. While we do have numbers on border apprehensions (as for example reported by Frontex, 2018) it is unclear how these numbers map to actual border crossings (in particular taking into account repeat attempts). Essentially we have no hard knowledge concerning the underlying processes: How likely is it for a migrant to be caught at the border? How much do migrants usually know about border controls? How do they use that knowledge in deciding where to go? What do migrants do that failed to cross a border?

For this reason the project has to put a strong emphasis on quantifying the uncertainty of any given modelling effort. In particular we need to test not only for numeric uncertainty resulting from the intrinsic stochasticity of the modelled processes but also for uncertainty resulting from our lack of knowledge of the processes themselves (Poile and Safayeni, 2016).

3.3 Methods

For most of the questions in the project we are unable to exclude that differences as well as interactions between individuals are an essential part of the dynamics we are interested in. At least as a starting point this commits us to agent-based modelling as the default architecture.

The advantages and disadvantages of this approach have been discussed at length elsewhere (Lomnicki, 1999; Bryson et al., 2007; Peck, 2012; Poile and Safayeni, 2016; Silverman, 2018). In the context of this project the method presents two major challenges. First, as mentioned in the previous section, many of the processes involved in our target system are not well defined. We will therefore have to be careful to take the uncertainty resulting from this lack of definition into account. This is not an easy task for a simple model, but even less so for a complicated agent-based model. Approximate Bayesian computation might be helpful in this context to pinpoint at least some of the missing information (Grazzini et al., 2017).

Second, large agent-based model tend to be computationally heavy which reduces the range of parameter values that can be tested and thus ultimately fine-grainedness of any result, including sensitivity analyses. One way around this that has been used successfully in the past is to train a Gaussian process emulator against the available results and use that as a stand in for further analysis (Bijak et al., 2013).

4 Model 1: Routes & Rumours

In the following we will detail the first full modelling effort currently underway in the project. In this part of the project we investigate the formation of transit routes based on the interactions between individuals' decisions and their environment.

Theoretical studies based on an economic optimality approach usually assume that migrants decide where to move by selecting the destination that optimizes economic (or other) outlook for them or their relatives/descendants in a rational and fully informed way. Furthermore it is usually assumed that the journey itself is of no interest and has little or no influence on the decision (Massey et al., 1993).

On closer look these assumptions appear rather problematic.

The criteria that people use to decide where to go seem to be influenced by, but by no means restricted to economic optimality (Castles, 2004; Simon et al., 2016). In many cases social and practical factors appear to be more important. For example, the existence of a social network in the destination country often seems to be an important criterion (De Haas, 2010; Dekker et al., 2018; Borkert et al., 2018).

Furthermore, information about destination countries as well as how to get there often appears to be rather limited (Borkert et al., 2018). Studies have shown that many migrants distrust "official" sources of information and primarily rely on information provided by trusted contacts (Dekker et al., 2018). This also explains how false or strongly exaggerated rumours about destination countries can spread and persist .

In addition there often seems to be little information about local conditions on the journey. Migrants therefore have to rely on contacts who have done the same journey before, inofficial sets of "instructions" or (voluntary or paid) local helpers in order to find the way to their destination (De Haas, 2010; Dekker et al., 2018).

Even assuming good information the travel routes of migrants are constrained by a huge number of practical factors such as local infrastructure, availability of transport, border controls, etc. We can also see that repeat migration along specific routes quickly leads to the establishment of an informal local infrastructure along these routes that provides migrants with food, shelter, transport etc. (De Haas, 2010)

Finally, which path migrants choose can impose constraints on the destination countries available to them. In the case of forced migration it could also be argued that the choice of destination will be secondary to the primary aim to get to a safe place.

All of this seems to suggest that migration is far from being a perfect optimization process. First, the route migrants take is constrained by a complicated set of factors (some of which involve feedback effects) and second, which routes migrants take does affect which destination country they will end up in.

In addition to these scientific considerations the issue also has significant humanitarian and political dimensions. Every year thousands die while attempting to cross the Sahara, the Mediterranean or the Sonoran desert, sometimes as a direct result of well-intentioned yet uninformed political decisions. Understanding the processes involved in the formation of migration routes might help to avoid at least some of these fatalities. Politically in particular in the case of migration to the EU migration routes are highly relevant. Due to the Dublin convention (EU, 1997) and, more generally, the structure of the EU's political decision making, peripheral countries as well as those that happen to be on the path of a main migration route tend to be primarily responsible for dealing with migration (with consequences varying between countries). On the other hand countries such as Hungary can get away with unilaterally rerouting migration along its borders. Where migrants arrive has therefore proven to be of high political significance.

It seems therefore worthwhile to take a closer look at migration routes and how they are determined by micro-processes. In particular we are interested in how availability and transfer of information - specifically within the migrant community - affects the migrants' journey and their choice of destination. In this we will for now focus on the journey itself and ignore the processes that lead up to migrants leaving their home as well as what happens after they have arrived at their (or a) destination.

The questions we will attempt to answer are:

- Are migration routes "optimal"?
- Are migration routes predictable?
- How quickly can changes in circumstances change routes?
- How difficult is it for individuals to find a new route?

We are currently in the process of implementing a spatially explicit agent-based model to answer these questions. While a full, detailed model description is beyond the scope of this report, we will in the following give a general overview of the current state of the model. We will then discuss some preliminary observations, some issues that have come up during model construction as well as future plans.

Model description

Space is represented as a two-dimensional grid. Grid cells differ in a number of properties such as (police) control, friction and availability of resources. Agents move stepwise through the grid cells, collecting information about the local conditions as well as from their network of social contacts. They decide where to move based on their knowledge of the landscape ahead while generally heading towards the other end of the simulated world.

Space

The world is represented as a square grid of location cells. Grid cells - which are thought to represent roughly $1km^2$ - have a number of properties that agents interact with and that they can collect information about:

friction How much time and effort it takes to cross that cell. High for rough landscape and cities, low for flat agricultural land and roads.

control How likely it is to be picked up by police or border patrols.

information The availability of information on other parts of the world.

resources The availability of food, shelter, etc.

In addition cells have an *opacity* that determines how much local expertise is required to make use of the local facilities.

At the beginning of the simulation the grid is filled with default values. Friction values are assigned at random with a controllable degree of spatial auto-correlation using the diamond-square algorithm. Then a network of cities and transport links connecting them is generated and property values of the respective grid cells are adjusted accordingly.

Agents

Agents can either stay at their current location, explore their surroundings and mingle with other agents or they move ahead to the next grid cell. Staying agents automatically gain some knowledge about their current as well as surrounding grid cells. They also meet other agents present in the same cell and - with a certain probability - add them to their list of contacts.

Moving agents assign a quality to all surrounding grid cells based on their knowledge and a set of (global for now) preferences. For example a cell with a high control has a lower quality than one with a low control value, high availability of resources improves quality, etc. Cell quality generally increases along one axis of the world (to ensure directed movement) as well as with proximity to a target (see below). Furthermore agents devalue uncertainty. After checking the quality of the available options agents move to the cell with the highest quality.

Next to moving or staying, agents also exchange information with their social network, by either gaining new knowledge or improving on their existing knowledge from a proportion of agents in their list of contacts.

Knowledge, orientation and planning

How agents gain knowledge of the world and how they use it to decide where to go are crucial elements of the model. We have to assume that most migrants know relatively little about the areas they travel through and that what they know (or think they know) is often based on unreliable second-hand information.

In the model we implement that by letting agents gain knowledge either by exploring a location on their own or by receiving information from another agent (who itself might again have received it somewhere else). To simulate the lack of reliability we assume that each piece of knowledge is associated with a trust value that increases the more information an agent collects about a location. A higher trust value increases the perceived quality of a location. It also makes it more likely that another agent will copy the piece of information in question.

Even if moving through a largely unknown landscape, people probably navigate using a combination of local decisions and long-term planning. There is also evidence that migrants often travel along specific waypoints, locations with a good availability of information for example or with an opportunity to rest and obtain food or transport (Kingsley, 2016). In a first rudimentary implementation of this process in the model we let agents assess locations according to a second set of quality criteria to determine whether they are suitable as long-range targets. The current list of targets is then used to inform movement decisions (see above).

Observations so far, issues and plans

In its current incarnation the model already shows a number of promising effects. The first few cohorts of agents departing from the origin act as a vanguard and - due to lack of information - move largely according to a directed random walk. Subsequent generations of agents, however, begin basing their decisions on the information collected by their precursors and start moving along paths of low friction. After a while enough information is available that agents start setting long-range targets. At this point migration routes emerge that are followed by most agents.

In terms of methods implementing imperfect spatial knowledge in a multi-agent system turned out to be an interesting challenge (see below). As far as we can tell this has never been attempted before (at least not on this scale) and might turn out to be a valuable methodological contribution in its own right.

Issues

The biggest problem during implementation turned out to be the agents' memory. In the first version of the model we naively implemented memory as a list of pieces of information where each location an agent had any information about was represented by one item. When two agents exchange information they have to essentially compare all locations they know about. With number of contacts per agent as well as number of agents increasing continuously over time as well this very quickly brings the simulation to a near halt.

We were able to increase efficiency substantially by assuming that agents only exchange information about interesting parts of the landscape. For this we let agents have an expectation concerning the properties of a grid cell. Grid cells that differ substantially from expectation are classified as "interesting", those that do not as "boring". During information exchange only detailed information concerning interesting grid cells is exchanged. Boring grid cells that agents hear about from another agent are marked as such in a separate area of memory.

This does not solve another issue of a grid-based representation of knowledge, however, which is that information about large-scale structures (e.g. roads) and semantic classification (e.g. city versus country side) is lost. In the next step we will therefore experiment with letting agents remember and communicate about discrete objects (e.g. city, village, road, border) instead of grid cells.

Plans

With the basic structure of the model implemented we can start adding the more interesting parts:

global information So far the only information entering the population originates in exploration by the agents themselves. It makes sense, however, to assume that agents have access to a certain amount of external information they can for example obtain online (Dekker et al., 2018).

wrong information One of the starting points of this part of the project was the observation that many migrants seem to hold exaggerated or even wrong beliefs about their country of destination. Existing results on the spread of different types of rumours in social networks could be used to add this effect to the simulation (e.g. Hu et al., 2018).

local feedbacks It has been observed that local, informal infrastructure tends to spring up along real-world migration routes (Dekker et al., 2018). This might lead to a positive feedback effect that could greatly increase the stability of routes once they are established (De Haas, 2010).

destinations So far we are only investigating the emergence of transit routes. If it turns out that route configurations are to a large degree emergent or stochastic it will be interesting to see how this affects the likelihood of agents to arrive at different destinations. On the other hand letting agents decide on a destination early on might substantially affect the development of routes.

changing circumstances Finally, with a full structure for the spread of information in place, it will be interesting to see how a change in circumstances (e.g. closing of a border, change of legislation in destination, new support centre by a charity) affects developing and established migration routes.

References

Azose, J. J. and Raftery, A. E. (2015). Bayesian Probabilistic Projection of International Migration. *Demography*, 52(5):1627–1650.

Barth, R., Meyer, M., and Spitzner, J. (2012). Typical Pitfalls of Simulation Modeling - Lessons Learned from Armed Forces and Business. *Jasss*, 15(2):5.

Bijak, J., Hilton, J., Silverman, E., and Cao, V. D. (2013). Reforging the wedding ring: Exploring a semi-artificial model of population for the united kingdom with gaussian process emulators. *Demographic Research*, 29(October 2013).

Bijak, J. and Wiśniowski, A. (2010). Bayesian forecasting of immigration to selected European countries by using expert knowledge. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 173(4):775–796.

Borkert, M., Fisher, K. E., and Yafi, E. (2018). The Best, the Worst, and the Hardest to Find: How People, Mobiles, and Social Media Connect Migrants In(to) Europe. *Social Media and Society*, 4(1).

Bryson, J. J., Ando, Y., and Lehmann, H. (2007). Agent-based modelling as scientific method: a case study analysing primate social behaviour. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 362(1485):1685–98.

Castles, S. (2004). Why migration policies fail. *Ethnic and Racial Studies*, 27(2):205–227.

Courgeau, D., Bijak, J., Franck, R., and Silverman, E. (2017). Model-Based Demography : Towards a Research Agenda. In Grow, A. and Van Bavel, J., editors, *Agent-Based Modelling in Population Studies*, chapter 2, pages 29—-51. Springer, Cham.

De Haas, H. (2010). The Internal Dynamics of Migration Processes: A Theoretical Inquiry. *Migration Studies*, 36(10):1587–1617.

Dekker, R., Engbersen, G., Klaver, J., and Vonk, H. (2018). Smart Refugees: How Syrian Asylum Migrants Use Social Media Information in Migration Decision-Making. *Social Media and Society*, 4(1).

Ekman, M. (2018). Anti-refugee Mobilization in Social Media: The Case of Soldiers of Odin. *Social Media and Society*, 4(1).

Epstein, J. M. (2008). Why Model? *Jasss*, 11(4):12.

EU (1997). Dublin Convention.

Frontex (2018). Risk Analysis for 2018. Technical report.

Godfrey-Smith, P. (2009). Models and Fictions in Science. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 143(1):101–116.

Grazzini, J., Richiardi, M. G., and Tsionas, M. (2017). Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control*, 77:26–47.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H.-H., Weiner, J., Wiegand, T., and DeAngelis, D. L. (2005). Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. *Science*, 310(5750):987–991.

Hu, Y., Pan, Q., Hou, W., and He, M. (2018). Rumor spreading model with the different attitudes towards rumors. *Physica A: Statistical Mechanics and its Applications*, 502:331–344.

Johnson, B. R. (2010). Eliminating the mystery from the concept of emergence. *Biology & philosophy*, 25(5):843–849.

Kingsley, P. (2016). *The new odyssey: The story of Europe's refugee crisis.* Guardian Books, London.

Leurs, K. and Smets, K. (2018). Five Questions for Digital Migration Studies: Learning From Digital Connectivity and Forced Migration In(to) Europe. *Social Media + Society*, 4(1):205630511876442.

Lomnicki, A. (1999). Individual-based models and the individual-based approach to population ecology. *Ecological Modelling*, 115(2-3):191–198.

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., and Taylor, J. E. (1993). Theories of International Migration: A Review and Appraisal. *Population and Development Review*, 19(3):431.

Peck, S. (2012). Agent-based Models as Fictive Instantiations of Ecological Processes. *Philosophy & Theory in Biology*, (March).

Poile, C. and Safayeni, F. (2016). Using computational modeling for building theory: A double edged sword. *JASSS*, 19(3):8.

Silverman, E. S. (2018). *Methodological Investigations in Agent-Based Modelling*. Springer.

Simon, M., Schwartz, C., Hudson, D., and Johnson, S. (2016). Illegal Migration as Adaptation: An Agent Based Model of Migration Dynamics. In *2016 APSA Annual Meeting & Exhibition*.

Strevens, M. (2016). How idealizations provide understanding. *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*, pages 37–49.

Weisberg, M. (2007). Three Kinds of Idealization. *Journal of Philosophy*, 104(12):639–659.