ABSTRACT

As soon as we will comprehend the inner mechanisms of human societies, then we will be able to answer pragmatical and existential questions about ourselves and the world we live in. Anthropological models allow us to understand these sophisticated concepts by identifying correlations. Artificial societies will explain and prove these correlations between "micro-level cognition and macro-level social behavior". The social simulation field currently focuses more on specific situations (evacuation, crowd control and traffic optimization) and less on general social simulations. While anthropological and archaeological simulations have gained momentum and support in the last years, it is still a relatively undeveloped and unpopular field.

We implement a custom agent-based system on a scalable and modular architecture that functions as the foundation for complex social simulations of artificial societies. The world is generated gradually by creating the environment, resources, and humans. It is then simulated by applying the general dynamics of the world and allowing the agents to take action. Using the designed system we conducted an experiment on the paleoecology of early hominids. The purpose was to create and simulate a population of early humans (Homo S. Sapiens) in the late Pleistocene of Europe (sometime between 40'000 - 12'000 years ago). The spatially explicit models provide accurate environmental attributes (temperature, altitude, geography, reconstructed ancient forests, and other resources) that are emulated based on a unique dataset inspired by satellite collections and scientific research. Human agents possess the abilities needed to perceive the environment, imagine scenarios, rationalize outcomes, decide what is best for them, and act upon the world or other agents. The decision-making processes are using decision trees trained on historical data extracted from the simulation world. The available actions are: moving, scavenging food, eating, and interaction (socialize, fight, copulate, share food). All micro-level attributes and parameters are modelled based on anthropological, biological, geographical, botanical, and zoological statistical knowledge. Validation is achieved by comparing the emergent phenomenon of population mortality to data from archaeological studies. It is also supported by the system which offers repeatability of experiments with slight variations and scalability by generating and emulating different resolutions of the world.

Contents

List of Figures

List of Tables

List of publications and presentations

- **Radu Galan** and Gabriela Czibula, *Generation and Simulation of Artificial Human Societies using Anthropologically Modelled Learning Agents*, IEEE Access journal, 2022, to be submitted
- **Radu Galan**. *Generation and Simulation of Artificial Human Societies using Anthropologically Modelled Learning Agents*. 2nd International Workshop on Applied Deep Learning, WeADL 2022 (`http://www.cs.ubbcluj.ro/weadl/program/`).

Chapter 1

Introduction

1.1 Key concepts

Human behavior modelling is the simplification process of a general human. Attributes, potential actions, sensory inputs, internal dynamics and cognitions are all important during this process. The goal is to create a viable computational model that can describe as many humans as accurately possible with as few parameters. Knowledge from social sciences is essential during this process in order to properly describe the essence of a human [Ken12].

Agent-based modelling is a computational design that creates a system where autonomous entities (the agents) can interact with the environment (through senses or actions) and with each other (through communication and actions). They usually possess some knowledge and take action based on a set of rules that can be fixed or dynamic (in which case they are considered to learn and adapt).

Artificial Societies represent an agent-based computational model that has the purpose of describing a complex system such as a real human society. They use simple attributes and rules to create complex emergent phenomenons [DF94].

Social Simulation will usually apply a set of constraints or new rules to an artificial society in order to prove or disprove theories from sciences like psychology, law, politics, economics, anthropology, linguistics, and others. While usual theories in social sciences rely strongly on arguments and descriptions, it does not offer formal proof like natural sciences. This could be fixed probabilistically by using simulations [Dre05].

1.2 Motivation

Simulations (and specifically Social Simulations) are a relatively old concept that has had enough time to gain maturity in the field of computer science. Whilst through the folklore they have achieved a supernatural reputation, the actual speciality literature has only cracked the surface of this field. The philosophical and theoretical formalizations of the concepts have not failed to be already thoroughly explored.

The goal would be to be able to run universal simulations that can answer any question we have about the real world (usually to improve decision-making processes). A few examples of universal questions are:

- What should be changed to have faster and safer transportation?
- How can we optimize the economical growth of a sector?
- How can we detect corruption in a political system?
- What is a weak spot in a city plan design?
- What is the best decision in a trial?
- What would happen if religion would disappear?
- Why did a specific species go extinct?
- How did the early humans develop certain traits?

Work in the field has already attempted answering a few problems that are contained in some of the questions above (but never to the entire question and with arguable accuracy nonetheless). The mentioned questions are of course very general and that makes them exponentially harder to answer (if not impossible altogether). Simplifying the simulation and focusing on a selection of parameters is a necessary evil in this situation. There will always be yet another factor in the equation that may seem irrelevant in analysis but in reality, there could be a hidden mechanic changing the outcome based on that very factor. We do not possess perfect data, absolute statistics, or unlimited computational power. Complete and correct simulations will always be out of reach but accuracy should constantly improve along with progress in other adjacent fields.

With a great increase of resources (computational and data-related) comes a great responsibility of extending the previous research by expanding the scope of the simulations, augmenting the attributes with updated real-world data, calibrating the parameters according to the new research, optimizing for the latest hardware innovations, and generally approach more universal solutions.

The motivation for this work and simulation-related work in general lies in the constant need for optimization and the drive for more accurate answers.

1.3 Project description

Given the previous arguments, this project has a relatively wide scope of creating motivation for further research and interest in the future of the field. However, it still intends to put forward practical proof that there still is exploitable potential.

The problem to be solved is: Discovering and proving correlations between the micro-level human cognition and macro-level social behavior in various societies. The definition could be extended from human sociology to zoology, botany, and general ecology.

The solution proposed uses social simulation implemented as an agent-based system with a scalable architecture, flexible modelling, ease of synchronisation with external data sources, and potential for improved correlation proving. We are going to focus on human correlations as a leading experiment due to interests, increased complexity of the model, and availability of the data. Extensions would only alter the data analysis component (but should inquire different emphasis during modelling for better accuracy) and not change the rest of the system.

The development will be presented in the same manner as it was constructed: based on iterations. The first stage was creating a reliable and flexible architecture that could withstand good scalability. Next, the modelling stage began that was designed based on the previous step all the specifications of each component (the environment agents, the resource agents and the human agents). The implementation and experimentation came again in several iterations. This stage would begin with a testing experiment for the environment and then would increase gradually the complexity until the end goal was reached (from environment data iteration, resource population and generation, to a non-intelligent human agents and finally intelligent human agents).

1.4 Original contributions

First of all, this work aims to cover an area of the research field that seems to be scarcely represented. Most public works in the field of agent-based simulations are specialized in solving very particular problems (fire evacuation, panic simulation, train traffic optimization, as discussed in Section 3.1). There are way fewer studies involving general social simulation (the few that were identified are discussed in Section 3.2). Whilst the very specific results with pragmatical conclusions are obviously easier to obtain with specialized simulation, many important questions are left unanswered. Fortunately, in the last years, many works started covering anthropological and archaeological simulations of hominins and other animals.

Secondly, it has been observed that even the more general social simulation usually lacks some very important features needed to gain credibility. Since the focus is usually on the creation of certain emergent mechanics through the right rules and dynamics, some other aspects are disregarded.

- The data used in attributes is often generated from scratch instead of iterating over real-world data
- Random attribute generation and random decision making are used instead of statistical approaches (based on social research studies)

The aforementioned aspects are considered in this work pillars for building better simulations: higher accuracy, bigger scale, and most importantly more reliable validation. We aim to start solving those issues and more. Some other specific contributions of this work are:

1. a unique accurate **spatially explicit** content
2. the **validation** through measured emergent phenomenon similarity, repeatability and scalability
3. the meticulously **modelled environment and agents**
4. **scalability** at different **resolutions** and agent counts,
5. ease of **designing and integrating any new agents** with various data sources easily
6. **modularity** to replace reliably logical functions or configurations

1.5 Thesis structure

This first chapter (1) presented the project idea and its description along with a simplified explanation of the key concepts and the general motivation for the field. We will continue (chapter 2) by introducing the history of the field with some important properties about agent-based systems, social simulations and other specific theoretical information used in the modelling phase. Much of the remaining relevant information will be described during the next chapter (3) by reviewing a selection of projects in the main formed trends of simulations.

The following chapters will delve into our work. Chapter 4 follows the design (planning, flow, architecture) and development stage (data gathering, modelling, and implementation). Modelling is rather important and will be discussed thoroughly. The experimentation stages are presented in chapter 5 where we build the target simulation from scratch by adding iterative layers: the environment, resources, and humans. In each stage, we present explanations, visualization and analysis. For the last stage, we also validate the process using proper metrics. We conclude by dissemination of the results and planning the future steps in chapter 6.

Chapter 2

Theoretical background

This chapter will organize all prerequisite knowledge that has been referenced during the thesis. General information about the field and the topics to which the work belongs will appear in the first sections. The following sections will contain information from adjacent fields related to the modelling phase especially. Information is used as modelling foundation from studies in many other fields: anthropology, archaeology, psychology, etc.

2.1 Agent-based simulations

Around the year 1998 interest started to spike for agent-based systems and many applications and implementations started to appear. All of them intended to combine a pleasant interface and visualizations with optimized algorithms easy to integrate in experiments. Some of these software for modelling and simulations are: NetLogo [TW04], StarLogo (still constantly updated today) [Res96], Repast [NM07], MASON [PL05], [AN11]. However, there are tens of new software applications and libraries in use today by scientists all over the world in many languages [PLPG20].

Based on publication frequency and quality that were approximated during this research there were noticed a few relatively distinct time periods. After the initial spike of interest, we can see many problems from different areas tackled by researchers using social simulations. Multiple considerable grants were allocated to the field [cor] [gena]. Afterwards followed a period of critique against simulations that gained traction (and with very good reasons). This critique never ceased and there are even opinions that simulations will never be realistic enough to correctly predict important phenomenons. Since 2010 there is a growing trend in the field being observed. The critiques and the research on simulation worked adversary with an overall positive impact. We witnessed the introduction of multiple new validation mechanisms and a developing good habit of replacing randomness with real-world data (data explicit simulations) and scientifically accurate probability distribution for attribute and agent generations.

2.2 Social simulations

2.2.1 Motivation

The motivation for this type of research in the 'Journal of Artificial Societies and Social Simulation' is justified and truly captivating when referring to this direction of research. It can be simply explained as **"artificial societies offer insight into the relationship between micro-level cognition and macro-level social behavior"** [Dre05]. The previous quote truly encapsulates so much regarding this work and the ways social simulations should work. This potentially great method can also be applied for understanding and predicting human behavior on a more essential level. We are referring to creating general simulations that are meant to enhance our understanding of humans, and their interactions with the world and with each other. This happens when the experiment's purpose is not optimizing a predefined metric (like the maximum number of cars on a street sector, the average time of exiting a building, etc) but analyzing the effects of altering different attributes (representing environmental factors or human features and reasonings) on the overall complex system (representing the society).

Other motivations appeared from the renowned Joshua M. Epstein. He argues that 'growing' a concept/phenomenon/society creates a whole new level of prerequisite understanding of matters. Only when we can reproduce it consistently, then we have understood it completely [Eps12]. The field of social simulations developed profoundly (but also laterally into other fields through symbiotic projects) in the last decades. He is often regarded as the populizer of the term: 'Artificial Societies'.

On a more practical note, this research's traction increase in recent years is strongly correlated with the growing computational capabilities and the rising interests of the government for political predictions (like predicting the reaction of a certain population as to avoid conflict and optimize prosperity).

2.2.2 History

The Progress from algorithms with rudimentary convolutions towards complex spatially explicit simulations with intelligent agents feels quite organic and natural. The motivations produce in time a constant increase in scale, complexity, and accuracy. Along with this evolution different branches of specialization and generalization formed to fit all the different classes of questions.

Simple specific simulations have been used for decades already to prove and solve different concepts. They require little data and cover small searching space and functions with limited dynamics. The purpose is usually to find an optimal dynamic for the system that minimizes a specific metric. There

General social simulations started to branch off in the field after some first experiments sparked interess. These experiments are Conway's game of life [Gam70] and Schelling's segregation model [HB12]. They managed to create very interesting and even controversial emergent phenomenons with simple rules integrated into the system. It is a simple environment with simple (but well modelled) rules and agents. After the spark created by these experiments many other works appeared. Libraries and tools were created, studies for multiple purposes got initiated, and funding started to be directed toward these areas. While at the beginning they did not show much potential for useful pragmatic solutions, in the last years that changed. We see now archaeologists and anthropologists constantly working with simulation tools.

Wild life spatially explicit simulations

The very precursor to a complete anthropological experiment seems to be research with simulations of the animal kingdom. The questions are very similar but they are addressed to different species. Many animals are monitored and taken care of all-around to globe and there is usually a lot of data available for simulations to integrate. The various simulations of this kind can be used to indicate migrations, movement trends, and even predict issues caused by resource insufficiency, demography changes or habitat deterioration/intervention. A beautiful way to counter (and hopefully prevent) some of the negative impacts humans have on the ecosystem.

Interest in visualization and open source projects also increased in recent years with the evolution of social platforms and especially in video games and the movie industry. We can see in the entertainment industry money being pumped towards complex animations picturing sophisticated societies. Some of them are real and some of them imaginary but all of them require a level of accuracy to receive credibility. In the last decade, we've witnessed the evolution of easy to use graphical engines that can animate such complex societies, tools that automate many of the tedious parts, and many artists and creators that create visualization and animations.

Complete anthropological simulations

Although there were a few early works around the 2000s in the field of paleoecology simulations focused on the early humans (and other hominins), not so many researchers ventured into this kind of experimentation. The ones that appeared developed simplistic simulations in generated environments with few attributes and limited dynamics. It only made sense to focus on specific effects, for example, foraging and food sharing. They are somewhat constrained by computational limitation (that are needed to meet spatially explicit and complex behaviours or learning) and inter-disciplinary research(for accurate traits and attribute) distributions. However, in the last years more ambitious work with considerate progress was made.

This class of simulations is by far the most relevant considering the similarity to our work. Although the literature is scarce, comparatively speaking, there still are a few impressive works of anthropological simulations. These are artificial human societies that are spatially explicit and use the latest methods of validation (data curating, expert modelling, experiment repetition, research-based statistical distributions, etc). They take into consideration a plethora of factors from calorie intake of different plant species to complex interaction and communication mechanisms. Without a doubt, from our scope of research, this can be considered the pinnacle of practical social simulations mainly due to their raw complexity and expert construction.

2.3 Practices in the field

A very important property of correctly constructed simulation is spatially explicitness. It started forming when scientists realised that the next step in simulations is to model real data into the simulation (usually as attributes). It refers to the fact that the environment is not generated anymore (randomly or based on statistics inspired from literature) but initialized or even iterated using a dataset collected from the real world (directly from sensors or indirectly through reconstructions). If data is iterated completely then it is called emulation.

Validation is normally an issue. There is no rigorous optimization process to find which parameters of the human model would perform best or worst because this would imply that there exists a valid metric that quantifies human progress. This formula does not exist since any supposition would be culturally imposed and finding such a formula is not in the scope of this research. However, we can identify correlations between cause (human attributes) and effect (social emergent behavior) using statistical data from the experiments. In order to ensure the validity of the conclusions, we must minimise the modelling error. In consequence, modelling with as much accuracy the human and environment is of the greatest importance.

2.4 Modelling

Population count at our selected time (40’000 - 11’000) in history is relatively uncertain. We have approximations from a few sources [MJ] [KBVD11] point to populations of size between 2 and 4 million about 10’000 years ago. Studies approximating population farther back usually rely on genetic differences of the mitochondria and they range from 0.1 to 0.5 million people. We can also take into consideration the mass extinction from 70’000 years ago (that may or may not have a connection to the eruption of the Toba super-volcano) which decimated the population to about 1’000 - 10’000 members. Extrapolating we can assume vaguely that the population in our chosen time period (sometimes between 40’000 to 11’000 years ago) would be somewhere in the interval of 10’000 to 100’000 worldwide. We will begin by simulating smaller numbers and increase up to this interval maximum.

The group size of early homo sapiens has been examined in many ways over the years. One interesting and notorious study [Dun92] correlates the neocortex size of primates with a limit in social group size. This is called the ’Dunbar’s number’ and is somewhere between 100-250. It must be mentioned that the early humans didn’t just live in one tribe for the entirety of their lives. The dynamics of their groups are not very well documented but based on modern existing tribes and other living primates certain behaviors are suspected. Tribes are formed and split with time usually due to the new generation of males (sometimes peacefully and sometimes with conflict). However, tribes are part of bigger bands that from time to time would interact for multiple reasons (trading, traditions, hunting, common foraging grounds, mating, etc). Many studies point to a very small tribe size (about 10) as being the most optimal hunter-gatherer groups but it has been proven that early sapiens would take care of wounded and elders (which would make the groups far from optimal in the measured metrics). Another factor worth considering is the fact that during the glaciation in Europe population plummeted and group sizes were reduced. Averaging out all the points mentioned we will approximate tribe size for our experiment at about 50 individuals.

Mortality is an important emergent property of early civilizations that leads to them be either thriving or ending. It will also help us validate the simulations using conventional metrics. The longevity of agents has been studied in multiple ways and received accuracies reasonable enough to be credible. Methods range from the calculation of the genealogical tree based on DNA differences between members over time to frequency and count approximations based on archaeological sites. We extracted our mortality of hunter-gatherers from the study provided by [GK07] that covers multiple categories of early human tribes. The numbers we needed are similar or correlated to multiple other sources [Fin10] [SSP+20] [wor].

Chapter 3

Related work

Comparing with other fields in computer science the simulation using agent-based models are relatively unpopular. Furthermore, comparing social simulations with the rest of the simulations they prove again to be less popular. The trend is understandable since it is much easier to receive results and, in the process, answer more specific questions. Smaller and simpler simulations that require little data and fewer assumptions are also more reliable.

We will begin by describing these specific simulations and then continue to present the evolution in the field of social simulations. They begin with rudimentary rule simulation and constantly add complexity to achieve great work of population simulation in spatially explicit environments with scientifically modelled agents.

3.1 Scientific works of specific simulation

By applying basic knowledge of human behavior to these agent-based simulations important experiments were created that intended to predict (and ultimately prevent) extreme scenarios. Some examples for this are:

- aircraft evacuation [SSP08a];
- building evacuation on fire emergency [THL15];
- general evacuation and crowd control [SSP08b] [MMKI02];
- traffic flow optimization on the roads [PPK02] or in the air [WJE+1];
- electronic market analysis [AL11];
- corruption detection [ZZVG20];
- epidemiological prevention [Dun05] [TAS+20] [HGX10].

Undoubtedly, these studies have constructed a very robust answer to their formulated question and many of them have provided useful improvements to their specific field. Simulations prove useful no matter the scale on which they are applied (if used correctly).

However, most of these simulations have very definitive differences from what we consider universal social simulation should mean:

- a very specific question (and in consequence goal);
- limited data inputs for realistic environments;
- rigid environment and simplistic dynamics;
- non-scalable agents;
- reduced simulation runs for statistical validation;
- and very importantly a predetermined formula of validation (which never really occurs naturally in social fields).

An ideal work should consider these points in developing better simulations for answering more general questions about humans and their societies.

3.2 General social simulations

An essential simulation was Schelling's experiment of segregation [HB12]. With a very simple rule and almost no attributes, it can be proven that residential neighbourhoods will tend towards ethnic segregation. If one individual is reticent about living next to a different ethnic group then one can move. Probably the simplicity of the experiment is in part responsible for the credibility that it received. However less specific questions will require more general simulations and this is a situation in which we knew exactly how to ask the question.

Sugarscape [RSZ07] is very interesting work because it has been developed in the Netlogo software (mentioned in Section 2.1) in a time of relative crisis for the world of agent-based simulation (as I debated in Section 2.1). Along with many other works, it implemented more complex systems with a wider variety of attributes and better visualizations in the data analysis methods. The project simulated a basic resource growth (sugar) and integrated an agent population (although not explicitly mentioned the pattern corresponds to KISS as described in Section 2.3. The wealth distribution is studied through these simulations and the consequences of inequality are proven.

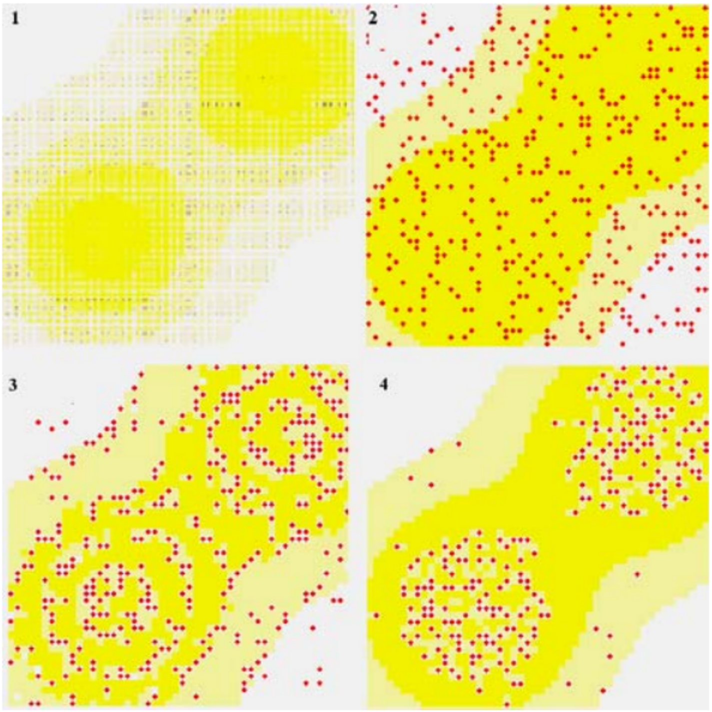


Figure 3.1: Sugarscape simulation [RSZ07]

One study [KAM20] created moving simulations by introducing an interesting concept for crowd decision making. Each turn a random agent is selected to be independent while the rest of the agents in the group are dependent. The latter class of agents only follow the independent agent. It was proven (using clustering) that their movements can be characterised by a Gaussian distribution. However, in a primate society, the independent agent should not be chosen randomly since it is usually the dominant member that makes the decisions. Also, they admit that the limited number of parameters makes certain assumptions (Figure 3.2).

Work conducted by UCL and the University of Leeds called the Genesis project [gena] focused on social policy and urban planning. While there is a chapter regarding their work available [genb] there are very few validations or visualization available since their actual outputs were in other adjacent fields (digital geography, geometry, environment mapping, etc). However, based on their code repository they managed to develop two simulations. The first one uses births, death, and migration probabilities to simulate daily population movements. The second one simulates traffic movements by the second.

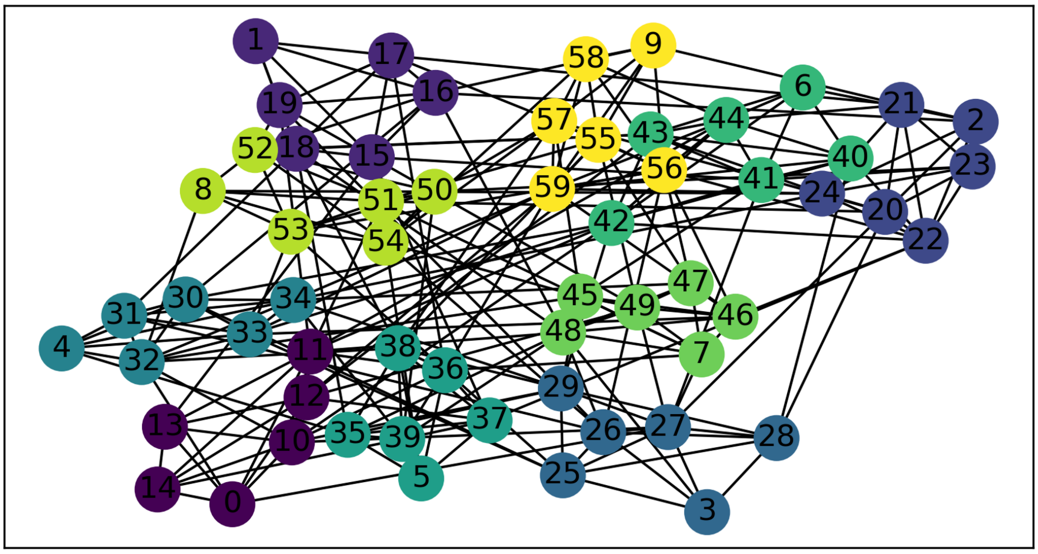


Figure 3.2: The estimated social network in the simulation [KAM20]

3.3 Wild life simulations

Wild life simulations have been attempted before but rarely with concrete data sources and a spatially explicit approach. We are going to mention two works that seemed relevant. A very good example of this is the 2011 work into capuchins populations in Brazil using agent-based modelling [BHP+11]. This geospatial simulation (which is almost spatially explicit) takes into consideration important factors from known research like: food resources, available areas to explore, foraging techniques, and social interactions. As validation, the information currently available about their population was used to compare with the simulation (Figure 3.3).

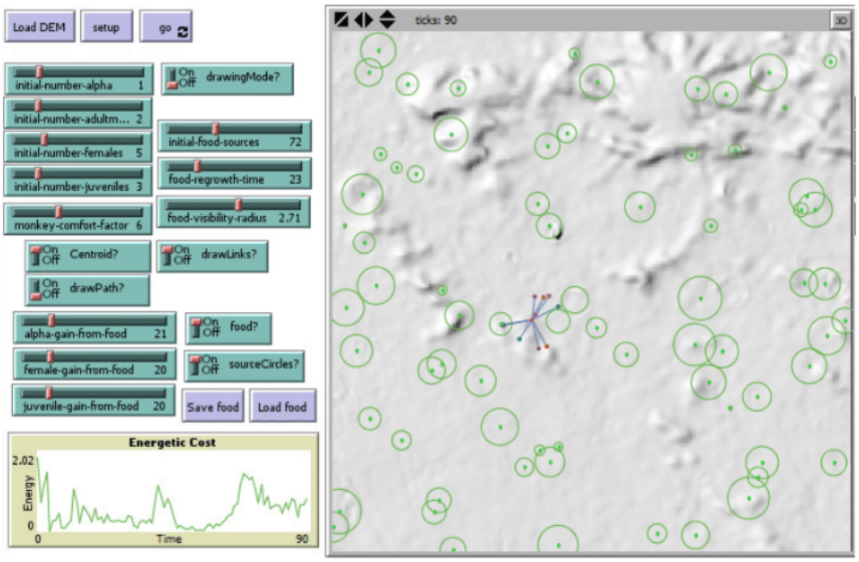


Figure 3.3: The visualization of resources and agents in a simulation of the capuchin tribe [BHP+11]

Another very interesting project studies jaguars in Central America (more exactly in forests and farmland areas) [WNF+15]. The purpose is the reproduce population dynamics and movement patterns. Relying on a spatially explicit approach they refer to data from camera traps and expert knowledge. This data is also used for validation of the jaguar approximated headcount and home range size. The most important mechanics simulated are: the feeding process, reproduction and death events. Using an accurate model it is possible in the future to test the impact on the jaguar population of any landscape changes.

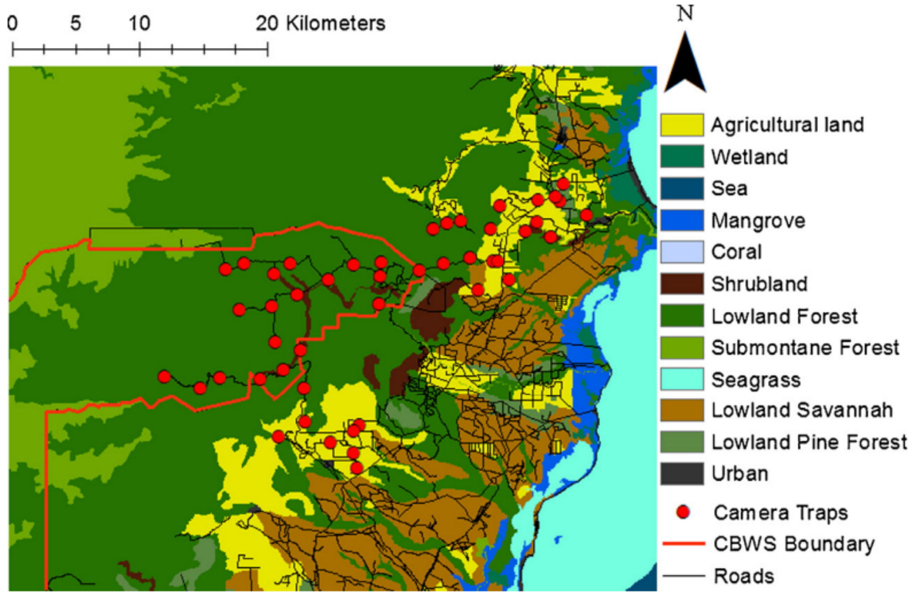


Figure 3.4: The jaguar habitat geography and points of data gathering [WNF+15]

3.4 Anthropological Social Simulations

The next works are the most relevant to our study and will pose an inspiration for any future work. The similarities and differences between these studies will further be presented in Chapter 6

The work of [GLS10] represents an agent-based simulation which is especially interesting being one of the first works of this complexity. The purpose is the simulate an area in East Africa (in the arid Turkana basin, the Voi River channel in Tsavo East) populated with Australopithecus boisei specimens (early hominids) approximately 2 million years ago. They designed multiple scenarios and ran simulations multiple times (for a total of 90 experiments). Empirical data has been used to model the spatially explicit landscape (as can be seen in Figure 3.5). These proto-humans have been modelled with relatively simple actions (forage, nest and sleep, eat) that are triggered based on a prescribed rule-based approach. They have a target calorie intake and until they reach it (or their belly is full) then they will continue to scavenge as long as the day is not over (12 hours split into one minute iterations). Although communication is minimum they can come together in the nest or work together to forage big carcasses (representing animal food sources). Other means of eating include: finding one of the many species of plants that offer fruits (scientifically modelled to only ripe in certain seasons and give a fixed caloric intake) or using digging tools to find root plants. There are experiments in 3 different phases: channel, flooded and unflooded; each with different particularities. They were able to properly simulate food resources and approximate trends and adaptations the populations needed to do in order to survive and thrive.

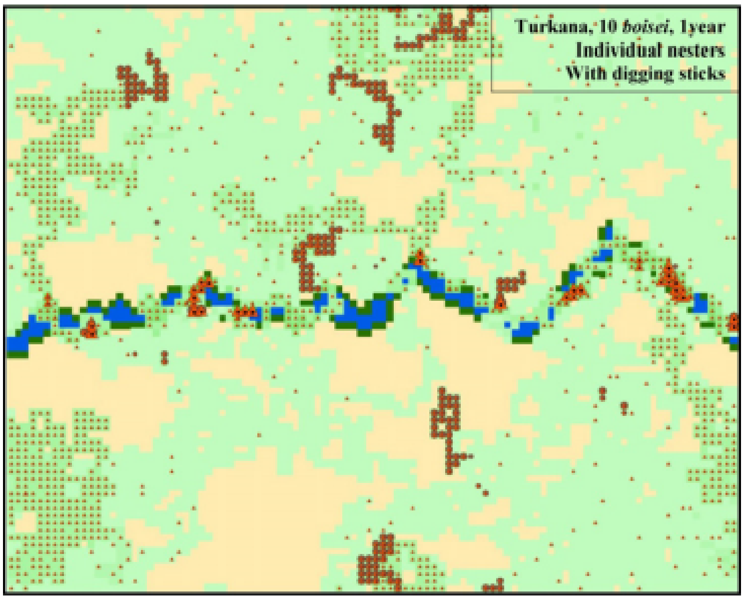


Figure 3.5: One of the habitats scenarios where they use digging with sticks [GLS10]

Last work we are including in this chapter is focusing on Neanderthals during the late Pleistocene in western Europe [Sch19]. This excellent study has been done in collaboration with archaeology professors which just goes to show how universally accepted and well regarded are simulations.

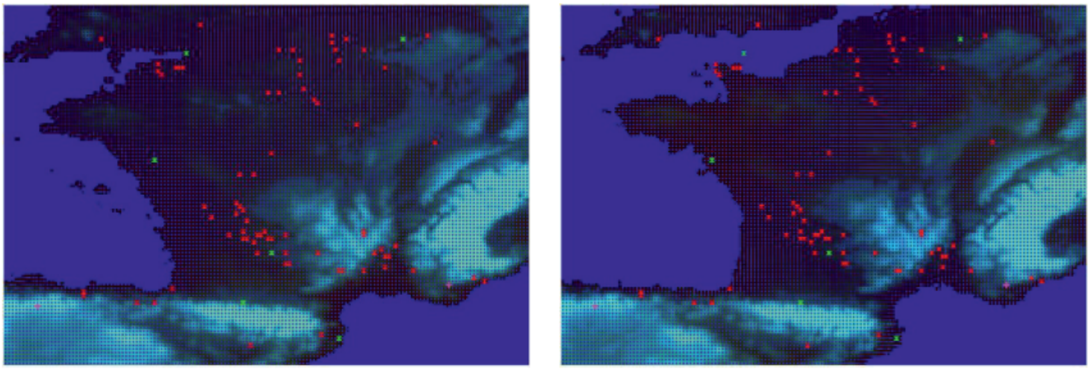
It is structured similarly to this work in multiple sections: construction of the land, creation of the actors, preparing and implementing the system, and experimenting. This agent-based simulation model is developed in a platform HomininSpace created by the author.

The work is spatially explicit as can be seen in figure 3.6. It adapts to topography changes considering the sea and ice level changes from the glaciation. The modelling is very important since it is intended to maximize validity by choosing data with the least bias and most impact.

The purpose of the experiments is to implement varying attributes that use the search space to detect the simulations with similar characteristics as the archaeological data. It is done this way to avoid using data that relies on ethnographic or paleoanthropological sources because they can always be considered false (in some measure).

Scenarios of the experiments are focused to identify the presence and absence of the neanderthals geographically. This model uses genetic algorithms to adapt and specialize in population dispersal prediction.

Complex visualization methods are used for debugging and general reconstruction (as we can see in figure 3.6).



131,000 years ago (-5.6°) 124,800 years ago (2.19°)

Figure 3.6: Visualisation of the habitat's topography in selected moments [Sch19]

Chapter 4

Methodology

This chapter will explain the development steps in detail. It will begin with the motivation for the chosen stages and continue by presenting each of them individually. First, there are the stages of conceptualization of the system: design and architecture. Then there is the modelling phase, arguably the most important in this work, that will present each layer of the simulations. On the way, there will also be subsections explaining decisions and sacrifices made.

4.1 Development stages

The development process receives as input the initial idea, plan and the researched theory in the field. The purpose is to use the knowledge acquired to materialize the plan in the best way.

It makes the most sense to split the work into development stages based on the complexity of the simulation as discussed in 4.5. We gradually construct a system that can be constantly tested and validated. The stages (both in modelling and experimentation) begin with an empty architecture and then add the components: environment, resources, and human agents.

For the actual implementation there were also some layers: the design flow, the architecture, the modelling and data gathering, and the validation.

Additional libraries or plugins that simulate agent-based systems were considered but the benefit they bring do not justify the disadvantages. Some important concepts that we need to be able to integrate are spatially explicit content, custom validations and dynamics, and ease to integrate new components that require intelligence, analysis or visualizations.

4.2 Experimental design

The experiments are designed to offer ease of **repeatability** that is needed considering the lack of deterministic results. Many components of simulations contain a factor of chance. This happens especially during the generation of the agents or the environment. This is beneficial because it extends the searching space of a viable solution to the problem (or the search space for correlations and certain emergent patterns). Repetition is also needed to improve the validation of the system against real-world data.

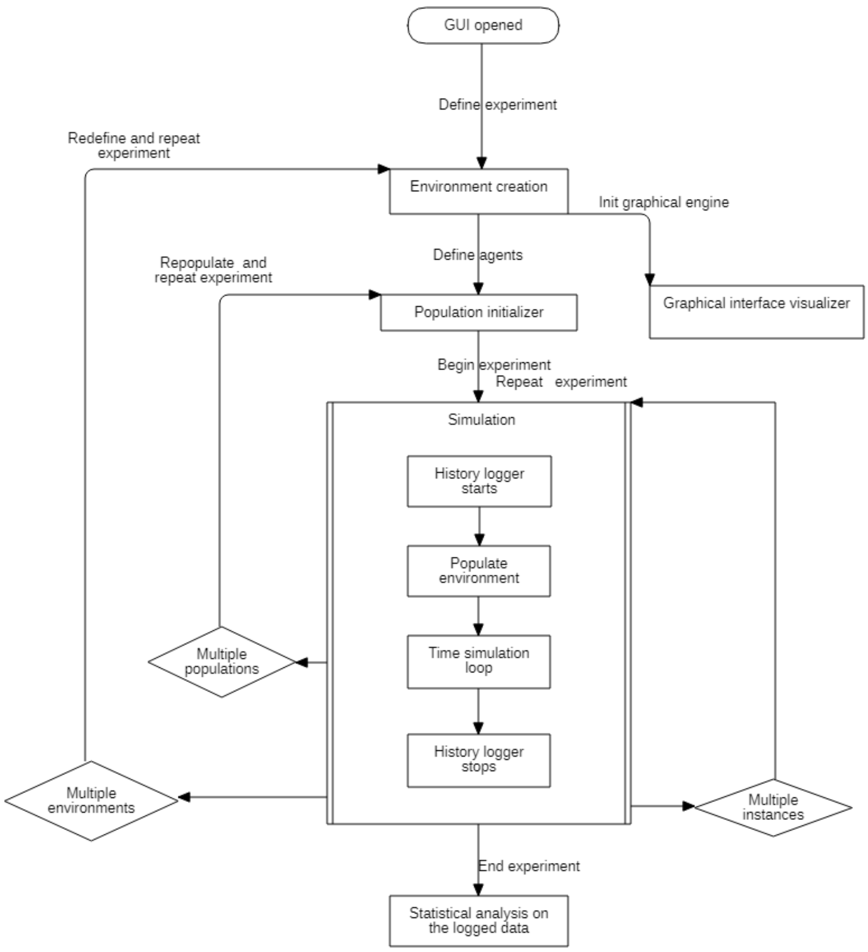


Figure 4.1: Experiment flow

We have a representation at Figure 4.1 that shows the experimentation flow. The organic way to divide an experiment is: **generation** and **simulation**.

The first phase will need to be repeated with different parameters (if there are no internal random mechanics) in every experiment. First, the **environment** will be generated, attributes will be populated with data from the database, and then the agent **populations** (both intelligent and non-intelligent) will be created strategically. To repeat a test we can choose to use the same environment and/or populations. For ideal results, they should all be reinitialized even if computational costs rise.

The actual simulation will have multiple components running independently but the overall concept can be simplified to: an initialization of adjacent tools (for analysis and visualization purposes), the integration of all populations and the environment, an iterative process through time, an end to all adjacent tools (that will analyse and save under different formats the logged data, statistics and visuals).

4.3 Architecture

The whole purpose of a custom architecture is to obtain a few desired advantages. Some of the most important reasons were: ease of integration of any custom visualization, scalability of the system (using different simulation resolutions and generated parameters), ease of data importing and attribute distribution generations. Although this is not the most optimal choice (as discussed in one of the following Section 4.3.3) it will facilitate the implementation of more complex simulations and maintain a logical flow in development.

The Agent is one of the elements around which everything revolves in social simulations. Naturally, the architecture should be constructed to facilitate dependencies on agents and resolve communication issues between agents and other important data. The agents were classified by their ability to make decisions based on an internal cognition as:

- **Environment Agent** (passive and only altered by dynamics or other agent's actions)
- **Developing Agent** (they learned to act upon the environment and other agents depending on circumstances)

As we can see in Figure 4.2 the architecture begins with the **Experimenter** that is responsible for running one unique experiment. Its job is to load all the parameters and initialize everything. The **Scheduler** is called to deal with actually running the simulations, instantiating the environment and the population, then simulating the flow of time and interpreting data for analysis and visualization (also partly processed by this class. The **Environment** and the **Population** are repository classes for the corresponding agents, they run basic functions during each iteration: triggering dynamics and cognition for each agent, assigning values to all attributes either from a file or a predefined distribution, collecting data and keep the environment updated. The **Agent** is the interface with basic attribute configurations that the other two agent classes inherit. Any **Agent** has dynamics that will be triggered periodically but a **Developing Agent** also has cognitions (which trigger actions), memories

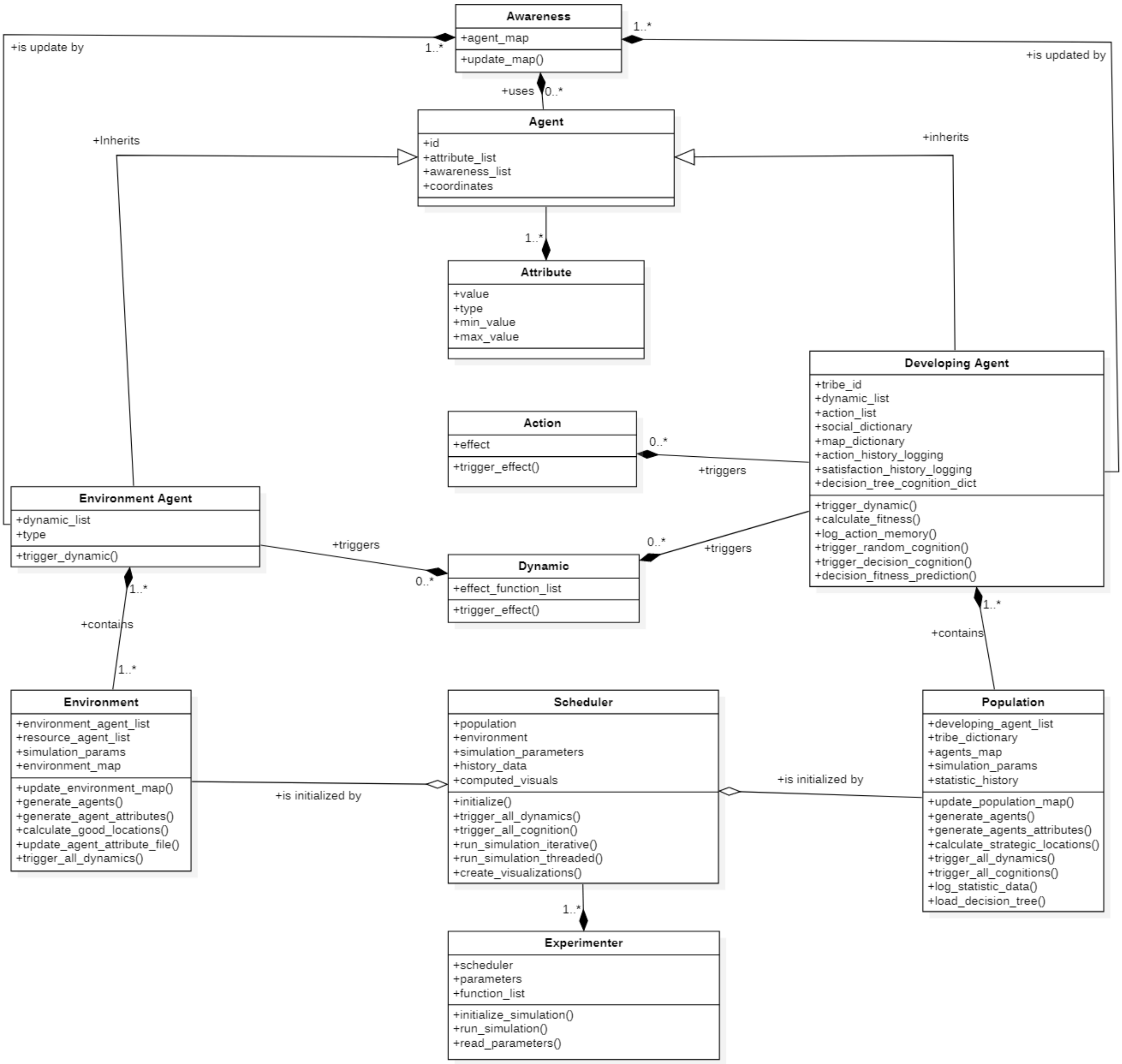


Figure 4.2: Project architecture

(of relations with other agents and of different places), and other auxiliary functions (mainly used for data logging, dataset construction). **Actions** and **Dynamics** are the classes that encompass a list of effects. The effects are custom functions that are triggered whenever something major happened in the environment. **Attributes** are important bricks that give values to all agents. These are modelled for each agent category and have different values for each of them. Some of them have static values and others have dynamism. **Awareness** is an important concept in communication between all agents. They are updated by the environment and the population regularly (before each new iteration) and then accessed by actions and dynamics (to interrogate and update).

4.3.1 Agent design

Different types of agents will have different requirements for functioning. The environment agent will be one of two types: geographical or resource. A geographical type agent will represent a square area of space with different attributes (temperature, relief, altitude, etc) and resources (food supplies). The population agent will represent an organism with superior cognitive functions (humans; and potentially animals in the future).

All agents have in common a few things. They have attributes (e.g: environment day temperature, plant caloric potential, human age) assigned in the first phase of generation. Most values are updated over the simulations, although some of them are static (human character predisposition). Another common feature is location: all of them have physical manifestation and this will be recorded as bi-dimensional matrix coordinates. All agents also have assigned awareness that represents the input to their affiliated dynamics and cognitions. They may differ in size from one attribute to another (as humans can see farther than they can smell) and they need constant updating (not of values but of pointers if there are spatial changes of any other agents).

4.3.2 Time and simultaneity

Discussion regarding causality and conflicts are somewhat controversial. In physics, for atomic events, the special theory of relativity of Einstein [Mil98] can comfortably motivate the lack of simultaneity. In philosophy, the validity of time has been contested multiple times, most notably by McTaggart's argument [McD20] where he judges the universe as time-less through a relatively well-constructed argument (but the matter is subjective and has been contested). Due to these reasons, it would not be too far fetched to not simulate the universe concurrently and instead create a causality based system with instantaneous effects and no conflicts. However, in real-life scenarios, temporal relationships of events are very relevant. This happens because events (including the ones we are going to simulate) are not instantaneous. There will be relations of overlapping between events and that is why concurrency is important.

Simulations would normally need to properly reflect events and the conflicts that are created when temporal overlapping occurs. However, this is valid only when events are simulated with small time periods. The time of one step in our simulation will be a minimum of **8 days** which means it is not going to be necessary for the events to be pinpointed in time (instead they are a statistical representation of reality that already takes into consideration conflicts and causality). During the time step, small events are going to be considered as **cumulative** (how much food was gathered by one human in 8 days, how much food was shared between two humans in 8 days, how many fruits were ripe in 8 days). Similar to instantaneous events our simulated event will be time-less and will not require conflict management.

As previously mentioned the time step will be 8 days (due to data limitations in climatic attributes) and a simulation will go on for at least one human generation life (about a maximum of 50 years).

4.3.3 Generalization vs Performance

One of the most important features of this project are:

- the ability to scale the simulation at different **resolutions**;
- the option to **design and integrate any new agents** with various data sources easily;
- the ability to replace any logical functions, or configurations easily.

For this development to happen with at most fluidity we need a simple and generalized architecture that can allow it. If this means more classes, more **modularization**, and more configuration options, then it will also pose a great threat to performance. This is a trade-off that has been considered and accepted.

4.4 Modelling and data gathering

This is, arguably, the most important step in the development of the project. Properly modelling and designing the agents and the environment is mandatory for ensuring a minimum error and accurate representation of a real scenario. We need to use as few attributes as possible that represent the environment as well as possible.

Micro-level attributes (plant population density, quantity of calories needed for a human, etc) will be generated based on the current knowledge in the literature. Rules, dynamics, and actions (based on agent cognitions: moving, eating, etc) will bring alteration to the environment. They should be designed based on knowledge or common sense. The Macro-level emergent properties/behavior (population density, average life-span, etc) should be in accordance with current literature. Data from social studies (especially psychological and anthropological) will be used to map and model the human behavior and attributes, and also for the environment and all the cycles present there. We seek to simplify the system whilst losing the least possible amount of information to preserve the integrity of the conclusions.

It feels natural and optimal that the modelling process goes hand in hand with the data gathering process. If data or statistics are not available for an attribute/dynamics/action, then we cannot include them in the model. If it does not make sense to include an attribute/dynamic/action into a model, then data should not be searched for it. Their relationship is inter-dependable.

In the following subsections we will go into more depth for each component beginning with a general picture, the environment and building up towards climate, resources and humans.

4.4.1 Context

We are now going to define the essential circumstances of this simulation. The settings for this proposed equilibrium are sometimes in the **Late Pleistocene Age** in **Europe** during the last great migration of the early humans: **Homo S. Sapiens** (100’000 – 70’000 years ago). The humans depicted must offer simplicity (few occupations, actions, and little culture) to be able to accurately create the equilibrium in computationally-viable time. Mass cultural and artistic development happened especially after the agricultural revolution (11’000 – 9’000 years ago) which means the most optimal period will be that of **hunter-gatherers**. We also prefer a steady and peaceful environment that will lack inter-species competition. This will happen preferably after the extinction of the other hominins (more precisely H. Heidelbergensis) like: neanderthalensis (until 50’000 years ago), denisovensis (until 50’000 years ago), florensiensis (until 100’000 years ago), rhodesiensis (until 120’000 years ago), and heidelbergensis s.s. (until 200’000 years ago). Based on anthropological literature this will mean **40’000 – 11’000 years ago ??**.

We are looking at Homo S. Sapiens between 40’000 and 11’000 years ago in the Late Pleistocene Age of Europe simulated with a step of 8 days for about 50 years.

4.4.2 Environment

The environment will be constructed as a matrix where each cell will represent a physical cube space of reality. We are defining the European continent as the space between -20 and 60 degrees of longitude, and between 30 and 70 degrees of latitude. The representation as a matrix describes it as being a rectangle but in reality is closer to a trapezoid with rounded edges (the edge dimensions are: 4450km, 4450km, 7530km, and 2830km). The base towards the equator will be more than double in size. To avoid extra calculations (based on area) for every attribute of the system in each particular cell we are going to approximate the sizes of all cells to the average based on resolution. The entire area is 6'171'438'737.53 ha or 24974926.49 km^2. By dividing by the number of cells we can obtain the average area of a cell:

- resolution 80x160: 1951 km^2 in one cell
- resolution 160x320: 487.8 km^2 in one cell
- resolution 320x640: 121.9 km^2 in one cell
- resolution 480x960: 54.2 km^2 in one cell
- resolution 800x1600: 19.5 km^2 in one cell

The previous resolutions are going to be essential for the following components because many of the data will have to be re-scaled each iteration (or pre-processed in multiple resolutions from the very beginning)

4.4.3 Geography

For this component we need attributes about the topography, relief, and general environment but it is required ot be relevant. Data from our selected time period only exists as projection and all viable resolutions are modern (at least 1980).

We decided to use modern datasets and alter the geography terrain only slightly. The main attributes are about the terrain topography on which we corrected deforested areas and added water bodies. Most maps extracted were in a cylindrical projection (which we also adopted) but some exceptions needed further editing (for which we used image warping editing). Data were combined from the following datasets:

- **Terrain type** obtained from global map dataset MCD12C1 v006 (MODIS dataset) [SMF18]. We combined the 16 available classes (and removed the city class) to result in 4 classes: field, forest, shrub-land, ocean/sea. The global map was cropped to our selected coordinates.

- **Prehistoric forest reconstruction** map from 1000 BC (oldest available data in a research) [KKZ] was used to correct deforestation that occurred in the last 3000 years. The map was converted through puppet warping operations from what seems to be Azimuthal map projection to cylindrical, scaled, and cropped. The continuous color map was converted through discretization to two values: old forest and young forest.

- **Hydrography** map was used from HydroRivers v1.0 [LG13]. The map was converted from what seems to be a pseudocylindrical Eckert map projection to cylindrical, cropped, and scaled. Through discretization with 3 classes, we obtained: shallow river, and deep river.

As it is also represented visually in Figure 4.3, the three different map datasets were combined into one. Upon the original terrain, map changes were made. All deforested areas that currently were shrubland or field are turned to forest. All terrain is turned into shallow river, deep river or sea/ocean if there are overlapping. Any unassigned terrain is converted to the predominant category in the vicinity (with a 3x3 neighbourhood) to cover all areas left empty by what was originally the city class.

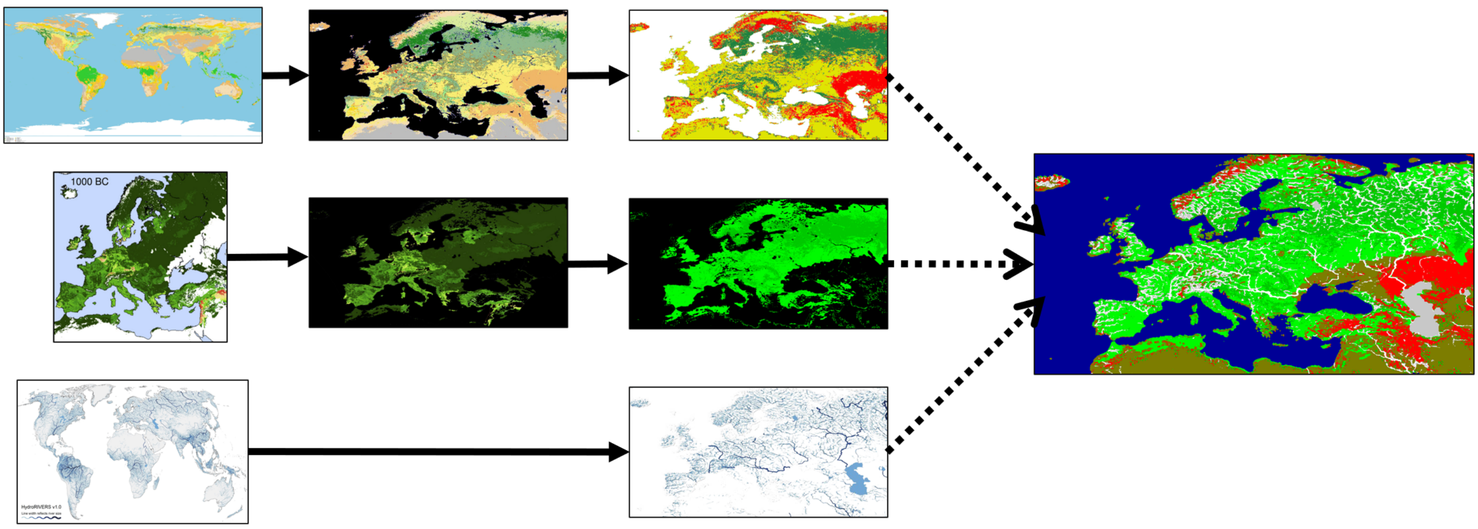


Figure 4.3: Data processing flow to obtain the environment geography of 6 classes: 1 — shrub, 2 — forest, 3 — field, 4 — deep water, 5 — shallow water, 6 — ocean seas

It is known that the topography in Europe was slightly different in our selected time frame due to the last glaciation. Water levels were low enough that the British islands were connected to the continent, the Scandinavian peninsula was mostly underwater and many elevations were different due to ice formations. We knowingly did not take that into consideration whilst building the environment due to the lack of viable public datasets.

One more important attribute that will be included in the system is altitude. We have integrated the data sets from World digital elevation model (ETOPO5) [Pet]. It was extracted from a ".tif" file and converted into custom file. The elevation attribute will be useful because certain terrain types and temperature ranges will appear at specific elevations. The data will be imported as an attribute of every environment agent. We can see a visual representation in figure 4.4.

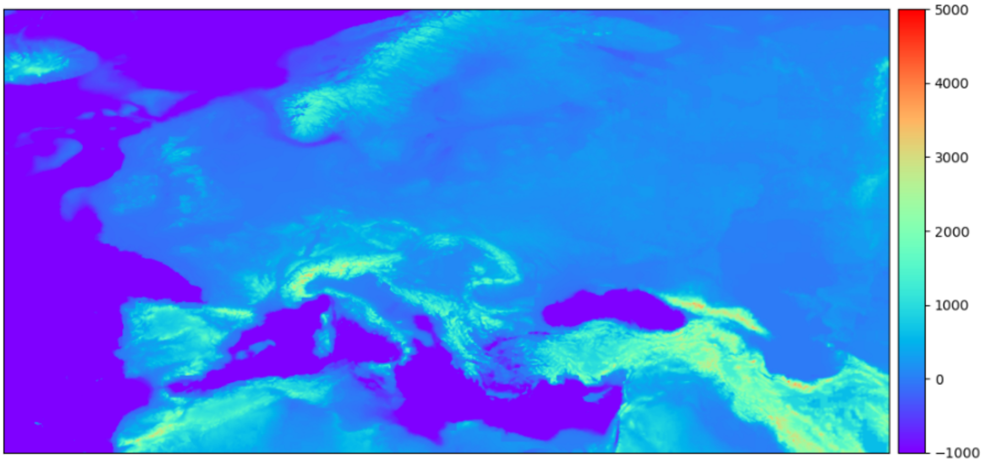


Figure 4.4: Visualization of the elevation attribute using a color map

4.4.4 Climate

Whilst the terrain and geography attributes of the environment are not suffering dramatic change over time (other than de-glaciation and tectonic plate movement) but the atmosphere does. To obtain an accurate model many attributes should be considered in the climate.

Normally, **simulating** the climate would not make sense because **emulation** would reduce general error in the system but for our selected time there are no datasets and, therefore, the adaptation of climate simulation would make sense. We did consider initially simulating completely the atmospheric processes and in consequence, we initiated viability tests. Two models were considered: the WRF model [SKD+08] and a custom variation from the 3-cell model (also called the Hadley cell model or the Global circulation model [HO07]).

The WRF model is very complex and even with optimizations (of parameters and formulas) it will still be bounded by the upper limit of a time frame. It is not computationally viable for large scale simulation since it would take years.

We did experiment with a simplified version of the 3-cel model variation in order to approximate o possible climate simulation. The climate model we designed included a relatively wide range of attributes (as can be seen in Table 4.1). The dynamics proposed are designed to cover as many natural phenomena with as few physics formulas as possible. They are inspired by reports and curriculum offered by the US National Weather Service [gov] and the University of Colorado [Joh].

Surface Attr.	Low Air Attr.	Stratosphere Attr.	Dynamics
Temperature Absorption Heat	Air temperature	Air temperature	Heat absorption Heat Convection
Humidity	Humidity	Cloud mass	Water Accumulation Water evaporation Precipitations
Height	Pressure	Pressure	Pressure creation
Material	Wind vector	Wind vector	Air movement Wind direction

Table 4.1: Experimental climate attribute

The experiment simulated the environment with the **Temperature** surface attribute and the **heat convection** dynamic. One epoch of the environment on a low resolution 250’000 agents took between 0.005-0.02 seconds (depending on computational unit) as described in 5.1.

We want to know if it feasible to also simulate the climate using this model. We multiply the computing time by the number of needed dynamics (about a total of 11 considering sub-functions of the dynamics in table 4.1), by the number of epochs needed in a year (especially for precipitation and wind we need updates on a frequency between 1-6 hours minimum which means 8760 - 1460 epoch) and by the minimum number of years needed (which is 50) and obtain 24’000 - 96’000 seconds per simulation. This means between 6 to 27 hours for the climate component alone. After the addition of resources and humans, the time needed for a singular simulation would be counted in days. Considering that we also need multiple simulations for validation purposes it is completely not feasible to attempt a simulation of the climate.

A simple solution to include climate data is by emulation. All climate dynamics and attributes boil down to a few that are experienced by humans: temperature, pressure, humidity, wind, and precipitation. Humidity and pressure both affect the human body in many ways but we do not simulate that level of detail. Wind and precipitation play a bigger role in the environment than for humans. We model society before the invention of many revolutionary discoveries and our time-step of 8 days makes all other attributes redundant. Of all the attributes only temperature is mandatory to create relevant climate cycles.

We included dataset MOD11C2 v061 from MODIS [WHH15] containing high density temperature infrared satellite data. We choose the 8-day time step with 0.05 arc degrees and extract two attributes: **”LST_Day_CMG”** , **”LST_Night_CMG”** .

After processing the "*.hdf" files (crop by coordinates using rioxarray [rio]) resulted are of 800x1600 resolution as can be seen in figure 4.5.

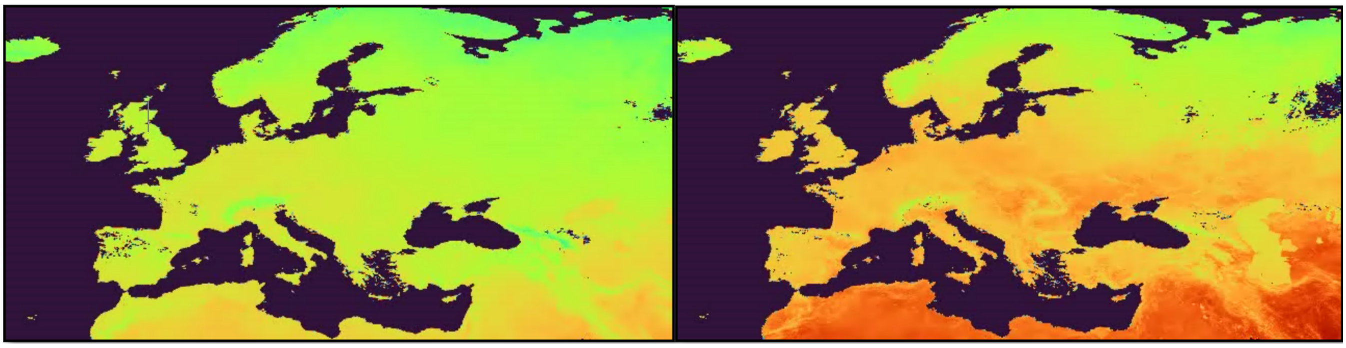


Figure 4.5: Geographic representation of the day and night temperature attributes (the only climate attributes used

4.4.5 Resources

Some of the resources originally considered were: plant produce (fruits/vegetables), water, animal meat (after hunting), and other tool resources (rocks, minerals, wood, etc). Water is already part of the geography of lakes and rivers. It is known that sapiens used other alternatives to find water as well. However, due to a lack of proper datasets of alternative water sources, it could not be done. Complex tool resources (regular tools are already being considered during the scavenge action as described in chapter 4.4.6) do not fit the time period of our simulation. Animal meat was discarded due to the complexity of the simulation for this stage and was replaced with an all year round alternative plant food source.

Plants have been modelled as resource agents (the only one at this point) contained by Environment as well and represented as "EnvironmentAgent". To best encompass a general population we design multiple types of plants and each will have a range of values for each attribute. These ranges will be used in the generation phase to create the agent's attributes. The types will be:

1. Fruits and nuts (mid - high-calorie plants that live for multiple years and are ripe annually)

2. Roots and vegetables (low - mid-calorie plants that live for less than a year and die until winter but spread faster)

3. Herbs and spices (zero - low-calorie plants that usually live only one year and spread faster)

Each of the mentioned plant types will have multiple particularities (regarding attributes) but all of them will suffer the same dynamics. The dynamics are:

population fluctuation and **food production**. We want the population of plants to change in each iteration because different plants have different life spans and different spreading abilities. Trees grow and are riping for tens of years while crops and vegetables live for a couple of months. It is essential for food to be produced by the plants as that is the only resource scavenged and consumed by humans. The attributes available to a plant will be either static or changing but all will contribute to the dynamics.

The dynamic attributes of plants are:

- **Population** of plant ([100,1000] * avg. area of cell(
- **Current food** in calories

The static attributes of plants are:

- **Color** (1,2,3)
- **Energy** ((0,700] – calories per 100g)
- **Heal** ([0,3] – health points)
- **Quantity** ([1,100] – multiple of 100g)
- **Probability** for each **terrain** ([0,100], . . . , [0,100])
- **Probability** preferred **temperature** for ripping ([-30,70])
- **Probability** referred **month** for ripping ([1,12])
- **Spread ability** ([0,100])
- **Sensibility** ([0,100])

All attributes are generated at the beginning of the simulation for all resource agents including plants. Each plant type will have different possible random values that are modelled to reflect their class. Examples: quantity of fruits (type 1) will be 5-10 kg and of vegetables will be 0.1-1 kg; the population of fruits will be 100-500 specimens on a km^2 and of vegetables will be 1000-3000 specimens on a km^2.

The first dynamic of population fluctuation is an iteration based formula:

$$population\_change = \frac{(spread - death) * population}{factor}$$

, where the spread and death are a percent of the population that changes (in interval [0,1]) and are calculated by the weighted formulas of attributes:

$$death = \frac{(1 - fact\_relief) * w + (1 - fact\_temp) * x + fact\_death * y + (1 - fact\_month) * z}{20}$$

$$spread = \frac{fact\_relief * w + fact\_temp * x + fact\_spread * y + fact\_month * z}{20}$$

, the calibrated weights are:

$$w = 3; x = 2; y = 7; z = 8;$$

There are two factors that are calculated using a linear distribution model having as input the interval values of the attribute (fact_temp and fact_month). The rest of the factors are simply scaled to [0,1] interval.

Locations on the map for plants agents are generated **randomly** with a single constraint: to be in the neighbourhood of a water source. They also have a higher probability to be generated in the preferred terrain type (described in section 4.4.3) specified by the attribute "Terrain Probability".

4.4.6 The Humans

The main **population size** of humans used in our experiments will be of 1'000 agents. This is limited mostly computationally but even if resources would be available it should not be raised over a maximum of 100'000 for our time period as it is discussed in section 2.4. Another interesting factor is population density (also available in section 2.4) or tribe count which will vary between 30-50 (even though there are some notable exceptions).

Locations on the map for agents are generated **randomly** with the single constraint to be in the neighbourhood of a water source.

Regarding **emotional modelling** current studies in the field have created very complex psychological computational models used in a plethora of domains and structurally synthesized by [KC16] [MGP]. At the beginning of our project, a complex emotional system was designed for the cognition component. However, the complexity was not justified by the potential output and usage relative to the input complexity of the learning component (which is discussed below). Another crucial argument is the fact that our time step (of 8 days) will not allow for enough input to the model. If a live learning component would be integrated, then an emotional input would be useful for learning and adapting. The agents will still have some emotional input in the relationship developing process between agents and in the overall satisfaction of the agents (also the fitness function used in training) as can be seen in the attribute list.

Another important reasoning regarded the **energy cost** of different actions. It was decided that applying cost to every action would not make sense due to multiple reasons. First of all, energy is consumed automatically every iteration regardless of a human's actions (which appears inaccurate because different activities require different amounts of energy). Agents might learn to exploit such a trade but since there is a limit of actions every epoch (which already works as a time cost) that could not happen. Secondly, our time step of 8 days (which is only a minimum) is too long for such costs to matter.

The **life length** of an individual will be an **entirely emergent phenomenon** through the dynamics and actions of multiple agents in an ever-changing environment. That is the reason for which validation based on mortality is empowered. We describe the source for the chosen mortality to which we compare at 2.4

There are a wide range of **attributes** mainly focused around human needs (which are inspired by Maslow's pyramid of needs [Blo11]):

- **Age** ((0,100))
- **Sex** (0,1 corresponds to male/female)
- **Health** ([0,100] where 0 is worse and 100 is best)
- **Fitness** ([0,100] where 0 is worse and 100 is best)
- **Hunger need** ([0,100] where 0 is not satisfied and 100 is completely satisfied)
- **Thirst need** ([0,100] where 0 is not satisfied and 100 is completely satisfied)
- **Reproductive need** ([0,100] where 0 is not satisfied and 100 is completely satisfied)
- **Safety need** ([0,100] where 0 is not satisfied and 100 is completely satisfied)
- **Emotional State** ([0,100] where 0 is worse and 100 is best)
- **Plant find ability** ([0,100] where 0 is worse and 100 is best)
- **Social interaction ability** ([0,100] where 0 is worse and 100 is best)
- **Food Inventory** (a list of counters for each resource)
- **Character** (a list of percents representing preferences for each action)
- **Offspring** (whether or not it has and what age)

Each attribute is generated at the beginning of the experiment by the "Population" with fixed chance and intervals of value. All needed attributes start with random values over 50 which is not founded in any way and it will take a while to reach the equilibrium or natural pace. All attributes are solely designed to play a role in the dynamics. 'Character' attribute is static and all other attributes are dynamic and may change during the experiment.

The dynamics represent the natural laws of the world and their effect on humans dealing with the process of degradation in all states. This is the drive to action for all agents. The dynamics are: **aging** and **need evolution**. They are happening before each epoch.

The aging formula is a simple constant subtraction of exponentially growing values that will basically cap the lifetime of all humans at 50-70. Of course, this is a very insignificant incentive for the agents as there are means of eating healing plants to gain health and also countless ways in which one might take damage:

$$health = current - exponential\_incentive/experiment\_factor$$

where the exponential_incentive is a fixed value based on age and experiment_factor is calculated based on epoch length, a number of days in a year, and incentive.

All other needs (thirst, hunger, reproductive) will be calculated in almost identical ways but with different incentives.

Thirst will decrease the need to 0 in 2-3 days (but minimum an epoch) and then health is drastically decreased. The food they consume does contain 74-96% water and this will also contribute to thirst need. If the human is not in the vicinity of a freshwater source it will be considered that thirst need is not satisfied at all. If water sources are nearby based on how close they are a coefficient is calculated. Thirst is designed to kill very fast if humans are not close to the water and also not eating well.

Hunger will decrease to a minimum in about 20-30 days and then draw from health directly. Food can be acquired in multiple manners but can be consumed only through eating and that is when hunger is replenished.

Reproductive need will decrease to a minimum in 3-5 years and can potentially stay there forever. Low reproductive need will not decrease health and kill agents. It can only be cured through copulation.

Emotional state also is altered using an incentive based on the weighted average of all other needs.

Safety need and **Offspring** does not bring contribution to this experiment.

In order to satisfy their never-ending craving modelled using constantly decreasing needs agents must act. They have at their disposal a range of actions and memories that can be recalled using their cognition. The actions are:

- Move
- Eat
- Scavenge
- Social Request

- Social response (automated reflex)

All actions will provide after execution a level of satisfaction. This is another incentive for the people to take action and not just decay. The satisfaction level is based on character. If an agent will have been born with an inclination towards scavenging or social behavior then it will receive more satisfaction for that action.

Moving can offer new opportunities by changing the neighbourhood. The target moving space is important in making this decision because basic attributes like terrain type and altitude are available to the agent.

Eating is a simple action that decreases the contents of the inventory and increases the hunger need satisfaction. The amount by which it is increased is calculated relative to a daily needed dose of calories of 2500 (which will be just a little higher if temperatures are higher) and increased if the agent is a male. The values will range between 1500 and 3000.

Scavenging for food will let the agent choose a type of plant and collect a quantity from the neighbourhood. The quantity is defined by the available space in the inventory and by the ″plant find ability″. Since it is limited by the inventory the maximum value will be important. It was chosen to be a maximum of 40′000 to 80′000 calories. It is modelled so that the agent can carry at most a few tens of kilos. It can stock up for the long run since it can hold food for about at most a month. The inventory is not necessary a bag or backpack but rather a place in which the agent stores food for himself.

Social request is a complex query to another agent that can be of two types: food sharing request or interaction request. If it is a food-sharing request then based on the quality of the relationship between the agents (based on the social dictionary), the hunger need, emotional status, available inventories, social similarities and a quantity will be selected to be shared. The bigger the quantity the better the interaction between the two and otherwise. If an interaction is called then depending on the relationship quality and similarities the outcome will be decided. The relationship can be neutral and then they will only slide slightly toward the direction of the last interaction. Slowly they will reach one of two extremes: positive or negative relationships. The negative ones will have a chance to result in a fight and then in a neutral state or copulation action (because it is proved that primates regulate hate inside a tribe through intercourse). The positive ones will only result in a chance of copulation. There is a higher chance of copulation between different sex (as suggested by studies both happen in many primates) and a higher chance of fighting between males. The similarities between attributes also contribute to both requests. This means that age and sex (if opposite then better) contribute to the intensity of the request.

4.5 Core implementation and iteration-based development

The project's development makes the most sense to be iterative. Complexity is overwhelming and could pose troubles in testing, validating, debugging or calibrating the system. We will first test the architecture's structure and functionality to ensure the foundation of the following experiments. We will do this by a rudimentary experiment to ensure that all components (empiric unit test) and interactions between them (empiric integration testing) are functioning as intended.

More experiments will follow this guarantee of basic functionality. They will be increasingly more complex and will integrate one by one the conceptual layers of the simulation as described in the previous section. We begin by emulating the environment, then proceed to add the resources (plants only), and the humans. At this point, we will have created a functioning simulation but the humans will not be final.

A very important characteristic of the human agents is that they are learning agents. Whilst they do not learn from experience (since our time frame is relatively limited) they must have an inherited knowledge (partly genetic and partly cultural). We will train the agents offline using data from the initial simulation (where agents make random choices) and then deploy them in one final experiment that is to be repeated for validation purposes.

4.6 Statistical validation and correlations

Currently in the field the most advanced work (as described in 3.3 and 3.4) encourage advanced validation processes based on explicit researches. This is important in order to approximate the quality and accuracy of the experiments because that will be the only manner of measuring confidence if predictions are also being made. When populations are being simulated the validation must be a relevant metric that does not measure an emulated process but rather an emergent phenomenon triggered by the dynamics of the experiments. These dynamics are generally the simulated components. We measure their validation in order to argue in favour or against the presumptions made in the simulations. The presumptions are represented (in our case) by: all implemented attributes and their generation distributions, all dynamics or action formulas and their support, and even the general architectural design of the experiment (implementation of time, chosen scales, chosen integration between agents and environment, etc). Considering how many presumptions exist in the creation of simulations it is essential to have relevant validations.

The focus of our simulation is on humans. We will validate the population evolution through mortality. As presented in 2.4 we will use the extracted data to compare with the experiment. We will use simulations with the same generation parameters and run multiple experiments for both scenarios. The first scenario will emulate the mortality through a death probability for all humans and nothing else. The second scenario is aiming to simulate the complete modelled experiment with all dynamics, actions and cognitions turned on. However, offspring will be turned off since that would imply a new level of validation and would require additional emulation information.

Chapter 5

Experiments

In this chapter we will go through the development and describe the experiments in each iteration (as they are described and justified in section 4.6). This experimentation phase represents only the implementation and visualization of the components and iteration described thoroughly in the previous chapter 4. For each iteration, we will have a section where we will **describe** the process, present some **visualizations** and analyses of performance, and in the end interpret the results.

All experiments have been developed in python 3.8 on an anaconda environment and took advantage of many processing libraries: NumPy, Pandas, Pickle, rasterio, rioxarray, earthpy, OpenCV, shutil, requests, matplotlib, sys, scipy, sklearn, math, time, threading, random.

All computations have been conducted on a personal computer with AMD Ryzen 5 3600 6 core and available 16 GB RAM. GPU processing was not used.

All experiments are computed at the minimum resolutions of 80x160. Training is computed at both 80x160 and 160x320

To visualize further information, validation, visualizations and analysis please visit: **https://github.com/RaduGalan1/AnthropologicalSocialSimulation**

5.1 Convection testing

5.1.1 Description

The first experiment has as sole purpose to ensure the architectural integrity and to motivate the modelling decision of climate (as described in section 4.4.4). The experiment is designed to be as simple as possible but at the same time, it needs to cover as many components (cognition is harder to test because we need more actions and parameters). We judged that the environment will contain agents with a unique attribute (temperature) and a unique dynamic (convection).

5.1.2 Visualization and data analysis

In figure 5.1 we can debug the results of the simulation. It represents the graphical representation of all attribute values on the 2D map of the environment based on a color map ("COLORMAP_HOT" from opencv [Bra00]). Lighter color means higher temperature and darker color means lower temperature.

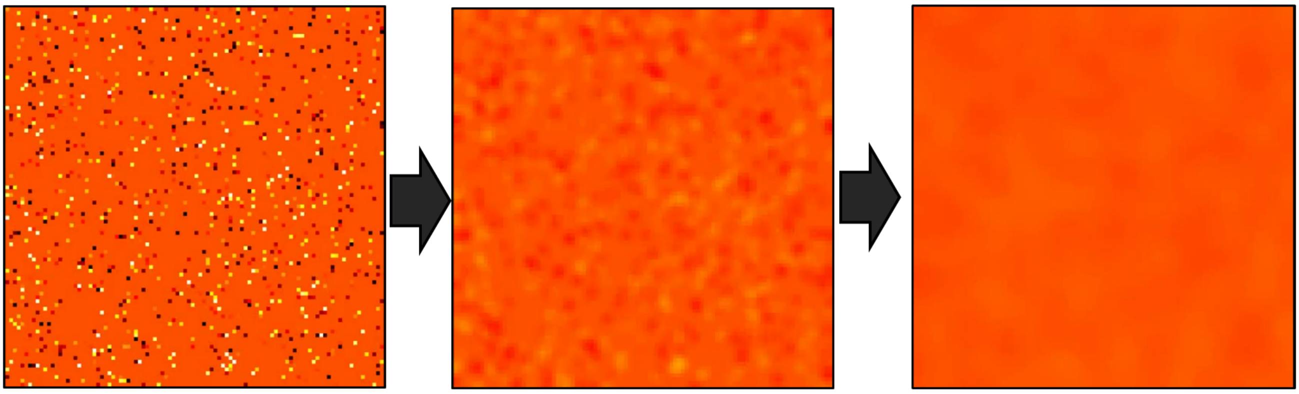


Figure 5.1: Three states of the convection experiment chronologically ordered

5.1.3 Results and Performance

The visualization proves that the process executes as intended. Tweaking the parameters to multiple values changes the results as expected and we can consider the experiment successful. Since the architecture is reliable enough we can pursue the following iterations safely.

One iteration/frame took an average of 0.005-0.02 seconds on our machine.

5.2 Environment simulation

5.2.1 Description

The very first phase of constructing the simulation is the environment. As described in the modelling phase we will have the following attributes: day temperature, night temperature, elevation, and terrain type. This step is exclusively an emulation but it is still important that it works.

5.2.2 Visualization and data analysis

The visualization is composed of 4 different videos computed at the end of the experiment. We can see in figure 5.2 how all of them look. The two on the right are static attributes and the two on the left are dynamic.

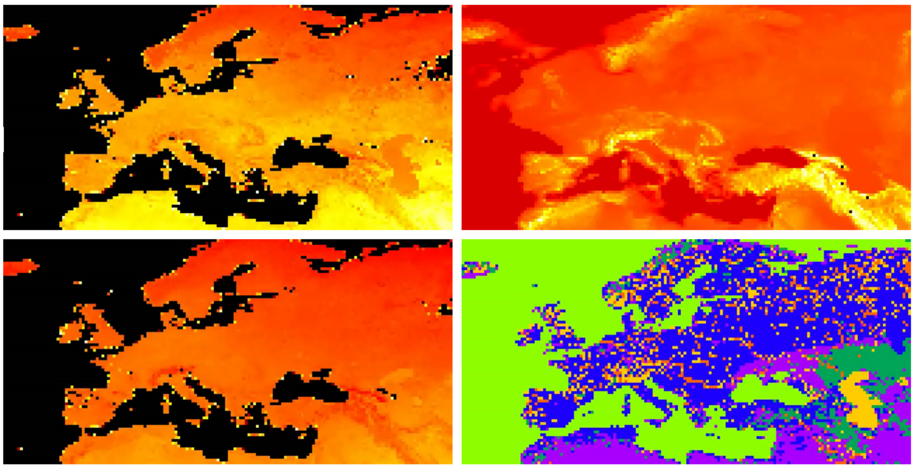


Figure 5.2: Snapshot of the visualization of attributes emulated for the environment

5.2.3 Results and Performance

No results are yet notable but the system works. Performance is very similar to the previous experiment at about 0.005 seconds an epoch. Most time is spent during the data reading. Optimization has already improved the loading time tenfold by separating dynamic attribute's origin files into thousands of small time-encoded files.

5.2.4 Validation and Interpretation

An important step in validating this was debugging the correctness of the yearly cycle created by the temperature attribute. This is the main attribute (also calendar date) that will contribute to the creation of cycles and the desired equilibrium.

5.3 Environment with resource agents

5.3.1 Description

This step brings a significant increase in complexity as these next agents will also be computing more complex dynamics. The only resource agents for this experiment are **Plants**. The resource agents are completely generated and maintained by the "Environment" component. What we need to obtain is a reliable environment that offers resources for human agents in realistic circumstances. The most important emergent phenomenon that we want to create are the seasonal changes (like plants having ripe).

5.3.2 Visualization and data analysis

The first visualization developed to debug the evolution of plants involves a collage of all the attributes representations (with different color maps). The collage is automatically created as it only receives a list of attributes and then generates collage images of different dimensions with different borders. This visualization was used in debugging during both the generation and the simulation phase. We can distinguish from the static attributes that there are variations which correspond to the different colors of plants.

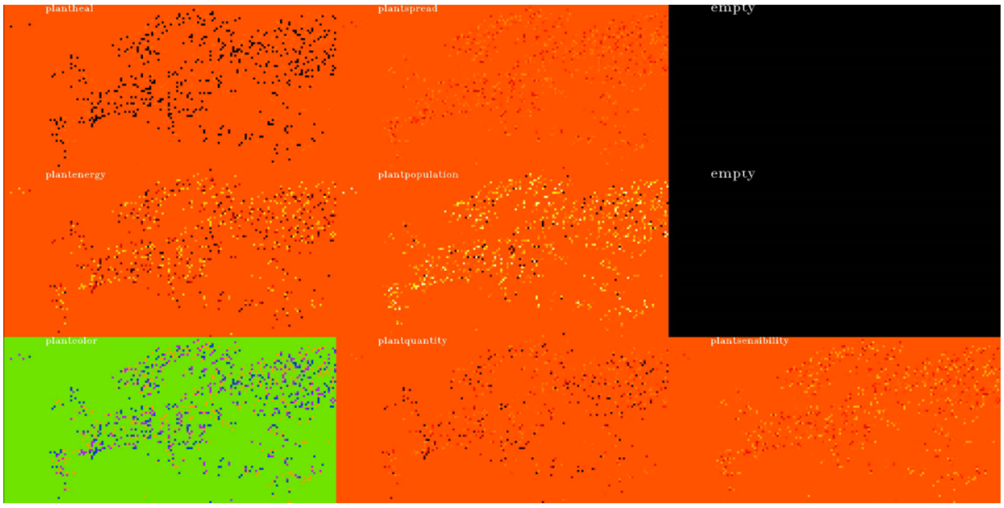


Figure 5.3: Visualization of all plant attributes using geographic color maps

The terrain visualization (labeled material) is updated to fit the standard colors for each class. Elevation and temperature ('surftempnight' and 'surftempday') are also adjusted to be more pleasant. Each plant attribute representation is updated to a combined image (also called overlapped) between the previous attribute color map (as seen in figure 5.3) and the terrain representation. All these changes can be seen in the general collage of the simulation (figure 5.4).

5.3.3 Results and Performance

One iteration now takes between 0.2 - 0.5 seconds. Optimizations have been attempted using threads but due to python's GIL component, it cannot achieve bet-

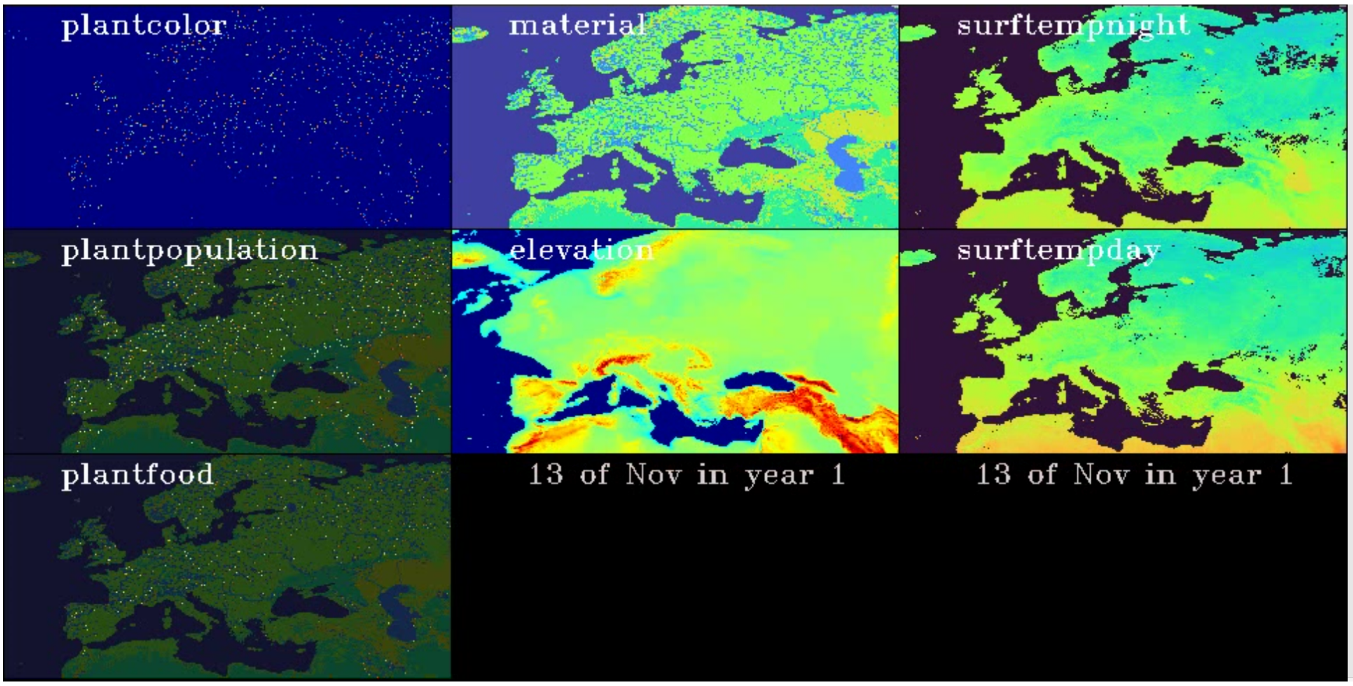


Figure 5.4: Snapshot of visualization during plant simulation

ter performance than the iterative version using our modelling (one thread for one agent).

5.3.4 Validation and Interpretation

Only two attributes are dynamically changing (as described during modelling): the plant population and the plant quantity. We need them to be changing properly with the seasonal cycles.

Plant population must decrease for all plants of type 2 and 3 (roots and vegetables, herbs and spices) during the cold months and increase back again. The population should achieve however a constant cycle (since we do not add any other natural forces that could allow plant species to thrive over other specie's extinction).

Food quantity must vary as well by decreasing during the cold months and increasing during the warm months.

All these requirements can be validated (as can also be seen in the snapshot at figure 5.5).

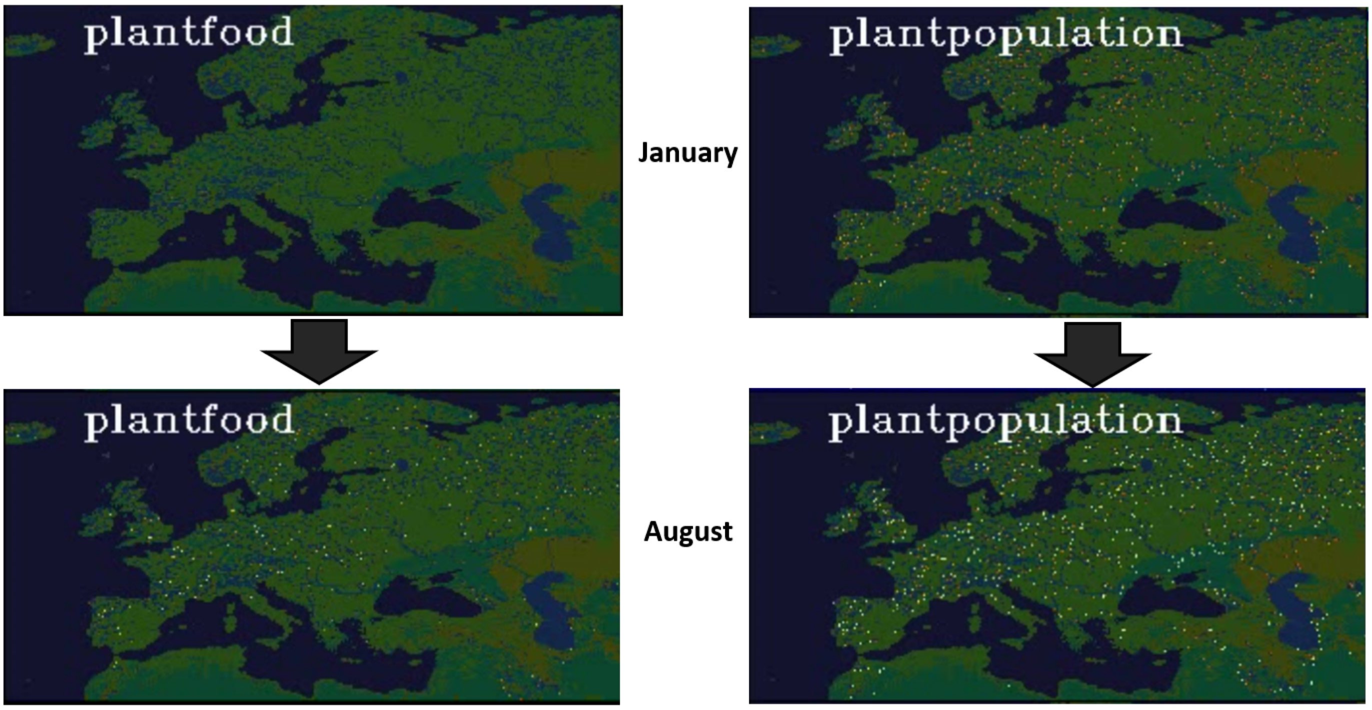


Figure 5.5: Visual proof of plant cycles

5.4 Environment with resources and non-intelligent human agents

5.4.1 Description

The first iteration with human agents integrated has as sole purpose to gather the data require in the training phase of intelligent agents. In this phase, agents will randomly act in the environment. They have a random chance to make an action which will be executed with random parameters. A specific functionality inside the ”Population” and ”Developing Agent” components deals with logging the information from actions.

This **dataset** is constructed with data gathered right before the execution of the action. Those represent the circumstances of the current time for that agent. The attributes used are:

- **Personal State**: human_count, emotional_state, human_reproductive_need, human_thirst_need, human_hunger_need, human_health, human_age, human_sex
- **Resources in neighbourhood**: plant_color
- **Target environment**: target_temp, elevation, target_relied_shrub, target_relied_forest, target_relied_field, target_relied_deepwater, target_relied_shallowwater, target_relied_ocean
- **Target human**: called_relation, called_sex, called_age, called_inventory

5.4.2 Training and Visualization

The fitness of the action will be calculated by averaging with predetermined weights all the needs (hunger, thirst, reproductive) and the emotional state. This will be calculated after the dynamics are computed so that some of the consequences of the action are known. This fitness will be converted to a percentage of change and then a labelled will be assigned (-20,-15,-10,-7,-5,-4,-2,-1,0,1,2,4,5,7,10,15,20) that is the closest in value. The labels will be classes used in training by decision trees.

One dataset will result for each action and one decision tree for each dataset. The number of entries (data points) for each dataset are: 1'210'400 (random food scavenge); 1'157'883 (random eat); 92'637 (random request); 92'637 (random move). We observe different levels of over-fitting from the decision trees as can be seen in table 5.1 We want to add variation between decisions and for this multiple possible decision trees will be generated with different depths (5,10,15,20,30,40).

Depth	Food Scavenge		Eat		Request		Move	
	Train	Test	Train	Test	Train	Test	Train	Test
5	0.392	0.402	0.344	0.355	0.328	0.318	0.415	0.421
10	0.411	0.417	0.356	0.363	0.349	0.317	0.447	0.413
15	0.440	0.411	0.383	0.358	0.434	0.294	0.558	0.383
20	0.533	0.380	0.472	0.329	0.617	0.263	0.728	0.339
30	0.853	0.313	0.822	0.259	0.942	0.212	0.961	0.294
40	0.986	0.290	0.980	0.240	0.998	0.205	0.998	0.291

Table 5.1: Training metrics of the decision trees

After we train the decision trees on the described datasets we will obtain something similar to figure 5.6.

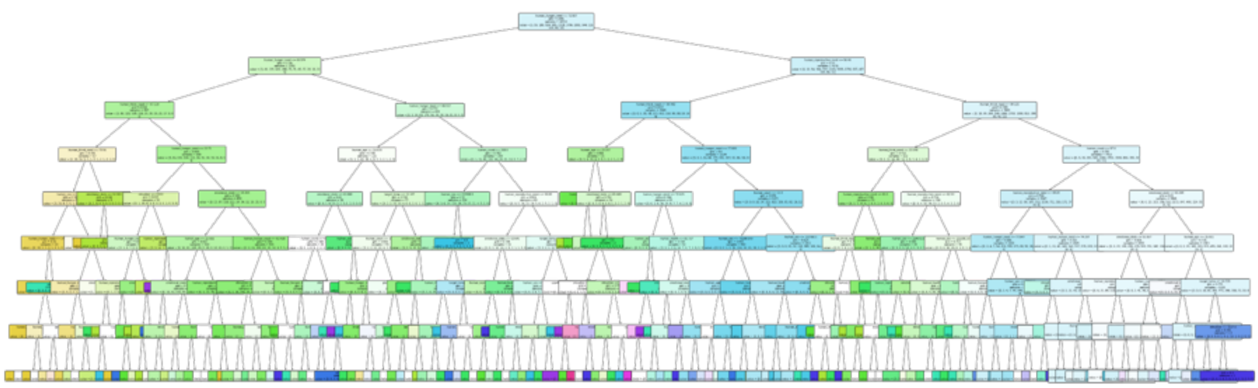


Figure 5.6: Representation of minimum depth decision tree

5.5 Environment with resources and trained intelligent human agents

5.5.1 Description

In the generation phase, each human agent will receive a random thinking 'depth' (from the available decision tree depths) and then the cognitions will be initialized using the corresponding decision trees. This will offer variation in individuals, diversity in groups and a more accurate general view compared to using identical models for all agents.

In the cognition phase on each agent's turn we will go through a **decision-making process** which is rather interesting. It will consider each action and then it will begin by **imagining** scenarios (moving to a certain square, requesting food from a certain agent, going to scavenge for a specific food, etc). For each imagined scenario, there will be an inference step using the available decision tree to **think** about the outcome. All these predictions will be compared and the agent will **decide** on executing the best one. There is an attribute in an agent's cognition called **standard** that will be lower for sad agents and higher for happy ones. This standard will offer a quick **spontaneous decision** if the predicted outcome exceeds standards. The more the agent is imagining and thinking about a good action the lower the standard gets (as desperation increases).

All agents will use this imagine-think-decide paradigm to act. If they found no decent action it is possible that they do not act at all as well.

5.5.2 Visualization and data analysis

One very important animation during the debugging process but also useful and satisfying to analyse after an experiment is the Attribute histogram (as seen in figure 5.7). This is created using the attributes of all alive agents in the population in a given time iteration. We can use them to identify trends and connections between different phenomena. For example, there are some obvious connections between the amount of food in the inventory and the hunger need; during the winter when inventories are depleted hunger need satisfaction decreases. Some more interesting connections regard the emotional state. During the winter it decreases slightly because people are hungry but also because people are having less copulation since most of the social requests regard food sharing instead of social interactions.

In figure 5.8 we can see the final collage that presents the environment, the main food resource, a time reference, and human attributes. The attributes are presented in two ways: by individual and by the tribe. The attribute "human_sex" depicts

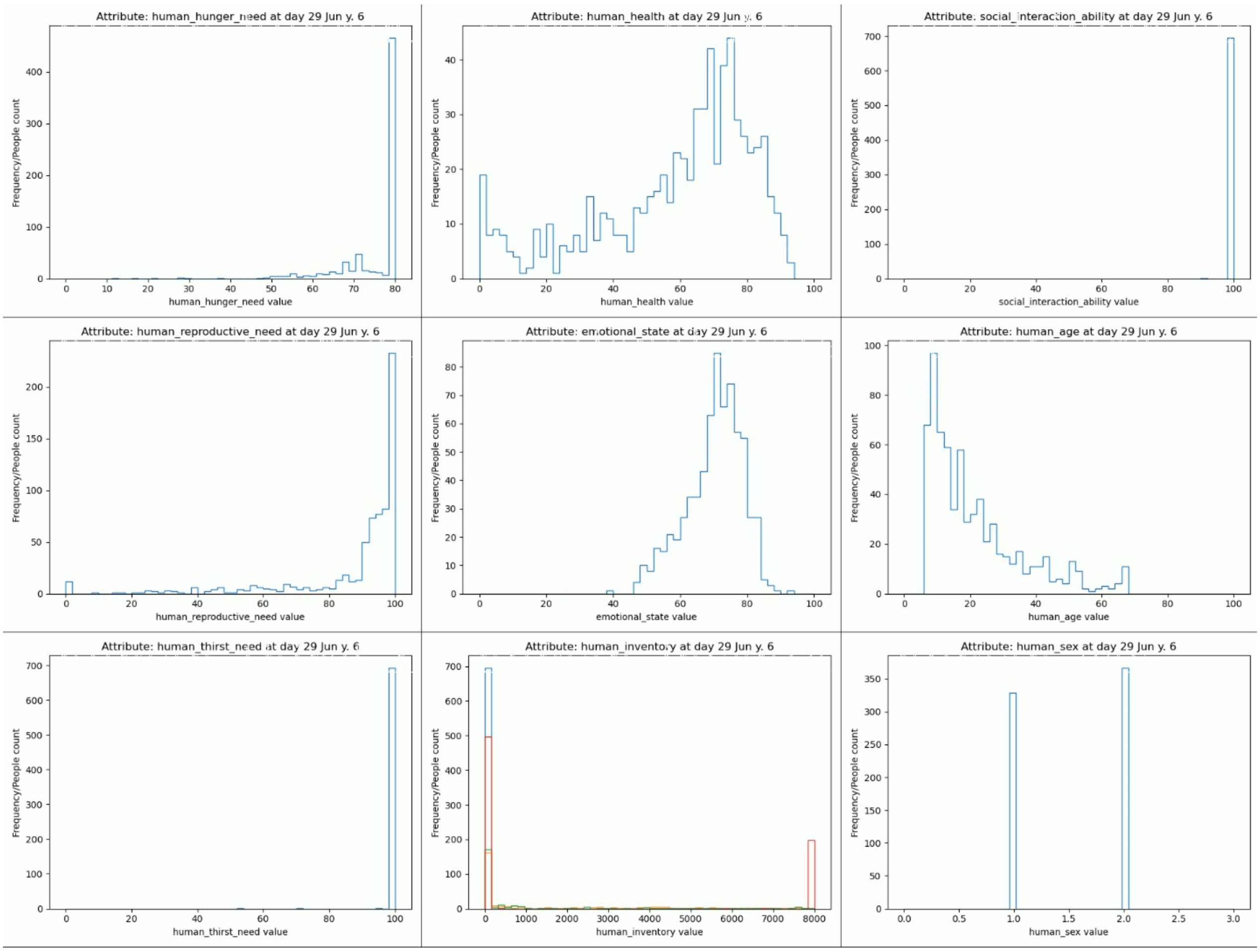


Figure 5.7: Collage of histogram for human attribute values

a dot for all agents while the "human_age" and "human_health" attributes tell the average value of the attributes calculated for each tribe.

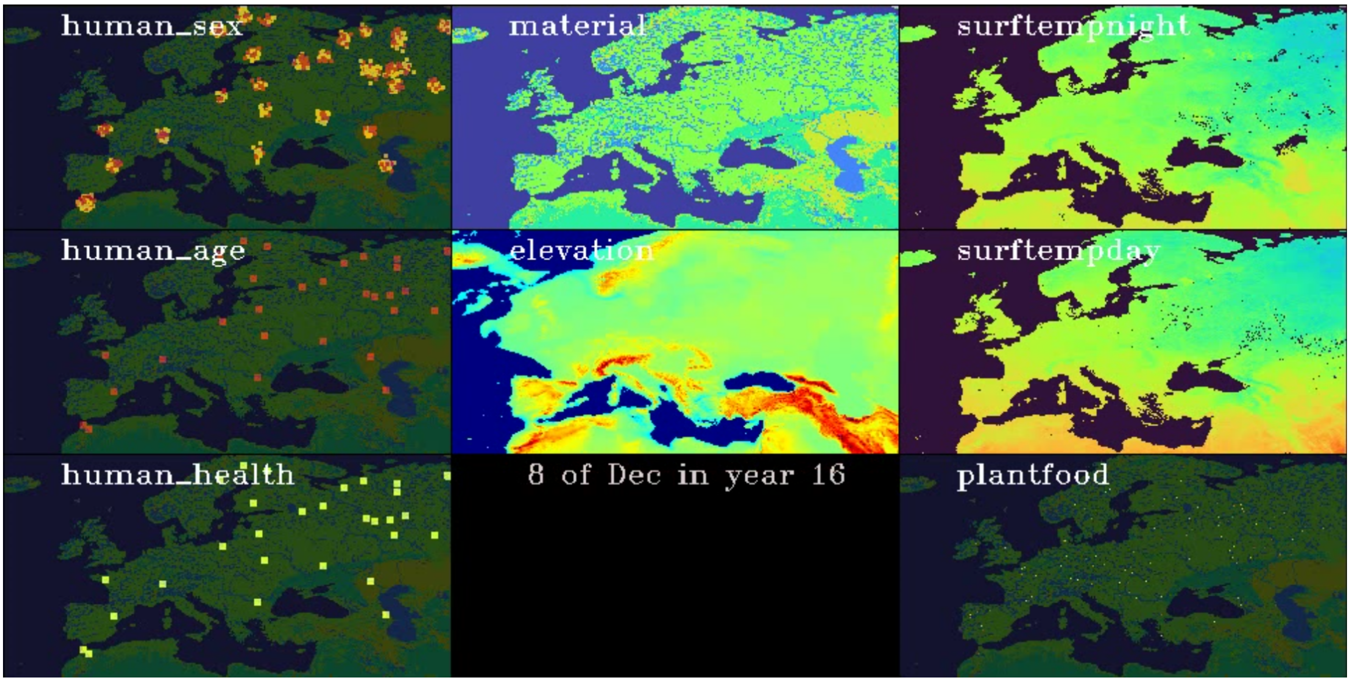


Figure 5.8: General visualization of human attributes geographically

5.5.3 Results and Performance

One epoch would take between 2.5 - 4 seconds which usually means that a simulation will take 3-5 hours. They run faster towards the end since the computationally intensive part takes place during cognition and at the end, there are very few agents left.

5.5.4 Validation and Interpretation

Validation will be done as discussed using the population evolution from a mortality rate based rudimentary simulation. We can see in 5.9 how the population is generally evolving in experiments. The simulation is quite close to the emulation. In the next table 5.2 we have the averages of some basic metrics for every two pairs of (simulation, emulation) / (simulation, mortality). The 3 metrics are: Root Mean Squared Percentage Error(RMSPE), Root Mean Squared Error(RMSE), and Mean absolute percentage error (MAPE). MAPE should be the most valid since we want the error to contribute equally regardless of the population size. The formulas used are:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(f_i - y_i)^2}{n}}$$

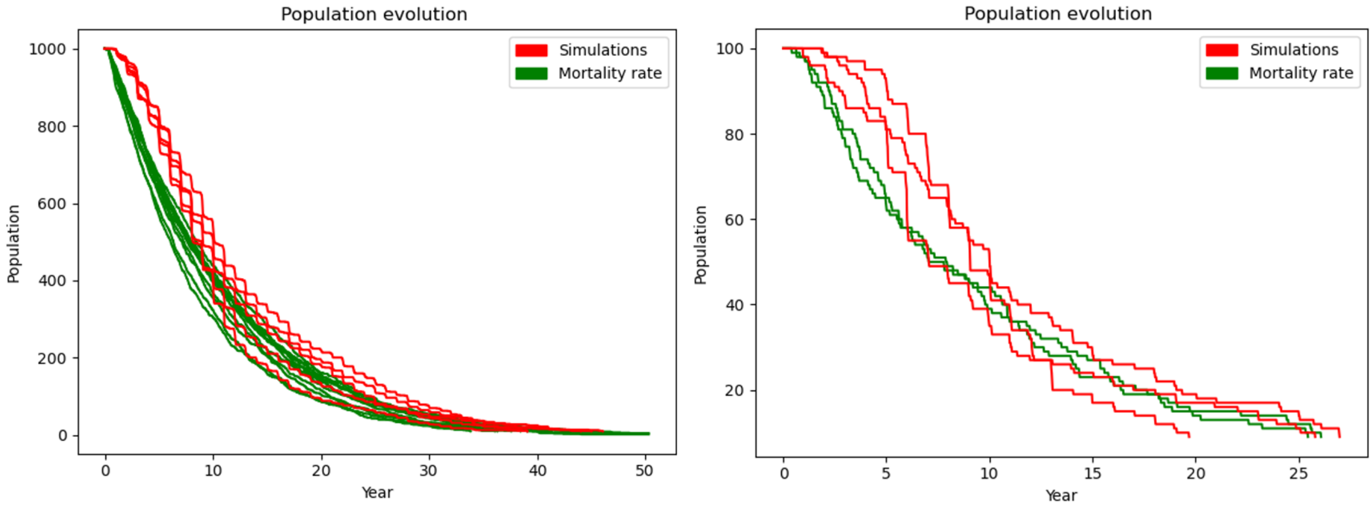


Figure 5.9: Line graph of experiment's population

$$RMSPE = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{f_i - y_i}{f_i}\right)^2}{n}}$$

$$MAPE = \frac{\sum_{t=1}^{n} \frac{|f_i - y_i|}{|f_i|}}{n}$$

Type	**MAPE**	**RMSE**	**RMSPE**
Average	0.4025	76.249	0.4794
Best	0.0767	42.790	0.0957
Worse	1.3246	160.651	1.9258
Average Spring	0.3750	68.874	0.4487
Best Spring	0.0693	37.499	0.0820
Worse Spring	1.2386	152.625	1.9108

Table 5.2: Metrics for all pairs by average best, worse

Putting all the metrics on a histogram (see figure 5.10) shows that the distribution is quite inclined towards better accuracy and also that there is an obvious difference in accuracy when calculated in Spring. This can also be seen in table 5.2.

5.6 Discussions

There are a couple of interesting issues that cause a drop in performance when measured with metrics.

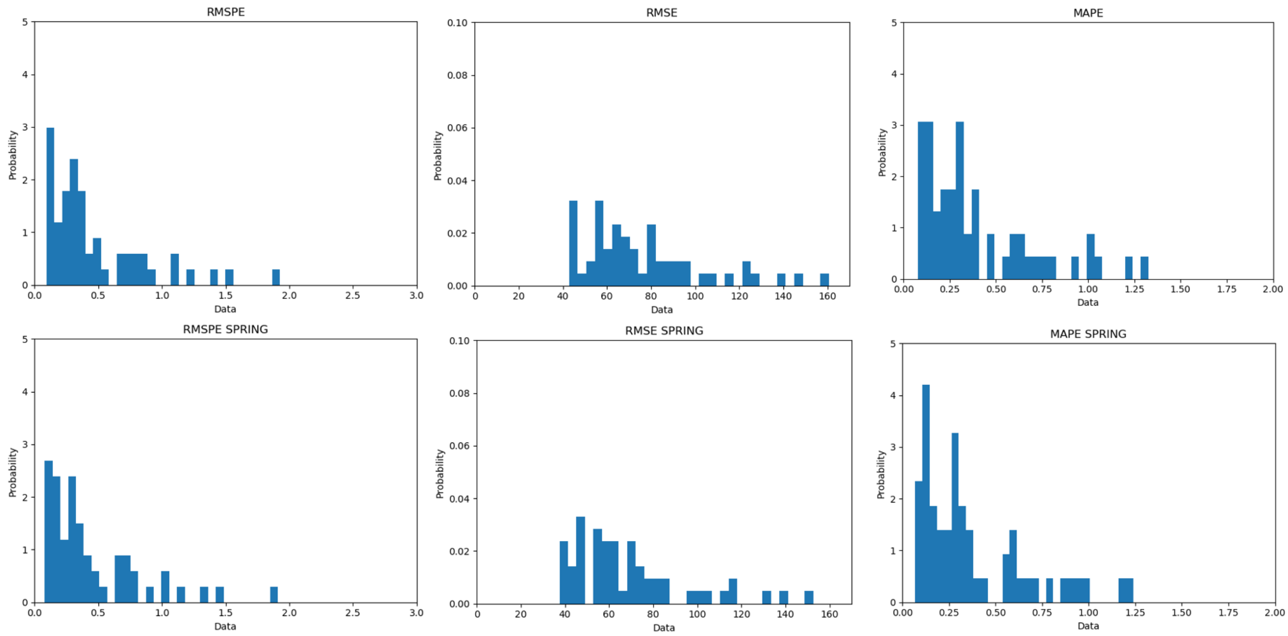


Figure 5.10: Histogram of validation metrics for all experiments

The fact that metrics are higher when calculated in the spring has the same reason as the 'stairs-like' appearance of the line graph evolution of population (as seen in 5.9). In the simulation, most agents die during the winter when food is scarce whilst in the emulation there is a constant death probability year-round.

It can be observed that at the beginning of the simulations is when the error is at its highest. This happens because in the simulation it is very hard to decide correctly what agents are going to be starting with low health and low need satisfaction without giving an unfair advantage to the rest. We initialize them all with average values which means that it will require time to adjust to the emergent trend of the simulation.

Validation proves that the simulation is very accurate in emergent metrics as the emulation which means it is likely that the simulation encapsulates correctly reality.

Chapter 6

Conclusions and future work

6.1 Strengths and challenges

Many issues created challenges for this project ranging from ideological to functional. Some of the more interesting and profound ones proved to also be the most useful. Many of the challenges encountered turned into strengths.

Data is currently an essential part of most scientific works. Regarding simulations, the complex and thorough use of data is relatively a modern habit. It's certainly been an existential step forward but the standards kept rising while data acquisition did not get easier. Any serious work in the this field uses spatially explicit components. The data for this emulation ranges from maps and resources to measurable metrics of emergent phenomenons (mortality, plant production, etc). Any new attribute addition requires: research to detect a reliable source with little bias but a big contribution to the effect, converting and processing of various data files (rarely standardized), and the proper integration into the system. However, this is very beneficial for the validity of the simulation and the process of parametrization as well.

The modelling is also connected to data gathering because the relationship between the theoretically needed attributes and the available data for attributes is bidirectional. That means that changes will constantly be made on each end to ensure evolution. Considering our selected context, more precisely our prehistoric time period, it was a real challenge to find any relevant data. A complicated wide view of the historical evolution of humans and of the ecosystem was needed (for example: there is a glaciation, intersection with other species of humans, different water levels, possible world disasters, etc).

6.2 Weaknesses and Threats

With complexity also comes **rigidity**. In the final stage of our project, it would not be an easy process to add new functionality (attribute, dynamic or agent). To scientifically validate its connections to all other dynamics and attributes should be considered (since the entire system is strongly inter-connected). Although the process is tedious, the results are worthwhile.

Another weakness is poor **performance**. This is a trade that was made intentionally to gain flexibility and more efficient modelling powers.

Modelling is an issue because it relies on other research and each new dependency adds another possible error. However, by fearing to add new research one would fail to create the work which would have improved the error. This creates a vicious cycle for science in general: taking a risk with other researchers in order to advance the field and eventually identify the errors.

There are also some more **specific issues** that we considered during our work. One would be the fact that health should not be dropping directly at all (artificially creating a fixed age barrier if you do not increase health through food); it should instead affect a fitness level that creates issues in activities. The topography should be changing if the time frame is bigger than a few decades/centuries because in our context there is a de-glaciation active.

A problem with social simulations, in general, is posed in the work [GY09]. It is argued that simulations should only be applied to specific cases and optimizations of different metrics but not to artificial societies. The implications of large-scale simulations pose a problem through their doubtful validation. Hopefully, some of the issues raised by this work were solved in the following years and as an example, we have spatially explicit components in simulations and more robust validation systems.

6.3 Ethical considerations and impact

An interesting argument comes from a more philosophical argument [SB07]. The fact that we are trying to create a simulation representing something from our world and this simulation is being created in our world poses multiple issues. This **inception** is a natural limiter. Our current understanding of the world and its mechanics tells us that with size comes complexity. To simulate properly a society we would require a space equally big, which is in turn impossible. Another issue is the bias from which we suffer. By being inside the target world to be simulated we cannot objectively perceive the reality good enough to create proper models.

The development of such simulations is a wonder but the potential usage of similar work can be frightening. In normal conditions, one would hope that the legislation would come before mass applications of such simulations in the society (by private companies or state societies). With increased complexity and computational power, the simulation could iterate the daily activities of modern humans to a certain measure. Manipulation for the wrong reasons can become a serious threat.

There is a really interesting topic that debates whether we really benefit from complexity or it doesn't add any value to the experiment. If a simple well-put decision (as in the segregation model [SB07]) can create the emergent effect that we are looking for then is really meaningful to increase complexity with many other non-relevant dynamics and attributes. Broadening the scope too much will convert the experiment into a complete simulation even if parts of it should have been emulated instead. Only the most relevant components should be simulated (in order to be proven or disproved). However, the issue here is that you will not be able to detect any errors in the existing system without specifically simulating the responsible components and there might still be some emergent phenomenons worth considering that are still out of our scope.

6.4 Future steps

This project presents endless opportunities for optimization by parallelization, better thread concurrency and memory allocation, and the addition of cloud computing. However, the more interesting developments are within the modelling and development area.

Through modelling more phenomenons could be simulated by the addition of agents, action, and dynamics. A wider range of actions and dynamics could mean the implementation of complex concepts of technologies (as craftable objects that can be discovered through experimentation), culture (identity as common myths), and intricate interactions between agents.

Many specific improvements could be integrated: dynamic topography of the geography, more water sources, more food sources (other plant species, animal populations to hunt), neanderthals and other hominin agents, and more educated generation of settlements.

Agent's cognition can easily be taken a few steps forward due to the modularity of the project. Different learning algorithms, long-term planning, a wider range of emotions, and reinforcement learning are just a few technologies that can be implemented.

Visualization could also be taken a few steps forward by integrating a 3D environment, dynamics animations and study of correlations between attributes. Validation can be augmented with metrics regarding home range areas and migration rates. Another type of measure quality could be introduced based on the concept of correlations. This can be explored by different graph algorithms that can optimize the searching in the parameter solution space based on real-world metrics.

Ultimately, the goal is to simulate a big segment of the population and environment and bring the context (time period, spatial environment) into the present and even the future.

6.5 Comparison to similar work

All simulations designed before contributed one way or another towards the current state of the art. However, there are a few specific simulation categories that affected more the field of social simulations. This reason motivates us to compare to the correct works. While specific simulations 3.1 are useful, it is not relevant to our work (especially due to different validation mechanisms). Simulations of different population (animals 3.3 or humans 3.4) are the most relevant to compare to.

There are many states of the art practices that our work and other simulations have in common. We have adopted the use of spatially explicit attributes to create reliable context. Proper validation of emergent phenomenons has been integrated. Rigorous modelling ensured minimum bias, minimum error, and maximum weight. Emulation was favoured to generation for non-target components and the exceptions benefited from generation based on scientific supported distributions.

Comparing our work with the most complex projects ([Sch19], [GLS10]) we can see some improvements that could be brought forward and integrated into this work. These are some of the specific improvements mentioned in the previous chapter 6.4.

We also brought original contribution by creating a unique dataset for the emulation of the environment and adapting state of the art methods to our specific context. No other work has been identified to simulate homo s. sapiens during the late Pleistocene in Europe. Human modelling using our architecture is again a new contribution to the field. Another essential contribution is the introduction of complexity scaling which allows different population sizes and different resolutions for the simulation. While most studies in this field end up creating their custom agent-based system since we also achieved that we can consider it a contribution as well.

6.6 Conclusion

We managed to bring many contributions to the field (as described in the previous section 6.5) by creating a custom agent-based system with meticulously modelled human agents, the possibility to scale experiments by resolution and agent-count, ability to use or integrate spatially explicit context, combine emulation with simulation, and eventually conduct an advance simulation of the human society.

Acknowledgements

The author acknowledges the support received from the professor Eugen Mihulet in the form of feedback and comments relative to the climate and geographic component's modelling.

Appreciations are also directed towards the family and the life partner of the author in the form of support, constant feedback, and encouragement.

Bibliography

[AL11] Scholarworks@umass Amherst and Ravindra R Lote. Agent Based Modeling of Electronic Markets To Analyze the Agent Based Modeling of Electronic Markets To Analyze the Sustainability of Mutual Cooperation Sustainability of Mutual Cooperation. 1911.

[AN11] Muaz Ahmed and Khan Niazi. Towards A Novel Unified Framework for Developing Formal, Network and Validated Agent-Based Simulation Models of Complex Adaptive Systems. 2011.

[BHP+11] Sergio Bernardes, Allison Howard, Andréa Presotto, Marguerite Madden, Thomas Jordan, Dorothy Fragaszy, Patrícia Izar, and Y. Tavares-Rocha. An agent based modeling approach for representing capuchin (sapajus spp.) behavior in brazil. *American Society for Photogrammetry and Remote Sensing Annual Conference 2011*, pages 41–47, 01 2011.

[Blo11] Michael Block. Maslow's Hierarchy of Needs. *Encyclopedia of Child Behavior and Development*, pages 913–915, 2011.

[Bra00] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[cor] New and Emergent World Models Through Individual, Evolutionary, and Social Learning — NEW TIES Project — Fact Sheet — FP6 — CORDIS — European Commission.

[DF94] Alexis Drogoul and Jacques Ferber. Multi-agent simulation as a tool for modeling societies: Application to social differentiation in ant colonies. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 830 LNAI, pages 3–23. Springer Verlag, 1994.

[Dre05] Michael Drennan. The Human Science of Simulation: a Robust Hermeneutics for Artificial Societies. jan 2005.

[Dun92] R. I.M. Dunbar. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6):469–493, jun 1992.

[Dun05] Jill Bigley Dunham. An agent-based spatially explicit epidemiological model in mason. *Journal of Artificial Societies and Social Simulation*, 9(1):3, 2005.

[Eps12] Joshua M. Epstein. Generative Social Science: Studies in Agent-Based Computational Modeling. dec 2012.

[Fin10] Caleb E. Finch. Evolution of the human lifespan and diseases of aging: Roles of infection, inflammation, and nutrition. *Proceedings of the National Academy of Sciences of the United States of America*, 107(SUPPL. 1):1718–1724, jan 2010.

[Gam70] Mathematical Games. The fantastic combinations of john conway's new solitaire game "life" by martin gardner. *Scientific American*, 223:120–123, 1970.

[gena] GENESIS PROPOSAL.

[genb] (PDF) GENESIS Social Simulation — Andy Turner - Academia.edu.

[GK07] Michael Gurven and Hillard Kaplan. Longevity among hunter-gatherers: A cross-cultural examination. *Population and Development Review*, 33(2):321–365, 2007.

[GLS10] Cameron S. Griffith, Byron L. Long, and Jeanne M. Sept. HOMINIDS: An agent-based spatial simulation model to evaluate behavioral patterns of early Pleistocene hominids. *Ecological Modelling*, 221(5):738–760, mar 2010.

[gov] Us government. Air pressure and wind.

[GY09] Till Grüne-Yanoff. The explanatory potential of artificial societies. *Synthese*, 169(3):539–555, aug 2009.

[HB12] Erez Hatna and Itzhak Benenson. The schelling model of ethnic residential dynamics: Beyond the integrated - segregated dichotomy of patterns. *JASSS*, 15(1), 2012.

[HGX10] Haitang Hu, Peng Gong, and Bing Xu. Spatially explicit agent-based modelling for schistosomiasis transmission: human-environment interaction simulation and control strategy assessment. *Epidemics*, 2(2):49–65, jun 2010.

[HO07] John Harrington and John E. Oliver. Understanding and Portraying the Global Atmospheric Circulation. *http://dx.doi.org/10.1080/00221340008978950*, 99(1):23–31, 2007.

[Joh] John Cassano. Athmospheric Basics and weather map analysis.

[KAM20] Hiroki Koda, Zin Arai, and Ikki Matsuda. Agent-based simulation for reconstructing social structure by observing collective movements with special reference to single-file movement. *PLoS ONE*, 15(12), dec 2020.

[KBVD11] Kees Klein Goldewijk, Arthur Beusen, Gerard Van Drecht, and Martine De Vos. The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography*, 20(1):73–86, jan 2011.

[KC16] Zdzislaw Kowalczuk and Michal Czubenko. Computational approaches to modeling artificial emotion - an overview of the proposed solutions. *Frontiers Robotics AI*, 3(APR):21, apr 2016.

[Ken12] William G. Kennedy. Modelling human behaviour in agent-based models. *Agent-Based Models of Geographical Systems*, pages 167–179, jan 2012.

[KKZ] Jed O Kaplan, Kristen M Krumhardt, and Niklaus Zimmermann. The prehistoric and preindustrial deforestation of Europe. *Quaternary Science Reviews*, 28:3016–3034.

[LG13] Bernhard Lehner and Günther Grill. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27(15):2171–2186, jul 2013.

[McD20] Kris McDaniel. John M. E. McTaggart. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2020 edition, 2020.

[MGP] Stacy Marsella, Jonathan Gratch, and Paolo Petta. Computational Models of Emotion.

[Mil98] Arthur I. Miller. *Albert Einstein's special theory of relativity : emergence (1905) and early interpretation (1905-1911)*. Springer, 1998.

[MJ] Colin. McEvedy and Richard Jones. Atlas of world population history. page 368.

[MMKI02] Y. Murakami, K. Minami, T. Kawasoe, and T. Ishida. Multi-agent simulation for crisis management. In *Proceedings - IEEE Workshop on Knowledge Media Networking, KMN 2002*, pages 135–139. Institute of Electrical and Electronics Engineers Inc., 2002.

[NM07] Michael J. North and Charles M. Macal. Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation. *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*, 9780195172119:1–368, sep 2007.

[Pet] Peter W. Sloss. ETOPO5 Global Surface Relief.

[PL05] Liviu Panait and Sean Luke. Cooperative Multi-Agent Learning: The State of the Art. *Autonomous Agents and Multi-Agent Systems 2005 11:3*, 11(3):387–434, nov 2005.

[PLPG20] Constantin-Valentin Pal, Florin Leon, Marcin Paprzycki, and Maria Ganzha. A Review of Platforms for the Development of Agent Systems. 2020.

[PPK02] Praveen Paruchuri, Alok Reddy Pullalarevu, and Kamalakar Karlapalem. Multi agent simulation of unorganized traffic. page 176, 2002.

[Res96] Mitchel Resnick. StarLogo. pages 11–12, 1996.

[rio] rioxarray README — rioxarray 0.11.0 documentation.

[RSZ07] Arash Rahman, Saeed Setayeshi, and Mojtaba Shamsaei Zafarghandi. AN ANALYSIS TO WEALTH DISTRIBUTION BASED ON SUGARSCAPE MODEL IN AN ARTIFICIAL SOCIETY. *IJE Transactions B: Applications*, 20(3), 2007.

[SB07] Eric Silverman and John Bryden. From Artificial Societies to New Social Science Theory. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4648 LNAI:565–574, 2007.

[Sch19] Fulco Scherjon. Title: Virtual Neanderthals : a study in agent-based modelling Late Pleistocene hominins in western Europe. 2019.

[SKD+08] W C Skamarock, J B Klemp, J Dudhi, D O Gill, D M Barker, M G Duda, X.-Y. Huang, W Wang, and J G Powers. A Description of the Advanced Research WRF Version 3. *Technical Report*, (June):113, 2008.

[SMF18] Damien Sulla-Menashe and Mark A Friedl. User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product. 2018.

[SSP08a] Sharad Sharma, Harpreet Singh, and Atul Prakash. Multi-agent modeling and simulation of human behavior in aircraft evacuations. *IEEE Transactions on Aerospace and Electronic Systems*, 44(4):1477–1488, 2008.

[SSP08b] Sharad Sharma, Harpreet Singh, and Atul Prakash. Multi-agent modeling and simulation of human behavior in aircraft evacuations. *IEEE Transactions on Aerospace and Electronic Systems*, 44(4):1477–1488, 2008.

[SSP+20] Vladimir P. Skulachev, Gregory A. Shilovsky, Tatyana S. Putyatina, Nikita A. Popov, Alexander V. Markov, Maxim V. Skulachev, and Victor A. Sadovnichii. Perspectives of Homo sapiens lifespan extension: focus on external or internal resources? *Aging (Albany NY)*, 12(6):5566, mar 2020.

[TAS+20] Mohammad Tabasi, Ali Asghar Alesheikh, Aioub Sofizadeh, Bahram Saeidian, Biswajeet Pradhan, and Abdullah AlAmri. A spatio-temporal agent-based approach for modeling the spread of zoonotic cutaneous leishmaniasis in northeast Iran. *Parasites and Vectors*, 13(1):1–17, dec 2020.

[THL15] Lu Tan, Mingyuan Hu, and Hui Lin. Agent-based simulation of building evacuation: Combining human behavior with predictable spatial accessibility in a fire emergency. *Information Sciences*, 295:53–66, feb 2015.

[TW04] Seth Tisue and Uri Wilensky. NetLogo: Design and implementation of a multi-agent modeling environment 1. *Conference on Social Dynamics: Interaction*, 2004.

[WHH15] Zhengming Wan, Simon Hook, and Glynn Hulley. Mod11a2 modis/terra land surface temperature/emissivity 8-day l3 global 1km sin grid v006. *Nasa Eosdis Land Processes Daac*, 10, 2015.

[WJE+1] Shawn R. Wolfe, Peter A. Jarvis, Francis Y. Enomoto, Maarten Sierhuis, and Bart-Jan van Putten. A Multi-Agent Simulation of Collaborative Air Traffic Flow Management. *https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-60566-226-8.ch018*, pages 357–381, jan 1.

[WNF+15] A. Watkins, J. Noble, R. J. Foster, B. J. Harmsen, and C. P. Doncaster. A spatially explicit agent-based model of the interactions between jaguar populations and their habitats. *Ecological Modelling*, 306:268–277, jun 2015.

[wor] Mortality in the past – around half died as children - Our World in Data.

[ZZVG20] Jana Zausinová, Martin Zoričak, Marcel Vološin, and Vladimír Gazda. Aspects of complexity in citizen–bureaucrat corruption: an agent-based simulation model. *Journal of Economic Interaction and Coordination*, 15(2):527–552, apr 2020.