J Archaeol Method Theory
https://doi.org/10.1007/s10816-017-9355-x

# Identifying Narrative Descriptions in Agent-Based Models Representing Past Human-Environment Interactions

**George L. W. Perry**[1] · **David O'Sullivan**[2]



**Abstract** There is a growing use of bottom-up simulation models to reconstruct past human-environment interactions. Such detailed representations pose difficult questions not only in their design (the generality-realism trade-off) but also about the inferences that are made from them. The historical sciences are faced with seeking to make robust inferences from limited, potentially biased and/or incomplete samples from uncontrolled systems, and as a result have sometimes employed narrative explanation. By contrast, simulation models can be used experimentally and can generate large amounts of data. Here, using an agent-based model of hunter-gatherer foraging in a previously unexplored ecosystem, we consider how narratives might be identified from the trajectories produced by simulations. We show how machine learning methods can isolate qualitatively similar types of model behaviour based on summaries of model outcomes and time series. We stand to learn from this approach because it enables us to answer two questions: (i) under what conditions (representations and/or parameterisations) do we observe in the model what is recorded in the archaeological and/or palaeoenvironmental record? and (ii) does the model yield unobserved dynamics? If so, are they plausible? Using models to develop narratives is a logical extension of the bottom-up approach inherent in agent-based modelling and has the potential, alongside conventional methods of model evaluation, to aid in learning from the rich dynamics of such simulations.

**Data and Code Availability** NetLogo model code is available at https://figshare.com/s/141c63b6bedc5332aba2 (doi: 10.17608/k6.auckland.5327944).

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s10816-017-9355-x) contains supplementary material, which is available to authorized users.

✉ George L. W. Perry
george.perry@auckland.ac.nz

1 School of Environment, University of Auckland, Auckland, New Zealand

2 Department of Geography, University of California, Berkeley, Berkeley, CA, USA

Published online: 13 November 2017

**Keywords** Agent-based models · Human-environment interactions · Machine learning · Narrative

## Introduction

> "While archaeologists will continue to craft compelling narratives, we look to a future where these narratives are based on theoretically informed, explicit, computational models that form the historical framework for a science of social dynamics." Barton et al. 2010 (p. 383)

Making robust inferences about contingent socio-ecological systems (as discussed in Barton et al. 2004) requires an approach that marries empirical advances with modern methods in simulation modelling and analysis (Barton et al. 2012; Kintigh et al. 2014; Perry et al. 2016). While in some areas of the socio-ecological sciences data-driven modelling is fundamental (*e.g.* data assimilation approaches, Niu et al. 2014), there is a need to make sure that models are not seen solely as sophisticated and useful, but largely opaque, predictive devices (Lorscheid et al. 2012; Peck 2004). To be clear from the outset, we are *not* arguing that simplicity in model formulation is inherently better (see Evans et al. 2013 for an interesting perspective on the simplicity-generality debate) or that simple models are inferentially privileged, but rather that there is much more to the use and evaluation of models than prediction supported by model-data confrontation. Likewise, we are not defending the *ad hoc* evaluation of agent-based models (ABMs) by loose pattern matching that seems prevalent in the fields in which they are applied. Rather, we consider how narrative approaches when coupled with modern computational analysis provide a way to illuminate and better understand the dynamics produced by complex bottom-up models (that is models in which macroscopic dynamics develop from finer-scale processes and interactions) and hence ultimately to learn from them.

A model's performance is often measured by comparing its predictions against some data from the system of interest (*e.g.* using the types of tools that Mayer and Butler 1993 describe). The model is then deemed acceptable if it reproduces the observed data to some level. This model versus data confrontation has been used to assess agent-based models in many contexts, including archaeological systems (*e.g.* Axtell et al. 2002; Kohler et al. 2012). However, while model-data confrontation is a necessary component of model evaluation, there are a number of concerns with a sole reliance on it. A constraint for any pattern-based model evaluation exercise is that a single pattern is unlikely to be sufficient to select between competing model parameterisation and structures (Grimm et al. 2005; O'Sullivan and Perry 2013). Different models may have the same or similar predictive capacity; this problem is called equifinality (Beven 1993) and is central to the difficulties in inferring dynamic process from snapshot patterns (McIntire and Fajardo 2009). While the evaluation of models using multiple patterns, such as in pattern-oriented modelling (Grimm and Railsback 2012; Perry et al. 2016), may partially reduce these concerns, they do not resolve them entirely. Another problem with the emphasis on pattern matching in model evaluation is that, insofar as it judges models on their ability to reproduce some observation set, it emphasises their role as predictive devices over the other purposes they serve. Models can play important roles in dialogue and learning (Epstein 2008; Oreskes et al. 1994), and it is not obvious that model-data confrontation is a suitable way to evaluate model performance in such contexts. It is in this light that we seek to develop a framework for seeing and using bottom-up simulation models as narrative devices.

The view that simulation models are most appropriately used heuristically is not new (Oreskes et al. 1994). However, there are challenges in communicating the learning derived from simulation-based (*in silico*) experiments and, indeed, the simulation models themselves (Grimm et al. 2010; Lorscheid et al. 2012). These challenges may be particularly acute when dealing with extended time scales (Boschetti et al. 2016). One way to facilitate such communication is to see and use models as narrative devices (McGlade 2014; Millington et al. 2012; Morgan 2001; Topping et al. 2015). A narrative perspective is particularly relevant for the historical natural and social sciences (Abell 2009; Biondi 2014; Carpenter 2002; Griffin 1992). For example, Griffin (1993, p. 1099) comments that:

> "Narrative explanation takes the form of an unfolding, open-ended story fraught with conjunctures and contingency, where what happens, an action, in fact happens because of its order and position in the story. Narrative therefore permits a form of sequential causation that allows for twisting, varied and heterogeneous time paths to a particular outcome. In narratives, we can see how the cumulative consequences of past actions increasingly constrain and limit future action."

Although Griffin (1993) is not discussing simulation, his arguments for narrative explanation are remarkably similar to those often put forward by advocates of complex bottom-up models. Likewise, historically focused social scientists have developed analytical narrative techniques. Bates et al. (1998, p. 10) describe the analytic narratives approach thus:

> "Our approach is narrative; it pays close attention to stories, accounts, and context. It is analytic in that it extracts explicit and formal lines of reasoning which facilitate both exposition and explanation."

Again we would argue that this perspective could fruitfully be adopted, alongside more conventional approaches, in the evaluation of ABMs. The so-called 'narrative analytics project' (Bates et al. 1998) is, at least in part, a response to the challenges of making inferences from one-off events such as case studies (Abell 2009)—this is a general problem for the historical sciences (Biondi 2014; Cleland 2001, 2011) and beyond (March et al. 1991). As an example of the quantitative analytic narrative approach, Abell (2007) describes a Bayesian network approach to narrative, which emphasises the importance of understanding the causal links between events in a chronology *via* a graph theoretical framework. Abell's approach is just one example of the many approaches used to make inferences from qualitative data (Gerring 2017). This style of narrative analysis is potentially applicable to simulation models such as ABMs. However, unlike the one-off cases that analytic narratives often consider, computational experiments may yield thousands of individual model realisations, so isolating different and interesting classes of model behaviour (or 'narratives') is a critical first step in applying narrative approaches to them. Here, we focus on finding narratives, which we define as similar trajectories of change, among Winsberg's (2010) 'pile of numbers' (or model outputs), and then using them to better understand the past dynamics of socio-ecological systems. Our approach is intended to complement, rather than supplant, more conventional approaches to model evaluation, and is an effort to develop the mixed qualitative-simulation approach argued for by Millington and Wainwright (2016). In fact, what we present could be seen as a special case of the recent emphasis on finding structure in large, complex datasets that has driven developments in machine learning and allied methods, with the difference that we are concerned with identifying qualitative, rather than quantitative, structure, and also with using the latter to identify the former.

Irrespective of their specific context, narratives are multilayered. At the highest level, there are overarching discipline-wide narratives such as those of environmental determinism versus environmental possibilism (the latter as imagined by Dalby 2016)—these are broad in reach and applicable in (m)any context(s) (*i.e.* they are not necessarily place-based). While models alone cannot resolve these, probably unresolvable, high-level debates, they can inform them. For example, the dramatic changes in ecosystems that accompanied late Holocene settlement of the Pacific archipelagos can be seen either as inevitable, irrespective of human behaviour, given the fragility of island ecosystems, or, alternatively, as the outcome of motivated and deliberate human action, and thus, depending on perspective, either as ecodisaster or ecotriumph (Anderson 2002, p. 375). ABMs provide a way to consider such competing narratives: for example, do certain outcomes, such as megafaunal extinction, occur regardless of model structure and parameterisation (*i.e.* representation of environmental conditions and human agency)? A second level of narrative is more place-bound and focuses on the collective dynamics of specific locations or systems. For example, Barton et al. (2010) use a socio-ecological ABM to explore how agricultural and land-use practices affect soil and hydrological dynamics during the Neolithic in the Wad Ziqlab catchment, Jordan. Such approaches allow an evaluation of how landscape-level patterns in biotic and abiotic conditions emerge from finer-scale processes and decisions. Most socio-ecological ABMs have focussed on this level of enquiry. At a yet lower level, narratives may consider individual agents or actors and their interactions with the environments they inhabit. Such individual narratives are important if we want to demystify the emergent dynamics so often canonised in bottom-up models such as ABMs (O'Sullivan and Haklay 2000). However, before we can appropriately depict and interrogate individual-level narratives, we need to isolate place- or context-bound model dynamics. Here, using an ABM of resource foraging on a previously unexplored ecosystem (*e.g.* an isolated island or a fraction of a larger landmass) as a test case, we address how narratives can be identified using multivariate statistical analysis and machine learning approaches.

## Methods

### The Hunter-Gatherer Foraging Agent-Based Model

We implemented an ABM representing hunter-gatherer behaviour on a newly discovered landscape. Agents on the virtual landscape, each representing a hunter-gatherer group, make decisions about resource acquisition, including the effort they put into somewhat risky hunting for energetically higher-value resources (*e.g.* large animals) relative to lower-value but more reliable local foraging for subsistence food sources (*e.g.* plant material). The group agents may also decide to relocate their 'home camp' when they have been doing poorly at resource collection, and as they seek to explore the landscape. Over time, if successful at resource collection, a group's population will grow, enabling it to do better still, and perhaps split into multiple groups that subsequently independently exploit resources over a wider area. Eventually, a lack of success in resource collection will lead to the complete abandonment of the landscape or island.

While we do not consider a specific ecosystem, the context we consider is typical of the late Holocene human settlement of the islands of the Pacific (Kirch 2010) and the wave of faunal extinctions that followed (Steadman 1995). However, our aim is not to reconstruct the dynamics of specific systems or cultures to inform debates about the colonisation of the Pacific. Instead, we use the model and our analyses to illustrate a means of analysing the types of bottom-up model becoming more widely used by archaeologists in a way that complements and extends traditional model-data confrontation approaches.

A schematic overview of the model is presented in Fig. 1a, along with a high-level flow chart depicting the model's operation in Fig. 1b. The model is implemented in NetLogo 5.3.0 (Wilensky 1999). Full details of the model

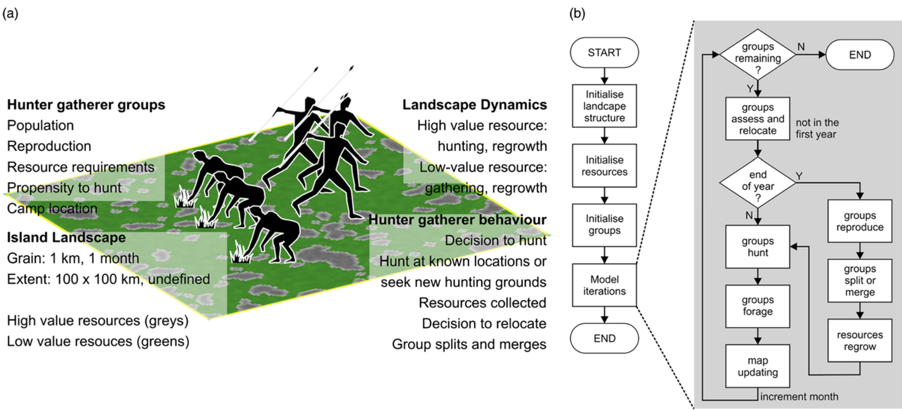


**Fig. 1** The agent-based model. (**a**) Human-environment interactions are represented in a reciprocal way as are the attributes and decision-making of the hunter-gatherer groups. In the landscape here the green areas are low-value subsistence resources, whereas the dark grey shows high-value resource (shading within each shows heterogeneity of value within each type). (**b**) A high-level flowchart of the overall model operation; shaded area shows the dynamics of the model during an individual time step (month). The details of the 'groups hunt' step are provided in Fig. 3. Figure (**b**) redrawn from O'Sullivan and Perry (2013)

implementation are provided in the supplementary materials following the ODD protocol (Grimm et al. 2010).

## Model Interrogation

Simulation models allow experimentation on aspects of system behaviour that are otherwise impossible to manipulate (Peck 2004). We conduct two model-based analyses here:

### *Same Conditions, Same Place, Same Narrative?*

We start by exploring whether the same environmental conditions consistently result in the same model dynamics (trajectories and endpoints). We do this with five analyses under baseline conditions (Table 1). These analyses consisted of a total of 10,000 model runs on 1, 10, 100, 1000 and 10,000 islands, respectively (*i.e.* in the first case all 10,000 runs were on the same islands, the second comprised of 1000 runs on each of 10 islands, and so forth). We use the term 'trajectory' to describe the individual time series arising from each model realisation and the term 'narrative' when referring to clusters of similar trajectories.

We analyse these simulations by clustering: (i) the trajectories that emerge from the model (*i.e.* the evolution of the system over time) and (ii) statistical summaries of those trajectories (*e.g.* average conditions over time or endpoints). We use a series of state variables to summarise the model's dynamics (Table 2, SM Section 2.11). These metrics describe the duration of settlement, size of human population and level of exploitation of high-value resources and are reasonably non-correlated (mean Spearman's $\rho = 0.07$; Fig. SM 1). Clustering the trajectories into narratives is ultimately a problem of time series classification. To achieve this, we use dynamic time warping (DTW), a method that measures the distance between time series on the basis of how much one time series needs to be deformed with respect to another for them to be optimally matched (Giorgino 2009). DTW can be applied to time series that differ in length, which was important in our case. The first step in the trajectory clustering was to generate a summary of the variables in each time series. We did this *via* principal components analysis (PCA) on population size, high-value resource take and kill, and local (low-value) resource use at each time step over all model replicates; PCA component one explained c. 75% of the variance in all cases. We then applied DTW to PCA axis one and so generated a distance matrix (elements being distances between time series) to which we applied standard multivariate ordination and clustering methods. We ordinate the trajectories and their summaries using metric multidimensional scaling (MDS, also known as principal coordinates analysis) with Euclidean distance (with data scaled) and then used *k*-means to assess clustering in that ordination space, with the Caliński-Harabasz criterion used to identify the optimal number of clusters (Caliński and Harabasz 1974). We fit vectors to the ordination space describing correlations with summary statistics for each model run. These analyses enable us to evaluate if different colonisation/settlement narratives can emerge under similar environmental conditions. The entire workflow is shown in Fig. 2.

### *What Controls the Narrative?*

In a second suite of analyses, we use machine learning approaches to identify the parameter conditions that lead to the emergence of different endpoints and narratives.

**Table 1** Baseline and uncertainty analysis (UA) parameter values and ranges used in the model evaluation.

| Component | Parameter | Code | Baseline value | UA range |
|---|---|---|---|---|
| Spatial structure (3) | | | | |
| | percolation.threshold | p.thresh | 0.5 | |
| | proportion.high.resource | phr | 0.15 | |
| | no.singleton.patches? | [C] | TRUE | |
| Human demography (5) | | | | |
| | min.viable.humanpop | min.*via* | 5 | |
| | max.group.size | mx.grp | 30 | |
| | r.humans | r.hum | 0.015 | |
| | r.humans.sd | r.hum.sd | 0.1 | |
| | max.birth.rate.multiple | mx.birth | 3 | |
| Resource demography (8) | | | | |
| | r.high | r.high | 0.05 | 0.04–0.06 |
| | r.high.sd | rh.sd | 0.1 | 0.08–1.2 |
| | max.high.K | mx.high.K | 2.5 | 2.0–3.0 |
| | min.sustainable.h | min.sus.h | 0.1 | 0.08–1.2 |
| | r.low | r.low | 0.2 | 0.16–0.24 |
| | r.low.sd | rl.sd | 0.2 | 0.16–0.24 |
| | max.low.K | ms.lo.K | 0.5 | 0.4–0.6 |
| | diffusion.rate | diff.rt | 0.1 | 0.08–0.12 |
| Resource exploitation (9) | | | | |
| | resource.per.head | res.ph | 1 | 0.8–1.2 |
| | hunt.kill.per.head | kill.ph | 5 | 2–10 |
| | hunt.take.per.head | take.ph | 0.1 | 0.01–0.2 |
| | hunt.party.size | h.size | 6 | 3–12 |
| | hunt.range | h.rng | 16 | 10–20 |
| | max.hunts.per.month | mx.hpm | 4 | 1–6 |
| | hunt.memory.length | h.mem | 15 | 8–20 |
| | gather.per.head | gat.ph | 0.05 | 0.01–0.1 |
| | nearby.range | nr.rng | 2.3 | 1–4 |
| Search behaviour (5) | | | | |
| | initial.search.tortuosity | init.tort | 0.1 | |
| | search.adjust | s.adj | 0.05 | |
| | max.tortuosity | mx.tort | 0.95 | |
| | min.tortuosity | mn.tort | 0.05 | |
| | relocate.near.hunting? | rnh [C] | TRUE | |

We do this by searching for contrasting model dynamics and then assessing the model parameterisations associated with them. We conducted three uncertainty analyses across plausible parameter ranges for components of the model relating to: (i) human

**Table 2** State variables used to evaluate model outcomes divided into those that emerge at model initialisation (*i.e.* are initial conditions arising from parametrisation rather than being pre-determined), human population and resource exploitation

| Component | Parameter | Variable code |
|---|---|---|
| Emergent initial conditions (1) | | |
| | Initial human population size | initial.pop |
| Human population (5) | | |
| | Length of occupation (y; ticks/12) | ticks |
| | Average population size (per year) | ave.pop |
| | Maximum population size reached | max.pop |
| | Tick of maximum population size | t.max.pop |
| | Max. number of groups during occupation | max.groups |
| Resource exploitation (12) | | |
| | Mean low-value resource used (per year) | mean.local |
| | Mean high-value resource taken (per year) | mean.taken |
| | Maximum high-value resource taken (per year) | max.taken |
| | Time maximum high-value resource taken (per year) | t.max.taken |
| | Mean high-resource kill (per year) | mean.kill |
| | Maximum high-value resource killed (per year) | max.kill |
| | Time maximum high-value resource killed (per year) | t.max.kill |
| | Max overkill (kill−taken) | max.overkill |
| | Medium ratio of low- to high-value resource use | med.local.ratio |
| | Maximum ratio of low- to high-value resource use | max.local.ratio |
| | Proportion of total available high-value resource used | prop.taken |
| | Time at which 50% of high-value resource exploited | t.prop50 |

exploitation of resources ($n$ = 10,000), (ii) island resource demography and dynamics ($n$ = 10,000), and (iii) exploitation behaviour and resource demography simultaneously ($n$ = 25,000). Each parameter combination was simulated just once, but the parameter space was swept in detail using Latin hypercube samples (Stein 1987)—this follows the protocol for the evaluation of stochastic models recommended by Prowse et al. (2016).

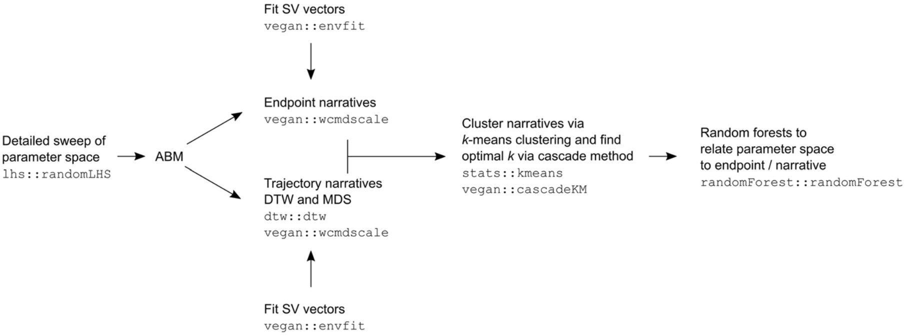


**Fig. 2** The analysis tool chain used to identify narratives. Text in `courier font` refers to R (R-Development-Core-Team 2016) packages and commands. *SV* = state variable

This analysis constitutes an uncertainty analysis on the model's qualitative dynamics as we evaluate which parameters are most influential in determining model outcome and over what ranges. As above, we clustered the model's outcomes (summaries and trajectories) using dynamic time warping, metric multidimensional scaling and *k*-means clustering.

Having classified model summaries and trajectories, we identified the parameters that influence whether a given model run yielded a given endpoint or narrative. In other words, we are evaluating the extent to which different parameter settings (or combinations thereof) give rise to the clusters. For this step, we used random forests (RF), an ensemble-based machine learning method for statistical classification (Breiman 2001). RFs are based on conventional decision trees in which some parameter space is recursively divided into units that are as homogeneous as possible in terms of the outcome being predicted. RF extends decision trees by using bagging (bootstrap aggregating), in which the final RF model is the average of many individual decision trees based on bootstrap samples of the data; this approach helps to reduce model variance (James et al. 2013). RFs have high classification accuracy and overcome some of the overfitting and instability issues that conventional decision trees can suffer from (Cutler et al. 2007). RF models do not produce significance tests, but rather identify the importance of predictors on the basis of how their exclusion affects model performance. A parameter's influence on model predictions can be visualised using partial dependency plots in which predictions are made while the parameter is varied across its range with all others held constant (the 'marginal effect'). In all analyses presented here, we constructed the final RF model on the basis of 5000 individual decision trees and randomly sampled three candidate variables at each split. We assessed the importance of each parameter in the RF classification on the basis of how much model performance was diluted by its omission. Classification error was assessed by confusion matrices (cross-tabulation of observations and predictions) and the out-of-bag (OOB) error rate, which is based on the error for each observation using the points omitted from the model training in the bagging process (James et al. 2013).

### Software

For the analyses presented here, we used R 3.3.1 and 3.4.0 (R-Development-Core-Team 2016) and the *RNetLogo* 1.0.2 library (Thiele 2014) and the *snowfall* library 1.84-6.1 (Knaus 2013) for local multicore use. The dynamic time warping distance analysis used the *dtw* library 1.18-1 (Giorgino 2009), the metric multidimensional scaling and *k*-means classification used the *vegan* 2.4-3 library (Oksanen et al. 2017) and the random forests were constructed using the *randomForest* R library 4.6-12 (Liaw and Wiener 2002).

## Model Experiments

### Same Conditions, Same Place, Same Narrative?

Even on the same landscape, different summaries and trajectories can emerge under the same parameter conditions (Fig. 3a). Using the statistical summaries to classify the simulations consistently resulted in four clusters and the classification into narratives *via* DTW of the trajectories consistently five, with one exception (the 10,000 replications on one island case). The vector fits suggest that the different classes in the summary analyses (Fig. 3, left column) lie along axes related to the proportion of resource taken and the length of the occupation. For the classifications based on trajectory, the classes relate to proportion of resource taken, initial population size and mean high-value resource kill per year (Fig. 3, right column). Thus, the state variables that explain a given simulation's position in the ordination space differ between the analysis of endpoints and trajectories; for example, length of occupation is only important in the former. These differences suggest that analysis of summaries and trajectories is not redundant and can reveal different, if complementary, model dynamics.

The classes that emerge from the use of summary statistics and the trajectories do not map onto each other (Fig. 4 and Fig. SM 2), with the DTW tending to split the classes related to less successful settlement histories into two classes but not discriminating between others (Figs. 4 and 5). That the two classifications do not directly map onto each other indicates that different information is contained in the summary statistics and in the entire trajectories. The narratives identified by the DTW approach comprise four prolonged, more successful settlements and two short-lived failures. The main difference between the failures (Fig. 5d, f) seems to lie in the rate of resource procurement in the first few years of the simulation; such subtle differences are likely very difficult to discern in archaeological or palaeoenvironmental records. The four successful narratives are more varied. Two of them are prolonged (multicentury; Fig. 5b, c) and two are shorter (multidecadal; Fig. 5a, e). Of the two prolonged occupations, one (Fig. 5b) culminates in a peak towards the end before abandonment and the other is more temporally consistent before tapering away (Fig. 5c). The two medium length occupations are more similar, but again vary in their ending (boom-bust vs. slow decline). The classification of the trajectories does not discriminate between the longer dynamics based on the summaries (Fig. 5b, c) and the classification of the summaries does not discriminate between the two short (failed) narratives (Fig. 5d, f). Neither the endpoint nor the narrative clusters relate to individual landscape geographies (Fig. 6).

### Using Machine Learning to Identify What Controls the Narrative

The analyses described above suggest that even in the same setting contrasting dynamics can arise. The archaeological record contains a variety of trajectories and narratives, not least because locations vary in their physical geography and were settled by different groups of people; our preliminary analysis suggests that contingency or 'chance' plays an important role in the narratives that emerge under a given set of conditions. However, it is informative to isolate the parameterisations that lead to the emergence of different narratives because it may, *via* surrogative reasoning (that is, applying learning from a model to some real system; Contessa 2007), yield some insight into the types of human behaviours and environmental conditions more likely to have been associated with them. Again, while our model is not intended to be a highly realistic representation of a specific system, the analysis approach we describe here is applicable to more targeted simulations.

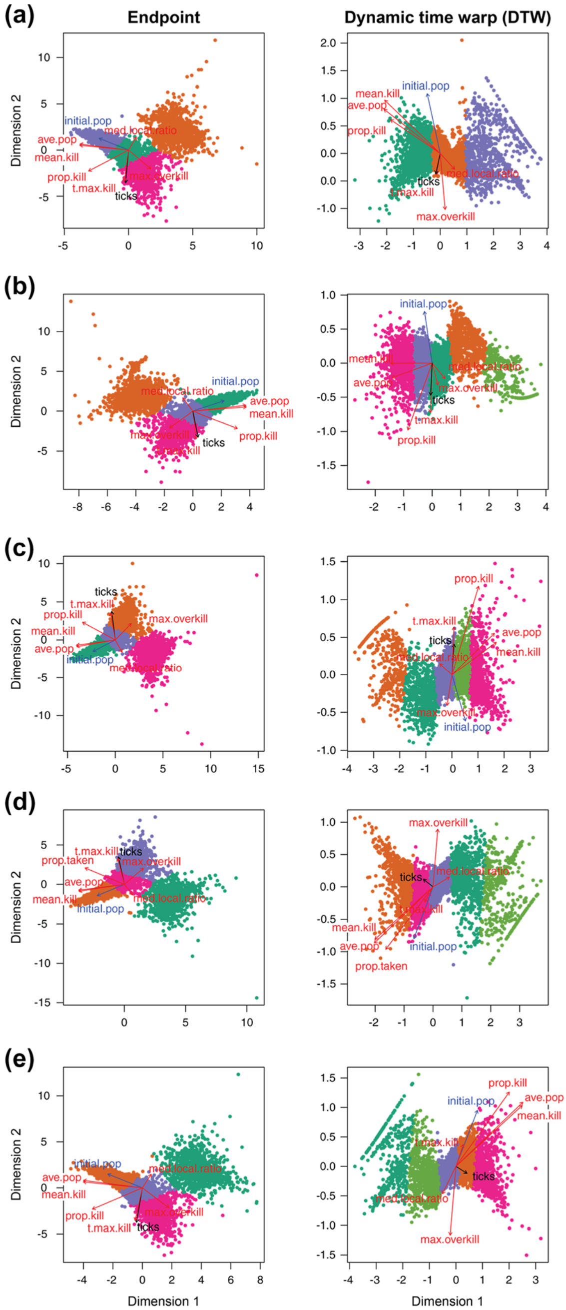


**Fig. 3** MDS ordination plots for the analyses under baseline conditions on islands; in all cases a total of 10,000 runs were conducted on (from [**a**] to [**e**]) 1, 10, 100, 1000, 10,000 different islands. Colours represent different clusters in ordination space as identified *via* *k*-means clustering and the optimal number of clusters *via* the Caliński-Harabasz index. Blue vectors are variables that are emergent initial conditions, red vectors represent resource exploitation related variables and black represent human population descriptors (see Table 2)

For all three parameter space explorations, the RF classifications suggest that there are three to five classes, which blend into each other (Figs. 7 and 8 and SM 3–6), falling on a continuum from short, low exploitation ('failed') settlement to prolonged, high exploitation ('successful') settlement. This is a similar classification to those shown in Figs. 4 and 5; again, the clusters identified on the basis of the summary conditions and the trajectories do not perfectly map onto each other (Fig. SM 7). The clustering analysis suggests that whereas there is just one way to 'fail', there may be multiple ways to 'succeed' in colonising an island (taking resource exploitation and settlement duration as indices of success).

The predictive performance of the model ranges from around 50 to 70% based on out-of-bag error estimation. While the RFs predict the classes at each end of the

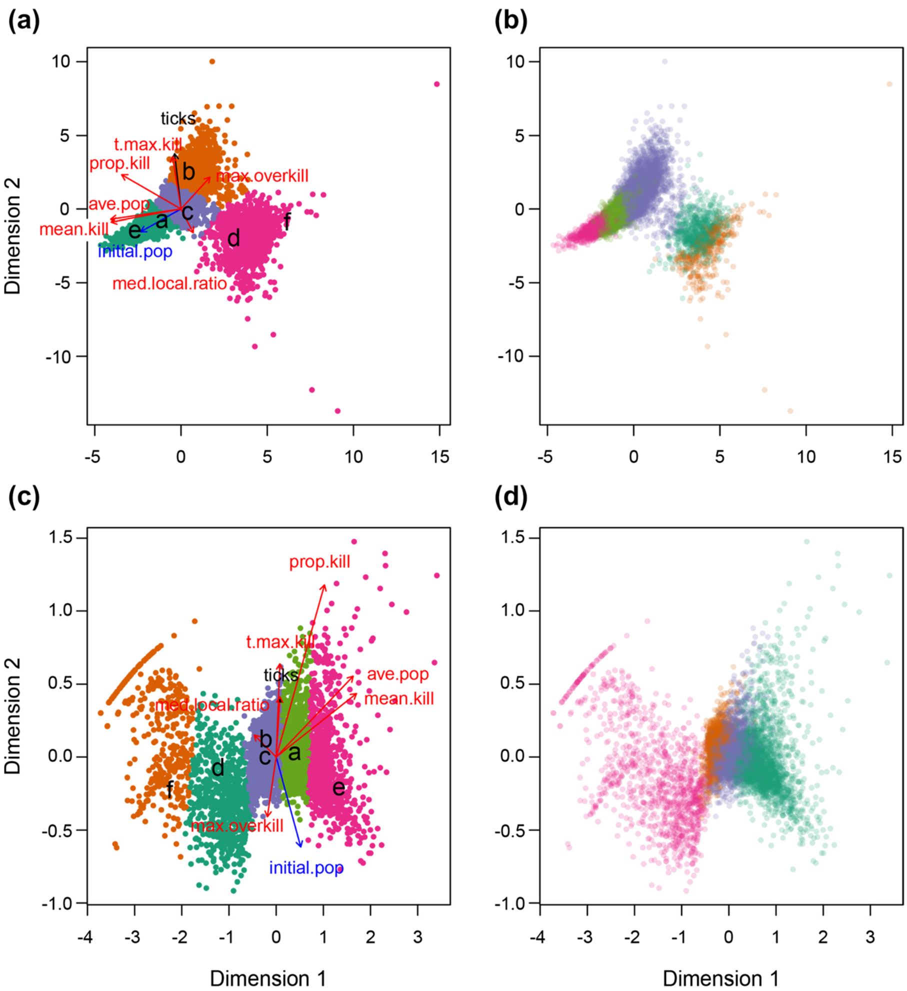


**Fig. 4** MDS ordinations for the scenario where 100 islands were simulated 100 times each, showing the clusters as identified by the statistical summaries (**a**) and the trajectories *via* DTW (**c**). In (**b**) and (**d**), the plots are coloured such that the location of each point in the ordination spaces is based on the summary and DTW, respectively, but coloured by the other. The letters on (**a**) and (**c**) relate to the time series plotted in Fig. 6. Vectors as per Fig. 3

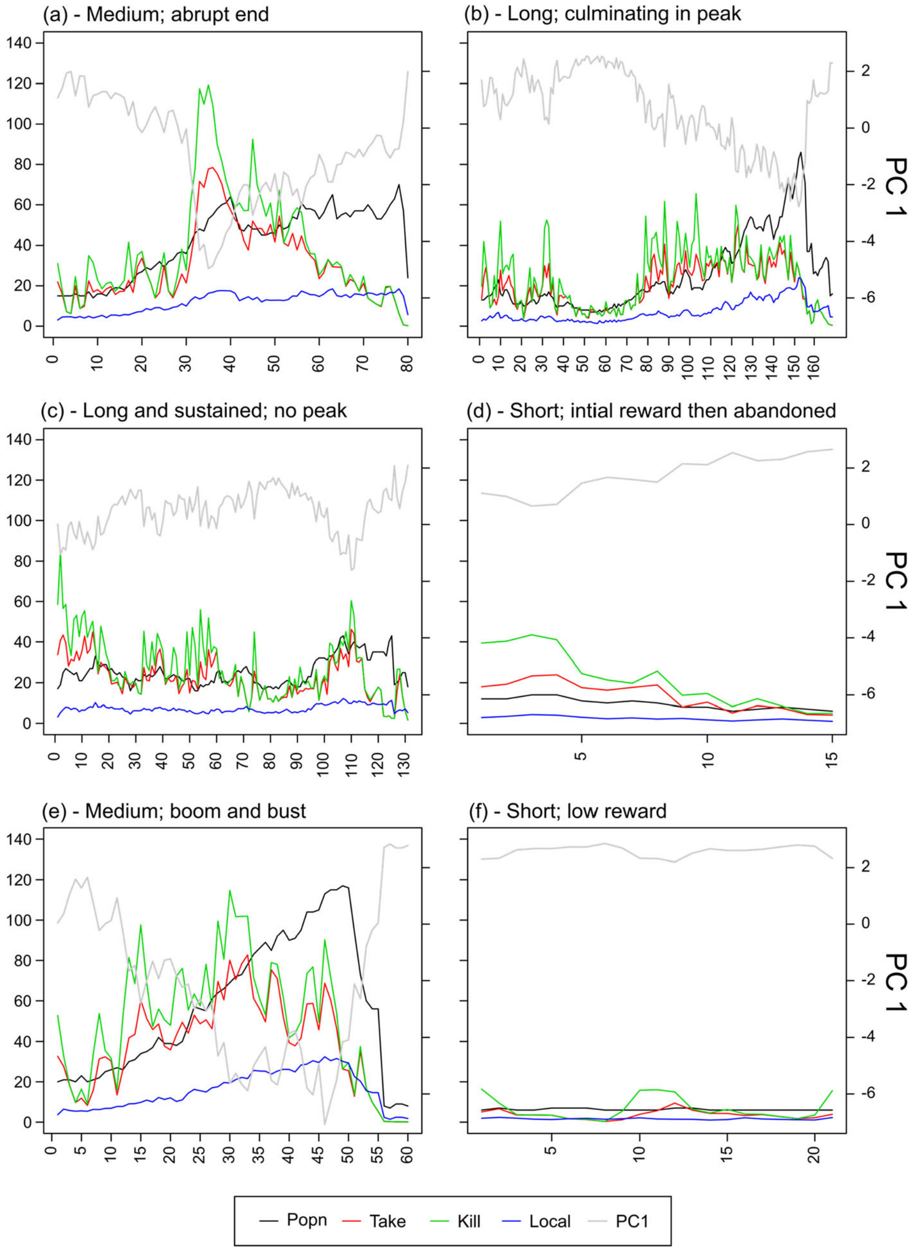


**Fig. 5** Representative time series (trajectories) for each of the clusters identified by MDS ordination of statistical summaries of the trajectory and the trajectories themselves *via* DTW. The position in ordination space of each of these time series is shown in Fig. 4

spectrum of dynamics (*i.e.* prolonged successful settlement vs. short failed settlement) reasonably well, the intermediate ones are less successfully predicted (Table SM 1). Some of this error may arise from the stochastic nature of the model we evaluate—similar parameterisations may result in quite different outcomes which violates the

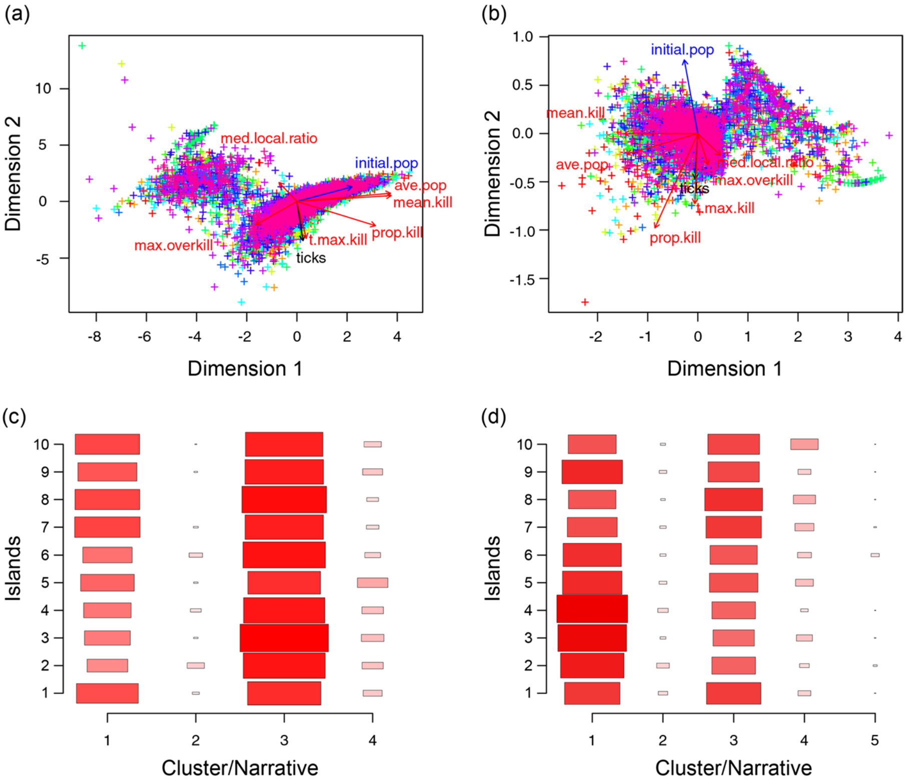


**Fig. 6** Top row (**a** and **b**) shows MDS ordination of 10,000 model runs with 1000 replicate runs were conducted on 10 islands. The different colours denote the islands, with (**a**) the MDS based on summary measures of the final condition and (**b**) that based on DTW of the trajectories. The bottom row (**c** and **d**) shows the corresponding Hinton plots with the size and colours of the squares representing the frequency of each cluster/narrative on each island. Although some narratives occur more frequently than others, there is no pattern in cluster frequency by island identity

smoothness (continuity) assumption inherent in classification (that is, that points near to each other in the parameter space will map to similar outcomes; Chapelle et al. 2006, p. 5). This issue could be explored by repeated simulation of the same parameterisations, but this would come at a computational cost that, assuming finite digital resources, would limit exploration of as broad a parameter space. When both island resource demography and human exploitation behaviour are evaluated simultaneously, a slightly different suite of predictors are identified as influential in the classifications based on the summary statistics and the trajectories. In the summary-based classification (Fig. 7d), the four most important parameters are the area over which hunting takes place (hunt.range), the number of hunts per month (max.hunts.per.month), the distance over which local foraging occurs (nearby.range) and the size of the hunting party (hunt.party.size). In the DTW-based classification (Fig. 8d), the four most important parameters are the area over which hunting takes place (hunt.range), the carrying capacity for high-value resources (max.high.K), how much resource can be gathered per head (gather.per.head) and the distance over which local foraging occurs

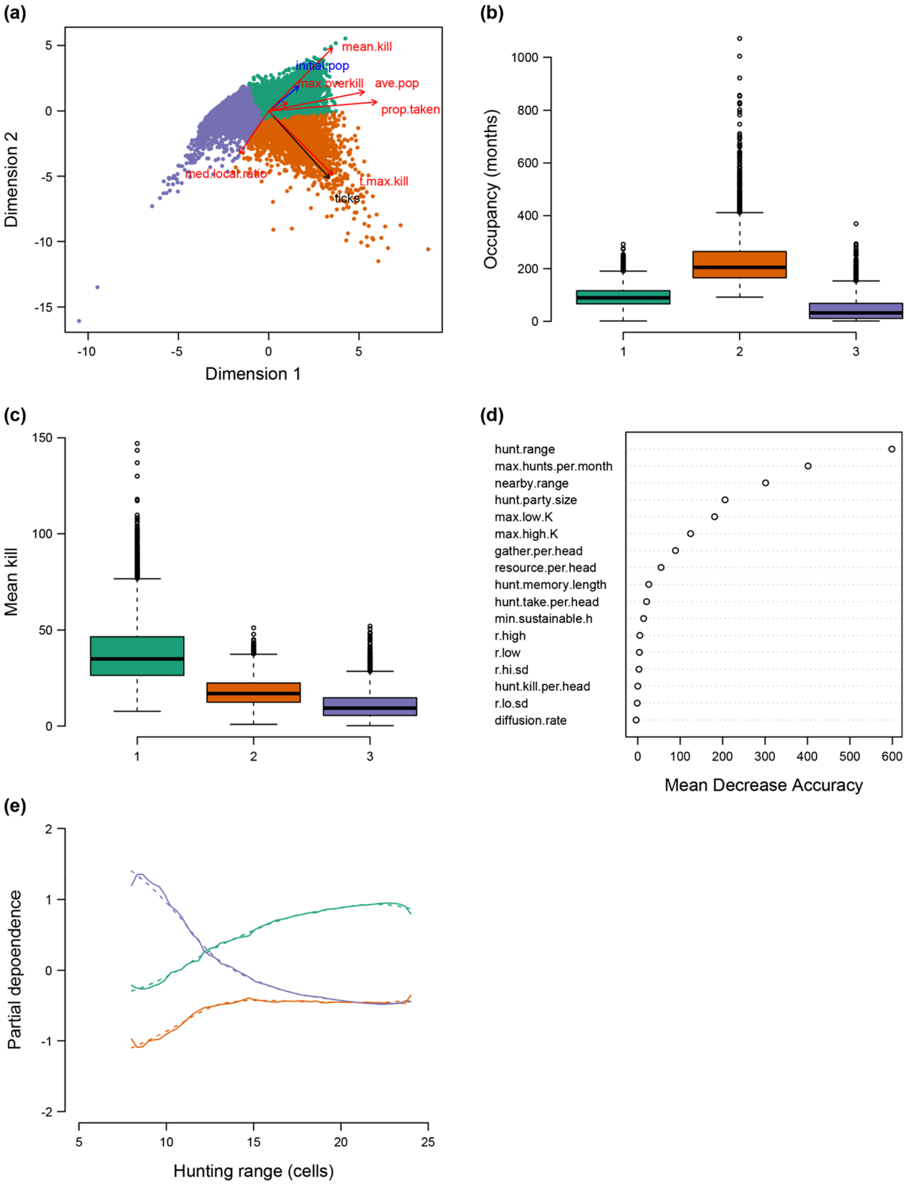


**Fig. 7** Summary of a model-based classification of 25,000 model runs, using statistical summaries of each, with island resource demography and human search simultaneously evaluated using a Latin hypercube, shown here reduced to two dimensions *via* MDS. **a** MDS ordination of the model runs with vectors as per Fig. 3, **b** and **c** boxplots showing length of occupation and mean take by cluster, respectively, **d** importance of predictor variables (model parameters) in splitting the clusters and **e** partial dependency plot for the most important predictor (hunting range), with dashed lines showing a smoothed fit. Colours represent clusters identified *via* *k*-means clustering and the optimal number of clusters *via* the Caliński-Harabasz index; clusters here are not the same as those in Fig. 8

(nearby.range). Thus, for the summary-based classifications, parameters related to human behaviour are the most important, whereas for the DTW-based trajectory classification, a combination of resource demography and human behaviour is

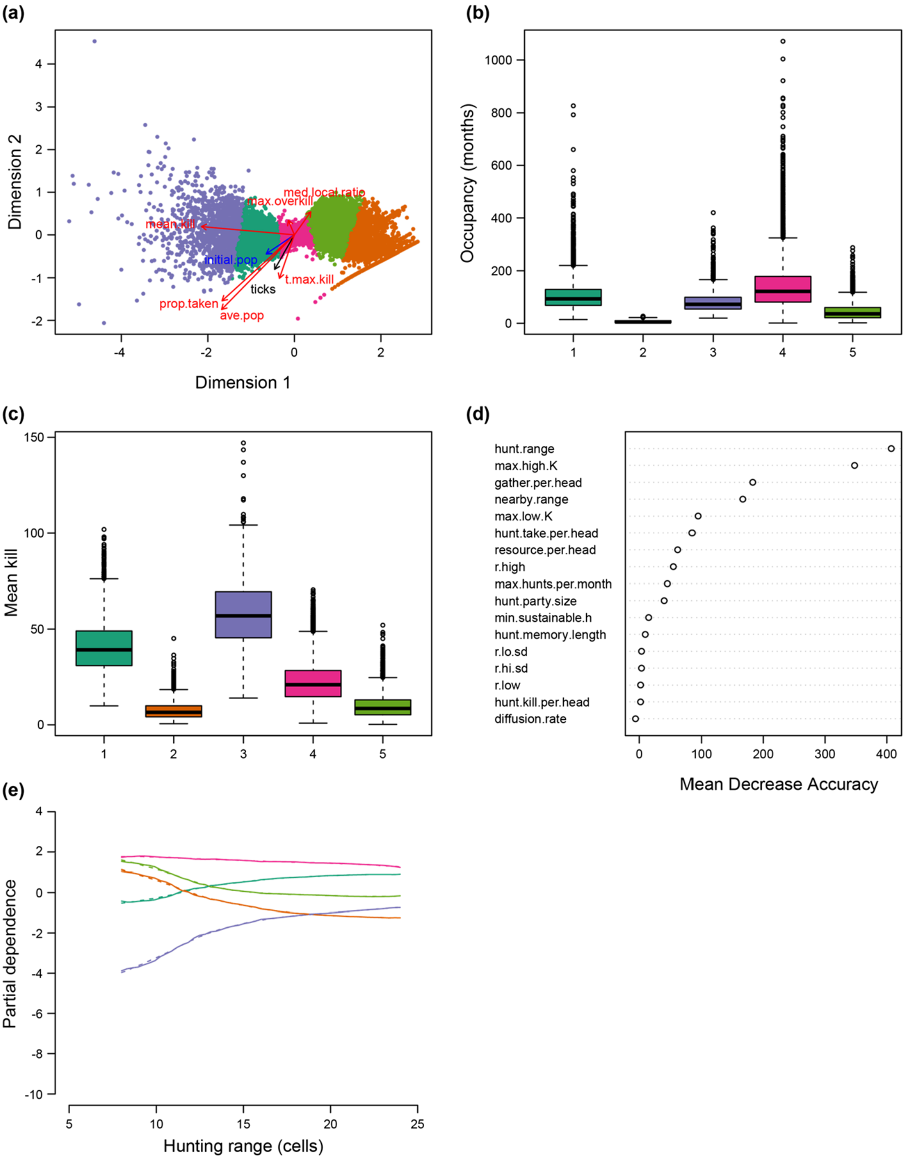


**Fig. 8** Summary of a model-based classification of 25,000 model runs, using DTW for each trajectory, with island resource demography and human search simultaneously evaluated using a Latin hypercube, shown here reduced to two dimensions *via* MDS. **a** MDS ordination of the model runs with vectors as per Fig. 5, **b** and **c** boxplots showing length of occupation (months) and mean kill by cluster, respectively, **d** importance of predictor variables (model parameters) in splitting the clusters and **e** partial dependency plot for the most important predictor (hunting range), with dashed lines a smoothed fit. Colours represent clusters identified *via* *k*-means clustering and the optimal number of clusters *via* the Caliński-Harabasz index; clusters here are not the same as those in Fig. 8

important. However, where only resource demography is evaluated, the high-value resource's carrying capacity is overwhelmingly important in both classifications (Fig. SM 3 and 4). When human search behaviour is considered alone then in the summary-based classification hunting range, the maximum number of hunts per month and the nearby range are the top three predictors, whereas for the DTW-based classification, hunting range, resource gathered per head and nearby foraging range are the most important. Interestingly when both sets of parameters are varied, parameters related to resource demography become relatively more important than when they are varied with fixed exploitation parameters; this suggests interactions between parameters related to humans and those related to the environment.

The RF analyses identify the importance of each parameter in isolation, but partial dependence plots (Fig. 7e and 8e) show the marginal effects of a given parameter (that is the effect of a predictor while holding all others fixed) on the probability of an outcome being placed into a specific class. For the most part, the partial dependence plots suggest the marginal effects are intuitive in their direction. They do not show parameter values at which the probability of belonging to different clusters (or narratives) change abruptly suggesting a lack of clear thresholds in outcomes—this reflects the way that the clusters intergrade into each other rather than being discrete.

## Discussion

### The Analytic Approach

Evaluation is central to any modelling exercise, and can take the form of model versus data confrontation or it might focus on evaluating how much we have learned or communicated with the model of interest (Barton 2014; Bennett et al. 2013). At its heart though, any model evaluation is concerned with closing the surrogative reasoning loop: that is, relating the model to the theoretical and/or empirical context(s) it is grounded in. There is little doubt that modern methods of statistical inference for stochastic processes have a valuable role to play in parameterising and analysing simulation models such as ABMs (Hartig et al. 2011; Perry et al. 2016; van der Vaart et al. 2015). In some settings, however, the concern is not solely with pattern matching but is, instead, with identifying typical model dynamics and making robust inferences about them to inform theory development and data collection. Furthermore, the targets that a model is evaluated against may be qualitative rather than quantitative. Our approach draws on aspects of the narrative analytic approach but in a way that is rather different from that articulated in the social science literature. The similarity is that we are seeking to establish quantitative links between individual case studies and formal theory—we are concerned with "what benefits are, or can be, secured by formalizing verbal accounts?" (Bates et al. 2000, p. 696)—but the difference is that our 'verbal accounts' are the outcomes of experiments conducted with bottom-up simulation models.

Of course, the use of models to generate narratives requires that the model itself is credible (or realistic), both in representation (the processes included and how they are mimicked) and in the parameterisation of that representation. Direct model-data evaluation is challenging for models of archaeological and palaeoenvironmental systems (Biondi 2013; Perry et al. 2016), and, as Barton (2014) argues, it is perhaps better in such settings to see the use of and archaeological and palaeoenvironmental data for model or theory refinement and model-based learning, rather than direct model-data confrontation. The use of models in complexity science has tended to turn the standard inferential chain on its head such that models are used to generate theory rather than *vice versa* (Dowling 1999; O'Sullivan 2004)—in this way we can potentially 'coax experiments from history' (Deevey 1969) using simulation models. Unfortunately in many ABM-based projects, the closure of the surrogative reasoning loop has tended to rely on pattern matching and has run the (seductive) risk of affirming the consequent (*i.e.* our model produces pattern *x* and the system of interest shows pattern *x*; therefore, the explanatory mechanisms in the model are potentially those driving the system). It is also the case that many evaluations of ABMs rely on emergent macroscale patterns (*e.g.* population size or settlement location) rather than agent-level phenomena such as narrative trajectories (see Millington et al. 2012). Visualisation is important in testing, evaluating and communicating any model (Dorin and Geard 2014; Grimm 2002; Lee et al. 2015) and can help to identify model trajectories and narratives. However, complex bottom-up models, such as ABMs, tend to be highly dimensional and show rich spatio-temporal dynamics. There are advantages in using developments in data science to identify qualitative classes of model behaviour (*e.g.* narratives) that emerge from the model; such classification provides a bridge to the *in silico* narratives advocated by Millington et al. (2012) and McGlade (2014).

## Inference from Narrative

As Griffin (1993) notes, narratives do not of themselves constitute causal understanding and, indeed, a narrative device need not contain causality to be successful as such. Earlier, we argued that in some model evaluation contexts, the "concern is not with quantitatively rigorous pattern matching but with identifying typical model dynamics and making robust inferences from them". Our position echoes the view of Barton et al. (2012) that the application of models to the types of systems we consider may be better suited to the development of generalities rather than site- or culture-specific reconstructions (although see Evans et al. 2013 for a well-argued case for more complicated site-specific models). While we have isolated the types of dynamics our ABM can generate, we have not broached the types of inferences we might make from this process. Developing these inferences will be challenging because it requires assessing the types and dynamics of settlement at a particular place and time. Clearly having identified a suite of candidate narratives and shown the types of conditions under which they might occur, the necessary next step is to ground them in relevant empirical information and, ultimately, crystallise them in theory.

While there is a rich literature on quantitative model-data comparison, it tends to give primacy to contemporary observational data and how (best) to confront qualitative model outcomes with qualitative data is less clear (Bennett et al. 2013); this is, again, an area where environmental modellers may need to turn to the social sciences for inspiration. If by focussing on trajectories not just endpoints or statistical summaries we consider the output of simulation models as narratives, then some of the approaches of comparative history become useful (Lange 2013). Mahoney (2003, p. 11) describes narrative analysis as an approach for comparative analysis of historical cases, which "is fundamentally concerned with explanation and the identification of causal configurations that produce major outcomes of interest". In a model evaluation context, pattern-oriented modelling (POM; Grimm and Railsback 2012) is concerned with using *multiple* observed patterns to act as filters on model structure and parameterisation (*i.e.* the model configurations that yield interesting dynamics); these patterns are typically quantitative. In a qualitative context, the filters used in a POM exercise will be crucial and must include not just the end of the story (*i.e.* the 'final' spatio-temporal outcome) but also the order and position of events in time and space. Developing narrative-type approaches to model evaluation will help to provide the link between simulation and qualitative methods advocated by Millington and Wainwright (2016).

As a shallow example of how a pattern-oriented comparative narrative analysis might proceed, we could compare the contrasting outcomes our model produces with the dynamics recorded in the palaeoenvironmental and archaeological records. Looking at the narratives in Figs. 3, 4 and 5, two broad types of settlement history emerge: (i) prolonged settlements with high-value resources exploited effectively and (ii) short-duration settlements with high-value resources left more or less unexploited. Examples of both dynamics can be found in the archaeological literature describing the settlement of the islands and landscapes of remote Oceania in the late Holocene. For example, southern New Zealand, where traditional Polynesian agriculture was not possible due to the cool climate, typifies the rapid, if short-term, exploitation of high-value food resources (Rawlence et al. 2015; Smith 2013), and the Pitcairns group (south-eastern Pacific) a less successful colonisation (Weisler 1995). However, the palaeoenvironmental and archaeological records suggest that failed or impermanent settlements tend to be associated with very specific circumstances such as extreme remoteness, depauperate pre-settlement ecosystems and/or unfavourable climatic conditions (Allen and Wallace 2007; Anderson 2002; Nunn and Britton 2001). These controls operate at spatio-temporal scales that our model does not consider. Furthermore on such islands, humans may have deliberately changed their behaviours and decision-making to avoid the risk of ecosystem collapse (Rolett 2008). There are few, if any, instances of 'suitable' landscapes not being successfully settled or exploited. So what might explain this discrepancy? Our model considers just a single settlement point in the landscape (or island), albeit one with some resources available, and it seems more likely that dispersing humans would systematically explore newly found landscapes; or, perhaps more charitably, we might interpret the failed settlements as representing advance parties, especially given that no virtual island proved 'immune' to settlement. This is a case where both the model and the data are under-determined. Certainly, however, model-based narrative analyses will need to be more rigorous and thorough than that we present here and must focus on the ordering and interdependency of events—both of which are central to a narrative approach and historical explanation—rather than just their outcome. Developing a framework for model-derived narrative analysis is challenging, not least because of the potentially large amounts of data involved, but approaches such as those described by Abell (2007) may represent useful starting points. If outcomes are the sole interest, then static data-driven models may be a more fruitful way to reconstruct the endpoints of ancient human-environment interactions (*e.g.* see the approach adopted by Perry et al. 2012).

### Notes on Computational Implementation

The computational workflow we adopted (Fig. 2) has multiple steps and at each of them the analytical methods we used could be substituted by others. We have adopted reasonably familiar methods (classical MDS, PCA, *k*-means clustering) to make our approach as accessible as possible. Central to our analysis is the clustering of massive distance matrices (up to $[2.5 \times 10^3]^2$ elements). Tools for clustering high-dimensional data are developing rapidly (James et al. 2013) and are applicable to the types of questions we consider. While methods such as random forests are becoming widely used to support supervised and semi-supervised classification (*e.g.* Menze and Ur 2012 is a recent archaeological example), they are more rarely used to evaluate the outcomes of ABMs. Dynamic time warping proved an effective way to compare and classify the temporal dynamics of the model experiments that we conducted. There are statistical alternatives to DTW, such as functional data analysis (Ramsay and Silverman 2005), which are worth exploring, not least because DTW is computationally demanding, for data sets of the size we consider.

## Conclusions

Narrative and related methodological approaches (*e.g.* narrative analytics and comparative narrative analysis) provide a way to interpret the dynamics of complex bottom-up simulation models. Narrative approaches are usually applied where there are few case studies to work with (small *n*), but in a model-based setting there may be massive amounts of data to sift through (large *n*). Machine learning approaches provide powerful ways to identify narratives from individual trajectories and to isolate the circumstances under which they might arise. The challenge then lies in developing causal inferences regarding those narratives and their plausibility. The approach we describe is likely most appropriate for empirically inaccessible (or difficult to access) cases in the medium- to far-past or future.

**Acknowledgements** This manuscript has benefited from discussions with James Millington, John Wainwright, Matt McGlone, Janet Wilmshurst, Jamie Wood, Tom Etherington and Ben Davies. We are grateful for funding support from the University of Auckland Faculty Research Development Fund (3702237). The authors wish to acknowledge the contribution of NeSI high-performance computing facilities to the results of this research. NZ's national facilities are provided by the NZ eScience Infrastructure and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation & Employment's Research Infrastructure programme. URL is https://www.nesi.org.nz.

## References

Abell, P. (2007). Narratives, Bayesian narratives and narrative actions. *Sociologica, 3*, 3/2007. https://doi.org/10.2383/25959.

Abell, P. (2009). A case for cases: comparative narratives in sociological explanation. *Sociological Methods & Research, 38*(1), 38–70. https://doi.org/10.1177/0049124109339372.
Allen, M. S., & Wallace, R. (2007). New evidence from the east Polynesian gateway: substantive and methodological results from Aitutaki, southern Cook Islands. *Radiocarbon, 49*, 1163–1179.
Anderson, A. J. (2002). Faunal collapse, landscape change and settlement history in Remote Oceania. *World Archaeology, 33*(3), 375–390.
Axtell, R. L., Epstein, J. M., Dean, J. S., Gumerman, G. J., & Swedlung, A. C. (2002). Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. *Proceedings of the National Academy of Sciences (USA), 99*, 7275–7279. https://doi.org/10.1073/pnas.092080799.
Barton, C. M. (2014). Complexity, social complexity, and modeling. *Journal of Archaeological Method and Theory, 21*(2), 306–324. https://doi.org/10.1007/s10816-013-9187-2.
Barton, C. M., Bernabeu, J., Aura, J. E., Garcia, O., Schmich, S., & Molina, L. (2004). Long-term socioecology and contingent landscapes. *Journal of Archaeological Method and Theory, 11*(3), 253–295.
Barton, C. M., Ullah, I. I. T., & Mitasova, H. (2010). Computational modeling and neolithic socioecological dynamics: a case study from southwest Asia. *American Antiquity, 75*(2), 364–386. https://doi.org/10.7183/0002-7316.75.2.364.
Barton, C. M., Ullah, I. I. T., Bergin, S. M., Mitasova, H., & Sarjoughian, H. (2012). Looking for the future in the past: long-term change in socioecological systems. *Ecological Modelling*. https://doi.org/10.1016/j.ecolmodel.2012.02.010.
Bates, R. H., Greif, A., Levi, M., & Rosenthal, J. P. (1998). *Analytic narratives*. Princeton: Princeton University Press.
Bates, R. H., Greif, A., Levi, M., Rosenthal, J. P., & Weingast, B. (2000). The analytical narrative project. *The American Political Science Review, 94*(3), 696–702. https://doi.org/10.2307/2585843.
Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., et al. (2013). Characterising performance of environmental models. *Environmental Modelling & Software, 40*, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011.
Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources, 16*, 41–51.
Biondi, F. (2013). The fourth dimension of interdisciplinary modeling. *Journal of Contemporary Water Research & Education, 152*(1), 42–48. https://doi.org/10.1111/j.1936-704X.2013.03166.x.
Biondi, F. (2014). Paleoecology—grand challenge. *Frontiers in Ecology and Evolution, 2*. https://doi.org/10.3389/fevo.2014.00050.
Boschetti, F., Walker, I., & Price, J. (2016). Modelling and attitudes towards the future. *Ecological Modelling, 322*, 71–81. https://doi.org/10.1016/j.ecolmodel.2015.11.009.
Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.
Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics, 3*(1), 1–27. https://doi.org/10.1080/03610927408827101.
Carpenter, S. R. (2002). Ecological futures: building an ecology of the long now. *Ecology, 83*, 2069–2083.
Chapelle, O., Schölkopf, B., & Zien, A. (Eds.). (2006). *Semi-supervised learning*. Cambridge: MIT Press.
Cleland, C. E. (2001). Historical science, experimental science, and the scientific method. *Geology, 29*(11), 987–990. https://doi.org/10.1130/0091-7613(2001)029%3C0987:hsesat%3E2.0.co;2.
Cleland, C. E. (2011). Prediction and explanation in historical natural science. *The British Journal for the Philosophy of Science, 62*, 551–582. https://doi.org/10.1093/bjps/axq024.
Contessa, G. (2007). Scientific representation, interpretation, and surrogative reasoning. *Philosophy of Science, 74*(1), 48–68. https://doi.org/10.1086/519478.
Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology, 88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1.
Dalby, S. (2016). Framing the Anthropocene: the good, the bad and the ugly. *The Anthropocene Review, 3*(1), 33–51. https://doi.org/10.1177/2053019615618681.
Deevey, E. S. (1969). Coaxing history to conduct experiments. *Bioscience, 19*, 40–43.
Dorin, A., & Geard, N. (2014). The practice of agent-based model visualization. *Artificial Life, 20*(2), 271–289. https://doi.org/10.1162/ARTL_a_00129.
Dowling, D. (1999). Experimenting on theories. *Science in Context, 12*, 261–273.
Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation, 11*(4), 12.
Evans, M. R., Grimm, V., Johst, K., Knuuttila, T., de Langhe, R., Lessells, C. M., et al. (2013). Do simple models lead to generality in ecology? *Trends in Ecology & Evolution, 28*(10), 578–583. https://doi.org/10.1016/j.tree.2013.05.022.
Gerring, J. (2017). Qualitative methods. *Annual Review of Political Science, 20*(1), 15–36. https://doi.org/10.1146/annurev-polisci-092415-024158.

Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in *R*: the dtw package. *Journal of Statistical Software, 31*(7). 10.18637/jss.v031.i07.

Griffin, L. J. (1992). Temporality, events, and explanation in historical sociology: an introduction. *Sociological Methods & Research, 20*(4), 403–427.

Griffin, L. J. (1993). Narrative, event-structure analysis, and causal interpretation in historical sociology. *American Journal of Sociology, 98*(5), 1094–1133. https://doi.org/10.2307/2781584.

Grimm, V. (2002). Visual debugging: a way of analyzing, understanding and communicating bottom-up simulation models in ecology. *Natural Resource Modelling, 15*, 23–38.

Grimm, V., & Railsback, S. F. (2012). Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1586), 298–310.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., et al. (2005). Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science, 310*, 987–991. https://doi.org/10.1126/science.1116681.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: a review and first update. *Ecological Modelling, 221*(23), 2760–2768. https://doi.org/10.1016/j.ecolmodel.2010.08.019.

Hartig, F., Calabrese, J. M., Reineking, B., Wiegand, T., & Huth, A. (2011). Statistical inference for stochastic simulation models—theory and application. *Ecology Letters, 14*(8), 816–827. https://doi.org/10.1111/j.1461-0248.2011.01640.x.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R*. New York: Springer.

Kintigh, K., Altschul, J., Beaudry, M., Drennan, R., Kinzig, A., Kohler, T., et al. (2014). Grand challenges for archaeology. *American Antiquity, 79*(1), 5–24. https://doi.org/10.7183/0002-7316.79.1.5.

Kirch, P. V. (2010). Peopling of the Pacific: a holistic anthropological perspective. *Annual Review of Anthropology, 39*(1), 131–148. https://doi.org/10.1146/annurev.anthro.012809.104936.

Knaus, J. (2013). snowfall: easier cluster computing (based on snow). http://CRAN.R-project.org/package=snowfall.

Kohler, T. A., Bocinsky, R. K., Cockburn, D., Crabtree, S. A., Varien, M. D., Kolm, K. E., et al. (2012). Modelling prehispanic Pueblo societies in their ecosystems. *Ecological Modelling*. https://doi.org/10.1016/j.ecolmodel.2012.01.002.

Lange, M. (2013). *Comparative-historical methods*. London: SAGE Publications.

Lee, J.-S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., et al. (2015). The complexities of agent-based modeling output analysis. *Journal of Artificial Societies and Social Simulation, 18*(4). 10.18564/jasss.2897.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news, 2*(3), 18–22.

Lorscheid, I., Heine, B.-O., & Meyer, M. (2012). Opening the 'black box' of simulations: increased transparency and effective communication through the systematic design of experiments. *Computational and Mathematical Organization Theory, 18*(1), 22–62. https://doi.org/10.1007/s10588-011-9097-3.

Mahoney, J., & Rueschemeyer, D. (2003). Comparative historical analysis: achievements and agendas. In J. Mahoney & D. Rueschemeyer (Eds.), *Comparative historical analysis in the social sciences* (pp. 3–41). Cambridge; New York: Cambridge University Press.

March, J. G., Sproull, L. S., & Tamuz, M. (1991). Learning from samples of one or fewer. *Organization Science, 2*(1), 1–13. https://doi.org/10.1287/orsc.2.1.1.

Mayer, D. G., & Butler, D. G. (1993). Statistical validation. *Ecological Modelling, 68*, 21–31.

McGlade, J. (2014). Simulation as narrative: contingency, dialogics, and the modeling conundrum. *Journal of Archaeological Method and Theory, 21*(2), 288–305. https://doi.org/10.1007/s10816-014-9201-3.

McIntire, E. J. B., & Fajardo, A. (2009). Beyond description: the active and effective way to infer processes from spatial patterns. *Ecology, 90*, 46–56.

Menze, B. H., & Ur, J. A. (2012). Mapping patterns of long-term settlement in Northern Mesopotamia at a large scale. *Proceedings of the National Academy of Sciences, 109*(14), E778–E787. https://doi.org/10.1073/pnas.1115472109.

Millington, J. D. A., & Wainwright, J. (2016). Mixed qualitative-simulation methods: understanding geography through thick and thin. *Progress in Human Geography*. https://doi.org/10.1177/0309132515627021.

Millington, J. D. A., O'Sullivan, D., & Perry, G. L. W. (2012). Model histories: narrative explanation in generative simulation modelling. *Geoforum, 43*, 1025–1034.

Morgan, M. S. (2001). Models, stories and the economic world. *Journal of Economic Methodology, 8*, 361–384.

Niu, S., Luo, Y., Dietze, M. C., Keenan, T. F., Shi, Z., Li, J., & Chapin III, F. S. (2014). The role of data assimilation in predictive ecology. *Ecosphere, 5*(5), art65. https://doi.org/10.1890/ES13-00273.1.

Nunn, P. D., & Britton, J. M. R. (2001). Human-environment relationships in the Pacific Islands around A.D. 1300. *Environment and History, 7*(1), 3–22. https://doi.org/10.3197/096734001129342388.

O'Sullivan, D. (2004). Complexity science and human geography. *Transactions of the Institute of British Geographers, 29*, 282–295. https://doi.org/10.1111/j.0020-2754.2004.00321.x.

O'Sullivan, D., & Haklay, M. (2000). Agent-based models and individualism: is the world agent-based? *Environment and Planning A, 32*(8), 1409–1425.

O'Sullivan, D., & Perry, G. L. W. (2013). *Spatial simulation: exploring pattern and process*. Chichester: John Wiley & Sons.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., O'Hara, R.B., Simpson, G.L., et al. (2017). vegan: community ecology package, v 2.4–3.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science, 263*, 641–646.

Peck, S. L. (2004). Simulation as experiment: a philosophical reassessment for biological modeling. *Trends in Ecology and Evolution, 19*, 530–534. https://doi.org/10.1016/j.tree.2004.07.019.

Perry, G. L. W., Wilmshurst, J. M., McGlone, M. S., & Napier, A. (2012). Reconstructing spatial vulnerability to forest loss by fire in pre-historic New Zealand. *Global Ecology and Biogeography, 21*, 1029–1041. https://doi.org/10.1111/j.1466-8238.2011.00745.x.

Perry, G. L. W., Wainwright, J., Etherington, T. R., & Wilmshurst, J. M. (2016). Experimental simulation: using generative modeling and palaeoecological data to understand human-environment interactions. *Frontiers in Ecology and Evolution, 4*, 109. https://doi.org/10.3389/fevo.2016.00109.

Prowse, T. A. A., Bradshaw, C. J. A., Delean, S., Cassey, P., Lacy, R. C., Wells, K., et al. (2016). An efficient protocol for the global sensitivity analysis of stochastic ecological models. *Ecosphere, 7*(3), e01238. https://doi.org/10.1002/ecs2.1238.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis (2nd ed.)*. New York: Springer.

Rawlence, N. J., Perry, G. L. W., Smith, I. W. G., Scofield, R. P., Tennyson, A. J. D., Matisoo-Smith, E. A., et al. (2015). Radiocarbon-dating and ancient DNA reveal rapid replacement of extinct prehistoric penguins. *Quaternary Science Reviews, 112*, 59–65. https://doi.org/10.1016/j.quascirev.2015.01.011.

R-Development-Core-Team. (2016). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org.

Rolett, B. V. (2008). Avoiding collapse: pre-European sustainability on Pacific Islands. *Quaternary International, 184*(1), 4–10. https://doi.org/10.1016/j.quaint.2007.10.016.

Smith, I. (2013). Pre-European Maori exploitation of marine resources in two New Zealand case study areas: species range and temporal change. *Journal of the Royal Society of New Zealand, 43*(1), 1–37. https://doi.org/10.1080/03036758.2011.574709.

Steadman, D. W. (1995). Prehistoric extinctions of Pacific Island birds: biodiversity meets zooarchaeology. *Science, 267*(5201), 1123–1131. https://doi.org/10.1126/science.267.5201.1123.

Stein, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics, 29*(2), 143–151.

Thiele, J. C. (2014). R marries NetLogo: introduction to the RNetLogo package. *Journal of Statistical Software, 58*(2), 1–41. 10.18637/jss.v058.i02.

Topping, C. J., Alrøe, H. F., Farrell, K. N., & Grimm, V. (2015). *Per aspera ad astra*: through complex population modeling to predictive theory. *The American Naturalist, 186*(5), 669–674. https://doi.org/10.1086/683181.

van der Vaart, E., Beaumont, M. A., Johnston, A. S. A., & Sibly, R. M. (2015). Calibration and evaluation of individual-based models using approximate Bayesian computation. *Ecological Modelling, 312*, 182–190. https://doi.org/10.1016/j.ecolmodel.2015.05.020.

Weisler, M. I. (1995). Henderson Island prehistory: colonization and extinction on a remote Polynesian island. *Biological Journal of the Linnean Society, 56*(1–2), 377–404. https://doi.org/10.1111/j.1095-8312.1995.tb01099.x.

Wilensky, U. (1999). *NetLogo*. Evanston: Center for Connected Learning and Computer-Based Modeling, Northwestern University.

Winsberg, E. (2010). *Science in the age of computer simulation*. Chicago: Chicago University Press.