Characterising behaviour of human agents

A. Smajgl[1], Daniel G. Brown[2], Diego Valbuena[3], and Marco G. A. Huigen[4]

[1] *CSIRO Ecosystem Sciences, Climate Adaptation Flagship, ATSIP, University Rd, Townsville 4810 QLD*
[2] *University of Michigan, School of Natural Resources and Environment, Ann Arbor, MI, USA*
[3] *International Livestock Research Institute, System-wide Livestock Programme, P.O.Box 5689, Addis Ababa, Ethiopia*
[4] *Universität Hohenheim*

alex.smajgl@csiro.au

Abstract: Agent-based modelling has advantages in modelling independent of disciplinary boundaries, explicitly simulating people's behaviour and allow for spatial referencing of dynamics processes. These features have fostered accelerated applications in empirical situations. However, the ability of simulating human behaviour comes with the need to specify assumptions on human behaviour and behavioural changes due to external changes. Such characterisation and parameterisation tasks are widely conducted in an ad-hoc fashion and without the required transparency. Recent work provided a framework for systematically structuring the characterisation and parameterisation. This paper explains the framework, which identifies twelve distinct sequences for the characterisation and parameterisation of human behaviours. In addition, this paper outlines the process of testing and revising the framework in the broader modelling community.

Keywords: *Agent-based modelling; empirical characterisation; parameterisation.*

1. INTRODUCTION

Environmental policy and management increasingly demand the integration of cross-disciplinary knowledge of socio-ecological systems. Understanding human responses to environmental and/or policy is critical for understanding socio-ecological systems and the outcomes they produce. As agent-based modelling (ABM) allows for simulating actual decision-making processes of individuals or groups of individuals this modelling technique is gaining importance (Matthews et al. 2007).

This paper is focused on ABM, in which human agents are represented at the scale of households or individuals. In empirical contexts, this class of ABM is mostly chosen for its ability to simulate explicit human decision making processes (mostly described in discontinuous functions). However, the main strength of ABM can be a key weakness. As the implementation of human decision making processes is the main strength of ABM, the agent attributes and behavioural response functions that represent these processes require knowledge support from qualitative and/or quantitative empirical sources. Unfortunately, there are no standard approaches to documenting and communicating the empirical support that underlies modelling and design decisions in agent-based models (Bohensky, Smajgl, & Herr, 2007).

Empirical characterisation and parameterisation of human decision making processes can involve a range of methods that include: expert knowledge, surveys, interviews, and participant observation. While we do not claim that the specific methods described here include all possible options, they include those that are most commonly used in practice and help define what we believe to be a robust set of sequences within which alternative specific methods can be applied. In many cases it is necessary to combine several empirical methods in a sequence to fully parameterise behavioural responses in ABM.

The development of an ABM that aims to simulate behavioural responses of humans requires two fundamental steps in which empirical data are required: the development of behavioural categories and the scaling to the whole population of agents. The effectiveness of different methodological sequences depends on the modelling context (Robinson, et al., 2007). Most significantly, the size of the human population in the system to be modelled helps to determine appropriateness of various methods, and can be anything from 20 households to many million households. Carrying out in-depth interviews with 20 households is realistic. With larger populations and more complex socio-ecological interactions, it becomes necessary to work with a sample of the population. Sampling often requires characterisation of the existing heterogeneity of agent attributes and behavioural responses in order to provide a simplified description of behavioural realities that can guide sampling and the realistic parameterisation of model agents.

The first challenge for the parameterisation of ABMs that represent larger populations is the identification of suitable empirical methods for eliciting behavioural data. Then, the sample-based data need to be translated into behavioural representations for the whole population (i.e. up-scaling). Up-scaling can be performed proportionally or disproportionally. Proportional up-scaling assumes that the sample is representative for the whole population. This can result in under-representation of behavioural minorities. Disproportional up-scaling is needed if these behavioural minorities are likely to determine system dynamics, such as free-riders in common pool situations. Disproportional up-scaling is based on the development of agent typologies. The inverse problem (i.e., down-scaling) occurs when data for a large collection of agents are acquired for aggregated sources (e.g., over census enumeration districts), and populations of agents are generated through some simulation technique that honours the aggregate data.

This paper aims to provide an initial structure that can be used within rigorous discussions of parameterisation methods for human behaviour in ABM. Thus, we developed a framework for this process followed by a discussion of methods available for what we call the parameterisation sequence. Discussing sequences instead of individual methods allows for a systematic understanding of existing options and their suitability across a range of modelling contexts. Then, we give examples demonstrating the application of the framework. The paper concludes with a discussion of the need to further tests of the generality of the framework.

2. PARAMETERISATION FRAMEWORK

ABM requires the systematic representation of three main phenomena: agents, their social networks and the agent environment (**Error! Reference source not found.**). This paper discusses only the parameterisation of human agents; we leave development of empirical frameworks for the other phenomena to others. To achieve agent parameterisation, six different methodological steps were identified (**Error! Reference source not found.**), each with multiple empirical methods available (**Error! Reference source not found.**).

The first step of the framework is to identify different classes of agents (M1, **Error! Reference source not found.**). Agent classes are defined as groups of agents that share the same sequence of actions. The sequences of actions represent the modelled behaviour of the agents and constitute the totality of their actions within the model. The identification of distinct agent groups and their sequences of actions can be based on expert knowledge (EK), participant observation (PO), laboratory experiments (LE), interviews and role-playing games (RPG). We group these methods as *Methods 1 (M1)* as listed in **Error! Reference source not found.**.
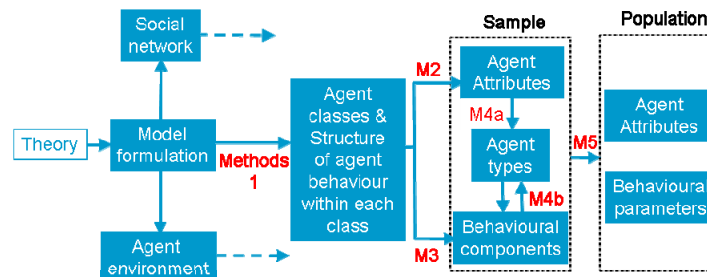

Figure 1. Framework for parameterisation of ABM

The second step of the framework can be either to specify the values for agent attributes (M2) or the parameters for the behavioural rules agents follow (M3). While agents in each class share the same sequence of actions, the parameters governing the magnitudes of the actions and the degree to which various sources of information influence those actions can vary. Agents with the same behaviours and similar or identical parameters of behaviour are of the same agent type (where types are the equivalent of sub-classes). Agent types can be developed from agent attributes or from data on agent behaviours. Thus, data have to be elicited to measure real-world attributes and behavioural responses. Behavioural responses can include observed relationships between qualitative or quantitative contextual variables and the actions or probabilities of actions carried out by agents. We group methods for obtaining agent attributes as *Methods 2* (*M2*) and those for obtaining behavioural responses as *Methods 3* (*M3*). **Error! Reference source not found.** lists surveys and census data as *M2* methods. Options for step *M3* include surveys, interviews, field experiments, participant observation, RPG, time series data, and expert knowledge.

Agent types can be developed from agent attributes or from observed behavioural responses. This means that the sequence of steps within the proposed framework (**Error! Reference source not found.**) can vary. If types are developed from agent attributes, *M4a* methods can be employed, through the combination of clustering and regression, the combination of correlation and expert knowledge, expert knowledge individually, or dasymetric mapping. The first three options can also be used if developing agent types from behavioural data. Additionally, participant observation can be conducted (method group M4b).

Table 1: Overview of methodologies relevant for the parameterisation of behavioural traits of human agents

M1	M2	M3	M4a ATB	M4b BT	M5
Expert Knowledge	Survey	Survey	Clustering& Regression	Clustering& Regression	Proportional
ParticObser	Census (incl GIS data)	Interviews	Correlation& Expert Knowledge	Correlation+ Expert Knowledge	Census/GIS based assignment
Lab Experiment		Field Experiment	Expert Knowledge	Expert Knowledge	MonteCarlo
Interviews		ParticObser	Dasymetric mapping	ParticObser	
RPG		RPG			
		Time series data			
		Expert Knowledge			

Once agent types are developed, and their attributes and behavioural response functions characterised, agents in the whole population need to be assigned to the appropriate agent type (M5). If data were elicited from the whole population, this step does not require any up- or down-scaling, as the sub-population (sample) and population (**Error! Reference source not found.**) are identical. If there are some assurances that the sample is representative of the population, the scaling can be conducted proportionally. Under such assumptions each sample is assumed to represent a number of other agents, determined by the proportion of the population sampled, and is then *cloned*, accordingly. Cloning does not mean that all agents of one type have to be identical, as ranges of values or measures of variability can be used to introduce some level of heterogeneity. If representativeness of the sample cannot be assumed, agent types can be up- or down-

scaled by means of census data or a Monte Carlo approach. These methods are grouped as M5 methods and listed in **Error! Reference source not found.**.

3. PARAMETERISATION SEQUENCES

What combinations of M1 to M5 methods make sense will depend on the context. In this paper, we characterise the empirical context using three variables: the size of the population (N), high/low; the behavioural diversity (BD), high/low; and the possibility of creating representative samples. Behavioural diversity describes the relative number of different types of actions that agents can take in a given context. If the options are relatively limited, e.g., to grow rice or not, then behavioural diversity is low. If the agents choose among a large set of options, behavioural diversity is said to be high. The possiblility to create a representative sample depends on, for instance, availability of funding for the modelling project and ease of accessing the people.

Table 2: Overview of techniques for the parameterisation of human agent behavioural responses

Case	When	M1	M2	M3	M4a	M4b	M5
1	N high Sample representative	EK or PO or Interviews (optional: Experiments or RPG	Survey	Interviews or Field Experiments or RPG or EK	-	-	(Cloning or MonteCarlo) an Spatial references
2	N high BD high Sample not representative	EK or PO or Interviews (optional: Experiments or RPG	Survey	Interviews or RPG	-	(EK or Clustering) and Regression	CD/GIS based
3	N high BD high Sample not representative	EK or PO or Interviews (optional: Experiments or RPG	Survey	Interviews or RPG	Clustering or (Statistics and EK)	-	CD/GIS based
4	Sample representative BD high	EK or PO or RPG	Census& Surveys (PO)	Interviews (PO) or RPG	PO	PO	CD/GIS based
5	N high or low BD low Access to agents low	EK or PO	Disaggre-gated Census	Time series data AND EK	-	EK	CD/GIS based
6	N high BD low Sample not representative	EK or PO or Interviews (optional: Experiments or RPG	Survey	Interviews or Field Experiments or RPG	-	(EK or Clustering) and Regression	CD/GIS based
7	N low (=sample)	EK or PO or Interviews (optional: Experiments or RPG	Survey	Interviews or Field Experiments or RPG	-	-	No up-scaling
8	N high BD low	EK or PO or Interviews (optional: Experiments or RPG	Survey OR Census	EK or Interviews or Field Experiments or RPG	Statistics and EK	-	CD/GIS based
9	N low BD low	EK or PO or RPG	Census	EK or Field Experiments or RPG	EK	EK	No up Scaling
10	N low BD high	EK or PO or RPG	Census	EK or RPG	EK	EK	No up Scaling
11		EK or PO	Aggreg. Census	Time series data	Dasymetric mapping	Clustering & Regression	CD/GIS based
12		EK or PO	Disaggreg. Census	Time series data	-	Clustering & Regression	CD/GIS based

[N: agent population; BD: Behavioural diversity; EK: Expert Knowledge; RPG: Role-playing games; PO: Participant observation]

We have identified twelve different sequences of methods (referring to those listed in Table 1) that can be followed to parameterise the behavioural responses of human agents in ABMs (Table 2). There may be other sequences, but we believe this set covers most empirical ABM contexts. This reasonably comprehensive list of cases and parameterisation sequences (even if this is only a first step) should allow for a structured testing of methods to identify which sequences prove to be effective under various modelling conditions. Such testing could aim for quantifying uncertainty and guide the ABM community

towards more robust empirical model development. The characterisation of sequences should facilitate description and documentation of ABM applications.

While model description is clearly important (Grimm, et al., 2006), we argue that good description of the empirical support for models in specific contexts is equally important in allowing decision makers to make good judgements about the value of the model results. We describe the modelling contexts within which the different cases make sense and emphasise different routes through the parameterisation processes (**Error! Reference source not found.**) depending on these contextual variables.

The first case represents a modelling context with a large population and the possibility to develop representative samples (Case 1). An understanding of agents and their actions (M1) can be developed using expert knowledge (EK), participant observation (PO), or interviews. Detailed agent attribute data (M2) can be elicited by conducting sample surveys, while behavioural data (M3) can be obtained from interviews, field experiments, role playing games (RPG), or expert knowledge (EK). Assuming that the sample used to generate attributes and behavioural parameters is representative, proportional up-scaling can be carried out, in which each data point is cloned to generate the whole population (M5). This step needs to include spatial referencing if agent locations are an element of the model design.

In cases where the behavioural diversity is high and assumptions on the representativeness of samples do not hold, variations to the first sequence have to be introduced. While steps M1 and M2 can remain unchanged, the behavioural complexity makes field experiments or expert knowledge less suitable in step M3. Instead, two alternative sequences that require the development of agent types (for disproportional up-scaling) can be used. Firstly, interviews can be conducted to generate behavioural data for development of agent types (Case 2). For each type, distributions of agent attribute values can be identified by surveys. This means that the actual sequence of this approach (following **Error! Reference source not found.**) is M1→M3→M4b→M2→M5.

Instead of developing agent types from behavioural data, types could be constructed from agent attributes (Case 3). This means that a survey could be conducted to elicit attributes and clustering techniques applied to develop agent types. Then, interviews can be targeted to people who match the profile of agent types. The behavioural data would then be dis-proportionally up-scaled to the whole population based on a process that assigns each agent to a type by mapping type-specific characteristics against census data. This approach sequences methods as M1→M2→M4a→M3→M5.

Cases 2 and 3 require census data to allow for up-scaling (M5). This type of up-scaling requires the identification of the types using census data, in order to determine the proportions of the population in each type. While both of these parameterisation sequences reduce uncertainty resulting from sample non-representativeness, both approaches introduce uncertainty from the development of agent types by clustering (i.e. disproportional up-scaling).

In situations where representative samples can be developed in spite of high behavioural diversity (Case 4), expert knowledge or participant observation can provide broad systems understanding (M1). Participant observation allows for developing agent types based on attribute data (census or surveys) and behavioural data (interviews or role playing games). The final up-scaling could be guided by census data.

Case 5 describes a situation where the actors are not accessible for interviews or surveys (e.g., because they are not responsive to surveys or due to funding limitations). In such cases agent classes can be identified based on expert knowledge or participant observation. This sequence requires the availability of disaggregated data to provide agent attributes. Assumptions on behavioural responses can be developed from statistical analysis of time series data or expert knowledge. Experts could then guide the mapping of behavioural assumptions into the agent population.

Case 6 is another variation of the first sequence. This approach assumes that instead of a proportional sample, the development of agent types is based on behavioural data (M4b). Disproportional up-scaling could be performed with a census-based mapping, as described for Case 2.

Case 7 involves relatively small populations. It is very similar to Case 1 but does not require any up-scaling (M5) because the whole population can be described.

Case 8 is another variation of Case 1 (and similar to Case 3). This case represents a large population with low behavioural diversity. It assumes that disaggregated census data are sufficiently available for providing agent attribute information. In such a case agent types could be statistically developed from census data or surveys (e.g., for disproportional up-scaling). Then, interviews could be conducted to obtain behavioural response data (M3) before finally behavioural data are mapped via census data into the agent population. The sequence would be M1→M2→M4a →M3→M5.

In cases with small populations and available census data, two different sequences can be used, depending on the behavioural diversity. In case of low behavioural diversity (Case 9) agent classes can be identified using expert knowledge, participant observation or role playing games. Disaggregated census data would provide agent attributes. Agent attributes could further be characterised by adding expert knowledge (M4a). Then, behavioural data can be developed for each agent based on expert knowledge, field experiments, or role-playing games. Experts could guide the process of refining behavioural assumptions for each agent. The case of high behavioural diversity (Case 10) only differs from Case 9 in that field experiments are likely to become unrealistic for M3 due to complications in designing experiments.

The final two cases start with expert knowledge or participant observation to design system properties. Behavioural parameters are then developed through analysis of time-series data, while attribute data are available in the form of aggregated or disaggregated census data. Statistical techniques such as clustering or regression methods can be employed to process the time-series data into agent types (M4b). Aggregated census data has to be disaggregated through dasymetric mapping (Case 11) before behavioural data can be assigned for the whole agent population (i.e., down-scaling). If disaggregated census data are available they can be directly used for proportional up-scaling assumptions on agent behaviour (Case 12).

4. DISCUSSION

This paper developed a framework for the parameterisation of attributes and behavioural response functions of human agents in agent-based models. The aim is to allow for systematic testing, documentation and communication of methods that can be compared across various modelling contexts. Such an effort to systematically understand the effectiveness of methodologies employed in ABM for the parameterisation of behavioural aspects is critical because the (technical) strength of ABM, which is its flexibility and capacity in representing human decision making, has to be based on a robust empirical support. Without translating this advantage into empirical contexts, the applied value of ABM remains limited and the focus remains on hypothetical and theoretical analyses. The framework for and describing different parameterisation sequences should also be tested in various modelling (Robinson, et al., 2007) and application contexts (Matthews, Gilbert, Roach, Polhill, & Gotts, 2007). Further refinement and testing of this framework would allow for reducing model uncertainty, for provision of systematic guidelines for building and applying empirical ABMs, and for facilitation of the model comparisons (Parker, Brown, Polhill, Deadman, & Manson, 2008; Dawn C. Parker, et al., 2008). The framework described in this paper is a first step in this process. The community of ABM researchers need to assess the framework and whether it is sufficient to cover all existing or potential approaches for empirical agent-based models. Opportunities for extension of the framework might appear as methods for the parameterisation for the social networks of agents and the agent environment are refined (**Error! Reference source not found.**). For example, the requirement for real-world data on social network might change the approach to obtaining behavioural data for individual agents, as individual behaviours might be dependent on how these individuals interact with each other.

ACKNOWLEDGEMENTS

REFERENCES

Bohensky, E., Smajgl, A., & Herr, A. (2007, 2007). *Calibrating behavioural variables in agent-based models: Insights from a case study in East Kalimantan, Indonesia.* Paper presented at the MODSIM 2007 International Congress on Modelling and Simulation.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., et al. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling, 198*(1-2), 115-126.

Matthews, R., Gilbert, N., Roach, A., Polhill, J., & Gotts, N. (2007). Agent-based land-use models: a review of applications. *Landscape Ecology, 22*(10), 1447-1459.

Parker, D. C., Brown, D., Polhill, J., Deadman, P. J., & Manson, S. M. (2008). Illustrating a new 'conceptual design pattern' for agent-based models of land use via five case studies - the Mr. Potatohead framework. In A. López Paredes & C. Hernández Iglesias (Eds.), *Agent-Based Modelling in Natural Resource Management* (pp. 23-51). Upper Saddle River, New Jersey: Pearson Education.

Parker, D. C., Entwisle, B., Rindfuss, R. R., Vanwey, L. K., Manson, S. M., Moran, E., et al. (2008). Case studies, cross-site comparisons, and the challenge of generalization: comparing agent-based models of land-use change in frontier regions. *Journal of Land Use Science, 3*(1), 41-72.

Robinson, D. T., Brown, D. G., Schreinemachers, P., Janssen, M. A., Huigen, M., Wittmer, H., et al. (2007). Comparison of empirical methods for building agent-based models in land use science. *Journal of Land Use Science, 2*(1), 31-55.