RESEARCH ARTICLE

Agent-Based Modeling, Scientific Reproducibility, and Taphonomy: A Successful Model Implementation Case Study

Molly Carney* and Benjamin Davies†

There is a growing use of agent-based model (ABM) simulations to reconstruct past human-environment interactions. ABMs are useful in that they offer scientists the opportunity to model processes and study phenomena and systems that may not be otherwise reproducible or testable. Replication or re-implementation studies of ABMs are, however, infrequently undertaken, and there are few examples within archaeology or other social sciences. This paper documents the process of a successful ABM replication study, as well as two additional modifications to the original model. Results corroborate the findings of the original geoarchaeological model and indicate that episodic geomorphic events significantly affect archaeological deposit formation and the inferences drawn from associated radiocarbon records. One revision of the model further demonstrates that episodic fluvial events can create highly varied radiocarbon distributions. The second modification illustrates that excavation data helps to fill in hiatuses in radiocarbon chronologies on depositional landforms, although there is no effect across landscapes subject to erosion. This successful replication exercise also illustrates the value of Open Access data and analyses in reproducing, testing, and expanding upon archaeological research and theory building.

Keywords: Agent-based modeling; simulation; replication; geoarchaeology; radiocarbon dating

1. Introduction

Reproducibility is considered a hallmark of the scientific method (e.g., Latour and Woolgar 2013; Moonesinghe, Khoury and Janssens 2007; Popper 2005; Simons 2014). It is general practice for scientists to publish the details of experiments and scientific research, with explicit documentation of the methods used, to enable future researchers to replicate and expand upon prior experiments. If a finding cannot be replicated, then current understanding of the study system in question and its variables or our methods of testing may be insufficient. Within archaeology, however, the process of studying the archaeological record is itself destructive, making reproducibility within research designs near impossible. Agent-based models (ABMs) are one tool which allow archaeologists to create representations of past reality and manipulate variables within a simplified system that can be replicated and falsified.

ABMs are a class of computational models that simulate the behavior and actions of autonomous agents (whether individuals, groups, natural phenomena, or other units

of interest) to investigate emergent phenomena within complex systems (e.g., Macal 2017; Railsback and Grimm 2012). ABMs are object-oriented computational models where individuals or agents are unique entities that interact locally with each other and/or their environment. Rather than describing overall, global phenomena, ABMs simulate system dynamics that arise from how the system's individual components interact with and respond to each other and their environment over time. This bottom-up, generative, and historically contingent exploration of emergent patterns is the most important feature of ABMs (Epstein 2008). Thus, ABMs are particularly suitable for the analysis of complex adaptive systems and emergent phenomena in social sciences, biology, paleoecology, taphonomy, and many other disciplines. While more frequently employed in evolutionary anthropology, human ecology, and complexity science, ABMs offer archaeologists and social scientists the opportunity to study systems which may be impossible to replicate or may no longer exist (Premo 2006a; Premo 2006b) or as a space in which to test archaeological data (Thiele, Kurth and Grimm 2014).

Agent-based modeling also offers new ways of ensuring research is reproducible and falsifiable. Model replication or re-implementation as a form of model validation is critical component in the modelling process and is encouraged in almost every ABM guide (Axtell et al. 1996; Grimm

* Washington State University, US

† University of Utah, US

Corresponding author: Molly Carney (molly.carney@wsu.edu)

et al. 2006, 2010; Macal 2017: 255; Popper and Pichler 2015; Railsback and Grim 2012: 295–296; Romanowska et al. 2019; Wilensky and Rand 2007, 2015). Among fields in which reproducing study systems is either impossible or unethical, ABMs offer a means of simulating or representing real-world phenomena. However, there are only a few published cases in which archaeological models are replicated (Janssen 2009; Kanters 2018; Oestmo, Janssen and Marean 2016; Pop 2016), and few if any attempts that use the same programming language or environment.

This study addresses Wilensky and Rand's (2007) and Grimm et al.'s (2006, 2010) propositions that models should be replicated. Our study re-created a geoarchaeological model of human-environment interactions, and the taphonomic processes which influence the archaeological record. Davies, Holdaway and Fanning (2016) created an ABM (HMODEL) in NetLogo (Wilensky 1999) to simulate the potential for recovering dateable radiocarbon material within arid, fluvial environments. In the present study, the importance and methods of model replication are discussed before summarizing the work completed by Davies, Holdaway and Fanning (2016). A replication of HMODEL is then described, including difficulties and successes encountered, and modifications to HMODEL are made to address additional research questions. The paper concludes with a discussion of the utility of replication for methodological testing and its potential for theory generation.

2. Reproducibility and Replication Background

Archaeologists are well-aware that archaeological excavation irrevocably changes the archaeological record (Lucas 2001). Given this property of archaeological research and recent discussions on the overarching reproducibility crisis in science (e.g., Baker 2016; Fanelli 2009; Fitzpatrick 2019), agent-based models (ABMs) offer the social sciences one way of maintaining that core principle of reproducibility. As Cegielski and Rogers (2016: 285) note, reproducible archaeological research can be quite difficult; ABMs with proper documentation provide archaeologists with a tool for evaluating the strength of their inferential narratives and move past "black box" critiques (Huggett 2004, see also Epstein 2008).

There are several standards for model documentation and reproduction. The ODD (Overview, Design Concepts, and Details) protocol was designed to standardize and structure published descriptions of ABMs (Grimm et al. 2006, 2010; Polhill et al. 2008). The ODD protocol, most commonly used in archaeological models, was designed to provide a clear framework to increase model understanding, utility, and replicability. Another method of model documentation, less common within archaeology, is the Unified Modeling Language (UML). UML documentation is comprised of a set of standardized diagrams resembling a flow chart that graphically represents the model processes through time (Bersini 2012). Others push for the publication of code and documentation in open-source database repositories as a means of scientific transparency (Müller et al. 2014; Rollins et al. 2014). In archaeological ABMs, the use of a documentation method and open-source archiving of computer code is designed to

increase the transparency of the process and allow others to evaluate and construct the same experiments (Cegielski and Rogers 2016).

There are two main methods of verifying ABMs by model reproduction. Following Wilensky and Rand (2015: 337–341), replication or re-implementation refers to the re-creation and execution of a model by one scientist or group of scientists (the replicated model) that has been previously described and already implemented (original model). Model replication verifies the functions and results by exactly duplicating the original model. Many ABM modelers encourage model replication as part of evaluating the "correctness" of a model. If two distinct implementations of a model can produce the same results, as modelers we can be confident that the implemented model accurately reflects some phenomena of interest. However, if differences are discovered, the replicated model may need to be adjusted or the original model may need to be revisited. Model alignment is an alternative method of replication in which the simulation is built using a different method, often a different coding language, than the original study (Axtell et al. 1996; Edmonds and Hales 2003). In a number of model reproduction exercises, researchers find that premises within the conceptual model or differences in programming decisions yield different results when attempting to replicate a model (Donkin et al. 2017; Wilensky and Rand 2007). Our study focuses primarily upon a model replication using the same coding language (NetLogo); we use the terms replication or re-implementation interchangeably to refer to the former reproduction process.

In this replication process, a new implementation of an original model should be compared to the previous model's results to evaluate the outcome. Axtell et al. (1996) and Wilensky and Rand (2007, 2015) suggest determining the success of the replication by criterion or replication standards. Axtell et al. (1996) developed three categories of replication standards for a replication experiment. The first category, numerical identity, requires showing that the original and replicated models produce the exact same numerical results. With the second standard, distributional equivalence, the two implemented models should be statistically indistinguishable to each other. The goal of the third category, relational alignment, is to show the results of the two implemented models have qualitatively similar relationships between input and output variables. Wilensky and Rand (2007) note that numerical identity is likely impossible due to the complexities of implementing simulations across different hardware and software platforms. Often, distributional equivalence is chosen as the replication standard in replication exercises (Miodownik, Cartrite and Bhavnani 2010). Relational alignment is perhaps the "weakest" of the three standards, yet may be appropriate to illustrate relationships that may not have robust statistical significance or when quantitative data is not available in certain situations. In this study, we rely on relational alignment as our primary replication criterion as the original model published by Davies, Holdaway and Fanning (2016) primarily provided graphical, qualitative results.

In recent years there have been an increasing number of attempts at agent-based model replication (Miodownik, Cartrite and Bhavnani 2010; Wilensky and Rand 2007; Pop

2016; Oestmo, Janssen and Marean 2016; Donkin et al. 2017; Janssen 2009; Kanters 2018). Many replication studies, however, are either implementations of conceptual models (Edmonds and Hales 2003) or are actually re-alignments of models coded in different languages (Bigbee, Cioffi-Revilla and Luke 2005; Donkin et al. 2017; Galán and Izquierdo 2005; Miodownik, Cartrite and Bhavnani 2010). Archaeologically, Brantingham's (2003) simple model of stone tool procurement has been replicated several times, although all exercises were undertaken either to critique the original or as a teaching tool (Crabtree et al. 2019; Davies et al. 2019; Oestmo, Janssen and Marean 2016; Pop 2016; Romanowska et al. 2019). With reference to ABMs and NetLogo, only two replication studies come to mind (Janssen 2009; Kanters 2018), and both studies were re-alignments rather than true replications.

As more ABMs are published and used to explore behavioral emergence within complex systems, researchers are increasingly aware that ABMs are difficult to replicate and that small differences in coding, explanations of model implementation, and translations between coding languages yield significant differences in numerical, distributional, and relational model equivalence (Fitzpatrick 2019). These findings and discourse illustrate the need for additional replication studies of archaeological and landscape-based simulations, a situation which Romanowska (2015: 179) calls "worrying."

Replication and reproducibility are widely perceived as beneficial for science (e.g. Marwick et al. 2017; McNutt 2014; Munafò et al. 2017; Nosek et al. 2015), but rarely positioned as having much role beyond quality control. This perception of replication as strictly a methodological check on existing research can discourage replication studies in favor of novel research designs. Nosek, Spies and Motyl (2012: 616–617), for example, argue that there is a disconnect between individual incentives for "publishability" of research outcomes and the global benefits of publishing replication studies and negative results, discouraging the latter in favor of the former (see also Martin and Clarke 2017). Yet replication, particularly replication of a computer simulation from a description, is akin to building a model of a model. Making (or remaking) a model in the medium used for its construction forces the builder (or replicator) to confront decisions regarding the inclusion or exclusion of particular elements in a model which are made based on assumptions about the relative importance of those elements to the nature of the entity in question (Giere 1988: 77). Therefore many of the heuristic benefits of explicit modeling (e.g., discovering new questions, suggesting analogies; see Epstein 2008; Miller and Page 2009: 78–89 for more comprehensive lists) can be likewise ascribed to a replication exercise. While replications are indeed tests of existing ideas, they can also provide pathways to theory building in similar ways to building models from scratch. This notion is reinforced in the replication exercise described below.

3. HMODEL Revisited

3.1. HMODEL Summary

The original model by Davies, Holdaway and Fanning (2016) was designed as an exploratory agent-based sim-

ulation to examine the effects of episodic erosion and deposition on archaeological palimpsest deposits. Davies, Holdaway and Fanning (2016) specifically sought to model the temporal distribution of chronometric data, including radiocarbon dates and optically stimulated luminescence dates, collected from surface hearth deposits in arid western New South Wales, Australia. Archaeologists commonly use radiocarbon dates to reconstruct population densities and demographic change across time (e.g., Rick 1987; Williams 2012). By creating hypothetical chronometric distributions, the authors sought to test the interplay between human behavior, environmental change, and taphonomic processes across a dynamic landscape. These simulated chronometric distributions were also used to explore the validity of chronologic pauses or perceived gaps in the archaeological record and test for natural or cultural influences on those pauses.

The original model (HMODEL) was implemented in NetLogo v.5.05 (Wilensky 1999). In the model, a cell-based, gridded world contains dated sedimentary layers. Over the course of the model's run, additional layers are either deposited or eroded by the model's processes. In this landscape, hunter-gatherer bands maintain a steady population, move randomly throughout the world, and create five hearths for every "year" (also known as a tick or a time step) in the model. These hearths are visible upon the surface and contain a date equivalent to the year they were constructed; when a geomorphic event occurs, there is a probability that surface hearths will be either buried or eroded. At the end of the simulation, 100 surface hearth dates are randomly chosen to mimic pedestrian survey collection of radiocarbon and luminescence material. For each parameter combination, the simulation was run 1000 times to create a hypothetical radiocarbon date distribution.

The authors ultimately argued that geomorphic processes affecting site preservation and visibility have the potential to change a static radiocarbon record into one biased towards the present, with periodic chronological gaps. By comparing results from HMODEL to the surficial radiocarbon record of New South Wales the authors also argued that patterns in the chronometric data from the study area are more consistent with episodic geomorphic change than interpretations of demographic change. Their findings not only illustrate complications associated with using chronometric material to estimate past demographics, but have been used to highlight other taphonomic processes which may affect our interpretations of antiquity (Crema, Bevan and Shennan 2017; Dibble et al. 2017; Douglass et al. 2018; Perry et al. 2016).

3.2. Replicating HMODEL

As HMODEL is a simple model of geoarchaeological landscape processes with tacit archaeological significance, it is a relatively straightforward case study for model replication. Being unaware of any ABM replication attempts which use the same programming language or environment, this study specifically sought to recreate HMODEL using the same coding language (NetLogo). The replication was undertaken by one of us (Carney), who contacted the other (Davies) who shared

the ODD protocol. The attempt to recreate HMODEL relied entirely on the ODD protocol, brief correspondence with the original author, and the results published in Davies, Holdaway and Fanning (2016). These resources were used to interpret HMODEL's initialization, process, and scheduling (**Figure 1**), and inform coding decisions and model replication.

All reproduced scientific research needs some standard against which to compare the replicated results. Axtell et al.'s (1996) third criterion for replication standards was used to assess model equivalence between HMODEL and this re-implementation. Relational alignment seeks to show qualitatively similar relationships between input and output variables between models (Axtell et al. 1996: 135; Wilensky and Rand 2015: 340). As the exploratory work conducted by Davies, Holdaway and Fanning (2016) did not include statistical data against which to compare, this study specifically sought to replicate one figure: the initial exploratory HMODEL output (Davies, Holdaway and Fanning 2016: 456, Figure 5). This replication standard

was chosen as a simple graphical means of measuring the qualitative equivalence of the basic model and dictated many of the modeling decisions outlined below.

In this reimplementation, which mirrors the original model, the simulation begins with an empty world of 32×32 cells in toroidal space. Each cell in the gridded world contains an ordered list of numbers representing the ages of sedimentary layers in years before present. Invisible agents (humans) move across the landscape, creating five hearths with dateable material at a random location every year in the simulation (represented by one time step, or tick in the model). These hearths contain a radiocarbon age in years before present, which is also stored as a list associated with a patch. Any hearth that has a radiocarbon age younger than or equal to the age of the most recent sedimentary layer is considered visible on the surface, while any that are older are hidden as part of a subsurface deposit.

At given intervals, a scheduled event occurs. This *Event Interval* parameter determines the frequency of

Figure 1: Flow chart illustrating the scheduling and processes in the re-implemented HMODEL, as interpreted from Davies' (2018) original ODD.

geomorphic events and catastrophic flooding on the landscape. The *Stability* parameter determined the probability that each patch will undergo a geomorphic event during an *Event Interval*. The chances of an event occurring are randomly drawn from a Bernoulli process; if a patch does experience an event, one of two effects are determined by the *Erosion/Deposition Proportion*, also randomly drawn from a Bernoulli process. If *Erosion* occurs, the most recent layer of sediment is removed, and any hearths visible on the surface are destroyed (simulating the dispersal of charcoal and hearth material). If *Deposition* occurs, a layer of sediment is added to the sedimentary layers list and any hearths currently visible on the surface are buried. During subsequent erosional events, hidden hearths may be re-exposed through the erosion of overlying sediments.

As the authors were interested in the archaeological record of late Holocene Australia, all simulations were run at a yearly time step from 2000 years BP to the present, with hearth building ceasing at 200 years before present

to account for a decline in the use of hearths among aboriginal populations after European contact. At the end of the simulation, 100 dates are randomly sampled from all surface hearths and indexed from oldest to youngest.

Each simulation varied either the *Stability* or the *Event Interval* parameters and was run 1000 times, following Davies, Holdaway and Fanning (2016: 455–456). All generated data were analyzed and graphically plotted in R 3.6.1. NetLogo files, ODD, code, and all generated .csv files used in this analysis are available online (Carney 2020).

4. Results and Discussion

4.1. HMODEL Reimplementation Results

As an initial test of the replication exercise, the *Erosion/Deposition* simulations by Davies, Holdaway and Fanning (2016: 457, Figure 5) were chosen, in which landscape *Stability* was held constant at 0.0 (0%) while the *Erosion/Deposition* parameter was varied by increments of 0.2 (20%) (Figure 2). Davies, Holdaway and Fanning

Figure 2: Hypothetical radiocarbon distributions from unstable landforms (*Stability* = 0.0). The x-axis represents radiocarbon dates in years BP, while y-axis represents indexed values. The dashed line indicates an unbiased, continual occupation. **a)** Reprinted Figure 5 (Davies, Holdaway and Fanning 2016: 456). **b)** Re-implemented results.

(2016: 456) explained that in their model, an unaltered and unbiased radiocarbon record would be visually represented as a diagonal line 200 years before present at bottom left to 2000 years before present at top right. They interpreted curves falling to the left of this line as biased towards the present, while curves falling to the right were interpreted as being biased towards the past. They interpreted their results for this simulation as clearly biased towards the present.

The re-implementation exercise was partially successful; the replicated model was successfully able to recreate half of the graphs (**Figure 2b**; see section 4.3). Notably, the re-implemented model produced all five of the 200-year *Event Interval* graphs, three of the 100-year graphs, and only the 50-year *Event Interval* at the 0.5 *Erosion/Deposition* proportion. This re-implementation of HMODEL did not always produce 100 surface hearths in certain scenarios, which are blank boxes in **Figure 2b**. Even successful simulations did not produce 100 surface hearths in all 1000 runs.

Despite these discrepancies, discussed in more detail below, the overall exercise was still successful. Half of the graphs in Figure 5 were replicated, and of those graphs, all are qualitatively similar and illustrate similar trends. On landscapes subject to drastic geomorphic events (*Stability* parameter = 0.0), the archaeological record and available radiocarbon data is affected. In these scenarios the combination of *Event Interval* and *Erosion/Deposition* parameters simulate the higher magnitude geomorphic events which completely alter landscapes either via sediment transport or deposition. The hypothetical radiocarbon distributions produced by the implemented model illustrate

significant gaps comparable to those produced by Davies, Holdaway and Fanning (2016). **Figure 3** illustrates boxplots of the generated radiocarbon data and substantiates the inferences drawn from the biplots in **Figure 2b**, as well as illustrates significant bias towards the present on highly erosional or highly depositional landforms. In places with such unstable landforms and geomorphic events, radiocarbon material collected from pedestrian survey is clearly biased toward the present day and is not reflective of other human behaviors, population migrations, or other demographic trends.

4.2. Expanding HMODEL

The above exercise demonstrated that HMODEL's results are repeatable and provides additional support to the conclusions Davies, Holdaway and Fanning (2016) drew. However, model replication is not only useful in assessing the validity of a model and the processes which it attempts to simulate. Wilensky and Rand (2015: 340) argue that a "side benefit" of a replication is the opportunity to use the replicated model to advance separate research agendas and further explore the phenomenon beyond the original publication. This is the primary context in which archaeological models are replicated (Oestmo, Janssen and Marean 2016; Pop 2016). In this study, our HMODEL re-implementation was built upon in two separate ways: one experiment expands upon the results of **Figure 2** by introducing excavation data to the model output. In a second experiment, the *Event Interval* parameter is changed from a set interval to a probability, to more accurately reflect real world sequences of fluvial processes. Results from both experiments are outlined below.

Figure 3: Boxplots illustrating data range for the simulated radiocarbon distributions produced by the re-implemented HMODEL. The x-axis represents the experiment; the EI acronym stands for the value for the *Event Interval* while ED denotes the value of the *Erosion/Deposition* proportion. The y-axis illustrates the simulated radiocarbon date values.

4.2.1. Introducing Excavations

To explore if and how subsurface excavation affects HMODEL's outputs, additional experiments were re-implemented using HMODEL. Small-scale archaeological excavations were simulated to see if the addition of excavation data affects the simulated radiocarbon distributions. A new parameter, *Collection Method*, was introduced to account for pedestrian survey or the addition of excavated material. In the excavation scenarios, either five or ten "test units" are randomly placed on the landscape. At the end of each simulation run, "archaeologists" collect data from both surface deposits and "excavate" an additional four sedimentary layers in either five or ten test units (i.e., five or ten cells in the simulation are randomly selected and an additional four sedimentary layers sam-

pled). Any dates recovered from these test units are added to the total dateable surface hearths population, and 100 dates are randomly drawn from this population, indexed, and plotted to create the radiocarbon distribution at the end of the run. In these simulations, the *Stability* parameter was held constant at 0.0 while *Collection Method*, *Event Interval*, and *Erosion/Deposition* parameters were varied. Each scenario was run 1000 times.

Figure 4a and **4b** illustrate the results of the excavations experiments. While the addition of excavation data helps to fill in gaps created by geomorphic events, they do not help to mitigate the inherent bias in radiocarbon dates toward the present. Additionally, erosion and deposition play a greater role in the chronometric record than when looking at primarily survey data. In erosional

Figure 4: Results of the Excavations experiment on unstable landforms (*Stability* = 0.0). The x-axis represents radiocarbon dates in years BP, while y-axis represents indexed values. The dashed line indicates an unbiased, continual occupation. **a)** Excavation-5 with five random test units across the landscape; **b)** Excavation-10 with ten random test units.

environments (i.e., *Erosion/Deposition* parameter is >0.5), archaeologists will be less likely to subsample radiocarbon dates to fill in the gaps created by geomorphic events (*Event Interval*). However, on depositional landforms (i.e., *Erosion/Deposition* parameter ≤ 0.5), there is a likelihood that radiocarbon dates will more closely approximate continuous, uninterrupted landscape use.

These qualitative relationships also have quantitative significance. **Table 1** illustrates the results of the F-test for equality of variances and the Student's t-test for sample population independence. The results from each excavation scenario was tested against results from the equivalent HMODEL reimplementation scenario. For both the Excavation-5 and Excavation-10 experiments, five out of nine scenarios produced sample populations that were not significantly different than the equivalent survey scenarios. These runs then were distributionally equivalent per Axtell et al.'s (1996) replication criterion. Of note, however, is that across all the experiment scenarios, *Erosion/Deposition* parameters of 0.1 and 0.3 (depositional simulations) produced statistically significantly ($\alpha = 0.05$) different radiocarbon distributions. Excavations placed upon depositional environments yield different radiocarbon distributions than survey-based radiocarbon date collection and may be more likely to yield continuous chronometric (and other) cultural materials. Thus, the addition of excavation data in these depositional environments helps to create a more complete record of

human occupation throughout time, if still biased toward the present. Across erosional landforms, however, excavation data does not significantly alter the hypothetical radiocarbon distributions, and these hypothetical samples are not different than the data generated by the survey experiment alone.

4.2.2. Events as a Probability

All models are simplifications of real-world phenomena, and the *Event Interval* parameter and *Excavations* experiment substantially abridged real-world geomorphic fluvial processes. Fluvial events do not occur at set intervals but are stochastic processes. The Annual Exceedance Probability (AEP) is a probability measure which estimates the likelihood of occurrence of a flood of a given size or larger occurring in any one year (Flynn, Kirby and Hummel 2006). For example, fluvial events are colloquially referred to as 100-year-floods. Statistically, however, these events are represented as a probability or percentage (i.e., a 100-year-flood has a 1% probability).

As a second expansion upon our HMODEL replication, the *Event Interval* parameter was edited to simulate the probability that an event occurs, rather than have events occur at set intervals during the simulation. At every tick or time step, a random number is drawn from a Bernoulli process. If that number is less than or equal to the *Event Interval* probability parameter, a geomorphic event occurs across the simulated landscape. All patches, or cells,

Table 1: F-test for equality of variance and Student's t-test results for Excavation experiments and re-implemented HMODEL survey data. Where F-tests indicate significant variance within sample populations at the 0.01 α level, t-tests were not conducted (Shennan 1997: 88–89).

Collection Method	F	p	t	P	Collection Method	Event Interval	Erosion/Deposition Proportion
Excavation 5	1.0292	0.7921	0.2883	0.7731	Survey	50	0.5
Excavation 5	1.0233	0.7677	2.4659	0.0138	Survey	100	0.3
Excavation 5	1.0383	0.5535	1.524	0.1277	Survey	100	0.5
Excavation 5	0.9894	0.8901	-0.9620	0.9234	Survey	100	0.7
Excavation 5	1.2399	0.0029	—	—	Survey	200	0.1
Excavation 5	0.9122	0.1985	7.6429	3.83E-14	Survey	200	0.3
Excavation 5	1.0733	0.2644	2.6217	0.0088	Survey	200	0.5
Excavation 5	0.9796	0.745	1.1742	0.2405	Survey	200	0.7
Excavation 5	1.0146	0.8405	0.2765	0.7822	Survey	200	0.9
Excavation 10	1.0185	0.867	0.7789	0.4363	Survey	50	0.5
Excavation 10	0.9968	0.9675	4.122	3.99E-05	Survey	100	0.3
Excavation 10	1.0368	0.5687	1.9929	0.464	Survey	100	0.5
Excavation 10	0.9859	0.8538	-0.2878	0.7652	Survey	100	0.7
Excavation 10	1.4017	2.97E-06	—	—	Survey	200	0.1
Excavation 10	1.1211	0.07128	10.179	2.20E-16	Survey	200	0.3
Excavation 10	0.9481	0.4001	2.5271	0.0116	Survey	200	0.5
Excavation 10	0.9829	0.785	1.4407	0.1498	Survey	200	0.7
Excavation 10	0.9905	0.8943	0.1950	0.8454	Survey	200	0.9

in the HMODEL simulated world then have a chance of undergoing an erosional or depositional event, with those probabilities set by the *Stability* and *Erosion/Deposition* parameters. This experiment iteration then followed the re-implemented HMODEL process scheduling for the remainder of the run (**Figure 1**).

Like the previous simulations, the probability of a 50, 100, or 200-year-flood on landforms subject to varying proportions of erosion or deposition were simulated. All

other parameters were held constant, and each of the 9 scenarios was run 1000 times.

Figures 5 and 6 illustrate results of the Event Probability experiments. Simulating the probability of a geomorphic event creates a much wider radiocarbon distribution with a large data spread. While the breaks or pauses visible within the record are not present in this experiment, the overall hypothetical radiocarbon date distribution is much more variable than those produced by the original model.

Figure 5: Results of the Probability experiment, in which Events are determined by a probability rather than a set interval. The x-axis represents radiocarbon dates in years BP, while y-axis represents indexed values. The red dashed line indicates an unbiased, continual occupation.

Figure 6: Boxplots of the Probability experiment. The x-axis represents the experiment; the EI acronym represents the value for the *Event Interval* while ED denotes the value of the *Erosion/Deposition* proportion. The y-axis illustrates the simulated radiocarbon date values. Note larger data range when compared to Figure 3.

Simulated radiocarbon distributions are still biased toward the present. Individual runs still illustrate gaps or pauses from erosional events, and thus these results confirm the interpretations of Davies, Holdaway and Fanning (2016). In addition, data generated by the probability experiment (.csv files available in Carney 2020) were not normally distributed, and when compared to the original HMODEL runs, had unequal variances (**Table 2**). The boxplots in **Figure 6** illustrate the considerable differences in data ranges across the experiments. This small modification, which more accurately reflects real-world visibility and preservation of surface archaeological material, illustrates that chronometric proxies derived from pedestrian survey alone may produce widely varied radiocarbon dates.

This result and the excavations experiment further substantiate calls for caution in using radiocarbon data in summed probability distributions in regions with differentially preserved landforms or depositional processes (Hiscock and Attenbrow 2016). These results also corroborate Contreras and Meadows' (2014) radiocarbon distribution simulation experiments and observations that patterns drawn from radiocarbon data may result from a variety of factors other than population fluctuations.

4.3. Digging into discrepancies

The outcomes described above are consistent with the logic of the model and expand upon it, but discrepancies remain between the replicated model and the original. Comparing simulation results archived from the original study (Davies, Holdaway and Fanning 2016; Davies 2018) with those produced in the present study lead to the conclusion that the difference lay not in the implementation of the reproduced model, but the sampling mechanism used to produce the plots. In the replication, 100 hearths were sampled after every run; however, as demonstrated above, some parameter configurations do not produce 100 surface hearths in single instances, especially under highly erosive or deposition conditions. This prevented the generation of plots for these configurations. In the original study, *all hearths* from the surface were recorded from *every run* of a parameter configuration, and this larger dataset was then resampled in lots of 100 to produce the plots. This would be like running a larger ver-

sion of the single-instance HMODEL world. Producing the same outcome within a single instance of HMODEL could be achieved by increasing the number of agents and/or the space: in other words, increasing the volume of the record to be sampled.

The discrepancy between the replication and original was undoubtedly the result of incomplete documentation. Although the sampling procedure is described in the original publication, versions of the HMODEL NetLogo code published after the original study included a plot generator to illustrate the outcomes of single model runs. This addition was not documented in the ODD or model code, but gives the impression that all parameter configurations should produce an adequate sample when in fact they do not.

At the same time, identifying the reasons for the discrepancy leads to a conclusion not considered in the original paper: to *see* the archaeological pattern produced by HMODEL under highly erosive or highly depositional conditions with frequent events may require study of a greater volume of the archaeological record. To apply these findings to a real archaeological study, this means that either the number of hearths constructed through time would need to be higher in highly erosive or depositional places than that produced under conditions that are more balanced or, more likely, the scale of the survey would need to be larger.

The expansions of HMODEL further illustrate the implications of the scale of observation for temporal patterning in dated surface features. In the original study, the timing of sedimentary events is uniform between runs (e.g. 10, 50, 100, or 200 years), resulting in pronounced gaps at these intervals. This implicitly assumes that the area under study shares a history of wetting and drying driving sedimentary events, which is reasonable for a single catchment like Rutherfords Creek, New South Wales. However, when the timing of sedimentary events is randomized between runs, single runs still show gaps and clustering of hearth ages in time, but these gaps are not visible when runs are aggregated. This is similar to combining hearth ages across sampling areas with different precipitation histories and topographies, where the record erased in one area is preserved in another. The outcome suggests

Table 2: F-test for equality of variance for Probability experiments and equivalent re-implemented HMODEL survey data.

Collection Method	Probability	Erosion/Deposition	F	p	Collection Method	Event Interval
Probability	0.02	0.5	5.5763	2.20E-16	Survey	50
Probability	0.01	0.3	9.6345	2.20E-16	Survey	100
Probability	0.01	0.5	9.9368	2.20E-16	Survey	100
Probability	0.01	0.7	8.446	2.20E-16	Survey	100
Probability	0.005	0.1	34.917	2.20E-16	Survey	200
Probability	0.005	0.3	18.936	2.20E-16	Survey	200
Probability	0.005	0.5	4.8876	1.17E-06	Survey	200
Probability	0.005	0.7	19.635	2.20E-16	Survey	200
Probability	0.005	0.9	36.156	2.20E-16	Survey	200

that in order to see gaps resulting from erosional history, the landscape needs to have been subjected to a common history of forces driving sedimentary processes. As the scale of observation increases, say from a creek catchment to a river basin, this becomes more difficult to control.

5. Conclusion

While the re-implemented HMODEL replicated only half of Davies, Holdaway and Fanning's (2016) results, the model replication exercise can be considered a successful endeavor. Results confirm the findings of the original model and indicate that episodic geomorphic events significantly affect archaeological deposit formation, radiocarbon chronologies, and human occupation reconstructions. Results were relationally and qualitatively similar to the original model (**Figure 2**). This study, then, is significant in that it demonstrates a successful replication of an ABM in which the original paper's findings are validated. The above exercises add to the small but growing number of successful archaeological ABM reproduction studies (Kanters 2018; Janssen 2009). Likewise, the probability experiment revision further demonstrates that episodic fluvial events can create highly varied radiocarbon distributions which may be problematic to extrapolate to past demographic and occupational patterns.

Moreover, the excavations experiment illustrates that excavation data helps to fill in hiatuses in radiocarbon chronologies on depositional landforms. This simulation, however, indicates that excavation data has no effect on radiocarbon distributions drawn from landscapes subject to erosion. As the use of summed probability distribution models to reconstruct occupational trends is currently in vogue, the simulated results presented here should be considered by those employing such techniques. This finding should be corroborated or "ground-truthed" and tested against actual archaeological data. Further research should also look into the interplay between geomorphology, chronometric data collection, and reconstructed occupational patterns across dynamic landscapes.

Returning to the question of whether replication will help to create better theories, the exercise illustrates pathways to achieve this. The incongruence between the original study and the replication derives from a miscommunication about sampling mechanisms, and speaks to the need for clearer and more thorough documentation of computational models as advocated in calls for reproducibility (e.g. Marwick et al. 2017; Romanowska 2015). Nevertheless, attempting the replication and identifying the reasons for discrepancies between the replication and original raised issues with theoretical implications that build on those raised by the original model. Most important is that under highly erosional or highly depositional conditions where the time between sedimentary events is short, the number of hearths remaining on the surface may be very few relative to a comparable area with more geomorphic balance and longer intervals between sedimentary processes. Clarke (1972: 1) describes models as "pieces of machinery that relate observations to theoretical ideas." In this case, by engaging with the "mechanics" that produced the inconsistency between model

and replica, the exercise produced a novel realization, extending its value beyond the important but somewhat less glamorous realm of quality control. This finding will require additional data to test for the effects of scale of observation under different geomorphic regimes, feeding into the "virtuous cycle" of knowledge generation (Smaldino 2019: 9).

Model replication or re-alignment is still a relatively uncommon practice across disciplines (Fitzpatrick 2019). Computational scientists working with ABMs have been urging their peers to practice model replication for over 20 years (Axtell et al. 1996). Alas, across scientific fields there are few to no incentives for scientists to reproduce their studies, experiments, models, or other forms of research. The "publish or perish" axiom is well known, yet academic and scientific communities do not consider reproduced or negative results as prestigious work with significant weight in an individual's overall scientific output (Fanelli 2012; Nosek, Spies and Motyl 2012). Génova and de la Vara (2019) see this phenomenon as a result of cultural pressures within scientific disciplines. We suspect that until such mindsets change, model reproductions will be slow to increase.

There are, however, some concrete ways in which computational scientists are pushing for model reproduction. Some researchers now call for changes in journal editorial standards as a means of ensuring computational reproducible research. This could entail requiring research compendiums with all associated data, files, analytic code, and programs be archived in online, Open Access digital repositories (Beale 2012; Nosek et al. 2015; Peng 2011). Others suggest that ABMs should be replicated or tested as part of the peer review process (Romanowska 2015). Model replication, however, is a time-consuming process with few benefits for reviewers or journals. Coding and reconstructing another's work is dependent on the degree of available open access data and can create significant time demands (Marwick 2017). Nevertheless, as computational reproducibility, training in reproducible research, and an ethos of sharing data across the social sciences continue to gain traction, perhaps model replication will become a more frequent practice (Marwick 2017; Marwick et al. 2019; Nosek et al. 2015). Such practices may continue to become commonplace as archaeologists work together to discuss the utility of ABM's in jargon-free language and create tutorials to increase accessibility (Crabtree et al. 2019; Davies et al. 2019; Romanowska et al. 2019).

As Popper and Pichler (2015) note, if models and simulations are to be useful in answering archaeological and anthropological questions, it is important that the model outputs useful and accurate results. Model replication or re-implementation is one way in which agent-based models may be tested, falsified, and expanded upon. By increasing the number of verified models and sharing analysis results and code openly, archaeologists can continue to incorporate and expand upon geoarchaeological and radiocarbon simulations and analyses of past human behavior. Such practice not only verifies the validity of published models but allows archaeologists to strengthen and expand their inferential and interpretive narratives.

Acknowledgements

We would like to thank L.S. Premo and the 2016 Models and Simulations class at Washington State University for their assistance, coding support, and overall encouragement of this project. We also wish to thank the two anonymous reviewers whose comments greatly improved this work.

Competing Interests

The authors have no competing interests to declare.

References

- Axtell, R, Axelrod, R, Epstein, JM and Cohen, MD.** 1996. Aligning simulation models: A case study and results. *Computational and Mathematical Organization Theory*, 1(2): 123–141. DOI: <https://doi.org/10.1007/BF01299065>
- Baker, M.** 2016. Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the crisis rocking science and what they think will help. *Nature*, 533(7604): 452–455. DOI: <https://doi.org/10.1038/533452a>
- Beale, N.** 2012. How community archaeology can make use of open data to achieve further its objectives. *World Archaeology*, 44(4): 612–633. DOI: <https://doi.org/10.1080/00438243.2012.743252>
- Bersini, H.** 2012. UML for ABM. *Journal of Artificial Societies and Social Simulation*, 15(1): 9. DOI: <https://doi.org/10.18564/jasss.1897>
- Bigbee, A, Cioffi-Revilla, C and Luke, S.** 2005. Replication of Sugarscape using MASON. In: Troitzsch, KG (ed.) *Representing Social Reality: Pre-proceedings of the Third Conference of the European Social Simulation Association, Koblenz, September 5–9, 2005*. Koblenz: Verlag Dietmar Fölbach, pp. 6–15.
- Brantingham, PJ.** 2003. A neutral model of stone raw material procurement. *American Antiquity*, 68(3): 487–509. DOI: <https://doi.org/10.2307/3557105>
- Carney, M.** 2020. mollyrcarney/HMODEL-replication: HMODEL Replication Code and Supplementary Files (Version 1.0.2) [Data set]. *Zenodo*. DOI: <http://doi.org/10.5281/zenodo.3601033>
- Cegielski, WH and Rogers, JD.** 2016. Rethinking the role of Agent-Based Modeling in archaeology. *Journal of Anthropological Archaeology*, 41: 283–298. DOI: <https://doi.org/10.1016/j.jaa.2016.01.009>
- Clarke, DL.** 1972. Models and paradigms in contemporary archaeology. In: Clarke, DL (ed.) *Models in Archaeology*. London: Methuen, pp. 1–60.
- Contreras, DA and Meadows, J.** 2014. Summed radiocarbon calibrations as a population proxy: a critical evaluation using a realistic simulation approach. *Journal of Archaeological Science*, 52: 591–608. DOI: <https://doi.org/10.1016/j.jas.2014.05.030>
- Crabtree, SA, Harris, K, Davies, B and Romanowska, I.** 2019. Outreach in archaeology with agent-based modeling: Part 3 of 3. *Advances in Archaeological Practice*, 7(2): 194–202. DOI: <https://doi.org/10.1017/aap.2019.4>
- Crema, ER, Bevan, A and Shennan, S.** 2017. Spatio-temporal approaches to archaeological radiocarbon dates. *Journal of Archaeological Science*, 87: 1–9. DOI: <https://doi.org/10.1016/j.jas.2017.09.007>
- Davies, B.** 2018. *HMODEL*, March 1, 2018. Available at <https://github.com/b-davies/HMODEL>. [Last accessed November 2019].
- Davies, B, Holdaway, SJ and Fanning, PC.** 2016. Modelling the palimpsest: An exploratory agent-based model of surface archaeological deposit formation in a fluvial arid Australian landscape. *The Holocene*, 26(3): 450–463. DOI: <https://doi.org/10.1177/0959683615609754>
- Davies, B, Romanowska, I, Harris, K and Crabtree, SA.** 2019. Combining geographic information systems and agent-based models in archaeology: Part 2 of 3. *Advances in Archaeological Practice*, 7(2): 185–193. DOI: <https://doi.org/10.1017/aap.2019.5>
- Dibble, HL, Holdaway, SJ, Lin, SC, Braun, DR, Douglass, MJ, Iovita, R, Mcpherron, SP, Olszewski, DI and Sandgathe, D.** 2017. Major fallacies surrounding stone artifacts and assemblages. *Journal of Archaeological Method and Theory*, 24(3): 813–851. DOI: <https://doi.org/10.1007/s10816-016-9297-8>
- Donkin, E, Dennis, P, Ustalakov, A, Warren, J and Clare, A.** 2017. Replicating complex agent based models, a formidable task. *Environmental Modelling & Software*, 92: 142–151. DOI: <https://doi.org/10.1016/j.envsoft.2017.01.020>
- Douglass, MJ, Lin, SC, Braun, DR and Plummer, TW.** 2018. Core use-life distributions in lithic assemblages as a means for reconstructing behavioral patterns. *Journal of Archaeological Method and Theory*, 25(1): 254–288. DOI: <https://doi.org/10.1007/s10816-017-9334-2>
- Edmonds, B and Hales, D.** 2003. Replication, replication and replication: Some hard lessons from model alignment. *Journal of Artificial Societies and Social Simulation*, 6(4): 1–25. <http://jasss.soc.surrey.ac.uk/6/4/11.html>
- Epstein, JM.** 2008. Why model? *Journal of Artificial Societies and Social Simulation*, 11(4): 12. <http://jasss.soc.surrey.ac.uk/11/4/12.html>
- Fanelli, D.** 2009. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PloS one*, 4(5): e5738. DOI: <https://doi.org/10.1371/journal.pone.0005738>
- Fanelli, D.** 2012. Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90: 891–904. DOI: <https://doi.org/10.1007/s11192-011-0494-7>
- Fitzpatrick, BG.** 2019. Issues in reproducible simulation research. *Bulletin of Mathematical Biology*, 81(1): 1–6. DOI: <https://doi.org/10.1007/s11538-018-0496-1>
- Flynn, KM, Kirby, WH and Hummel, PR.** 2006. *User's manual for program PeakFQ, annual flood-frequency analysis using Bulletin 17B guidelines. Techniques and Methods*. Virginia: U.S. Geological Survey. DOI: <https://doi.org/10.3133/tm4B4>
- Galán, JM and Izquierdo, LR.** 2005. Appearances can be deceiving: Lessons learned re-implementing

- Axelrod's 'Evolutionary Approach to Norms'. *Journal of Artificial Societies and Social Simulation*, 8(3). <http://jasss.soc.surrey.ac.uk/8/3/2.html>.
- Génova, G and Vara, JL de la.** 2019. The problem is not professional publishing, but the publish-or-perish culture. *Science and Engineering Ethics*, 25: 617–619. DOI: <https://doi.org/10.1007/s11948-017-0015-z>
- Giere, R.** 1988. Laws, theories, and generalizations. In: Grunbaum, A and Salmon, W (eds.), *The Limits of Deductivism*. Berkeley: University of California Press, pp. 37–46.
- Grimm, V, Berger, U, Bastiansen, F, Eliassen, S, Ginot, V, Giske, J, Goss-Custard, J, Grand, T, Heinz, SK, Huse, G, Huth, A, Jepsen, JU, Jørgensen, C, Mooij, WM, Müller, B, Pe'er, G, Pious, C, Railsback, SF, Robbins, AM, Robbins, MM, Rossmanith, E, Rüger, N, Strand, E, Souissi, S, Stillman, RA, Vabø, R, Visser, U and Deangelis, DL.** 2006. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1–2): 115–126. DOI: <https://doi.org/10.1016/j.ecolmodel.2006.04.023>
- Grimm, V, Berger, U, Deangelis, DL, Polhill, JG, Giske, J and Railsback, SF.** 2010. The ODD protocol: A review and first update. *Ecological Modelling*, 221(23): 2760–2768. DOI: <https://doi.org/10.1016/j.ecolmodel.2010.08.019>
- Hiscock, P and Attenbrow, V.** 2016. Dates and demography? The need for caution in using radiometric dates as a robust proxy for prehistoric population change. *Archaeology in Oceania*, 51(3): 218–219. DOI: <https://doi.org/10.1002/arco.5096>
- Huggett, J.** 2004. Archaeology and the new technological fetishism. *Archeologia e calcolatori*, 15: 81–92.
- Janssen, MA.** 2009. Understanding artificial Anasazi. *Journal of Artificial Societies and Social Simulation*, 12(4): 13. <http://jasss.soc.surrey.ac.uk/12/4/13.html>.
- Kanters, J.** 2018. Replication of agent-based models in archaeology: a case study using Brughmans and Poblome's MERCURY model. Unpublished MSc thesis. Leiden: University of Leiden.
- Latour, B and Woolgar, S.** 2013. *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press. DOI: <https://doi.org/10.2307/j.ctt32bbxc>
- Lucas, G.** 2001. Destruction and the rhetoric of excavation. *Norwegian Archaeological Review*, 34(1): 35–46. DOI: <https://doi.org/10.1080/00293650119347>
- Macal, CM.** 2017. Everything you need to know about agent-based modelling and simulation. *Journal of Simulation*, 10(2): 144–156. DOI: <https://doi.org/10.1057/jos.2016.7>
- Martin, GN and Clarke, RM.** 2017. Are psychology journals anti-replication? A snapshot of editorial practices. *Frontiers in Psychology* 8: 523. DOI: <https://doi.org/10.3389/fpsyg.2017.00523>
- Marwick, B.** 2017. Computational reproducibility in archaeological research: Basic principles and a case study of their implementation. *Journal of Archaeological Method and Theory*, 24(2): 424–450. DOI: <https://doi.org/10.1007/s10816-015-9272-9>
- Marwick, B, D'Alpoim Guedes, J, Barton, CM, Bates, LA, Baxter, M, Bevan, A, Bollwerk, EA, Bocinsky, RK, Brughmans, T and Carter, AK.** 2017. Open science in archaeology. *SAA Archaeological Record*, 17(4): 8–14. DOI: <https://doi.org/10.31235/osf.io/72n8g>
- Marwick, B, Wang, L, Robinson, R and Loiselle, H.** 2019. How to use replication assignments for teaching integrity in empirical archaeology. *Advances in Archaeological Practice*. DOI: <https://doi.org/10.1017/aap.2019.38>
- McNutt, M.** 2014. Reproducibility. *Science*, 343(6168): 229–229. DOI: <https://doi.org/10.1126/science.1250475>
- Miller, JH and Page, SE.** 2009. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton: Princeton University Press. DOI: <https://doi.org/10.1515/9781400835522>
- Miodownik, D, Cartrite, B and Bhavnani, R.** 2010. Between replication and docking: "Adaptive Agents, Political Institutions, and Civic Traditions" revisited. *Journal of Artificial Societies and Social Simulation*, 13(3). DOI: <https://doi.org/10.18564/jasss.1627>
- Moonesinghe, R, Khoury, MJ and Janssens, ACJ.** 2007. Most published research findings are false – but a little replication goes a long way. *PLoS medicine*, 4(2): e28. DOI: <https://doi.org/10.1371/journal.pmed.0040028>
- Müller, B, Balbi, S, Buchmann, CM, De Sousa, L, Dressler, G, Groeneveld, J, Klassert, CJ, Le, QB, Millington, JD and Nolzen, H.** 2014. Standardised and transparent model descriptions for agent-based models: Current status and prospects. *Environmental Modelling & Software*, 55: 156–163. DOI: <https://doi.org/10.1016/j.envsoft.2014.01.029>
- Munafò, MR, Nosek, BA, Bishop, DVM, Button, KS, Chambers, CD, Percie du Sert, N, Simonsohn, U, Wagenmakers, EJ, Ware, JJ and Ioannidis, JPA.** 2017. A manifesto for reproducible science. *Nature Human Behaviour*, 1(1): 1–9. DOI: <https://doi.org/10.1038/s41562-016-0021>
- Nosek, BA, Alter, G, Banks, GC, Borsboom, D, Bowman, SD, Breckler, SJ, Buck, S, Chambers, CD, Chin, G, Christensen, G, Contestabile, M, Dafoe, A, Eich, E, Freese, J, Glennerster, R, Goroff, D, Green, DP, Hesse, B, Humphreys, M, Ishiyama, J, Karlan, D, Kraut, A, Lupia, A, Mabry, P, Madon, T, Malhotra, N, Mayo-Wilson, E, McNutt, M, Miguel, E, Paluck, E L, Simonsohn, U, Soderberg, C, Spellman, BA, Turitto, J, Vandenbos, G, Vazire, S, Wagenmakers, EJ, Wilson, R and Yarkoni, T.** 2015. Promoting an open research culture. *Science*, 348(6242): 1422–1425. DOI: <https://doi.org/10.1126/science.aab2374>
- Nosek, BA, Spies, JR and Motyl, M.** 2012. Scientific Utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on*

- Psychological Science*, 7(6): 615–31. DOI: <https://doi.org/10.1177/1745691612459058>
- Oestmo, S, Janssen, MA and Marean, CW.** 2016. Testing Brantingham's neutral model: The effect of spatial clustering on stone raw material procurement. In: Barceló, JA and Del Castillo, F (eds.), *Simulating Prehistoric and Ancient Worlds*. Cham: Springer International Publishing, pp. 175–188. DOI: https://doi.org/10.1007/978-3-319-31481-5_4
- Peng, RD.** 2011. Reproducible research in computational science. *Science*, 334(6060): 1226–1227. DOI: <https://doi.org/10.1126/science.1213847>
- Perry, GL, Wainwright, J, Etherington, TR and Wilmschurst, JM.** 2016. Experimental simulation: Using generative modeling and palaeoecological data to understand human-environment interactions. *Frontiers in Ecology and Evolution*, 4: 109. DOI: <https://doi.org/10.3389/fevo.2016.00109>
- Polhill, JG, Parker, D, Brown, D and Grimm, V.** 2008. Using the ODD protocol for describing three agent-based social simulation models of land-use change. *Journal of Artificial Societies and Social Simulation*, 11(23): 1–35. <http://jasss.soc.surrey.ac.uk/11/2/3.html>.
- Pop, CM.** 2016. Simulating lithic raw material variability in archaeological contexts: A re-evaluation and revision of Brantingham's neutral model. *Journal of Archaeological Method and Theory*, 23(4): 1127–1161. DOI: <https://doi.org/10.1007/s10816-015-9262-y>
- Popper, K.** 2005. *The Logic of Scientific Discovery*. Abingdon-on-Thames: Routledge. DOI: <https://doi.org/10.4324/9780203994627>
- Popper, N and Pichler, P.** 2015. Reproducibility. In: Wurzer, G, Kowarik, K and Reschreiter, H (eds.), *Agent-based Modeling and Simulation in Archaeology*. Cham: Springer, pp. 77–98. DOI: https://doi.org/10.1007/978-3-319-00008-4_4
- Premo, LS.** 2006a. Agent-based models as behavioral laboratories for evolutionary anthropological research. *Paper presented at the 2006 Computer Applications and Quantitative Methods in Archaeology Conference*.
- Premo, LS.** 2006b. Exploratory agent-based models: towards an experimental ethnoarchaeology. In: Clark, JT and Hagemeister, EM (eds.), *Digital Discovery: Exploring New Frontiers in Human Heritage: CAA 2006*. Budapest: Archaeolingua Press, pp. 29–36.
- Railsback, SF and Grimm, V.** 2012. *Agent Based and Individual Based Modeling: A Practical Introduction*. Princeton and Oxford: Princeton University Press.
- Rick, JW.** 1987. Dates as data: an examination of the Peruvian preceramic radiocarbon record. *American Antiquity*, 52(1): 55–73. DOI: <https://doi.org/10.2307/281060>
- Rollins, ND, Barton, CM, Bergin, S, Janssen, MA and Lee, A.** 2014. A computational model library for publishing model documentation and code. *Environmental Modelling & Software*, 61: 59–64. DOI: <https://doi.org/10.1016/j.envsoft.2014.06.022>
- Romanowska, I.** 2015. So you think you can model? A guide to building and evaluating archaeological simulation models of dispersals. *Human Biology*, 87(3): 169–193. DOI: <https://doi.org/10.13110/humanbiology.87.3.0169>
- Romanowska, I, Crabtree, SA, Harris, K and Davies, B.** 2019. Agent-based modeling for archaeologists: Part 1 of 3. *Advances in Archaeological Practice*, 7(2): 178–184. DOI: <https://doi.org/10.1017/aap.2019.6>
- Shennan, S.** 1997. *Quantifying archaeology*. 2nd Edition. Edinburgh: Edinburgh University Press.
- Simons, DJ.** 2014. The value of direct replication. *Perspectives on Psychological Science*, 9(1): 76–80. DOI: <https://doi.org/10.1177/1745691613514755>
- Smaldino, PE.** 2019. Better methods can't make up for mediocre theory. *Nature*, 575(7781): 9. DOI: <https://doi.org/10.1038/d41586-019-03350-5>
- Thiele, JC, Kurth, W and Grimm, V.** 2014. Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using NetLogo and R. *Journal of Artificial Societies and Social Simulation*, 17(3): 11. <http://jasss.soc.surrey.ac.uk/17/3/11.html>. DOI: <https://doi.org/10.18564/jasss.2503>
- Wilensky, U.** 1999. NetLogo 6.0.2 Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University. Available at: <http://ccl.northwestern.edu/>.
- Wilensky, U and Rand, W.** 2007. Making models match: Replicating an agent-based model. *Journal of Artificial Societies and Social Simulation*, 10(42): 1–25. <http://jasss.soc.surrey.ac.uk/10/4/2.html>.
- Wilensky, U and Rand, W.** 2015. Verification, validation, and replication. *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. Boston, MA: MIT Press.
- Williams, AN.** 2012. The use of summed radiocarbon probability distributions in archaeology: a review of methods. *Journal of Archaeological Science*, 39(3): 578–589. DOI: <https://doi.org/10.1016/j.jas.2011.07.014>

How to cite this article: Carney, M and Davies, B. 2020. Agent-Based Modeling, Scientific Reproducibility, and Taphonomy: A Successful Model Implementation Case Study. *Journal of Computer Applications in Archaeology*, 3(1), pp. 182–196. DOI: <https://doi.org/10.5334/jcaa.52>

Submitted: 17 January 2020

Accepted: 29 May 2020

Published: 01 July 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[

Journal of Computer Applications in Archaeology, is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS