Agent-based Modeling for Hunter-Gatherer Communities - Day1: Implementation and Visualization of ABM Mobility

Johannes Zonker

Hackathon on Small Data Analysis

January 2024

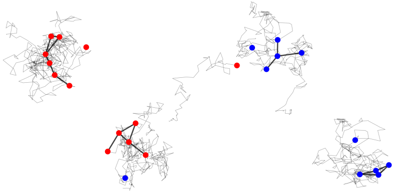
Overview

1. Introduction to agent-based modeling for hunter-gatherers
   - General ABM Definition
   - Cultural Evolution of Hunter-Gatherer Societies
2. Simulation Methods for Day 1
   - Euler-Maruyama Scheme
   - Finite Difference Method
3. Tutorials for ABM Mobility
   - Movements in a Double Well Landscape
   - Mobility for Hunter-Gatherer ABM

General ABM Definition

ABM Definition[1]:

- We consider a system of n_a agents, that have a **continuous** state variable (position) and a **discrete** state variable (status).
- We denote the **position** of all agents in a vector X with entries from $\mathbb{X} \subset \mathbb{R}^d$.
- We denote the **status** of all agents in a vector S with entries from $\mathbb{S} \subseteq \mathbb{N}$. Usually $\mathbb{S}$ is a finite set with n_s elements.
- The **system state** space is denoted by $\mathbb{Y} := \mathbb{X}^{n_a} \times \mathbb{S}^{n_a}$.



Change of variables:

- The movements are modeled by **diffusive dynamics** described by a generator L.
- The adoption is modeled by a **jump process** with rates depending on the current system state.

Status change

We define for each agent α and each possible change from status i to status j an **adoption rate function**

$$f_{ij}^{(\alpha)} : \mathbb{Y} \to [0, \infty)$$

and set $f_{ij}^{(\alpha)} = 0$ for $i = j$.
We will especially consider adoption rate functions defined by

$$f_{ij}^{(\alpha)}(X, S) = \delta_i(s_\alpha)\gamma_{ij}(x_\alpha),$$

for **first-order adoptions**, where $\gamma_{ij} : \mathbb{X} \to [0, \infty)$ and

$$f_{ij}^{(\alpha)}(X, S) = \delta_i(s_\alpha) \sum_{\substack{\beta=1 \\ \beta\neq\alpha}}^{n_a} \delta_j(s_\beta)\gamma_{ij} d_r(x_\alpha, x_\beta),$$

for **second-order adoptions**, where $\gamma_{ij} > 0$ and $d_r(x_\alpha, x_\beta) := \chi_{[0,r]}(\|x_\alpha - x_\beta\|)$.

Change of probability mass

For the adoption dynamics we define the operator G by

$$Gp(X,S,t) := -\sum_{i,j=1}^{n_s}\sum_{\alpha=1}^{n_a} f_{ij}^{(\alpha)}(X,S)p(X,S,t)$$
$$+\sum_{i,j=1}^{n_s}\sum_{\alpha=1}^{n_a} f_{ij}^{(\alpha)}(X,S+ie_\alpha-je_\alpha)p(X,S+ie_\alpha-je_\alpha,t),$$

where e_α denotes the αth unit vector of $\mathbb{R}^{n_a}$. The change of the probability mass function $p(X,S,t)$ for the system state is given by the set of differential equations

$$\partial_t p(X,S,t) = Lp(X,S,t) + Gp(X,S,t).$$

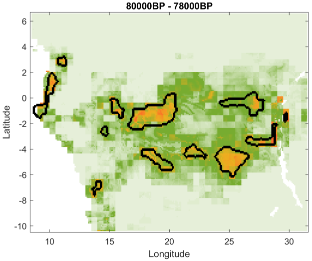
ABM for Hunter-Gatherer Societies in Central Africa

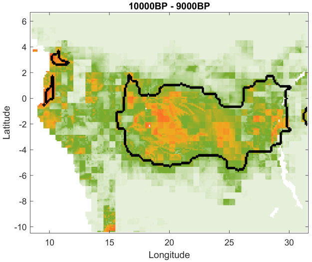
Model assumptions[2]:

- Each agent α represents a **hunter-gatherer camp** and has a **position** $x_\alpha \in \mathbb{X} \subset \mathbb{R}^2$, **population** $d_\alpha \in \mathbb{R}$ and a discrete status $s^{(\alpha)} \in \mathbb{S}$ representing multiple **cultural features**
- We model the **mobility** with a diffusion process, the **demographic** changes as deterministic and the **cultural evolution** with a jump process.

Residential Mobility:

- The mobility process of the agents is composed of three major influences:
  1. A *time dependent* suitability landscape $V : \mathbb{X} \times \mathbb{T} \to \mathbb{R}$ for environmental influences incorporating real-world data
  2. An interaction force $U_\alpha : \mathbb{X}^{n_a} \to \mathbb{R}$
  3. A Brownian motion $B(t)$ with a scaling function $\sigma : \mathbb{X} \to \mathbb{R}$ depending on friction



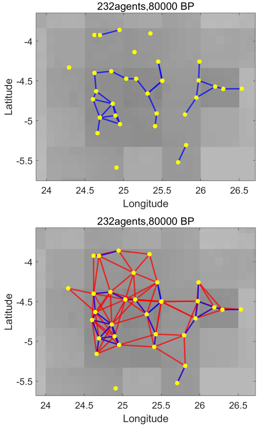


Cultural Evolution Dynamics

- We distinguish between
  1. Progressive cultural features with incremental gain and loss of information
  2. Non-progressive cultural features similar to opinion dynamics
- We consider an interaction network defined by

$$A_{\alpha\beta}(Y(t)) := \begin{cases} \varphi_1, & \text{if } \|x_\alpha - x_\beta\| \leq r_1 \\ \varphi_2, & \text{if } r_1 < \|x_\alpha - x_\beta\| \leq r_2 \\ 0, & \text{else} \end{cases}$$

with $\varphi_1, \varphi_2 > 0$ being short and long range interaction rates corresponding to interaction radii $r_1, r_2 > 0$.



Non-Progressive Cultural Features

Status variable:

- We consider for a non-progressive feature i a finite set $\mathbb{S}_i$ of n_s traits.
- We assume the trait values to be unordered and that the switch from one arbitrary trait $k \in \mathbb{S}_i$ to any other trait $l \in \mathbb{S}_i$ is possible.

First order events:

- Spontaneous trait value change for agent α:
  - State change vector: $v_{ikl}^{(\alpha)} := (-k + l)e_i^{(\alpha)}$ encoding the switch from trait k to l in feature i for agent α.
  - Adoption rate function: $g_{ikl}^{(\alpha)}(Y) := \gamma_i \delta_k(s_i)$ with $\gamma_i > 0$ and $k \neq l$.

Second order events:

- Transmission of a trait value to agent α:
  - State change vector: $v_{ikl}^{(\alpha)}$
  - Adoption rate function:

$$f_{ikl}^{(\alpha)}(Y) := \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^{n} \delta_k(s_i^{(\alpha)}) \delta_l(s_i^{(\beta)}) \cdot A_{\alpha\beta}(Y), \tag{1}$$

Progressive Cultural Features

Status variable:

- We encode the status of all agents in a vector S with the entry $s_i^{(\alpha)} \in \mathbb{N}$ being the trait value of agent α for feature i.

First order events:

- **Gain of knowledge** for agent α:
  - State change vector: $e_i^{(\alpha)}$, i.e. 1 at entry (i, α) and zero else.
  - Adoption rate function: $g_i^{(\alpha)}(Y(t)) := \gamma_i$ with $\gamma_i > 0$.
- **Loss of knowledge** for agent α:
  - State change vector: $-e_i^{(\alpha)}$, loss rate constant $\lambda_i > 0$
  - Adoption rate function:

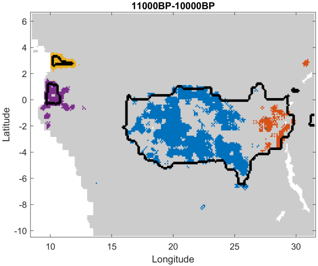
$$l_i^{(\alpha)}(Y(t)) := \begin{cases} \lambda_i, & \text{if } s_i^{(\alpha)}(t) > 0 \\ 0, & \text{else} \end{cases}$$

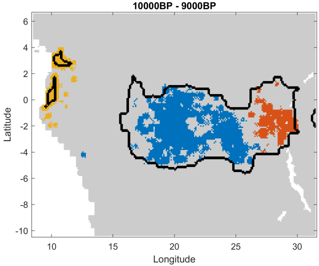
Second order events:

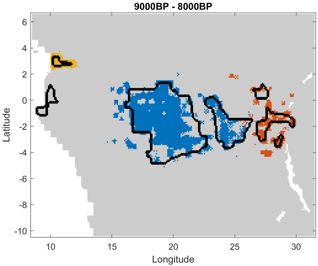
- **Knowledge transmission** to agent α:
  - State change vector: $e_i^{(\alpha)}$
  - Adoption rate function:

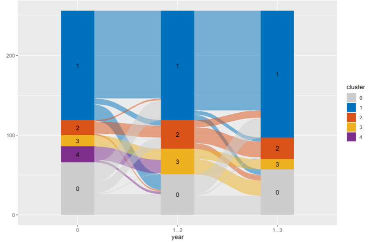
$$f_i^{(\alpha)}(Y(t)) := \sum_{\substack{\beta=1 \\ \beta\neq\alpha}}^{n} \max\left\{0, s_i^{(\beta)}(t) - s_i^{(\alpha)}(t)\right\} \cdot A_{\alpha\beta}(Y(t)) \qquad (2)$$

Spatio-Temporal Clustering









Euler-Maruyama Scheme

- We want to create a realization of a diffusion process $(\boldsymbol{X}(t))_{t\in\mathbb{T}}$ defined by a stochastic differential equation of the type

$$d\boldsymbol{X}(t) = b(\boldsymbol{X}(t), t)dt + \sigma(\boldsymbol{X}(t), t)d\boldsymbol{B}(t). \tag{3}$$

- Choosing a time step Δt we only evaluate the process at chosen discrete points assuming that the derivative of the deterministic part and the scaling of the Brownian motion does not change between time steps.
- For the Brownian motion it holds that

$$\boldsymbol{B}(t+\Delta t) - \boldsymbol{B}(t) \sim \mathcal{N}(0, \Delta t),$$

  so we can approximate an increment of the Brownian motion by $\sqrt{\Delta t}\xi$, whith $\xi \sim \mathcal{N}(0, 1)$.
- The Euler-Maruyama scheme can be defined as

$$\hat{\boldsymbol{X}}(t+\Delta t) = \hat{\boldsymbol{X}}(t) + b(\hat{\boldsymbol{X}}(t), t)\Delta t + \sigma(\hat{\boldsymbol{X}}(t), t)\sqrt{\Delta t}\xi_t$$

  with step size Δt and $\xi_t \sim \mathcal{N}(0, 1)$ independent random variables.

Finite Difference Method

- For a function f on $\mathbb{R}$ we can approximate $f'(x)$ by $\frac{f(x+h)-f(x-h)}{2h}$ for suitable $h > 0$
- For $h \to 0$ the approximation converges against the derivative and we can estimate the error by the Taylor expansion of f
- Similarly, for $f : \mathbb{R}^2 \to \mathbb{R}$ we can approximate $\frac{\partial f(x,y)}{\partial x}$ by

$$\frac{f(x+h, y) - f(x-h, y)}{2h}$$

and $\frac{\partial f(x,y)}{\partial y}$ by

$$\frac{f(x, y+h) - f(x, y-h)}{2h}$$

General Recommendations

- Write a new script for each task
- Reuse parts from old scripts that do not need to be changed for a new task
- Clearing the workspace and initializing variables at the beginning of a script can help avoiding indexing errors
- When using Matlab, vectorized operations can be faster than loops
- Use optimized pre-built methods when available (e.g. find, rangesearch in Matlab)
- For scripts that run for a longer time print the progress regularly
- Comment your code
- Communicate with each other and do not hesitate to ask questions when stuck
- Have fun! :)

Movements in a double well landscape

We consider a system of $n_a = 100$ agents that have a 2-dimensional position. The agents move in a landscape given by a function describing a double well potential.

State space for positions:

- $\mathbb{X} = \mathbb{R}^2$

Mobility dynamics:

- Diffusion in $\mathbb{X}$ given by the SDE

$$d\boldsymbol{x}(t) = -\nabla V(\boldsymbol{x}(t))dt + \sigma d\boldsymbol{B}(t),$$

for $\sigma > 0$ and $V(x, y) = (x-1)^2 + 3.5y^2$

Euler-Maruyama scheme for diffusion in a potential landscape:

$$\hat{\boldsymbol{x}}(t + \Delta t) = \hat{\boldsymbol{x}}(t) + -\nabla V(\boldsymbol{x}(t))\Delta t + \sigma\sqrt{\Delta t}\xi_t$$

Double well implementation

Task 1:

- Implement the agent system for the double well potential with $n_a = 100$ agents with Matlab/Python using the Euler-Maruyama scheme. Initialize the agents with (uniform) random positions within the area $[-2, 2] \times [-1.5, 1.5]$.
- A step size of $\Delta t = 0.01$ is recommended for simulating the system.
- Simulate the system for a fixed amount of $n_t = 10^4$ time steps and save a system snapshot for every unit of time (100 time steps). Save the snapshots and the corresponding simulation times in a file at the end of the simulation.
- For the initial simulation run choose the noise parameter to be $\sigma = 0.7$.

Visualization

Task 2:

- Write a script for generating a video (e.g. .avi format) from the snapshot data.
- Run the system for different parameters σ. What do you observe for very large and very small values of σ?

Step-by-step guide:

- Find a suitable representation for the landscape (e.g. contour lines) in the background of the plots.
- Use a scatter plot with filled circles to visualize the agent positions.

Double well given by a data grid

Motivation:

- Usually the landscape is based on real world data and not an analytical function. Environmental data often is given by values (e.g. for elevation) on a grid of lat/long coordinates.

Task 3:

- Adapt the script for simulating the double well ABM to use a grid of suitability values on the area $\mathbb{X} := [-2, 2] \times [-1.5, 1.5]$ instead of an analytical function to generate the landscape. Initialize the agents to start at random grid points. For convenience you can work with positions on the scale of the grid cell indices instead of using the original scale.
- Consider an expanded SDE

$$dx(t) = -c_{\text{pot}} \nabla V(x(t)) c_{\text{time}}^2 dt + \sigma c_{\text{time}} d\boldsymbol{B}(t)$$

  with parameters $c_{\text{pot}} > 0$ for scaling the potential and $c_{\text{time}} > 0$ for scaling the time units. Consider reflective boundary conditions (i.e. reject movements that would lead an agent to leave the area $\mathbb{X}$).
- You can follow the step-by-step guide on the next slide.

Step-by-step guide for Task 3

- Construct a grid of the area $[-2, 2] \times [-1.5, 1.5]$ with resolution 0.01×0.01 and evaluate the function V at each grid point
- Define a variable for defining the boundary of the grid area with the boundary of the grid being assigned 0 and the inside 1
- Calculate an approximation of the partial derivatives on the grid points within the boundary using the finite difference method (hint: writing a separate function for generating a matrix with finite difference derivatives is recommended)
- Find suitable parameter values for $c_{\text{pot}}, c_{\text{time}}$ that generate a similar output to the first setting. What do you observe if you keep one scaling parameter fixed and only variate the other?

Suitability data for the Hunter gatherer ABM

Information about the data[3]:

- The area that is covered is a region in Central Africa (coordinates: 8.45 to 31.55 degrees longitude and -10.45 to 6.5 degrees latitude)
- Mobility data from contemporary hunter-gatherers has been used to estimate an **environmental niche model** (ENM) that estimates the likelihood of hunter-gatherers to be present at a location based on environmental and climatic variables (e.g. elevation, temperature, humidity)
- Through paleoclimatic reconstructions and using the ENM an estimation for the presence of hunter-gatherers has been estimated up to 120000 years BP

Data format:

- Values at grid cells range between 0 and 1 and can be interpreted as a probability of hunter-gatherers to be present at the grid cell
- Areas without estimations (e.g. ocean) have NaN values

Construct Suitability Landscape from Data

Task 4:

- Import the four data landscape data sets (e.g., "landscape1.csv") and construct variables for the suitability landscape, a boundary indicator and the partial derivatives of the landscape. Save the new variables in a new file.
- Scale the landscape such that values remain between 0 and 1 and assign to NaN entries the value 2. This way the partial derivatives at the area boundary push the agents back from the boundary and we need to apply the boundary conditions less often.
- Hint: A high probability indicates a suitable area, however, dynamically the attractive areas are the valleys of the potential landscape.

Movements in a Suitability landscape

Task 5:

- Use the variables derived from "landscape1.csv" to build a script for mobility of agents in a suitability landscape.
- Consider the same SDE and boundary conditions as in Task 3, but with the data based suitability landscape replacing the double well.
- Set the number of agents to $n_a = 300$, $\Delta t = \frac{1}{12}$ (representing one month of time) and initialize the agents at random grid positions within the boundary of the area.
- Run the system for 10000 simulation years and save a system snapshot every 10 simulation years
- Visualize the ABM and experiment with the scaling parameters.

Movements in a time-changing suitability landscape

Task 6 (final task for day 1):

- Adapt the mobility model for hunter-gatherers to switch to the next landscape after 2500 simulation years and save the ID of the current landscape for each snapshot.
- Adapt the visualization to always show the current landscape in the background.

References I

- Stefanie Winkelmann, Johannes Zonker, Christof Schütte, and Nataša Djurdjevac Conrad.
  Mathematical modeling of spatio-temporal population dynamics and application to epidemic spreading.
  Mathematical Biosciences, 336:108619, 2021.

- Johannes Zonker, Cecilia Padilla-Iglesias, and Nataša Djurdjevac Conrad.
  Insights into drivers of mobility and cultural dynamics of african hunter–gatherers over the past 120 000 years.
  Royal Society Open Science, 10(11):230495, 2023.

- Cecilia Padilla-Iglesias, Lane M Atmore, Jesús Olivero, Karen Lupo, Andrea Manica, Epifanía Arango Isaza, Lucio Vinicius, and Andrea Bamberg Migliano.
  Population interconnectivity over the past 120,000 years explains distribution and diversity of central african hunter-gatherers.
  Proceedings of the National Academy of Sciences, 119(21):e2113936119, 2022.