11-9-2015

So You Think You Can Model? A Guide to Building and Evaluating Archaeological Simulation Models of Dispersals

Iza Romanowska
University of Southampton, i.romanowska@soton.ac.uk

Recommended Citation
http://digitalcommons.wayne.edu/humbiol_preprints/79

This Article is brought to you for free and open access by the WSU Press at DigitalCommons@WayneState. It has been accepted for inclusion in Human Biology Open Access Pre-Prints by an authorized administrator of DigitalCommons@WayneState.
So You Think You Can Model? A Guide to Building and Evaluating Archaeological Simulation Models of Dispersals

Iza Romanowska

1University of Southampton.

Correspondence to: Iza Romanowska, Institute for Complex Systems Simulation and Centure for the Archaeology of Human Origins, University of Southampton, UK. E-mail: i.romanowska@soton.ac.uk.

Short title: So You Think You Can Model? A Guide to Archaeological Simulation

KEY WORDS: SIMULATION, AGENT-BASED MODELLING, ARCHAEOLOGY, HUMAN DISPERSAL.

Abstract

With the current surge of simulation studies in archaeology there is a growing concern for the lack of engagement and feedback between modellers and domain specialists. To facilitate this dialogue I present a compact guide to the simulation modelling process applied to a common research topic and the focus of this special issue of *Human Biology*—human dispersals. The process of developing a simulation is divided into nine steps grouped in three phases. The
conceptual phase consists of identifying research questions (step 1) and finding the most suitable method (step 2), designing the general framework and the resolution of the simulation (step 3) and then by filling in that framework with the modelled entities and the rules of interactions (step 4). This is followed by the technical phase of coding and testing (step 5), parameterising the simulation (step 6) and running it (step 7). In the final phase the results of the simulation are analysed and re-contextualised (step 8) and the findings of the model are disseminated in publications and code repositories (step 9). Each step will be defined and characterised and then illustrated with examples of published human dispersals simulation studies. While not aiming to be a comprehensive textbook-style guide to simulation, this overview of the process of modelling human dispersals should arm any non-modeller with enough understanding to evaluate the quality, strengths and weaknesses of any particular archaeological simulation and provide a starting point for further exploration of this common scientific tool.

All archaeologists start modelling the moment they step out of the excavation trench. We interpret the individual finds (e.g., pots, skeletons, buildings etc.) within certain frameworks (e.g., pottery typologies, bone taxonomies, architectural types etc.) and analyse sets of finds to detect population level patterns (e.g., cultural similarities, age profiles of bone assemblages, urban
development, etc.) with strict analytical rigour. However, the interpretations of these patterns in terms of human behaviour and causality are predominantly built using natural language, i.e. they are constructed in the form of conceptual models that hypothesise which causal mechanisms might have led from actions of individual actors (the owners of pots, users of skeletons and inhabitants of buildings) to the detected population-level patterns. These causal relationships are often described as, for example, “culture A influenced culture B,” “the dispersal reached area C,” or “population D outcompeted population E.” Although inferences like these are made on the basis of rigorously collected and analysed data, and are commonly built upon extensive research and a good understanding of multiple strands of evidence, they nevertheless represent a thought experiment and are therefore limited to the researcher’s ability to mentally manipulate the information in his or her head (Evans et al. 2013; Davidsson and Verhagen 2013). In recent years we have witnessed a surge of studies using qualitative formal methods, in particular simulation, which try to fill this gap (for overviews see Kohler 2012; Lake 2014; Barton 2014) (Figure 1).

Simulation, understood as a scientific, computational methodology, is a technique in which researchers build a simplified model of a real-world system and observe its behaviour over a defined time period in order to explore its properties in a formal environment. Whilst formal models (which can be created in a multitude of ways, e.g., using statistics, GIS or game theory) lie at the heart of
the approach, simulation adds the dynamic dimension, therefore enabling researchers to investigate how different scenarios unfold over time. Simulation is a common scientific technique used across diverse disciplines ranging from physics, engineering and medicine to ecology and the social sciences. Simulation approaches are used for a variety of purposes, such as prediction, theory building, data exploration and integration, model testing and selection, and they have proven to be an indispensable tool for investigating the complexities of a wide variety of modern physical, biological and social systems (Grimm 1999; Mitchell 2009; Chattoe-Brown 2013).
Figure 1. The relationship between data analysis tools and simulation approaches. The former allow archaeologists to detect patterns in the data but are not designed to test different scenarios of how these patterns might have emerged. Simulation enables researchers to construct conceptual models of past interactions in a formal environment and compare them to the archaeological record (see also Thomas 1973: fig. 1; Gilbert and Troitzsch 2005: fig. 2.1; Raub et al. 2011: fig. 1).

Although simulation techniques have been applied in archaeological research since the early days of the method (e.g., Doran 1970; Thomas 1973; Wobst 1974), they have not occupied such a central position in archaeology as they have done in other branches of science. The reasons behind this slow
adoption are unclear. One contributing factor might be that the rapid rise of simulation techniques in other disciplines (Gilbert and Troitzsch 2005: 6-9; Hollocks 2008; Heath and Hill 2010), coincided with the period when “post-processual” critiques were introduced into the archaeological discipline, which led to a re-thinking of the epistemological foundations of the discipline and a criticism of the dependence on the scientific method. Furthermore, the process of simulation modelling differs from the common in archaeology research process consisting of data collection and analysis followed by proposing the most plausible conceptual model to explain the patterns detected in the data. In simulation modelling this process is sometimes reversed, as the model-building phase often precedes data collection and analysis.

As a result, simulation techniques are not commonly taught as part of the standard archaeological curriculum at university level, and the discipline lacks comprehensive textbooks or introductory texts aimed specifically at archaeologists. Also, compared to other disciplines, simulation techniques are infrequently used to approach archaeological research questions (Lake 2010; 2014). This has led to a very limited engagement by archaeologists with the few existing archaeological models and to limited discussion and feedback between modellers (themselves often archaeologists) and domain specialists - experts in the Palaeolithic paleoenvironments, Bronze Age pottery, Roman trade, etc.
This paper presents a guide for archaeologists to build and evaluate simulation models, with a particular focus on the case studies of human dispersals. Its main aim is to support more critical engagement between modellers and domain specialists in order to encourage more and better applications of simulation modelling in archaeology. It is written for those archaeologists who have limited previous experience with simulation and would like to get to know the technique well enough to feel confident to critically engage with previous and future models. While not aiming to be a comprehensive textbook-style guide to simulation (e.g., Pidd 1998; Chung 2004; Gilbert and Troitzsch 2005; Miller and Page 2007; Railsback and Grimm 2011; Squazzoni 2012; Edmonds and Meyer 2013; O’Sullivan and Perry 2013), this overview of the simulation development process with a strong focus on challenges and examples of best practice should arm any non-modeller with enough understanding to evaluate the quality, strengths and weaknesses of any particular archaeological simulation. For modellers, this overview may encourage self-reflection on the process we go through when developing models and the pitfalls one can easily fall into at almost every stage of the simulation modelling process.

The large-scale movements of people, objects, and ideas, described in a multitude of terms from dispersal, colonisation, and peopling, to demic diffusion, is a particularly rewarding topic for simulation studies (Young 2002; Steele 2009). Hence the particular focus of this paper on modelling human dispersals.
Past human dispersals offer good examples of processes in which interactions between individuals and groups and the environment give rise to meso- and macro-scale patterns that can be relatively easily detected in the archaeological record. It has inspired a relatively high number of simulation studies, from models of the first “Out of Africa” dispersal ca. 1.8 Mya (e.g., Mithen and Reed 2002; Nikitas and Nikita 2005; Romanowska 2015), and simulations of the spread of the Neolithic (e.g., Wirtz and Lemmen 2003; Ackland et al. 2007; Fort et al. 2012; Isern and Fort 2012), to models of the more recent peopling of Oceania (e.g., Di Piazza et al. 2007; Davies and Bickler 2015). Although no two dispersal events were the same, they share enough common characteristics to be pulled out of their “time periods” and be treated as one “research problem” for the purposes of this review. Throughout the paper I will focus on these common characteristics and their representation in a simulation approach, and illustrate the different challenges and methods for dealing with them.

A number of authors have described and discussed the process of simulation building in other disciplines before (e.g., Shannon 1998; Chung 2004; Becker et al. 2005) and in most cases the steps they recognised coincide with the ones described here. However, there are a number of differences affecting the process of model building in archaeology, including (but not restricted to) a strong focus on explanation rather than prediction (Epstein 2008), uncertainties in input data limiting the scales at which the models can be built and affecting the
resolution of results, the common need to rely on proxies in parameterization, the specific nature of archaeological data limiting its use in validation, the epistemological issues of applying models of modern human behaviour to past societies or the different focus of the research questions asked by archaeologists. These are only a few challenges affecting the archaeological modelling process in ways other branches of science would be unfamiliar with and which have not been fully discussed in our discipline yet (McGlade 2014). These particularities strongly affect how models are constructed in archaeology and they will be accounted for throughout the paper.

![Figure 2. The model development sequence.](image-url)
Figure 2 illustrates the process of building and evaluating simulation models, and serves as an outline of this paper. After the principal steps of identifying research questions (step 1) and finding the most suitable method to tackle them (step 2) the modeller designs the simulation, first by sketching out the general framework and the resolution of the simulation (step 3) and then by filling in that framework with the modelled entities and the rules governing interactions between them (step 4). The technical phase of coding and testing (step 5) is followed by literature review style research aimed at finding realistic value ranges for the model’s parameters (step 6) as well as deciding on which scenarios to run (step 7). Finally, the key phase of analysing and re-contextualising the results of the simulation (step 8) naturally concludes in the dissemination of the findings of the model in publications and sharing the code through online repositories (step 9). Each step will first be described with a special emphasis on the key issues and then followed by examples of their relevance for the study of human dispersals.

The stages of the model development process presented here follow rather arbitrary divisions and represent an ideal scenario. In reality, the different phases of building a simulation are often closely intertwined and performed in parallel or even reversed. For example, it is not uncommon to revisit the research questions when it turns out that it is impossible to obtain the necessary input data for parameterization. For the sake of clarity, we will assume in this paper that in this instance the simulation project smoothly progresses from step one to step nine.
The Purpose of the Model and Research Questions

As with all research, every modelling exercise starts with identifying a research topic and constructing research questions. Although one could build a simulation model without clear research questions (often termed “exploratory modelling”) there are two major disadvantages to this approach. The first is the risk that one’s research will not reveal anything interesting or new. Although this is a possibility for any type of scientific investigation, the risk increases if one tries to do research without identifying any specific research questions. Secondly, knowing one’s research questions and the purpose of the model significantly aids its construction (Grimm and Railsback 2012; Norling et al. 2013). Sterman (2000, p.89-90) compares the research questions to a knife used to cut the simulation model into the desired shape and size. Almost all subsequent design decisions are made in light of the research questions, such as the choice of the simulation method and the level of detail, the decision of what to include and what to leave out of the model, or which parameters to test and which to leave fixed. These points are all decided on with respect to what the modeller believes is crucial to understanding and solving the research problem.

Although the range of research questions that can be approached with various simulation techniques is vast, they share certain characteristics. In general,
simulation tools were developed to test hypotheses, provide a formal environment for theory building, attach probabilities and conditions to conceptual models, and identify areas where future research effort should be focused. At the same time, there are certain research questions which computational modelling cannot solve as they belong to one of the two other legs of the scientific method: the data and the theory. There is no simulation model, no matter how skilfully developed, that can tell us, for example, when the earliest Neolithic farmers arrived in Europe as this kind of knowledge depends on empirical data, e.g., a discovery of a new site predating the already known ones. Equally, simulation, being by definition a computational instantiation of a conceptual model, cannot produce new and original but still contextually plausible hypotheses in a way the creative process of conceptual modelling can.

Previous simulation studies of dispersal and diffusion processes looked at a variety of research topics and questions. They often focused on determining the rate of advance of the dispersal front and factors that increase or decrease that rate. A classic example is the study of the Neolithic spread into Europe through equation-based models (Wirtz and Lemmen 2003; Hazelwood and Steele 2004; Ackland et al. 2007; Lemmen et al. 2011; Baggaley et al. 2012; Fort 2012; Fort et al. 2012; Isern and Fort 2012). Other common research topics include the Palaeolithic dispersals (Mithen and Reed 2002; Scherjon 2013; Callegari et al.
2013; Wren et al. 2014; Romanowska 2015), the peopling of the Americas (Steele et al. 1996; 1998; Rouly and Crooks 2010) or the spread of the Bantu languages (Russell et al. 2014). Some studies investigated interdependencies between different factors within an abstract theoretical framework (e.g., Wren et al. 2014) while other focused on finding explanations for particularly startling patterns in the archaeological record (e.g., Nikitas and Nikita 2005), still others tested a specific hypothesis (e.g., Scherjon 2013) or contrasted two alternative models (e.g., Fort 2012).

**Choosing the Method**

Once the purpose of the model and research questions are defined, the natural next step is to choose the method with which to tackle them. However, the question of whether we actually need a simulation to answer the research question posed in step one is not a trivial one (O’Sullivan et al. 2012). Simulation should only be used if there are no suitable analytical techniques for addressing the research question, for example GIS analysis or statistics. Data analysis techniques are both quicker and more reliable as they were developed for specific tasks. For example, there is no need for an elaborate simulation model if one wants to infer from a dataset of C14 dates the average speed of the colonisation wave as this can
be done using standard analytical tools and, in some cases, even simple arithmetic.

However, often there are good arguments for a simulation approach. For example, the research may concern causality or explanation of non-intuitive data patterns, or aims to test one or contrasts several hypotheses for which predictions are unclear. Simulation can tackle system complexity thus its common application when multiple factors are believed to influence the system and when the modelled interaction and processes are known to be non-linear (see Bonabeau 2002). However, different simulation techniques are most suited for different research problems. In archaeology the three most popular simulation approaches applied to dispersal case studies are: Equation-based modelling (EBM), Cellular Automata (CA), and Agent-based modelling (ABM) (Figure 3) (Steele and Shennan 2009).
Figure 3. The three most common types of simulation techniques used in archaeological dispersal case studies. The top panel shows the structure of the underlying algorithms: (EBM) an equation describing the dispersal as a function of population growth and diffusion, (CA) rules describing probabilities determining how each cell in a matrix is updated, and (ABM) rules describing how individuals interact. The bottom panel shows example visualisations for each type.

Equation-based Modelling

Sometimes referred to as system dynamics\(^1\) or mathematical simulations, equation-based models (EBMs\(^2\)) are probably one of the most common tools in science (Gilbert and Troitzsch 2005; Epstein 2006; Steele and Shennan 2009). They consist of one or a set of equations that can be solved analytically or numerically. An example of such a model, used commonly in modelling.
dispersals, is the Fisher-Skellam model, derived and modified from equations used to model chemical reaction-diffusion processes. It consists of two terms (see Figure 3): one describing population growth and the other simulating diffusion from the denser to sparser occupied regions (Steele 2009). The equation can be extended by other terms, for example a competition with another population (e.g., Isern and Fort 2012) or different factors affecting the speed of the diffusion wave (Steele et al. 1998). Equation-based modelling is known as a “top-down” approach (Klüver et al. 2003; Macal and North 2010) since the population is treated as one homogenous entity and only the population-level characteristics are modelled, e.g., the overall population size, the growth rate, or general rules for competition between groups. The entities and the relationships between them are described mathematically, and the model is then run for a range of parameter values (Epstein 2006; Premo 2006).

The representation of the population as one homogenous entity is the main drawback of the method. Although, it makes for a fast development and good tractability of EBMs, the homogeneity of the modelled entities means that EBM usually does not deal well with modelling variability within the population (e.g., variable levels of particular characteristics among individuals), chance processes and behaviours dependent on individual circumstances and chance events (e.g., innovation, adaptation, and learning), or changes to the population occurring during the dispersal, such as evolutionary and cultural trends (e.g., shifts in the
adaptations, changes in social and cultural behaviour, etc.). Another disadvantage of EBM related specifically to archaeological case studies is the mathematical skill needed to develop them. For the same reason, they are also difficult to communicate to an archaeological audience, which is usually not accustomed to seeing past human behaviour represented in the non-intuitive language of equations. However, there are many benefits of using equation-based models as they are easy to analyse, understand and interpret, and their better mathematical tractability makes them virtually impervious to code mistakes or misinterpretations of the results as compared to other simulation methods (Steele and Shennan 2009). Most of the models of the spread of the Neolithic are developed as EBMs (e.g., Ackland et al. 2007; Fort et al. 2012).

3.2. Cellular Automata. Cellular automata models (CAs) reached the height of their popularity in the early 2000s, even inspiring some researchers to declare them to be “a new type of science” (Wolfram 2002). CAs represent a virtual world in the form of a grid in which each cell is modelled as a numerical state value (usually boolean 0-1 or True/False) (Gilbert and Troitzsch 2005; Bithell and Macmillan 2007), which in dispersal studies usually represents the state of being “occupied” or “empty.” The value may be a function of the states of the neighbouring cells as in the famous Game of Life (Poundstone 1985), but in dispersal applications it is usually a probability that describes the odds that the cell becomes occupied or empty at each time step. Although these types of models
are relatively easy to develop and communicate, this advantage comes at the cost of a restricted range of applications. For example, CA cannot be used in any models where population dynamics need to be explicitly modelled. Moreover, contrary to EBMs the models require specific techniques to analyse results. Examples of the use of CA for studying human dispersals include the SteppingOut model (Mithen and Reed 2002) of the first “Out of Africa” dispersal, and the related model by Nikitas and Nikita (2005).

Agent-based modelling

Sometimes referred to as individual-based or multiagent simulations, agent-based models (ABMs) are becoming increasingly popular in archaeology (Lake 2014). ABMs comprise individual heterogeneous agents interacting with each other and with their environments. These interactions are governed by rules of behaviour defined by the modeller and often modelled stochastically (i.e., using a random number generator) (Gilbert and Terna 2000; O’Sullivan and Haklay 2000; Macal and North 2010; Davidsson and Verhagen 2013). Occasionally these simple rules lead to unexpected population-level patterns, a phenomenon referred to as “emergence” (Epstein 2006: p. 31-33). ABM is often described as a “bottom-up” approach, because population level patterns are expected to emerge from the local interactions between agents. For example, population growth may be modelled not as a function of time (like in EBM) but as a result of interactions between male and female agents under conditions imposed by the modeller (e.g., both
agents need to be in spatial proximity of each other, reach a certain age or have enough energy, etc.). ABMs can suit a very wide range of applications, and they are intuitively easy to build and understand because of the agents and their behaviour being explicitly represented and easy to relate to everyday experiences (Bonabeau 2002) as well as thanks to the existence of user-friendly modelling platforms (Wilensky 1999).

However, this flexibility comes at a price. As much as EBMs tend towards elegant, but generalising and, as a result, sometimes overly simplistic representations of past systems, in ABMs it is only too easy to keep on adding new variables, additional rules of interactions and different types of entities. This may easily lead to over-engineered models littered with complex representations of aspects that are not relevant to the research questions. More realism in a simulation is not a problem in itself. However, the higher the number of variables the better chance of introducing unintentional errors, the more difficult it is to check the code and the more challenging it is to recognise which elements of the model caused the “interesting” results (Railsback and Grimm 2011; Bullock 2014). Although this issue is not inherent to ABM, it is nevertheless more common than in other techniques (Ropella et al. 2002; O’Sullivan et al. 2012; Galán et al. 2013). The “SteppingIn” model (Scherjon 2013) and the SHEEP model (Romanowska 2015) are examples of the archaeological use of ABMs for the study of Pleistocene human dispersals.

Pre-print version. Visit digitalcommons.wayne.edu/humbiol after publication for the final version.
If interactions between entities within a heterogeneous population are important to address the research questions then ABM is probably the right choice, but when population-level dynamics are key then EBM or CA are more suitable (see also Siebers et al. 2010). Models can also be built using a combination of these three techniques, for example the QHG (Quantitative Human Geophylogeny) model by Callegari and colleagues (2013) uses both EBM elements and agents. Although the decision of which simulation technique to use should be determined by the research questions and the characteristics of the modelled system, rather than personal preference or skills, there is a clear preference for ABMs among archaeologists. However, the reliability, speed and clarity of EBMs are likely to make it a highly attractive tool once simulation becomes better established in archaeology.

**Establishing the Scale and Resolution**

In this step the modeller determines an appropriate scale of analysis and resolution of the entities and inputs of the simulation model. Some models are purely abstract and do not need to be translated into real world terms (Premo 2010). However, most simulations in archaeology, especially those that will be compared to real world data, need at least some definition. In such cases the modeller has to be explicit about what the different entities in the model represent:
are agents individuals, households, or groups? Is the time step a year, a generation or 1000 years? Deciding on these questions early on is important as consistency and compatibility of analytical scales is key: for example, if the agents represent individuals then using a time step of 1000 years would make the model meaningless.

The research questions defined in step one and the methods chosen in step two will to a large extent determine the analytical scale of the model (Grimm and Railsback 2012). If the researchers are interested in the general patterns of the peopling of entire continents it is safe to assume that individual decisions and life trajectories can be aggregated into larger entities and modelled in a probabilistic or even deterministic way. However, if they want to pursue a more fine-grained research problem, for instance assessing the hypothesis that following seasonal migrations of large game provided a boost to the speed of the peopling, then going down to the level of at least groups is unavoidable. Evidently, the method chosen in step two will depend on and impact the level of detail presented in the model. For example, when using EBM or CA the mechanisms of interactions are modelled on the population scale, hence any questions invoking variability in individual or even group-level behaviour would be difficult to address.

The final factor to consider in this step is the resolution of different types of input the model is reliant on, often called “exogenous data,” for example, soil
quality maps, the Pleistocene temperature curve or detailed data used to
classify individual behavior. Jason Noble (pers. comm. 2013) aptly
summarized: “even a very exact number multiplied by a random number is still a
random number,” meaning that if the simulation depends on any external data
input (e.g., environmental data), then the resolution of that input is also the
maximum resolution of any reliable results produced by the model. Thus, trying
to, for example, use environmental data of a resolution of hundreds of kilometres
to a model that treats agents as individuals, would be equivalent to multiplying a
very exact number (agent behaviour) by a random number (the environment).
Therefore, the rule arising from Noble’s observation is to treat the element with
the lowest resolution as a guide to what is a feasible framework for the whole
simulation. This rule forms the basis of a methodological framework commonly
applied in ecology called Pattern-oriented Modelling (POM)(Grimm et al. 2005).

Most examples of simulations of the earliest, Pleistocene dispersals use
very coarse chronological and spatial resolutions: hundreds of years for a time
step, and 0.5-1 degrees for distances as they inherit the resolutions of the available
environmental data. As a result, in these models group or individual level
interactions are treated in an aggregated manner. Usually the closer to the present,
the higher the resolution of the input data and of the whole simulation. For
example, in the “SteppingOut” model of the first Out of Africa (Mithen and Reed
2002) each step represents 250 years, while Steele and colleagues (1996; 1998)
were able to use a chronological resolution of 1 year in their model of the Paleoindian peopling of the Americas.

**Entities and Rules of Interaction ("Defining the Ontology")**

In the previous step, we determined the general characteristics of the simulation. In this step it is time for a more detailed description of the modelled entities and the rules of interaction governing their actions. A model is always a simplified representation of the system of interest (Klüver et al. 2003). However, how it will be represented depends on the modeller and their hypothesis of what constitutes the essential elements and mechanisms of the system and which ones are necessary to answer the research questions defined in the first step (Ahrweiler and Gilbert 2005; McGinnis et al. 2011). This conceptualisation of the modelled system into a set of entities and rules of interaction is known as an “ontology” (Gruber 1993).

Building the conceptual framework of a simulation is a difficult task. On the one hand, it has to be as specific as possible to enable translation into computer code in the next step, on the other hand ontologies are descriptive in nature as they represent the modeller’s belief of how the “real-world” system works (Hofmann et al. 2011). Therefore, multiple simulations of the same system may be built on the basis of different conceptual models comprising distinct
entities and the rules that govern their behaviour. In the social sciences there are two schools of thought that disagree as to how conceptual models should be represented in simulation models: the KISS, or “Keep It Simple, Stupid,” and the KIDS, or “Keep It Descriptive, Stupid” (Edmonds and Moss 2005; Premo 2010; Railsback and Grimm 2011; David 2013). The KISS approach works on the premise that one should build models from the bottom up, starting from the simplest possible representation and only including additional entities, processes or rules if it is absolutely necessary. An example would be an agent-based model in which the agents’ dispersal is fuelled solely by one mechanism: population growth (e.g., Romanowska 2015) or choosing the most attractive “next-step” (e.g., Scherjon 2013). One can subsequently add further elements, like including a friction map with topographic barriers to see if the shape of the dispersal wave changes (e.g., Steele et al. 1996), or differentiating the population growth depending on which environment the agents occupy. By implementing such incremental changes the modeller experiments with the increasing complexity of the simulation, trying to achieve the simplest yet accurate representation of the system in question.

On the other hand, the proponents of the KIDS methodology (Edmonds and Moss 2005; Edmonds and Meyer 2013) argue that it is unwise to discount the knowledge we already have of the system or judge what types of data should take primacy (e.g., qualitative, “anecdotal” accounts may be just as good a source of
inspiration as formal social theory). Thus, it may be better to try to model the phenomenon in question in as much detail as possible and then gradually strip the model of everything that turns out not to be significant. This approach removes the need for strong theoretical arguments for why certain elements were included in the model while others were left out, but it also assumes a high degree of knowledge of the system we try to model, which may limit its utility for archaeological applications.

The KISS rule is more commonly followed than the KIDS approach in simulation modelling because each additional line of code brings the possibility of introducing an unintentional error into the code and each extra algorithm makes the behaviour of the simulated system more difficult to track, understand and interpret. This is particularly true for models containing non-linearity, feedback loops, and stochasticity, which constitute the vast majority of all simulations concerned with human societies. The previously mentioned issue of “overdesigned” ABMs is directly correlated with the potential for code errors, intractability and difficulties related with describing and interpreting the results as well as the computational cost of exploring the parameter space of the model (discussed further on). This is a critical distinction between the KISS and KIDS approaches, and the main reason why KISS is more effective than KIDS in data-poor research areas. Unless there is an incredible high-quality knowledge of the
data and the process – a rare occurrence in archaeology – the KISS approach is likely to be more appropriate.

Regardless of whether the KISS or KIDS strategy is followed, the final shape of the model as presented in the final publication is always the result of a long series of experiments and modifications, where various elements and algorithms are continuously added to and removed from the simulation. It is common for modellers to be accused of leaving some elements out of their models that others argue to be key in the functioning of the simulated system or using rules of interaction that the critic disagrees with (Waldherr and Wijermans 2013). These arguments often stem from the misunderstanding of the scope and capacity of simulation techniques (see Chattoe-Brown 2013). Rather than trying to mimic the full complexity of the past or present reality, a simulation’s aim is to evaluate hypotheses, answer research questions or provide predictions only in terms of the implemented ontology and within the tested ranges of parameter values. Although in theory it may be possible to fully replicate an existing system, this might render the model useless since the more complex the model gets the higher the risk that the simulation contains errors or becomes as impenetrable as the original system. Like in a virtual laboratory, experimenting with different elements and their combinations, and testing them one or a few at a time is a valid scientific method (Moss 2008; Premo 2010).
When constructing dispersal models in archaeology the most commonly used entity is a human population. In the case of EBMs, human groups are treated as a homogenous population and usually described as a number representing the density of occupation. In CAs, the human population is represented as a Boolean value indicating whether the cell is “occupied” or “empty.” Finally, ABMs use agents as representations of individuals, households, human groups, or larger populations.

Almost universally, the second key entity is the environment over which the population disperses, usually represented in the form of a friction map. The environmental characteristics included in the friction map depend on the modelled system. For example, what is considered a fast-track corridor in the first “Out of Africa” dispersal may be an area of slower diffusion for Neolithic populations. Therefore, the environmental layer of the simulation may be any combination of topography (e.g., Nikitas and Nikita 2005), biomes (e.g., Hughes et al. 2007), soil quality (e.g., Ackland et al. 2007), proximity to rivers (e.g., Scherjon 2013) or even a social network if the model concerns cultural transmission.

The rules of interaction within and between population(s) as well as with the environment are usually grounded in a particular theoretical framework or academic consensus among the research community (Ahrweiler and Gilbert 2005), although occasionally attempts are made at testing a wider set of
mechanisms driving the studied phenomenon (e.g., Scherjon 2013). For example, the notion of heavy dependence of early hominins on the environment they inhabited, generally accepted among Palaeolithic researchers (see Holmes 2007), has been translated into strong environmental determinism in the models of Pleistocene dispersals (e.g., Mithen and Reed 2002; Romanowska 2015). Other commonly used mechanisms in modelling human dispersal include: population growth (e.g., Steele 2009), competition with an indigenous population (e.g., Isern and Fort 2012), and varying modes of transport (e.g., Fort et al. 2012).

**Coding and Testing (“Implementation” and “Verification”)**

Once the modelled entities and rules of interactions are defined and placed within the framework established in step 4, the next phase is to implement them as computer code. There are a number of simulation software platforms, designed to make this step easier, such as NetLogo, Repast or Swarm for ABMs (Railsback et al. 2006; Nikolai and Madey 2009), MATLAB (The MathWorks 2014) for EBMs and AnyLogic for combining EBMs and ABMs (AnyLogic 2014). Sometimes, modellers write their models from scratch in a general-purpose programming language such as Python, Java, C++, or R (Downey 2012).

There are two common pitfalls at this stage (Galán et al. 2013). The first error that can occur during the coding step is when the computer code does not
correctly represent the conceptual model. This may be the fault of the modeller who misrepresented the conceptual model or of the author of the conceptual model who did not formulate their model in an unambiguous manner. Many researchers build their models in an explicit manner using a formal mathematical description of their hypotheses, which facilitates the translation of conceptual models into simulations. However, perhaps largely due to the aforementioned lack of reciprocal feedback between domain specialists and modellers, examples of such mathematical definitions of hypotheses are rare in archaeology, leaving the step of formalising conceptual models in the hands of modellers, who may or may not correctly understand the intentions of the authors of the conceptual models (Waldherr and Wijermans 2013).

Translating conceptual models from natural language into computer code is one of the best methods for identifying the assumptions underlying the modelled theory, including those that are not immediately recognisable as being assumptions at all (Norling et al. 2013). For example, almost all simulations of the first “Out of Africa” dispersal start from East Africa (e.g., Mithen and Reed 2002) and most of the Neolithic diffusion models take Jericho (or its environs) as their point of origin (e.g., Fort, Pujol and Linden 2012). Although rarely questioned even by archaeologists, these are assumptions and it is useful to consider them as such. Similarly, although it is generally agreed that early hominins would spread faster over grassland-type biomes similar to the environment in which they
evolved, no conceptual model has ever ventured to specify how much faster? Was it 20% faster than over forested biomes? Or twice as fast? While conceptual modelling allows us to construct models with such underspecified elements, the need to translate them into computer code forces one to describe the researcher’s assumptions in a more formal manner. It is likely, that one of the more lasting legacies of the current increase of simulation case studies in archaeology will be a wider recognition of the advantages of formal modelling over building hypotheses in natural language.

The second common error at the coding stage is more difficult to recognise and occurs when the computer code does not do what the modeller intends it to do (David 2013; Galán et al. 2013). This may derive from simple errors such as a typing mistake or incorrectly placed brackets. Even more difficult to identify are misconceptions regarding the deeper levels of the code’s architecture, for example drawing random numbers from an inappropriate distribution or floating-point number rounding errors (Izquierdo and Polhill 2006; Banks and Chwif 2011). Therefore, the coding step is usually divided into two phases: code building (“implementation”), and consolidation/testing (“verification”) in which each part of the code is repeatedly checked and tested or “debugged” in computer science jargon (Norling et al. 2013). The latter phase is often as long or longer as the former, since the more thoroughly the code is tested,
the more confidence we can have that the results represent the true dynamics of
the simulation.

Since every simulation is in fact a form of software, commercial software
development techniques and best practices can be employed (Baxter et al. 2006;
Joppa et al. 2013; Rossiter 2015). A large number of techniques have been
developed in computer science for building reliable code, such as testing
frameworks, reimplementation of the code in another programming language, unit
testing, defensive programming, test-driven development, metamodelling
languages such as the Universal Modelling Language (UML), or CoSMoS Pattern
Language etc. (Ropella et al. 2002; Stepney 2012; Collier and Ozik 2013; Galán

Nevertheless, it is unrealistic to assume that, even if extensive testing was
undertaken and best practice followed, all simulations are error free.

EBMs are usually simpler to code than ABMs and CAs and, since they are
more mathematically explicit, less likely to contain errors. On the other hand,
ABMs are particularly prone to coding errors as the interactions and processes are
often stochastic and irregularities do not stand out as much as they would in
EBMs (Ropella et al. 2002). The experiences gained in other disciplines (Macy
and Sato 2010; Railsback and Grimm 2011; David 2013; Norling et al. 2013)
show that unintentional errors are far more common than one may wish them to
be. Hence the importance of another verification technique: replication (Wilensky and Rand 2007; Galán et al. 2009; 2013). Replication is a specific form of peer-review, in which a group of scholars, independent of those who published the model, build the model themselves to check whether they obtain the same results as the original implementation (Edmonds and Hales 2003; Wilensky and Rand 2007; David 2013). Worryingly, the number of replication studies of archaeological simulations is low, especially in real geographies, and not a single one of a dispersal case study has been performed so far (cf. Barton and Riel-Salvatore 2012; Premo 2012).

Choosing Parameter Ranges (“_parametrisation”)

Once the simulation is built and tested, one needs to decide on appropriate settings and ranges of values for the parameters. In EBMs these are the parameter values used in equations, in ABMs they describe the features of the agents and their environment, while in many CAs parameters are used to define probabilities of various events. In simulation models of human dispersal, common parameters include population growth rate, the size of the initial population, and the carrying capacity of different environments.

In theory one could test the full parameter space, that is, run the model with all possible values for all parameters and their combinations. In practice this
is not feasible even for relatively simple models\(^4\) and a more efficient strategy of setting at least some of them at a specific value and only testing specific ranges of values is deployed. The most common approach to finding such values is to derive them from empirical data. In archaeology where such data is not always readily available, one may use proxies representing the most parsimonious approximation of the real values (Hazelwood and Steele 2004; Robson and Wood 2008; Fort et al. 2012), for example, statistical aggregates of values derived from observations of contemporary groups of hunter-gatherers or sedentary agriculturalists (e.g., Binford 2001). Sometimes it is reasonable to significantly extend their range, for example, by doubling the maximum and minimum values, therefore ensuring that the “real” value is likely to be somewhere within the tested range even if it is impossible to pinpoint it exactly.

The results of the simulation are rarely equally sensitive to all tested parameters and the models are often surprisingly unresponsive to even a high variation of some of them. This can have different reasons: sometimes certain processes have a much stronger impact on the key elements of the system than others, while in other cases built-in feedback loops balance things out. For example, the rate of population growth under conditions of low competition typical for the initial phase of settling a new area is important to the system dynamics. However, once the population reaches its local carrying capacity, the value of the growth rate becomes negligible as the competition over resources is
the main determinant of the absolute size of the group. Thus, depending on the model and its implementation, varying the population growth rate may or may not change the results.

Modellers commonly use two techniques to explore the range of parameter values in a model (Richiardi et al. 2006; Evans et al. 2013; David 2013). First, to test the impact of different parameters on the final results, modellers perform a sensitivity analysis (e.g., Fasham 1995). In this procedure, each parameter is varied independently, for example by 10, 50 and 100%, to see if the results of the simulation change in a predictable fashion. By highlighting only those parameters that have a real impact on the results, the number of runs can be reduced, saving significant amounts of time and computing power. The second technique is calibration (Richiardi et al. 2006; Moss 2008). In simple terms this means checking under what parameter values the simulation will produce results consistent with a subset of data not used in the development of the simulation. This produces a benchmark, which allows testing only some parameters while keeping most of the values fixed. There are many statistical techniques for comparing the model output with data developed for calibration, from simple t-tests to more complex approaches such as maximum likelihood estimation or the method of simulated moments (Windrum et al. 2007; David 2013).
In simulation models of human dispersals, the most common approach to calibration is to select a few sites at which the arrival of humans or cultural markers is particularly well dated, and vary the parameters until the simulated dispersal front arrives at their locations at the time consistent with their dating (e.g., Mithen and Reeds 2002; Nikitas and Nikita 2005; Baggaley et al. 2012). Calibration is an important step in itself in model building, as it tests the underlying conceptual model: if the simulation fails to replicate the training data, it indicates that an important factor or mechanism was not accounted for in the simulation.

**Running the Simulation (“Experiment Design” or “DoE”)**

After coding and parameterizing the model, the modeller will design the experiments that should allow them to address the research questions specified in step one. The selection of the most appropriate strategy at this stage depends, again, on the nature of the research questions: are they concerned with how the system performs under a wide range of parameter values or is the researcher’s goal to evaluate conflicting hypotheses?

The first step is always to decide on the start and stop conditions. The start conditions, commonly termed “initialisation,” describe the state of the model at time zero. In the case of dispersal models this usually involves decisions on, for
example, where the dispersal will start from, the size and the composition of the original population, the initial values of any specific traits of the population or their distribution among agents and, if the simulation is pegged to any real data input that changes over time (e.g., the environmental record), what date will be used as the starting time of the dispersal. Equally the stop condition has to be determined in advance. The simulation may be halted after a certain number of steps, which may relate to concrete time units (years, decades etc.) or may be determined statistically (Chen 2012). Alternatively, conditional stops can be used, for example the arrival of humans at a specific island could be considered a conclusive step in models of the peopling of the Pacific Islands (e.g., Fitzpatrick and Callaghan 2013). In abstract models the time step usually does not correspond to any particular real time unit and there is no obvious stopping point, hence an arbitrary (usually very high) number of steps is used. In some cases the modeller waits until the simulation “converges” or attains “fixation,” which can be loosely translated as “nothing interesting happens any more.” For example, in a hypothetical simulation, which investigates the interaction between an original population and an incoming one, there would be little point in continuing the run once one of them dies out.

There are two main techniques for running the simulation in ways that will answer the research questions: the parameter sweep and the testing of alternatives (Hoad et al. 2014). In the first technique the modeller changes all parameter
values gradually one by one in order to understand the relationship between the parameters and the model behaviour (e.g., Premo 2006). This technique identifies dependencies such as, for example, the higher the population growth the faster the dispersal. It may also help to detect thresholds and feedback loops, for example, the higher the availability of resources the larger the population, as long as the consumption of resources does not exceed the population’s regeneration rate.

Another experimental design approach is to test alternative scenarios, usually representing conflicting hypotheses, by running the simulation with specific sets of parameter values that fundamentally change the model/behavioural structure (this is termed “theoretical plurality” as opposed to the “parametric variation of a single model” tested in the parameter sweep). For example, one can use different starting locations for the dispersal, or run the model with certain passages, like narrow water crossings or mountain passes, first blocked and then unblocked while keeping all the other parameters fixed (e.g., Romanowska 2015). Additionally, model selection is being increasingly used instead of classical null-hypothesis testing. The model selection approach uses statistical tools, derived from Bayesian statistics, which give a value denoting how well each model/hypothesis corresponds with the data (e.g., Crema et al. 2014).

The final question when running simulation models is how many times each parameterization of the model needs to be run. The answer to this question depends on the model being stochastic (like most ABMs and CAs) or
deterministic (like many EBMs). Deterministic models are run only once as repeating the run with the same parameter values would simply produce exactly the same results. In stochastic models some of the mechanisms are probabilistic, i.e., require drawing random numbers that will differ from one execution of the simulation to another. Therefore each scenario has to be re-run a number of times to find the average and the distribution of the results. Although there is no established minimum of how many iterations should be run, the number rarely drops below 20 runs. In general, the higher the variance the more runs should be performed, the exact number of which can be calculated using appropriate statistical techniques, for example convergence tests (Ritter et al. 2011).

Once the start button is pressed the simulation starts performing all the calculations coded-in as algorithms and the modeller can do little besides grabbing a coffee and patiently waiting for the results to start coming in.

**Analysing and Interpreting the Results**

The results of individual simulation runs rarely produce instant eureka moments, but instead mostly consist of long columns of numbers. Moreover, once these numbers are analysed and reveal interesting patterns it may not be immediately obvious which of the elements of the model caused the interesting results, highlighting the need for more work to tease out the full sequence of causality in
the model. The analysis and interpretation of the output of a simulation therefore usually consist of three stages: recognising patterns in the results, understanding what mechanisms caused them, and quantifying the uncertainties of the simulation (Evans et al. 2013).

In the first stage, often termed “exploratory data analysis,” modellers use a range of statistical and visualisation techniques to recognise patterns in the results (Kornhauser et al. 2009; Hoad et al. 2011; Evans et al. 2013; Dorin and Geard 2014). The outcome of each run is usually described as one (or a set) of summary statistics, for example in dispersal models, the timing of the first arrival at a particular location. However, the modeller could choose other types of “artificial data.” It is sometimes possible to define a simulation-specific index, which combines a number of outputs into one number and can be plotted over time to observe changes. Visual analysis of the results is often a useful approach, especially for recognising general patterns or so-called “stylised facts,” i.e., broad generalisations which are often qualitative in nature but nonetheless widely accepted as correct by the community. A good example of a “stylised fact” would be a particularly early arrival or a delay of the dispersal wave in a given region, e.g., speedy dispersal into Southeast Asia combined with late colonisation of Europe in the first Out of Africa dispersal (Mithen and Reed 2002) or a specific shape of the dispersal front, e.g., comparable with the known extent of one of the early Neolithic cultures (Ackland et al. 2007). In some models, especially if they
involve evolutionary mechanisms, the simulated cultural record (for example, density, spatial distribution or variability in cultural assemblages) is not described by summary statistics but instead sampled to mimic the time-averaging and other postdepositional biases that affect archaeological data (Madsen 2012; Premo 2014). The question of how to represent the archaeological consequences of the modelled processes in a way that is directly comparable with the archaeological record is an issue particularly relevant to archaeology. However, this topic has so far received little discussion in the archaeological simulation literature.

Once the interesting patterns are identified, the second stage is to understand the mechanisms that gave rise to them. In EBMs, the role of each factor is known as it appears in the equations, so the interpretation of the model’s results is usually straightforward (Bullock 2014). However, as the model departs from the analytical rigour of mathematics and strives for realistic representation of the studied real-world system, the results become increasingly opaque because “the behaviour of a simulation is not understandable by simple inspection, on the contrary, effort towards the results of a simulation must be expended, since there is no guarantee that what goes on in it is going to be obvious” (Di Paolo et al. 2000: p. 501). This has led researchers to call simulations opaque thought experiments. In some cases, particularly in the social sciences where simulations are built to inform policies, some of them become the infamous black boxes, as they are developed only to provide predictions (Epstein 2008). However, in the
explorative, heuristic models more common in archaeology the modeller has to tease out which particular mechanisms in the simulation caused the interesting patterns and reconceptualise it to see what insight this causality may give us into the real-world system (Gore and Reynolds 2008; 2010).

The final stage in the analysis and interpretation of simulation results is to deal with the problem of validity: how do we know that our simulation is a valid representation of the real system? This is a common concern in any type of modelling and one not taken lightly by the simulation community (e.g., Oreskes et al. 1994; Troitzsch 2004; Küppers and Lenhard 2005; Windrum et al. 2007; David 2013). The phase of the model development that deals with this issue is called “validation” and can be defined as: “an assessment of the extent to which [the simulation] M[odel] is a good representation of the (unknown) process that generated a set of observed data” (Windrum et al. 2007: 1.7). In all modelling exercises, both conceptual and computational, the validation is performed by comparing the model predictions (in the case of simulations, the model results) with the available empirical data. A good fit between the two indicates at least that this set of system entities, the rules governing their behaviour, and the relationships between them can lead to outcomes which match (to some given tolerance) the empirical data.
However, like with every model (that includes conceptual hypotheses), this is no proof that the absolute “truth” about the past has been revealed. Results coinciding with few data points may just be a coincidence and it is often not entirely clear how close to the patterns observed in real data the model results need to be in order to declare it “valid.” This issue is especially strong in cases where the aim is to replicate qualitative stylised facts. Also, even if the model matches a large number of data points, one can never be certain whether an alternative model of the same system would not be able to align with the data even better. This problem is known as “equifinality” (Premo 2010; Evans et al. 2013; Graham and Weingart 2015). For these reasons, the validity of any modelling work is closely interlinked with the amount and quality of data as well as the strength of conceptual models.

Some types of models only need to be verified (i.e., checked for any internal errors), as they are not designed to produce point predictions but rather elucidate general patterns and mechanisms. For example, subjunctive models test abstract “if-then” questions (e.g., Wren et al. 2014), feasibility models aim to demonstrate if something was or was not possible (e.g., Davies and Bickler 2015), whilst other types test the internal coherence (but not the correctness) of abstract conceptual models (e.g., Young 2002). In these cases researchers use only verification techniques: replication (discussed in section 6) or model alignment –
an alternative method of replication in which the simulation built using a different method or ontology is compared to the original study (Axtell et al. 1996).

In those models, which have to show that they correctly relate to the past reality, i.e., need to be “validated,” there are a few ways around the problem of uncertainty (Chwif et al. 2008; Sargent 2013). The most commonly used one, especially in disciplines researching modern societies, is to provide predictions that could be compared to future data. This is obviously not an option for archaeological case studies. Another approach is to compare the results to multiple independently collected sets of empirical data. For example, in the classic Artificial Anasazi model (Dean et al. 2000; Axtell et al. 2002) simulating the rise and decline of ancestral Pueblo peoples in the American Southwest, the results replicated two independent sets of data which were not included in the model: population fluctuations through time as well as the spatial distribution of sites. Although in theory one could argue that, again, it may be possible to construct a model that fits the data even more closely, the probability of that happening (especially when one considers that the model was built in an informed way) decreases significantly (Grimm 2005).

Although the inherent uncertainty of all scientific models is widely recognised, the process of producing new hypotheses and testing their validity against the available data using simulation techniques is a standard approach
across most scientific disciplines. There is therefore a strong notion of using simulation as a heuristic tool, which allows researchers to improve their models and “validate” them against the continuous stream of new data and changing academic consensus (Oreskes et al. 1994; Ahrweiler and Gilbert 2005; Windrum et al. 2007). In most cases the real value of a simulation study depends on a subjective answer to the question: was the model useful? The answer is “yes” if the model told us something new, challenged some deeply entrenched assumptions, showed surprising implications of a conceptual model, suggested a new way of approaching a research question or highlighted gaps in our knowledge.

In the archaeological case studies of ancient dispersals, the most common validation method is to compare the simulation results against the first arrival dates derived from archaeological data (e.g., Baggaley et al. 2012; Scherjon 2013), although replication of “stylised facts” was also used (Mithen and Reed 2002; Nikita and Nikitas 2005). In other cases, simulations investigated abstract mechanisms behind dispersals and therefore did not need to be validated against real world data sets. For example, the model by Wren et al. (2014) explored the evolutionary dependency between a specific cognitive ability of early hominins (spatial foresight) and their propensity to migrate. In a similar vein, Young (2002), tested different types of movement in an abstract space, concluding that even very simple behaviours of the simulated populations can result in large-scale
spatial patterns indistinguishable from the ones we observe in the archaeological record.

**Feeding Back into the Discipline**

After the results produced by a simulation are identified, checked and understood, it is common to share these findings with other researchers. Although the dissemination usually happens in a standardised form by presenting and publishing research papers, authors of simulation studies expect and hope that their simulations will be discussed by two very different sets of researchers: the domain specialists and the modellers. The goals of their publications are therefore also twofold: to inform the interested archaeologists of the findings of their research and to facilitate replication and reuse of the model.

To enable critical assessments and replication, modellers follow three standards: providing a detailed description of the model, following standardised protocols, and making the code freely available for peer review. First, the published model description often contains technical details, comprehensive explanations of the algorithms, and discussions of the decisions taken during model development. These technical descriptions may be off-putting to non-modellers, but they are necessary to fully understand the scope of the model and its possible weaknesses. Second, many models use a standardised protocol for
model description, for example the ODD (Overview, Design Concepts, and Details) for ABMs (Grimm et al. 2006; 2010). It allows researchers to describe their models using the same categories, thus facilitating the comparison between models and guiding readers through not only the technical details of the simulation but also the intentions of the authors. The ODD format has been commonly used to describe archaeological ABMs (e.g., Callegari et al. 2013; Scherjon 2013; Davies and Bickler 2015). Finally, to ensure reproducibility, modellers are urged to give full access to the source code of their simulations (Ince et al. 2012; Collberg et al. 2014) and are encouraged to have the code peer-reviewed and placed in online code repositories. There are a number of such model libraries, for example the CoMSES library (OpenABM 2014, e.g., Wren 2014) and GitHub (GitHub 2014). Models deposited in such libraries often use specific licenses that are explicitly designed for sharing the source code (e.g., the GNU General Public License). Thanks to these measures, replication studies as well as repurposing and reuse of existing models are becoming more commonplace.

Returning to Any of the Points 1-9

Simulation models differ from many conceptual hypotheses in their narrow focus on one, or a few aspects of the complex system they study. Although all modellers hope that their models are an accurate approximation of the past reality, they also
recognise that they give only selective insight into particular aspects of the studied system rather than solving the “whole problem.” Similarly, a published simulation model is never considered the “final word” but rather a virtual experiment within well-defined limits of specific ontology and the range of the tested parameter values. Therefore, the simulation model building enterprise is iterative in nature as there are always more factors that can be included and different parameter values that can be tested. Debugging, testing and discarding individual factors, whole segments and even full models are all steps in the cumulative process of gaining knowledge. However, each new iteration does not replace the previous versions of the model but builds upon them, reflecting a gradual progress towards a better, more accurate and more secure understanding of past systems (Figure 4)(Neiman 1995; David 2013; Norling et al. 2013). Thus, the final step in the model development cycle is almost always a return to any of the steps 1 to 9 (Sterman 2000).
Figure 4. An overview of the development and refinement of models, illustrating the iterative nature of the process. The conceptual model expressed in natural language is formalised and translated into algorithms and then computer code which, after parameterisation with available data, produces results. These results are subsequently used to improve and refine the original model.

A good example of this process is offered by the “SteppingOut” model (Mithen and Reed 2002), the first published simulation of the early hominin “Out of Africa” dispersal. After the publication of the initial implementation in 2002,
the model was challenged by another study (Nikitas and Nikita 2005),
subsequently enhanced with a better quality environmental data layer (Hughes et
al. 2007) and further expanded and used to test a number of hypotheses (Hughes

Discussion and Conclusion

In this paper I have described the process of developing an archaeological
simulation and illustrated it using examples of models of past human dispersals.
The range of simulation studies mentioned in the text shows that simulation is a
mature, versatile scientific tool, which can be successfully used to infer causal
mechanisms behind patterns observed in empirical data. However, as shown in
multiple examples its application is not a straightforward process, instead it
requires creativity, academic rigour and a high level of peer scrutiny. As
archaeological simulation is still a relatively young subdiscipline, it suffers from
the limited interaction and feedback between archaeologists and modellers. This
paper is only a step in the process of bridging this gap but by providing a compact
guide to the modelling process it hopes to encourage archaeologists to offer
critical, informed and constructive responses to previous and future published
models. In some way, the late adoption of simulation techniques in archaeology is
a fortunate delay, as many of the lessons learnt in other disciplines in the last two
decades and the resulting guidelines to best practice are available to researchers interested in simulations. However, some of the issues related to the particularities of archaeology remain. In the final few paragraphs I will briefly discuss the few areas in which more discussion between domain specialists and modellers could benefit both communities.

It is likely that archaeological models will become increasingly complex in the pursuit of more realistic representations of the complexities of past systems (see Chwif and Paul 2000), but equally there will always be space for elegant, abstract models illustrating how simple relationships may lead to complex patterns. In both cases there is scope for a wider debate between modellers and domain specialists on such topics as: what to simulate, what do we consider as the fundamental building blocks, where to derive parameter proxies from or how to validate models using archaeological data. Currently, most simulation studies concentrate on explaining population-level patterns or testing the feasibility of archaeologically attested events. Although the archaeological interest is often focused on individuals or groups, only a few models (e.g., Wren et al. 2014) work on this level, as a result not addressing the most topical archaeological questions (Moncel 2010), for example, what incite people to disperse? Which factors enable dispersal into new habitats? Or what are the characteristics of those who moved and those who stayed? Equally, the attempts to validate the model results with archaeological data have been limited so far, likely due to the complexity of the

Pre-print version. Visit digitalcommons.wayne.edu/humbiol after publication for the final version.
processes involved in the creation of the archaeological record, not encountered by researchers in other disciplines. However, regardless of its difficulty this issue needs to be addressed if simulation is to become a more common tool in archaeology.

As the number of voices advocating a wider use of simulation techniques in archaeology increases, there is also a growing awareness among archaeologists not using modelling techniques that the currently common analytical and theoretical methods simply cannot deal with the complexity of non-linear and multiscalar past interactions (Van der Leeuw 2004; Premo 2010; Barceló 2012; Barton 2014). This is also true for the study of human dispersals – a process comprising a multitude of factors, from local and global environmental conditions to social interactions between individuals and groups. Simulation techniques provide a method to tackle the challenge of understanding such past complexities and a tool for combining different sources of evidence (archaeological, environmental, genetic records etc.) into one coherent framework.
Figure 5. A schematic depiction of different confidence thresholds. Conceptual models expressed in natural language are the least secure. Formalising them allows one to identify all the assumptions and provide more transparency. Thanks to simulation they can be tested and validated against empirical data. However, the highest level of certainty is attainable only through repeated replication and testing against new data and new hypotheses.
However, the main advantage of formal methods including simulation is that they allow us to build our understanding in an explicit and cumulative manner. Instead of being swamped with different explanations, scenarios and theories about the past, whose strength depends on one’s personal beliefs and poorly defined common sense, it gives the community a tool to build an increasingly stronger framework where new ideas are thoroughly tested before they are incorporated while new data is used in the continuous process of validation (Neiman 1995; Barceló 2012; Barton 2014). The iterative nature of the modelling cycle constitutes part of a larger scientific process of increasing confidence in the models of the past (Figure 5). Formalising the conceptual models pushes them above a higher certainty threshold, where entities and the rules governing their behaviours are well defined and can be examined and discussed in detail by the research community with a much higher degree of clarity and transparency. The next step, a simulation, allows us to test whether the model is logically coherent and plausible as well as conforming to the existing evidence. It is not the final step though, as it is only by replicating simulation studies, constructing libraries of tested models and reusing them, and continuously challenging the models with new data and new hypotheses that a high level of certainty can be obtained (David 2013). Given the current hype around simulation models in archaeology (Lake 2014), and the relative scarcity of replication studies, we may expect a turbulent but necessary period of questioning.
the existing models to follow soon. However frustrating, finding out that a model is wrong is a useful exercise. Paraphrasing Carl Sagan (1990, 33:20) there are many models in science, which are wrong. That’s perfectly alright: it’s the aperture to finding out what’s right.

A closer collaboration between modellers and domain specialists has the potential to move the focus of future models closer to issues at the heart of archaeological debates. Better communication between the two groups will mean closer and more comprehensive scrutiny of the published models and their alignment with the current research consensus. The potential of the technique in becoming the “epistemological engine of our time” (Küppers et al. 2006) is extremely high as simulation provides archaeologists with a much needed “virtual lab” or “tool to think with” for testing their ideas, even the seemingly implausible ones, and combining the swathes of data with conceptual models in a critical way. However, the specific theoretical implications of the epistemology of simulation modelling in archaeology (cf. Frank and Troitzsch 2005; Chattoe-Brown 2013; Tolk et al. 2013) still need to be thoroughly discussed within the discipline. It is only when simulation enters the mainstream of archaeological practice and university curricula that the full potential of the technique can be realised and better adapted to the specifics of our discipline.
Notes

1. System dynamics is a type of Equation-based modeling, so the terms are not fully synonymous.

2. It is customary to use the “EBM” acronym to denote Equation-based modelling and the “EBMs” acronym for Equation-based models. The same applies to ABM/ABMs and CA/CAs.

3. Although some researchers recognize subtle differences between “individual-based,” “agent-based” and “multiagent” simulations, equally often the terms are used interchangeably (e.g., Railsback and Grimm 2011). The first is more common in ecology, the second in social sciences and the third in robotics and computer science.

4. With each new parameter the number of simulation runs increases exponentially. Imagine an unusually simple simulation with only 2 parameters—if their range is tested in five places (for example: -10, -5, 0, 5, 10) 25 runs need to be performed. Adding a third parameter would require 125 runs. If the simulation is stochastic (i.e., includes probabilities and randomly drawn numbers) each of these runs has to be repeated at least 20 times, which means the simulation would need to be run 2500 times in total.

5. A closely related issue, the problem of identifiability, is that without testing the full parameter range it is impossible to say whether a simulation output that compares favourably with empirical data could not have been achieved with a
different combination of parameter values not tested in this particular experiment (Evans et al. 2013).

**Literature Cited**


http://jasss.soc.surrey.ac.uk/8/4/1.html.


doi:10.1016/j.ecolmodel.2010.08.019.


Pre-print version. Visit digitalcommons.wayne.edu/humbiol after publication for the final version.


Pre-print version. Visit digitalcommons.wayne.edu/humbiol after publication for the final version.

