Modelling Migration: Decisions, Processes and Outcomes

Jakub Bijak[1✉], Philip Higham[1], Jason Hilton[1], Martin Hinsch[1], Sarah Nurse[1], Toby Prike[1], Oliver Reinhardt[2], Peter W. F. Smith[1], Adelinde Uhrmacher[2]

[1] University of Southampton, UK; [2] University of Rostock, Germany
✉ Corresponding author: j.bijak@soton.ac.uk

1. Introduction

Human migration is uncertain and complex, and some of its distinct features, such as migration routes or responses of flows to the underlying drivers, can emerge in a very short time and are characterised by very high volatility, as witnessed during the 2015–16 Syrian asylum crisis (Kingsley, 2016). One of the key drivers of this complexity, and one of the main reasons behind the inefficiency of attempts to control migration, is the agency of various actors involved – migrants, institutions, intermediaries, smugglers, and so on (Castles, 1994). This agency also remains one fundamental reason why migration typically eludes attempts at its theoretical description, explanation and prediction, the efforts for undertaking which typically remain scattered across various disciplines (Arango, 2000).

To address migration challenges, a unified and interdisciplinary approach is needed that would formally reflect this underlying agency, and describe micro-level migration decisions and the resulting macro-level processes in a coherent way. In this paper, we present four building blocks of such an approach, based on the methodology of agent-based modelling. These building blocks include: construction of an agent-based simulation model of migration, in this case illustrating the formation of migrant routes; a unified framework for assessment of the existing data and their quality; results of psychological experiments on human decision making under uncertainty; and the choice of an appropriate programming language and modelling formalisms. These elements are subsequently brought together by the means of statistical (Bayesian) meta-modelling, using Gaussian processes, which enable the analysis of the uncertainty of the model outputs and their sensitivity to various parameters.

2. Building blocks of the proposed modelling approach

2.1. Agent-based modelling of migration routes

At the core of the proposed approach is an agent-based model. We start from the question of how the migrant routes, clearly observed for real migration processes, are formed and sustained. Migrants attempting to reach a safe destination often have to make their navigation decisions based on very limited information that is to a large degree sourced from other migrants that have made the journey before (see e.g. Kingsley, 2016). Therefore, communication between migrants could be a key factor in determining the dynamics of flows, especially in such rapidly-changing processes as Syrian asylum migration (Dekker et al., 2018). In this example, we study the effect of information transfer on variability and optimality of migration routes by using an agent-based model with explicit representation of geography, resources and agents' knowledge, where the agents need to navigate a partially unknown landscape (see Figure 1). The model is described in more detail in Hinsch and Bijak (2019) and the key innovations towards the state of the art (see Klabunde & Willekens, 2016 for a review) include an explicit modelling of agent knowledge, their social networks, as well as information exchange.

We find that unless agents very quickly acquire objective information from the environment, a higher degree of social information exchange leads to less predictable and less optimal migration routes. This indicates that if a high proportion of information is socially received, routes are the result of self-organization rather than optimization. We suggest that similar effects should occur in all situations where individuals have to make complex decisions under limited information but in a social context.

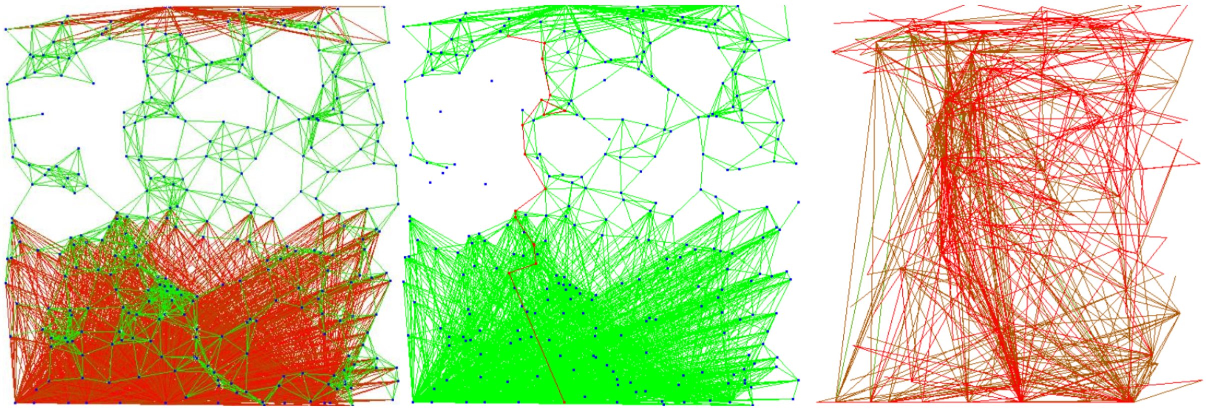

Figure 1. Sample screenshot from the model monitoring: map of the simulated world with cities and routes (left), a random agent's perception of the world and their route plan (centre) and their social network connections (right)

2.2. Data sources and their quality evaluation

Migration data are notorious for their problems with completeness, quality and comparability (Poulain et al., 2006); problems which are only exacerbated in the case of asylum migration (Singleton, 2016). To make the various aspects of data quality explicit and useful for inclusion in modelling, a common and transparent framework for assessment of different data sources is required (for an example, see e.g. Vogel & Kovacheva, 2008), which would enable supplementing data with meta-information on quality.

Specifically, for recent Syrian asylum migration to Europe, we have proposed such a unified framework of analysis (Nurse & Bijak, 2019), comprised of eight dimensions: purpose of collection; timeliness of data; trustworthiness; detailed disaggregation; definition of population under study and associated definitions; transparency of the sources; their completeness; as well as sample design (for surveys). Each of these dimensions, as well as a global quality score for a given data source, is classified into one of three categories: green, amber or red (with two in-between classes, green-amber and amber-red), reflecting the various aspects of variation and bias inherent in the source, which need to be adequately described in any modelling endeavour using the data, ideally by using probability distributions. As a general rule of thumb, we posit that in the agent-based modelling exercise, contextual data, as well as data on micro-level drivers of migration processes should be used as model input, whereas macro-level information on the features of these processes can be used to benchmark model output.

2.3. Eliciting decisions under uncertainty

Utility functions for financial decisions have been widely studied. However, it is less clear how these findings generalise to decisions in other domains, such as migration. Developing a better understanding of decision-making is crucial for improving our knowledge of migration processes, and yet previous work has often assumed that migrants simply maximise their utility, which is reflected in how decisions are typically described in existing agent-based models of migration (Klabunde & Willekens, 2016). This assumption is called into question by cognitive research on decision-making, such as the prospect theory (Kahneman & Tversky 1979), and has not previously been empirically tested (although see Czaika, 2014).

In our work, we conducted two pre-registered experiments, respectively with 130 lab-based and 403 online participants, who were tasked to choose between gambles presented in either a migration or financial context. A financial context is commonly used for studying risky decision-making and allowed our findings to be compared with previous research. We elicited non-parametric utility functions following Abdellaoui et al. (2016) and tested whether they differed depending on the context. Loss aversion was calculated based on the inflection point of the utility function at the reference value, as well as by regressing the points of the utility function elicited for gains on those elicited for losses.

The results suggest that despite there being many similarities in risky decision-making across financial and migration contexts, there may also be some differences, particularly in relation to losses, although this result proved somewhat sensitive to the analytical strategy. This highlights the need to examine migration decision-making specifically, rather than simply relying on assumptions from other areas without testing them in a migration context.

2.4. Programming languages and formalisms

As the final building block of the proposed modelling approach, we elucidate the differences between the alternative formalisms through which a simulation model can be realised by implementing the model in parallel in a general-purpose programming language (Julia) and in a domain-specific language specifically developed for demographic agent-based modelling applications (ML3, Warnke et al., 2017). In particular, the granularity of time and formalisation of the stochastic processes governing the events in the model – whether as continuous or discrete – matters for the results obtained. Tentative findings indicate that seemingly innocuous decisions made by the modellers may have important downstream effects for the simulation results, which reinforces the need for the highest levels of transparency of model description. In addition, the parallel model realisations also help identify crucial features and trade-offs for the future implementation of the model, as well as providing important guidelines for the further development of the domain-specific language.

3. Statistical analysis of model properties

The four elements of the proposed approach are brought together through a unified statistical analysis of the migration model and its outcomes. As argued in Reinhardt et al. (2018), the existing methods for analysing the results of agent-based models have been so far largely descriptive and were not utilising the possibilities offered by contemporary statistical theory and principles of experimental design. To fill this gap, we apply the methods of Bayesian meta-modelling based on Gaussian Process assumptions (Oakley & O'Hagan, 2004). This allows us to combine the different elements of the modelling approach (data, models, decision mechanisms and languages), analyse the uncertainty of model outputs and their sensitivity to a range of input parameters, and to calibrate of the model to the extent allowed by the available data. This approach allows for formally identifying the input parameters that matter for a given output – in our case, seven parameters related to information exchange, information errors, social networks and exploration. Figure 2 shows an example analysis of a response surface for one selected output, the share of agents following a plan, obtained by using the GEM-SA package (Kennedy & Petropoulos, 2016). The model is currently undergoing calibration to align it with empirical data.

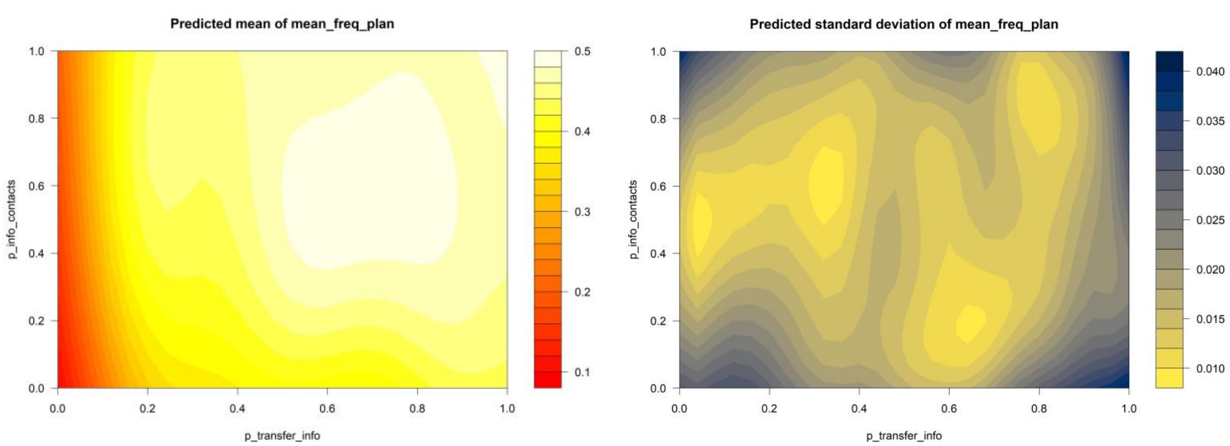

Figure 2. Estimated response surface of the proportion of time the agents follow a plan vs two input parameters, probabilities of information transfer and of communication with contacts: mean proportion (left) and its SD (right)

4. Preliminary conclusions

The proposed approach, coupling agent-based modelling with statistical analysis, is a natural way to address the dual challenges of complexity and uncertainty of contemporary migration processes. It does so by shedding light on the micro-level mechanisms underpinning the observed macro-level processes, and at the same time describing the probabilistic properties of the migration system under study. Each of the four building blocks provides a contribution in its own right, but combining them together by the means of a statistical analysis allows for formally exploring the behaviour of complex migration systems in a more rigorous and transparent way than has been the case before.

More generally, by framing the analysis in such a multi-perspective way, we hope to strengthen the case for model-based demography, which is intended to fill the widely acknowledged theoretical gaps in population studies (Courgeau et al., 2017; Burch, 2018). At the same time, as migration is particularly lacking a formal discussion of causal mechanisms, which would enable making testable predictions (Willekens, 2018), we hope that the proposed approach will go some way towards achieving this goal.

Acknowledgements: Funding of the European Research Council for the project *Bayesian Agent-Based Population Studies: Transforming Simulation Models of Human Migration* (CoG-2016-725232), is gratefully acknowledged.

References

Abdellaoui M, Bleichrodt H, L'Haridon O and van Dolder D (2016) Measuring loss aversion under ambiguity: A method to make prospect theory completely observable. *Journal of Risk and Uncertainty*, 52(1):1–20.

Arango J (2000) Explaining Migration: A Critical View. *International Social Science Journal*, 52, 283-296.

Burch T (2018) *Model-Based Demography: Essays on Integrating Data, Technique and Theory*. Cham: Springer.

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: A Grow and J Van Bavel (eds.), *Agent-Based Modelling in Population Studies*, pp. 29-51. Cham: Springer.

Czaika M (2014) Migration and economic prospects. *Journal of Ethnic and Migration Studies*, 41(1), 58-82.

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).

Hinsch M and Bijak J (2019) Rumours lead to self-organized migration routes. Paper for the Agent-based Modelling Hub, Artificial Life conference, ALife 2019. Via: https://www.southampton.ac.uk/baps/publications.page

Kahneman D and Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.

Kennedy MC and Petropoulos GP (2016) GEM-SA: The Gaussian Emulation Machine for Sensitivity Analysis. In: GP Petropoulos & PK Srivastava (eds), *Sensitivity Analysis in Earth Observation Modelling*. Elsevier. (pp. 341-361).

Kingsley P (2016) *The New Odyssey: The Story of Europe's Refugee Crisis*. London: Faber & Faber.

Klabunde A and Willekens F (2016) Decision making in agent-based models of migration - State of the art and challenges. *European Journal of Population*, 32(1), 73-97.

Nurse S and Bijak J (2019) Meta-information on data sources on Syrian migration into Europe, v.1.1. Mimeo, University of Southampton. Via: https://www.southampton.ac.uk/baps/inventory/data-sources.page

Oakley JE and O'Hagan A (2004) Probabilistic sensitivity analysis of complex models: A Bayesian approach. *Journal of the Royal Statistical Society: Series B*, 66(3), 751–769.

Poulain M, Perrin N and Singleton A (2006) *THESIM: Towards Harmonised European Statistics on International Migration*. Louvain-la-Neuve: Presses Universitaires de Louvain.

THESIM: Towards Harmonised European Statistics on International Migration

Reinhardt O, Hilton J, Warnke T, Bijak J and Uhrmacher AM (2018) Streamlining Simulation Experiments with Agent-Based Models in Demography. *Journal of Artificial Societies and Social Simulation* 21 (3), art. 9.

Singleton A (2016) Migration and Asylum Data for Policy-Making in the European Union – The Problem with Numbers. CEPS Paper 89, Centre for Europe and Policy Studies, Brussels.

Warnke T, Reinhardt O, Klabunde A, Willekens F and Uhrmacher AM (2017) Modelling and simulating decision processes of linked lives: An approach based on concurrent processes and stochastic race. *Population Studies*, 71(S1), 69–83.

Vogel D and Kovacheva V (2008) Classification report: Quality assessment of estimates on stocks of irregular migrants. Report of the CLANDESTINO project, Hamburg Institute of International Economics, Hamburg.

Willekens F (2018) Towards causal forecasting of international migration. *Vienna Yearbook of Population Research*, 16, 199–218.

■