Finding Environmental Opportunities for Early Sea Crossings:
An Agent-Based Model of Middle to Late Pleistocene Mediterranean Coastal Migration

by

Bailey Baumann

A Thesis Presented to the
Faculty of the USC Graduate School
University of Southern California
In Partial Fulfillment of the
Requirements for the Degree
Master of Science
(Geographic Information Science and Technology)

August 2017

Table of Contents

List of Figures

List of Code Chunks

List of Abbreviations

ABM   Agent-based model/ modeling

ASCII   American Standard Code for Information Interchange

BCE   Before the Common Era

CE   Common Era

GIS   Geographic information System

HSL   High sea level

LSL   Low sea level

MSL   Median sea level

NaN   Not a number

OAT   One (variable) at a time

ODD   Overview, Design concepts, and Details

POM   Pattern-oriented modeling

SA   Sensitivity analysis

YBP   Years before present ("present" is 1950)

Abstract

This research hypothesizes that a data-rich, geographically explicit agent-based model can provide context for archaeological finds when the archaeological record itself is too incomplete or damaged to do so. It specifically seeks to address the problem posed by disparate but mounting evidence of earlier than expected sea crossings in the Mediterranean. Hundred-thousand-year-old lithic evidence of human presence on islands encourages the revisionist view that the Pleistocene Mediterranean was less of a barrier and more of a facilitator for travel than previously thought. Nevertheless, it fails to answer Mediterranean archaeologists' questions about how and why. This research shows how an agent-based model can be designed to allow archaeologists to formulate and test theories about the ways the environment could have created opportunities for early sea crossings. It demonstrates the process of designing and building this model in R and NetLogo. Preliminary results show that this model can be used to help archaeologists better understand the revisionist conceptual model of sea crossings in the Pleistocene Mediterranean.

Chapter 1 Introduction

Mediterranean archaeology faces the challenge of disparate but mounting evidence of earlier than expected human presence on islands in the region. Unfortunately, the archaeological record is currently too damaged and incomplete to offer sufficient context for these surprising finds. Computer simulations, specifically agent-based models, can help archaeologists formulate and test theories when the archaeological record alone cannot.

Agent-based models (ABM) allow complex systems to be modeled from the bottom-up. Data-rich, geographically explicit ABM are able to represent past systems even more faithfully. The output of these models can be leveraged to draw conclusions about the real systems they represent. In this way, agent-based models can inform archaeology's discussion of the cognitive abilities of these early seafarers. However, this study takes on a more pressing and immediate question: "How can we find places where the environment created opportunities for early sea crossings?" This study develops a process for implementing a data-rich, geographically explicit ABM of coastal migration in the Middle to Late Pleistocene Mediterranean that can be used to formulate and test theories about the ways past environments created opportunities for sea crossings.

To set the stage for this investigation of opportunities for early sea crossings, this chapter provides an overview of the evidence for Pleistocene seafaring in the Mediterranean. It explains the importance of lithic evidence from oceanic islands like Crete. It also describes environmental patterns from the Pleistocene that could have created opportunities for migration. Finally, to situate the research method in context, the growing role of agent-based modeling in archaeology and related research is discussed.

1.1. Words Matter

Often, word choice carries undefined assumptions about the nature of an activity or objective. For example, it remains unclear whether ancient voyages, especially longer ones, were deliberate or accidental; the product of *navigation* or *drift*. Cyprian Broodbank is among the archaeologists who have contrasted "tentative seagoing" with "more adept seafaring" (Broodbank 2006, 200). Importantly, while in some cases the term "seafaring" might suggest greater navigational abilities, in this research, movement over both land and sea is represented by random walkers (agents) who do not plan voyages. These agents base their decisions to move from one cell to the next on the favorable and unfavorable qualities of their immediate surroundings.

A distinction should be made between evidence of human *presence* in certain places and evidence of more permanent *settlement*, even, potentially deliberate, *colonization*. For example, there may be evidence suggesting only occasional human presence, but it is still possible that the Middle to Late Pleistocene did see sustained colonization on Mediterranean islands (Leppard 2014). Accordingly, this project focuses on contextualizing evidence of human presence without attempting to address the duration of that presence or the question of how more permanent settlements might have been established.

A final term that requires parsing is *coastal migration*. Thematically, the focus of this analysis is earlier than expected sea crossings. However, the term coastal migration is used in this study to refer to movement over both land and sea, since this project does consider movement over land as it relates to possible sea crossings.

1.2. Human Migration During the Middle to Late Pleistocene

The Pleistocene was a climatic phase that saw frequent fluctuations in global temperatures and, as a result, significant milestones in early human migration and adaptation, including some of the first sea crossings. This project pays special attention to the reconstruction of Pleistocene environments, as

previous archaeological research may have not sufficiently considered the role of environmental factors in human dispersal (Palombo 2010).

The Pleistocene Epoch divides into three Sub-Epochs: Early (2,588,000 – 781,000 ybp), Middle (781,000 – 126,000 ybp), and Late (126,000 – 12,000 ybp) (Subcommission on Quaternary Stratigraphy 2010). The divisions of the Pleistocene are informed by glacial periods, but the end of the Pleistocene does roughly correspond with the end of the Paleolithic in archaeology. The Pleistocene is followed by the Holocene and, in archaeology, the Paleolithic transitions into the Neolithic.

Over the last three million years, climate change has significantly affected the course of both human evolution and migration. Around 800,000 ybp, during transition from the Early to Middle Pleistocene, the final of three major shifts in climate corresponded with a third wave of human migration out of Africa. Possibly responding to environmental pressures, such as the increased aridity in Northeast Africa, people moved toward Eurasia, which saw only comparatively moderate deteriorations in climate during this time (Almogi-Labin 2011).

Changes in cycles of aridity can be read in the content of sediment samples. For example, Larrasoana et al. (2003) provides three million years of North African monsoon records using magnetic properties indicative of Aeolian dust levels in sediment collected off the coast of Cyprus (Lat/Lon: 34.07 N/32.72 W). With an average temporal resolution of 400 years, these records show how closely climate and vegetation were connected, and can lead to a greater understanding of how changes in climate and vegetation may have together driven human adaptation and migration. For this model, the Larrasoana data were used to indicate the environmental pressures affecting levels of motivation for migration in the Mediterranean. Larrasoana et al. (2003) was selected because of its temporal extent (three million years) and temporal resolution (400 years). Because these sediment samples were collected off the coast of Cyprus, this is also a spatially relevant data set for this research.

Lithic evidence and physical remains dating to the late Early Pleistocene have been identified at sites in Spain, Italy, and France. They support the theory that the earliest populating of Europe began roughly 1,300,000 – 1,500,000 ybp (Palombo 2010). Barranco Leon, in southeast Spain, is particularly interesting because patterns of occupation at this site suggest early humans favored warmer and more humid conditions. Especially considering the proximity of water to this site, it can be used to argue for the importance of climate and environment (not just physiology and culture) in the migration of hominins into Europe (Agustí et al. 2015).

The archaeological record supports the theory that the first major cold stage of the Pleistocene (~870,000 ybp) and the concurrent changes in aridity may have prompted a wave of human migration into Europe. The dating of hominin remains and lithic evidence from sites in Southern Europe suggests a significant period of hominin presence approximately 990,000 – 780,000 ybp (Muttoni et al. 2010). Major changes in climate affected the size and distribution of fauna. Accordingly, the transition from the Early to Middle Pleistocene saw changes in temperature and aridity and a faunal renewal that provided a greater abundance of suitable prey for hominins in Europe. The "significant peopling" of the continent began during this period (Palombo 2010, 179). Moreover, using elephant remains at key archaeological sites dating to this time, it can even be argued that these hunter-gatherers could have followed the large the herbivores out of Africa, across the Levant, and around the Mediterranean Basin (Muttoni et al. 2010).

1.3. Earlier Than Expected Seafaring in the Mediterranean

The revisionist model of early seafaring in the Mediterranean considers how the sea may have been a facilitator rather than a barrier for early humans (Leppard 2015). This is a reversal of previous thinking. For example, Broodbank had used nearly the same language (i.e. "bridge" versus "barrier") to characterize a far more strained relationship between humans and the sea during the Pleistocene (Broodbank 2006, 208). Mediterranean archaeologists had widely thought seafaring during the

Pleistocene would have been rare. If sea crossings occurred at all, they may have been across the Strait of Gibraltar (Straus 2001). Now, different ways of thinking about the relationship between humans and the sea have gained traction due to an increase in evidence—particularly lithic evidence—that can be used to support earlier timelines for the beginnings of seafaring in the region.

The discovery of lithic evidence indicating human presence on Crete as early as 130,000 years ago (Strasser et al. 2010) surprised Mediterranean archaeologists because it made the strongest case yet that humans were capable of seafaring before the Holocene. Previously, in the absence of securely dated archaeological evidence on islands that could not have been connected to land in the last three million years, it had been difficult to argue that seafaring began in the Mediterranean as early as it had begun in the Pacific (Bednarik 2003; Broodbank 2006; Cherry 1990; Erlandson 2010).

Experimental archaeology provides insight into possible designs of some of the earliest watercraft (Cherry and Leppard 2015; Bednarik 1998). However, in the absence of wrecks dated to this time, the presence of artifacts dating to the Pleistocene on islands that were separated from the mainland during that time can be used as proxy for evidence of seafaring (Howitt-Marshall and Runnels 2016). The Aegean in particular offers such evidence on islands suggesting early sea crossings (Howitt-Marshall and Runnels 2016, Papoulia 2016, Laskaris et al. 2011, Runnels 2014). The Aegean does present special challenges, as it is possible some of its islands, such as the Cyclades, could have been connected to land at various times during the Pleistocene. Evidence from Crete, Gavdos, and Cyrpus is therefore especially important because these are oceanic islands that were not connected to land at any point during the Pleistocene (Papoulia 2016). Despite encouraging signs from the lithic evidence, further evidence and more secure dating of the current evidence is needed to better establish a Pleistocene presence of humans on islands in the Mediterranean (Strasser et al. 2010; Runnels 2014).

Much of the evidence for Pleistocene presence on Mediterranean islands comes from the Aegean, suggesting this landscape may have had some particular quality that favored seafaring.

Broodbank (2013) considers how long straight coastlines are less likely to encourage travel by sea than

denser coastlines and offers that the mingling of land and sea provides clues to the locations of early

seagoing hotspots. He builds on the work of Colin McEvedy and John Woodcock (2002) who illustrate

how the concentration of coastline in the Aegean could have formed an important "littoral community."

The work for this project and others like it is to determine how to correct the oversight

characterized by Jon McVey Erlandson below:

> Reading the archaeological literature of the twentieth century, one could easily conclude that
> fishing, seafaring, and maritime migrations played little or no role in human history for the first
> 99.9 percent of the 2.5 million years our genus has existed. Anthropological theory was
> dominated by the notion that our ancestors largely ignored the oceans and other aquatic
> habitats until the end of the Pleistocene (Erlandson 2010, 19).

While avoiding assumptions requiring skillful navigation or deliberate colonization occurring

earlier than the data suggest, it is necessary to allow the sea a greater role in our understanding of

Pleistocene coastal migration. The work of this research was to assist in this endeavor by simulating

Mediterranean coastal migration that treats the sea as a facilitator rather than a barrier. This study

seeks to answer the question, "How can we find places where the environment created opportunities

for early sea crossings?" It does so by developing a process for designing a data-rich, geographically

explicit ABM of coastal migration in the Middle to Late Pleistocene Mediterranean that can be used to

formulate and test theories about the ways past environments created opportunities for sea crossings.

1.4. Agent-Based Modeling and Archaeology

The growing number of publications featuring archaeological agent-based modeling indicate the

increasing importance of this methodology to the field. Agent-based modeling has the potential to

change the way archeologists understand their data and think about space. By simulating the actions of

individual agents within systems, this type of modeling can allow complex spatial phenomena to be

understood in terms of individual agents. The full impact of agent-based modeling on archaeology

remains to be seen, and it may be tempered by factors such as the strength of the relationship between these models and empirical data (Cegielski 2016).

This study and the question it seeks to answer sit at the nexus of several disciplines. Agent-based models can balance the tendency of disciplines like geography, biology, computer science, and climate studies to reach for simple explanations, with the preference of archaeology and anthropology for complex ones. As models, they offer simplified representations of real-world systems, the analysis of which can afford a better understanding of the variables involved and the relationships between them. They allow for the study of systems from the bottom up, for example from the decision of single person to the migration dynamics of whole populations.

Chapter 2 sets out the theoretical framework for this model. It describes the anthropological theory and spatial science thinking that informed the design of this model. It also details related research. Chapter 3 focuses on the design of the model. It describes the software and data used. It explains how the agents' decision-making algorithm was developed and tested. It then demonstrates the process for creating and testing the modeled environments over which the agents move. Chapter 4 discusses preliminary results and directions for future work. Chapter 5 concludes with the contributions of this research and final thoughts on agent-based modeling software.

Chapter 2 Theoretical Framework

This chapter locates this study within the theoretical traditions of both anthropology and spatial science, with a special interest in discussions within the growing field of early seafaring. This model follows the example of early seafaring research as a whole by borrowing from other traditions and employing more specific paradigms when appropriate.

2.1 Insights from Anthropological Theory

While archaeologists previously limited sea crossings to *Homo sapiens* (Erlandson 2001), newer evidence suggests *Homo neanderthalensis* –possibly even *Homo habilis*—would have been capable of seafaring (Simmons 2014; Erlandson 2010). In fact, for some archaeologists, the current lithic evidence is enough to support the theory that early humans had the technology and skills needed for open sea-crossings (Howitt-Marshall and Runnels 2016). What, then, were the adaptive abilities that hominins would have needed to successfully cross seas? Cognitively, these early humans would have needed to weigh the costs and benefits of crossing distances by sea, possibly to reach places they could not directly observe. They may have needed to communicate, even coordinate with each other. Most importantly, they would have needed to build the watercraft for these journeys (Leppard 2015).

Rather than attempting to determine which species of early humans would have been evolutionarily advanced enough to accomplish different journeys by sea, this project follows d'Errico and Banks (2013) in examining more closely the relationship between cultural adaptation and environmental change. They resist the idea that behavioral modernity is limited to *Homo sapiens* and argue against using it as a primary explanation for the differences in the complexity within the material record. Increasingly, archaeological evidence has shown that *Homo neanderthalensis* exhibited signs of modern behavior even if they were not necessarily physically or cognitively modern. D'Errico and Banks

(2013) explain their approach allows for greater focus to be placed on the mechanisms that lead to different kinds of cultural adaptations instead of the causes.

Anthropological theory, especially hunter-gatherer studies, might benefit from a similar change in focus from causes and their necessary consequences to the mechanisms that govern the underlying processes. Bettigner et al. (2015) make this point in their work on hunter-gatherer research in archaeological theory: "The problem with much of contemporary hunter-gatherer archaeology is not that it lacks a general theory of culture but that the theory rests on generalizations about consequences" (Bettigner et al. 2015, 282).

Assumptions about causes and consequences do not allow processes to work independently and be sufficiently understood. For Bettinger et al. (2015) the most successful general theory for hunter-gatherer research is not even a strictly anthropological one. Darwinian models of culture allow for a plurality of outcomes to result from the same causes, because in Darwinian theory, processes are governed by opportunity rather than a larger design.

2.1.1 Thinking Locally and Globally

Because the goal of this research is to produce a model that uses and helps explain empirical evidence as well as ways of thinking about large-scale spatial phenomena, careful consideration was given to the anthropological theory that grounds it. The growing area of seafaring research, like the rest of anthropological theory, experiences a tension between global theories of seafaring and local variations in its development.

Anderson (2010) maintains that to study time, space, agency, and interaction in seafaring, it is best to examine them in terms of broad, even global trends, in accordance with the conceptual model of "macrohistory." And yet, the development of seafaring saw regional variations in timing and technology that should be also accounted for. In the same volume as Anderson, Erlandson cautions that the early history of seafaring saw significant enough variations in environments and the ways people responded

to them to make "effective global or universal generalizations" difficult (Erlandson 2010, 23). For example, seafaring in warm regions with many islands cannot on its own prove that equally early and extensive maritime activity occurred in less accommodating environments.

Similarly, in hunter-gatherer studies, more limited theories tend to be closer to the evidence and easier to operationalize, but they are not always well connected to more general theories about what culture is and how it evolves (Bettinger et al. 2015). For example, optimal foraging theory is one of the limited theories that has been specifically applied to seafaring research but it is not as easily applied to the various circumstances in which early seafaring arose. It makes assumptions about the outcomes of Darwinian processes being necessarily adaptive, which does not always hold true (Anderson 2010; Bettinger et al. 2015).

Therefore, this model follows Bettinger et al. (2015) in primarily relying on a more Darwinian theory of culture instead of general anthropological theory. In the model described in this document, early humans are able to undertake early seafaring when their environment makes it immediately beneficial to them. There is no larger rule written into this model that makes early seafaring more likely for *Homo sapiens*.

Even when studying hunter-gathers and seafaring separately, anthropology must deal with a tension between local and global ways of thinking about these areas of research. Therefore, when studying both areas together, it becomes important to remain aware of the scales on which different theories and evidence exist. This was a consideration throughout the process of designing and testing this model.

2.1.2 Understanding Past Humans in Their Own Terms

It is possible to make inferences about the cognitive and adaptive abilities of early humans using their tools. Different milestones in human evolution can be better understood through the development of stone tool industries (Ambrose 2001; Nowell and Davidson 2010). In this way, archaeology and

anthropology have studied successive groups by comparing them to one another. This approach can be useful. For example, the social evolutionary model of early hunter-gatherers does help us think about how and why human societies changed over time. Different species of early humans likely had different adaptive abilities that enabled them to successfully inhabit various types of environments.

However, as much as possible, this study seeks to understand past humans on their own terms rather than in comparison to their successors. It resists assumptions about the outcomes of evolutionary processes and focuses on simulating the specific types of circumstances in which past environments would have created opportunities to travel, whether by land or sea.

This project focuses first on the past environments of early Mediterranean seafarers. This is one way of attempting to limit the bias of modern scholars, which constrains our ability to understand our data and, at its worst, can be called "temporal chauvinism" (Cherry 1990, 201; Broodbank 2006, 200).

As this project is grounded in environmental processes more than evolutionary trajectories, time is referred to by climate phase or in years before present rather than by stone age. *Homo sapiens* and species of humans that predated them are primarily referred to as hominins or early humans rather than as archaic or pre-modern humans. Where greater precision is needed the actual species is used, for example: *Homo sapiens* rather than Anatomically Modern Humans, and *Homo neanderthalensis* rather than Neanderthals. The beginnings of cognitive and behavioral modernity are well beyond the scope of this work. Language that implies the adaptive abilities of the people under study were somehow pre-modern is therefore avoided.

2.2 Perspectives from Spatial Science

Building a model that represents a geographically-bounded system requires an awareness of the kinds of thinking that guide similar spatial science projects. In spatial science, patterns and processes are codependent aspects of a system. Patterns are the "discernable outcomes or signatures" of spatial

processes, while spatial processes are identifiable through their patterns and can be anything that would "cause a system to change its state" (O'Sullivan and Perry 2013, 30-31).

Patterns are fundamental to translating a real system into a modeled one. Bottom-up models like agent-based models can risk including too many variables in order to achieve greater realism. An alternative approach, fully outlined in Grimm et al. (2005), is pattern-oriented modeling (POM). Pattern-oriented modeling encourages developers to focus on the observed patterns in the real system that they are trying to represent in the model. They focus on theories explaining patterns resulting from agent decision and build them into the model. Then alternative theories can be "tested by how well they reproduce the patterns." (Grimm et al. 2005, 988). If multiple patterns are represented, then the model's parameters should be calibrated so that the greatest number of patterns can be reproduced simultaneously. Grimm et al. make an observation about complex systems that is analogous to the Darwinian direction of anthropological hunter-gatherer theory that Bettinger et al. (2015) describe:

> Theories of complex systems may never be reproducible to simple analytical equations, but are more likely to be sets of conceptually simple mechanisms (e.g., Darwinian natural selection) that produce different dynamics and outcomes in different contexts. POM thus may lead us to an algorithmic, rather than analytical approach to theory (Grimm et al. 2005, 991).

Accordingly, the geographically explicit agent-based model described in this study was designed to help identify an algorithm for agent movement in these environments rather than an equation. This algorithm can contribute to a greater understanding of the system and opportunities for sea crossings but it is not designed to identify variables that will always directly lead to sea crossings.

2.1.1. Similar Spatial Science Studies

This study marks the first time that an agent-based model has been applied to early seafaring in the Middle to Late Pleistocene Mediterranean. However, similar simulations have been created by archaeologists to allow for a better understanding of related problems, and the decisions they made with their models informed the design of this one.

One of the main uses for agent-based modeling in archaeology has been to simulate spatially explicit "what-if" scenarios that are constrained by archaeological data and can be used to better understand a social and environmental context. In this case, a successful model would produce output that closely resembles the actual archaeological record (Cegielski 2016, 287).

Archaeologists have used computer simulations to better understand the spatial dynamics of large-scale migrations, for example, the arrival of humans in the Americas or the Neolithic transition in Europe. Hazelwood and Steele (2004) demonstrated how archaeological data, specifically radiocarbon-dated samples, could be used in conjunction with models of past human dispersals. Entwisle et al. (2016) created an agent-based model to simulate the relationship between climate shock and migration. One computer simulation of human migration out of Africa during the Pleistocene comes from Mithen (2002) who used a simulation to test theories about the environmental influences (glacial/interglacial cycles, land bridges) on human migration. Mithen's model was largely terrestrial.

Especially in the Mediterranean, computer simulations of human migration are mostly limited to travel over land. Using NetLogo, Romanowska (2013) created an agent-based model of human migration out of Africa during the Early Pleistocene. Romanowska's model tested how access to different routes might have affected Lower Paleolithic (approximately Early to Middle Pleistocene) site distribution. She found that none of her tested routes affected the distribution patterns of European Lower Paleolithic sites. She recommends creating a model with better environmental data and looking outside of the dispersal routes conceptual model. Romanowska shows how agent-based modeling can be used to quantify and test certain conceptual models and hypotheses in archaeology; however, it does not allow for travel over the sea.

Some models have considered whether ancient seafaring could be better described by a voyage model or a drift one. Fitzpatrick and Callaghan (2008) used a computer simulation to answer questions about voyaging and drifting during one particular trip. They used historical records in combination with a

computer simulation based on oceanographic, anemological (wind), and climatological data to better

understand how Magellan successfully sailed around the tip of South America (despite usually harsh

conditions) but passed the Moluccas to end up as far north as Guam.

Davies and Bickler (2013) used R, NetLogo, OPeNDAP, and Google Earth to create a model of

early seafaring. With the assistance of ethnographic research, they considered the capabilities of ancient

watercraft and allowed for several varieties in their model, including vessels equipped with sails. In

addition, drifting and navigating vessels were able to move in different ways, for example drifting

vessels may have lacked the capacity to row or drop an anchor. Davies and Bickler used an equation to

calculate the visibility of each point in their study area from the sea. They started with modern

bathymetric data and then adjusted the sea level to approximate past environments. They also used

modern winds and currents as proxies for past ones. To demonstrate their model, they applied it to a

particular case study of seafaring between Tahiti and Hawaii. This model does not take into account the

environmental pressures that might have prompted sea-crossings.

Davies and Bickler have elements of their methodology in common with similar projects. For

example, Voris (2000) created Pleistocene maps of Southeast Asia, another site of early seafaring, by

using bathymetric depth contours and past sea level data. In his maps,

> present day bathymetric depth contours [were] used as proxies for previous shore lines at
> particular times in the past 250,000 years before present (Voris 2000, 1155).

Furthermore, classicists have long used available wind data from their own times to

contextualize their analysis of Ancient Greek and Roman authors. To test the validity of this approach,

Murray (1987) compared the descriptions of winds recorded by Aristotle and his student Theophrastos

with modern observations. He noted largely "good" and "good to fair" degrees of agreement between

ancient and modern observations for much of the Mediterranean. The observations near the Strait of

Gibraltar and Sardinia were "poor" and "fair to poor," but they improved as one approached Greece

from the west. This is likely because the ancient observers themselves were Greeks. Murray's work

supports the possibility that modern winds are "essentially the same" as Classical ones and can

therefore be used to study Classical antiquity (Murray 1987, 159).

Murray's findings allowed Leidwanger (2013) to model early seafaring using present-day wind

data for the Mediterranean. In his model, Leidwanger uses GIS to look more closely at how ancient

sailing times might have affected connectivity in the Aegean. Using reconstructed environments, he

examines the journey of a sixth century BCE merchant vessel that sunk off the coast off Turkey.

2.2.1 Complexity and Agent-Based Modeling

In the last fifty years, modeling software has advanced enough to allow models to represent

very detailed systems. Agent-based models, especially, are able to contain more complexity, allowing

research conducted with them to be more informative though less strongly predictive.

The idea that a simpler hypothesis with fewer assumptions is better, often referred to as

Ockham's razor, is challenged by these newer simulation models. O'Sullivan et al. (2012) caution against

broadly applying this idea to agent-based models, especially considering the real systems which they

represent are far from simple. While the role of increasingly complex agent-based models might be to

inform more than predict, it is still worth noting that they tend to generate more successful predictions

when their geographic contexts are specified and constrained (Heppenstall et al. 2012).

The amount of complexity to build into this agent-based model was carefully considered. The

simulation should be simple enough to be studied and tested, but at the same time grounded enough in

real environmental data so that it might contribute to discussion about Middle to Late Pleistocene sea

crossings in the Mediterranean. While building and evaluating this model, it was important to keep in

mind where additional detail might limit the predictive ability of the model but add to its ability to

describe the spatial dynamics being studied. With hunter-gather anthropology's growing shift in focus

from cause-and-effect to governing processes, the informative role of complex agent-based models is

especially valuable.

2.2.2 Validating Agent-Based Models

Agent-based models are validated to ensure that they function as they should. Structural validation involves verifying that the output is appropriate and that the model operates in a way that closely represents the processes being studied. One way the latter can be studied is indirectly, through methods such as sensitivity analysis. Sensitivity analysis involves systematically adjusting model parameters and examining the ways the output changes as a result. In this way, sensitivity analysis can help with model calibration, which determines the ideal range of parameter settings that will allow the model to best simulate real-world processes. The output of the model should also be tested as a part of structural validation to make sure it closely matches real data (Ngo and See 2012).

Cost surfaces created using the Principle of Least Effort offer a way of generating early estimates of whether or not the distribution of the agents made sense considering the difficulty of traveling through different parts of the model environment. The Principle of Least Effort assumes individuals would make choices that minimize the amount of effort required to accomplish something (Zipf 1949). Cost distance is calculated in the amount of effort (or time or resources) required to traverse a distance.

This is a part of structural validation referred to as face validation. Face validation is conducted in the early stages of model development and it involves studying the animated model runs, examining a single agent over time, and gauging whether or not the initial output falls within an acceptable range (Ngo and See 2012).

Model validation occurs throughout the model building process, and while some validation methods focus on input and output, visualization offers valuable insight into what happens during model runs. Visually examining the simulation as well as the output is a valuable way to catch errors at different stages of the modeling process (Heppenstall et al. 2012). For this project, it was important to choose a modeling software that offered animated simulations.

The archaeological record, especially the spatially and temporally relevant sites identified in Broodbank (2013), Howitt-Marshall and Runnels (2016), and Papoulia (2016), can be used to help validate the model at this stage. However, care should be taken to avoid overfitting the model to an incomplete archaeological record, especially when modeling coastal migration as far back as the Pleistocene. Erlandson argues that the role of the sea in the course of human history was overlooked for so long because too much weight had been given to a "deeply flawed" terrestrial record. Over time, much of the coastal evidence of early seafaring in the Mediterranean was destroyed and possibly submerged by rising seas (Erlandson 2010, 20).

Therefore the model should only be "weakly constrained" (Hazelwood and Steele 2004, 673) by these data. Instead, sensitivity analysis and early validation of this model should focus on the appropriate functioning of the model given the input more so than the similarities or differences between the output and current archaeological record.

2.1.2. Best Practices

In "Best Practices for Scientific Computing," Wilson et al. (2014) offer guidance for developers of programs to be used for scientific research. A program should be developed for people. The work the users have to do should be minimized and the code should be written so that it can be easily understood by people. A computer will run ugly code but a person might not read it. The design of the program should be documented and apart from the code and the mechanics. Programs should be developed incrementally to facilitate greater control of the process and easier debugging. Unnecessary repetition of code as well as human work should be avoided. Lastly, programming should be done in a way that facilitates collaboration. As much as possible, these best practices were followed in the design of the R spatial analysis and in the development of the agent-based model.

18

Chapter 3 Model Design

The research question for this study is: "How can we find places where the environment created opportunities for early sea crossings." The working hypothesis is that a data-rich, geographically explicit agent-based model could help Mediterranean archaeologists find these places. An agent-based model can allow archaeologists to formulate and test theories about the ways past environments created opportunities for sea crossings. Accordingly, the purpose of this research was to determine how such a model could be designed.

3.1. Required Software

This model was written on Mac OS X 10.12.3. The language R (version 3.3.3) and software NetLogo (version 6.0.1) were used to create this model. R is a programming language and software environment used for statistical analysis and visualization. Its flexibility lends it to a wide range of applications, including spatial science. R is especially useful for managing large datasets and repeated tasks, which was an important consideration for this project given its temporal and spatial scale. NetLogo is the agent-based software used in this project. Its versatility, the abundance of its documentation and the quality of its visualizations made it a good fit. Often the inner workings of agent-based models can be a black box to researchers trying to understand them. There is input and output, and what occurs between must be inferred. NetLogo's visualizations and its monitors, which update as the model runs, help mitigate this problem.

The use of R and NetLogo also make this project easier to share and reproduce. R and NetLogo are both open source and freely available online. The data and code used in this project have been made available on GitHub (Baumann 2017a). The NetLogo models used in this study are both available on Modeling Commons (Baumann 2017b). An R Markdown HTML document allows code chunks and plots generated from them to be included within a text. R Markdown versions of Chapters 3 and 4 have

been created and posted to the GitHub repository for this project as well. Code chunks included

throughout the remainder of this document match those that appear in the R Markdown versions of

these chapters.

Code Chunk 1 lists the main R packages used in this project. Among these, one stands out as

particularly important. The package "raster" allows for the manipulation of spatial data that is used to

determine the changing environmental conditions in the model cells.

```
install.packages('raster')
install.packages('rdrop2')
install.packages('gdalUtils')
install.packages('rgdal')
install.packages('proj4')
install.packages('gdistance')
install.packages('RColorBrewer')
install.packages('psych')
```

Code Chunk 1: R Packages Needed to Run This Model (R)

3.2. Conceptual Model

A geographically explicit agent-based model was designed to show where environmental

pressures could have created opportunities for sea crossings in the Middle to Late Pleistocene

Mediterranean. This model involves recreating past landscapes and representing agent movement over

those landscapes based on spatially and temporally varying levels of cost as well as temporally varying

levels of motivation.

Below is the conceptual model diagram that guided the design process. Environmental factors,

such as slope, wind, and vegetation were processed in R to generate a cost and motivation value for

each cell in one of the reconstructed environments. Agent movement over these reconstructed

environments was modeled in NetLogo. The results of this model can be read in terms of the agent

distribution after a run.
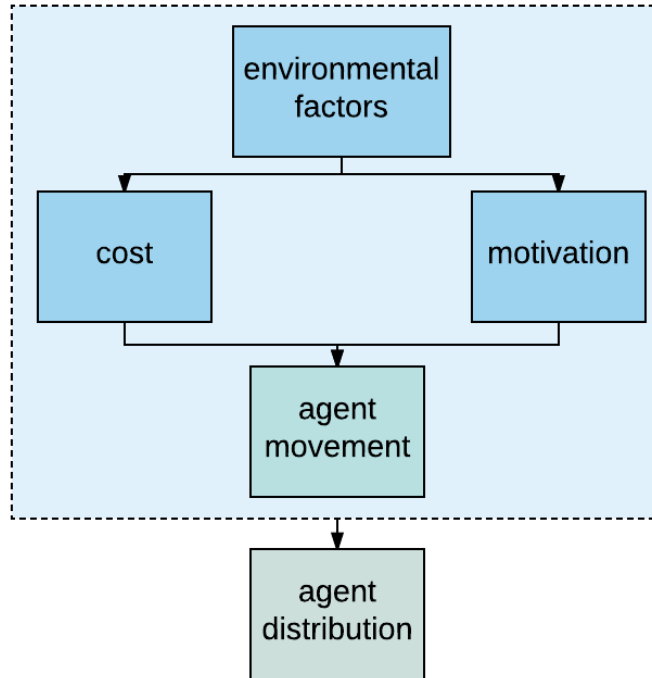
Figure 1: Conceptual Model Diagram

For land cells, the cost of moving from one cell to the next is calculated in terms of slope. For

sea cells, the cost of moving to the next cell is represented by wind cost. In order to keep this model as

straightforward and easy to run as possible, there are no directional variables. In this model, higher

slope indicates more difficult terrain and higher winds indicate choppier seas. This model focuses on the

role of decreasing vegetation as a driver of migration.

Agents make decisions about how to move across the modeled landscapes based on spatially

and temporally varying levels of motivation and cost. In this model, an agent represents one person or

group of people. Rather than attempting to precisely estimate the number of people and size and

location of their groups, this model simply works with 100 agents that begin at random points within the

study area. This was a large enough number to test agent responses to different parts of the study area

over a 1,000 or even 500 year run. Moreover, the use of 100 agents makes it easier to translate readings

from monitors into percentages (e.g. Only 3% of agents select high cost cells during the conditions of

this time slice).

A standard way to systematically describe the design of an agent-based model is the Overview,

Design concepts, and Details (ODD) protocol, which was first published in 2006 (Grimm et al. 2006). This

model is described below using guidance from the updated ODD (Grimm et al. 2010) which outlines the

required components of the protocol. These components include descriptions of the dimensions of

variables within the modeled environment. They also describe the agents themselves and how they

function within this system. While code chunks are helpful to those familiar with the programming

languages used in a particular study, the ODD is a better tool for describing a model to developers who

prefer other programming languages or to people who prefer not to read code at all.

Overview

Purpose: This model shows that an agent-based model can be used to answer the question, "How can

we find places where the environment created opportunities for early sea crossings." In doing so

it operationalizes a revisionist conceptual model in which the sea was less of a barrier and more

of a facilitator for early human migration in the Pleistocene Mediterranean, setting it apart from

other, more terrestrial models of Pleistocene migration through the region.

Entities: Agents representing migrating people or groups of people (100 during each model run) and grid

cells. The

State variables: The cell variables are land.or.sea, cost, motivation, boundary, difference, times.here.

Scales: Temporal scale is 769,000 years. In the full model, the temporal resolution is 1 year with

environmental (cell variable) changes every 1,000 years. Cell resolution is 34 km X 34 km. Model

environment is 82 x 125 cells.

Process and scheduling: If in the set of neighboring cells, there exist cells for which motivation is greater

than cost, the agent will move to the cell in which the difference between motivation and cost is

the greatest. In other words, the agents must determine whether they have enough reason to

move to a neighboring cell to make it worth the effort it would take to get there. If competition

is enabled, agents will not move to cells that are already occupied by more than the allowed

number of agents (adjustable by slider). Difference and times.here are updated every tick.

Land.or.sea, cost, motivation, and boundary are updated every 1,000 years. Time is represented

through discrete steps (ticks). A run of this model requires that the agents move over one

environment for a set number of ticks (representing years). In the full model, a run takes 1,000

years or ticks and there are 769 runs for the entire time range included.

Design Concepts

Basic principles: In this model, the sea is a facilitator and not a barrier for early human migration. Higher

slope indicates steeper terrain, which is more costly to traverse. Likewise, higher winds indicate

choppier seas, which is more costly to cross. The main driver (motivation) is decrease in

vegetation as indicated by levels of Saharan dust measured from a sea core are which are

indicative of aridity at different time steps. These measures are used as proxies for general

environmental pressures for migration.

Emergence: In the test run described in Chapter 4, agents appeared to cluster in certain sections of the

modeled environment. This is an encouraging sign that a model like this might be generating

patterns that were not directly programmed. However, no definitive claims can be made at this

time.

Adaptation: Indirect adaptation can be modeled through sea crossings, which happen because of

chance and immediate environmental conditions. Agents can cross the sea if the cost of crossing

the land in their immediate environment is greater than the cost of crossing the sea and the

amount of motivation in that time slice makes the cost of crossing the sea worth it.

Objectives: Agents want to be in the cell with the greatest difference between motivation and cost. At

times when motivation is low or equal to cost, they will not move at all. At times when

motivation is higher, they will have more reason to traverse cells with higher costs (either due to slope or wind speed).

Learning: Agents cannot learn or remember.

Prediction: There is minimal prediction. Agent decision-making horizons are temporally limited to the current tick (year) and spatially limited to the surrounding four cells.

Sensing: Agents are able to examine four proximal cells (rook case) and determine which one has the greatest difference between cost (slope or wind speed) and motivation (vegetation level, uniform throughout the whole environment for each time step).

Interaction: If competition is enabled, agents will not be able to move to a cell that has more than the allowed number of agents.

Stochasticity: The starting points of the agents are randomly determined. At 781,000 ybp (the lower temporal limit of the model) humans could have been present in Europe, not just in Africa and the Levant. Allowing agents to begin at any point in the model allows us to better understand how people would have moved about the environment as a whole and it does not artificially limit people to starting points within Africa.

Collectives: Agents can represent one person or groups of people. However, these agents will still operate in this environment the same way. Cooperation isn't represented in this model, though it might be in future versions (for example larger vessels would require more people to build and operate).

Observation: The results of dummy data and real data runs are interpreted using the distribution of agents on the NetLogo interface and the levels of key monitors indicating the number of agents on land and sea as well as the number of agents on high cost cells. Agents hatch trails so their paths can be traced as the model runs. Though not used in this study, the cell variable, times.here, can be exported using the GIS extension and examined as a heat map.

Details

Initialization: At years=0, the cells have been set up according to the land.or.sea, cost, motivation, and

boundary variables for that time slice. Additionally, 100 agents have been randomly scattered

throughout the environment. Times.here is 0. Initialization is always the same. Currently, the

model cannot remember agent locations or cell variables from previous time slice runs.

Input data: The patch variables land.or.sea, cost, motivation, and boundary are imported using the GIS

extension and four ASCII files per time slice or run.

Submodels: This model contains 769 time slices for 769 model runs, each meant to be run for 1000

years (ticks). Thus the agent decision making scheme can be run in 769 different environments.

3.3. Agent Decision-Making Algorithm

Taking after Bettinger et al. (2015) and Grimm et al. (2005), as described in Section 2.2 above,

this model has at its core a decision-making algorithm. There is a simple set of rules that the agents

must follow when deciding how they are going to move around the study area. These rules are based on

the concept of cost distance and the Principle of Least Effort (Zipf 1949). If in the set of neighboring cells,

there exist cells for which motivation is greater than cost, the agent will move to the cell in which the

difference between motivation and cost is the greatest. In other words, the agents must determine

whether they have enough reason to move to a neighboring cell to make it worth the effort it would

take to get there. If competition is enabled, agents will not move to cells that are already occupied by

more than the allowed number of agents (adjustable by slider). The decision-making scheme of the

agents is summarized in Figure 2 below.
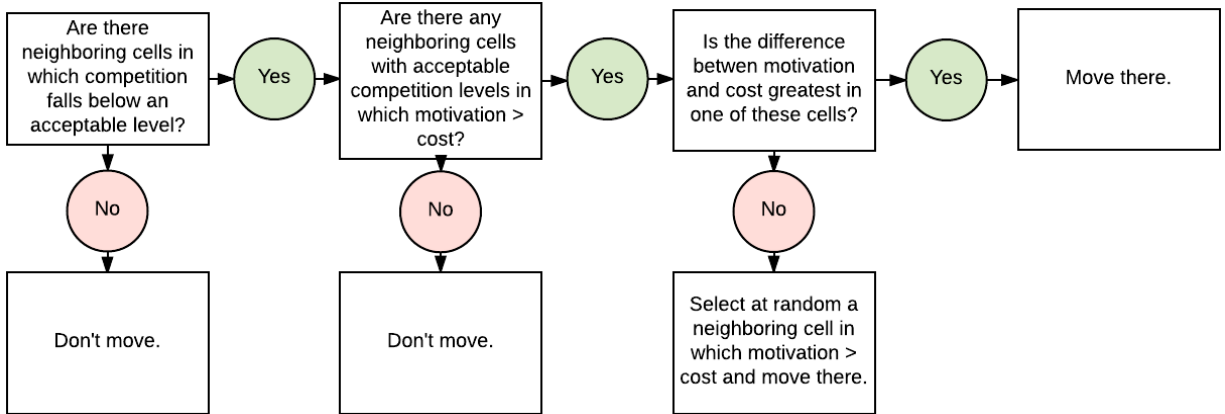
Figure 2: Agent Decision-Making Algorithm

This decision-making algorithm was then translated into NetLogo code. Code Chunk 2 shows the

NetLogo code that controls how the agents move considering different levels of competition and

different coefficients for cost and motivation. The hatch-trails procedure allows each agent to leave a

visual path in the interface. However, for more systematic analysis of the distribution of the agents after

a run, the times.here patch variable can be used. There are two separate move-persons procedures in

Code Chunk 2. One is run if competition is enabled and one is run if competition is not enabled. The

maximum number of agents per cell can be set before the run using a slider in the interface.

```
to move-persons-with-competition
  if competition-value < max-persons-per-cell [
    ask patches [
      set motivation (motivation * motivation.coefficient)
      set cost (cost * cost.coefficient)
      set difference (motivation - cost)
        ask persons-here [
          uphill4 difference
        ]
      ]
    ]
  ask persons[
    hatch-trails 1 [set color 16]
  ]
end

to move-persons
  ask patches [
    set motivation (motivation * motivation.coefficient)
    set cost(cost * cost.coefficient)
    set difference (motivation - cost)
    ask persons-here [
      uphill4 difference
    ]
  ]
  ask persons[
    hatch-trails 1 [set color 16]
  ]
end
```

Code Chunk 2: NetLogo Move Persons Procedure (NetLogo)

The model allows cost and motivation to be weighed differently relative to one another. As

shown in Figure 3, NetLogo interface includes sliders so that a cost coefficient and a motivation

coefficient can be set after the user sets up the run, but before they click "go." These coefficients range

from 1-3 so, for example, cost can be equally important as motivation, slightly less important than

motivation, or slightly more important than motivation. A greater range in coefficient values can be

created for future versions of this model. However, fewer options allow greater control of the model.

Figure 3: Cost and Motivation Coefficient Sliders

3.3.1. Testing the Agent Decision-Making Algorithm

To test the decision-making algorithm and ensure that it was correctly translated into NetLogo

code, a version of the model was created to run on dummy data. It is available on Modelling Commons

(Baumann 2017b). In the dummy version of the model, there are an even number of land and sea cells

and they are randomly assigned. The motivation for all cells is set using a slider in the interface. Cost is

randomly assigned values as well, and a histogram in the interface can be used to ensure that roughly

the same number of cells are assigned each cost value. However, range of cost and motivation values is

kept the same as it is in the real version of the model (values: 1-5, coefficients: 1-3). The randomness in

this dummy model allowed it to be used as a control. This dummy model was used to confirm that

nothing was accidentally coded into the model that would force agents to cluster in one part of the

study area or favor sea cells over land cells. This way, when the model was run with real data, this kind

of emergence could be attributed to the data rather than the model itself.

This dummy model was used for initial sensitivity analysis, to make sure the model functioned as

it should before the real data were introduced. The sensitivity analysis (SA) was designed as a

compromise between one at a time (OAT) SA and global SA. While global sensitivity analysis is more

comprehensive, OAT is less computationally rigorous. The worksheet used for the sensitivity analysis of

this model can be found on the GitHub repository for this study (Baumann 2017a). Figure 4 shows an

image of the dummy model interface after a run. The blue cells represent sea and the green cells

represent land. The red points are trails agents left behind after moving over a cell.
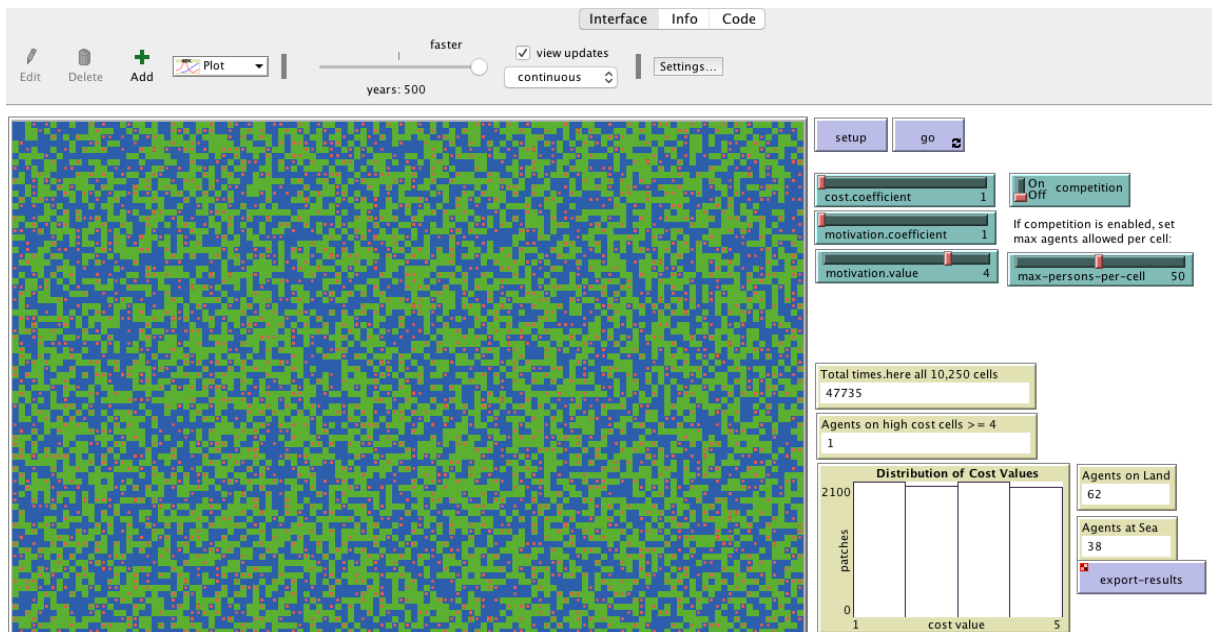
Figure 4: Dummy Model Interface after a Run

In the sensitivity analysis, first, every combination of cost.coefficient, motivation.coefficient, and motivation.value were run. During these runs, competition was left off. Next, every value of max.persons.per.cell was tested while the other values were left at 1. To complete all combinations, there were a total of 145 runs and the values of the monitors were recorded after each run.

For the most part, the distribution of agents over land cells and sea cells was about even at the beginning and ends of runs. This is means that any major difference in the number of agents on sea cells in the real model is more likely to be caused by the data than by the NetLogo code itself. As expected, increasing the number of agents allowed per cell increased the amount of movement during the runs. Having monitors that changed in the interface during the runs proved especially useful, especially since increasing the cost coefficient increased the number of high cost cells in general. Interestingly, most agent movement seemed to occur at the beginning of the runs. This was the justification for reducing the number of ticks for a dummy data run to 500 ticks instead of 1,000 after an error began to occur in the interface. As a result, in the dummy model, 1 tick is equivalent to 2 years. This error does not occur

in the real model, but it was observed over the course of this study that the longer NetLogo was run,

and the more complex the input became, the more likely interface errors were to occur.

In general, the expectation was that the greater motivation was relative to cost, the more

movement there would be over the modeled environment and the more agents would appear on high

cost cells. This mostly held true: increasing motivation (either the coefficient or the value itself) did

increase movement over the modeled environment. However an important find was that raising the

cost coefficient kept agents moving for a longer period of time than just raising the motivation

coefficient did. This is likely because of the randomness with which cost cells were distributed in the

model. There are likely many instances where low cost cells are directly situated next to high cost cells.

These situations are less frequent in the real model.

These initial runs with dummy data indicate how important local variation in cost can be when

understanding the effects of a large regional driver like decreasing vegetation. Therefore, the next step

in developing this model was devising a reliable way to prepare reconstructed environments over which

to model agent movement.

3.4. Modeled Environment

The elevation raster used in this study is the ETOPO1, a one arc-minute Global Relief Model that

includes both bathymetric and topographic data (Amante and Eakins 2009). This dataset, acquired in the

(unprojected) WGS 1984 datum, was used to establish the raster framework for the model. Because of

the size and location of the study area, as well as the types of spatial questions this model seeks to

explore, a compromise global projection, the Winkel Tripel, was required. The Winkel Tripel attempts to

minimize the distortion of shapes, areas, and distances. The National Geographic Society has used the

Winkel Tripel in place of the Robinson projection since the late 1990s (Esri 2016).

The map was examined, projected, and cropped in ArcMap. This version of the map is available

in the GitHub repository for this study and is labeled "raster0" (Baumann 2017a). Then the map was

imported into R. With accessibility in mind, the resolution of the model was deliberately lowered to 34

km x 34 km so that the model could be run on personal computers without encountering Java heap size

errors. Aggregating the data for NetLogo meant agents were considering larger sections of the study

area when making their decisions. Though not preferred, it did not harm the overall ability of the model

to indicate large regional trends in agent distribution.

Code Chunk 3 shows how this raster was imported, aggregated, plotted, and examined in R.

While the raster was already projected in ArcMap, it was necessary to identify the projection of this

raster in R. The aggregate function of the R raster package was used to reduce the cell resolution of the

modeled environments to approximately 34km x 34km. This level of aggregation was selected by trial

and error. Different environments were loaded into the NetLogo model until one could be found that

both loaded reasonably well and faithfully represented the shapes of the landforms in the study area.

The projected, aggregated study area with modern sea levels is shown in Figure 5 below.

```
require(raster)
require(rgdal)
require(rdrop2)
require(gdalUtils)
require(proj4)

#import & project base raster
base.raster <- raster("~/Desktop/seacrossings/raster0")
projection(base.raster) <- "+proj=wintri"
res(base.raster)
raster0 <-aggregate(base.raster, fact=20, fun=mean, expand=FALSE)
res(raster0)

#examine then plot the base raster
summary(raster0)
breakpoints <- c(-4966.0925, -1, 1, 2643.2900)
colors <- c("blue", "white", "green")
plot(raster0, breaks=breakpoints, col=colors)

#determine the world settings and transformation for NetLogo
ncell(raster0)
nrow(raster0)
ncol(raster0)
res(raster0)
nl.trans <- c(xmin(raster0), xmax(raster0), ymin(raster0), ymax(raster0))
```

Code Chunk 3: Import, Aggregate, Plot, and Examine Base Raster (R)
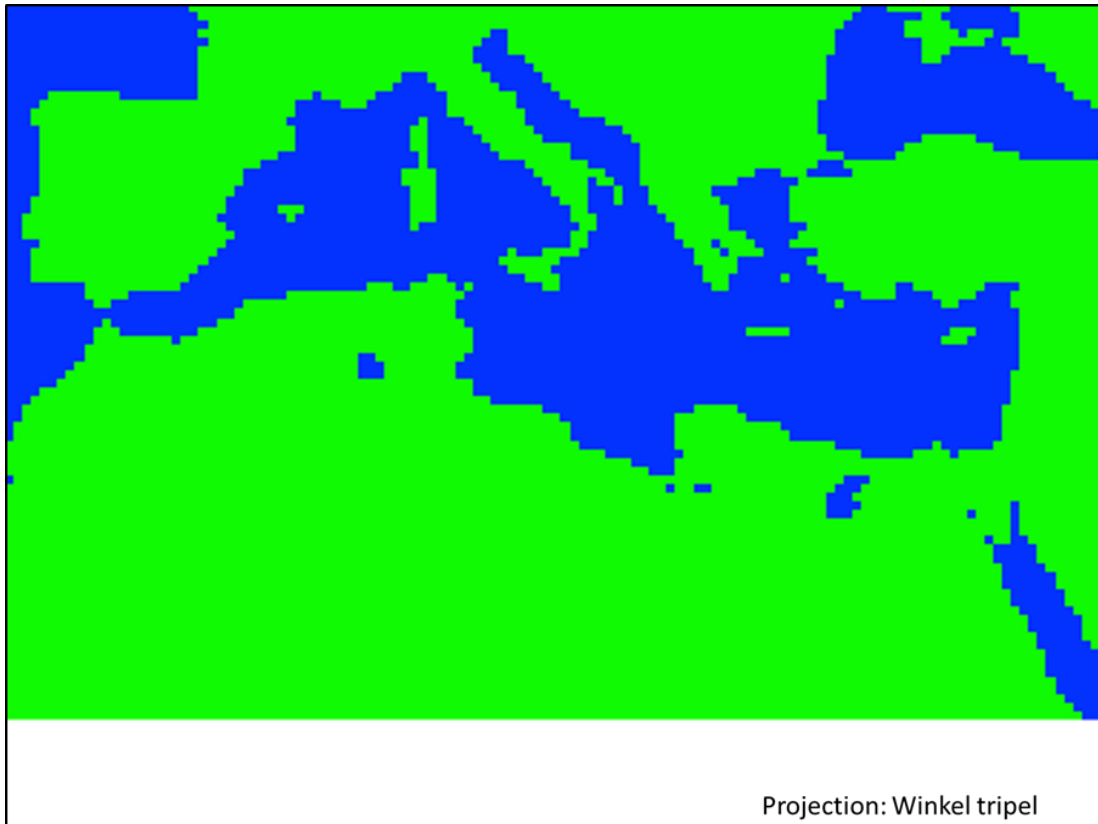
Figure 5: Aggregated Study Area with Modern Sea Levels

The NetLogo settings were configured so that the world was finite and did not wrap around itself. This was done to keep agents from being unrealistically bound to the study area or from ending up at one edge of the model after wandering off the opposite edge (the default setting in NetLogo). Instead, a procedure called remove-persons was written so that agents would disappear (or in NetLogo language, "die") if they reached an edge. A procedure called replace-persons was written so that an agent would then appear at a random point in the model to replace the agent who left the modeled environment. These procedures can be found in the NetLogo models posted on Modeling Commons and GitHub (Baumann 2017a; Baumann 2017b).

3.4.1. Reconstructing Past Environments

When covering such a large period, especially in the Mediterranean, it is important to account for varying environments so that the decisions agents make based on environmental pressures can be

shown against the corresponding environments at those times. By dividing the temporal range of 769,000 years into 1,000 year slices, a collection of 769 raster stacks was created with each stack representing the environment for a specific 1,000 year slice. The next step was to find a driver and a landscape for each time slice, in other words, a dust level and a sea level.

Using a technique similar to that used by Voris (2000) and Davies and Bickler (2013), described in Section 2.1.1 above, the base raster was reclassified 769 times using past sea levels to approximate paleocoastlines. Despite the availability of global sea level data and air temperature data for the last 3 million years (Bintanja and van de Wal 2008) as well as Red Sea level data for the last 150,000 years (Grant et al. 2012), it proved difficult to find Mediterranean Sea level data for the last 800,000 years. Fortunately, Rohling et al. (2014) used a new method of calculating past sea levels and this method both relied on Mediterranean samples (Wang, Tian and Lourens 2010) and extended the past sea level estimates to 5.3 million years ago. This was the dataset used to reclassify the raster maps using past sea levels. A chart from Rohling et al. (2014) was digitized using a program called GraphClick. The chart was loaded into the program, the x and y axes were specified, and then the sea levels included in that chart were individually digitized with mouse clicks. The output was a table with x and y values. To simplify the data to one sea level for every 10,000 years, some averaging and linear interpolation were needed. The section of the Rohling et al. chart that was used in this study is shown in Figure 6. The sea levels used in this study are shown in Figure 7.
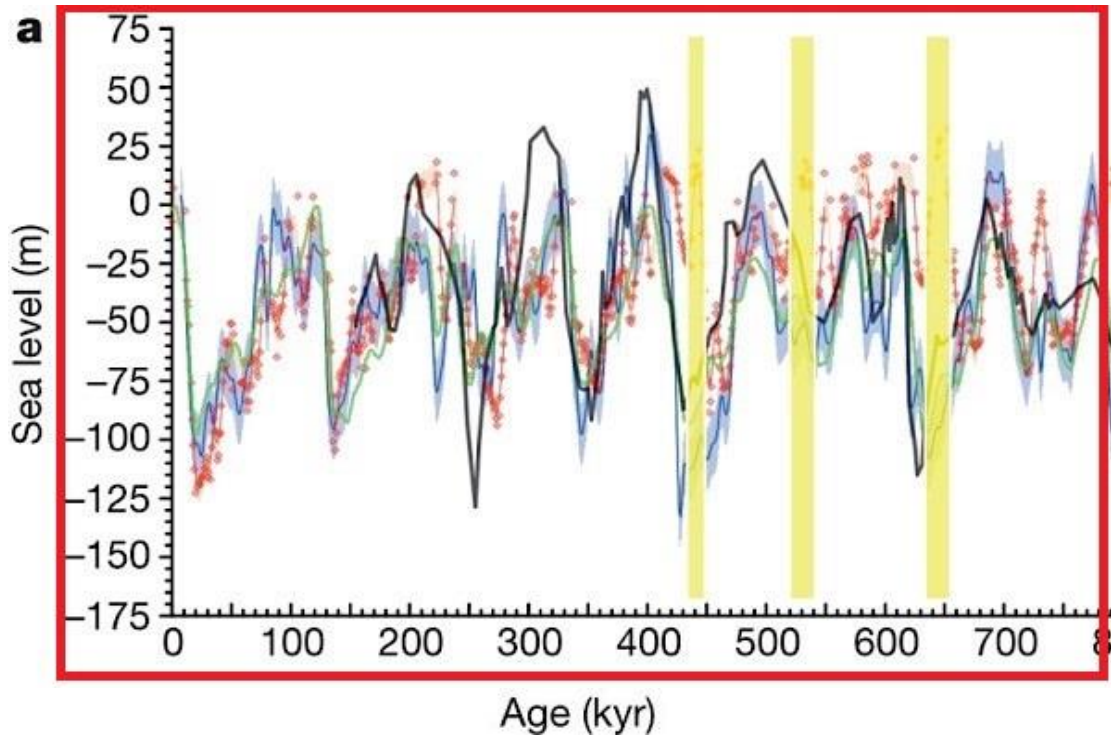
Figure 6: Digitized Portion of Rohling et al. (2014) Sea Levels Chart
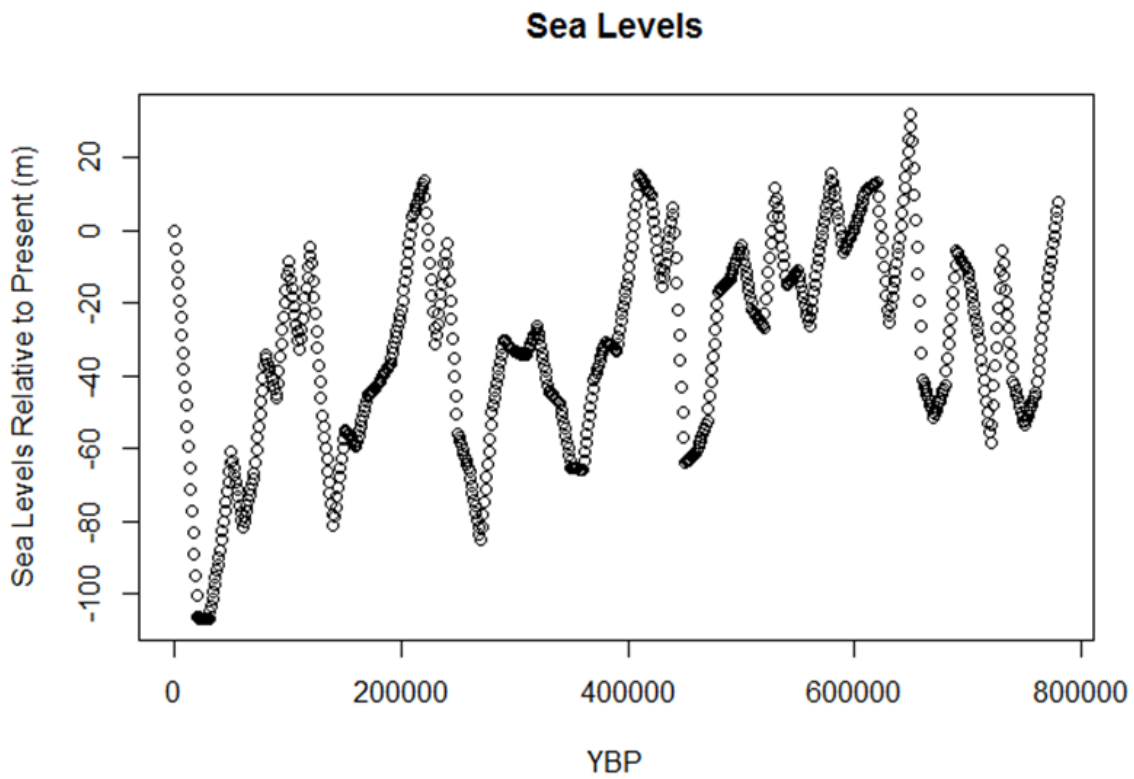

Figure 7: Sea Levels Used in this Model. Data: Rohling et al. 2014

While past sea levels alone have been used to map estimates of exposed land (e.g. Voris 2000), additional criteria would be needed to more precisely reconstruct coastlines, especially in the Aegean. Sea levels alone cannot be used to determine which islands were once connected to land. For that, subsidence rates (Lykousis 2009) would need to be accounted for. Accordingly, while this model only accounts for changes in past sea levels at each time step, the work of those like Lykousis (2009) and Lambeck et al. (2011) should be considered when examining the model's results. For example, it is possible that the Cyclades were connected to the mainland at times during the Pleistocene, though this model did not show it.

In light of the literature described in Section 1.2, decrease in vegetation is conceptually the main driver in the model. Since no readily available data on vegetation over the temporal range of the model is available, Aeolian dust levels, as calculated by Larrasoana (2003) and discussed above, were used to indicate temporally varying levels of regional environmental pressures serving as motivation for migration. These data indicate levels of aridity in the Sahara. They can be read as a proxy for vegetation and they also speak to larger climatic trends in the region. Therefore these data do not vary spatially. They indicate the overall environmental pressures for migration in the region during a time slice.

These data were already in a tabular format so no digitizing was necessary. However one dust level was needed for every 1,000 years (the data had a temporal resolution of 400 years). Because a similar data wrangling process was required for both sea and vegetation levels, a test was devised to determine whether this process was faster in R or in Excel. The process of preparing sea level data as described above was saved in an Excel Workbook which can be found on the GitHub repository for this model (Baumann 2017a). The process of similarly preparing vegetation data in R is shown in Section 3.5.4. The R method was faster for such a large dataset. A calculation could be applied to the entire dataset at once without the need for highlighting, copying and pasting, or even pivot tables. The dust levels from Larrasoana (2003) are shown in Figure 8 below.
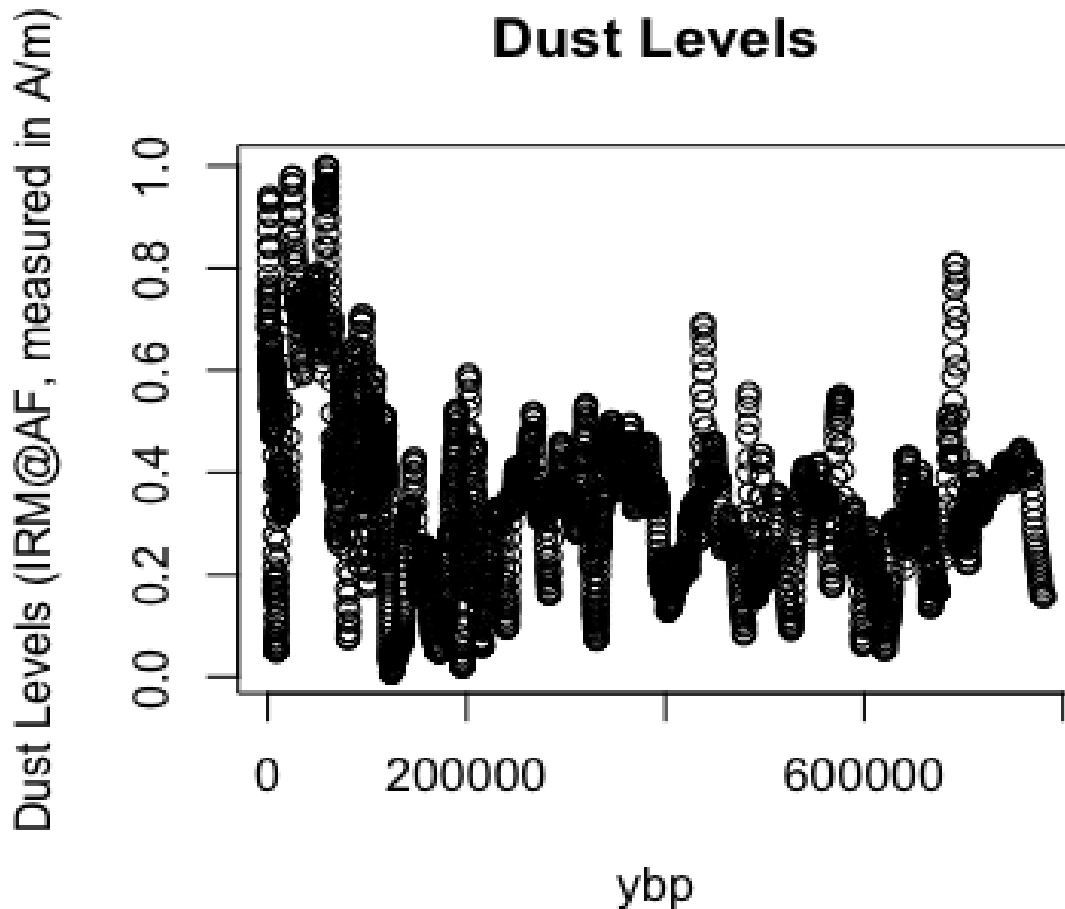
Figure 8: Dust Levels (Vegetation Proxy & Main Driver). Data: Larrasoana et al. 2003

To account for the cost of moving over sea cells, a wind speed dataset was acquired and built

into the model. As summarized earlier, previous work has shown how present-day wind patterns might

be used in simulation models in place of past wind data (Murray 1987; Leidwanger 2013). As shown in

Figure 9, The Modern Era-Retrospective Analysis for Research and Applications (MEERA) 100 m layer

from the Global Wind Atlas provides mean wind speeds aggregated from simulations spanning 1979 –

2013 (DTU Wind Energy, 2017). This was an image that was georeferenced in ArcMap and then

converted to a single-band raster before being imported into R where it was aligned and resampled to

match the dimensions and resolution of the base raster.
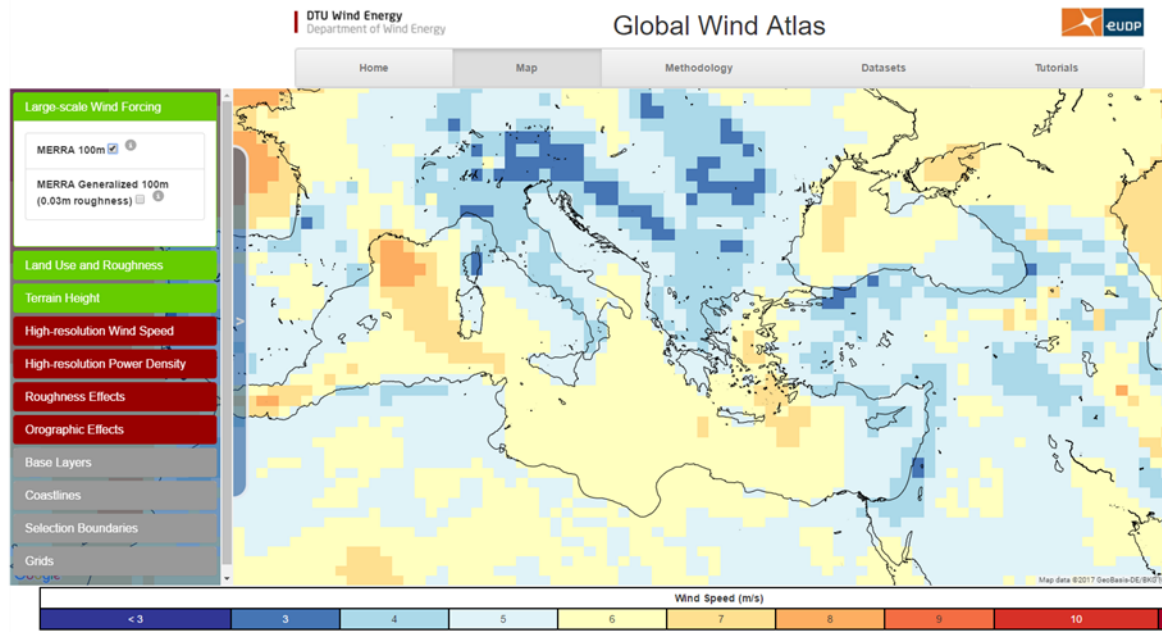
Figure 9: MEERA 100m Layer of the Global Wind Atlas

The Wind and Wave Atlas of the Mediterranean Sea is another dataset that might have been used in this model. It provides aggregated wind and wave data for the Mediterranean and it uses samples spanning a 10-year period (Cavaleri 2005). However, the continuous surface model of the Global Wind Atlas was easier to adapt to this model than the vectors provided by the Wind and Wave Atlas.

3.5. Raster Layers

 Four ASCII files are prepared in R to be run in NetLogo for each time slice. The first ASCII file indicates whether each cell is a land cell (1) or a sea cell (0). The second indicates the cost to traverse each cell (values range from 1-5). The third indicates the motivation for all cells during that particular time slice (values range from 1-5, but are spatially constant in each time slice). The last ASCII file indicates whether the cell is on the edge of the modeled environment (1 or 0).

To capture the range of environments possible, the time slices with the lowest, median, and highest sea levels (LSL, MSL, HSL) during the Middle to Late Pleistocene were the focus when designing

the model. The code below describes how the ASCII files for these time slices were generated, and the same method can be used to generate ASCII files for any of the other time slices in this study.

The raster package in R allows related rasters to be stacked on top of one another so that cell values in one raster can be calculated based on the values of corresponding cells in related rasters. Layers in a raster stack can be called for different functions using their position within that raster stack. This is analogous to the way components in a vector are handled. Figure 10 shows the components of the raster stacks used for spatial analysis in R, and for reading different environments into NetLogo. Layers within a raster stack are indicated by their position.

For Processing in R		
raster stack name	purpose	layers
raster0	base	LSL, MSL, HSL
raster1	sealevel	LSL, MSL, HSL
raster2	land.or.sea	LSL, MSL, HSL
raster3	cost	LSL, MSL, HSL
raster4	motivation	LSL, MSL, HSL
raster5	boundary	LSL, MSL, HSL

For Reading into NetLogo	
raster stack name	layers
environmentLSL	land.or.sea [1]
	cost [2]
	motivation [3]
	boundary [4]
environmentMSL	land.or.sea [1]
	cost [2]
	motivation [3]
	boundary [4]
environmentHSL	land.or.sea [1]
	cost [2]
	motivation [3]
	boundary [4]

Figure 10: Raster Layer Names

3.5.1. Raster 1 Sea Level

In the code chunk below, the base raster is reclassified using past sea levels in order to approximate the appropriate paleocoastlines. These sea levels come from the Rohling et al. (2014) data set after the data preparation described in Section 3.4.1.

```
#raster1 (partial)
#sealevel
raster1LSL <- (raster0 - 107.02)
raster1MSL <- (raster0 - 30.84)
raster1HSL <- (raster0 + 31.67)
```

Code Chunk 4: Raster 1 Sea Levels (R)

3.5.2. Raster 2 Land or Sea

Once the different sea levels can have been accounted for, the rasters from the previous step

can again be reclassified to create new layers that identify all cells as either land (1) or sea (0). To do this

a three column from-to-becomes matrix is created and then used to identify which range of elevations

should be designated as land and which should be designated as sea. For example, in Code Chunk 5, six

values are concatenated and then converted into a three column matrix. The first value is the lowest

elevation that could be considered sea. The second value is the highest elevation that could be

considered sea. The third value "0" is assigned to any cells that have elevations within this range. Next is

the lowest elevation that could be considered land and then the highest elevation which could be

considered land. Elevations within this range are assigned a "1." A lower sea level and a higher

elevation than are likely to be found in any of the possible environments are used in this matrix as a

precaution. This way, if any cells have an elevation that does not fit into the matrix, they will receive an

NA value, and NetLogo can encounter errors when it is asked to perform a mathematical operation on a

non-number. This check prevents NA values.

```
#raster2 (partial)
#land.or.sea
m <- c(-6000,-1,0, -1,4000,1)
rclmat<-matrix(m, ncol=3, byrow=TRUE)
raster2LSL <- reclassify(raster1LSL, rclmat)
raster2MSL <- reclassify(raster1MSL, rclmat)
raster2HSL <- reclassify(raster1HSL, rclmat)
```

Code Chunk 5: Raster 2 Land or Sea (R)

3.5.3. Raster 3 Sea Cost (Wind) and Land Cost (Slope)

This next step requires the wind layer to be imported and aligned to the base raster. The

resolution of the wind data was 100m, but wind raster had to be resampled so that it had the same

number of cells as the base raster. Nearest neighbor ("ngb" in Code Chunk 6) is the resampling method

used. A reclassification matrix is then used to sort the wind raster values into five classes. As mentioned

above, five classes were used for slope, wind, and dust levels. These classes were manually set using

histograms of each data set. The values in the reclassification matrix in Code Chunk 6 correspond to values from the single-band, grayscale image, but the actual range of values for the study area is 4-8 m/s.

Next, slope is calculated using the terrain function from the raster package. The unit used is degrees and only four neighbors (rook case) are used in the calculation. The NetLogo model is also limited to four neighbors to allow it to be more easily run and understood. However, future versions of the model have the option of eight neighbors both at this step and within the NetLogo move-persons procedure.

For cost and motivation to be comparable and then simultaneously scalable in the weighted versions of the model, wind, slope, and vegetation all needed to be sorted into the same number of classes. To do this, histograms were created in R for each dataset. The distribution of the data was examined and five classes were selected because it was a low number that still accounted for the variety in the data.

Lastly, a function is used to assign a slope value to sea cells and a land value to raster cells. Because these raster layers are called like vectors, *ifelse* must be used rather than *if else*. Code Chunk 6 shows the process of assigning wind and slope costs to the cost rasters.

```
#raster3 (partial)
#land cost (slope) & sea cost (wind)

#import windraster
require(raster)
windraster <- raster("~/Desktop/seacrossings/windraster")
projection(windraster) <- "+proj=wintri"
alignExtent(windraster, raster0, snap= "near")
windraster.resampled <- resample(windraster, raster0, method ="ngb")

#classify windraster
m.wind <- c(0, 42, 1, 42, 96, 2, 96, 163, 3, 163, 222, 4, 222, 250, 5)
rclmat.wind <- matrix(m.wind, ncol =3, byrow= TRUE)
classified.wind <- reclassify(windraster.resampled, rclmat.wind)

#calculate slope
raster3LSL0 <- terrain(raster1MSL, opt='slope', unit="degrees", neighbors=4)
raster3MSL0 <- terrain(raster1MSL, opt='slope', unit="degrees", neighbors=4)
raster3HSL0 <- terrain(raster1MSL, opt='slope', unit="degrees", neighbors=4)

#remove NAs
raster3LSL0[is.na(raster3LSL0)] <- 0
raster3MSL0[is.na(raster3MSL0)] <- 0
raster3HSL0[is.na(raster3HSL0)] <- 0

#classify slope
m.slope <- c(0, 0.5, 1, 0.5, 1.5, 2, 1.5, 2.5, 3, 2.5, 3, 4, 3, 4.5, 5 )
rclmat.slope <- matrix(m.slope, ncol =3, byrow= TRUE)
raster3LSL0c <- reclassify(raster3LSL0, rclmat.slope)
raster3MSL0c <- reclassify(raster3MSL0, rclmat.slope)
raster3HSL0c <- reclassify(raster3HSL0, rclmat.slope)

#assign slope or wind
costLSL <-stack(raster2LSL, raster3LSL0c, classified.wind)
costMSL <-stack(raster2MSL, raster3MSL0c, classified.wind)
costHSL <-stack(raster2HSL, raster3HSL0c, classified.wind)

raster3LSL <- calc(costLSL, fun= function(costLSL)
  ifelse((costLSL[1] =1), (raster3LSL <- costLSL[2]), (raster3LSL <- costLSL[3])))

raster3MSL <- calc(costMSL, fun= function(costMSL)
  ifelse((costMSL[1] =1), (raster3MSL <- costMSL[2]), (raster3MSL <- costMSL[3])))

raster3HSL <- calc(costHSL, fun= function(costHSL)
  ifelse((costHSL[1] =1), (raster3HSL <- costHSL[2]), (raster3HSL <- costHSL[3])))
```

Code Chunk 6: Raster 3 Cost (R)

3.5.4. Raster 4 Motivation

Code Chunk 7 shows how the vegetation proxy values (Aeolian dust levels from Larrasoana et al.

2003) are imported and processed. Unneeded data are removed, NA values are omitted, and the data

are aggregated temporally so there are only values every 1,000 years. (The original data had a resolution

of 400 years.) The data are then grouped into five classes. Five indicates a lot of dust, suggesting greater

aridity, lower vegetation, and therefore high regional pressures for migration. One indicates a low level

of dust, a high level of vegetation and much less motivation to move. The motivation level will be the

same for all cells in a time slice as these values are indicative of broad environmental trends at each

time period.

```r
# raster4
#motivation (vegetation)

#import data
larrasoana2003 <- read.csv("larrasoana2003.csv", header=TRUE)

#remove data later than the middle to late pleistocene
mlpleist.na<-larrasoana2003[!(larrasoana2003$Age..kyr.>781),]

#remove any na values
mlpleist <- na.omit(mlpleist.na)

#convert from kyr to yr
dust <- mlpleist$IRM.0.9.AF.120mT..A.m.
dust.kyr <- mlpleist$Age..kyr.
dust.yr <- (dust.kyr * 1000)
dust.yr.trunc <- trunc(dust.yr)

#limit significant digits - we only need values every 1,000 years
dust.round <-ifelse((dust.yr.trunc <= 1000), 1000,
                ifelse((dust.yr.trunc >1000 & dust.yr.trunc <= 10000),
                    signif(dust.yr.trunc, digits=1),
                    ifelse((dust.yr.trunc >10000 & dust.yr.trunc <= 100000),
                        signif(dust.yr.trunc, digits=2),
                        signif(dust.yr.trunc, digits=3))))

#aggreage dust measurements
dust.df <- data.frame(dust.round, dust)
options(scipen=10)
dust.agg <- aggregate(dust.df, by= list(dust.df$dust.round), FUN=mean, na.action= na.omit)

#sort data into classes 1-5
#level 5 vegetation = a lot of dust, therefore, regionally, a lot of pressure to move
#level 1 vegetation = not that much dust, not that much reason to move
levels.matrix <- matrix(c(0, .2, 1,    .2, .4, 2,    .4, .6, 3,    .6, .8, 4,    .8, 1, 5),
                    nrow= 5, ncol=3, byrow=TRUE)
levels.df <- as.data.frame(levels.matrix)

dust.levels <- transform(dust.agg, Type=levels.df$V3[findInterval(dust, levels.df$V1)])


#reclassify raster using vegetation data
LSL.m.veg <- c(1, 4, 0, 0)
LSL.rclmat.veg <- matrix(LSL.m.veg, ncol= 2, byrow= TRUE)
raster4LSL <-reclassify(raster2LSL, LSL.rclmat.veg)

#HSL and MSL have same dust level so we can make one reclassification matrix for both
MHSL.m.veg <-c(1, 2, 0, 0)
MHSL.rclmat.veg <-matrix(MHSL.m.veg, ncol= 2, byrow= TRUE)
raster4MSL <- reclassify(raster2MSL, MHSL.rclmat.veg)
raster4HSL <- reclassify(raster2HSL, MHSL.rclmat.veg)
```

Code Chunk 7: Raster 4 Motivation (R)

3.5.5. Raster 5 Identify Boundaries of Study Area

As mentioned earlier, NetLogo procedures were written to remove and replace agents that wander to an edge of the study area. They were written so that agents would not be artificially bound to the study area but also so that there could be the same number of agents within the study area at every tick in the run. This section of R code creates raster layers that indicate whether a cell is a boundary cell (1) or not (0). Because there are 125 columns and 82 rows in every raster, it is possible to identify cells that should receive a boundary value (1) based on their location.

```
#raster5 (partial)
#boudary
require(raster)
edge.cells <- c(cellFromRow(raster0, 1), cellFromRow(raster0, 82),
                cellFromCol(raster0, 1), cellFromCol(raster0, 125))
raster5 <-raster0
raster5[] <- 0
raster5[edge.cells] <- 1
```

Code Chunk 8: Raster 5 Boundary (R)

3.5.6. Compile and Export Environments

The code below was used to stack these layers into environments. A different environment is needed for each time slice. However, Raster 5 is the same for every run because the boundary cells are always in the same place. The writeRaster function creates a separate ASCII file for each layer in the environment raster stack and distinguishes between them by adding "_1", "_2", "_3", or "_4" after the rest of the environment name.

```
#save rasters as ascii files for gis extension
writeRaster(environmentLSL, filename= "environmentLSL.asc", bylayer=TRUE)
writeRaster(environmentMSL, filename= "environmentMSL.asc", bylayer=TRUE)
writeRaster(environmentHSL, filename= "environmentHSL.asc", bylayer=TRUE)
```

Code Chunk 9: Stack and Export Environments (R)

3.5.7. Import Environments into NetLogo for Setup

The NetLogo GIS extension was used to import these rasters into NetLogo. After the ASCII files were loaded into NetLogo, the set-patch-values procedure converted each of these to different cell (or

43

"patch") variables. Land.or.sea is used to tell the agents whether they are on land or sea cells. The

values from the cost and motivation variables are used in the move-persons procedure (Code Chunk 2).

The boundary variable is used to remove and replace agents when they move too close to the edge of

the modeled environment. The gis:apply-raster command is used here instead of the gis:raster-sample

command. While both are suggested in the GIS General Examples (Wilensky 2017), gis:raster-sample

produced a not real number (NaN) error.

```
to setup-gis
  set land.or.sea-dataset gis:load-dataset "environment000L_1.asc"
  set cost-dataset gis:load-dataset "environment000_2.asc"
  set motivation-dataset gis:load-dataset "environment000_3.asc"
  set boundary-dataset gis:load-dataset "environment000_4.asc"
  gis:set-world-envelope (gis:envelope-union-of (gis:envelope-of land.or.sea-dataset)
                                                (gis:envelope-of cost-dataset)
                                                (gis:envelope-of motivation-dataset)
                                                (gis:envelope-of boundary-dataset))
end

to set-patch-values
  gis:apply-raster land.or.sea-dataset land.or.sea
  gis:apply-raster cost-dataset cost
  gis:apply-raster motivation-dataset motivation
  gis:apply-raster boundary-dataset boundary
end
```

Code Chunk 10: Import Environments into NetLogo (NetLogo)

This chapter described how agents' decision-making scheme was designed and tested. It also

described the process for creating the reconstructed environments over which the agents moved. The

next step was to bring everything together and test how the model worked with real data.

Chapter 4 Results and Future Directions

The preliminary results of a test run with Low Sea Level data are discussed below. These results show that a data-rich geographically explicit model built using the process detailed above can produce patterns in the distribution of agents. A model developed using this method can be used to formulate and test theories about the ways past environments created opportunities for sea crossings. This chapter also discusses the roads not taken in the development of this model which future developers might explore when trying to understand different aspects of this system.

Before proceeding with analysis of the model results, it should be reiterated that this kind of model, like much of archaeological spatial analysis, should only be "weakly constrained by surviving control data" (Hazelwood and Steele 2004, 673). The archaeological record of this activity is especially incomplete and damaged. Care should be taken to avoid overfitting this model to the archaeological record. Additionally, because this model is so complex, it is better used to inform our understanding of the system being represented than to strongly predict behavior of past people.

4.1. Preliminary LSL Results and Face Validation

Figure 11 shows the NetLogo interface after a 1,000 tick or year run of the model with the ASCII files for the environment with the lowest sea level. Red dots indicate where the agents landed during the run. They can be read like a dot density map. While the data appeared to load correctly (there were no errors and the model ran all the way through 1,000 ticks), the interface behaved strangely. It shows all cells as blue instead of some green (which indicates a difference between land and sea cells). This was a pattern throughout the modeling process. The longer the model ran and the more complex the input data were, the more likely interface issues became. Nevertheless, some emergence is still visible in this interface.
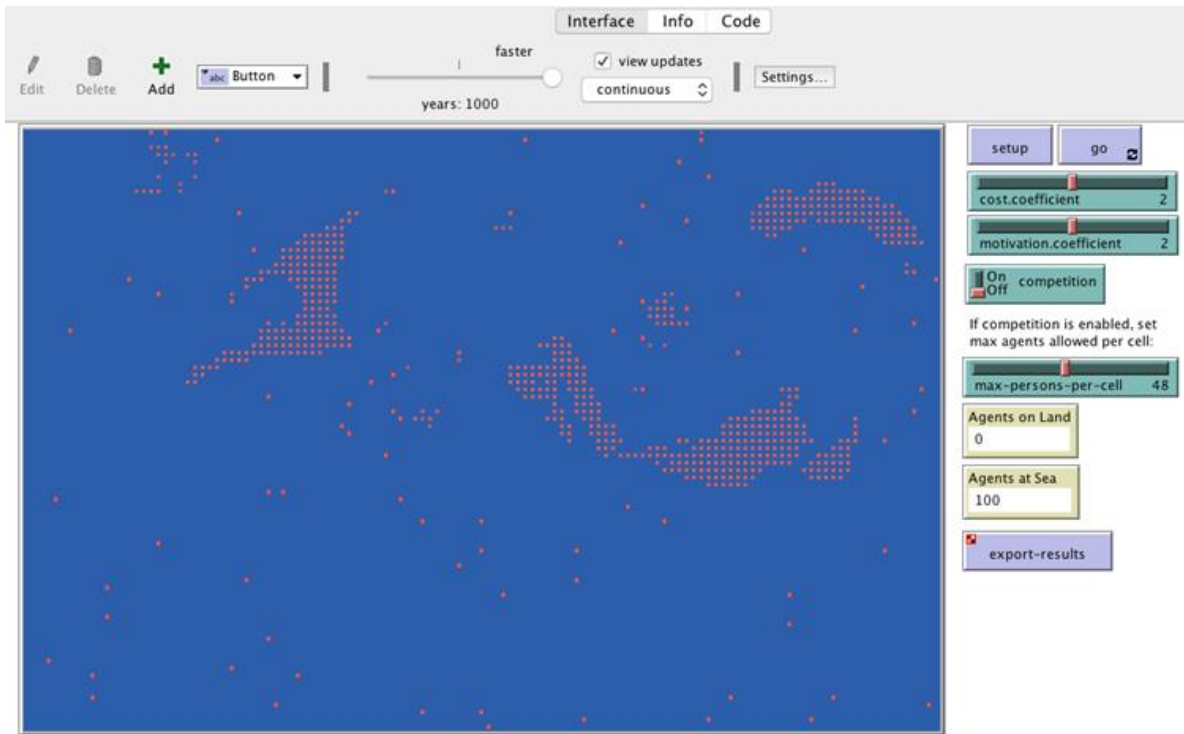
Figure 11: Preliminary Run over LSL

Clearly some non-random pattern has emerged. By comparing this pattern to a cost surface generated from the Low Sea Level data (Code Chunk 11), it is clear that areas of high cost provided barriers to travel, regardless of whether they were on the land or sea. Figure 12 shows the preliminary run results along with the distribution of land and sea and the cost surface for this model run. Agents (the red dots) appear to be clustering in the Eastern Mediterranean, Western Europe, and the Black Sea. The large clusters of agents in the NetLogo interface could be outlined by the dark green regions in the cost surface (where cost – motivation was between 2 and 4).

```
##Cost surface
require(gdistance)

#cost-motivation
LSLcost <- (raster3LSL - raster4LSL)
plot(LSLcost, main="Low Sea Level Cost Surface")
```

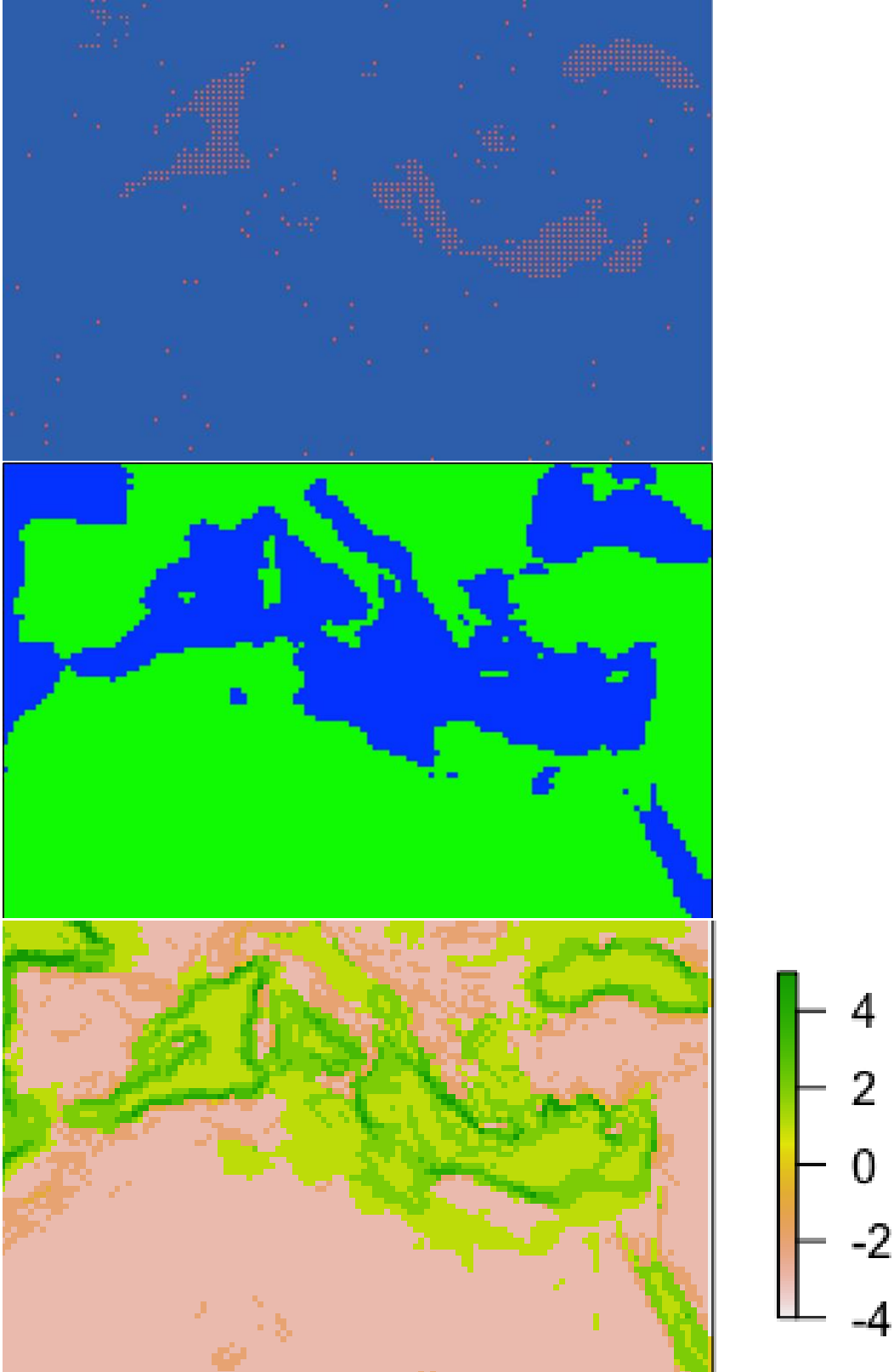Code Chunk 11: Cost Surface for LSL Face Validation (R)

Figure 12: A side-by-side comparison of LSL Preliminary Results, Base Map, and LSL Cost Surface

These preliminary results can be used to understand the implications of the revisionist conceptual model. In this model, the sea itself is not a barrier to travel. Instead this model shows where areas of high slope and high wind may have been a barrier to travel.

4.2. Possible Variations

A model does not need to include every variable that exists in the real system. It only needs to capture the core of the system, the driver(s) that really make it work. Because they were not essential to this model, greater complexity and more variables were not added. However future researchers might be interested in exploring different routes in related models. This section provides guidance for adding new variables and greater complexity to this model.

4.2.1. Adding More Complex Data

Future versions of this model might use directional variables such as aspect in addition to slope, or wind direction as well as wind speed. They might substitute data with finer spatial and temporal resolutions. In this model, rook case (4 proximal cells) was used for spatial analysis and NetLogo decision-making. Future work might use queen case (8 proximal cells). The cells in this agent-based model are aggregated to 34 km X 34 km but if a future model is able to be run on a computer with more RAM or a version of NetLogo not yet released, these cells might be set smaller. Computing power limited this first version of the model but it might not necessarily limit future versions.

While the archaeological record of this activity is too incomplete and damaged to help systematically validate this model, the fossil record might be able to help. As described in Section 1.2, humans might have been pursuing better flora and fauna as well as better climates when they migrated. Fossils of plants and animals they would have relied upon are a relevant dataset that could help validate future version of this model.

4.2.2. Adding Other Variables

In this version of the model, vegetation, in the form of the Aeolian dust proxy, is the main driver. This is justified by our current understanding of the importance of vegetation in Pleistocene migration (described in Section 1.2). Nevertheless, there are still other variables that affect sea crossings and these might be the focus of alternate versions of this model. One of those variables is coastline density. As McEvedy (2002) suggests, denser, more complex coastlines could have increased the interaction between humans and the sea, creating littoral communities and even hotspots for early sea crossings (Broodbank 2013). Conversely, there would be fewer sea crossings where the coastline is longer and straighter. This could help explain why there is more evidence of early sea crossings in the Aegean compared to the rest of the Mediterranean. In the current version of the model, agents only cross the sea where local cost from slope is greater than local cost from wind and vegetation motivation is high enough. Instead, maybe agents should be slightly more likely to cross the sea when they are in a region of relatively dense coastline. The R code that can be used to calculate coastline density is shown below in Code Chunk 12.

```
#coastline draw
raster0.coast <- raster("~/Desktop/seacrossings/raster0")
projection(raster0.coast) <- "+proj=wintri"
summary(raster0.coast)

coast.pol <- rasterToPolygons(raster0.coast, fun=function(raster0.coast){raster0.coast = 0})

fun=function(x){x>3 & x<6}

m.coast <- c(-5300, -1, 0, -1, 1, 1, 1, 4200, 0)
rclmat.coast <- matrix(m.coast, ncol =3, byrow= TRUE)
classified.coast<- reclassify(raster0.coast, rclmat.coast)

hist(classified.coast)
breakpoints.coast <- c(-1, 0, 1)
colors.coast <- c("white", "black")
plot(classified.coast, breaks=breakpoints.coast, col=colors.coast)


#coastline density
m.coast <- c(-5300, -1, NA, -1, 1, 1, 1, 4200, NA)
rclmat.coast <- matrix(m.coast, ncol =3, byrow= TRUE)
classified.coast<- reclassify(raster0.coast, rclmat.coast)
```

Code Chunk 12: Coastline Density (R)

Another variable that might be considered is visibility, especially shore visibility. Seafaring where the shore is not visible would have to involve greater navigation and planning, or else very lucky drifting. This variable was not included in this version of the model because the places where environments created opportunities for sea crossings needed to be represented first before the varying cognitive abilities of agents could be added. However, this model could be adapted to include shore visibility. At the time of this writing, R does not have any viewshed or visibility functions that do not rely on proprietary software. To keep this process open source and easily repeated for large datasets, a work around, shown in Code Chunk 13, was developed. Here the visibility of land from sea is calculated as a function of distance.

```
#visibility

require(raster)
require(gdistance)
require(RColorBrewer)

#import & project base raster, using unaggregated version
raster0.na <- raster("~/Desktop/seacrossings/raster0")
projection(raster0.na) <- "+proj=wintri"

#sea values become NA
raster0.na[raster0.na <= 0] <- NA

#calculate distance for all NA cells (sea) to nearest non NA cell (land)
raster0.distance <- distance(raster0.na)
c <- brewer.pal(6, "Blues")
plot(raster0.distance, col= c)
```

Code Chunk 13: Visibility of Land from Sea (R)

Altering the decision-making horizon might offer the most interesting path for developing this model. If agents can look further ahead spatially and temporally, they might be able to make better decisions about where to move. Variables like visibility offer an indirect way to represent agent planning in an agent-based model. On the other hand, directly changing the decision horizons of agents allows this planning to be explicitly represented in the model. It would be important to see how giving agents the ability to plan further might affect the amount of agent presence on sea cells.

From the perspective of Mediterranean archaeologists and anthropologists in general, this model's decision-making horizon can be an excellent tool for testing theories about voyaging (or drifting) and Neanderthal cognitive abilities. The first iteration of this model deliberately sidesteps this debate to focus more on environmental opportunities. However, it could be further adapted to weigh in on more abstract questions around evolutionary timelines.

4.2.3. Variation in the Agents Themselves

Agent-based modeling offers a powerful tool for exploring how individual decisions can produce the patterns that define a system. In this version of the model, all the individuals make decisions the same way. Yet it might be worth introducing some variation in the decision-making schemes of agents in future versions of this model. Additional data would be needed so that the number of agents in the system and the range of starting points could be more closely represented. Agents could then be programed to group together and cooperate. If bigger, more sea-worthy vessels existed in the Middle to Late Pleistocene, it may have taken more than one person to build and sail them. Agents might also be given a range of decision-making horizons or some could be made to be more risk averse while others could be less risk averse. This modeling effort only begins to explore what an agent-based model of coastal migration during this time could do. Agent-based modeling can offer insights into the gaps in the archaeological record that could previously have only been filled with speculation from competing theories.

Chapter 5 Conclusions and Contributions

This research shows how a data-rich, geographically explicit agent-based model can be designed to allow archaeologists to formulate and test theories about an activity for which there is limited archaeological evidence. Preliminary results show how this model can be used to operationalize the revisionist understanding of human migration in the Pleistocene Mediterranean. While the first version of any one model is unlikely to change a field of research, cleverly designed, well justified, and thoroughly documented code chunks can make a contribution. They can outlive the current limitations of software and they can be useful outside of their original models. Because it is well documented and modularized, this method for designing a model of early sea crossings provides a foundation for future work in this area.

There were computational limitations on this project that might not be present for future researchers. Additionally, this kind of model design is iterative and increasingly the work of teams rather than individuals. With this in mind, steps were taken to ensure this work could be transparent and reproducible and therefore more easily reused. All of the code and data have been and saved to a GitHub (Baumann 2017a). The NetLogo models are also on Modeling Commons (Baumann 2017b). An ODD was written for developers who prefer to work in other programming languages.

Additionally, steps were taken in the writing of this research to modularize the code, so that future researchers could adapt parts of this project to their own work. In this document, R and NetLogo code chunks were described and annotated. Future developers might prefer to only use parts of this model and instead develop their own models. Section 3.5 above breaks apart the creation of the raster layers into separate steps so that future code can selectively borrow from part of this process. Code for optional variables such as coastline density and visibility is also provided separately.

Two code chunks may prove especially useful for future work. Code Chunk 2 is the NetLogo move-persons procedure. It is the translation of the decision-making algorithm into NetLogo code and

the core of this agent-based model. It allows for movement with and without competition. It allows

agents to hatch trails so that their paths can be visualized. This code also allows cost and motivation to

be scaled relative to one another. The sensitivity analysis described in Section 3.3.1 shows that this

algorithm can work and that, when it does, local variation in cost becomes especially important when

trying to understand the effects of a large regional driver like decreasing vegetation.

Appropriately, the second component that might be adapted for related models is Code Chunk

6, which captures some of the most important steps taken when recreating past environments for

agents to traverse. It shows how a new raster should be imported, aligned, and resampled to match an

existing raster. It also includes the important step of setting a projection in R. This code chunk shows

how to reclassify a raster using a reclassification matrix. It demonstrates how to calculate slope and then

remove any NA values (which helps prevent NaN errors during NetLogo math operations). Then the

process of selectively assigning either a slope or a wind cost based on whether the cell lines up with a

land or sea cell in a related raster is demonstrated. Lastly, this Code Chunk shows how raster layers can

be stacked and then called by position like components in a vector. This feature of raster stacks

facilitates the repeated tasks required throughout this model.

5.1.   Final Thoughts on ABM Software

Originally, this model was going to be created using the Esri Agent Analyst Extension which

allows for agent-based modeling in ArcMap. However this extension is not compatible with ArcMap

after version 10.0. Moreover, it would have also been difficult to prepare all 769 rasters in ArcMap.

The next plan was to build this model for R and NetLogo and use the RNetLogo package (Thiele

2014) in R as a bridge. RNetLogo would have allowed NetLogo to be run from R. Environmental input

would have been sent more easily to NetLogo and output would have been much easier to send to R.

Moreover, this package would have allowed this model to run continuously by looping the RNetLogo

commands in R. In hopes that the bridge would function, an intermediate set-up procedure was written into the original NetLogo model so the agent positions would remain the same between runs.

Unfortunately, but not unexpected for open source software, R, NetLogo, and RNetLogo were updated during this project. Challenges with compatibility between the versions might be traced to the rJava package, on which RNetLogo depends. It appears rJava currently runs Java 1.6 but recent versions of NetLogo require Java 1.8. There are other R-NetLogo bridges, including Rserve. However, none had the same capabilities as the RNetLogo bridge.

Ultimately, a more manual approach was used. The environmental variables were exported as ASCII files and saved to the sea-crossings GitHub. Following the GIS General Examples created by Uri Wilensky (2017), these data were imported using the NetLogo GIS extension. Since these data are already projected, a projection did not need to be set in NetLogo. Importing environmental input through the GIS extension is a slower and less reliable process but, at this stage, it is sufficient to demonstrate the model's value as a tool.

Additionally, NetLogo also seemed to struggle with longer, more complex runs even after the settings of the program were changed to allow it to use more RAM. The unexpected behavior of the interface during more computationally complex runs posed a limitation for this research.

All of these issues are likely to be resolved in time. As spatially explicit agent-based models become more popular, software that allows for *both* robust spatial analysis and agent-based modelling will arise. As modeling with big data increases, software to accommodate greater computational complexity in ABM runs will be developed. Because of the pace at which this software moves, the strategies used to design this model and its well-documented, reusable code chunks will be the real contribution of this research. Countless software updates from now, archaeologists will still be asking questions about what people were thinking when they attempted sea crossings a hundred thousand

years ago. When they do, agent-based modeling will still be one of the best ways they can find answers

to their questions and this process and these code chunks will be able to help them.

References

Agustí, Jordi, Hugues-Alexandre Blain, Iván Lozano-Fernández, Pedro Piñero, Oriol Oms, Marc Furió, Angel Blanco, Juan Manuel López-García, and Robert Sala. 2015. "Chronological and Environmental Context of the First Hominin Dispersal into Western Europe: The Case of Barranco León (Guadix-Baza Basin, SE Spain)". *Journal of Human Evolution* 87: 87-94. doi:10.1016/j.jhevol.2015.02.014.

Almogi-Labin, Ahuva. 2011. "The Paleoclimate of the Eastern Mediterranean during the Transition from Early to Mid Pleistocene (900 to 700Ka) Based on Marine and Non-Marine Records: An Integrated Overview". *Journal of Human Evolution* 60 (4): 428-436. doi:10.1016/j.jhevol.2010.03.007.

Amante, C. and B.W. Eakins. 2009. "ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24.". *National Geophysical Data Center, NOAA*. doi:10.7289/V5C8276M.

Ambrose, S. H. 2001. "Paleolithic Technology and Human Evolution". *Science* 291 (5509): 1748-1753. doi:10.1126/science.1059487.

Anderson, Atholl. 2010. "The Origins and Development of Seafaring: Towards a Global Approach". In A. Anderson, J. H. Barrett, K. V. Boyle (Eds.), *The Global Origins and Development of Seafaring*, 1st ed., 3-18. Cambridge, UK: McDonald Institute for Archaeological Research, University of Cambridge.

Baumann, Bailey. 2017a. "Bailey-B/Sea-Crossings". Github. https://github.com/Bailey-B/sea-crossings.

Baumann, Bailey. 2017b. "Models Belonging To Bailey Baumann -- Netlogo Modeling Commons". Modelingcommons.Org. http://modelingcommons.org/account/models/2324.

Bednarik, Robert G. 2003. "Seafaring in the Pleistocene". *Cambridge Archaeological Journal* 13 (1): 41-66. doi:10.1017/s0959774303000039.

Bednarik, Robert G. 1998. "An Experiment in Pleistocene Seafaring". *International Journal of Nautical Archaeology* 27 (2): 139-149. doi:10.1111/j.1095-9270.1998.tb00797.x.

Bettinger, Robert L, Raven Garvey, and Shannon Tushingham. 2015. *Hunter-Gatherers: Archaeological and Evolutionary Theory*. Boston, MA: Springer US.

Bintanja, R. and R. S. W. van de Wal. 2008. "North American Ice-Sheet Dynamics and the Onset of 100,000-Year Glacial Cycles". *Nature* 454 (7206): 869-872. doi:10.1038/nature07158.

Broodbank, Cyprian. 2006. "The Origins and Early Development of Mediterranean Maritime Activity". *Journal of Mediterranean Archaeology* 19 (2). doi:10.1558//jmea.2006.v19i2.199.

Broodbank, Cyprian. 2013. *The Making of the Middle Sea*. Oxford University Press.

Cavaleri, L. 2005. "The Wind and Wave Atlas of the Mediterranean Sea - The Calibration Phase".
    Advances in Geosciences 2: 255-257. doi:10.5194/adgeo-2-255-2005.

Cegielski, Wendy H. and J. Daniel Rogers. 2016. "Rethinking the Role of Agent-Based Modeling in
    Archaeology". *Journal of Anthropological Archaeology* 41: 283-298. doi:10.1016/j.jaa.2016.01.009.

Cherry, J.F. 1990. "The First Colonization of the Mediterranean Islands: A Review of Recent Research".
    Journal of Mediterranean Archaeology 3 (1): 145-221. doi:10.1558/jmea.v3i1.145.

Cherry, John F. and Thomas P. Leppard. 2015. "Experimental Archaeology and the Earliest Seagoing: The
    Limitations of Inference". *World Archaeology* 47 (5): 740-755.
    doi:10.1080/00438243.2015.1078739.

Davies, Benjamin and Simon H. Bickler. 2013. "Sailing The Simulated Seas: A New Simulation for
    Evaluating Prehistoric Seafaring". In A. Traviglia (Ed.), *Across Space and Time. Papers from the 41st
    Conference on Computer Applications and Quantitative Methods in Archaeology*. Amsterdam
    University Press.

d'Errico, Francesco and William E. Banks. 2013. "Identifying Mechanisms behind Middle Paleolithic and
    Middle Stone Age Cultural Trajectories". *Current Anthropology* 54 (S8): S371-S387.
    doi:10.1086/673388.

DTU Wind Energy, Department of Wind Energy. "Global Wind Atlas - Map". 2017. Globalwindatlas.Com.
    http://globalwindatlas.com/map.html.

Entwisle, Barbara, Nathalie E. Williams, Ashton M. Verdery, Ronald R. Rindfuss, Stephen J. Walsh,
    George P. Malanson, and Peter J. Mucha et al. 2016. "Climate Shocks and Migration: An Agent-
    Based Modeling Approach". *Population and Environment* 38 (1): 47-71. doi:10.1007/s11111-016-
    0254-y.

Esri, Inc., 2016. "Winkel Tripel—Help | Arcgis for Desktop". *Desktop.Arcgis.Com*.
    http://desktop.arcgis.com/en/arcmap/10.3/guide-books/map-projections/winkel-tripel.htm.

Erlandson, Jon McVey. 2001. "The Archaeology of Aquatic Adaptations: Paradigms for a New
    Millennium". *Journal of Archaeological Research* 9 (4): 287-350. doi:10.1023/a:1013062712695.

Erlandson, Jon McVey. 2010. "Neptune's Children: The Evolution of Human Seafaring". In A. Anderson, J.
    H. Barrett, K. V. Boyle (Eds.) *The Global Origins and Development of Seafaring*, 1st ed., 19-28.
    Cambridge, UK: McDonald Institute for Archaeological Research, University of Cambridge.

Fitzpatrick, Scott M. and Richard Callaghan. 2008. "Magellan's Crossing of the Pacific: Using Computer
    Simulations to Examine Oceanographic Effects On One of the World's Greatest Voyages". *The
    Journal of Pacific History* 43 (2): 145-165. doi:10.1080/00223340802303611.

Grant, K. M., E. J. Rohling, M. Bar-Matthews, A. Ayalon, M. Medina-Elizalde, C. Bronk Ramsey, C. Satow, and A. P. Roberts. 2012. "Rapid Coupling between Ice Volume and Polar Temperature over the Past 150,000 Years". *Nature* 491 (7426). doi:10.1038/nature11593.

Grimm, V. 2005. "Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology". *Science* 310 (5750): 987-991. doi:10.1126/science.1116681.

Grimm, Volker, Uta Berger, Donald L. DeAngelis, J. Gary Polhill, Jarl Giske, and Steven F. Railsback. 2010. "The ODD Protocol: A Review and First Update". *Ecological Modelling* 221 (23): 2760-2768. doi:10.1016/j.ecolmodel.2010.08.019.

Grimm, Volker, Uta Berger, Finn Bastiansen, Sigrunn Eliassen, Vincent Ginot, Jarl Giske, and John Goss-Custard et al. 2006. "A Standard Protocol for Describing Individual-Based and Agent-Based Models". *Ecological Modelling* 198 (1-2): 115-126. doi:10.1016/j.ecolmodel.2006.04.023.

Hazelwood, Lee and James Steele. 2004. "Spatial Dynamics of Human Dispersals". *Journal of Archaeological Science* 31 (6): 669-679. doi:10.1016/j.jas.2003.11.009.

Heppenstall, Alison J., Andrew T. Crooks, Michael Batty, and Linda M. See. 2012. "Reflections and Conclusions: Geographic Models to Address Grand Challenges". In A. Heppenstall, A. T. Crooks, L. M. See, M. Batty (Eds.) *Agent-Based Models of Geographic Systems*, 1st ed. Dordrecht: Springer Netherlands.

Howitt-Marshall, Duncan and Curtis Runnels. 2016. "Middle Pleistocene Sea-Crossings in The Eastern Mediterranean?". *Journal of Anthropological Archaeology* 42: 140-153. doi:10.1016/j.jaa.2016.04.005.

Lambeck, Kurt, Anthony Purcell, Nicholas. C. Flemming, Claudio Vita-Finzi, Abdullah M. Alsharekh, and Geoffrey N. Bailey. 2011. "Sea Level and Shoreline Reconstructions for the Red Sea: Isostatic and Tectonic Considerations and Implications for Hominin Migration Out of Africa". *Quaternary Science Reviews* 30 (25-26): 3542-3574. doi:10.1016/j.quascirev.2011.08.008.

Larrasoana, J. C., A. P. Roberts, E. J. Rohling, M. Winklhofer, and R. Wehausen. 2003. "Three Million Years of Monsoon Variability over the Northern Sahara". *Climate Dynamics* 21 (7-8): 689-698. doi:10.1007/s00382-003-0355-z.

Laskaris, N., A. Sampson, F. Mavridis, and I. Liritzis. 2011. "Late Pleistocene/Early Holocene Seafaring in the Aegean: New Obsidian Hydration Dates with The SIMS-SS Method". *Journal of Archaeological Science* 38 (9): 2475-2479. doi:10.1016/j.jas.2011.05.019.

Leidwanger, Justin. 2013. "Modeling Distance with Time in Ancient Mediterranean Seafaring: A GIS Application for the Interpretation of Maritime Connectivity". *Journal of Archaeological Science* 40 (8): 3302-3308. doi:10.1016/j.jas.2013.03.016.

Leppard, Thomas. 2014. "Modeling the Impacts of Mediterranean Island Colonization by Archaic Hominins: The Likelihood of an Insular Lower Palaeolithic". *Journal of Mediterranean Archaeology* 27 (2): 231-254. doi:10.1558/jmea.v27i2.231.

Leppard, Thomas P. 2015. "The Evolution of Modern Behaviour and Its Implications for Maritime Dispersal during the Palaeolithic". *Cambridge Archaeological Journal* 25 (04): 829-846. doi:10.1017/s0959774315000098.

Lykousis, V. 2009. "Sea-Level Changes and Shelf Break Prograding Sequences during the Last 400Ka in The Aegean Margins: Subsidence Rates and Palaeogeographic Implications". *Continental Shelf Research* 29 (16): 2037-2044. doi:10.1016/j.csr.2008.11.005.

McEvedy, Colin and John Woodcock. 2002. *The Penguin Atlas of Ancient History*. 2nd ed. Harmondsworth, Middlesex, England: Penguin Books.

Mithen, S. 2002. "Stepping Out: A Computer Simulation of Hominid Dispersal from Africa". *Journal of Human Evolution* 43 (4): 433-462. doi:10.1016/s0047-2484(02)90584-1.

Murray, William M. 1987. "Do Modern Winds Equal Ancient Winds?". *Mediterranean Historical Review* 2 (2): 139-167. doi:10.1080/09518968708569525.

Muttoni, Giovanni, Giancarlo Scardia, and Dennis V. Kent. 2010. "Human Migration into Europe during the Late Early Pleistocene Climate Transition". *Palaeogeography, Palaeoclimatology, Palaeoecology* 296 (1-2): 79-93. doi:10.1016/j.palaeo.2010.06.016.

Ngo, The An and Linda See. 2012. "Calibration and Validation of Agent-Based Models of Land-Cover Change". In A. Heppenstall, A. T. Crooks, L. M. See, M. Batty (Eds.) *Agent-Based Models of Geographic Systems*, 1st ed. Dordrecht: Springer Netherlands.

Nowell, April and Iain Davidson. 2010. *Stone Tools and the Evolution of Human Cognition*. Boulder, Colo.: University Press of Colorado.

O'Sullivan, David and George L. W Perry. 2013. *Spatial Simulation*. Wiley-Blackwell.

O'Sullivan, David, James Millington, George Perry, and John Wain Wright. 2012. "Agent-Based Models - Because They're Worth It?". In A. Heppenstall, A. T. Crooks, L. M. See, M. Batty (Eds.) *Agent-Based Models of Geographic Systems*, 1st ed. Dordrecht: Springer Netherlands.

Palombo, Maria Rita. 2010. "A Scenario of Human Dispersal in the Northwestern Mediterranean throughout the Early to Middle Pleistocene". *Quaternary International* 223-224: 179-194. doi:10.1016/j.quaint.2009.11.016.

Papoulia, Christina. 2016. "Seaward Dispersals to the NE Mediterranean Islands in the Pleistocene. The Lithic Evidence in Retrospect". *Quaternary International*. doi:10.1016/j.quaint.2016.02.019.

Rohling, E. J., G. L. Foster, K. M. Grant, G. Marino, A. P. Roberts, M. E. Tamisiea, and F. Williams. 2014. "Sea-Level and Deep-Sea-Temperature Variability over the Past 5.3 Million Years". *Nature* 508 (7497): 477-482. doi:10.1038/nature13230.

Romanowska, Iza. 2013. "Agent-Based Modeling and Archaeological Hypothesis Testing: The Case Study of the European Lower Palaeolithic". In A. Traviglia (Ed.), *Across Space and Time. Papers from the 41st Conference on Computer Applications and Quantitative Methods in Archaeology*. Amsterdam University Press.

Runnels, Curtis. 2014. "Early Palaeolithic On the Greek Islands?". *Journal of Mediterranean Archaeology* 27 (2): 211-230. doi:10.1558/jmea.v27i2.211.

Simmons, Alan H. 2014. *Stone Age Sailors*. Walnut Creek: Left Coast Press.

Strasser, Thomas F, Eleni Panagopoulou, Curtis N Runnels, Priscilla M Murray, Nicholas Thompson, Panayiotis Karkanas, Floyd W McCoy, and Karl W Wegmann. 2010. "Stone Age Seafaring in the Mediterranean: Evidence from The Plakias Region for Lower Palaeolithic and Mesolithic Habitation of Crete". *Hesperia* 79 (2): 145-190. doi:10.2972/hesp.79.2.145.

Straus, Lawrence Guy. 2001. "Africa and Iberia in the Pleistocene". *Quaternary International* 75 (1): 91-102. doi:10.1016/s1040-6182(00)00081-1.

Subcommission on Quaternary Stratigraphy. "Formal Stratigraphical Definitions". 2010. http://quaternary.stratigraphy.org/definitions/pleistocenesubdivision/.

Thiele, Jan C. 2014. "R Marries Netlogo: Introduction to The Rnetlogo Package". *Journal of Statistical Software* 58 (2). doi:10.18637/jss.v058.i02.

Voris, Harold K. 2000. "Maps of Pleistocene Sea Levels in Southeast Asia: Shorelines, River Systems and Time Durations". *Journal of Biogeography* 27 (5): 1153-1167. doi:10.1046/j.1365-2699.2000.00489.x.

Wang, Pinxian, Jun Tian, and Lucas J. Lourens. 2010. "Obscuring of Long Eccentricity Cyclicity in Pleistocene Oceanic Carbon Isotope Records". *Earth and Planetary Science Letters* 290 (3-4): 319-330. doi:10.1016/j.epsl.2009.12.028.

Wilensky, Uri. 2017. "GIS General Examples, (Model ID 2354) -- Netlogo Modeling Commons". *Modelingcommons.Org*. Accessed April 28. http://modelingcommons.org/browse/one_model/2354#model_tabs_browse_procedures.

Wilson, Greg, D. A. Aruliah, C. Titus Brown, Neil P. Chue Hong, Matt Davis, Richard T. Guy, and Steven H. D. Haddock et al. 2014. "Best Practices for Scientific Computing". *PLoS Biology* 12 (1): e1001745. doi:10.1371/journal.pbio.1001745.

Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. 1st ed. Cambridge (MA): Addison-Wesley Press.

Appendix A: GitHub Table of Contents

This section lists of files that can be found in the GitHub repository for this project (Baumann, 2017a).

https://github.com/Bailey-B/sea-crossings

- ASCII-environments
    - (Contains ASCII environments for low, median, and high sea level time slices)
- Chapter 3 R Markdown
- Chapter 4 R Markdown
- NetLogo
    - seacrossings000.nlogo (which can be used with ASCII files)
    - seacrossingsdummydata.nlogo (which can be run without the ASCII files)
    - (Note that these models are also hosted on Modeling Commons: http://modelingcommons.org/account/models/2324)
- RMarkdown-images
    - Conceptual Model Diagram
    - NetLogo Decision Schemes
    - Cost and Motivation Sliders
    - Dummy Data Output
    - Real Data Output
- raster0
    - dblbnd.adf
    - hdr.adf
    - metadata.xml
    - prj.adf
    - sta.adf
    - vat.adf
    - w001001.adf
    - w001001x.adf
- windraster
    - dblbnd.adf
    - hdr.adf
    - metadata.xml
    - prj.adf
    - sta.adf
    - vat.adf
    - w001001.adf
    - w00100xadf
- README.md
- larrasoana2003.csv
- sea level worksheet.xlsx
- sealevels.csv
- worksheet for initial SA.xlsx